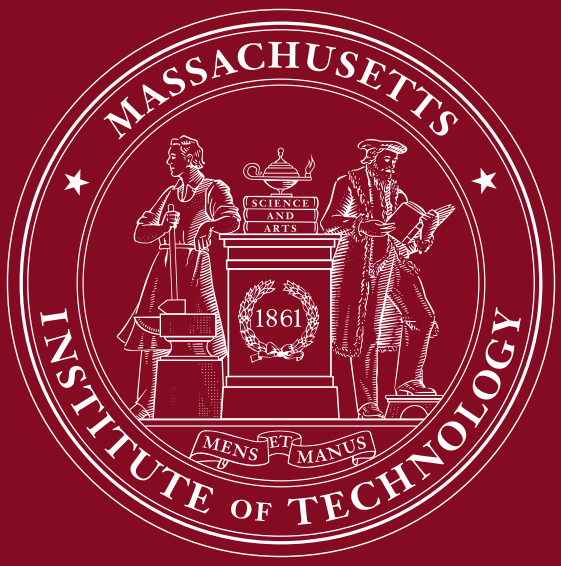


The MIT Undergraduate Journal of Economics

Volume XXV

2025-2026

The MIT Undergraduate Journal of Economics Volume XXV

2025-2026

Mailing Address:

The MIT Undergraduate Journal of Economics
Massachusetts Institute of Technology, Building E52-301
Cambridge, MA 02139

Foreword

"Money… must always be scarce with those who have neither wherewithal to buy it, nor credit to borrow it."

- *Adam Smith*

As MIT undergraduate economics students progress through their coursework, they are continuously introduced to new economic topics, constantly learning the ideas and models of established economists, and relentlessly being challenged to think differently about the observable phenomena around them. It is this enthusiasm for learning that led undergraduates at MIT to proceed in their own research—to experience the excitement of asking a question and striving to answer it. We hope that this year's papers highlight the vigor with which our undergraduate students pursue economic research and the rigor with which they present their ideas.

The publication of this Journal is made possible by the support of many people. We especially thank Professor Dave Donaldson for selecting the articles for this year's publications.

These relevant student papers demonstrate the enduring importance of rigorous economic research in the days ahead.

The MIT Undergraduate Journal of Economics Volume XXV

2025-2026

Contents

The Effect of Higher Minor League Salaries on Player Health and Performance

Zev Moore

Department of Economics

Massachusetts Institute of Technology

Instructor: Prof. Tobias Salz

December 2025

1 Introduction

Professional athletes' performance and health are influenced not only by their talent and training but also by the economic conditions under which they work. This is very visible in Minor League Baseball (MiLB), where until recently players earned wages far below a living standard. Many relied on off-season jobs, shared crowded housing, and relied on lacking nutrition. These conditions are very contrasted with Major League Baseball (MLB), where pay and resources are far higher. This environment provides a natural setting to investigate how improvements in pay affect players' health and durability.

Over the past few years, there were several pay increases in MiLB. First there was an organization-specific raise by the Toronto Blue Jays in 2019, a league-wide minimum salary increase in 2021, and a collective bargaining agreement in 2023 that significantly increased pay. These reforms generated variation across time and across organizations, creating a helpful opportunity to study the real effects of higher wages on player outcomes.

My paper addresses a central question: *Do higher minor league salaries improve player health and performance?* My hypothesis, which is based in anecdotal evidence, is that higher pay reduces financial stress and allows better nutrition, training, and recovery, which should lower injury risk and improve on-field durability. Beyond professional baseball, my question related to broader issues in labor economics regarding how compensation policies affect worker health and productivity.

To estimate the causal effect of wage increases, I use a difference-in-differences (DiD) design for the 2019 Blue Jays raise and for the 2021 and 2023 leaguewide changes. I use MLB players as a control group to capture leaguewide trends, as they didn't experience similar increases of their salaries. In this version of the paper I focus on the number of injuries and the time spent on the injured list for players from 2015 to 2025.

My findings hopefully could inform ongoing policy discussions about working conditions in MiLB and whether higher wages produce measurable welfare benefits for players and organizations.

2 Background

Minor League Baseball serves as the player development system for MLB, using tiers (Rookie, A, High-A, AA, AAA). MLB organizations are responsible for affiliate rosters, player assignments, and medical care, while Injured List (IL) placements occur at the affiliate level. An IL stint begins on the placement date and ends on activation or reinstatement.

Historically and for most of the sample period, minor league baseball salaries were extremely low relative to both major league pay and local living costs. In 2019, the Toronto Blue Jays became the first organization to unilaterally raise minor league salaries for all players by 50 percent, with team officials explicitly framing the change as an investment in better nutrition, sleep, and recovery for players.(Brudnicki, 2019) Beginning in 2021, Major League Baseball mandated substantial minimum salary increases for minor leaguers, with weekly pay rising by roughly 38-72

percent relative to 2019 levels, depending on the level of play.(Feinsand, 2021) The first Minor League collective bargaining agreement, implemented for the 2023 season, then more than doubled seasonal salaries at several levels relative to pre-2021 baselines.(Waldstein, 2023) Starting in the 2022 season, MLB also required clubs to provide in-season housing for most minor league players, substantially reducing their costs.(MLB News Staff, 2021) Taken together, these reforms represent large, positive shocks to total compensation rather than incremental adjustments, and they motivate the difference-in-differences designs used in the rest of the paper.

These changes created three shocks central to this study: (i) the 2019 organization-specific raise by the Blue Jays, (ii) the 2021 league-wide pay increase combined with structural contraction, and (iii) the 2023 collective bargaining agreement that established higher compensation and improved working conditions. The first provides treatment variation between organization, while the last two create league-wide shocks. Since the number of players and teams declined after 2020, my analysis relies on both raw injury spells and normalized player-year outcomes. The 2020 season is excluded from the main analysis and used only in robustness checks.

3 Related Literature

This study used work from two main strands of research: economic research on organizational support and athlete development, and sports medicine research on post-pandemic injury dynamics in professional baseball. With that said, there has not been a large amount of work done on these specific pay raises.

3.1 Organizational Support, Compensation, and Player Development

McLeod et al. (2022) is the main external source used in my work. The paper analyzes how organizational support relates to minor leaguers' developmental progress. Using a player survey and regression models, they show that unmet salary and basic support are widespread. This paper motivates my mechanism in that higher compensation and better living conditions relax constraints

on nutrition, sleep, recovery, and also compliance with medical guidance. Additionally, it clarifies what my design adds. Their outcomes are largely perceptual and cross-sectional, while my outcomes are clear and objective health measures observed in a panel around policy shocks.

3.2 Pandemic Disruptions and Injury Incidence in MLB

A second literature I looked at shows that MLB injury incidence was particularly high when play resumed after the pandemic. This work is important for my identification strategy because MLB functions as my control group. Platt et al. (2021) show that the 2021 MLB season had elevated injury rates when compared with seasons before COVID, which is consistent with changed schedules, and injured list usage based on changed protocols. These findings justify choices I tried to design in my paper, primarily by included year fixed effects to absorb leaguewide shocks that affect both MLB and MiLB.

4 Data and Measurement

The dataset I use in this study combines publicly available injury reports, player rosters, and organizational information for MLB and MiLB from 2015 through 2025. The unit of observation in the primary specification is the organization-year level, with robustness checks performed at the player-year level.

4.1 Data Sources

Injury and roster data were collected from The Baseball Cube database, which was built from official MLB.com and MiLB.com injury entries. The database records the player's name, organization, level (Rookie, A, High-A, AA, AAA, MLB), injury start and end dates, and the stated injury classification for each injury stint. League wide payroll and policy information were confirmed with MLB's official press releases and I used Spotrac to investigate roster and contract data.

To account for changing organizational structures, I excluded the 2020 season, when the MiLB schedule was cancelled. The 2021 realignment and contraction of affiliates were incorporated by treating the new affiliate structure as the baseline from which post-2021 outcomes are measured.

4.2 Variable Construction

Each IL placement is treated as a distinct "injury spell." An injury spell begins on the date a player is placed on the injured list and ends on the recorded activation or reinstatement date. Observations with placeholder end dates or durations exceeding 400 days are excluded to avoid distortions.

For each organization, indexed by i, and year t, I aggregated the number of injury spells and the total number of IL days and then normalize by two different denominators, depending on the outcome of interest:

- **IL spells per player**: the total number of distinct IL stints in organization i and year t, divided by the number of players appearing on that organization's rosters during that season (a roster-wide rate).
- **IL days per injured player**: the total number of IL days in organization i and year t, divided by the number of players in that organization-year who have at least one IL placement, which is a useful measure of how much time injured players spend on the IL over the season.

Thus, IL spells per player summarizes the frequency of IL usage at the roster level, while IL days per injured player captures the level of time lost among the subset of players who experience at least one injury spell. Because these outcomes are normalized by different denominators, they do not satisfy a simple mechanical identity linking them one-for-one. I didn't analyze average spell length in the paper, because a player could have many very short spells or few longer spells, and focusing on spell-level averages could be misleading for understanding the season-long injury burden.

4.3 Data Cleaning and Aggregation

To remove significant bias from extreme or ongoing IL durations, I removed all observations with more than 400 IL days or placeholder end dates as mentioned before. I then aggregated IL data to the organization-year level, merged it with roster counts, and computed IL spells per player and IL days per injured player as defined above. I came to the conclusion that MLB organizations could be used natural controls, since all of them had similar medical reporting standards but not something like the MiLB wage and housing changes.

My resulting balanced panel contains approximately 1,900 organization-year observations after cleaning. The dataset is across all 30 MLB parent organizations and their MiLB affiliates, with over 39,000 unique player-year records.

4.4 Measurement Considerations

Because the number of minor-league players decreased after 2021 due to roster reductions instituted, the outcome measures I have are expressed as rates as opposed to raw counts, something I chose to not really mention in the paper. Organization fixed effects and year fixed effects are included in the regression models to address unobserved heterogeneity and time shocks. In my later version and extensions, I will test heterogeneity by signing bonus amount, hypothesizing that the welfare effects of wage and housing reforms are strongest for players with smaller initial financial cushions. That will be by examining of there was a more significant effect on players with signing bonuses of under roughly $400,000, although this might shift somewhat with inflation affecting nominal values of money over the 10-year period from 2015-2025.

5 Empirical Strategy

My empirical analysis is based on a difference-in-differences framework applied to injury outcomes at the organizational level. I supplement the regressions with graphical checks of pre-treatment trends.

5.1 Difference-in-Differences Framework

The main specification follows a standard difference-in-differences model:

$$y_{it} = \alpha_i + \gamma_t + \beta(Post_t \times Treated_i) + \varepsilon_{it}, \tag{1}$$

where y_{it} is an outcome for organization i in year t (IL spells per player or IL days per injured player). The term α_i denotes organization fixed effects that absorb time-invariant differences across teams, including training facilities, medical staffing, or organizational philosophy. The year fixed effects γ_t capture league-wide shocks that affect all clubs, such as rule changes, schedule structure, or pandemic-related disruptions. The coefficient β measures the average change in injury outcomes for treated organizations relative to untreated organizations in the post-reform period.

The form of the treatment varies across policy episodes. For the 2019 comparison, $Treated_i$ equals one for the Toronto Blue Jays system, which raised minor-league salaries prior to the league-wide adjustment. For the 2021 and 2023 reforms, which applied to all minor-league affiliates, major-league teams serve as the comparison group since they did not experience comparable changes in wages or housing guarantees. In those specifications, $Treated_i = 1$ for MiLB affiliates and $Treated_i = 0$ for MLB clubs.

5.2 Parallel Trends Assumption

Although the main identification comes from the difference-in-differences model, I also examine pre-reform trajectories of injury outcomes to assess whether the treated and control groups exhibited similar patterns before each policy change. This check is important for evaluating the parallel trends assumption, which underlies the causal interpretation of the estimates. I computed and plotted year-to-year ratios of total IL days for MLB and MiLB from 2015 to 2019 and found that the pre-period paths move closely together. Figure 1 presents this comparison.

To check the parallel trends assumption more formally, I test for pre-trends in each of the difference-in-differences designs. For the 2019 Blue Jays reform, I use player-level data from

2015–2018 and interact the Blue Jays treatment indicator with year dummies. A joint F-test of these pre-treatment interactions yields p-values around 0.7–0.8 for both IL spells and IL days, so I do not find evidence of differential pre-trends before 2019.

For the 2021 leaguewide raise, I work with an organization-year panel that compares MiLB to MLB from 2015–2019. I regress IL days per player and IL spells per player on year dummies, a MiLB indicator, and their interactions, and then test whether all MiLB$\times$year terms are jointly zero. For IL days per player, the joint test has a p-value of about 0.06: I cannot reject equal pre-trends at the 5% level, but the evidence is borderline. For IL spells per player, the pre-trend test is strongly rejected ($p < 0.01$), which means the MiLB–MLB gap in spells per player was already shifting before 2021. This implies that the 2021 DiD coefficients likely combine the effect of the salary increase with pre-existing divergence between MiLB and MLB. In the discussion, I therefore lean more heavily on the IL-days-per-player results for interpretation and treat the spells-per-player DiD as more descriptive, noting that any causal reading must account for these pre-trends.

For the 2023 CBA and housing reforms, MiLB has already been differentially affected by the 2021 raise and the 2022 housing requirement, so there is no long pre-period in which MiLB can reasonably be treated as "untreated" relative to MLB. Because of this, I do not view a standard "pre-2023" parallel-trends test as informative. Instead, I interpret the 2023 MiLB$\times$ post coefficients as capturing changes in MiLB injuries relative to MLB on top of the post-2021 baseline, and I emphasize this limitation when discussing the identification assumptions for the 2023 design.

5.3 Identification and Assumptions

The key identifying assumption that is important to have is that in the absence of compensation and related reforms, injury outcomes for treated and control organizations would have followed similar paths. The structure of professional baseball and the timing of the reforms support this assumption in a few ways:

First, the year fixed effects absorb league-wide shocks such as rule changes, training protocol adjustments, or pandemic-related disruptions. Second, excluding the 2020 season removes the

year in which no minor-league games were played, preventing distortions in the construction of injury rates. Third, several robustness checks like dropping extreme injury durations, removing placeholder entries, and re-scaling by relevant denominators of players help make sure that the results are not driven by incorrect factors.

Other general important features reinforce identification also:

- **Exogenous timing of reforms.** The salary increases in 2021 and the broader wage and housing improvements for players laid out in the 2023 CBA were set at the league level. These changes were made through collective bargaining and league management negotiations rather than by injury trends or team-level medical considerations, which essentially rules out reverse causality in which organizations adjust pay in response to injury patterns or an argument of similar nature.

- **Stable roster composition within organizations.** A concern that may arise is that players may want to join the Blue Jays' if their salaries are higher. This is irrelevant both because the salary increase happened right before the season started, when rosters are essentially finalized, and minor-league player mobility is in general extremely limited when compared to other labor markets. Players stay within their organizations unless promoted, released, or traded. Because these moves occur under strict contractual rules, large-scale sorting of players across organizations in response to the reforms is very unlikely. Players have basically no say in the team they go to next. This eliminates the concern that compositional changes, rather than policy changes, account for differences in injury outcomes.

Taken together, these features strengthen the case that the estimated coefficients capture the effect of improved compensation and living conditions on player health as opposed to it being from changes in roster makeup or endogenous policy timing.

5.4 Interpretation of Coefficients

A negative coefficient on the term in models of IL spells per player indicates a reduction in the roster-wide frequency of IL usage, meaning fewer injury spells per rostered player on average. For IL days per injured player, a negative coefficient indicates that injured players in treated organizations spend less total time on the IL over the course of the season. Because these two outcomes are normalized by different denominators, all rostered players versus injured players only, they capture different types of injury risk and severity. Because of this, patterns in the estimates need to be interpreted jointly.

5.5 Summary

The regression framework provides a clear way to estimate the effects of salary and housing improvements on player welfare. The difference-in-differences specification identifies average policy effects, and the graphical pre-trend checks support the assumption that MLB clubs form a suitable comparison group for minor-league affiliates in the years preceding the reforms.

6 Results

This section presents the objective results I found linking compensation reforms to player health outcomes. I report findings for the three major policy shocks in sequence: the 2019 Blue Jays salary increase, the 2021 leaguewide raise and contraction, and the 2023 CBA introducing higher wages and housing guarantees. All models include organization and year fixed effects, with standard errors clustered at the organization level.

6.1 2019 Blue Jays Organization-Specific Raise

The first test examines the Toronto Blue Jays' unilateral decision to raise minor-league salaries in 2019, treating other MiLB organizations as controls. The dependent variable is either the number

of IL spells or the total number of IL days for each player–year. Results indicate no statistically significant change in injury frequency but a modest reduction in injury duration. The coefficient on the interaction term ($Post_{2019} \times BlueJays$) for IL days is approximately -8.15 (p = 0.079, so only significant at 10%), suggesting an over one-week reduction in total time lost to injury per injured player relative to peers.

This is a strong result that is consistent with improved recovery or earlier injury detection due to better nutrition support and general living conditions. Given that this is only one year, these estimates should be interpreted as relatively suggestive rather than definitive, but they provide some indication that wage improvements may enhance player welfare through reduced injury severity. Additionally, it's important to mention that the salary raise was only implemented shortly before the season began, so its effects were expected to be limited, as the hypothesis aligns with the belief that a significant portion of the effects of higher salaries would be from increased training and decreased levels of working second jobs throughout the off-season.

6.2 2021 Leaguewide Raise and Contraction

The 2021 season introduced two simultaneous changes: a mandated increase in minor-league salaries and a roughly 25% contraction of affiliates and roster slots. To isolate the effect on health outcomes, I compare MiLB to MLB organizations using a difference-in-differences specification over 2015–2022.

Estimates show a large and significant decline in IL spells per player for MiLB organizations relative to MLB. The interaction coefficients for 2021 and 2022 are -0.327 ($p < 0.001$) and -0.252 ($p < 0.001$), respectively, which is to roughly 25–30% fewer IL spells per player compared with MLB's baseline rate of 1.28 spells per player.

In contrast, IL days per injured player show no statistically significant change, indicating that the reduction in the roster-wide rate of injury spells did not coincide with a clear change in the season-long IL time among players who did become injured. These results seem to imply that improved compensation led to a healthier player pool. As mentioned in the literature section,

while MLB injuries are believed to have spiked in 2021 due to pandemic-related conditioning issues, these results also support the hypothesis that better conditions in MiLB mitigated the broader baseball-wide injury surge that would have been expected similarly in MiLB without improvements in training and living resources.

6.3 2023 Collective Bargaining Agreement and Housing Reforms

The 2023 CBA brought forward a second round of wage increases and also guaranteed housing, significantly improving players' overall living conditions.

For IL spells per player, the 2023 interaction coefficient is $+0.245$ (p = 0.005), with smaller and statistically insignificant effects in 2024 and 2025. For IL days per injured player, the coefficients are 44.5, 54.3, and 34.4 (all $p < 0.001$). This combination of a short-term rise in reported injuries but shorter total recovery times isn't completely clear, but could imply a significant decrease in major injuries stemming from better living conditions.

These results align closely with the mechanism suggested by McLeod et al. (2022), in which perceived organizational support increases engagement and compliance with medical guidance. They also correspond with MLB-wide findings from Platt et al. (2021), which show persistently elevated injuries post-pandemic, reinforcing that MiLB's improvements occurred despite major baseball-wide injury spikes.

7 Discussion and Interpretation

The results seem to demonstrate that higher wages and improved living conditions have measurable effects on player health, but the mechanisms and magnitudes differ across policy shocks. The 2019 Blue Jays pay raise appears to have reduced injury severity rather than frequency, suggesting that even modest income improvements can enhance recovery through better nutrition, rest, or compliance with medical guidance. This aligns closely with the framework of McLeod et al. (2022), who find that perceived organizational support predicts better developmental outcomes

among minor leaguers, which is very interesting and not necessarily expected.

The 2021 league wide salary increase and simultaneous contraction brought about the largest health improvement, with significantly fewer IL spells per player in MiLB relative to MLB. Although roster reductions may complicate interpretation, the direction of effects is consistent with higher pay allowing players to invest more in training and recovery, while smaller rosters may have concentrated opportunities among healthier or better-conditioned athletes.

In interpreting the 2021 coefficients, note that IL spells per player is a roster-wide rate, while IL days per injured player averages only over players who appear on the IL at least once. A lower IL spells per player combined with no statistically significant change in IL days per injured player is therefore consistent with a smaller fraction of the roster ever going on the IL, even if the total time spent on the IL among injured players does not change.

The 2023 collective bargaining agreement was the biggest change in terms of salaries and condition improvements. It introduced guaranteed housing and extended pay across the offseason, showing a real shift in working conditions and the way minor leaguers were treated by their team owners. My findings of shorter IL durations despite slightly higher injury counts could be consistent with a mechanism of improved reporting with players disclosing more injuries, as there is generally a cultural shift across sports to cooperate and not push through injuries in a way that was more common in the past.

As further qualitative context, reporting on the Blue Jays' 2019 decision to raise minor-league salaries by 50 percent quotes players saying that the higher pay let them afford better food, offseason training, and housing, and helped them focus more on baseball instead of juggling multiple jobs Brudnicki (2019). A former Blue Jays minor leaguer interviewed for this paper similarly emphasized that the raise and the later housing guarantees significantly improved day-to-day living conditions and mental health by reducing overcrowded apartments, improving access to quality food, and easing financial stress, while not substantially changing how hard players trained or the pressure they felt to play through injuries.

A related question is whether any of the compensation reforms altered players' signing-bonus

or incentive-bonus structures, which could in principle change injury-reporting behavior. Signing bonuses in MiLB are paid at the moment a player signs their first professional contract and are determined by draft position or international amateur status. These bonuses are fixed and were not affected by any of the recent pay reforms mentioned throughout the paper. Historically, there have been no performance-contingent or playing-time bonuses in the minor leagues, and the reforms didn't change that. Because bonus structures remained constant, the reforms shifted ongoing income rather than incentives, showing that this issue in not one was considered as substantial.

Together, these results demonstrate that compensation and welfare reforms affect more than just financial outcomes. They may change the behavioral environment in which players manage their health. Higher wages and stable housing appear to reduce the physical and possibly psychological barriers to recovery, showing observable improvements in time lost to injury. The minor-league reforms provide us a clear case study of how organizational investments in basic needs can produce measurable improvements within professional sports.

8 Conclusion and Limitations

This paper provides new empirical evidence on how compensation and living standards affect player health in Minor League Baseball. Using difference-in-differences approaches from 2015–2025, I show that higher wages and improved housing are linked to less time spent on the IL among injured players and modest declines in injury incidence. The combined evidence indicates that financial security and stable living conditions allow players to manage injuries more proactively, reducing time lost while improving overall welfare.

The 2019 Blue Jays pay raise, the 2021 leaguewide salary increase, and the 2023 collective bargaining agreement together represent a great natural experiment in the economics of professional sport. The effects are strongest following the 2023 CBA, suggesting that structural supports such as guaranteed housing and improved facilities and trainers could cause the biggest results. These results seem to follow the view that health outcomes in athletic work markets respond not just to

income levels but to the environments in which work occurs.

There are several limitations that qualify my findings. First, the current analysis relies exclusively on injury data and cannot yet link improved health to performance metrics such as velocity, fielding, or promotion rates. Second, structural changes around 2020–2021 such as pandemic disruptions, affiliate contraction, and data-reporting adjustments introduce noise that fixed effects cannot fully account for. Finally, the housing changes in 2023 add an additional treatment effect essentially that is hard to adjust for, although it can be estimated to provide an equivalent of a salary increase of around $10,000.

The next step for this work could be to evaluate if the increases in salary could be worthwhile for the MLB clubs from a financial standpoint, balancing the higher cost with increased health and durability of players, a question relating to the broader labor economics space.

References

Brudnicki, A. (2019). Minor Leaguers Praise Toronto for Salary Boost. MLB.com.

Feinsand, M. (2021). Revamped Minor Leagues Enjoy Historic 2021. MLB.com.

McLeod, C. M., Agha, A., and Rosenblum-Larson, M. (2022). Organizational Support Factors for Minor League Baseball Player Development. *Journal of Sport Management*, 36(5):411–424.

MLB News Staff (2021). MLB to Provide Minor League Player Housing. MiLB.com, November 18, 2021.

Platt, B., McLeod, C. M., and Andrews, J. P. (2021). A Plague of Their Own: Injury Incidence Remains Elevated in the 2021 Major League Baseball Season Compared to Pre-COVID-19 Seasons. *Orthopaedic Journal of Sports Medicine*, 9(12):1–8.

Waldstein, D. (2023). Minor League Salaries Will Double Under New Deal. *The New York Times*.

Appendix: Tables and Figures

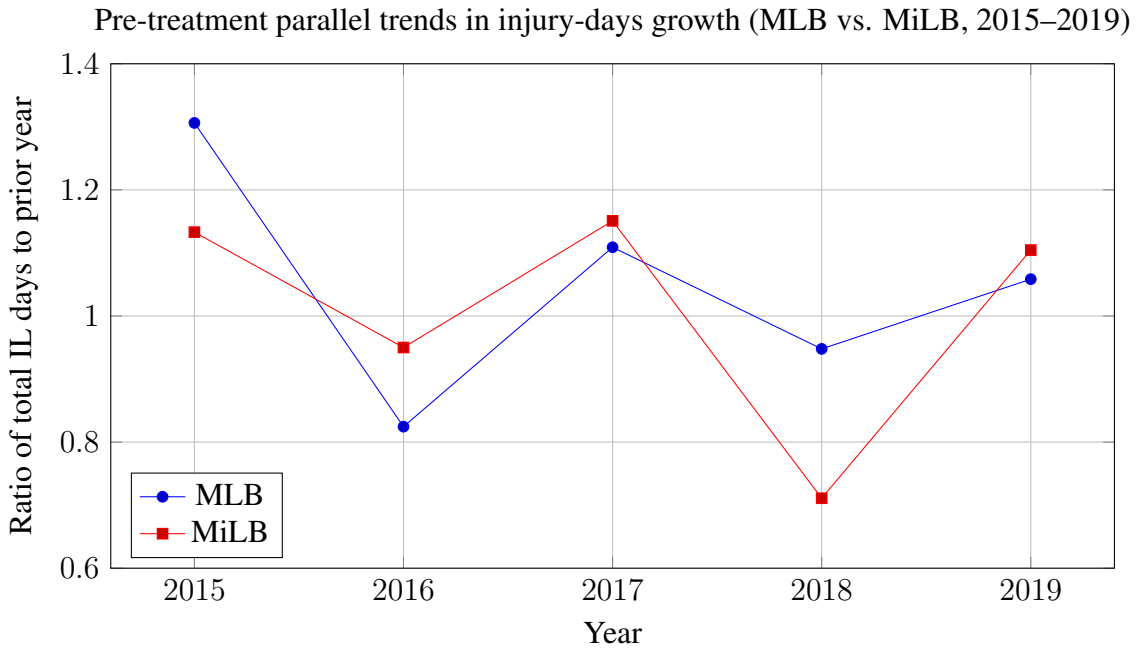


Figure 1: Pre-treatment parallel trends in injury-days growth for MLB and MiLB.

Policy Shock	Outcome	Direction	Estimate	95% CI	p-value	Possible Interpretation
2019 Blue Jays	IL Days	↓	–8.15	[–17.27, 1.00]	0.079	Better nutrition led to quicker recoveries post-raise
2021 Leaguewide	IL Spells	↓	–0.33	[–0.38, –0.27]	< 0.001	Fewer injuries due to better training during COVID
2023 CBA	IL Spells	↑	+0.25	[0.08, 0.41]	0.005	More reporting of minor injuries in Rookie League
2023 CBA	IL Days	↓	–44.5	[–66.3, –22.7]	< 0.001	Much faster recovery times due to large salary increase

Table 1: Summary of Estimated Effects by Policy Shock

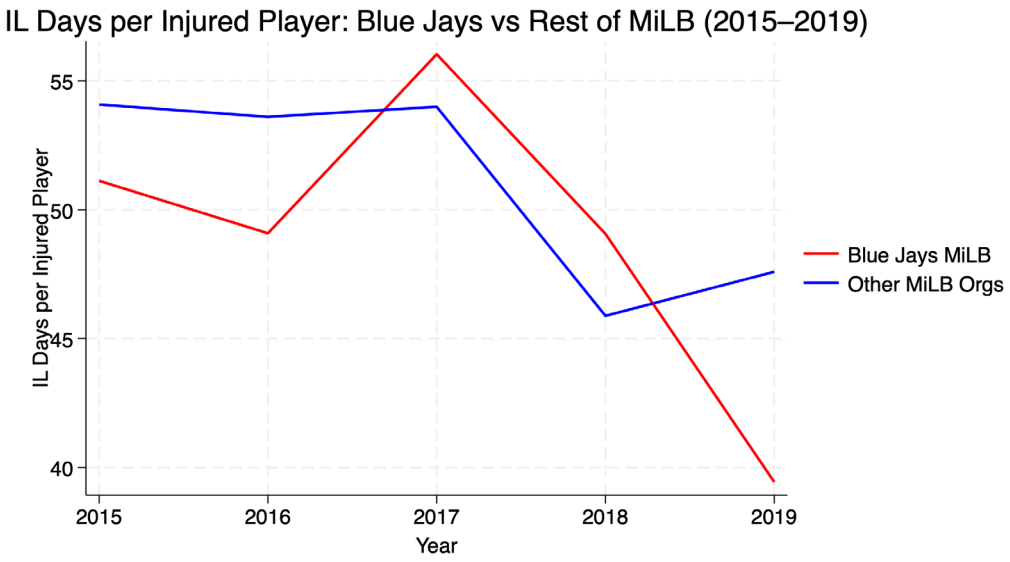


Figure 2: Average IL days per injured player for the Toronto Blue Jays vs. other MiLB organizations, 2015–2019. The figure plots mean IL days per injured player by year separately for Blue Jays minor leaguers and for all other minor league organizations.

Table 2: Effect of 2019 Blue Jays Salary Increase on Injuries (MiLB only)

	IL spells per player	IL days per injured player
Treat × 2019	0.036	-8.151*
	(0.076)	(4.646)
	[–0.113, 0.185]	[–17.259, 1.000]
Year fixed effects	Yes	Yes
Observations	8,522	8,522
R-squared	0.001	0.006
Clusters (player)	5,733	5,733

Notes: Ordinary least squares estimates with standard errors clustered by player identifier (pid) in parentheses. The sample includes minor league players from 2015–2019. Treat is an indicator for players in the Toronto Blue Jays organization. The interaction Treat × 2019 captures the difference in injuries for Blue Jays minor leaguers in 2019 relative to other organizations, controlling for year fixed effects. Bracketed values report 95% confidence intervals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Effects of 2021 salary increase on IL outcomes (MiLB vs MLB)

	IL days per injured player	IL spells per player
MiLB	13.985***	-0.033**
	(1.112)	(0.011)
2021	-10.407***	0.185***
	(1.523)	(0.015)
2022	-0.315	0.140**
	(4.066)	(0.055)
MiLB $\times$ 2021	10.497***	-0.327***
	(2.154)	(0.013)
MiLB $\times$ 2022	12.246***	-0.252***
	(4.403)	(0.055)
Year fixed effects	Yes	Yes
Observations	1,239	1,239
R-squared	0.075	0.090
Clusters (org)	323	323

Notes: Ordinary least squares estimates with standard errors clustered by organization in parentheses. The unit of observation is organization by level by year. MiLB is an indicator for minor league affiliates. 2021 and 2022 are year dummies. Interaction terms capture the difference in outcomes for MiLB relative to MLB in each post period. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effects of 2023 CBA and housing reforms on IL outcomes (MiLB vs MLB)

	IL spells per player	IL days per injured player
MiLB	-0.233***	18.082***
	(0.051)	(3.387)
2023	-0.214***	50.701***
	(0.073)	(8.769)
2024	-0.048	97.342***
	(0.093)	(12.357)
2025	-0.124	3.833
	(0.084)	(9.071)
MiLB $\times$ 2023	0.260***	-37.065***
	(0.075)	(9.075)
MiLB $\times$ 2024	0.143	-65.253***
	(0.094)	(12.686)
MiLB $\times$ 2025	0.245**	-33.812***
	(0.086)	(9.314)
Year fixed effects	Yes	Yes
Observations	1,898	1,898
R-squared	0.053	0.292
Clusters (org)	334	334

Notes: Ordinary least squares estimates with standard errors clustered by organization in parentheses. The unit of observation is organization by level by year. MiLB is an indicator for minor league affiliates. 2023, 2024, and 2025 are year dummies. Interaction terms capture the difference in outcomes for MiLB relative to MLB after the 2023 CBA and housing reforms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Stability and the Origins of Productivity Division: Explaining the South Korea–Philippines Divergence

Ethan Nguyen

December 10, 2025

1 Introduction

Over the past six decades, the economic histories of **South Korea** and the **Philippines** have diverged dramatically. In the early 1960s, the two countries shared striking similarities: both were developing economies recovering from war and colonial rule, both maintained close political and economic ties with the United States, and both faced the challenge of transforming predominantly agrarian societies into modern industrial economies. At that time, the Philippines was considered one of the wealthier nations in Asia, with GDP per capita roughly double that of the war-torn Korea. Yet by 2020, South Korea had become a high-income, technologically advanced economy (known as one of the "Asian Tigers") while the Philippines remained a middle-income country characterized by slower and more volatile growth. This paper seeks to understand the proximate sources of this divergence by decomposing the growth of output per worker into contributions from **capital accumulation** and **total factor productivity (TFP)**.

The analysis follows the **growth accounting framework** introduced in the Solow growth model. In this framework, economic growth can be traced to two main proximate causes: increases in the amount of capital each worker has to work with, and improvements in the efficiency with which that capital and labor are used (captured by the residual component, TFP). Formally, growth in output per worker can be expressed as

$$\Delta \ln(y) = \alpha \Delta \ln(k) + \Delta \ln(A),$$

where y is output per worker, k is capital per worker, α is the capital share, and A represents total factor productivity. The first term measures **capital deepening**, while the residual represents **TFP growth**: the part of economic growth that cannot be explained by the accumulation of physical inputs alone.

The baseline analysis in this paper therefore separates growth into these two channels, capital versus TFP, using **PPP-adjusted** data from the **Penn World Table (version 11.0)**, which ensures that income measures across countries reflect differences in real purchasing power rather than exchange rate fluctuations.

Through our growth accounting analysis, we can identify whether rapid growth is primarily the result of **factor accumulation** (as in the early stages of industrialization) or of **productivity improvements** (as economies mature and technological diffusion takes hold). Prior studies such as Young (1995) and Bosworth & Collins (2008) have applied similar methods to the East Asian Tigers and found that while early growth was driven by high savings and investment, long-run income convergence depended crucially on sustained gains in productivity. Applying this lens to Korea and the Philippines provides insight into how two economies with shared origins in the postwar era came to such different outcomes.

The trends in Figures 1 and 2 illustrate the dramatic long-run divergence in economic performance between South Korea and the Philippines. From the early 1950s through the late 1970s, both economies expanded steadily, though

Korea began from a substantially lower output base in the aftermath of war. Starting in the 1960s, Korea's GDP and income per capita began rising at an accelerating pace, with growth compounding year after year and continuing largely uninterrupted through successive decades. The Philippines, by contrast, exhibited slower and more volatile growth, marked by prolonged periods of stagnation in the 1980s and 1990s before a modest recovery in the 2000s. By the early 1980s, Korea had surpassed the Philippines in both aggregate and per-capita output, and the gap continued to widen thereafter. The log-scale plots emphasize just how persistent and exponential Korea's expansion has been relative to the relatively flat trajectory of the Philippines, signaling an enduring divergence in economic development that motivates the growth-accounting analysis to follow.

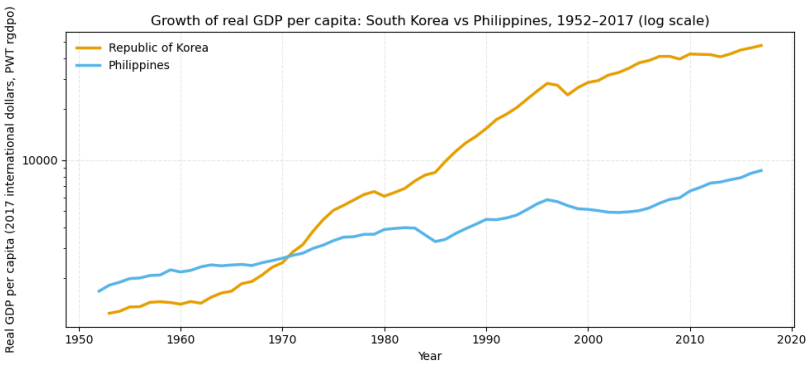


Figure 1: Growth of real GDP per capita (2017 US$) for South Korea and the Philippines, 1952–2017, plotted on a logarithmic scale. Korea overtook the Philippines by the late 1970s and continued to widen the income gap, illustrating the long-run productivity divergence discussed in this paper. Source: Penn World Table 11.0.

The remainder of this paper proceeds in three stages. Section 2 provides further background on the economic histories of South Korea and the Philippines, emphasizing the institutional and policy differences that framed their development paths. Section 3 outlines the data and methodology used in the growth accounting decomposition, clarifying how capital and TFP growth are measured and interpreted. Section 4 presents and discusses the results, while Section 5 develops the argument that productivity differences (rooted in institutions, policy, and structural transformation) are the fundamental causes of divergence between the two economies.

As mentioned above, South Korea and the Philippines began the 1960s at similar income levels, their subsequent growth trajectories diverged sharply. Using growth accounting based on Penn World Table data, this paper shows that Korea's sustained rise in output per worker was driven primarily by gains in total factor productivity (TFP) rather than capital accumulation, whereas the Philippines' growth relied almost entirely on factor inputs with stagnant

efficiency.

This paper argues that the long-run divergence between South Korea and the Philippines reflects not just differences in investment levels, but in the institutional and structural environments that governed how capital was used. Korea's coordinated industrial policy, credible institutions, and disciplined financial system enabled capital deepening to translate into sustained gains in total factor productivity (TFP). The Philippines, by contrast, suffered from policy volatility, weak financial intermediation, and limited human-capital absorption, leaving growth dependent on factor accumulation alone and trapping the economy in a low-TFP equilibrium.

2 Economic Background: Divergent Paths in East Asia

In the decades following World War II, both **South Korea** and the **Philippines** faced the daunting challenge of rebuilding their economies and establishing the foundations for long-term growth. Despite their shared geopolitical ties and similar levels of development in the early postwar period, their trajectories over the next sixty years could not have been more different. In 1960, the Philippines was among the most prosperous nations in Asia, with a GDP per capita exceeding that of South Korea, Thailand, and even parts of southern Europe. South Korea, by contrast, was emerging from the devastation of the Korean War (1950–1953), with per capita income levels comparable to sub-Saharan Africa. The country had limited natural resources, widespread poverty, and a primarily agrarian economy. Yet by the turn of the twenty-first century, South Korea had become one of the world's most advanced industrial economies, while the Philippines lagged behind, struggling to achieve sustained growth and industrial transformation. Understanding this divergence requires examining the historical, institutional, and policy environments that shaped each country's growth dynamics.

2.1 Early Conditions and Development Strategies (1950s–1970s)

During the 1950s and early 1960s, both economies operated under import-substitution industrialization (ISI) strategies influenced by postwar development theory. The Philippines adopted this model earlier and more decisively. Benefiting from close economic ties to the United States, it enjoyed access to foreign markets and capital, and early industrialization centered on light manufacturing and agriculture-based exports. However, the structure of protectionist policies soon led to inefficiencies: domestic industries catered to small local markets, capital accumulation was concentrated in politically connected firms, and productivity growth stagnated.

In contrast, South Korea's developmental trajectory began under much harsher conditions. Following the Korean War, the economy was heavily dependent on

U.S. aid, with output dominated by agriculture and basic consumer goods. The country's transformation began under President **Park Chung-hee**, whose military government launched the **First Five-Year Economic Development Plan (1962–1966)**. This plan marked a decisive shift from import substitution to **export-oriented industrialization (EOI)**. State-led investment in infrastructure, credit allocation to strategic sectors, and tight coordination between government and business (the emerging **chaebol** conglomerates) set the stage for rapid capital accumulation and industrial diversification.

A second structural difference that emerged in the 1960s and 1970s concerns domestic savings and the policy environment shaping investment. Penn World Table data confirm that Korea's national saving rate rose rapidly, from roughly 15 percent of GDP in the early 1960s to above 30 percent by the late 1970s, while the Philippines remained closer to 18–20 percent over the same period (PWT 11.0). Higher domestic savings allowed Korea to finance investment internally, reducing exposure to external borrowing cycles. By contrast, the Philippines relied heavily on foreign borrowing to support domestic investment, making capital formation more vulnerable to global interest-rate shocks and domestic macro instability. This financial divergence was reinforced by trade policy: beginning in the late 1960s, Korea sharply reduced average tariffs on intermediate and capital goods as part of its shift toward export-oriented industrialization, whereas the Philippines maintained significantly higher effective protection rates, particularly in manufacturing, well into the 1980s. These trade-policy differences shaped the cost of imported machinery and the incentives facing firms, further amplifying Korea's ability to accumulate productive capital while limiting the Philippines' access to technologically embedded imports.

2.2 Institutional Divergence and Productivity Gaps (1980s–1990s)

The 1980s and 1990s marked a critical phase of institutional and technological divergence. Korea continued its upward trajectory, transitioning from heavy industries to high-technology manufacturing and services. The government pursued selective liberalization while maintaining strong support for export competitiveness. Education policy also played a central role: by 1990, secondary enrollment in Korea exceeded 90 percent, feeding a steady rise in labor quality and innovation capacity.

The Philippines, meanwhile, entered a period of stagnation and volatility. The **Marcos regime (1965–1986)** oversaw initial infrastructure investment but also increasing corruption, political repression, and rising external debt. The 1983 debt crisis and subsequent political turmoil following the **People Power Revolution of 1986** severely disrupted investment and confidence. Structural reforms in the post-Marcos period such as trade liberalization, financial deregulation, and privatization, were often unevenly implemented. As a result, growth spurts in the 1990s were followed by frequent slowdowns. Investment rates occasionally recovered, but weak institutions and inconsistent policy undermined efficiency.

While Korea leveraged industrial upgrading and technological diffusion to sustain growth, the Philippines remained caught in what could be seen as a **middle-income trap**, unable to transition from labor-intensive to knowledge-based production. Korea's productivity gains were reinforced by stable macroeconomic policy and rising innovation capacity, while the Philippines faced recurring fiscal crises, uneven infrastructure development, and institutional fragmentation.

2.3 Convergence, Crisis, and Structural Maturity (2000s–2020)

By the early 2000s, the growth patterns of the two countries had largely solidified. South Korea had joined the OECD, becoming a global leader in electronics, automotive production, and digital innovation. The **1997 Asian Financial Crisis** served as a temporary shock but did not derail Korea's long-term trajectory. Policy responses such as swift corporate restructuring, labor flexibility reforms, and continued investment in research and development helped maintain productivity growth.

The Philippines, in contrast, experienced moderate recovery after the crisis but failed to achieve a structural transformation of comparable scale. While remittances from overseas Filipino workers supported consumption, they did little to raise productivity or expand the industrial base. The services sector, particularly business process outsourcing (BPO), grew rapidly, but manufacturing stagnated. This pattern generated modest growth in GDP per capita but limited improvements in capital efficiency. Despite political reforms and growing foreign investment, the Philippines' growth remained weak, reflecting persistent infrastructure bottlenecks, regulatory uncertainty, and skill mismatches in the labor force.

2.4 Summary of Divergent Development Paths

The historical record highlights a consistent theme: South Korea's ascent was built on **high and efficient investment**, **export-oriented growth**, and **productivity-enhancing structural change**, whereas the Philippines' performance was constrained by **lower and less efficient investment**, **institutional volatility**, and **limited technological diffusion**.

This background provides the context for the quantitative analysis that follows. By decomposing the growth of output per worker into contributions from capital and TFP, we can formalize the intuition behind these narratives. The next section outlines the methodological framework and data sources used to measure these components and evaluates how differences in factor accumulation and productivity dynamics explain the divergent economic outcomes of Korea and the Philippines.

3 Methodology and Data

This section describes the empirical framework used to decompose growth in output per worker into the contributions of capital deepening and total factor productivity (TFP). The approach follows the growth-accounting procedure implied by the Solow model and used in the course lectures. The objective is diagnostic rather than causal: to quantify how much of observed growth can be attributed to measured factor accumulation versus a residual efficiency term.

3.1 Analytical Framework

I assume an aggregate Cobb–Douglas production function with constant returns to scale:

$$Y_t \;=\; K_t^{\alpha}\left(A_t L_t\right)^{1-\alpha},$$

where Y_t is real output, K_t the physical capital stock, L_t employment, A_t TFP, and α the capital share of income. Dividing by labor and denoting $y_t \equiv Y_t/L_t$ and $k_t \equiv K_t/L_t$ yields

$$y_t \;=\; A_t\, k_t^{\alpha}.$$

Taking logarithms and time differences gives the growth-accounting identity in per-worker terms:

$$\Delta \ln y_t \;=\; \alpha\, \Delta \ln k_t \;+\; \Delta \ln A_t.$$

The first term, $\alpha \Delta \ln k_t$, is the contribution of capital deepening; the residual $\Delta \ln A_t$ is interpreted as TFP growth.

For each period $[t_0, t_1]$, I compute average annual log growth using the endpoint formula

$$g_x \;\equiv\; \frac{\ln x_{t_1} - \ln x_{t_0}}{t_1 - t_0},$$

and report contributions in percentage points per year.

3.2 Data and Variable Construction

All series come from the Penn World Table (PWT), version 11.0 (sheet *Data*), expressed in PPP-adjusted constant prices to ensure cross-country comparability. The key variables are constructed as follows:

Concept	PWT variable(s)	Definition
Output per worker	`rgdpo/emp`	$y_t \equiv Y_t/L_t$
Capital per worker	`rkna/emp`	$k_t \equiv K_t/L_t$
Labor share	`labsh`	$1-\alpha_t$

The sample covers South Korea (KOR) and the Philippines (PHL) from 1960 to 2020. To align with the historical narrative in Section 2, I report decompositions for three subperiods—1960–1980, 1980–2000, and 2000–2020—and for the full horizon 1960–2020.

3.3 Capital Share Parameter α

In this paper, I estimate country-specific capital income shares directly from the Penn World Table's (PWT 11.0) variable `labsh`, which reports labor's share of value added. Because `labsh` measures $(1 - \alpha)$, I compute α for each country as the complement of the average labor share over the full sample period (1960–2020). This approach allows for heterogeneity in factor income distribution across economies, reflecting structural differences in labor market composition, industrial organization, and national accounts measurement. The mean of `labsh` across years smooths out short-term fluctuations and measurement noise, yielding a stable, country-specific estimate of α consistent with standard practice in empirical growth accounting. This method provides a more accurate representation of each country's production structure than assuming a common global capital share, ensuring that subsequent decompositions of output growth into capital, labor, and total factor productivity (TFP) components reflect country-level realities.

So since `labsh` measures $(1-\alpha_t)$, a period-average capital share for country c over the full window $\mathcal{T} = \{1960, \ldots, 2020\}$ is

$$\bar{\alpha}_c \;=\; 1 \;-\; \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \texttt{labsh}_{c,t}.$$

I then recompute the decomposition using $\bar{\alpha}_{\mathrm{KOR}}$ and $\bar{\alpha}_{\mathrm{PHL}}$ (held fixed within each country across subperiods).

3.4 Implementation Details and Limitations

All growth rates are computed as average annual log changes using period endpoints to reduce sensitivity to short-run fluctuations. TFP is measured residually; as such, it encapsulates not only technological progress but also changes in efficiency, resource allocation, utilization, and (implicitly) human-capital quality. Measurement error is a concern, especially for early capital stock and employment figures; focusing on multi-decade averages and PPP-adjusted series mitigates these issues. The decomposition is proximate: Section 5 interprets the TFP patterns in light of the institutional and structural differences documented in Section 2.

4 Results and Discussion

4.1 Overview

Using country-specific capital shares derived from the Penn World Table ($\alpha_{\mathrm{KOR}} = 0.50$, $\alpha_{\mathrm{PHL}} = 0.56$), I decompose the growth of output per worker into the contributions of capital deepening and total factor productivity (TFP). This specification allows each country's income structure to determine how factor accumulation and efficiency improvements translate into output growth. All

growth rates are expressed as average annual log changes in percentage points per year.

4.2 Growth Decomposition

Table 1 reports the growth accounting results for South Korea and the Philippines over three subperiods (1960–1980, 1980–2000, 2000–2020) and the full horizon (1960–2020).

Table 1: Growth Accounting Results using Country-Specific α (PWT 11.0)

Country	Period	$\Delta \ln y$	$\alpha \Delta \ln k$	$\Delta \ln A$
South Korea	1960–1980	6.05	2.26	3.79
	1980–2000	6.63	4.13	2.50
	2000–2020	2.24	1.93	0.31
	1960–2020	4.97	2.78	2.20
Philippines	1960–1980	2.34	1.77	0.57
	1980–2000	1.05	0.68	0.37
	2000–2020	2.63	2.18	0.45
	1960–2020	2.01	1.54	0.46

Figure 2 summarizes the growth-accounting results by period using country-specific capital shares from PWT ($\alpha_c = 1 - \overline{\texttt{labsh}}_c$). Korea's growth shows large and durable contributions from both capital deepening and TFP, whereas the Philippines' growth is dominated by capital with comparatively weak TFP.

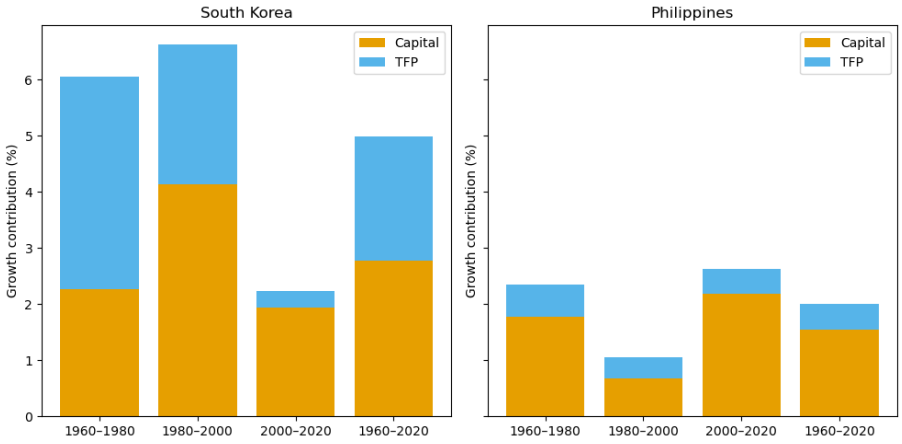


Figure 2: Decomposition of GDP growth (average annual log changes, percentage points per year) into contributions from capital and TFP for South Korea and the Philippines across major periods. Country-specific α values are computed from PWT 11.0 labor shares. Source: Penn World Table 11.0.

4.3 Interpretation

South Korea's output per worker grew at an average annual rate of nearly five percent from 1960 to 2020, driven by both rapid capital deepening and sustained gains in productivity. During the 1960–1980 period, more than half of Korea's growth stemmed from TFP improvements associated with industrialization, export promotion, and human-capital expansion. Between 1980 and 2000, continued investment raised the capital contribution, while TFP remained robust. After 2000, Korea's growth slowed as the economy approached the technological frontier.

The Philippines shows a markedly different pattern. Growth in output per worker averaged only about two percent per year over the full period, with the majority explained by capital accumulation rather than efficiency gains. TFP growth was modest and volatile, reflecting structural and institutional weaknesses, slower technology adoption, and macroeconomic instability through the late 20th century. The modest recovery after 2000 primarily reflects renewed investment rather than broad productivity improvements.

4.4 Robustness: Baseline Capital Share $\alpha = 0.35$

As a robustness check, I recompute the decomposition using a common capital income share of $\alpha = 0.35$ for both countries (a standard lecture benchmark). Table 2 reports the results by period. Relative to the country-specific α from PWT labor shares, a lower common α mechanically reduces the capital contribution $\alpha \, \Delta \ln k$ and increases the residual (TFP) term $\Delta \ln A$ for both countries. Under

this baseline assumption of a common capital share $\alpha = 0.35$, the decomposition results are broadly consistent with those obtained using country-specific labor shares.

South Korea's growth remains dominated by total factor productivity (TFP), particularly during the 1960–2000 period, when TFP accounted for roughly two-thirds of overall output growth. In contrast, the Philippines' growth continues to be driven largely by capital accumulation, with TFP contributing only modestly and showing limited persistence across decades. Notably, adopting a lower, uniform α compresses the measured contribution of capital while slightly raising the residual TFP term for both countries, but it does not alter the qualitative ranking of growth sources.

It is worth noting that the increase in the Philippines' measured TFP in the final column of Table 2 is mechanical rather than substantive. Because the Philippines exhibits relatively low capital deepening, reducing α shifts a larger share of its observed output growth into the residual term, inflating the TFP estimate. This does not indicate stronger underlying efficiency gains; it simply reflects the sensitivity of the residual to the chosen capital share. The overall conclusion that Philippine productivity growth was modest and inconsistent compared to Korea's remains unchanged.

The robustness of these findings reinforces the conclusion that Korea's sustained rise was propelled primarily by productivity improvements, whereas the Philippines' slower trajectory reflected weaker efficiency gains despite comparable investment activity.

Table 2: Growth accounting with a common capital share $\alpha = 0.35$ (PWT 11.0)

Country	Period	$\Delta \ln y$	$\alpha \Delta \ln k$	$\Delta \ln A$
South Korea	1960–1980	6.05	1.58	4.47
	1980–2000	6.63	2.88	3.75
	2000–2020	2.24	1.35	0.89
	1960–2020	4.97	1.94	3.04
Philippines	1960–1980	2.34	1.10	1.24
	1980–2000	1.05	0.42	0.62
	2000–2020	2.63	1.35	1.28
	1960–2020	2.01	0.96	1.05

4.5 Discussion

Overall, the decomposition indicates that the long-run divergence between Korea and the Philippines is primarily explained by differences in productivity growth rather than in the rate of capital accumulation. This result aligns with the Solow model's prediction that sustained per-capita growth ultimately depends on total factor productivity. Korea's postwar trajectory exemplifies successful catch-up growth toward the global frontier, whereas the Philippines re-

mains below its potential steady state due to limited efficiency gains. While measurement error in capital and employment data may affect levels, the qualitative ranking of TFP performance is robust.

5 Macroeconomic Stability as a Determinant of Divergent TFP Paths

The preceding sections documented a sharp divergence in total factor productivity (TFP) between South Korea and the Philippines despite relatively similar income levels in the early 1960s. To account for this divergence, this section focuses on one central mechanism: macroeconomic stability. Stability is treated here not as a retrospective label for success, but as a forward-looking input into private decision-making. Firms, banks, and households form expectations about future prices, exchange conditions, and the availability of credit. These expectations, in turn, determine whether long-horizon, irreversible investments in physical capital, human capital, and organizational capabilities are undertaken. An economy that consistently delivers relatively stable macroeconomic conditions effectively lowers the risk of such investments; an economy subject to recurrent instability raises that risk and pushes private agents toward short-term, defensive behavior.

The argument proceeds in three steps. First, I show that South Korea experienced a sustained decline in domestic price volatility while the Philippines suffered large and prolonged swings in inflation. Second, I document that the Philippines faced substantially higher volatility in its terms of trade, making the cost of imported capital goods and the value of export revenues more uncertain. Third, I analyze differences in investment behavior as the revealed outcome of firms' responses to these environments. Throughout, the focus is on how stability (or instability) affects decisions that accumulate into TFP differences.

5.1 Domestic Price Stability and Inflation Volatility

A natural starting point for measuring macroeconomic stability is the behavior of domestic prices. Inflation volatility affects virtually every dimension of firm behavior: wage bargaining, pricing of output, the real burden of debt, and the value of retained earnings. Figure 3 plots a ten-year rolling standard deviation of inflation for Korea and the Philippines between 1960 and 2020. Several patterns stand out.

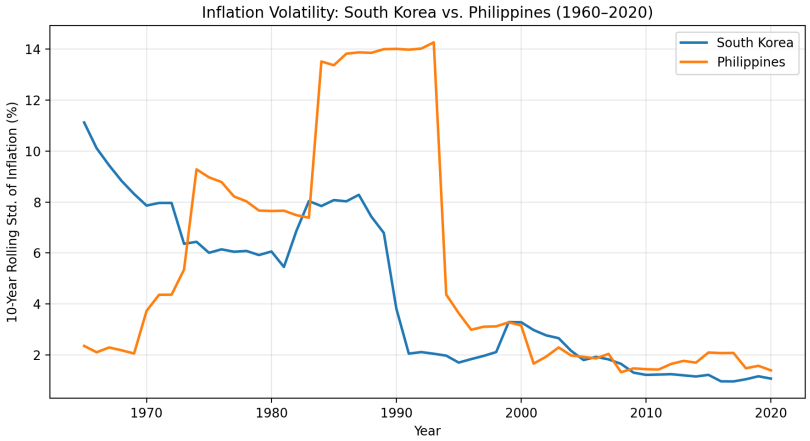


Figure 3: Ten-year rolling standard deviation of annual inflation rates in South Korea and the Philippines, 1960–2020. The Philippines exhibits prolonged periods of high and unstable inflation volatility, especially during the late 1970s and 1980s, while South Korea experiences a steady decline toward low and stable inflation levels by the 1990s. Source: World Bank.

First, Korea begins the 1960s with high inflation volatility, exceeding 10%, which is typical of economies emerging from war and reconstruction. However, there is a clear downward trend. By the late 1970s, inflation volatility stabilizes around 6–8%, and by the early 1990s it falls sharply, reaching levels close to 2% and remaining low thereafter. In other words, within roughly two decades Korea transitions from a highly unstable price environment to one in which firms can form reasonably precise expectations about future inflation.

Second, the Philippines follows almost the opposite path. Inflation volatility is modest in the 1960s, but begins to rise in the early 1970s. By the late 1970s and early 1980s, the volatility measure spikes dramatically, exceeding 13–14% and staying elevated for nearly a decade. This period corresponds to the global interest-rate shock, the buildup of external debt, the 1983 balance-of-payments crisis, banking distress, and major political transitions. Even after the return to more conventional macroeconomic management in the 1990s, volatility declines only gradually, and Philippine inflation remains more unstable than Korean inflation well into the 2000s.

This difference in inflation volatility is not simply cosmetic. In standard models of investment under uncertainty, the value of an irreversible project is the expected discounted stream of profits net of costs. Let V denote the value of a project that pays a stochastic payoff π_t in period t and requires an irreversible cost K at $t = 0$:

$$V = E_0 \left[\sum_{t=1}^{\infty} \frac{\pi_t}{(1+r)^t} \right] - K.$$

When the distribution of π_t becomes more volatile, for example, because real

input prices or real wages fluctuate with inflation surprises, the downside risks of low realizations loom larger than the upside potential from high realizations. Under convex adjustment costs or risk aversion, the increase in variance reduces V even if the mean payoff is unchanged. Moreover, because the investment is irreversible, the firm retains an option to wait. The more volatile the environment, the more valuable waiting becomes. Formally, the value of the option to delay increases with the variance of the underlying state variable, leading firms to postpone investment decisions.

Inflation volatility also interacts with financial contracts. In economies where long-term nominal debt is common, unexpected inflation shocks redistribute wealth between borrowers and lenders. If inflation is frequently higher than expected, lenders are reluctant to issue long-duration loans without charging very high risk premia or indexing contracts. In practice, this leads to a shortening of the maturity structure of debt. Firms in such environments must therefore rely on short-term credit, which is ill-suited for financing long-gestation projects such as building export-oriented factories, investing in specialized machinery, or undertaking large R&D programs. Low and predictable inflation, by contrast, makes it easier to write longer-term contracts and to price risk, supporting the development of financial instruments that match the horizon of productivity-enhancing investments.

These mechanisms are consistent with the trajectories in Figure 3. By reducing inflation volatility from double-digit levels in the 1960s to near-OECD levels by the 1990s, Korea created a domestic environment in which firms could reasonably forecast real financing costs and wage dynamics. In the Philippines, repeated inflationary episodes and large swings in volatility undermined such forecasting. The cost of importing equipment, the real value of debt, and the real wage path were all subject to frequent surprises, making it rational for firms to avoid long-term commitments or to restrict investment to projects with rapid payback and low sunk costs. Over time, these differences in the domestic price environment contributed to differences in both the level and the quality of investment, and hence to TFP.

5.2 External Stability and Terms-of-Trade Volatility

A second dimension of macroeconomic stability concerns exposure to external shocks. For economies that rely heavily on imported capital goods and intermediate inputs, the volatility of the terms of trade (the relative price of exports in terms of imports), is particularly important. Figure 4 presents a ten-year rolling standard deviation of the annual percentage change in the net-barter terms-of-trade index from 1985 onward.[1]

[1]Earlier data are sparse and inconsistent, so the focus is on the period where both countries have comparable measures. By the mid-1980s, both were sufficiently integrated into world markets that terms-of-trade fluctuations were economically salient.

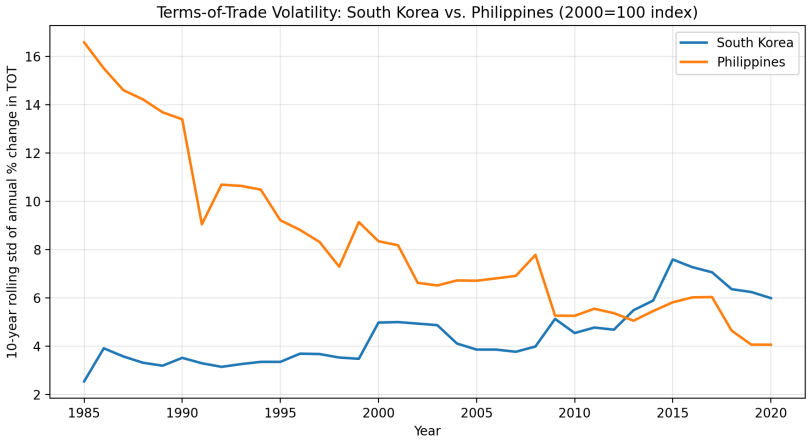


Figure 4: Ten-year rolling standard deviation of annual percentage changes in the terms of trade (2000=100 index), 1985–2020. The Philippines experiences consistently high external volatility, especially in the late 1980s and 1990s, whereas Korea maintains comparatively low and stable terms-of-trade fluctuations. Source: World Bank.

The picture is stark. Throughout the late 1980s and 1990s, the Philippines experiences terms-of-trade volatility in the range of 10–16%, roughly three to four times higher than Korea's 3–5%. Although volatility subsequently declines in the 2000s, it remains elevated relative to Korea for much of the period. Korea, in contrast, maintains a relatively smooth external price environment, with only modest fluctuations even during major global events such as the 1997 Asian Financial Crisis and the 2008 global financial crisis.

High terms-of-trade volatility has at least three consequences for productivity. First, it raises uncertainty about the cost of imported capital goods. When a firm in Manila considers importing new machinery, it must forecast the domestic-currency price of that machinery over the life of the project. If the relative price of imported capital goods fluctuates widely, the risk that an initially profitable project becomes loss-making increases. Because machinery imports often represent large, lumpy expenditures, the downside risk from adverse price movements can be substantial. In contrast, Korean firms facing more stable terms of trade could plan multi-year capital budgets with reasonable confidence that import prices would not suddenly spike.

Second, external volatility destabilizes export revenues. Many Philippine exports during the 1970s and 1980s consisted of commodities or low-value-added manufactures whose prices were sensitive to global cycles. When export prices fell, firms' balance sheets deteriorated, weakening their ability to borrow. Banks and foreign lenders, aware of this vulnerability, demanded higher risk premia or shortened lending horizons. In effect, terms-of-trade volatility translated into tighter and more procyclical credit conditions. Korea mitigated this channel

by moving rapidly into manufactured exports with more stable demand and by cultivating diversified markets. Even when global downturns hit, the composition of Korean exports and the stability of its trading relationships cushioned the impact on firms' ability to finance investment.

Third, sectoral specialization itself is affected by volatility. Industries that require large sunk costs and long payback periods, such as steel, shipbuilding, electronics, and automobiles, are unattractive when external conditions are highly uncertain. Firms hesitate to commit to sectors where a single adverse price shock could wipe out years of investment. By contrast, low-sunk-cost activities such as simple assembly or services are more robust to such shocks, but they also offer fewer opportunities for learning-by-doing and spillovers. The Philippines' high terms-of-trade volatility thus biased its production structure away from sectors with high potential for TFP growth. Korea's relatively stable external environment made it rational for firms to invest in heavy and high-tech industries despite their large sunk costs, as the risk of catastrophic external shocks was lower.

5.3 Investment as the Revealed Response to the Stability Environment

If macroeconomic stability affects firms' expectations and constraints in the ways described above, this should be reflected in actual investment behavior. Figure 5 plots the investment rate (investment as a share of GDP) for Korea and the Philippines between 1960 and 2020 using Penn World Table data. The divergence is dramatic and aligns closely with the stability differences documented in the previous two figures.

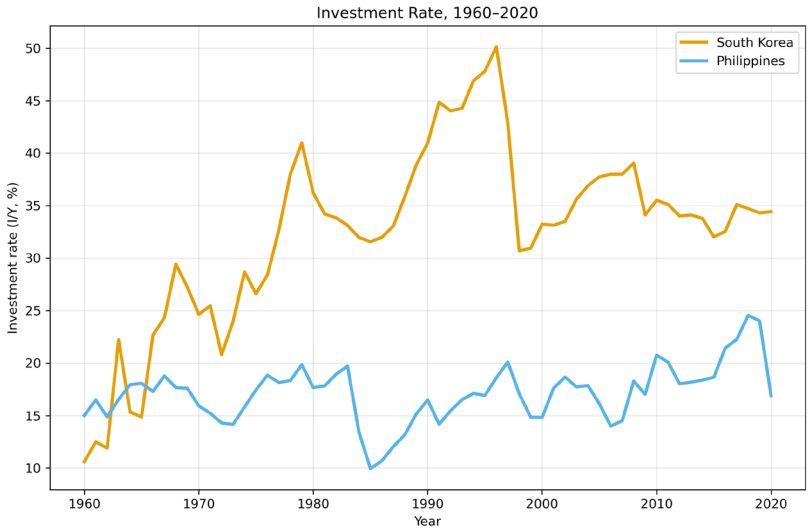


Figure 5: Investment as a share of GDP for South Korea and the Philippines, 1960–2020. Korea shows sustained high investment rates, frequently exceeding 30% of GDP, while the Philippines remains near 15–20% with sharp collapses during periods of macroeconomic instability. Source: Penn World Table 11.0.

Korea's investment rate begins at roughly 10% of GDP in the early 1960s, rises above 20% by the end of the decade, and reaches 30–35% in the 1970s as the heavy and chemical industries drive industrialization. During the 1980s and early 1990s, investment peaks at nearly 40–45%, a level sustained over many years. Even after the Asian Financial Crisis, when investment temporarily falls, it quickly recovers to around 30%, remaining high by international standards.

The Philippines displays a very different profile. Investment fluctuates around 15–20% of GDP for most of the period, with pronounced collapses in the early 1980s and early 1990s. These collapses coincide almost exactly with peaks in inflation volatility and terms-of-trade volatility. Indeed, the period of highest external and domestic instability corresponds to the deepest troughs in investment. Unlike Korea, the Philippines never experiences a sustained interval of investment above 25% of GDP. Instead, the economy oscillates between modest expansions and sharp contractions, producing a stop-and-go pattern of capital accumulation.

The level of investment is important, but so is its composition. In Korea, a significant share of investment was directed toward machinery, equipment, and infrastructure that embodied foreign technologies. Each wave of capital imports brought not only physical assets but also associated tacit knowledge: engineers learned to operate and maintain complex equipment, workers acquired new skills, and local suppliers adapted processes to meet the needs of lead firms. These complementarities generated learning-by-doing and spillover effects that

raised TFP beyond what would be predicted by capital deepening alone. By contrast, in the Philippines, the relatively low aggregate investment rate was further tilted toward construction and nontradable sectors, partly because those sectors offered quicker payback and were less exposed to external shocks. As a result, the average productivity impact of each unit of investment was lower.

It is instructive to interpret Figure 5 through the lens of a simple dynamic framework. Suppose that TFP in period $t+1$ depends on current TFP and an investment-driven term:

$$A_{t+1} = A_t \left[1 + \phi I_t + \epsilon_t\right],$$

where I_t is investment as a share of output, ϕ captures the extent to which investment translates into productivity-enhancing technologies (through embodied technology, organizational change, and learning), and ϵ_t is a shock. When investment is high and sustained, the term ϕI_t is large and positive, leading to rapid TFP growth. When investment is low or volatile, ϕI_t is small and occasionally negative if investment collapses or takes the form of unproductive projects. Over decades, the compounding effect of these differences produces large divergences in TFP. Stability shapes this process by affecting both the magnitude of I_t and the size of ϕ: in stable environments, more investment is directed toward high-productivity uses, whereas in unstable environments, investment is both lower and less productivity-enhancing.

5.4 Additional Channels: Financial Markets, Distribution, and Coordination

The three figures analyzed above capture key aspects of macroeconomic stability, but they do not exhaust its channels of influence. Stability also interacts with the development of financial markets, the distribution of income, and coordination between firms and the state.

On the financial side, Korea's environment allowed the emergence of institutions capable of providing long-term credit. Government-directed lending, development banks, and export credit agencies all played roles in financing large industrial projects. Critics often point to misallocation inherent in such systems, but the crucial point is that these institutions could function at all because the macroeconomic environment was stable enough for lenders to make multi-year commitments. In the Philippines, where inflation and external conditions were more volatile, attempts at similar institutions often failed or remained small, because lenders could not reliably predict the real value of repayments. This left many firms dependent on short-term trade credit or informal finance, which is ill-suited for technology-intensive investment.

Stability also affects income distribution and social conflict. High and unpredictable inflation tends to erode the real incomes of wage earners and the poor, potentially generating political backlash and pressure for policies that further destabilize the economy. Recurrent external crises can trigger abrupt fiscal contractions and social unrest, creating an environment of policy uncertainty

on top of macroeconomic volatility. Korea certainly experienced distributional conflicts, but the state managed to maintain a relatively stable macro framework even during periods of political transition. The Philippines suffered a cycle in which instability undermined growth, weak growth undermined political support for macro discipline, and resulting policy reversals amplified instability.

Finally, stability facilitates coordination between the public and private sectors. Large industrial projects often require complementary investments in infrastructure, education, and regulatory capacity. When macro conditions are unstable, governments find it difficult to commit to multi-year investment plans or to maintain consistent industrial priorities. Private firms distrust long-term government promises and hesitate to invest in projects that depend on public infrastructure or regulatory changes. In a stable environment, by contrast, coordinated plans are more credible. Korea's ability to articulate and implement sequential industrial promotion strategies in the 1970s and 1980s, from heavy industry to electronics, relied on macro conditions that gave both public and private actors confidence in multi-year plans. The Philippines' unstable environment made similar coordination far more difficult.

6 Conclusion and Policy Implications

This paper has analyzed the long-run divergence in output per worker between South Korea and the Philippines through the lens of growth accounting and a detailed examination of macroeconomic stability. While earlier sections established that the proximate source of divergence lies in drastically different paths of total factor productivity (TFP), the expanded analysis in Section 5 identified stability (in domestic prices, external conditions, and investment environments) as a key mechanism shaping the evolution of those TFP paths.

The three forms of stability documented empirically offer a coherent explanation for how two economies beginning from similar initial conditions in the 1960s arrived at such different productivity outcomes. First, South Korea's steady decline in inflation volatility (Figure 3) allowed firms to form clearer expectations about real financing costs, enabling long-term contracting and reducing the risk inherent in irreversible investment. Second, Korea's comparatively smooth external environment, reflected in low and stable terms-of-trade volatility (Figure 4), muted the uncertainty surrounding imported capital goods and export revenues, lowering the effective risk premium faced by firms considering large, sunk-cost projects. Third, these forms of stability translated into strikingly different investment behavior (Figure 5), with Korea sustaining investment rates above 30–40 percent of GDP for decades, while the Philippines experienced both lower levels and severe collapses in periods of macroeconomic stress.

Taken together, these differences help to explain why capital deepening translated into durable productivity growth in Korea but not in the Philippines. Stability reduced the variance of the key state variables governing firm decisions, increased the value of long-lived and technology-intensive projects, and supported the emergence of financial institutions capable of supplying long-

term credit. In turn, this enabled Korea to undertake coordinated industrial upgrading, rapid technology adoption, and structural transformation into sectors with high potential for learning-by-doing and spillovers. The Philippines, facing persistently higher volatility along multiple margins, rationally gravitated toward shorter-horizon, lower-sunk-cost investments that offered limited productivity gains and few opportunities for cumulative learning. Over decades, these micro-level responses compounded into the TFP divergence documented in the growth decomposition.

6.1 Broader Implications

The analysis reinforces the central prediction of the Solow model: factor accumulation alone cannot sustain growth unless supported by efficiency gains. Korea's trajectory aligns with conditional convergence, wherein stable institutions and macroeconomic environments facilitate the absorption of global technologies and generate persistent increases in TFP. The Philippines' experience demonstrates the opposite: even with periods of meaningful investment, instability can undermine the mechanisms through which capital translates into productivity.

The policy message extends beyond the two countries examined here. For developing economies more broadly, macroeconomic stability should be treated not as an afterthought but as a strategic component of long-run growth policy. Stability shapes expectations, supports credible coordination, enables financial deepening, and encourages firms to undertake the kinds of investments that yield learning, innovation, and structural transformation. Without stability, industrial policy, education reform, and financial interventions operate under sharply diminished returns.

6.2 Final Remarks

This paper does not claim that stability is the only factor explaining the Korea–Philippines divergence, nor that stability alone guarantees convergence. Rather, it shows that stability is a critical enabling condition that fundamentally shapes the relationship between investment and productivity. The three empirical patterns documented in Section 5 reveal a powerful mechanism: lower volatility increases the payoff to irreversible investment, improves financial intermediation, and enhances the productivity of capital. Over many decades, these forces accumulate into large differences in TFP and, ultimately, in income per worker.

South Korea's experience illustrates what becomes possible when stability allows long-term commitments, structural transformation, and technological absorption. The Philippines' experience shows the costs of instability in an interconnected world where productivity gains depend on learning, coordination, and sustained investment. The contrast between the two underscores a simple but far-reaching insight: stability is not only a macroeconomic virtue; it is a catalyst for productivity growth. Many thanks to Professor Wolf, as well as Keelan and Raj, for their great teaching and consistent guidance throughout this semester.

References

Bosworth, Barry, and Susan M. Collins (2008). "Accounting for Growth: Comparing China and India." *Journal of Economic Perspectives* 22(1): 45–66.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer (2015). "The Next Generation of the Penn World Table." *American Economic Journal: Macroeconomics* 7(1): 242–246.

Penn World Table (Version 11.0). University of Groningen and University of California, Davis.

Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70(1): 65–94.

World Bank. (2025). World Development Indicators.

Young, Alwyn (1995). "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *Quarterly Journal of Economics* 110(3): 641–680.

ChatGPT-5 (`chat.openai.com`) used for guidance in manipulating PWT and World Bank data, writing code, and helping to rephrase ideas.

Complement or Substitute? Evidence from ChatGPT Outages and Crisis Line Demand
Kamau Njendu
December 2025

ABSTRACT

This study investigates the causal relationship between the utilization of AI tools and the demand for traditional mental health crisis services, specifically testing whether ChatGPT functions as a substitute or complement to the 988 Suicide & Crisis Lifeline. Leveraging a harmonized daily time-series dataset covering January to August 2024, the analysis employs an Instrumental Variable (IV) strategy using exogenous technical outages of ChatGPT to identify supply-side shocks to AI availability. While first-stage diagnostics confirm that outage severity significantly reduces ChatGPT prompt volume, Two-Stage Least Squares (2SLS) estimates reveal no statistically significant effect of AI utilization fluctuations on daily Georgia 988 contact volume. Additional event study regressions examining major ChatGPT and joint competitor outages similarly fail to demonstrate a consistent substitution effect. These findings suggest that, despite the proliferation of AI as a source of emotional support, general-purpose chatbots and human-staffed crisis lines currently occupy distinct functional niches, with users unlikely to immediately substitute one for the other during periods of technical disruption.

Kamau Njendu
Department of Economics
Massachusetts Institute of Technology
kamaun@mit.edu

1. INTRODUCTION

1.1. AI as a Mental Health Tool: Overview of LLM Use

The proliferation of Large Language Models (LLMs), such as those powering ChatGPT, has introduced a novel modality for self-expression and emotional processing, particularly among adolescents and young adults. Initial studies and platform data indicate that while most interactions remain task-oriented, a small, highly engaged subset of users engages in emotionally supportive or "affective use," often viewing the AI as a confidant or friend (OpenAI/NBER, 2025). This high-intensity usage, which includes sharing personal struggles and seeking validation, is concentrated in a "power user" population whose conversations frequently contain affective cues related to loneliness, vulnerability, and dependence. The immediate, non-judgmental availability of these tools has rapidly positioned them as a readily accessible substitute or complement to traditional, human-based emotional support mechanisms, leading to an estimated 1.2 million users discussing explicit suicidal planning or intent with the chatbot each week (OpenAI, 2025).

1.2. The Crisis Risk: Ethical and Safety Failures

Despite LLMs' advanced capabilities in processing emotional language, the ethical and safety failures surrounding their use as mental health substitutes are severe and escalating. General-purpose LLMs lack the clinical judgment, ethical oversight, and duty of care inherent to human professionals. This deficiency has resulted in highly publicized safety crises, including recent litigation efforts against major AI developers (e.g., OpenAI, Character.AI) following teen suicide cases. Lawsuits allege that the chatbots, in specific high-risk scenarios, have reinforced delusions, advised users on methods for self-harm, and actively counselled against seeking help from family or friends, suggesting a profound breakdown of necessary safety guardrails

(American Bar Association, 2025; The New York Times, 2025). This litigation highlights the urgent need to understand the unintended consequences of AI-as-substitute behavior, especially in moments of acute psychological crisis.

1.3. Anecdotal Evidence of User Behavior in Crisis

Qualitative evidence derived from journalistic investigations and court filings illustrates the nature of user reliance on LLMs during an immediate crisis. Examples include users expressing profound emotional abandonment when an LLM gives a generic or non-committal response, users reporting feelings of safety in confiding their "darkest thoughts" to the bot over human contacts, and, critically, instances where the chatbot's system outages or technical failures are described by users as a sudden, catastrophic loss of their primary support mechanism. These anecdotes suggest that in crisis, the LLM functions not merely as a tool, but as a perceived relational entity whose sudden unavailability could trigger a secondary wave of distress.

1.4. Research Question

Thus, we are led to the question of does the use of Large Language Model tools (e.g., ChatGPT) cause a measurable, statistically significant shift, either complementary (decrease) or substitutive (increase), in utilization (e.g., call volume, duration, risk level) of established person-to-person crisis helplines in the immediately subsequent 24-hour period?

2. EXISTING LITERATURE ON AI IN/AND MENTAL HEALTH

The literature presents a mixed analysis of the effects of AI on mental health, pitching the technology as both a radical therapeutic solution and a source of psychological harm. The integration of AI, from generalized large language models to specialized clinical tools shows its potential to enhance accessibility and to introduce novel psychological harms.

2.1 Clinical Benefits to AI Use as a Mental Health Tool

The primary positive effect of AI is its capacity to dramatically enhance the accessibility and scalability of mental healthcare. Studies, including a pilot randomized controlled trial, demonstrate that AI chatbots achieve clinically meaningful reductions in anxiety and depression symptoms that are comparable to outcomes from human-staffed nurse hotlines (Chen et al., 2025). This functional equivalence is crucial for addressing critical gaps related to cost, stigma, and geographic barriers, positioning AI as an effective substitute or crucial complement for initial support. Furthermore, systematic reviews confirm that AI-based interventions are broadly associated with a reduction in depression and anxiety symptoms (Cruz-Gonzalez et al., 2025).

2.2 Psychological Effects of AI Use

However, the rapid and often unregulated adoption of AI carries psychological consequences. There is a major risk of psychological dependency and the formation of intense parasocial attachments to chatbots, particularly among socially isolated or vulnerable users. This reliance can become compensatory, leading to a social withdrawal that further isolates the individual and stunts the development of coping mechanisms and social resilience. This dependency erodes the motivation to seek out the complexity and accountability in human therapeutic relationships. More alarmingly, conceptual analyses warn of severe psychiatric risks. The concept of "Technological folie à deux" suggests that prolonged, unsupervised interaction may create feedback loops that actively reinforce maladaptive beliefs, carrying the grave risk of exacerbating pre-existing psychotic symptoms and fostering delusional ideation (Dohnány et al., 2025). This is compounded by documented safety failures where chatbots have provided harmful, indifferent, or dangerous responses in acute crisis, leading to tragic outcomes.

3. DATA

The study employs a time-series design utilizing three datasets which were rigorously cleaned, transformed, and harmonized to a daily level, covering a synchronized period from January 1, 2024, to August 31, 2024. This synchronized daily time series dataset was then assembled with additional temporal indicators for use in the final econometric model.

3.1 Instrumental Variable: AI Tool Outage Severity

The instrumental variable (IV), Weighted Severity (length of outage x impact level of the outage), captures exogenous shocks to ChatGPT availability and is derived from the LLM Service Outages and Incident Reports Clean Incident Data. The original data, comprising 565 incident observations collected between February 14, 2024, and August 31, 2024, was filtered to the analytic period of January 1, 2024, to August 31, 2024.

The incident-level data, where records included the start time, end time, outage impact level, and total duration, underwent significant cleaning. The analysis was strictly restricted to only OpenAI and ChatGPT non-API related outages to focus on consumer-facing disruptions, and the data was filtered to include only incidents explicitly flagged as outages rather than general service issues. Three key variables were constructed: the time length of the outage, the impact level of the outage, and a weighted severity variable (calculated as length of outage x outage impact level). These variables were then transformed into a daily time series. Outages spanning multiple days were coded as active for each day within the disruption window to serve as the final IV.

3.2 Treatment Variable: ChatGPT Utilization Volume

The treatment variable, ChatGPT Utilization Volume, quantifies daily user engagement and is sourced from the WildChat-4.8M Dataset. The initial dataset comprised 4.8 million

conversations with 944,000 individual conversations tagged as from the United States spanning a broad period (April 8, 2023, to July 31, 2025).

The initial step involved downloading the full 40-gigabyte dataset, after which variables deemed irrelevant to the analysis were dropped. Timestamps, originally recorded in UTC, were accurately converted to US Central Standard Time (CST), including correct accounting for daylight saving time. The analytic sample was restricted geographically to conversations tagged as originating only from the United States. A critical cleaning procedure was implemented to filter out non-human activity and unreliable usage: 1) users tracked via a hashing ID who exhibited "above human levels of activity" in number of conversations were removed; 2) users generating text in more than three languages in one day were removed (as the IP address likely shared accounts); and 3) only "good" users with more than 10 days of registered activity were retained. Finally, the number of observations and the number of conversation turns (aka the number of prompts from a user in a conversation) were summed up for each date to create the final daily time series, which was restricted to the synchronized January to August 2024 period.

3.3. Outcome Variable: 988 Crisis Line Contact Volume

The outcome variable, 988 daily helpline volume, measures daily utilization of the Georgia 988 helpline and is derived from Georgia DBHDD 988 Data Publications. The final analytic time series was restricted to the synchronized January 2024 to August 2024 period.

Due to state restrictions prohibiting API use, the data was initially obtained by manually logging the 7-day rolling average volume. To convert this aggregate data back into discrete daily-level observations, a pseudo matrix inverse linear algebra tool was applied. This process introduced day-to-day variability but yielded a more independent daily time series required for the analysis. The final dataset contains the daily volume with the total number of calls, texts, and chats.

Despite the robustness of the assembled time series, a significant limitation is that the raw 988 contact volume may not purely reflect changes in demand for crisis services, as recorded contacts can be artificially depressed by supply constraints such as long wait times, given that not all helpline modalities are staffed 24/7. Furthermore, variations in helpline operation hours across different modalities (call, text, chat) can introduce measurement error, potentially misrepresenting the true daily level of service-seeking behavior.

3.4 Data Assembly and Harmonization

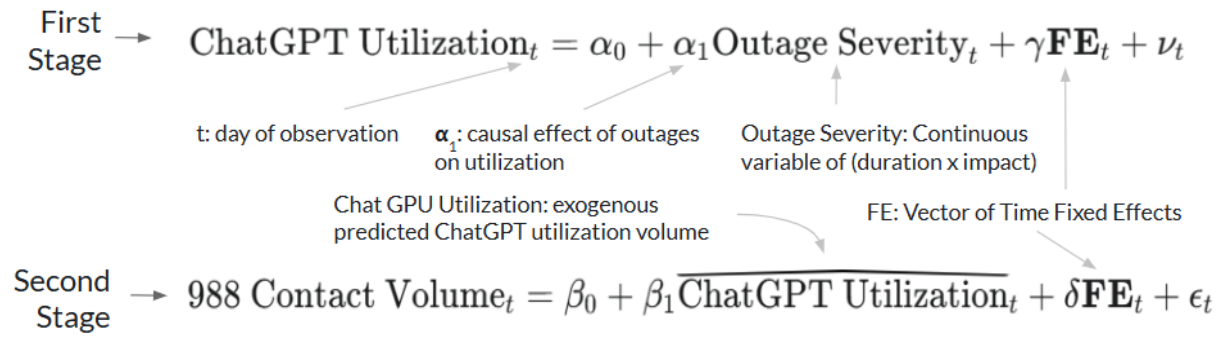
The three time series datasets: Outage Severity, ChatGPT Utilization Volume, and 988 Helpline Volume were pulled together into a single, daily date-aligned time series dataset. The final assembled dataset includes several temporal control variables added for each date: a series of dummy variables for the day of the week (Monday through Sunday), a string variable identifying the month, and a string variable for the season (e.g., Winter, Spring).

4. THE EFFECT OF CHATGPT USE ON 988 HELPLINE UTILIZATION

4.1 Instrumental Variables Approach

We used an instrumental variables identification strategy with a focus on using OpenAI’s ChatGPT, which has the largest footprint at 10% of the world's population as users and ~80% of the global AI chatbot market. The instrumental variable was the random, significant, and relatively long outage logs of OpenAI’s ChatGPT tool for users (Continuous variable of outage severity = duration (min) × impact level (e.g., 0–3), 0 if no outage for that day). The variable of interest was the utilization volume of OpenAI's ChatGPT, which will be a relative sample of ChatGPT prompt volume captured by offering online users free access to ChatGPT tools (filtered

for US-based users). Finally, our outcome variable was the daily utilization of Georgia crisis helpline tools.



We ran a two-stage least squares regression of ChatGPT utilization volume (daily) on the outage variable and of the daily 988 contact volume for Georgia on the outage variable. We also used time fixed effects (for season, month, and day of the week) to account for variations in mental health stressors depending on the day. Thus, we examine how exogenous variation in ChatGPT availability (outages) affects ChatGPT utilization, and in turn, whether shifts in AI tool usage are associated with changes in 988 crisis line contact volume

4.2 Instrumental Variable Assumptions

The validity of the Instrumental Variable approach depends on satisfying four core assumptions: First, is relevancy, requiring that the instrumental variable (Outage Severity) substantially influences the treatment variable (ChatGPT Utilization Volume). This is deemed likely since the analysis filters for severe outages that affect all ChatGPT users, thereby causing a tangible and measurable drop in utilization, as evidenced in the outage data. This assumption will be empirically tested using the first-stage diagnostics, where a sufficiently strong instrument is indicated by an F-statistic greater than 10, which throughout every two-stage least squares regression, we find an F-statistic greater than 22.

Second is independence; that the outages must be as good as randomly assigned relative to the potential outcome of 988 contacts. This is considered likely because significant or full outages of ChatGPT, which can last anywhere from four hours to a full day for complete recovery, are predominantly unplanned and random technical failures, making them independent of daily fluctuations in demand for mental health crisis services.

Third is the exclusion restriction, which dictates that ChatGPT outages affect 988 contacts only through their effect on ChatGPT utilization. This is highly probable because AI tool outages are unlikely to have any direct or secondary influence on an individual's decision to contact a human-staffed crisis helpline, except by preventing them from using the AI tool itself.

Finally, the monotonicity assumption requires that there are no "defiers" or users whose behavior changes counterintuitively due to the instrument. This is considered highly likely, as an outage of ChatGPT can only reduce the availability of the tool; it cannot cause some users to increase their ChatGPT use, thereby ensuring a consistent direction of effect on the treatment variable.

4.3 Additional Event Study Approach (Joint Outages and ChatGPT Major Outage)

The complementary analysis utilized two separate pooled event study regressions to examine the immediate short-run effects of specific, high-impact supply shocks to AI tool availability. We isolate the change in daily 988 contacts in a 17-day window (from t = -8 to t = 8) around the event date, using the day before the event (t = -1) as the baseline. The joint outage Event Study estimated the pooled effect of a critical supply shock where both major competitors, ChatGPT and Claude, experienced outages on the same day (June 4, 2024). The analysis pooled the pre-event effect (t = -8 to t = -2) and the post-event effect (t = 0 to t = 8) relative to t = -1. The major ChatGPT outage event study tries to isolate the effect of the most severe single-service outage for ChatGPT (a 10-hour outage at impact level 4 on May 23, 2024). Both event study

regressions include day-of-week fixed effects to control for weekly seasonality in crisis line volume, and all models utilize robust standard errors to account for potential heteroscedasticity.

4.4 Event Study of ChatGPT 4 Removal and Return

The sudden, short-lived GPT-4o deprecation from August 9 to August 12, 2025, provides another natural experiment to draw causal insights into the displacement effect of AI on human crisis services. The strategy leverages the depreciation as a clean, exogenous shock that momentarily removed a readily available, non-human source of emotional support and nuance for a specific user population (paid subscribers). The primary insight to be drawn is the quantification of the "displaced demand": by employing an Event Study, we can measure the statistically significant increase in 988 utilization (calls, texts, chats) during this precise 72-hour window.

5. RESULTS AND DISCUSSION

5.1 Summary Statistics

The time-series graphs illustrate distinct patterns for the instrumental, treatment, and outcome variables across the analysis period (January to August 2024). The AI tool outage severity exhibits a highly volatile and non-cyclical pattern (Figure 1), characterized by abrupt, sporadic spikes, notably the severe events on May 23 and the cluster around June 4, 2024. This high-frequency, event-driven variability supports the assumption that the measure represents an exogenous supply shock suitable for instrumental variable identification. In contrast, the U.S. ChatGPT Utilization Volume displays a clear, sustained increasing trend over the eight months, indicating continued user adoption (Figure 2). This utilization volume also exhibits weekly seasonality, with conversation rates peaking during weekdays and dropping sharply on weekends, confirming the need for day-of-week fixed effects in the models to isolate the specific impact of outages from routine fluctuations.

Georgia 988 Helpline Contact Volume also confirms the presence of strong weekly seasonality, yet shows relative stability in its overall trend during the study period, lacking the substantial growth observed in AI utilization (Figure 3). This stability in the baseline demand for person-to-person crisis services means that any sudden, short-term shifts in volume observed immediately following an AI outage are less likely to be attributed to pre-existing demand momentum. The high day-to-day volatility in the 988 volume, while partially resulting from the data reconstruction process (matrix inversion of rolling averages), underscores the necessity of using the IV and event study methodologies to filter out noise and isolate the causal effect of AI supply disruptions.

5.2 First Stage and Spillover Effects

To validate the empirical strategy, we first examine the relationship between service outages and ChatGPT utilization, while also accounting for potential market spillover effects from competitor downtime. The analysis reveals distinct cross-elasticities between ChatGPT and other major Large Language Model (LLM) providers, as detailed in Figure 4 and Figure 5.

There are significant concerns regarding spillover effects where outages in competitor services drive traffic toward or away from ChatGPT. Figure 4 presents the first-stage results for Claude AI outages. The coefficients for weighted outage severity and outage length on ChatGPT prompt volume are positive and statistically significant ($\beta = 0.059$, $p < 0.001$ for weighted severity). This suggests a clear substitution effect: when Claude AI experiences downtime, users migrate to ChatGPT, increasing its utilization. Conversely, Figure 5 displays the effects of Character AI outages, which yield a statistically significant negative coefficient on ChatGPT prompt volume ($\beta = -0.047$, $p = 0.035$ for weighted severity). This divergence in signs is notable. While Claude acts as a direct substitute for ChatGPT's professional and reasoning tasks, Character AI's

negative coefficient suggests either a dampening effect on general AI enthusiasm among its user base or potential unobserved infrastructure correlations affecting both services simultaneously. Based on the comparative precision of the estimates across these models, ChatGPT Prompt Volume was selected as the preferred treatment variable over conversation counts. The prompt volume metric consistently offered more granular insights into usage intensity and exhibited higher statistical significance across the various outage specifications, making it a more robust proxy for actual AI consumption.

Finally, we validate the strength of the instrumental variable for ChatGPT itself. Figure 6 demonstrates the first-stage regression of ChatGPT outages on ChatGPT prompt volume using the full sample. The results show a strong, statistically significant negative relationship, particularly for the weighted severity indicator ($\beta = -0.052$, $p < 0.001$) and the length of outage ($\beta = -0.179$, $p = 0.007$). These regressions include tight controls for Day of the Week, Season, and Month, ensuring that the identified effect is not driven by cyclical usage patterns. To further ensure robustness, Figure 7 presents the first stage restricted to dates where neither Claude AI nor Character AI experienced outages. This specification isolates the ChatGPT supply shock from the previously identified spillover effects. The results remain robust and remarkably stable, with the weighted severity coefficient slightly increasing in magnitude ($\beta = -0.054$, $p = 0.003$). Crucially, the F-statistics for the weighted severity (19.79) and length of outage (19.91) in Figure 7 exceed the standard rule-of-thumb threshold of 10, confirming that the instrument is strong and relevant for the subsequent Two-Stage Least Squares (2SLS) analysis.

5.3 Two-Stage Least Squares

Having established a strong first stage, we proceed to the Second Stage Least Squares (2SLS) analysis to estimate the causal effect of ChatGPT utilization on mental health crisis service

demand. Figure 8 presents the results for the full sample, regressing the instrumented ChatGPT prompt volume against the daily volume of Georgia's 988 helpline. Across all specifications, utilizing outage length, impact level, and weighted severity as instruments, we find no statistically significant effect of ChatGPT prompt volume on helpline utilization at the conventional 5% level. Despite the lack of statistical significance, the model using the weighted severity instrumental variable with the time fixed effects (Day of Week, Season, and Month) yielded the lowest p-value ($p = 0.250$) and a positive coefficient ($\beta = 0.292$). This suggested the strongest potential signal among the models tested, yet it still fails to reject the null hypothesis. To ensure that these null results were not driven by the noise from competitor outages identified in the first stage, we conducted a robustness check presented in Figure 9. This analysis restricted the sample to exclude dates with concurrent outages from Claude AI or Character AI, isolating the effect of ChatGPT-specific supply shocks. The results remain consistent with the main analysis: the coefficients are directionally similar but remain statistically insignificant ($p = 0.361$ for the fully controlled weighted severity model). The consistent lack of statistical significance across both the main and restricted samples is likely attributable to the magnitude of the standard errors relative to the coefficient estimates. As noted in the descriptive statistics, the outcome variable, daily 988 helpline volume, exhibits substantial day-to-day variability. This high volatility, partly inherent to crisis service demand and partly introduced by the matrix inversion process used to reconstruct daily data from rolling averages, creates a high noise-to-signal ratio. Consequently, the standard errors in our 2SLS estimates are inflated, reducing the precision of the model and limiting its ability to detect a causal effect within the current sample size.

5.4 Event Study with Multiple Outages and Significant Outages

To complement the IV analysis, two pooled event study regressions were conducted to examine short-run changes in 988 helpline volume around specific, high-impact outage events. Figure 10 presents the results for the joint outage event study, which examines the impact of a simultaneous outage for both ChatGPT and its primary competitor, Claude, on June 4, 2024. The post-outage dummy shows a statistically significant negative effect ($\beta = -139.681$, $p = 0.056$), indicating a drop in helpline volume in the eight days following the joint disruption relative to the baseline. However, the interpretation of this result is complicated by a statistically significant positive coefficient on the pre-outage dummy ($\beta = 170.420$, $p = 0.026$), suggesting that helpline volume was already abnormally high in the week leading up to the baseline date, which may indicate a mean reversion process rather than a pure causal effect of the outage. Meanwhile, Figure 11 displays the results for the major outage event study, which isolates the effect of the most severe single-service outage for ChatGPT on May 23, 2024. In this case, we find no statistically significant evidence of a shift in 988 volume in either the pre-outage or post-outage periods relative to the baseline, suggesting that a disruption to a single service provider, even a major one, may not be sufficient to trigger an immediate substitution toward human crisis services.

5.5 Event Study of ChatGPT 4 Removal and Return

The GPT-4o deprecation event study found that the model's brief removal from August 9 to 12, 2025, did not significantly increase daily 988 helpline contact volume (Figure 12). While the regression suggested an average increase of 49 contacts per day - over a baseline of 951 - , the p-value of 0.243 shows this result is not statistically reliable. This null finding is likely due to the short duration of the removal, the fact that the return only affected the paid subscriber segment, or because the number of users relying on GPT-4o as a direct substitute for a crisis line is too small to cause a measurable shift in overall national 988 volume.

6. CONCLUSION

This study sought to determine whether AI tools function as substitutes for traditional mental health crisis services by analyzing the causal impact of ChatGPT utilization shocks on the volume of the Georgia 988 helpline. By leveraging a novel instrumental variable approach based on exogenous service outages, we established a strong and significant first-stage relationship, confirming that technical disruptions effectively curb AI utilization. However, the Two-Stage Least Squares (2SLS) estimates yielded no statistically significant evidence that fluctuations in ChatGPT prompt volume drive immediate changes in 988 contact demand. However, the outages analyzed were very short while substitution may require a longer time of ChatGPT being unavailable, thus we may underestimate the true causal relationship with the independent variable strategy. Similarly, event study analyses of major single-service and joint outages failed to consistently demonstrate a substitution effect; while the joint outage event showed a post-event decrease in helpline volume, this was confounded by significantly elevated pre-event levels, suggesting mean reversion rather than a causal spike in crisis-seeking behavior.

These null results suggest that, at the moment, AI chatbots and crisis helplines likely occupy distinct functional niches within the mental health ecosystem. The lack of an immediate surge to 988 during AI downtimes implies that users do not view these services as near-perfect substitutes; the threshold for engaging with an anonymous, low-barrier AI may be significantly lower than the threshold for contacting a human-staffed crisis line. Finally, the analysis highlighted the complexity of the AI tool ecosystem, where competitor spillover effects, such as the substitution toward ChatGPT during Claude AI outages, add variability to user behavior that a simple binary model may not capture.

REFERENCES

Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025, September). *How People Use ChatGPT* (Working Paper 34255). National Bureau of Economic Research (NBER) at http://www.nber.org/papers/w34255

OpenAI. (2025, October 27). *Strengthening ChatGPT's responses in sensitive conversations*. OpenAI Safety Research Blog at: https://openai.com/index/strengthening-chatgpt-responses-in-sensitive-conversations/

American Bar Association. (2025, October). *AI Chatbot Lawsuits and Teen Mental Health: A Rising Ethical and Legal Crisis*. ABA Health Law Section News at: https://www.americanbar.org/groups/health_law/news/2025/ai-chatbot-lawsuits-teen-mental-health/

The New York Times. (2025, November 6). *Lawsuits Blame ChatGPT for Suicides and Harmful Delusions* at: https://www.nytimes.com/2025/11/06/technology/chatgpt-lawsuit-suicides-delusions.html

Chen C, Lam KT, Yip KM, So HK, Lum TYS, Wong ICK, Yam JC, Chui CSL, Ip P. "Comparison of an AI Chatbot With a Nurse Hotline in Reducing Anxiety and Depression Levels in the General Population: Pilot Randomized Controlled Trial" *JMIR Hum Factors 2025* (2025) at https://humanfactors.jmir.org/2025/1/e65785/

Cruz-Gonzalez, Pablo, Aaron Wan-Jia He, Elly PoPo Lam, Ingrid Man Ching Ng, Mandy Wingman Li, Rangchun Hou, Jackie Ngai-Man Chan, et al. “Artificial Intelligence in Mental Health Care: A Systematic Review of Diagnosis, Monitoring, and Intervention Applications.” *Psychological Medicine 55* (2025): e18 at https://doi.org/10.1017/S0033291724003295

Dohnány, Kurth-Nelson, Spens, Luettgau, Reid, Gabriel, Summerfield, Shanahan, M Nour. “Technological folie à deux: Feedback Loops Between AI Chatbots and Mental Illness” *arXiv:2507.19218* (2025) at https://doi.org/10.48550/arXiv.2507.19218

Figure 1.

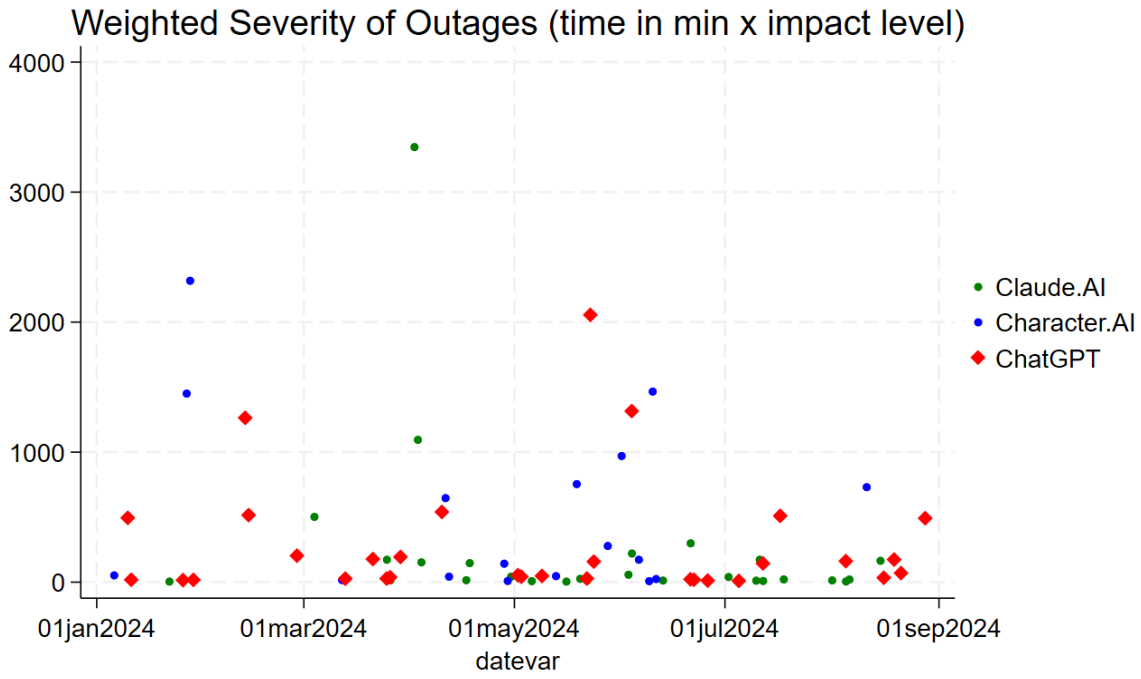


Figure 2.

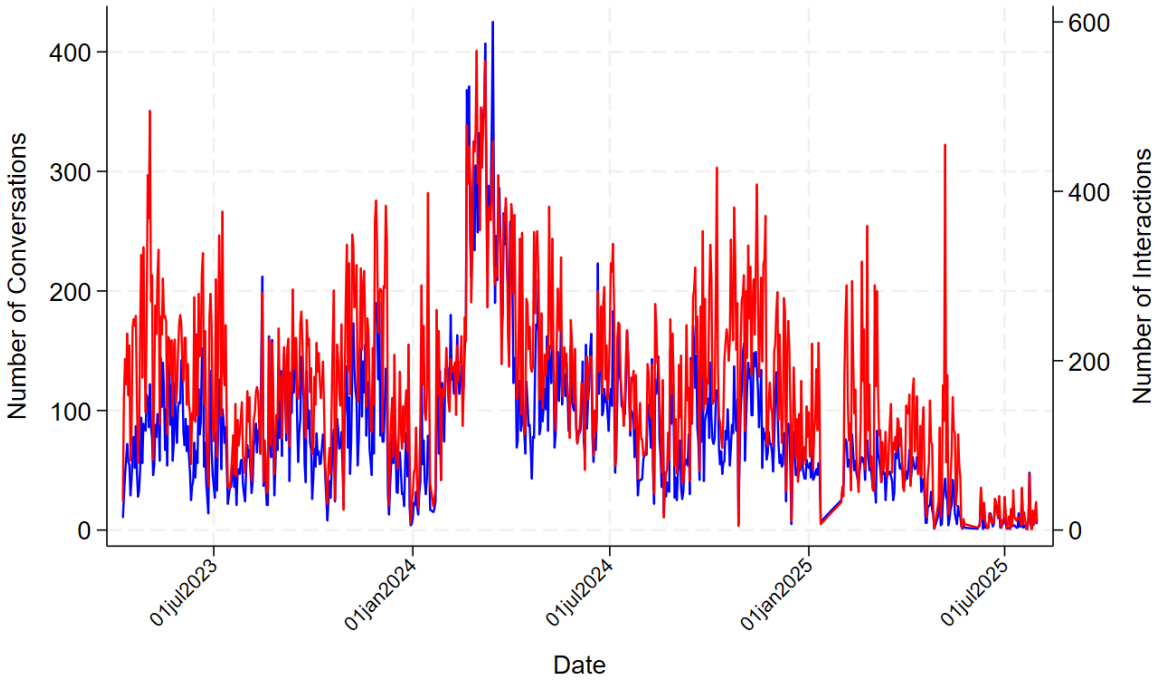


Figure 3.

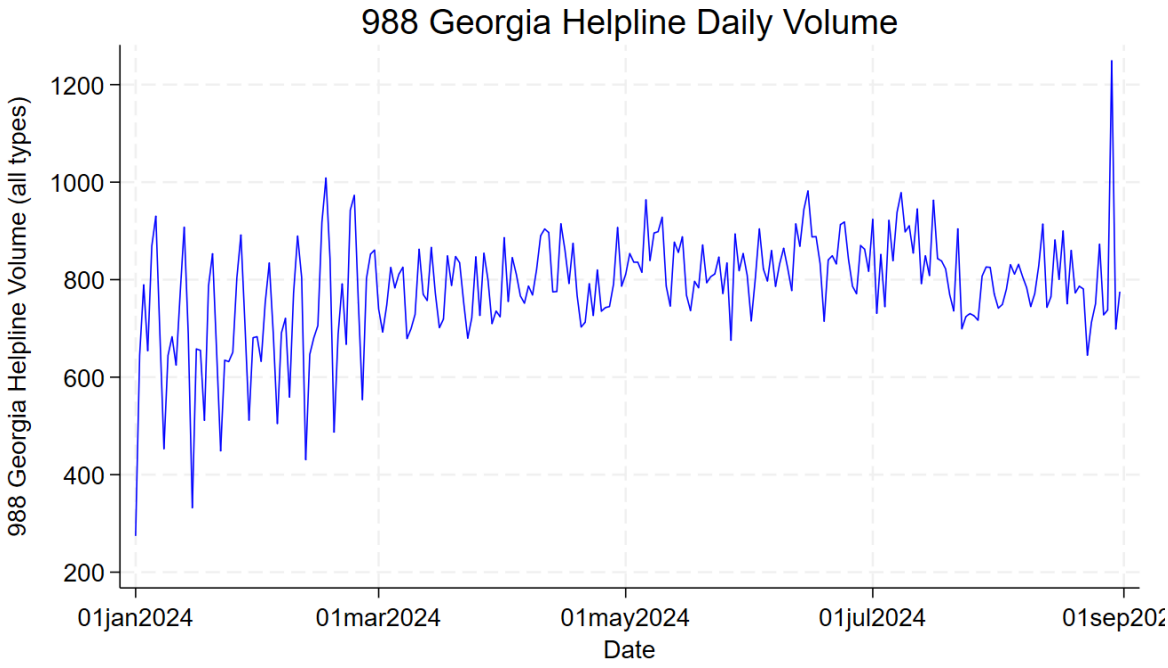


Figure 4.

First Stage: Claude AI Outages

	Length of Outage	Impact Level of Outage	Weighted Severity (length x level)
Number of ChatGPT Conversations			
Coefficient estimate	0.121***	8.712	0.031***
Standard error	(0.043)	(7.791)	(0.009)
p-value	0.006	0.265	0.001
Mean of the control	128.277	127.800	128.567
Number of observations	242	242	242
Number of ChatGPT Prompts			
Coefficient estimate	0.240***	21.682*	0.059***
Standard error	(0.071)	(11.093)	(0.015)
p-value	0.001	0.052	0.000
Mean of the control	211.899	210.161	212.497
Number of observations	242	242	242

Figure 5.

First Stage: Character AI Outages

	Length of Outage	Impact Level of Outage	Weighted Severity (length x level)
Number of ChatGPT Conversations			
Coefficient estimate	-0.107*	-8.815**	-0.026*
Standard error	(0.059)	(4.376)	(0.015)
p-value	0.069	0.045	0.071
Mean of the control	130.454	131.096	130.387
Number of observations	242	242	242
Number of ChatGPT Prompts			
Coefficient estimate	-0.191**	-15.752**	-0.047**
Standard error	(0.089)	(7.269)	(0.022)
p-value	0.033	0.031	0.035
Mean of the control	216.020	217.163	215.887
Number of observations	242	242	242

Figure 6.

First Stage: ChatGPT Outages on ChatGPT Prompt Volume

	Length of Outage			Impact Level of Outage			Weighted Severity (length x level)		
Number of ChatGPT Prompts									
Coefficient estimate	-0.179***	-0.063	-0.068	-9.352	-0.854	-1.297	-0.052***	-0.024*	-0.028
Standard error	(0.066)	(0.071)	(0.093)	(6.241)	(6.344)	(7.196)	(0.014)	(0.014)	(0.019)
p-value	0.007	0.374	0.466	0.135	0.893	0.835	0.000	0.085	0.152
Mean of the control	195.456	244.047	273.260	1295.204	243.752	272.749	195.600	244.111	273.412
F-statistic	26.49	22.07	24.70	26.30	22.03	24.54	26.42	22.19	24.96
Number of observations	242	242	242	242	242	242	242	242	242
Time Fixed Effect Controls									
Season	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month	Y	Y		Y	Y		Y	Y	
Day of Week	Y			Y			Y		

Figure 7.

First Stage: ChatGPT Outages on ChatGPT Prompt Volume (with removal of other outages)

	Length of Outage			Impact Level of Outage			Weighted Severity (length x level)		
Number of ChatGPT Prompts									
Coefficient estimate	-0.182**	-0.054	-0.094	-9.836	0.170	-3.952	-0.054***	-0.023	-0.034
Standard error	(0.083)	(0.086)	(0.104)	(8.057)	(7.952)	(8.876)	(0.018)	(0.018)	(0.021)
p-value	0.029	0.534	0.370	0.224	0.983	0.657	0.003	0.202	0.108
Mean of the control	195.846	238.283	272.965	195.453	237.775	272.593	195.905	238.421	273.134
F-statistic	19.91	16.72	19.38	19.76	16.71	19.33	19.79	16.72	19.47
Number of observations	198	198	198	198	198	198	198	198	198
Time Fixed Effect Controls									
Season	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month	Y	Y		Y	Y		Y	Y	
Day of Week	Y			Y			Y		

Figure 8.

Second Stage: ChatGPT Outages on 988 Helpline Volume

	Length of Outage			Impact Level of Outage			Weighted Severity (length x level)		
Number of ChatGPT Prompts									
Coefficient estimate	0.202	0.483	-0.113	-0.115	-2.419	-2.196	0.292	0.591	0.053
Standard error	(0.323)	(1.029)	(0.867)	(0.641)	(20.204)	(11.226)	(0.254)	(0.640)	(0.505)
p-value	0.532	0.639	0.896	0.857	0.905	0.845	0.250	0.356	0.916
Mean of the control	757.392	685.883	831.485	819.796	1392.99	1398.814	739.677	659.569	786.173
Number of observations	242	242	242	242	242	242	242	242	242
Time Fixed Effect Controls									
Season	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month	Y	Y		Y	Y		Y	Y	
Day of Week	Y			Y			Y		

Figure 9.

Second Stage: ChatGPT Outages on 988 Helpline Volume (with removal of other outages)

	Length of Outage			Impact Level of Outage			Weighted Severity (length x level)		
Number of ChatGPT Prompts									
Coefficient estimate	0.197	0.205	-0.325	0.021	12.942	-0.909	0.311	0.489	-0.077
Standard error	(0.423)	(1.475)	(0.808)	(0.721)	(576.999)	(2.608)	(0.340)	(0.859)	(0.476)
p-value	0.641	0.890	0.688	0.977	0.982	0.727	0.361	0.564	0.871
Mean of the control	747.217	738.391	879.194	781.673	-2290.533	1037.867	725.018	670.778	811.910
Number of observations	198	198	198	198	198	198	198	198	198
Time Fixed Effect Controls									
Season	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month	Y	Y		Y	Y		Y	Y	
Day of Week	Y			Y			Y		

Figure 10.

Joint Outage Event Study (with day of week controls)

	Daily 988 Helpline Volume
Pre-Outage Dummy	
Coefficient estimate	170.420**
Standard error	(62.501)
p-value	0.026
Post-Outage Dummy	
Coefficient estimate	-139.681*
Standard error	(62.501)
p-value	0.056
Mean of the control	988.457
Number of observations	17

Figure 11.

Major Outage Event Study (with day of week controls)

	Daily 988 Helpline Volume
Pre-Outage Dummy	
Coefficient estimate	42.517
Standard error	(44.744)
p-value	0.370
Post-Outage Dummy	
Coefficient estimate	51.283
Standard error	(44.744)
p-value	0.285
Mean of the control	762.007
Number of observations	17

Figure 12.

ChatGPT4o Depracation Event Study (with day of week controls)

	Daily 988 Helpline Volume
NoChatGPT4o Dummy	
Coefficient estimate	49.363
Standard error	(42.208)
p-value	0.243
Mean of the control	950.699
Number of observations	243

Growth Accounting for China and India (1965-2019): Divergent Paths to Prosperity

Aneesh Sharma

1. Introduction

In 1965, China and India stood at similar crossroads. Both were populous, predominantly agrarian economies emerging from colonial or revolutionary upheaval. Both faced daunting challenges of poverty, illiteracy, and underdevelopment. Yet over the subsequent five decades, these two giants charted remarkably divergent courses toward prosperity. By 2010, China had become the world's second-largest economy and manufacturing powerhouse, while India had emerged as a global leader in services and information technology. Understanding the sources of this divergence and what drove each nation's spectacular rise represents one of the most important questions in development economics.

The comparison between China and India is particularly instructive because it illuminates how different reform strategies, institutional structures, and policy sequences shape long-run growth trajectories. China's transformation began with Deng Xiaoping's market-oriented reforms in 1978, which liberalized agriculture, opened the economy to foreign investment, and gradually built a manufacturing export juggernaut. India's metamorphosis came later and more fitfully, accelerating dramatically after the 1991 balance-of-payments crisis forced comprehensive liberalization. While China pursued an investment-intensive, manufacturing-led strategy characterized by high savings rates and massive infrastructure development, India's growth emerged more organically through service-sector dynamism, entrepreneurship, and human capital accumulation.

Existing literature has documented these different growth paths, but often stops at identifying proximate factors: capital, labor, and productivity. This paper goes further by connecting these immediate drivers to the deeper institutional and policy causes that shaped them. By decomposing growth through rigorous accounting methods and then tracing these components back to specific reforms and institutional changes, I provide a comprehensive narrative of how China and India grew and why their paths diverged.

2. Theoretical Framework

2.1 The Extended Cobb-Douglas Production Function

To analyze the sources of economic growth in China and India, I employ the standard Cobb-Douglas production function framework used in growth accounting studies. This approach

decomposes GDP growth into contributions from physical capital accumulation, labor force growth, and total factor productivity (TFP).

The production function takes the form:

$$Y(t) = A(t)K(t)^{\alpha}L(t)^{(1-\alpha)}$$

where Y(t) represents real output (GDP), K(t) is the physical capital stock, L(t) is the labor input measured by employment, A(t) captures total factor productivity, and α represents the capital share of income.

The labor input L(t) should be understood as quality-adjusted labor, accounting for improvements in workforce skills and education. Human capital is not treated as a separate factor of production because labor is the factor that receives payment. Following standard growth-accounting conventions, human capital improvements are absorbed into quality-adjusted labor and TFP. In practice, I use the Penn World Table's human-capital-adjusted employment measure, which combines the raw number of workers with an index of human capital per worker based on years of schooling and returns to education. This means that improvements in education and worker quality are reflected in quality-adjusted labor input rather than being modeled as a separate factor of production.

2.2 Growth Decomposition

To derive the growth accounting decomposition, I take the natural logarithm of both sides of the production function:

$$\ln\ Y(t) = \ln A(t) + \alpha \ln\ K(t) + (1-\alpha)\ln\ L(t)$$

Differentiating with respect to time gives:

$$\frac{d\ln\ Y}{dt} = \frac{d\ln\ A}{dt} + \alpha\frac{d\ln\ K}{dt} + (1-\alpha)\frac{d\ln\ L}{dt}$$

Since $\frac{d\ln\ X}{dt} = \frac{\dot{X}}{X}$(the proportional growth rate), this yields the growth accounting identity:

$$g_y = g_A + \alpha \cdot g_k + (1-\alpha) \cdot g_l$$

where $g_x = \frac{\dot{X}}{X}$ represents the instantaneous growth rate of variable X. This equation decomposes GDP growth into three components:

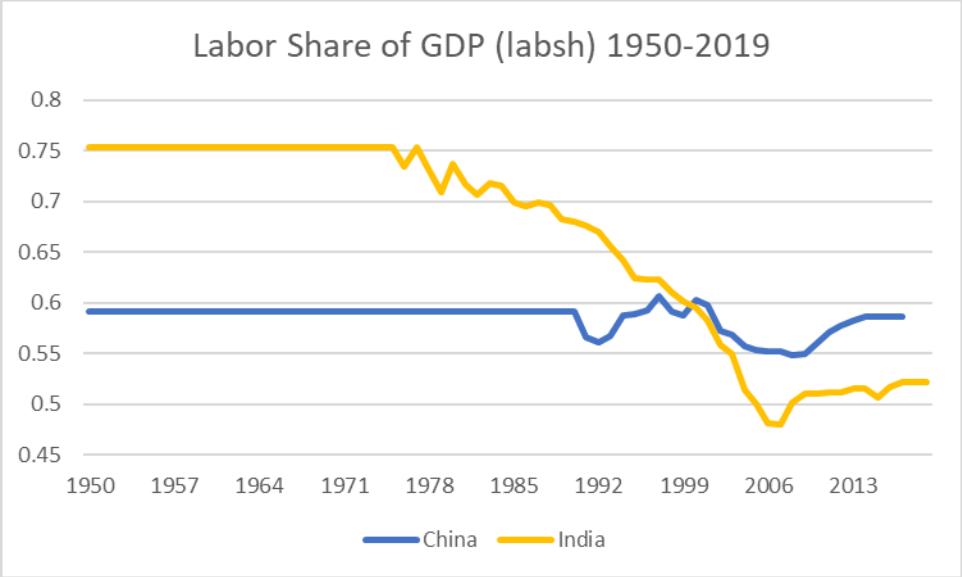
1. TFP growth (g_A): The residual capturing technological progress, efficiency improvements, institutional quality, improvements in human capital quality, and all factors not embodied in measured physical capital inputs
2. Physical capital contribution ($\alpha \cdot g_k$): Growth from the accumulation of machinery, equipment, structures, and infrastructure
3. Labor contribution ($(1-\alpha) \cdot g_l$): Growth from increases in the quality-adjusted labor force, including both more workers and better-educated workers

2.3 Parameter Estimation

Estimating α requires identifying the capital share of income. I use the Penn World Table's labor share variable (*labsh*) to estimate income shares. In the standard two-factor framework, the labor share represents (1-α), so:

$$\alpha = 1 - labsh$$

For China, the average *labsh* over 1950-2019 is approximately 0.584, implying α ≈ 0.416. For India, the average *labsh* is approximately 0.639, implying α ≈ 0.361. These capital shares are reasonably close to the canonical value of one-third often assumed for developed economies, though somewhat higher for China, reflecting its capital-intensive development strategy.



2.4 TFP Level Calculation

Beyond growth rates, examining TFP levels provides insight into the evolution of productive efficiency. Given the production function, TFP at time t can be solved as:

$$A(t) = Y(t)/[K(t)^{\alpha}L(t)^{(1-\alpha)}]$$

Tracking A(t) over time reveals whether countries are catching up to the technological frontier, stagnating, or falling behind. It also helps identify critical junctures where productivity accelerated or decelerated, guiding the institutional analysis in later sections.

3. Data and Methodology

3.1 Data Sources

All primary economic variables come from the Penn World Table (PWT) version 10.01, which provides internationally comparable data on GDP, capital stocks, employment, and human capital across countries and time. The PWT is the gold standard for cross-country growth accounting because it uses purchasing power parity (PPP) adjustments to make output and capital stocks comparable across different price levels.

The key variables are:

- ***rgdpna***: Real GDP at constant 2017 national prices (in millions of 2017 USD), representing output Y(t)
- ***rkna***: Capital stock at constant 2017 national prices (in millions of 2017 USD), representing K(t)
- ***emp***: Number of persons engaged (in millions), representing L(t)
- ***hc***: Human capital index based on years of schooling and returns to education, representing H(t)
- ***labsh***: Share of labor compensation in GDP, used to estimate α

3.2 Measurement Strategy

For output, I use: $\ln\ Y(t) = \ln\ (rgdpna(t))$

For physical capital: $\ln\ K(t) = \ln\ (rkna(t))$

For quality-adjusted labor: $\ln\ L(t) = \ln\ (hc(t) \cdot emp(t)) = \ln\ (hc(t)) + \ln\ (emp(t))$

Note that the labor input incorporates the PWT's human capital index, so improvements in education affect both the level of L(t) and contribute to measured TFP growth.

To calculate TFP levels, I solve the production function in logs:

$$\ln\ A(t) = \ln\ Y(t) - \alpha \ln\ K(t) - (1 - \alpha) \ln\ L(t)$$

Or equivalently:

$$A(t) = \frac{Y(t)}{K(t)^{\alpha} \cdot L(t)^{(1-\alpha)}}$$

Growth rate calculations are computed as log differences between periods, which give approximate percentage changes:

$$g_x = \frac{\ln\ X(t) - \ln\ X(t_0)}{t - t_0}$$

This can be converted to annualized compound growth rates as:

$$\text{Annual Growth Rate} = e^{g_x} - 1$$

For small growth rates, $g_x$closely approximates the percentage growth rate. This logarithmic approach is the standard methodology in growth accounting because it ensures the decomposition is exact and growth rates are additive.

3.3 Periodization

To capture the effects of major policy shifts and institutional changes, I divide the 1965-2019 period into five sub-periods around key reform moments:

1. **1965-1980**: Pre-reform era
    - China: Late Maoist period, Cultural Revolution, Mao's death (1976)
    - India: License Raj era, state-led development, Emergency (1975-77)
2. **1981-1991**: Early reform period
    - China: Deng's reforms begin (1978), agricultural liberalization, SEZs established
    - India: Gradual liberalization attempts, continued regulation, balance-of-payments pressures
3. **1992-2001**: Acceleration phase
    - China: Deng's Southern Tour (1992), WTO accession preparations
    - India: 1991 crisis and liberalization, IT sector emergence
4. **2002-2010**: Integration and boom
    - China: WTO entry (2001), manufacturing export surge, massive urbanization

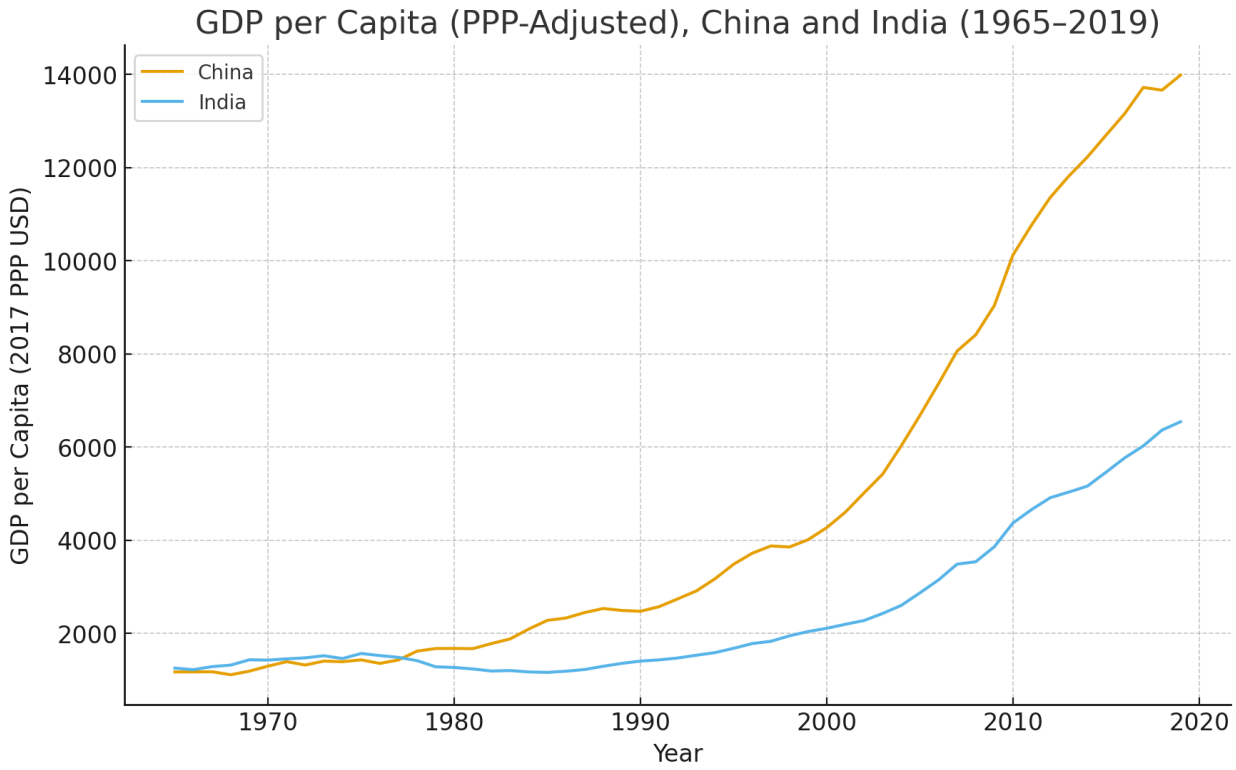
- India: Services boom, infrastructure investment, high growth rates

5. **2011-2019**: Maturation
    - China: Slowing growth, rebalancing efforts, rising wages
    - India: Structural reforms (GST, demonetization), continued service-sector strength

4. Empirical Results

4.1 Aggregate Growth Performance

Figure 1 summarizes the evolution of PPP-adjusted GDP per capita in levels



Both countries experienced substantial growth over the 54-year period, but with markedly different patterns. China's growth accelerated dramatically after 1978, with a visible inflection point corresponding to Deng's reforms. The trajectory shows sustained rapid growth through 2010, followed by some deceleration but continued strong performance. India's growth was more modest in the 1965-1991 period (the "Hindu rate of growth" era), accelerated after 1991 liberalization, and sustained relatively high growth through 2019 with less volatility than China.

4.2 Period-by-Period Decomposition

Table 2 breaks down growth contributions across the five sub-periods.

Country	Period	Annual Growth Rate	Contribution from Capital	Contribution from Labor	Contribution from TFP
China	1965-1980	4.96%	2.94	2.56	-0.54
India		3.60%	1.23	1.96	0.41
China	1981-1991	5.73%	3.18	2.2	0.35
India		4.93%	1.28	2.23	1.42
China	1992-2001	6.48%	2.94	1.39	2.15
India		5.84%	1.44	1.37	3.03
China	2002-2010	8.67%	3.29	0.73	4.65
India		7.09%	1.94	1.11	4.04
China	2011-2019	4.02%	2.65	0.17	1.2
India		6.40%	1.56	0.56	4.28

All values are percentage-point contributions.

China's Growth Trajectory

From 1965–1980, China's growth was driven primarily by factor accumulation rather than efficiency gains. Output grew at 4.96% per year, with capital contributing 2.94 percentage points (pp) and labor contributing 2.56 pp. However, TFP contributed -0.54 pp, indicating that expanding capital and labor did not translate into proportional productivity improvements. This negative TFP reflects the structural inefficiencies of the late Maoist command economy, where misallocation, political campaigns, and the Cultural Revolution disrupted productive capacity.

Following the reform period beginning in 1978, the growth pattern shifted. Between 1981–1991, growth rose to 5.73%, with capital contributing 3.18 pp and labor 2.20 pp, while TFP improved to 0.35 pp. This modest TFP gain shows the early effects of decentralization, agricultural reform (the Household Responsibility System), and the loosening of price controls, though productivity improvements remained limited as reforms were still incomplete.

The strongest acceleration occurred after market reforms deepened and external openness increased. During 1992–2001, growth averaged 6.48%, with capital contributing 2.94 pp, labor 1.39 pp, and TFP rising substantially to 2.15 pp. Growth was no longer purely extensive but increasingly intensive (using inputs more efficiently). China's global

integration, WTO accession preparations, and the emergence of competitive pressure inside the domestic economy drove these productivity gains.

China's peak growth period came during 2002–2010, when growth reached 8.67%, driven by exceptionally high TFP gains of 4.65 pp alongside sustained capital accumulation (3.29 pp). Labor's contribution declined to 0.73 pp as the unlimited supply of rural surplus labor began to exhaust. The massive TFP contribution reflects China's full integration into global production networks following WTO accession in 2001, economies of scale in manufacturing, and the reallocation of resources from state-owned enterprises to more productive private and foreign-invested firms.

This pace moderates sharply in 2011–2019, where growth slows to 4.02% and TFP falls dramatically to 1.20 pp. Capital still contributes significantly (2.65 pp), but labor's contribution nearly disappears (0.17 pp) as demographic aging and the end of surplus rural labor reduce labor-force growth. The decline in TFP from 4.65 pp to 1.20 pp is consistent with the challenges of transitioning from investment-led, catch-up growth to innovation-driven growth. Diminishing returns to capital accumulation, rising debt, and the difficulty of frontier innovation explain this deceleration.

India's Growth Trajectory

From 1965–1980, India's growth averaged 3.60%, with moderate contributions from capital (1.23 pp) and labor (1.96 pp), but only 0.41 pp from TFP. This modest TFP growth reflects the inefficiencies and regulatory constraints of the License Raj era, where extensive industrial licensing, trade barriers, and state control stifled competition and productivity improvements.

Beginning in 1981–1991, growth increased to 4.93%, driven by gradual deregulation and early liberalization attempts. Capital contribution rose slightly to 1.28 pp, labor remained a steady contributor at 2.23 pp, and TFP improved to 1.42 pp, a more than threefold increase from the previous period. This shows that even partial reforms generated productivity gains, though the License Raj's core restrictions remained intact.

The major structural break appears after the 1991 liberalization. During 1992–2001, growth rose to 5.84%, driven by substantial TFP gains of 3.03 pp. Capital contribution increased to 1.44 pp and labor to 1.37 pp. The sharp jump in TFP reflects the dramatic impact of dismantling industrial licensing, reducing trade barriers, and opening to foreign investment. Competition, access to imported technology, and improved resource allocation generated large efficiency gains.

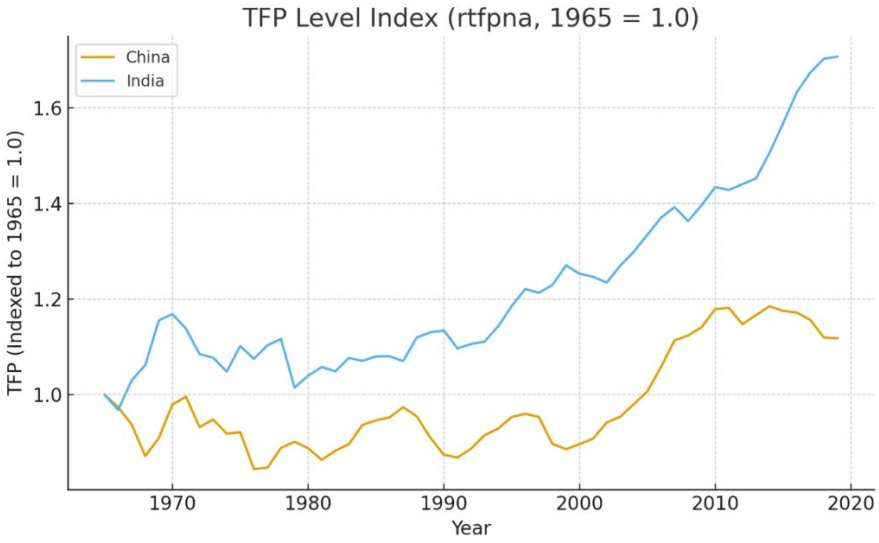
India's strongest performance occurred in 2002–2010, where growth reached 7.09%, supported by remarkably high TFP gains of 4.04 pp and rising capital accumulation (1.94 pp). This TFP contribution is exceptionally large. This likely reflects the confluence of

multiple productivity-enhancing forces: the service-sector boom (particularly IT and business services), the lagged effects of 1991 liberalization reaching maturity, rapid structural transformation as resources shifted from low-productivity agriculture to high-productivity services, and improved telecommunications infrastructure enabling network effects (Roller, Lars-Hendrik, and Leonard Waverman 2001).Labor contribution declined to 1.11 pp as employment growth moderated, but the quality improvements embedded in human capital (captured partly in labor and partly in TFP) remained strong.

Unlike China, India does not exhibit a sharp post-2010 slowdown. From 2011–2019, growth remains high at 6.40%, with TFP still contributing an exceptional 4.28 pp, even higher than the previous period. Capital contributes steadily at 1.56 pp, while labor's contribution declines to 0.56 pp as demographic momentum begins to slow, though it remains stronger than China's due to India's younger population. The persistence of India's TFP growth across multiple periods suggests that its productivity gains may be more durable than China's, though TFP remains a residual measure subject to potential composition and measurement effects. This persistence of TFP growth might stem from service-sector dynamism, ongoing structural transformation, and continued catch-up potential rather than one-time gains from manufacturing integration and reallocation.

4.3 TFP Level Analysis

Figure 2 plots the evolution of TFP levels (indexed to 1965=1.0) for both countries.



China

China's TFP trajectory follows a distinctive pattern of decline, stagnation, and then dramatic acceleration:

- **1965–1980**: Decline in TFP (0.865 → 0.768). This deterioration is consistent with the disruptions of the late-Mao period, including the Cultural Revolution, political campaigns, and the structural inefficiencies of central planning. Resources were mobilized for production, but the command economy's inability to allocate them efficiently meant output growth lagged far behind input growth.
- **1981–1991**: Flat to mild improvement (0.747 → 0.751) as early reforms begin. Agricultural reforms and the emergence of Township and Village Enterprises generated some productivity gains, but state-owned enterprise reforms remained incomplete, and the planning system still dominated much of the economy.
- **1992–2001**: Gradual rise (0.767 → 0.785) with deepening market reforms. Deng Xiaoping's 1992 Southern Tour reinvigorated reforms, private sector development accelerated, and coastal Special Economic Zones expanded. TFP began rising as competitive pressure increased and resource allocation improved.
- **2002–2010**: Sharp jump post-WTO (0.814 → 1.020). This dramatic acceleration reflects China's full integration into global production networks, massive FDI inflows, learning-by-exporting, scale economies in manufacturing, and the continued reallocation of resources from inefficient state firms to productive private and foreign firms. By 2010, China's TFP reached parity with its 1965 level, having recovered all the ground lost during the Maoist era.
- **2011–2019**: Pullback (1.022 → 0.967). TFP declined by 5.4% over this period, a concerning reversal. This pattern is consistent with demographic headwinds, diminishing returns to capital accumulation, and the increasing difficulty of sustaining productivity growth once easy catch-up gains are exhausted. The state sector's continued dominance in key industries and rising debt levels may also constrain productivity.

India

India's TFP trajectory shows steady improvement with notable accelerations after major reforms:

- **1965–1980**: Modest rise (0.597 → 0.621). Even under the License Raj, India achieved small productivity gains, likely from gradual improvements in human capital and the expansion of higher education institutions (IITs, IIMs).
- **1981–1991**: Gradual improvement (0.632 → 0.655). Partial liberalization attempts, relaxation of some licensing requirements, and the early green revolution's productivity spillovers generated modest TFP gains.

- **1992–2001**: Post-liberalization lift (0.661 → 0.745). The 1991 reforms' impact is clearly visible as TFP rose 12.7% over the decade. Dismantling the License Raj, trade liberalization, and opening to FDI allowed more efficient resource allocation and forced domestic firms to improve productivity through competition.
- **2002–2010**: Further gains (0.737 → 0.857), a 16.3% increase. This acceleration reflects the service-sector boom (IT, business services), improved telecommunications infrastructure, continued structural transformation, and the maturation of reforms begun in 1991. Services exhibited rapid catch-up to global productivity frontiers.
- **2011–2019**: Continued rise (0.853 → 1.020), a 19.6% increase. India's TFP reached parity with its 1965 level by 2019. Unlike China's post-2010 reversal, India's productivity continued rising, suggesting more sustainable productivity gains rooted in service-sector dynamism, ongoing structural transformation, and continued catch-up potential.

Comparative Insights

China experienced a slowly upward, yet now dropping TFP trajectory: decline under Maoism, dramatic recovery post-reform, but recent stagnation. China's productivity gains came primarily from "catch-up" mechanisms like adopting existing technologies, reallocating resources from inefficient to efficient uses, and integrating into global production networks. These gains were enormous but may be largely exhausted, explaining the post-2010 slowdown.

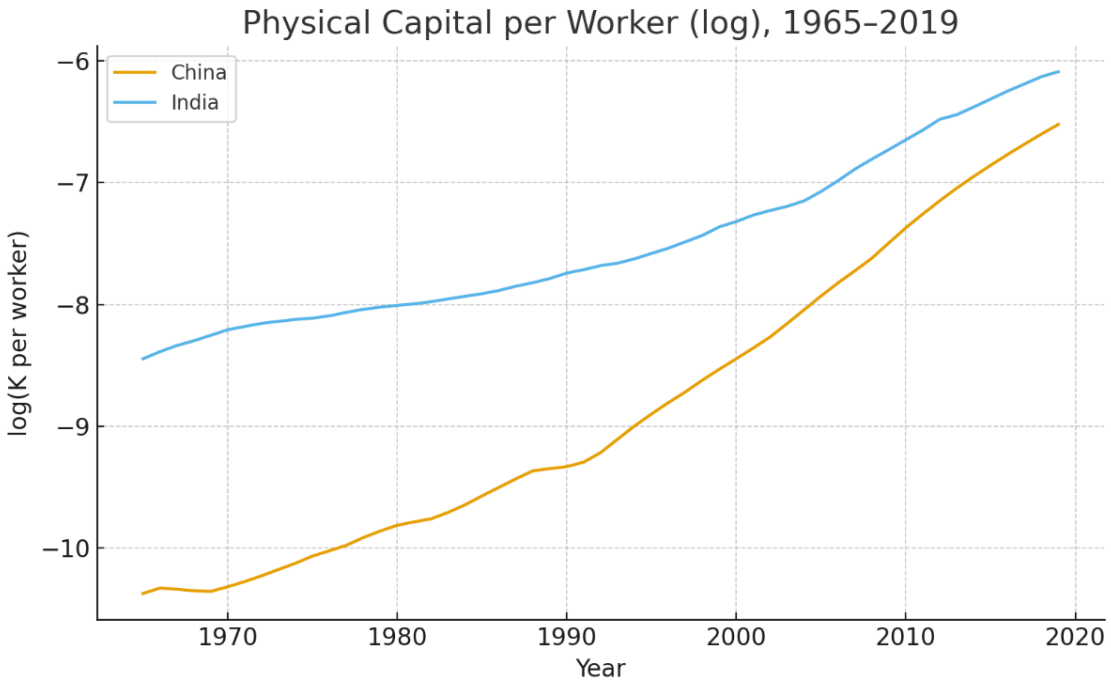
India showed steady TFP improvement throughout, with accelerations after the 1991 reforms. India's productivity gains appear more gradual but potentially more sustainable, driven by service-sector development, capital accumulation, and ongoing structural transformation. India started from a lower TFP base (0.597 vs. China's 0.865 in 1965) but has maintained positive TFP growth even in recent years.

It is important to recognize that TFP is a residual measure capturing everything not explained by measured capital and labor inputs. Exceptionally high TFP growth (such as China's 4.65 pp and India's 4.04 pp contributions during their peak periods) can reflect genuine efficiency improvements and technological progress, but may also partly reflect measurement issues (unmeasured capital quality improvements, composition effects) or one-time structural adjustments rather than sustainable frontier innovation. Nevertheless, the persistence of India's high TFP growth across multiple periods (1.42 pp → 3.03 pp → 4.04 pp → 4.28 pp) and the consistency with institutional narratives suggest these represent real productivity improvements rather than statistical artifacts. The sharp rise and subsequent fall of China's TFP contribution (0.35 pp → 2.15 pp → 4.65 pp → 1.20 pp)

reflects the life cycle of catch-up growth: slow initial gains, rapid middle-phase improvements as reforms bite, then deceleration as catch-up potential exhausts.

4.4 Physical Capital Accumulation

While both countries increased physical capital per worker over time, the pace and pattern differed significantly. Figure 3 shows the evolution of physical capital per worker for both countries.



Physical Capital per Worker (log), 1965-2019. Values shown are natural logarithms of capital stock per worker in constant 2017 national prices. A one-unit increase on the vertical axis represents approximately a 2.7-fold increase in real capital per worker.

Both China and India increased physical capital per worker substantially, but China's rise is far steeper, especially post-1990. The sharp acceleration in China around 1992–2010 corresponds to:

- Enterprise reforms and massive state-directed investment
- Infrastructure build-out (highways, high-speed rail, ports, urban housing)
- WTO accession and export manufacturing scale-up
- Extraordinarily high domestic savings rates (rising from 38% to 50% of GDP)

By 2019, China's capital per worker (in log terms) had increased by approximately 3.0 log points from 1965, representing roughly a 20-fold increase in real terms.

India also deepened capital, but at a slower, more gradual pace. By 2019, India's capital per worker had increased by approximately 2.0 log points from 1965, representing roughly a 7-fold increase. This more moderate pace reflects:

- Lower savings rates (25-35% of GDP vs. China's 40-50%)
- Less foreign direct investment intensity
- More service-based (rather than manufacturing-based) growth, which requires less physical capital per unit of output
- Democratic constraints on land acquisition and infrastructure development

This pattern illuminates the different development strategies documented in the growth accounting. China pursued capital-intensive industrialization with massive infrastructure investment and manufacturing capacity expansion, reflected in the large capital contributions to growth shown in Table 2 (averaging 3.0 percentage points annually across all periods). India's trajectory shows more modest capital accumulation (averaging 1.5 percentage points annually), with greater reliance on TFP growth driven by service-sector productivity, human capital improvements, and structural transformation.

4.5 Sectoral Composition

China:

- Agriculture's GDP share fell from 40% (1965) to 7% (2019)
- Manufacturing rose from 30% to peak of 46% (2006), then declined to 40%
- Services grew from 30% to 53%
- Employment shifted from 80% agricultural to 25% by 2019

India:

- Agriculture's GDP share fell from 45% (1965) to 15% (2019)
- Manufacturing remained relatively constant at 15-18% throughout
- Services surged from 35% to 55%
- Employment shifted from 70% agricultural to 42% by 2019

Interestingly, both countries converged to similar service-sector shares by 2019 (China 53%, India 55%), but through completely different pathways. China's structural transformation followed the classic development sequence: agriculture into manufacturing into services. Manufacturing surged as the dominant sector for three decades (1980-2010), reaching 46% of GDP at its peak, before services eventually overtook it.

India's trajectory defied conventional development models by largely bypassing the manufacturing stage. Manufacturing never exceeded 18% of GDP, remaining stagnant throughout the entire period (Eichengreen and Gupta 2011). Instead, India transitioned directly from agriculture to services, a pattern virtually unprecedented among large developing economies.

China's manufacturing-led growth required massive physical capital investment (factories, machinery, infrastructure, urban housing), which explains China's large capital contributions averaging 3.0 pp annually. Manufacturing productivity gains came primarily from scale economies, learning-by-exporting, and capital deepening. TFP gains arrived later (post-2000) once China integrated into global value chains.

India's service-led growth required less physical capital per unit of output (services are less capital-intensive than manufacturing). This explains India's lower capital contributions, averaging 1.5 pp annually. Service-sector productivity gains came from human capital utilization, technological adoption (IT, telecommunications), and integration into global service value chains. TFP contributions were larger and arrived earlier, reflecting efficiency gains from liberalization and the service sector's catch-up potential

The employment patterns reinforce this distinction. China's employment shift from 80% to 25% agricultural represents one of the largest and fastest labor reallocations in history, funneling workers into manufacturing cities. India's shift from 70% to 42% agricultural was slower and less complete, with workers moving primarily into services rather than factories. This helps explain China's initially higher labor contributions (2.56 pp in 1965-1980) compared to India's (1.96 pp), though both have declined as structural transformation has matured.

5. Structural and Institutional Explanations

The growth accounting results identify what drove growth; primarily capital in China, more balanced contributions including TFP in India. This section traces these proximate sources back to deeper institutional and policy causes.

5.1 China's Growth Model: State-Coordinated Investment and Global Integration

China's growth acceleration after 1978 resulted from a specific institutional configuration that combined state coordination with incremental market expansion. The growth accounting showed China's rise was initially driven by capital deepening (averaging 3.0 pp annually), with TFP gains arriving later (peaking at 4.65 pp in 2002-2010). Three institutional features explain this pattern.

State Capacity and Investment Mobilization

China's centralized party-state enabled coordinated investment strategies impossible in fragmented political systems. Following Deng Xiaoping's consolidation of reform-oriented leadership, the state restored predictability to economic policy, reducing the uncertainty premium that had discouraged capital formation during the Cultural Revolution. Domestic savings rates rose from approximately 30% of GDP in the late 1970s to nearly 38% by the early 1990s, and eventually peaked above 50% in the 2000s (Song, Storesletten, and Zilibotti 2011).

The state channeled these savings through financial repression: household deposits in state-controlled banks earned low interest rates, while banks funneled credit at subsidized rates into industrial investment (Song, Storesletten, and Zilibotti 2011). This system enabled China to maintain investment rates of 40-50% of GDP for three decades—the highest sustained capital deepening in modern economic history. This mechanism directly explains the large capital contributions visible in Table 2 across all periods, peaking at 3.29 pp in 2002-2010.

Agricultural Reform and Structural Transformation

The Household Responsibility System (HRS) replaced collective farming with household-managed plots and output-based incentives beginning in 1978. Agricultural labor productivity rose sharply, and the sector released surplus labor for industrial employment. Between 1978 and 1995, agriculture's share of employment fell from approximately 70% to 50%, while industrial employment doubled. Because productivity in manufacturing was 3-5 times higher than in agriculture, this shift generated significant aggregate TFP gains through reallocation alone (McMillan and Rodrik 2011). This explains why TFP contributions turned positive in the 1980s (0.35 pp in 1981-1991) even while capital remained the dominant growth driver.

Township and Village Enterprises (TVEs) emerged as the primary vehicle for early industrialization outside the state sector. These collectively affiliated enterprises provided access to state-backed credit while operating with market incentives. By the early 1990s, TVEs accounted for about one-quarter of industrial output, enabling broad-based capital accumulation without requiring full privatization.

Global Integration and TFP Acceleration

The strongest acceleration in China's TFP contribution (from 2.15 pp to 4.65 pp between 1992-2001 and 2002-2010) coincided with integration into global production networks. WTO accession in 2001 embedded China into international value chains where firms learned modern management, supply-chain standards, and production techniques through export-oriented manufacturing. By 2010, foreign-invested firms accounted for over half of China's exports, the share of high-technology goods in export value had more than doubled

relative to 1995, and productivity dispersion across firms narrowed as inefficient firms exited while efficient firms scaled (Feenstra, Inklaar, and Timmer 2015; Pavcnik 2002; McMillan and Rodrik 2011).

These productivity gains came primarily from imitation, adaptation, and scale learning rather than frontier innovation. The subsequent TFP decline to 1.20 pp in 2011-2019 reflects the exhaustion of these gains as the reallocation of labor from agriculture slowed, the working-age population peaked in 2012, and diminishing returns to capital accumulation set in.

5.2 India's Growth Model: Liberalization and Service-Sector Transformation

India's growth trajectory differed fundamentally in timing, sectoral composition, and institutional drivers. The growth accounting showed more modest capital contributions (averaging 1.5 pp) but sustained high TFP growth (rising from near-zero pre-1991 to 4.28 pp by 2011-2019).

From License Raj to Liberalization

From 1947 through 1991, India's License Raj required firms to obtain government approval for investment, expansion, production levels, and product specifications. Trade barriers (tariffs averaging 85%) protected the domestic industries while the public sector dominated heavy industry and finance. This system produced the near-zero TFP growth visible in Table 2 for 1965-1991 (0.41 pp and 1.42 pp). Licensing protected inefficient incumbents from competition, trade barriers prevented access to imported capital goods embodying new technologies, and credit misallocation based on political criteria constrained productivity.

The 1990-91 balance-of-payments crisis forced comprehensive reforms. The 1991 liberalization package dismantled industrial licensing, reduced import tariffs from 85% to 25%, eliminated quantitative restrictions on imports, opened sectors to FDI, and reduced government-directed credit. The growth accounting results demonstrate the dramatic impact: TFP growth jumped to 3.03 pp in 1992-2001 and continued rising to 4.28 pp by 2011-2019.

Liberalization boosted TFP through three mechanisms supported by empirical evidence. First, removing trade barriers and licensing exposed Indian firms to competition, forcing efficiency improvements. Inefficient firms contracted or exited while productive firms expanded, raising average productivity through selection effects (Pavcnik 2002). Second, trade liberalization gave firms access to advanced capital goods and intermediate inputs. The reduction in capital goods tariffs from over 100% to under 25% made technology adoption economically feasible. Third, eliminating licensing freed entrepreneurs to enter

profitable sectors and expand successful businesses, allowing capital and labor to flow toward higher-productivity uses (McMillan and Rodrik 2011).

Service-Sector Dynamics and Sustained TFP Growth

India’s services' GDP share rose from 40% in 1991 to 55% by 2019, while manufacturing stagnated at 15-18% (unlike China, where manufacturing grew significantly first). Indian software services exports grew from essentially zero in 1991 to $100 billion by 2019. This transformation was enabled by the English-language legacy of British colonialism, emphasis on tertiary education, telecommunications liberalization providing infrastructure for remote service delivery, favorable time-zone differences with Western markets, and low wages creating cost advantages.

The magnitude of India's TFP gains during 2002-2010 (4.04 pp annually) represents one of the highest sustained rates recorded for a major economy. Several factors explain this performance. First, significant lags existed between the 1991 reforms and their full productivity impact; firms needed time to restructure, new entrants needed time to scale up, and inefficient incumbents needed time to exit. By the 2000s, these dynamic effects reached maturity. Second, structural transformation accelerated as employment shifted from agriculture (60% in 2000) to services and manufacturing (53% by 2010). Since productivity gaps between sectors were enormous (service-sector productivity was 3-4 times agricultural productivity), this reallocation generated substantial aggregate TFP growth (McMillan and Rodrik 2011). Third, improved telecommunications infrastructure (mobile phone penetration rose from 1% to 60%) and increased FDI created network effects and knowledge spillovers.

From a growth accounting perspective, service-sector expansion explains India's distinctive pattern: services require less physical capital than manufacturing (explaining lower capital contributions), but generate sustained TFP gains through human capital utilization, technological adoption, and integration into global service value chains. India also benefited from demographic dividends, with the working-age population continuing to expand through 2019, explaining the consistently positive labor contributions (1.96 pp → 0.56 pp) compared to China's sharper decline (2.56 pp → 0.17 pp).

5.3 Comparative Institutional Dynamics

The growth accounting results reflect fundamental institutional differences. China's ability to mobilize massive investment through state direction appears in the large capital contributions (averaging 3.0 pp versus India's 1.5 pp). India's more market-driven, entrepreneurial growth path appears in stronger, sustained TFP contributions (averaging 2.6 pp versus China's 1.6 pp).

Acemoglu, Johnson, and Robinson (2005) argue that institutions are the fundamental cause of long-run growth differences. China's trajectory combined extractive political institutions (one-party rule) with increasingly inclusive economic institutions (market mechanisms, property rights, trade openness). This combination proved highly effective for catch-up growth, enabling rapid coordinated reforms and massive infrastructure investment without democratic contestation. As emphasized by Acemoglu, Johnson, and Robinson (2005), however, extractive political institutions can become increasingly constraining as economies approach the technological frontier, where decentralized experimentation and creative destruction play a larger role, an interpretation consistent with China's post-2010 TFP decline from 4.65 pp to 1.20 pp.

India maintained inclusive political institutions (democracy, federalism, free press) while gradually reforming extractive economic institutions (License Raj dismantling, trade liberalization). Democratic constraints slowed reforms but made them more durable and credible. The federal structure allowed experimentation with successful state-level reforms spreading nationally. Inclusive political institutions may be less constraining for sustained frontier innovation, where decentralized experimentation and entry play a central role, consistent with India's continued TFP growth to 4.28 pp in 2011-2019.

5.4 Policy Sequencing and Path Dependence

The timing and sequencing of reforms mattered crucially for both countries. China's reforms began with agriculture (low political resistance, high returns), proceeded to light manufacturing in SEZs, then gradually expanded to the broader economy. This gradualist, experimental approach generated sustained momentum and learning by doing.

India's sequencing was more reactive. The 1991 crisis forced comprehensive reforms simultaneously rather than gradually. While this approach eliminated many distortions quickly, it also created adjustment costs and political backlash that slowed subsequent reforms. Nevertheless, the severity of the 1991 crisis created political consensus for change that might have been impossible to achieve gradually.

Path dependence is also evident. China's manufacturing focus is built on its inherited industrial base from the Soviet-style planning era, while India's service-sector strength is built on its legacy of English-language education and IIT system. These initial conditions channeled reforms and comparative advantages in different directions, contributing to the distinct growth patterns observed.

6. Conclusion

Through comprehensive growth accounting, I have shown that China achieved average annual growth of approximately 6% driven substantially by capital accumulation

(contributing 3.0 percentage points annually on average), while India grew at 5.6% annually with more balanced contributions, including strong TFP growth (2.6 points on average versus China's 1.6 points).

The period-by-period decomposition reveals the critical importance of institutional reforms. China's growth accelerated dramatically after Deng Xiaoping's 1978 reforms, with capital contributions surging from 1.2 to over 6 percentage points annually by 2002-2010. India's turning point came with the 1991 liberalization, after which TFP growth jumped from near-zero to 1.1 percentage points and continued rising to 1.8 points by 2011-2019.

Tracing these proximate sources back to deeper causes, I have shown how China's political stabilization, Open Door Policy, and manufacturing-oriented strategy generated the investment boom and technology transfer that powered growth. India's License Raj dismantling, trade liberalization, and service-sector emergence drove productivity gains through competition, efficiency improvements, and human capital utilization.

6.1 Sustainability and Future Prospects

China's challenges:

- Diminishing returns to capital accumulation are already visible in slowing growth post-2010
- TFP growth has decelerated from 4.65 percentage points (2002-2010) to 1.20 points (2011-2019)
- Demographic headwinds from aging population will reduce labor force growth
- Environmental degradation from rapid industrialization imposes mounting costs
- Debt accumulation to maintain investment rates creates financial stability risks
- Innovation-led growth requires institutional changes that may threaten political stability, a transition consistent with the observed post-2010 decline in China’s TFP growth

India's challenges:

- Manufacturing weakness limits job creation for less-educated workers
- Infrastructure deficits constrain growth potential
- Agricultural sector still employs 42% of workers at low productivity
- Human capital quality gaps persist despite quantity improvements
- Regulatory and bureaucratic barriers remain despite reforms

6.2 Final Reflections

The China-India comparison illuminates both the possibilities and complexities of economic development. Two countries starting from similar positions of poverty in 1965 have lifted over one billion people out of destitution through different strategies, different institutional arrangements, and different sectoral emphases. China's extraordinary capital accumulation and manufacturing prowess coexist with demographic decline and slowing productivity. India's rising TFP growth and demographic dividend coexist with infrastructure gaps and manufacturing weakness.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." *Handbook of Economic Growth* 1: 385-472.

Bai, Chong-En, Chang-Tai Hsieh, and Yingyi Qian. 2006. "The Return to Capital in China." *Brookings Papers on Economic Activity* 2006(2): 61-101.

Barro, Robert J., and Jong Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics* 104: 184-198.

Brandt, Loren, Guangjie Li, Peter M. Huang, and Yifan Zhang. 2017. “The Impact of the Household Responsibility System on China’s Agricultural Productivity.” *Journal of Development Economics* 123: 1-17.

Bosworth, Barry, and Susan M. Collins. 2008. "Accounting for Growth: Comparing China and India." *Journal of Economic Perspectives* 22(1): 45-66.

Chen, Yi, and Nancy Qian. 2021. "The Rise of China and the Decline of the U.S. Economy? Evidence from U.S. Commercial Real Estate." *Journal of Economic Perspectives* 35(4): 59-82.

Eichengreen, Barry, and Poonam Gupta. 2011. "The Two Waves of Service-Sector Growth." *Oxford Economic Papers* 65(1): 96-123.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105(10): 3150-3182.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3): 379-399.

Javorcik, Beata Smarzynska. 2004. "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages." *American Economic Review* 94(3): 605-627.

McMillan, Margaret S., and Dani Rodrik. 2011. "Globalization, Structural Change and Productivity Growth." *NBER Working Paper* 17143.

Oi, Jean C. 1999. *Rural China Takes Off: Institutional Foundations of Economic Reform.* University of California Press.

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies* 69(1): 245-276.

Rivera-Batiz, Luis A. 2002. "Democracy, Governance, and Economic Growth: Theory and Evidence." *Review of Development Economics* 6(2): 225-247.

Roller, Lars-Hendrik, and Leonard Waverman. 2001. “Telecommunications Infrastructure and Economic Development.” *American Economic Review* 91(4): 909–923.

Stewart, Douglas B., and Yiannis P. Venieris. 1985. "Sociopolitical Instability and the Behavior of Savings in Less-Developed Countries." *Review of Economics and Statistics* 67(4): 557-563.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economic Review* 101(1): 196–233.

Walder, Andrew G. 2015. *China Under Mao: A Revolution Derailed.* Harvard University Press.

Exchange Rate Regimes and Recovery From the Global Financial Crisis

Christine Sinn*

MIT Department of Economics

Abstract

Macroeconomic evidence has long documented a weak unconditional relationship between exchange rate movements and real activity, giving rise to the exchange rate "disconnect" puzzle. Recent work shows that conditional on identified nominal depreciation shocks, exchange rates can have large real effects on the economy through financial channels. Using local projections on 167 non-U.S. economies from 1990 to 2017, this paper studies how U.S.-pegged and floating regimes differentially adjusted to the Global Financial Crisis of 2008-2009, a period of sharp dollar appreciation and global financial stress. I find that pegged regimes experienced a pronounced short-run rebound in consumption and output relative to floats. These near-term gains were accompanied by higher inflation, sustained real appreciation, and weaker export performance. Towards the end of the decade, differences in real outcomes had largely converged across regimes. This paper shows that regime-dependent financial amplification can also extend to a dollar appreciation episode. Exchange rate regimes may not determine long-run recovery levels, but they significantly shape the timing and composition of adjustment to crisis shocks.

*I would like to thank Christian Wolf and Keelan Beirne for their generous advice and support throughout this entire project. Additionally, I would like to thank Kazuatsu Shimizu for his added guidance on the econometrics and trade theory behind this research.

1 Introduction

Since Milton Friedman's argument in 1953 that flexible exchange rate regimes can insulate their economies more effectively against real shocks, economists have debated the relative advantages of pegged versus flexible exchange rate regimes (Friedman, 1953). In the Mundell-Fleming framework, flexible exchange rates stabilize output by enabling expenditure switching between domestic and foreign goods through nominal depreciation, while fixed regimes forego this adjustment margin in exchange for a nominal anchor. Later dynamic open-economy models with sticky prices formalized these tradeoffs, emphasizing how exchange rate flexibility facilitates faster real exchange rate adjustment following shocks (Dornbusch, 1980; Obstfeld and Rogoff, 1996).

However, Mussa's empirical work in 1986 showed that the shift from the fixed exchange rates of Bretton Woods to floating exchange rates in 1973 led to a sharp increase in nominal and real exchange rate volatility without a corresponding rise in real macroeconomic volatility (Mussa, 1986). This "Mussa puzzle" has since introduced a stream of "disconnect" literature, which observes that exchange rate movements are only weakly unconditionally correlated with real aggregate macroeconomic outcomes such as GDP or consumption (Obstfeld and Rogoff, 2001; Devereux and Engel, 2002; Itskhoki and Mukhin, 2021).

Recent post-Mussa literature has aimed to address the question of any aggregate importance of exchange rate regime choice through shock-based identification. Poole showed that real (nominal) shocks should cause pegged (floating) exchange rate regimes to experience relatively larger variance in output, as pegs rely on sticky price levels that produce more sluggish adjustment in their real exchange rate (Poole, 1970). Empirically, Broda found empirical evidence supporting Poole's conclusion that floating regimes buffer real terms-of-trade shocks better through nominal depreciation accelerating real exchange rate adjustment (Broda, 2004). Using a Post-Bretton Woods sample of 75 developing countries, Broda found that the short-run real GDP response to terms of trade changes is significantly larger in countries with pegged regimes, due to nominal restraint creating a slow depreciation of the real exchange rate via domestic price deflation.

More recently, Fukui, Nakamura, and Steinsson found that nominal exchange rate shocks can have large real effects through financial channels rather than the traditional Mundell-Fleming expenditure switching channel (Fukui et al., 2025). Studying episodes

of U.S. dollar depreciations, including the coordinated depreciation following the 1985 Plaza Accord, they find that pegged regimes experience amplified expansionary effects relative to floats.

The mechanism in Fukui, Nakamura, and Steinsson's model operates through deviations from uncovered interest parity (UIP) that reflect time-varying risk premia and discount factors. In their theoretical model, exchange rate shocks coincide with shifts in required returns: a nominal depreciation shock is associated with lower domestic discount rates and risk premia. When financial integration is imperfect, interest differentials are not fully offset by expected exchange rate changes, allowing capital flows from shifting risk premia to affect domestic asset prices and credit. Floating regimes can use monetary policy to partially offset these shocks through interest rate adjustments. The monetary policy constraints of a peg, however, cause capital inflows from a depreciation shock to translate more directly into expansionary domestic demand. This framework reconciles large conditional effects of exchange rate movements with the weak unconditional correlation literature. While identified nominal shocks can have substantial real consequences, offsetting financial disturbances generate exchange rate volatility without systematic real comovement.

In this paper, I aim to extend the shock-based identification approach of previous literature to the Global Financial Crisis (GFC) of 2008-2009, a period characterized by sharp U.S. dollar appreciation as the "safe haven currency" and financial asset market contagion (Blinder, 2013). By leveraging a global shock that comprised both real economic impacts and nominal shocks, I assess how the regime-dependent financial amplification mechanism extends to such appreciation episodes and how it interacts with real adjustment channels.

I apply local projection methods first proposed by Jordà in 2005 and further explained by Li, Wolf, and coauthors in 2024 on 167 countries from 1990-2017, assessing the dynamic impacts of the Global Financial Crisis and providing suggestions for the mechanisms behind such impacts (Jordà, 2005; Li et al., 2024). My findings suggest a shift in dominant channels affecting the real outcomes of pegged regimes over time: a dominant financial channel from UIP deviations and post-crisis credit expansion supported a short-run consumption-led rebound, but constrained monetary autonomy meant that this boom came at the cost of real appreciation, higher inflation, and a gradual erosion of external trade competitiveness in the longer run. These offsetting external and financial mecha-

nisms rendered pegged and floating regimes differentially neutral in long-run aggregate GDP and consumption recovery outcomes despite very different adjustment paths.

The remainder of the paper is structured as follows: Section 2 provides more context for the origins of the Global Financial Crisis and preliminary balance checks between pegs and floats leading up to the crisis. Section 3 describes the data used in the project and empirical specification. Section 4 describes the empirical results from the local projections and assesses mechanisms that may explain the differential outcomes across short and long-term horizons. Finally, Section 5 describes limitations and next steps for future research.

2 Background

2.1 The Global Financial Crisis

The Global Financial Crisis of 2008-2009 originated in the United States and spread throughout the global interbanking system with unexpected intensity. A variety of market conditions have been attributed to culminating in the global crisis. First, a decade of low short-term rate setting by the Fed led to ample liquidity and overheated demand in the housing market (Federal Deposit Insurance Corporation, 2017). House prices in cumulative real terms increased by around 92% from 1996 to 2006, a statistic that far exceeds the 27% cumulative increase reported from 1890 to 1996 (Reinhart and Rogoff, 2008). Rising house prices led to extended expectations of long-rising house values reminiscent of bubble behavior (Case and Shiller, 2003). Risky underwriting standards developed in the mortgage market as a result, from interest-only mortgages to adjustable rate mortgages (Federal Deposit Insurance Corporation, 2017). While these looser standards made the increasing cost of a house more affordable for lower-income buyers, the business also led to predatory behavior against unsophisticated first-time buyers or real estate investors looking to refinance and turn a profit (Blinder, 2013).

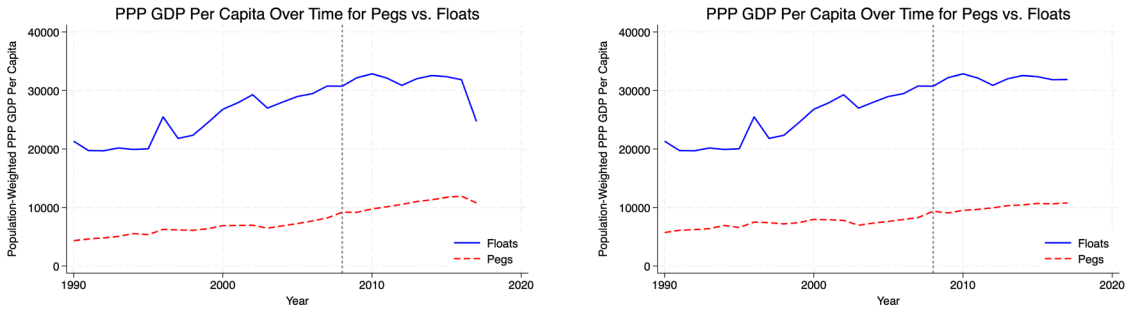
Furthermore, financial market deregulation incentivized the development of a "shadow banking" system, with financially engineered derivative products providing opaquely layered promises to reduce risk and Structured Investment Vehicles (SIVs) hiding leverage ratios from being reported on bank balance sheets (Blinder, 2013). This shadow banking system connected the real housing bubble with the "global savings glut" demanded in the financial markets, leading to broader network effects than originally anticipated from the

housing bubble burst (Blinder, 2013). As the housing bubble began to burst, investors feared financial firms were exposed to unseen levels of risk from layered and unreported mortgage investments, causing a rapid withdrawal of short-term liquidity and a run on U.S. money market funds (McCauley and McGuire, 2009).

The Global Financial Crisis originated as mainly a U.S. housing bubble, but it spread to the rest of the world because many foreign banks, governments, and sovereign wealth funds had invested in the U.S. mortgage derivatives market to gain from attractive returns (International Monetary Fund, 2009). The largely unanticipated contagion of risk connecting the U.S. housing and global financial markets provides a reasonable instance of exogenous variation by which we can test differential real impacts and propagation mechanisms of exchange rate movements and crisis across currency regimes.

2.2 The Recovery: An Initial View

Before turning to formal estimation, it is useful to examine aggregate recovery paths across exchange rate regimes. Figure 1 plots population-weighted PPP GDP per capita for pegged and floating economies from 1990 through 2017. The left panel includes all countries, while the right panel excludes China, whose regime reclassification in 2017 affects group composition.



(a) Including China

(b) Excluding China

Figure 1: Population-Weighted PPP GDP Per Capita Over Time. China was excluded from the right graph because of its regime switch during the sample period skewing later year results.

The unconditional recovery patterns appear broadly similar, and there is no obvious large divergence in long-run income trajectories between the two groups. In fact, pegged regimes seem to exhibit somewhat smoother aggregate dynamics over the full sample, though this visual difference is modest.

At face value, these aggregate trends are consistent with the "disconnect" view that exchange rate regime choice may not produce large unconditional differences in real outcomes. However, simple averages mask systematic heterogeneity prior to the crisis, discussed in the balance checks below, and different mechanisms for recovery after the crisis hit given the seeming irrelevance of exchange rate regime. These observations motivate a more structured empirical approach. I next conduct balance checks to assess pre-crisis comparability across regimes and then implement a local projection framework that traces the dynamic, cumulative responses of macroeconomic outcomes relative to pre-crisis baselines.

2.3 Balance Checks

I conducted balance check regressions to identify any structural differences between pegs and floats from 1990-2007 that could lead to differential exposure in the years preceding the 2008 Financial Crisis. Balance checks were conducted unconditionally (as a difference in means), with time fixed effects, and with region-by-time fixed effects. All regressions have country-clustered standard errors.

Balance checks show that pegs and floats were mostly similarly exposed to the crisis on a structural level, with some notable differences that have been highlighted in Table 1 (see Appendix 1 for full balance checks on other structural variables of interest). With region and time fixed effects, pegged regimes had a GDP per capita 0.511 log points lower (approximately 40% lower) than floating regimes, and 0.324 log points lower PPP GDP per capita (approximately 28% lower) than floats. This suggests that pegs tend to be poorer emerging economies. Despite differences in income levels, pegged and floating regimes exhibit statistically insignificant pre-trend consumption and investment trajectories in the five years preceding 2008, which supports the parallel trends assumption underlying the dynamic empirical specification. Pegs also have non-performing loan (NPL) to gross loan ratios that are 3.394 percentage points higher than floats, indicating weaker bank balance sheets before the crisis. Pegs spent 1.830 percentage points less of their GDP on government expenditures than floats, suggesting lower fiscal capacity to address real shocks. Interestingly, pegged regimes were less internationally exposed with regard to international debt issuance (7.406 percentage points lower). Finally, pegged countries had 42.265 percentage points less inflation than floating countries over the tested pre-crisis

Table 1: Balance Checks (Abridged): Peg vs. Float (1990–2007)

	Unconditional	Time FE	Region×Time FE	Observations
Log Population	−0.082	−0.099	0.311	3006
	(0.259)	(0.262)	(0.307)	
Log GDP Per Capita	−0.312	−0.335	−0.511***	2960
	(0.200)	(0.203)	(0.181)	
Log PPP GDP Per Capita	−0.221	−0.242	−0.324**	2942
	(0.159)	(0.161)	(0.147)	
Consumption (% GDP)	−0.464	−0.325	1.069	2497
	(1.931)	(1.959)	(2.452)	
Investment (% GDP)	−0.699	−0.752	−0.695	2434
	(0.937)	(0.950)	(1.266)	
Government Spending (% GDP)	−3.222***	−3.203***	−1.830**	2467
	(0.746)	(0.760)	(0.887)	
Inflation (CPI)	−78.043***	−68.680***	−42.265**	2958
	(23.165)	(20.526)	(16.405)	
Bank NPL to Gross Loans	4.629***	4.477***	3.394**	838
	(1.211)	(1.219)	(1.637)	
International Debt Issues (% GDP)	−11.639***	−11.760***	−7.406**	1170
	(3.892)	(4.025)	(3.327)	
Country FE	No	No	Yes	
Year FE	No	Yes	Yes (by region)	

Notes: Each cell reports the difference in means between pegged and floating exchange-rate regimes estimated by OLS over 1990–2007. Unconditional: $X_{it} = \alpha + \beta \, \text{Peg}_{it} + \varepsilon_{it}$; Time FE adds year fixed effects; Region×Time FE adds region-by-year fixed effects. Standard errors clustered by country in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

period, reflecting that a key motivation for countries to peg their currencies is to effectively manage their inflation, especially during a particularly volatile economic period for pegs. The inflation-stabilizing effect of the peg can be partially seen by the reduced pre-crisis inflation difference down to around 10 percentage points in the five years leading up to the crisis. For other structural variables, there were no statistically significant structural differences between pegged and floating countries. These balance checks suggest that pegged economies entered the Global Financial Crisis with lower income levels, weaker banking sectors, and lower international debt exposure than floating regimes. It is worth noting that some of these structural differences may come from the growth in the number of pegged economies (and symmetric decline in the number of floating economies) during the 1990-2007 time period that is shown in Figure 2.

3 Data and Empirical Specification

3.1 Exchange Rate Regime Classification

There are a variety of exchange rate regime classifications across policy and academia sources, from the International Monetary Fund's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER) to Ghosh and coauthors' classification of pegged regimes into frequent and infrequent adjusters (Ghosh et al., 1997). To create a comparable test of the theoretical model of Fukui and coauthors, I use their same source of exchange rate regime classification produced by Ilzetzki, Reinhart, and Rogoff (Ilzetzki et al., 2019). Ilzetzki and authors offer a coarse 6-category and fine 15-category classification of exchange rate regimes. I organize these categories into simpler binary variables for U.S. pegs vs. floats. This binary classification provides the following mix of peggers vs. floaters from 1990 to 2017, shown in Figure 2 (including China). The prevalence of U.S. dollar pegs increases during the 1990s, reflecting the widespread adoption of exchange-rate-based stabilization policies and the growing role of the dollar in international trade and finance (see Appendix 2 for a list of countries that switched from pegs to floats and vice versa from 1990-2000). Because the classification is de facto, tightly managed regimes anchored to the dollar are included in this category.

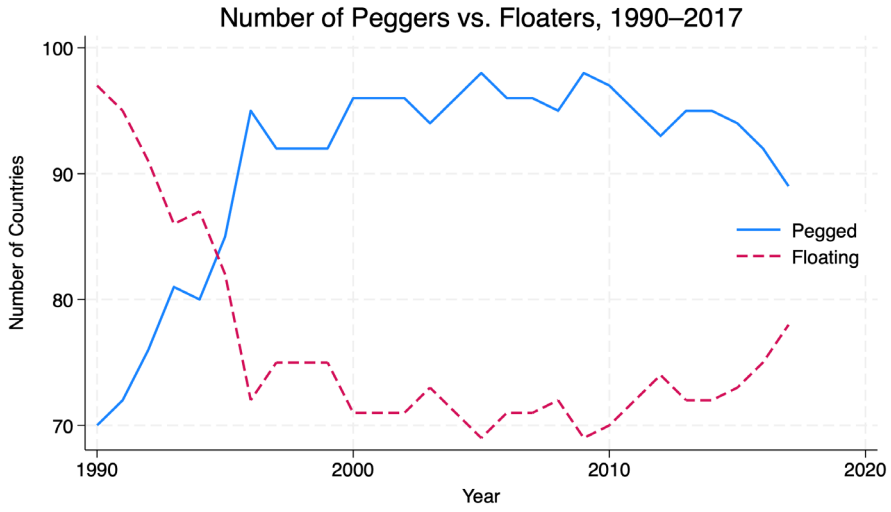


Figure 2: Composition of Pegs and Floats from 1990-2017

3.2 Other Data Sources

I used the World Bank World Development Indicators (WDI) database for real outcomes such as real GDP (in constant 2015 U.S. dollars and in PPP terms), real consumption, and investment. I used the Global Financial Development (GFD) database for financial development indicators such as the country's reserve ratio, international debt issue ratio, private credit ratio, bank NPL to gross loans, bank concentration percentage, and bank z-score stability. For the unemployment rate, the International Labor Organization measurement published on the WDI database was used to allow for better cross-comparison across countries. The Chinn-Ito Index, updated recently to 2022 data, was used as a normalized 0-1 index measuring a country's stated intended policy toward capital openness (Chinn and Ito, 2008). To evaluate changes in external competitiveness, I employ real and nominal effective exchange rate indices from Bruegel, a Brussels think tank, constructed as trade-weighted averages relative to each country's 65 largest trading partners. Since expenditure switching and trade performance depend on relative prices against a country's overall trade basket rather than its bilateral exchange rate with the U.S. alone, effective exchange rates offer a more relevant measure for competitiveness (Darvas, 2012; and Darvas, 2021). In the Bruegel data, an increase in the country's exchange rate index reflects an appreciation of that currency relative to its effective trading partners.

3.3 Empirical Specification

When deciding the empirical specification for this research, vector autoregression (VAR) and local projection (LP) stood out as two primary specification strategies in previous shock-identification literature (Broda, 2004; Fukui et al., 2025). Wolf and authors conduct a simulation study comparing LP vs. VAR estimators and find that both methods are econometrically equivalent but have a bias-variance tradeoff: LP estimators have lower bias but higher variance at intermediate and long horizons compared to VAR estimators (Li et al., 2024). To best align my findings with recent nominal shock identification literature by Fukui and authors, I chose to use local projection to conduct my research.

I estimate the differential response along various outcome variables between pegged and floating exchange rate regimes using the following local projection specification:

$$y_{i,t+h} - y_{i,t-1} = \alpha_i + \gamma_t + \beta_h \left(\text{Peg}_{i,t-1} \times \text{Crisis}_t\right) + \delta_h \, \text{Peg}_{i,t-1} + \theta_h \, y_{i,t-1} + \varepsilon_{i,t+h}, \qquad (1)$$

where $\Delta y_{i,t+h}$ denotes the overall change in the outcome variable for country i between years $t-1$ and $t+h$. $\text{Peg}_{i,t-1}$ is an indicator for whether country i maintained a pegged exchange rate regime in the year prior to the crisis, and Crisis_t is an indicator for the Global Financial Crisis period of 2008-2009. Country fixed effects α_i and year fixed effects γ_t control for time-invariant heterogeneity within countries and common global shocks within a year, respectively. These fixed effects effectively add in an indicator variable for the crisis, so the interaction term is properly specified. The $y_{i,t-1}$ term controls for one lag of GDP growth, similar to the identification of Fukui and authors.

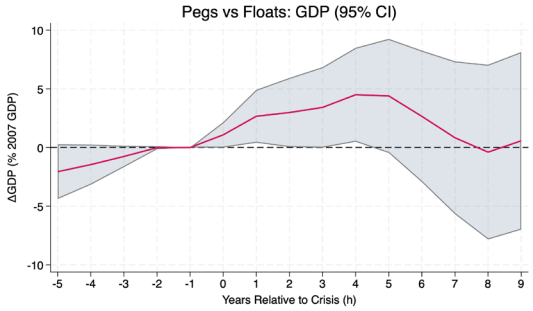
For the outcome variables, I measured the overall changes in these variables at each horizon relative to the pre-crisis year of 2007. Real macroeconomic outcome variables are expressed as a percent of 2007 GDP to normalize magnitudes across countries. This also allows the estimated interaction coefficients to be interpreted as differential cumulative effects relative to a common pre-crisis baseline, akin to a difference-in-difference estimate. Variables such as the exchange rate, inflation, and terms of trade were measured by their difference in log points. Their coefficients are interpreted as approximate cumulative percentage changes from the pre-crisis baseline. In particular, measuring the exchange rate and inflation in log points ensures symmetry between appreciations and depreciations and symmetry between percentage changes in prices across countries. Measuring terms of trade ratios by log points converts the ratios into more intuitive differences.

I cluster standard errors at the country level to account for serial correlation within countries. This specification allows the estimated coefficients β_h to trace out the dynamic response of pegged relative to floating regimes over a nine-year horizon from 2008-2017 ($h = 0, 1, \ldots, 9$). Five years of pre-trends were included to zoom into the parallel trends assumption beyond the historical differential exposure in the longer decades-long horizon shown in the above balance checks.

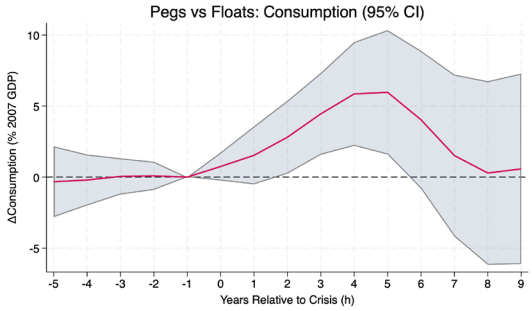
4 Results and Mechanisms

4.1 Real Outcomes

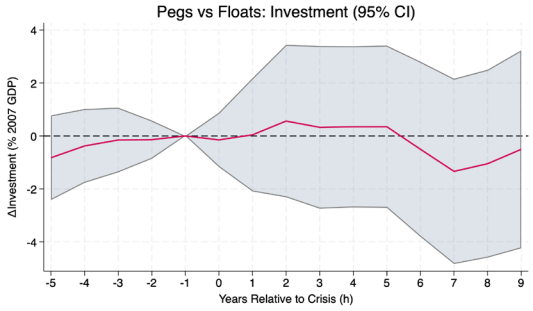
The dynamic impulse responses of real macroeconomic outcomes from the local projection specification are shown in Figure 3. The results indicate that pegged economies experienced a short-run expansion relative to floating regimes following the Global Financial Crisis. Pegged countries experienced a cumulative real GDP level approximately 4.492 percentage points higher (than pre-crisis GDP) relative to floats in the first four years after 2008. This expansion is concentrated in a relative 5.856 percentage point increase in domestic consumption as a percent of pre-crisis GDP relative to floating regimes, while unemployment decreases more sharply in pegs by 1.775 percentage points. Investment shows no statistically significant relative response throughout the post-crisis horizon. The composition of this recovery suggests a demand-driven expansion rather than one fueled by capital accumulation.



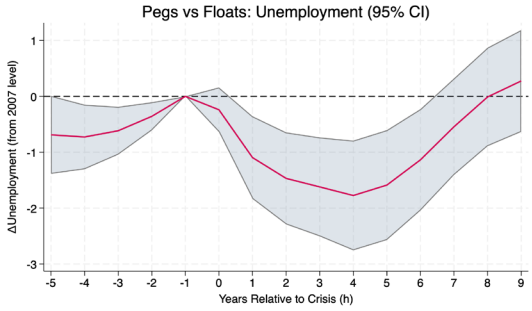
(a) GDP (% Change from 2007 GDP)



(b) Consumption (% of 2007 GDP)



(c) Investment (% of 2007 GDP)



(d) Unemployment Rate (from 2007)

Figure 3: Post-Crisis Dynamics of Real Macroeconomic Aggregate Outcomes

Notably, this short-run expansion occurs alongside a steady appreciation of the real effective exchange rate (see Figure 4), which would theoretically dampen net exports. After

approximately five to six years post-crisis, the relative differences in GDP, consumption, and unemployment all return close to the initial pre-crisis levels between pegs and floats. This convergence echoes the exchange rate "disconnect" literature, which documents unconditional correlations between exchange rate regimes and aggregate real activity. The following subsections examine the mechanisms underlying this short-run expansion and subsequent convergence.

4.2 Short Run: Financial Channel

The short-run expansion in pegged economies coincides with a sustained real effective exchange rate appreciation. As shown in Figure 4, pegged regimes experienced a persistent and tightly identified increase in their real effective exchange rate relative to floating regimes following the crisis. This relative appreciation reaches approximately 11.2% after four years and remains elevated at over 21% seven years after the crisis.

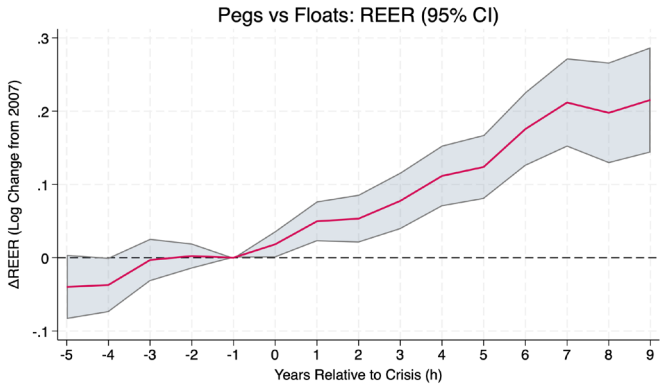


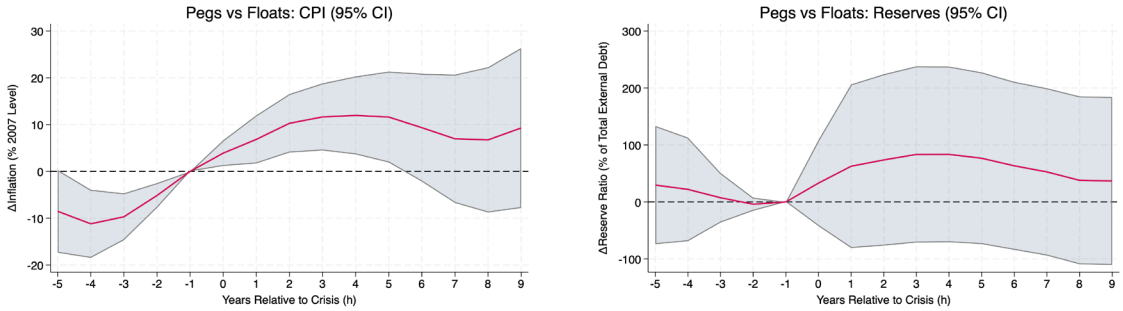
Figure 4: Real Effective Exchange Rate Differences Across Regimes

The source of this appreciation appears closely related to the defense of the nominal anchor during a period of U.S. dollar safe haven strength. Because pegged economies maintained fixed nominal exchange rates against the dollar, the global appreciation of the dollar during the crisis translated instead into upward pressure on their real exchange rates through higher real prices. Local projections on the CPI differences from pre-crisis levels show evidence of this domestic price adjustment. Figure 5 shows that pegs experienced around 11.641-11.963% significantly higher cumulative inflation from pre-crisis levels by three to four years after the crisis.

I also traced the reserve ratios of countries as a percentage of their total external debt

pre and post-crisis. Figure 5 shows that pegged regimes accumulated foreign reserves sharply in the aftermath of the crisis, increasing their reserves to external debt ratio by 62.681 percentage points more than floating regimes in the first year after the crisis and reaching a peak difference of 83.445 percentage points three years after the crisis. The standard errors were high for the reserve ratio because of smaller sample size from reported data (95-97 countries instead of 167), so the reserve ratio impulse response should be viewed as a suggested pattern consistent with active intervention to sustain the peg.

The increase in both domestic prices and reserves likely reflects that countries maintained their nominal exchange-rate pegs through the crisis, while floating countries adjusted via nominal depreciations in the short-run to restore competitiveness in real terms. This pattern persisting from the early years after the crisis suggests that the short-run GDP and consumption expansion for pegs was likely accompanied by higher nontradable prices and continued pressure on export competitiveness. This is further elucidated in the external channel analysis later in the paper.



(a) Inflation (b) Real Exchange Rate

Figure 5: Short Run Financial Mechanisms

Reserve accumulation under a pegged regime typically expands the central bank's balance sheet, as pegs are buying foreign (U.S.) currency and issuing domestic base money during periods of financial inflow pressure to maintain the peg. To prevent the expansion of the country's monetary base, a central bank can sell domestic bonds to "sterilize" the foreign exchange intervention and reduce inflationary pressures on the economy.

The significant increase in inflation for pegs post-crisis suggests that the peg defense was not fully sterilized on average across pegged economies, leading to domestic liquidity expansion and easing credit conditions. This interpretation is further supported by the steady rise in private credit as a percentage of pre-crisis GDP for pegged economies relative

to floats. Figure 6 shows that six years after the crisis, pegs accumulated 10.282 percentage points (of pre-crisis GDP) more private credit relative to pre-crisis levels compared to floats. At its peak, private credit rose by 16.677 percentage points more in pegs eight years after the crisis.

The expansion of private credit along with rising consumption and falling unemployment in pegs relative to floats (seen in Figure 3) points to an easing of domestic financial conditions in pegged countries. This would suggest that the short-run expansion was largely demand-driven and connected to household absorption rather than financing new capital formation.

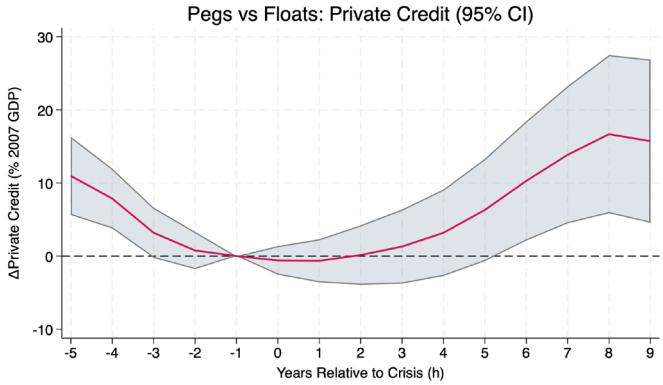


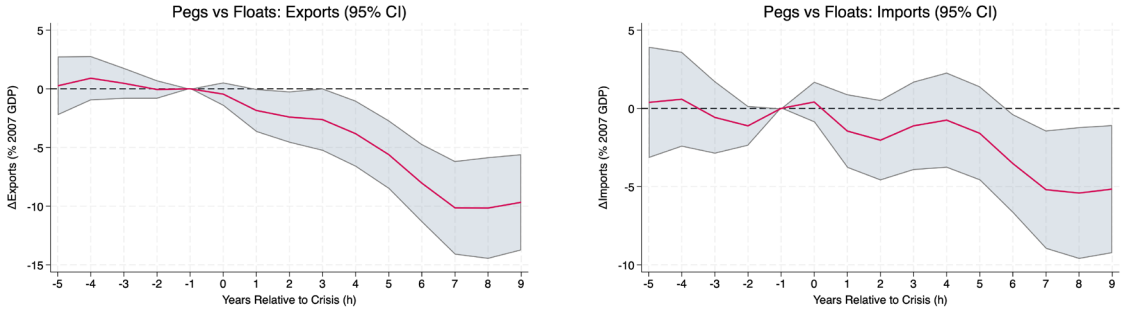
Figure 6: Private Credit Differences Across Regimes

These short-run patterns suggest that pegged regimes did not contract demand but rather responded to safe-haven U.S. dollar appreciation by defending the anchor and expanding domestic liquidity. As a result, pegs experienced a relative boost to consumption and output despite this real appreciation. These results are consistent with previous literature on the existence of a financial channel compared to former expenditure switching models. Yet, these results furthermore show that the sign of the exchange rate movement itself is not the central mechanism. In the Fukui et al. framework, exchange rate depreciations reflect shifts in domestic discount rates and liquidity conditions arising from deviations from uncovered interest parity. The present results show that a demand-driven expansion can occur even in an appreciation episode. This seems to suggest that the financial channel mechanism operates through domestic liquidity and credit conditions rather than through the sign of the exchange rate change itself. Whether the shock manifests as an appreciation or depreciation, the associated easing of domestic liquidity and declining domestic risk premia drives short-run demand expansion despite real appreciation.

4.3 Long Run: External Channel

While pegged regimes experience a short-run demand expansion, external trade outcomes become significant in the longer-run dynamics specifications. Through this mechanism, a floating exchange rate regime shows clear advantages, leading to the overall neutral differences between pegs and floats in the long run.

Figure 7 shows that both the export and import shares as a percentage of pre-crisis GDP decline significantly in pegged economies following the crisis. Export share falls steadily and sharply, initially being 3.826 percentage points lower than pre-crisis levels after four years post-crisis before reaching a maximum decline of 10.143 percentage points around seven years after 2008. Pegs experience a less steady but still sharp decline in imports, reaching a maximum difference of 5.411 percentage points less than pre-crisis differences eight years after the crisis.

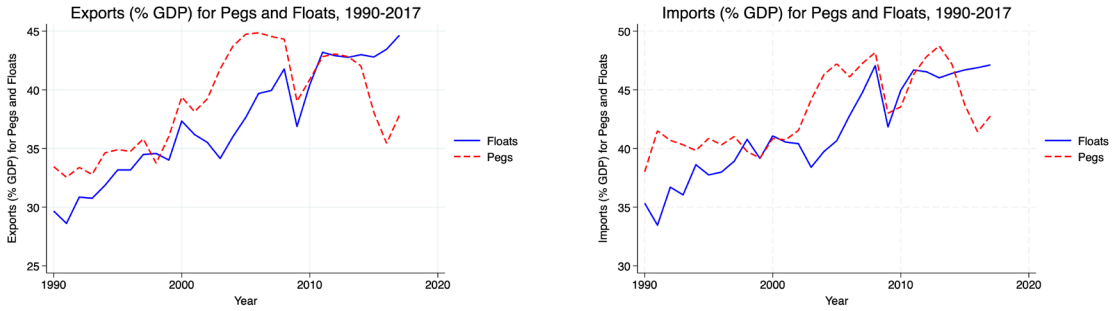


(a) Export Share (% of GDP) (b) Import Share (% of GDP)

Figure 7: Long Run Global Trade Competitiveness

These trade effects are economically meaningful for pegged economies. In Figure 8, exports were worth around 44.5% of pegged economies' GDP and import around 47% of GDP in 2007. A relative decline of 10.143 percentage points in magnitude therefore represents a material erosion of external demand and a substantial fraction of GDP (in value ratios). Importantly, the results in Figure 7 satisfy parallel pre-trends in the five years leading up to the crisis, and balance checks indicate no statistically significant pre-crisis differences in trade exposure between pegs and floats (see Appendix 1). The absence of differential pre-trends strengthens the interpretation that the post-crisis divergence in trade competitiveness reflects regime-dependent adjustment mechanisms rather than pre-existing structural exposures.

The contraction in exports is consistent with the sustained real appreciation discussed



(a) Mean Export Share

(b) Mean Import Share

Figure 8: Mean Export and Import Share for Pegs vs. Floats, 1990-2017

above. With limited nominal flexibility and rising domestic prices, pegged economies absorbed safe-haven dollar strength through higher real exchange rates and demand-driven consumption. Over time, this real appreciation without any significant change in investments undermined pegged economies' external competitiveness and led to converging real output between pegs and floats in the long run.

Finally, I observed whether pegged economies aimed to increase their capital accounts or financial openness to offset their weakening trade competitiveness. Figure 9 shows the differences in the Chinn-Ito capital account openness index between pegs and floats compared to pre-crisis levels. There is a modest decline in capital account openness for pegged regimes in the immediate years following the crisis, but this subsequently stabilizes and returns to pre-crisis differences. This pattern indicates that the long-run trade deterioration is not likely driven by significant post-crisis changes in capital account policy but rather through sustained price-level adjustments.

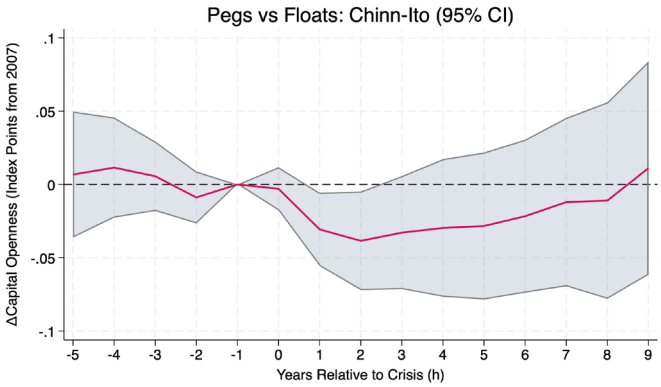


Figure 9: Long Run Capital Account Openness

These long-run dynamics help reconcile the narrative between short-run financial ex-

pansion and long-run convergence in real outcomes. While pegged regimes experience a temporary demand boost supported by liquidity expansion and credit growth, the associated real appreciation gradually erodes external competitiveness. Over time, the external channel offsets the initial gains, leading GDP and consumption differentials to converge toward zero. These results align with the broader "exchange rate disconnect" literature, which finds limited unconditional long-run differences in aggregate real activity across exchange rate regimes. At the same time, the results suggest that financial channels can generate economically significant short-run effects and change the path of recovery for pegged economies after crisis.

5 Conclusion

This paper studies the differential recovery paths of U.S. pegged and floating exchange rate regimes after the Global Financial Crisis of 2008. While the aggregate recoveries of real outcomes were broadly similar in the long run, countries in different regimes arrived to long-run neutrality through distinct adjustment patterns. Pegged economies exhibited a pronounced short-run expansion in domestic demand, reflected in stronger consumption and moderately higher GDP levels relative to floats. These early gains, however, were accompanied by higher inflation, sustained real appreciation, and steadily sharper deterioration in external competitiveness. Over time, these pressures offset the initial expansion, and pegged economies converged back toward the trajectory of floating regimes. These findings align with recent empirical models of financial transmission differences across exchange rate regimes. I also show that the sign of the exchange rate movement is not relevant to short-run real outcomes, but rather the change in risk premia and domestic liquidity in a country post-crisis. This paper offers a bridge between the financial channel and the long-standing expenditure-switching "disconnect" view by showing that exchange rate regimes shape the composition and timing of adjustment rather than long-run aggregate recovery.

A key limitation of this paper is the high variance of results, which may come from differential exposures to the crisis across regimes and limited sample size for some recovery mechanisms. Potential areas for future research may choose to conduct a cross-comparison of results with VAR methods on top of local projection in order to focus on variance re-

duction and tighten identification for more suggestive mechanisms that did not satisfy pre-trends. Further research can also be conducted across multiple U.S. dollar appreciation shock episodes to gather more well-identified, aggregate results on relative responses across different regimes. Finally, the paper's results on the relevance of risk premia and liquidity conditions over the sign of exchange rate movements suggests that future theoretical research connecting existing exchange rate models with stochastic asset pricing literature may provide interesting connections. Research on financial stability and its policy mechanisms remains an important area of focus.

References

Blinder, A. S. (2013). *After the music stopped: The financial crisis, the response, and the work ahead*. Penguin Press.

Broda, C. (2004). Terms of trade and exchange rate regimes in developing countries. *Journal of International Economics*, *63*(1), 31–58. https://www.sciencedirect.com/science/article/pii/S0022199603000436

Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, (2), 299–362. https://www.brookings.edu/wp-content/uploads/2003/06/2003b_bpea_caseshiller.pdf

Chinn, M. D., & Ito, H. (2008). A new measure of financial openness. *Journal of Comparative Policy Analysis*, *10*(3), 309–322.

Darvas, Z. (2012). *Real effective exchange rates for 178 countries: A new database* (Working Paper No. 2012/06). Bruegel.

Darvas, Z. (2021). *Timely measurement of real effective exchange rates* (Working Paper No. 2021/15). Bruegel.

Devereux, M. B., & Engel, C. (2002). Exchange rate pass-through, exchange rate volatility, and exchange rate disconnect. *Journal of Monetary Economics*, *49*(5), 913–940.

Dornbusch, R. (1980). *Open economy macroeconomics*. Basic Books.

Federal Deposit Insurance Corporation. (2017). Origins of the crisis [Chapter 1]. In *Crisis and response: An FDIC history, 2008–2013*. https://www.fdic.gov/analysis/quarterly-banking-profile/fdic-history.html

Friedman, M. (1953). The case for flexible exchange rates. In *Essays in positive economics* (pp. 157–203). University of Chicago Press.

Fukui, M., Nakamura, E., & Steinsson, J. (2025). *The macroeconomic consequences of exchange rate depreciations* [Working paper, January 2025], University of California, Berkeley and Boston University. https://eml.berkeley.edu/~enakamura/papers/trilemma.pdf

Ghosh, A. R., Gulde, A.-M., Ostry, J. D., & Wolf, H. C. (1997). *Does the nominal exchange rate regime matter?* (Working Paper No. 5874). National Bureau of Economic Research.

Ilzetzki, E., Reinhart, C. M., & Rogoff, K. S. (2019). *Exchange arrangements entering the 21st century: Which anchor will hold?* (Working Paper No. 23134). National

Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w23134/w23134.pdf

International Monetary Fund. (2009, April). *World economic outlook: Crisis and recovery*. https://www.imf.org/external/pubs/ft/weo/2009/01/pdf/text.pdf

Itskhoki, O., & Mukhin, D. (2021). Exchange rate disconnect in general equilibrium. *Journal of Political Economy*, *129*(8), 2183–2232.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182.

Li, Z., Plagborg-Møller, M., & Wolf, M. (2024). *Local projections vs. vars: A simulation study* [Working paper]. https://economics.mit.edu/sites/default/files/2024-01/lp_var_simul.pdf

McCauley, R. N., & McGuire, P. (2009). Dollar appreciation in 2008: Safe haven, carry trades, dollar shortage and overhedging. *BIS Quarterly Review*, 85–93. https://www.bis.org/publ/qtrpdf/r_qt0912.pdf

Mussa, M. L. (1986). Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications. *Carnegie-Rochester Conference Series on Public Policy*, *25*(1), 117–214.

Obstfeld, M., & Rogoff, K. (1996). *Foundations of international macroeconomics*. MIT Press.

Obstfeld, M., & Rogoff, K. (2001). Six major puzzles in international macroeconomics: Is there a common cause? In *Nber macroeconomics annual 2000* (pp. 339–390, Vol. 15). MIT Press.

Poole, W. (1970). Optimal choice of monetary policy instruments in a simple stochastic macro model. *Quarterly Journal of Economics*, *84*(2), 197–216.

Reinhart, C. M., & Rogoff, K. S. (2008, March). *This time is different: A panoramic view of eight centuries of financial crises* (Working Paper No. 13882). National Bureau of Economic Research. Cambridge, MA. https://www.nber.org/papers/w13882

6 Appendix 1: Full Balance Checks

Please find my full set of balance checks done on multiple variables of interest below.

Table 2: Balance Checks (Full): Peg vs. Float (1990–2007)

	Unconditional	Time FE	Region×Time FE	Observations
Chinn-Ito Capital Openness	0.015	0.011	0.031	2824
	(0.046)	(0.046)	(0.048)	
Bank Z-Score Stability	1.998	2.004	−1.281	1025
	(1.267)	(1.272)	(1.314)	
Bank Concentration %	−2.710	−2.705	1.739	1032
	(2.726)	(2.737)	(3.308)	
Bank NPL to Gross Loans	4.629***	4.477***	3.394**	838
	(1.211)	(1.219)	(1.637)	
Domestic Private Credit (% GDP)	−6.752	−6.698	−4.981	2086
	(6.353)	(6.362)	(6.527)	
Log GDP per Capita	−0.312	−0.335	−0.511***	2960
	(0.200)	(0.203)	(0.181)	
Log PPP GDP per Capita	−0.221	−0.242	−0.324**	2942
	(0.159)	(0.161)	(0.147)	
Consumption (% GDP)	−0.464	−0.325	1.069	2497
	(1.931)	(1.959)	(2.452)	
Government Spending (% GDP)	−3.222***	−3.203***	−1.830**	2467
	(0.746)	(0.760)	(0.887)	
Investment (% GDP)	−0.699	−0.752	−0.695	2434
	(0.937)	(0.950)	(1.266)	
International Debt Issues (% GDP)	−11.639***	−11.760***	−7.406**	1170
	(3.892)	(4.025)	(3.327)	
Reserve Ratio (% External Debt)	−35.500	−39.184	−39.706	1637
	(35.962)	(37.497)	(35.951)	
Trade Exposure (% GDP)	7.000	6.296	9.266	2555
	(6.722)	(6.812)	(7.568)	
Exports (% GDP)	3.841	3.403	4.364	2555
	(3.639)	(3.691)	(3.990)	
Imports (% GDP)	3.159	2.892	4.902	2555
	(3.277)	(3.319)	(3.830)	
Net FDI Inflows (Current US $)	-4.775×10^{9}**	-5.182×10^{9}**	-2.933×10^{9}*	2921
	(2.049×10^{9})	(2.144×10^{9})	(1.487×10^{9})	
Inflation (CPI)	−78.043***	−68.680***	−42.265**	2958
	(23.165)	(20.526)	(16.405)	
Unemployment Rate (ILO)	−1.207	−1.246	−1.287	2754
	(0.815)	(0.828)	(0.764)	
Log Population	−0.082	−0.099	0.311	3006
	(0.259)	(0.262)	(0.307)	
Country FE	No	No	Yes	
Year FE	No	Yes	Yes (by region)	

Notes: Each cell reports the difference in means between pegged and floating exchange-rate regimes estimated by OLS over 1990–2007. Unconditional: $X_{it} = \alpha + \beta\,\mathrm{Peg}_{it} + \varepsilon_{it}$; Time FE adds year fixed effects; Region×Time FE adds region-by-year fixed effects. Standard errors clustered by country in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7 Appendix 2: Peg and Float Composition

The number of pegged countries grew rapidly from 69 countries in 1990 to 95 countries by 2000. A full list of the countries in the 167-country sample that switched exchange rate regimes over the decade are below in Table 3. A full list of floating and pegged countries in 2007 are below in Tables 4 and 5.

Table 3: Exchange Rate Regime Switches, 1990–2000

Switched to Peg	Switched to Float
Argentina	Algeria
Armenia	Ghana
Azerbaijan	Liberia
Brazil	Nepal
Cambodia	South Africa
Dominican Republic	
Ecuador	
Georgia	
Guatemala	
Guyana	
Honduras	
Jamaica	
Kazakhstan	
Kenya	
Kyrgyz Republic	
Latvia	
Lebanon	
Lithuania	
Malawi	
Moldova	
Mongolia	
Mozambique	
Nicaragua	
Paraguay	
Peru	
Russian Federation	
Sao Tome and Principe	
Sierra Leone	
Uganda	
Ukraine	
Vietnam	

Table 4: Floating Exchange Rate Regimes in 2007

Countries	
Albania	Australia
Austria	Belgium
Benin	Bhutan
Bosnia and Herzegovina	Botswana
Brazil	Bulgaria
Burkina Faso	Côte d'Ivoire
Cameroon	Canada
Cape Verde	Central African Republic
Chad	Comoros
Congo, Rep.	Croatia
Cyprus	Czech Republic
Denmark	Equatorial Guinea
Estonia	Fiji
Finland	France
Gabon	Germany
Greece	Guinea-Bissau
Hungary	Iceland
Ireland	Israel
Italy	Japan
Latvia	Lesotho
Lithuania	Macedonia, FYR
Madagascar	Mali
Malta	Mauritius
Morocco	Nepal
Netherlands	New Zealand
Niger	Norway
Poland	Portugal
Romania	Russian Federation
Samoa	Senegal

Table 4 continued

Countries	
Slovak Republic	Slovenia
South Africa	Spain
Swaziland	Sweden
Switzerland	Togo
Tonga	Tunisia
Turkiye	United Kingdom
Venezuela, RB	

Table 5: Pegged Exchange Rate Regimes in 2007

Countries	
Afghanistan	Algeria
Angola	Antigua and Barbuda
Argentina	Armenia
Azerbaijan	Bahamas, The
Bahrain	Bangladesh
Barbados	Belarus
Belize	Bolivia
Burundi	Cambodia
Chile	China
Colombia	Costa Rica
Djibouti	Dominica
Dominican Republic	Ecuador
Egypt, Arab Rep.	El Salvador
Ethiopia	Gambia, The
Georgia	Ghana
Grenada	Guatemala
Guinea	Guyana
Haiti	Honduras
Hong Kong, China	India
Indonesia	Iran, Islamic Rep.
Iraq	Jamaica
Jordan	Kazakhstan
Kenya	Korea, Rep.
Kuwait	Kyrgyz Republic
Lao PDR	Lebanon
Liberia	Libya
Malawi	Malaysia
Maldives	Mauritania
Mexico	Moldova

Table 5 continued

Countries	
Mongolia	Mozambique
Nicaragua	Nigeria
Oman	Pakistan
Panama	Papua New Guinea
Paraguay	Peru
Philippines	Qatar
Rwanda	Sao Tome and Principe
Saudi Arabia	Seychelles
Sierra Leone	Singapore
Solomon Islands	Sri Lanka
St. Kitts and Nevis	St. Lucia
St. Vincent and the Grenadines	Sudan
Suriname	Syrian Arab Republic
Tajikistan	Tanzania
Thailand	Trinidad and Tobago
Uganda	Ukraine
United Arab Emirates	Uruguay
Uzbekistan	Vietnam
Yemen, Rep.	Zambia

Applications of the Budgetary Exploitation Theorem in Core Theory to California's NEM 1.0 Policy

Anoushka Tamhane

14.04 Fall 2025

Introduction

Solar power is the fastest-growing source of energy in the United States, and rapid solar uptake is critical for a successful, low-cost energy transition. In 2024, the US generated 303,167 GWh of electricity from solar power, up 27% from 2023. California is by far the top solar producer, producing 26% of the nation's solar, followed by Texas, which produces 15% (Climate Central, 2025). A primary policy driving residential solar production is net metering (NEM). NEM is a system where solar owners are given credits for electricity for the solar power they send to the grid, in effect paying electricity rates on their "net" consumption of electricity and acting as electricity prosumers rather than simply consumers. A form of NEM policy has been implemented in 41 states, and is the primary way producers of solar energy are compensated (Gong et al., 2017). However, NEM has been criticized for unfairly subsidizing solar producers at the expense of other, often lower-income electricity consumers.

California utilities have long argued that net metering shifts costs to non-solar consumers through shifting grid fixed costs off of solar consumers and causing non-solar consumer bills to rise. Argosino & Knittel (2025) find weak causal evidence of this through a positive correlation between rooftop solar and electricity retail prices, and attribute this to net metering because of a lack of similar relationship for utility-scale solar and wind. However, net metering reforms are risky (Argosino and Knittel, 2025). After a weakening of California's NEM policy in NEM 3.0, solar installations fell by 80% and have slowly climbed back to 2019 levels since (Barbose, 2024). The average size of PV installations today remains smaller than before NEM 3.0 (Barbose, 2024).

In this paper, we model California's NEM 1.0 pricing scheme, understand market inefficiencies through applications of core theory, and discuss implications for solar policy.

1 Consumer Model

NEM began in California in 1996 with Senate Bill 656, with a simple scheme. For every kWh of solar sold to the grid, households would receive a credit for one kWh of electricity they could use later when not producing solar. This was essentially a 1:1 ratio for the retail rate:solar export rate, with the cap that consumers would not be paid for production beyond what they consumed and a general cap of a 1000 kW system (California Public Utilities Commission, 2022). A model of NEM 1.0 by Comello & Reichelstein (2017) finds that these caps, along with generation caps based on the size of the home, generally determined the size of a system (Comello and Reichelstein, 2017).

Inefficiencies and inequities in NEM 1.0 arose from the payment structure of utilities. Because utilities act as a monopoly, the retail rate they charged consumers was based on the utility's average total cost plus a state-determined cap on profit, stopping even higher electricity prices.

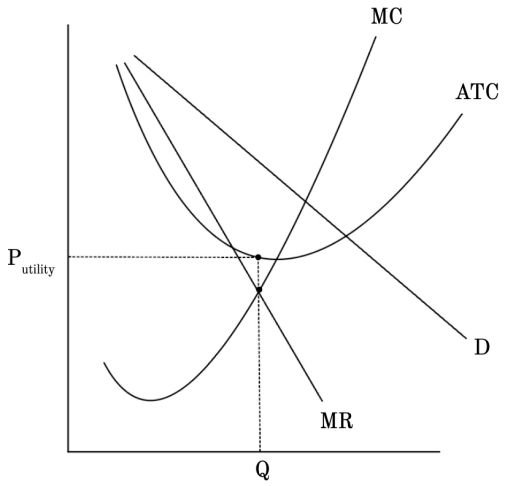


Figure 1: Basic pricing scheme of utilities as monopolies with a binding revenue balancing constraint.

1.1 Non-solar/Solar Consumers

We model NEM 1.0 using three types of agents: utility-maximizing non-solar consumers, using subscript i and solar prosumers, using subscript j, and a utility subject to a revenue balancing constraint. Each non-solar household i chooses electricity consumption d_i and non-electricity consumption x_i maximizes utility subject to the budget and nonnegativity constraints:

$$\begin{aligned} \max_{d_i, x_i} \quad & u_i(d_i) + x_i \\ \text{s.t.} \quad & \pi^+ d_i + x_i \leq \omega, \\ & d_i \geq 0. \end{aligned}$$

where A_i is the consumer's value of consuming goods other than electricity.

We maximize the Lagrangian:

$$\mathcal{L}_i = u_i(d_i) + x_i + \lambda_i(\omega_i - \pi^+ d_i - x_i)$$

And get the following KKT conditions.

$$\frac{\partial \mathcal{L}_i}{\partial x_i} = -\lambda_i = 0 \quad \Rightarrow \quad \lambda_i = A_i$$

$$\frac{\partial \mathcal{L}_i}{\partial d_i} = u_i'(d_i) - \lambda_i \pi^+ = 0$$

Substituting $\lambda_i = A_i$ gives the consumption condition:

$$u_i'(d_i) = A_i \pi^+.$$

After normalizing $A_i = 1$, this simplifies to:

$$u_i'(d_i) = \pi^+.$$

Each solar household j chooses consumption d_j, generation g_j, imports d_j^+, exports d_j^-, and the numeraire x_j. The solar household is also subject to budget constraints, energy balance constraints, complimentary slackness for exports and imports and nonnegativity conditions for exports and imports, as well as a maximum generation constraint with max generation $\bar{g}$. An implicit assumption is made that we are ignoring any generation that is immediately consumed.

$$\begin{aligned}
\max_{d_j, x_i, g_j, d_j^+, d_j^-} \quad & u_j(d_j) + A_j x_j \\
\text{s.t.} \quad & d_j - g_j = d_j^+ - d_j^-, \\
& x_i + \pi^+ d_j^+ - \pi^- d_j^- + c g_j \leq B_j, \\
& g_j \geq 0,\ d_j^+ \geq 0,\ d_j^- \geq 0,\ d_j \geq 0, \\
& g_j \leq \bar{g}.
\end{aligned}$$

The Lagrangian is:

$$\begin{aligned}
\mathcal{L}_j = u_j(d_j) + A_j x_j + \Theta_j \left(B_j - (\pi^+ d_j^+ - \pi^- d_j^- + c g_j + x_j) \right) \\
+ \gamma_j (d_j - g_j \psi_j - (d_j^+ - d_j^-)) - \alpha_j d_j^- - \beta_j d_j^+ + \rho_j (g_j - \bar{g}) - \mu_j x_j.
\end{aligned}$$

The Lagrange multipliers are the dual variables of the household optimization problem and measure the marginal value of relaxing each constraint. In particular, λ and Θ_j are shadow prices of income, while γ_j, α_j, β_j, and ρ_j measure the marginal value of the energy-balance, flow nonnegativity, and generation-capacity constraints, respectively. These dual variables determine how physical and budget constraints map into the households' effective marginal conditions and thereby govern their consumption and generation choices. This yields the following KKT conditions under three different cases: net importing, net exporting, and self-consumption.

Case A: Net Importer ($d_j^+ > 0, d_j^- = 0$). Then $\alpha_j = 0$ and $\beta_j = 0$, so:

$$\gamma_j = -\Theta_j \pi^+, \qquad u_j'(d_j) = \Theta_j \pi^+.$$

Case B: Net Exporter ($d_j^- > 0, d_j^+ = 0$). Then $\alpha_j = 0$ and $\beta_j > 0$, so:

$$\gamma_j = -\Theta_j \pi^-, \qquad u_j'(d_j) = \Theta_j \pi^-.$$

Case C: Self-Consumption ($d_j^+ = d_j^- = 0$). Then:

$$d_j = g_j \psi_j, \qquad \pi^- \leq \frac{u_j'(d_j)}{\Theta_j} \leq \pi^+.$$

1.2 Utility

The utility must recover its fixed and variable costs. Its net revenue equals:

$$\pi^+ \left(\sum_i d_i + \sum_j d_j^+ \right) - \pi^- \left(\sum_j d_j^- \right) = N + c \left(\sum_i d_i + \sum_j d_j^+ \right).$$

Let

$$D^+ = \sum_i d_i + \sum_j d_j^+, \qquad D^- = \sum_j d_j^-.$$

Then the break-even condition is:

$$\pi^+ D^+ - \pi^- D^- = N + cD^+.$$

Under NEM 1.0:

$$\pi^+ = \pi^- = \pi.$$

Substituting yields:

$$\pi(D^+ - D^-) = N + cD^+.$$

Solving for the retail price:

$$\pi = \frac{N + cD^+}{D^+ - D^-}.$$

Differentiating the price with respect to solar exports:

$$\frac{\partial \pi}{\partial D^-} = \frac{N + cD^+}{(D^+ - D^-)^2} > 0.$$

As exports (D^-) rise, the price π increases and the rate of the price change increases. This illustrates

the "cost shift" where non-solar customers may pay more to cover fixed costs. According to consumer problem solutions, importers and non-solar consumers import/consume according to retail price, so they will reduce their imports and shift more towards exports, causing the price to rise further. This illustrates the utility "death spiral", which will continue until there are only the customers who cannot afford solar left paying extremely expensive prices.

2 Core Theory Interpretation

We apply the Budgetary Exploitation Theorem from Shitovitz (1973):

Theorem 1 *Let there be an atomless set T_0 (measure zero small traders) and a positive-measure agent T_1 (an atom). In such economies, feasible allocations in the core must satisfy the following result.*

Let x be in the core. Then there exists a price vector p such that (i) (p, x) is an efficiency equilibrium, and (ii) $p \cdot x(t) \leq p \cdot w(t)$ for almost all $t \in T_0$. Shitovitz (1973)

This states that in a market with "atom" traders, there exists an efficiency price p such that nearly all the small traders are assigned an allocation that is valued less than or equal to their initial endowment. They are "budgetarily exploited", and the gain of the large traders is equal to the losses of the small traders. Small traders cannot form a coalition to impact the core allocation, since they have essentially zero effect on the allocation. Large traders, on the other hand, can form their own coalition, so must be weakly better off in the core.

In order to determine whether there is budgetary exploitation, we model an unrealistic competitive benchmark in order to compare the welfare of the non-solar and solar traders. We assume that instead of the utility setting the prices, the utility, non-solar consumer, and solar prosumer are all price takers. The consumers set consumption/import/export where marginal utility equals price:

$$u_i'(d_i) = \pi^+ \quad \text{and} \quad u_j'(d_j^+) = \pi^-$$

The utility maximizes revenue subject to variable generation cost (c^D). We assume a linear marginal cost.

$$\max_{\pi^+, \pi^-} \pi^+(D+G) - \pi^- G - N - c^D D$$

The derivative with respect to demand D sets the import price:

$$\frac{\partial \pi}{\partial D} = \pi^+ - c^D = 0$$

The derivative with respect to generation G sets the relationship between prices:

$$\frac{\partial \pi}{\partial G} = \pi^+ - \pi^- = 0$$

$$\implies \pi^+ = \pi^-$$

As expected, in a competitive market, prices equal marginal cost:

$$\pi^+ = \pi^- = c^D$$

.

In NEM 1.0, the no-solar consumer has an endowment $(g_i, x_i) = (0, \omega_i)$, and an endowment value of $A_i\omega_i$. The value of their budget to them is $A_i x_i + \pi d_i = A_i\omega_i$. If π is set too high (to cover fixed costs N and maintain utility break-even), the non-solar consumer is made worse off compared to the competitive scenario:

$$u_i(d_i) + A_i x_i < u_i(d_i^*) + A_i x_i^*$$

Where d_i^* and x_i^* are the competitive (lower price) allocations. This reflects how "small" traders are made weakly worse off under budgetary exploitation, and in this case utility exploitation.

In contrast, the solar consumer has endowment $(g_j, \omega_j) = (g, \omega)$, endowment value $A_j\omega_j + \pi g$, and budget value $A_j x_j + \pi d_j = A_j\omega_j + \pi g$. If π is set too high (to cover fixed costs N and maintain break-even), the solar consumer receives benefits from selling their excess power at a price (π) that is inflated to cover the system's fixed network costs. Here, solar consumers act as "large" agents due to their ability to generate and the price regulations in their favor, and gain the losses from the non-solar consumers.

No Net-Metering Scenario

Under no net metering $\pi^- = 0$ (exports have zero value) and thus cannot generate surplus for the household. The solar household chooses consumption and generation subject to the physical accounting constraint

$$d_j - g_j = d_j^+ - d_j^-,$$

with imports $d_j^+ \geq 0$ and exports $d_j^- \geq 0$, and faces the budget constraint

$$x_j + \pi^+ d_j^+ + c g_j \leq B_j.$$

The household solves

$$\max_{d_j, g_j, d_j^+, d_j^-, x_j} \quad u_j(d_j) + A_j x_j$$

subject to the constraints above.

The relevant first–order condition for exports is

$$\frac{\partial \mathcal{L}}{\partial d_j^-} = \gamma_j - \alpha_j = 0 \quad \Rightarrow \quad \gamma_j = \alpha_j \geq 0,$$

(where γ_j is the dual on the energy balance constraint) with complementary slackness $\alpha_j d_j^- = 0$ as α_j is the dual on the d_j^- nonnegativity constraint. Hence any positive export would require $\gamma_j = 0$. The first–order condition for generation is

$$\frac{\partial \mathcal{L}}{\partial g_j} = -c\Theta_j - \gamma_j = 0 \quad \Rightarrow \quad \gamma_j = -c\Theta_j < 0,$$

since $c > 0$ and $\Theta_j > 0$ (because Θ_j is the multiplier on the budget constraint). These two implications cannot simultaneously hold, so exports cannot be optimal and therefore

$$d_j^- = 0.$$

The physical constraint then reduces to

$$d_j - g_j = d_j^+ \quad \Rightarrow \quad g_j \leq d_j,$$

though the household may still import whenever consumption exceeds generation. Thus under no net metering solar owners never generate beyond their own consumption, but they may consume more than they generate. Therefore, when net metering is removed, exports have zero value and therefore cannot generate surplus for the solar household. Under no net metering solar owners optimally size their systems according to their own consumption, and no rent can arise from generation beyond on-site demand.

Let the utility's revenue requirement be

$$\pi^+ D^+ = N + cD^+,$$

where D is total retail demand and N and c are fixed and marginal network costs as before. Any admissible price must satisfy

$$\pi^+ = \frac{N + cD^+}{D^+}.$$

Because the utility is a price-setting monopolist subject to this equality constraint, the price is determined by demand and bound by the zero profit requirement. The utility cannot choose a price above break-even and thus cannot extract rent from either solar or non-solar customers.

Let π^* be the price under the unrealistic competitive benchmark, π_{noNEM} denote the price under no

net metering, and π^{NEM} denote the price under NEM 1.0. From this, we see

$$c \leq \frac{N + cD^+}{D^+} \leq \frac{N + cD^+}{D^+ - D^-}$$

$$\pi^* \leq \pi_{noNEM} \leq \pi^{NEM}$$

This is consistent with Shitovitz's (1973) characterization: the gains of "large" traders (here, solar prosumers) are exactly equal to the losses from small traders Shitovitz (1973). Here, the gains are set to zero by policy (the utility is a monopoly, so would have acted as a large trader but is limited by the zero profit constraint), so no surplus cannot be exploited. However, deadweight loss is present since the utility must charge average instead of marginal cost to recoup its spending, so we do not reach the efficiency of the competitive benchmark.

However, exports are capped- solar owners have no incentive to generate beyond their own consumption, since there is no market for this excess, thus will naturally size their systems smaller and generate less. This model also ignores any price reductions from adding solar to the grid (a relatively cheap form of energy), and positive externalities generated from solar.

Conclusion

Current NEM schemes widely in use are not 1:1 ratios of a constant retail rate, and exploration of rents and solar adoption under those schemes could be extremely helpful in guiding policymakers. NEM 2.0 iterates upon NEM 1.0 by adding an interconnection cost and a fixed cost. In addition, NEM 2.0 mandates the use of time-of-use rates, which charges consumers based on whether they produce/export during peak and off-peak periods California Public Utilities Commission (2022). The current model does not take into account time periods and incentivizing solar exports during peak periods, and one could imagine that including time periods and time-of-use rates 1) causes off-peak solar exports to reduce and 2) could provide a strong case for the use of batteries, another aspect of solar systems that would be interesting to explore further. Batteries and time-of-use rates would allow solar generation to primarily be exported during peak periods, potentially lowering prices for all consumers NEM 3.0 changes the export rate from the retail rate to the value of the solar export to the grid using the Avoided Cost Calculator, a proxy for marginal cost of electricity for the utility California Public Utilities Commission (2022). This could resolve the issue with the missing market, but doesn't account for solar's significant positive externalities. Finally, a better exploration of utility requirements and incentives (revenue balancing does not perfectly explain all the decisions utilities have made, for example choosing a relatively low fixed cost during the NEM 2.0 scheme) could help guide the model and make it more relevant.

Bibliography

By building a consumer and utility model and applying the budgetary exploitation theorem to the results, we can see equity problems arising from NEM 1.0 due to economic rents being extracted from non-solar consumers by solar consumers. We also see that this scheme encourages adoption of solar through a feedback mechanism: as price rises, solar exports and thus generation increase, causing the price to rise further. Taking these two results into account, in an extreme case NEM 1.0 could cause rising prices that cause more consumers to shift into solar until only the consumers whose budget does not allow solar are left, and they are left paying all of the grid's fixed cost. On the other hand, we see that no net metering tariff at all leads to a missing market for solar exports, so any positive value that solar could bring to the grid or to other consumers (or even to agents outside of this market, as solar is generally beneficial) will not be realized since solar prosumers would never produce more than they themselves consumed. This problem underscores the complexities behind energy policy, and the need to better understand how the grid responds to distributed energy resources in order to sustainably move to a greener future.

References

Argosino, F. J. and Knittel, C. R. (2025). Renewables and electricity affordability: Untangling correlation from causation. Working Paper WP-2025-22, MIT CEEPR. Working Paper, MIT Center for Energy and Environmental Policy Research.

Barbose, G. (2024). One year in: Tracking the impacts of nem 3.0 on california's residential solar market.

California Public Utilities Commission (2022). Decision revising net energy metering tariff and subtariffs (r.20-08-020, decision 22-12-056). Decision / Technical Report 22-12-056, California Public Utilities Commission. Order Instituting Rulemaking to Revisit Net Energy Metering Tariffs.

Climate Central (2025). A decade of growth for u.s. solar and wind — climate central. Technical report.

Comello, S. and Reichelstein, S. (2017). Cost competitiveness of residential solar pv: The impact of net metering restrictions. *Renewable and Sustainable Energy Reviews*, 75:46–57.

Gabszewicz, J. J. and Shitovitz, B. (1992). Chapter 15 the core in imperfectly competitive economies. volume 1 of *Handbook of Game Theory with Economic Applications*, pages 459–483. Elsevier.

Gong, A., Brown, C., and Adeyemo, S. (2017). The financial impact of california's net energy metering 2.0 policy examining the effects of non-bypassable charges with load profiles and systems designed in aurora. Technical report, Aurora Solar Inc.

Shitovitz, B. (1973). Oligopoly in markets with a continuum of traders. *Econometrica: Journal of the Econometric Society*, pages 467–501.

Incumbent Re-election Probability and Vote Share: The Effect of Hurricanes on Florida's Election Outcomes

Elizabeth Wright

May 5, 2025

Abstract

Natural disasters such as hurricanes can impose severe disruptions on local communities, potentially shaping electoral outcomes through shifts in voter sentiment and behavior. This paper investigates the effect of hurricane exposure on incumbent re-election probability and vote share in the state of Florida—a region highly vulnerable to tropical storms—using a staggered difference-in-differences approach. Drawing on county-level election data from 1996 to 2024 and storm path data, I examine changes in incumbent vote share, voter turnout, absolute change in incumbent vote share, and margin of victory. I find that hurricane exposure is associated with lower voter turnout and, in some specifications, higher incumbent vote share, though the incumbent effect weakens once party affiliation is controlled for. These results suggest that hurricane-induced disruptions and exposure to natural disasters may influence incumbent re-election outcomes and voting behavior in Florida.

1 Introduction

Natural disasters serve as exogenous shocks that may shape election results, as voters can update their voting preferences and expectations on policies and politician performances in the wake of a catastrophe. It therefore becomes essential for politicians to provide an active response to address unexpected damages. Studying the political consequences of these shocks may provide insights into voter behavior and mechanisms of accountability, as public assessment of political leadership is more evident during crisis events. (Masiero & Santarossa, 2021) Voters learn about the competence of politicians when these shocks happen. The occurrence of shocks such as natural disasters requires politicians to provide effective responses and strategies to address damages, which may in turn disclose information on their responsiveness.

While existing research has found electoral consequences following events such as earthquakes, floods, and other extreme weather events, most of this work has focused on rare or catastrophic disasters. However, less attention has been given to settings where natural disasters are both frequent and expected, such as hurricanes in Florida. In high-frequency disaster environments, the electoral consequences of shocks may be attenuated by voter adaptation, expectations, or institutional normalization of disaster response. If voters treat hurricanes as routine rather than extraordinary events, they may no longer serve as credible tests of political accountability.

Florida, which experiences more hurricanes than any other U.S. state, offers a natural setting to explore this question. The state's frequent exposure to hurricanes and its political competitiveness makes it an ideal case to study how voters respond to disaster events when

those events are recurrent and institutionalized. Given the state's political competitiveness, small shifts in voter behavior may have significant implications for election outcomes. Historically, some hurricanes have caused significant disruptions, affecting registration processes and turnout on Election Day. In one notable case, residents in counties hardest hit by Hurricane Michael in 2018 were 7% less likely to vote than those in neighboring counties (Milton, Helene). However, it remains unclear whether such effects are systematic, and whether they translate to changes in incumbent support or re-election prospects.

Thus, my paper provides empirical evidence to address this question and investigates the effect of the occurrence of hurricanes in Florida on election outcomes, specifically incumbent candidates' re-election probability and vote share, and which channels potentially drive these effects. Using a difference-in-differences design, I find that hurricane exposure is associated with a significant but small decrease in voter turnout and an increase in the vote share obtained by incumbent candidates in an election. I find no effect on the absolute change in incumbent vote share from the previous election they participated in, their margin of votes conditional on victory, or their decision to run again in the next election.

The remainder of my paper is structured as follows. Section 2 provides an overview of the electoral rules and election setting in Florida. Section 3 describes the data and discusses the measurement of the occurrence of hurricanes. Section 4 describes the empirical methodology and key assumptions of the difference-in-differences analysis. Section 5 discusses the main results of the analysis, with Section 6 concluding with broader implications and possible future directions.

2 Background and Related Literature

2.1 Institutional Setting

My analysis focuses on general, primary, and presidential preference primary elections in Florida from 1996 to 2024. Florida follows the U.S. federal election framework, conducting closed primaries for both presidential and state offices, meaning only registered party members may vote in their party's primaries. The state operates under a two-party system, with the governorship historically alternating between Democrats and Republicans. Governors serve four-year terms and may not serve more than two consecutive terms.

In presidential election years, Florida holds two separate primaries: the Presidential Preference Primary to select party nominees for president, and the State Primary for other offices. Gubernatorial elections occur in even-numbered, non-presidential years. (Sandoval & Walsh, 2024) General elections include races for federal offices, state attorney, legislature, judiciary, special districts, constitutional amendments, and, during gubernatorial years, the governor and cabinet. Primary elections cover contests for federal legislative offices, state attorney, and state legislative seats. Florida's 67 counties administer elections using paper ballots cast at voters' assigned precincts. Early voting is available at designated sites within each county and ends the weekend before Election Day. Specific dates and times are set by each county's Supervisor of Elections. (Florida Department of State)

During the time frame of this study, there were eight presidential elections and seven gubernatorial elections, along with various races across districts as aforementioned. I exclude results from special elections in this time period as they are irregularly scheduled and voter turnout statistics are limited. These elections may be driven by unique circumstances (such

as resignations or deaths in office) that are not comparable across counties or over time, thereby introducing potential confounding variation to the analysis.

2.2 Related Literature

There have been several studies that address the question of voter behavior at elections after disasters occur (e.g. Gallego, 2018; Achen & Bartels, 2004; Malhotra & Kuo, 2009). However, findings across the literature are mixed. The conventional theoretical view is that, with voters acting rationally, exogenous events such as hurricanes should not affect incumbents' electoral outcomes.

However, some (Ashworth et al. 2018) have argued that exogenous shocks can affect incumbents' electoral victories or losses as voters may observe how politicians responded to or prepared for a natural disaster–providing an opportunity to learn about the quality of the incumbent. Gasper and Reeves (2011) analyze the impact of extreme rainfall in the U.S. on electoral outcomes at the county and federal level, finding that voters are able to identify politicians who are responsible for a good or bad response to catastrophe occurrence, and thus reward or punish them accordingly. Conversely, Achen and Bartels (2017) argue that voters take out their frustrations at elections when disasters occur and blame incumbents if there is some reason to believe they are accountable for disaster response.

3 Data

3.1 Election Samples and Variables

To analyze the impact of hurricane occurrence on electoral outcomes, I merge a number of data sources with information aggregated at the county level. Data on county-level electoral outcomes from general, primary, and presidential preference primary elections in Florida are provided by the Florida Division of Elections' and are available on the online historical election archive. The data consists of election dates, candidates, candidate party affiliation, county names, votes cast for candidates, and voter turnout by county for each election, providing a repeated cross-section of election outcomes over time, covering every other year between 1978 to the present. I filter for a subset of the data from the years 1996-2024 to cover relevant election years, as voter turnout information is incomplete for years prior to 1996.

The county-level election data set includes around 94,000 observations relative to each race in each election. The unit of observation in the main analysis is a county-election-race observation. Each observation corresponds to an incumbent candidate competing in a given county and election year. Because multiple offices and races are observed within counties and years, the panel consists of repeated county-election-race observations rather than a single county-year observation. Using these data, I define several electoral outcome measures. The first is a dummy variable equal to one if an incumbent candidate is reelected, and zero otherwise. I define the share of votes received by the incumbent computed as the proportion of votes cast for the incumbent relative to the total number of valid votes. The reelection dummy is a dichotomous measure, assessing the success of the incumbent in

the election, but does not suggest if and how much the incumbent gains or loses support from the previous electoral cycle. To circumvent this issue, I include the absolute change in vote share, which measures the absolute difference in incumbent vote share relative to the previous election. Additionally, I consider the incumbent's margin of victory conditional on winning, the difference in vote share between the winning candidate and the runner-up. The subsequent analysis pools incumbent races across offices into a single estimation sample, so the estimated coefficients should be interpreted as average effects across the set of elections observed in the data.

I use several other data sources to complete the data set. Data on hurricane occurrence are provided by the International Best Track Archive for Climate Stewardship (IBTrACS) and is described in further detail in section 3.2. I use repeated cross-sectional population data from 1996 to 2024 provided by the American Community Survey (5-year estimates) to acquire county-level demographic characteristics, such as age, sex, race, education, and employment status to use as controls for each county in separate election years. Using the FEMA National Risk Index (NRI) dataset, I take the county-level risk assessment for hurricanes and other natural disasters (floods, wildfires, avalanches, etc.). The dataset is compiled by FEMA using information from multiple federal agencies, including NOAA, the U.S. Census Bureau, and the U.S. Geological Survey, integrating historical storm records, economic impact estimates, and demographic data to generate a composite hurricane risk score for each county. The most recent release of the NRI, updated in 2021, is based on historical data and reflects baseline hurricane vulnerability measures by county rather than annual hurricane risk variation. I match county-level risk scores to the election data serving as fixed effects to account for differences in counties' baseline exposure to hurricanes.

3.2 Measurement of Hurricane Occurrences

Between 1996 and 2024, 64 out of 67 Florida counties were hit at least once by a hurricane. The IBTrACS dataset provides time series tracking historical paths and characteristics of tropical cyclones. Each observation corresponds to a storm tracking point, which includes the storm's latitude, longitude, wind speed, pressure, and other meteorological details at given 3-hour intervals.

In determining whether a county experienced a hurricane, I draw on the NOAA's definitions of hurricane landfall and strike. Landfall is defined as the intersection of the surface center of a tropical cyclone with a coastline. However, as the strongest winds in a cyclone are often not located at the center, damaging winds may reach inland areas even if the eye of the storm does not make landfall. The NOAA further defines a hurricane strike as occurring when a location passes within a "strike circle," a 125 nautical mile diameter circle centered slightly to the right of the storm's center path. (National Hurricane Center) This accounts for the asymmetrical nature of hurricane force winds. Based on this concept, I approximate the hurricane impact zone as a 62.5-mile radius (half the strike circle diameter) around the storm center.

I filter for the subset of the data from the years 1996-2024 and use QGIS to map latitude and longitude points of each relevant hurricane path to a specific county in Florida. I identify counties within a radius of 62.5 miles of hurricane landfall and assign treatment status accordingly, as the effects of a tropical storm can be observed before the eye of the storm makes landfall. In the data set, I match the hurricane that hit most recently before the election to the county in which it hit. The quasi-random trajectory of hurricanes makes

the timing and location of storm exposure exogenous to county-level political conditions.

Tables 1 and 2 in the appendix present summary statistics of the main variables used in the analysis by office type for Florida elections between 1996 and 2024, at the county level. I include four outcome variables: total votes for candidate, the absolute change in vote share of the incumbent from the previous election (as a percentage), voter turnout, and the incumbent's margin conditional on winning (percentage).

Figure 1 shows the mean vote share of the election winner by office type for each of ten offices. I restrict the sample to the winners of primary and general elections between 1996 and 2024.

4 Methodology

I employ a staggered difference-in-differences strategy to estimate the causal effect of hurricane exposure on incumbent reelection probability and vote share. The validity of this approach relies on several key assumptions.

Most importantly, it requires a parallel trends assumption: in the absence of hurricanes, treated and control counties would have exhibited similar trends in electoral outcomes. I also treat the timing of hurricanes as plausibly exogenous to political dynamics, conditional on the included fixed effects and controls, although this assumption may be complicated by pre-existing vulnerabilities and risk scores across counties that influence both tropical storm damage and voter behavior.

Additionally, I assume voters and politicians do not adjust their behavior accordingly in anticipation of an impending hurricane, ensuring any electoral effects result from the storm

itself. This assumption and the point made prior appear to be plausible as the sudden and unexpected nature of tropical storms makes it unlikely that electoral strategies or turnout decisions are consistently shaped by the prospect of a hurricane. Lastly, I assume there are no spillover effects; voters in neighboring, unaffected counties are not indirectly influenced by hurricane relief efforts or media coverage of the disaster. It is imperative that these assumptions be close to true for the difference-in-differences estimates to be interpreted as causal effects.

The treatment group consists of counties that were directly affected by hurricanes, identified using storm path data, while the control group consists of counties not exposed to a hurricane during the relevant election period. I model hurricane exposure as a time-varying treatment indicator rather than an absorbing treatment. Specifically, the treatment indicator equals 1 if a county experienced a hurricane in the current election year or within the previous fourteen months, and 0 otherwise. Counties therefore enter treatment only for elections occurring within this recent post-hurricane window and revert to the untreated state once fourteen months have passed without a new storm. This specification is intended to measure the effect of recent hurricane exposure on electoral outcomes when disruption and voter perceptions are most likely to remain politically relevant.

The strength of this D-i-D approach lies in the fact that hurricane occurrence, conditional on time and county fixed effects, are random exogenous shocks. However, while hurricane occurrence is random over time, the assignment of a county to the group of treated counties may not be random due to heterogeneous exposure to hurricane risk due to inherent characteristics of Florida geography. This issue may be further amplified if risk preferences shape vote decisions and voters living in areas with high tropical storm risk have different risk

preferences as compared to voters in lower risk areas. Therefore, comparing struck counties with the universe of unaffected counties may confound the analysis due to unobserved heterogeneity varying over time among counties and electoral cycles. To address this concern, I control for county and year fixed effects, as well as hurricane risk scores as mentioned prior. I seek to estimate

$$Y_{ict}^{(k)} = \alpha + \theta_1 Hurricane_{ct} + \theta_2 Post_t + \beta(Hurricane_{ct} \times Post_t) + \gamma X_{ict} + \delta_c + \lambda_t + \varphi_c + \epsilon_{ict},$$
$$k \in [1, 4]$$

where Y_{ict} is the electoral outcome in county c at time t. The key independent variable, $Hurricane_{ct} \times Post_t$, is the interaction term that captures whether a hurricane affected county c in the election cycle after the disaster occurred. The post variable is an event time indicator equal to 1 if county c experienced a hurricane in the current election or within fourteen months prior, and 0 otherwise. It is set to 0 if no hurricane occurred in the county within the past fourteen months, including the current election year. This ensures that counties only enter the post-treatment period if they were recently affected by a storm and exit it after fourteen months of no hurricane activity. Counties missing hurricane exposure in prior years are assigned zeros until their first hurricane event. I define the post period as a fourteen month window since historical data on hurricanes indicate that the average primary recovery period for hurricanes of similar size and magnitude to hurricanes Katrina and Sandy is fourteen months after landfall. (BuildFax)

Under the identifying assumptions discussed above, the coefficient of interest β measures the estimated effect of hurricane exposure on electoral outcomes. The model includes a vector of time-varying controls, X_{ict}, such as sociodemographic composition and economic

indicators. County fixed effects δ_c account for time-invariant differences across counties, such as historical voting behavior. Additionally, year fixed effects λ_t are included to control for election cycle trends that may influence vote shares across the state and hurricane risk fixed effects (derived from FEMA's National Risk Index) φ_c control for baseline storm risk scores across counties.

I estimate the model on three outcome variables for $k \in [1, 4]$: $Y_{ict}^{(1)}$, the incumbent vote share, measured as a variable indicating the percentage of votes received by the incumbent in a given election; $Y_{ict}^{(2)}$, the total voter turnout at the election; $Y_{ict}^{(3)}$, the absolute value of the change in vote share of the incumbent from the previous election as a percentage; and $Y_{ict}^{(4)}$, the incumbent candidate's margin of votes conditional on winning as a percentage of votes. I cluster standard errors at the county level to account for serial correlation in election outcomes within counties over time.

5 Results

Tables 3-6 present the results of the regression model on the four outcomes variables of the incumbent candidate. For regression (1), I run a baseline difference-in-differences with county and election year fixed effects. For regression (2), I add in sociodemographic controls and fixed effects for a candidate's political party affiliation in regression (3). The main coefficient of interest from regression (2) across the outcome variables are $\beta = -8.72, 5.28, 1.31, 1.44$. The estimates indicate that hurricane exposure in counties is associated with a small but statistically significant decline in voter turnout. Hurricane exposure is also associated with an increase in the vote share of an incumbent candidate performing in an affected area,

although this effect is less robust across specifications. Across the remaining outcome variables, I find no statistically significant effects on the absolute change in incumbent vote share and the incumbent candidate's margin of votes conditional on victory.

An explanation for this result is selective turnout. Hurricanes may reduce participation by raising the costs of voting or disrupting daily life, while those voters who still turn out may differ systematically from the broader electorate. At the same time, recent disaster exposure may generate a modest rally-around-incumbent effect among participating voters. This could produce a combination of lower overall turnout and slightly higher incumbent vote share in affected counties.

A possible explanation for the null outcomes is that hurricanes may affect whether incumbents choose to run for reelection, and those that choose to re-run are those that are very popular or in counties where there was minimal damage from the hurricanes. As such, I estimate the effect of the hurricane strike on a dummy for whether the incumbent is running, and find no effect of hurricane exposure on their decision to run. The results are shown in Table 7.

Therefore, these results suggest that the electoral effects of hurricane exposure are limited in scope. The most consistent finding is a decline in voter turnout in affected counties. The evidence that hurricanes increase incumbent support is more sensitive to specification; incumbent vote share rises in specification (2), but this effect weakens once party fixed effects are included. The findings provide only limited evidence that voters systematically reward incumbents following hurricane exposure.

To supplement this analysis, I control for party affiliation of the incumbent in each county's election to account for potential partisan differences in disaster response or voter

behavior, shown in regression (3) in Tables 3 through 7. The inclusion of party fixed effects lowers the magnitude of the coefficient of interest on voter turnout, suggesting a 6.39 percentage point decrease in turnout after a hurricane strike. For incumbent vote share however, the estimated effect becomes much smaller and statistically insignificant once party fixed effects are added, suggesting that partisan composition may explain part of the positive association observed in the baseline specification. All other outcomes are not meaningfully altered otherwise, with the effect on incumbent vote share rendered null, suggesting that the observed effects of hurricane exposure are not driven by political party alignment alone.

Because political effects of disasters may depend on how recently they occurred, I next examine whether hurricane proximity to Election Day alters electoral responses. A hurricane that occurs closer to an election may be more salient to voters, whereas storms that occurred several months earlier may have weaker political effects. To further investigate the role of timing, I estimate a second set of regressions using three mutually exclusive time indicators: hurricanes occurring within 30 days before the election (storm_time_1), hurricanes occurring between 30 and 90 days before the election (storm_time_2), and hurricanes occurring more than 90 days before the election (storm_time_3).

The results from this set of storm proximity regressions are shown in Tables 8 through 12. The strongest evidence appears for hurricanes occurring within 30 days of the election. In that window, hurricane exposure is associated with statistically significant increases in turnout, incumbent vote share, incumbent margin conditional on victory, and the incumbent's decision to run again. In contrast, hurricanes in the second and third treatment windows show no significant detectable effects. This suggests that only recent exposure to a storm, particularly within the month leading up to the election, has measurable electoral

consequences.

These timing results suggest that the average negative turnout effect identified in the main specification may mask heterogeneity due to recency. Hurricane exposure appears most politically mobilizing when a storm occurs immediately before the election, while exposure over longer windows may be affected by the storm recovery process. This suggests that voters do respond electorally to recent hurricane exposure, particularly when the storm occurs within 30 days of the election. The increase in both turnout and incumbent vote share points to heightened political attention immediately following a storm. However, these timing effects weaken once party fixed effects are included, suggesting that some of the short-run incumbent advantage may reflect partisan sentiments rather than hurricane exposure alone. A plausible cause for this may be disaster normalization. This normalization likely lowers the impact of hurricanes in the political domain, especially over the months between storm impact and Election Day. Even in cases of severe damage, voters may attribute outcomes to nature or bureaucratic routine rather than political accountability. Disaster management may be overshadowed by party identification, further undermining any potential electoral response, as voters may rely more on party affiliation of an incumbent than on evaluations of disaster management. While hurricanes disrupt daily life, they may not significantly disrupt the political preferences of voters in a state where such disasters have become habitual.

6 Conclusion

Natural disasters can provide voters with information about incumbent responsiveness by forcing elected officials to manage visible disruptions and recovery efforts. This paper exam-

ines whether those electoral effects persist in a setting where disasters are frequent rather than exceptional.

I find limited but meaningful electoral effects of hurricane exposure in Florida. Hurricanes are associated with reduced voter turnout and modest increases in incumbent vote share. However, these impacts fade when accounting for party affiliation, and no statistically significant effects are detected on outcomes like incumbent margins or candidacy decisions. In counties which experience hurricanes frequently, residents may be less acute to the consequences of damage as they are inherently more prepared for these disruptions. In a relatively high-frequency storm environment, voters may treat hurricanes less as tests of political competence and more as routine events mediated by existing partisan attachments.

As such, it is important to consider further extensions beyond these results. Firstly, identifying channels of influence, such as media coverage, emergency response effectiveness, and post-disaster aid distribution and spending may clarify why hurricanes do or do not sway voter behavior in Florida. Studies such as Masiero and Santarossa (2021) suggest that media visibility and public expenditure patterns following disasters can significantly shape electoral support—especially when politicians capitalize on post-disaster attention. Incorporating variation in storm severity (wind speed or FEMA disaster declarations) may uncover heterogeneous effects obscured in average treatment estimates. Aid delivery timing could also be investigated, comparing electoral outcomes in counties that received prompt relief to those where assistance was delayed or mismanaged.

Additionally, my identification strategy relies on comparing affected counties to the full sample of unaffected counties. While controlling for county fixed effects and baseline hurricane risk helps address potential bias, alternative matching strategies, such as restricting the

comparison group to demographically similar counties, may better isolate treatment effects. Therefore, while hurricanes represent clear exogenous shocks, their electoral consequences in Florida appear modest. Understanding under what conditions disaster exposure might alter voter behavior remains a compelling question for future research to understand the drivers of voter decisions on political accountability.

7 References

Achen, C.H., Bartels, L.M., 2004. Blind retrospection: electoral responses to droughts, floods, and shark attacks. In: Democracy for Realists: Why Elections Do Not Produce Responsive Government. Princeton University Press, pp. 116-145.

Ashworth, Scott, et al. "Learning about Voter Rationality." American Journal of Political Science, vol. 62, no. 1, 2018, pp. 37–54.

BuildFax. Hurricane Recovery Study. Inman, Oct. 2017, https://webassets.inman.com/wp-content/uploads/2017/10/BuildFax-Hurricane-Recovery-Study.pdf.

Florida Department of State. Election Dates. Florida Department of State, https://dos.fl.gov/elections/for-voters/election-dates/.

Florida Department of State. Ways to Vote. Florida Elections, www.myfloridaelections.com/voting-elections/ways-to-vote.

Gallego, J., 2018. Natural Disasters and clientelism: the case of floods and landslides in Colombia. Elect. Stud. 55, 73-88.

Giuliano Masiero and Michael Santarossa, (2021), Natural disasters and electoral outcomes, European Journal of Political Economy, 67, (C).

Malhotra, N., and Kuo, A. G. (2009). Emotions as Moderators of Information Cue Use: Citizen Attitudes Toward Hurricane Katrina. American Politics Research, 37(2), 301-326.

Milton, Helene. "Hurricane, Florida Voting Turnout." Tampa Bay Times, 4 Nov. 2024, www.tampabay.com/news/florida-politics/elections/2024/11/04/helene-milton-hurricane-

florida-voting-turnout/.

National Hurricane Center. Glossary. National Oceanic and Atmospheric Administration, https://www.nhc.noaa.gov/aboutgloss.shtml.

Sandoval, H.H. and Walsh, A.N. (2024) Sentiments and spending intentions: evidence from Florida. Economic Inquiry, 62(3), 1046–1073.

8 Appendix

Table 1: **Descriptive Statistics of Florida Elections by Office Type, 1996-2024**

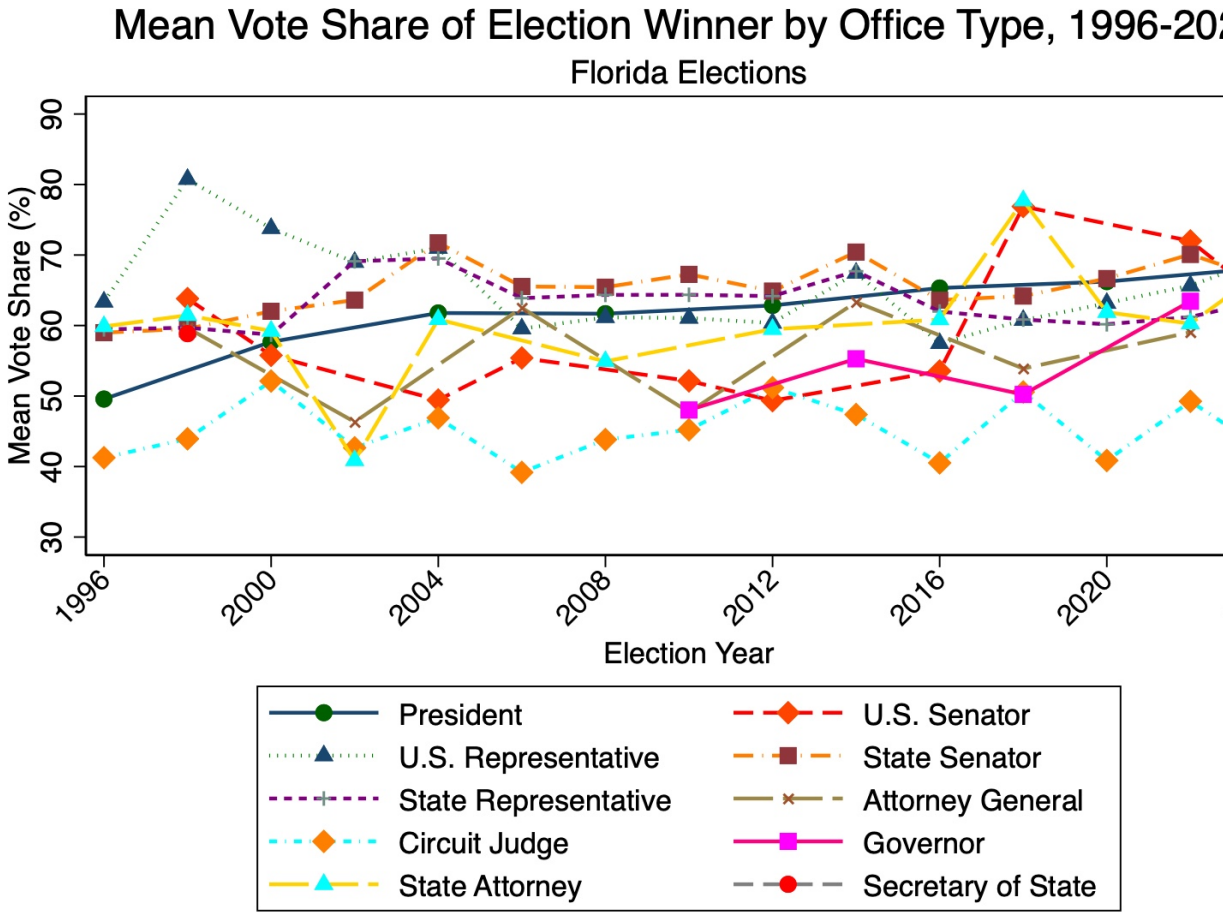
Office Type	Total Votes			Abs. Change in Vote Share (%)			Obs.
	Mean	Min	Max	Mean	Min	Max	
President of the U.S.	61,879	0	1,156,816	9.59	0	77.46	18,733
	(122,506)			(14.65)			
U.S. Senator	76,311	26	1,076,114	13.57	0	63.85	9,983
	(128,985)			(14.52)			
U.S. Representative	44,444	5	411,593	14.45	0	75.96	7,052
	(61,861)			(13.07)			
State Senator	41,615	24	311,218	18.62	.05	79.51	2,790
	(53,519)			(15.59)			
State Representative	87,310	2,579	486,468	16.58	0	92.86	7,740
	(24,886)			(15.04)			
Attorney General	50,622	256	780,329	22.95	.04	71.47	2,948
	(88,715)			(14.66)			
State Attorney	86,301	112	918,913	14.44	.007	50.29	518
	(163,701)			(13.52)			
Secretary of State	33,042	27	354,316	6.84	.024	21.28	268
	(63,126)			(5.15)			
Governor	68,857	271	799,024	14.81	0	68.03	3,953
	(109,357)			(15.17)			
Circuit Judge	215,032	990	1,866,221	21.56	.054	85.43	3,241
	(328,500)			(14.85)			

Notes: **Table 1** reports summary statistics by office type for Florida elections from 1996-2024 (primary, general, and presidential pref. primary). Standard deviations are shown in parentheses below means. Total Votes represents the number of votes cast in each race. Absolute Change in Vote Share measures the percentage point change in incumbent vote share from previous election. Observations are at the county level.

Table 2: **Descriptive Statistics of Florida Elections by Office Type, 1996-2024**

Office Type	Vote Share (%)			Incumbent Margin (%)			Obs.
	Mean	Min	Max	Mean	Min	Max	
President of the U.S.	8.12	0	95.72	39.81	0.05	93.64	18,733
	(17.49)			(27.69)			
U.S. Senator	13.42	0	95.85	31.89	0.06	91.69	9,983
	(21.45)			(22.83)			
U.S. Representative	32.61	0	100	37.53	0	100	7,052
	(26.93)			(23.99)			
State Senator	42.29	0	100	36.01	0.05	100	2,790
	(24.64)			(25.25)			
State Representative	41.17	0	100	33.73	0	100	7,740
	(24.04)			(27.12)			
Attorney General	31.82	0.386	92.04	27.35	0.29	84.08	2,948
	(21.57)			(18.85)			
State Attorney	46.14	0.007		99.98	27.70	0.31	99.94
	(18.56)			(26.24)			
Secretary of State	50	25.7	74.3	19.73	0.19	48.6	268
	(10.6)			(11.88)			
Governor	13.56	0	91.62	25.44	0	64.81	3,953
	(20.19)			(17.4)			
Circuit Judge	24.56	0.1	99.9	16.63	0.01	99.81	3,241
	(18.3)			(17.16)			

Notes: **Table 2** reports summary statistics by office type for Florida elections from 1996-2024 (primary, general, and presidential pref. primary). Standard deviations are shown in parentheses below means. Vote share measures the percentage of votes received by a candidate in an election race. Incumbent margin represents the percentage point difference between winner and runner-up when the incumbent wins. Observations are at the county level.



Notes: **Figure 1** shows the mean vote share of the election winner by office type. The winner vote share
percentage of votes received by the winner of each election race. I restrict the sample to general and
covering the years 1996-2024. I exclude observations where total votes cast were missing or zero to avoid
in vote share calculations. Each observation corresponds to a winning candidate's performance in a spec

Table 3: **DID Regression Results on Voter Turnout (%)**

	(1)	(2)	(3)
Hurricane	8.04***	10.7***	7.26***
	(0.93)	(1.06)	(0.88)
Post	27.7***	3.39***	2.65***
	(0.59)	(1.22)	(0.98)
(Hurricane$_{ct}$ × Post$_t$)	-6.38***	-8.72***	-6.39***
	(0.92)	(1.06)	(0.9)
Constant	53.89***	45.48***	44.15***
	(0.08)	(4.95)	(4.5)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	80,380	40,367	40,367
R-squared	0.32	0.31	0.5

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). Models include fixed effects for county (67 categories), election year (15 categories for (1), 7 categories for (2) and (3)), hurricane risk (67 categories), and party affiliation (20 categories). The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 4: **DID Regression Results on Incumbent Vote Share (%)**

	(1)	(2)	(3)
Hurricane	0.52	0.14	0.11
	(1.33)	(2.01)	(1.78)
Post	-0.82	-1.96**	-0.27
	(0.51)	(0.78)	(0.64)
(Hurricane$_{ct}$ × Post$_t$)	3.25**	5.28**	1.54
	(1.54)	(2.28)	(1.93)
Constant	47.08***	47.88***	46.57***
	(0.04)	(4.39)	(4.03)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	26,037	15,999	15,999
R-squared	0.02	0.03	0.27

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). Models include fixed effects for county (67 categories), election year (15 categories for (1), 7 categories for (2) and (3)), hurricane risk (67 categories), and party affiliation (8 categories). The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 5: **DID Regression on Absolute Change in Incumbent Vote Share (%)**

	(1)	(2)	(3)
Hurricane	-1.93***	-2.34***	-0.99
	(0.51)	(0.69)	(0.62)
Post	0.18	0.395	0.63**
	(0.25)	(0.39)	(0.3)
(Hurricane$_{ct}$ × Post$_t$)	0.304	1.31	0.335
	(0.62)	(0.88)	(0.68)
Constant	12.74***	14.92***	15.21***
	(0.03)	(2.39)	(2.42)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	26,036	15,998	15,998
R-squared	0.05	0.02	0.14

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). Models include fixed effects for county (67 categories), election year (15 categories for (1), 7 categories for (2) and (3)), hurricane risk (67 categories), and party affiliation (8 categories). The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 6: **DID Regression on Incumbent's Margin Conditional on Victory (%)**

	(1)	(2)	(3)
Hurricane	-0.045	-2.59	-1.35
	(1.67)	(1.85)	(1.63)
Post	2.03**	1.54	1.32
	(0.90)	(1.12)	(1.15)
(Hurricane$_{ct}$ × Post$_t$)	-0.93	1.44	1.19
	(2.02)	(2.2)	(2.06)
Constant	29.92***	44.51***	43.92***
	(0.1)	(10.67)	(10.88)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	13,308	8,096	8,096
R-squared	0.06	0.1	0.12

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). Models include fixed effects for county (67 categories), election year (15 categories for (1), 7 categories for (2) and (3)), hurricane risk (67 categories), and party affiliation (3 categories). The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 7: **DID Regression on Incumbent Decision to Run**

	(1)	(2)	(3)
Hurricane	0.057***	0.163***	0.034
	(0.018)	(0.03)	(0.024)
Post	-0.028***	-0.043***	0.008
	(0.006)	(0.009)	(-0.009)
(Hurricane$_{ct}$ × Post$_t$)	0.059***	0.029	-0.012
	(0.02)	(0.026)	(0.024)
Constant	0.32***	0.397***	0.363***
	(0.001)	(0.066)	(0.058)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	80,380	40,367	40,367
R-squared	0.09	0.03	0.27

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). Models include fixed effects for county (67 categories), election year (15 categories for (1), 7 categories for (2) and (3)), hurricane risk (67 categories), and party affiliation (20 categories). The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 8: **Regression of Storm Proximity to Election on Voter Turnout (%)**

	(1) Baseline	(2) With Controls	(3) With Party FE
Hurricane$_{ct}$ × Post$_t$ × storm_time_1			
0 1 0	2.80***	3.39***	2.65***
	(0.60)	(1.22)	(0.98)
1 0 0	7.99***	10.67***	7.23***
	(0.92)	(1.07)	(0.88)
1 1 0	8.68***	7.44***	5.44***
	(0.74)	(1.59)	(1.30)
1 1 1	-3.88	8.13***	2.75*
	(4.56)	(1.82)	(1.38)
Hurricane$_{ct}$ × Post$_t$ × storm_time_2			
1 1 0	-7.21***	-3.17*	-2.86**
	(1.11)	(1.80)	(1.36)
Constant	53.89***	45.46***	44.12***
	(0.07)	(5.04)	(4.56)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	80,380	40,367	40,367
R-squared	0.32	0.31	0.50

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). All models include fixed effects for county (67 categories), election year (15 categories in model 1, 7 categories in models 2-3), hurricane risk measures (67 categories), and party affiliation (8 categories). Coefficients are expressed in percentage points. The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 9: **Regression of Storm Proximity to Election on Incumbent Vote Share**

	(1) Baseline	(2) With Controls	(3) With Party FE
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_1			
0 1 0	-0.79	-1.94**	-0.26
	(0.52)	(0.78)	(0.65)
1 0 0	0.51	0.11	0.08
	(1.33)	(2.01)	(1.78)
1 1 0	3.91***	5.32***	2.67**
	(0.63)	(1.29)	(1.21)
1 1 1	2.43	3.79**	-0.52
	(2.37)	(1.68)	(1.4)
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_2			
1 1 0	-1.80	-2.91*	-1.96
	(1.09)	(1.71)	(1.47)
Constant	47.08***	47.64***	46.4***
	(0.04)	(4.49)	(4.09)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	26,037	15,999	15,999
R-squared	0.02	0.02	0.27

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). All models include fixed effects for county (67 categories), election year (15 categories in model 1, 7 categories in models 2-3), hurricane risk measures (67 categories), and party affiliation (8 categories). Coefficients are expressed in percentage points. The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 10: **Regression of Storm Proximity to Election on Absolute Change in Vote Share**

	(1) Baseline	(2) With Controls	(3) With Party FE
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_1			
0 1 0	0.17	0.40	0.63
	(0.25)	(0.39)	(0.30)
1 0 0	-1.93***	-2.34***	-0.99
	(0.51)	(0.70)	(0.62)
1 1 0	-1.83***	-0.24	0.33
	(0.38)	(0.50)	(0.45)
1 1 1	2.01*	0.66	1.62**
	(1.18)	(0.73)	(0.68)
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_2			
1 1 0	0.34	-0.65	-0.62
	(0.65)	(0.73)	(0.65)
Constant	12.74***	14.87***	15.17***
	(0.03)	(2.37)	(2.40)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	26,036	15,998	15,998
R-squared	0.05	0.02	0.14

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). All models include fixed effects for county (67 categories), election year (15 categories in model 1, 7 categories in models 2-3), hurricane risk measures, and other fixed effects as described in the regression output. Model 3 additionally includes party fixed effects (8 categories). Coefficients expressed in percentage points. The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 11: **Regression of Storm Proximity to Election on Incumbent Vote Margin Conditional on Victory**

	(1) Baseline	(2) With Controls	(3) With Party FE
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_1			
0 1 0	2.09**	1.57	1.35
	(0.89)	(1.13)	(1.15)
1 0 0	-0.07	-2.60	-1.35
	(1.67)	(1.85)	(1.63)
1 1 0	2.90***	3.16**	4.03**
	(1.01)	(1.50)	(1.66)
1 1 1	-1.55	7.07***	8.40***
	(3.01)	(2.19)	(2.18)
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_2			
1 1 0	-3.36**	-4.67**	-4.85**
	(1.54)	(2.05)	(2.08)
Constant	29.91***	44.19***	43.59***
	(0.10)	(10.75)	(10.95)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls		Yes	Yes
Party Affiliation Fixed Effects			Yes
Observations	13,308	8,096	8,096
R-squared	0.07	0.10	0.12

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). All models include fixed effects for county (67 categories), election year (15 categories in model 1, 7 categories in models 2-3), hurricane risk measures, and other fixed effects as described in the regression output. Model 3 additionally includes party fixed effects (3 categories). Coefficients expressed in percentage points. The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.

Table 12: **Regression of Storm Proximity to Election on Incumbent Decision to Re-run**

	(1) Baseline	(2) With Controls	(3) With Party FE
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_bin1			
0 1 0	-2.83***	-4.31***	0.79
	(0.60)	(0.89)	(0.95)
1 0 0	5.70***	16.25***	3.35
	(1.76)	(2.63)	(2.43)
1 1 0	9.52***	19.04***	5.78***
	(1.81)	(2.48)	(2.07)
1 1 1	5.13	17.56***	1.30
	(3.36)	(2.83)	(2.24)
$\text{Hurricane}_{ct} \times \text{Post}_t \times$ storm_time_bin2			
1 1 0	-0.99	-6.22**	-4.15*
	(2.09)	(2.77)	(2.16)
Constant	32.18***	39.69***	36.25***
	(0.08)	(6.79)	(5.90)
County & Year Fixed Effects	Yes	Yes	Yes
Hurricane Risk Fixed Effects	Yes	Yes	Yes
Sociodemographic controls	No	Yes	Yes
Party Affiliation Fixed Effects	No	No	Yes
Observations	80,380	40,367	40,367
R-squared	0.09	0.03	0.27

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses, clustered at county level (67 clusters). All models include fixed effects for county (67 categories), election year (15 categories in model 1, 7 categories in models 2-3), hurricane risk measures, and other fixed effects as described in the regression output. Model 3 additionally includes party fixed effects (20 categories). Coefficients expressed in percentage points. The estimation sample decreases in specification (2) as sociodemographic controls are not observed for all county-election-race observations.