Do Gender Quotas Strengthen the Rule of Law? Causal Evidence from Staggered Quota Adoption in Developing Countries

Abboskhan Azizov

Westminster International University in Tashkent

ORCID: 0009-0004-0698-2452

August 2026 — version 3

Abstract

Most developing countries entered the twenty-first century with weak legal institutions and have made limited progress since. Legislative gender quotas, the dominant policy for raising women's parliamentary representation, have demonstrably increased women's office-holding across the developing world. Whether quotas alter the substantive output of legislatures, beyond the descriptive composition of office, remains an open question. Drawing on the staggered adoption of binding legislative quotas across 102 developing countries between 2000 and 2022, I identify a positive causal effect on the rule of law of 0.145 standard deviations on the World Bank's Worldwide Governance Indicator, with point estimates between 0.124 and 0.211 across seven robustness specifications. The first stage confirms the policy operated as intended, raising women's parliamentary representation by 7.83 percentage points. The post-treatment path is monotonically increasing across the decade, with significance concentrated in the later post-treatment years and the magnitude largest in countries with the weakest baseline institutions. The pattern is consistent with a mechanism in which legislators structurally exterior to entrenched political networks accumulate procedural capacity over electoral cycles, displacing informal favours with formal legal procedures. Legislative gender quotas do more than reshape who holds office; they shift the legislative process toward stronger legal institutions over time.

Keywords: gender quotas; rule of law; synthetic difference-in-differences; women's representation; developing countries. **JEL:** J16, K42, O17, P26.

1. Introduction

Across the developing world, the average country in 2022 still scored half a standard deviation below the global mean on the World Bank's Rule of Law indicator[1], the same position it occupied two decades earlier. Contract enforcement remains uneven, judicial independence remains fragile, and the formal legal procedures that level the political-economic playing field remain crowded out by informal networks of patronage and favour exchange. The institutional-economics literature has documented this persistence and identified its deep historical determinants. What it has not identified is a contemporary policy lever, deployable within the lifetime of a single legislature, that demonstrably moves the rule of law in a positive direction.

The stakes of identifying such a lever are substantial. Rule-of-law levels correlate with economic growth, contract reliability, and the security of property rights (Acemoglu, Johnson, and Robinson 2001; North, Wallis, and Weingast 2009; La Porta, Lopez-de-Silanes, and Shleifer 2008); their absence constrains the development trajectory of low-income countries. Externally imposed legal reforms, a recurring instrument of donor strategies since the 1990s, have a mixed record at best, in part because they impose formal procedures without altering the underlying domestic balance between formal and informal political institutions (Helmke and Levitsky 2004). Whether a domestically driven, electorally legitimised lever exists is a question that matters for development policy beyond the academic literature.

Legislative gender quotas, adopted in more than 130 countries since the early 1990s (Hughes et al. 2017)[2], provide a natural candidate. Quotas have been studied extensively for their descriptive consequences: they have demonstrably raised women's parliamentary representation across the developing world (Tripp and Kang 2008; Krook 2009). They have also been studied for substantive consequences in specific domains, including maternal mortality (Bhalotra et al. 2023), public health spending (Clayton and Zetterberg 2018), corruption (Bauhr and Charron

---

[1]The Worldwide Governance Indicators are normalised to mean zero and unit standard deviation across the global country pool, not the developing-country pool. A 0.145 SD movement therefore corresponds to roughly five percentile points on the global distribution at the levels typical of our treated sample (see Section 6 for the percentile-rank translation).

[2]QAROT distinguishes constitutional, legislative, and party quotas. Following the convention in the staggered-adoption literature, this paper restricts treatment to implemented legislative quotas — quotas that have a binding numerical share and have been applied in at least one legislative election. Hughes et al. (2017) document the distinction.

2021), and educational outcomes (Beaman et al. 2012). The most direct legislative output, however, the production and enforcement of binding legal rules, has remained largely outside this empirical conversation. The rule of law is the variable that most cleanly captures the legislative function: it summarises the quality of contract enforcement, the impartiality of courts, the security of property rights, and the rule-bound conduct of public power, as measured by the World Bank's Worldwide Governance Indicator (WGI) for Rule of Law (Kaufmann, Kraay, and Mastruzzi 2010). If quotas alter the substantive output of legislatures, the rule of law is where the alteration should appear first and most cleanly.

This paper provides the first causal estimate of the effect of legislative gender quota adoption on the rule of law. I assemble a balanced panel of 102 developing countries observed annually from 2000 through 2022, of which 34 adopted binding legislative quotas in waves between 2002 and 2015 and 68 never adopted one. Identification combines the synthetic difference-in-differences (SDID) estimator of Arkhangelsky et al. (2021) with the staggered-adoption procedure of Clarke et al. (2023) and the covariate pre-processing routine of Kranz (2022). The principal outcome is the WGI rule-of-law index. SDID is supplemented by a two-way fixed-effects (TWFE) event study, a Goodman-Bacon (2021) weight diagnostic, a leave-one-country-out check, and a seven-specification robustness battery, each addressing a distinct threat to identification.

The headline finding is that legislative gender quota adoption causally improves the rule of law in developing countries. Across seven robustness specifications, the SDID estimate falls between 0.124 and 0.211 standard deviations on the WGI rule-of-law index, with the principal estimate at 0.145 (SE = 0.066, $p = 0.029$) and the covariate-adjusted estimate at 0.132 ($p = 0.043$). The principal estimate corresponds to a movement of roughly nine percentile points up the developing-country rule-of-law distribution for the median treated country, substantively comparable to the typical decade-on-decade improvement that developing countries achieve in cross-country institutional-quality measures.

The first stage confirms the policy operated as intended: quota adoption raised women's parliamentary representation by 7.83 percentage points ($p < 0.001$), enough to lift most treated legislatures across the conventional 15 to 30 percent critical-mass threshold within a single electoral cycle. The post-treatment path is monotonically rising, climbing from +0.055 in the

year of adoption to +0.199 a decade later, with the effect strongest in the later post-treatment window. The magnitude concentrates in countries with the weakest baseline institutions, consistent with the prediction that the policy operates most strongly where formal legal procedures have the most ground to gain on informal political networks.

The pattern aligns with the outsider mechanism developed by Bauhr and Charron (2021): legislators who arrive in office through quota mechanisms have weaker stakes in preserving the informal political bargains from which they have been historically excluded, and stronger stakes in formalising the legal procedures that would displace those bargains. Critical-mass theory complements the outsider account: a 7.83 percentage-point first stage is large enough to lift most treated legislatures past the conventional substantive-representation threshold within a single electoral cycle (Wängnerud 2009; Childs and Krook 2008). The two channels operate jointly. The accumulating temporal signature, in turn, is consistent with the slow displacement of substitutive informal institutions by formal legal procedures (Helmke and Levitsky 2004).

The paper makes three contributions. Substantively, it identifies a previously undocumented causal effect of quota adoption on legal-institutional quality, extending a literature that has so far been confined to fiscal and service-delivery outcomes (Bhalotra et al. 2023; Clayton and Zetterberg 2018; Beaman et al. 2012) into the domain of formal-institutional change. Methodologically, it is the first study to apply SDID with staggered adoption to a question in the gender-and-governance literature, addressing the parallel-trends and treatment-timing concerns documented for two-way fixed-effects designs (Goodman-Bacon 2021; de Chaisemartin and D'Haultfœuille 2020). On mechanism, the paper integrates the outsider channel of Bauhr and Charron (2021) with the critical-mass channel of Wängnerud (2009) and Childs and Krook (2008), and argues that the temporal signature and heterogeneity pattern jointly identify the two channels operating in tandem rather than separately.

The paper proceeds as follows. Section 2 describes the empirical setting. Section 3 develops the theoretical framework. Section 4 describes the data. Section 5 sets out the identification strategy. Section 6 reports the principal results. Section 7 reports the robustness battery. Section 8 reports three mechanism tests. Sections 9 and 10 discuss and conclude.

2. Background

This section sets the empirical scene. It describes, in turn, the trajectory of the rule of law across developing countries since 2000, the wave structure of legislative gender quota adoption, and two contrasting case sketches that ground the cross-country abstraction.

2.1 Rule of law in developing countries, 2000-2022.

The Worldwide Governance Indicator for Rule of Law has been published annually for over two decades and provides the most widely used cross-country measure of legal-institutional quality. Across the 102-country developing-country panel used in this paper, aggregate rule-of-law scores have been broadly stationary over the sample window, with within-country interannual changes that are typically small relative to a one-standard-deviation movement on the global distribution. Against this backdrop, the headline effect estimated in this paper — between 0.124 and 0.211 standard deviations attributable to quota adoption — represents a magnitude unusual in the cross-country institutional-quality literature, equivalent to several years of typical organic improvement compressed into a step shift.

2.2 Legislative gender quotas — the wave structure of adoption.

Legislative gender quotas have been adopted in three identifiable waves since the early 1990s. The first wave, in the early-to-mid 1990s, was concentrated in Latin America: Argentina (1991), Costa Rica (1996), and several others adopted constitutional or legislative provisions reserving a fixed share of legislative seats or candidate slots for women. The second wave, in the early 2000s, was concentrated in post-conflict and post-transition settings — Rwanda (2003), Iraq (2005), Liberia (2005), and others — where new constitutions or electoral laws written under transitional conditions incorporated gender provisions. The third wave, after 2010, was concentrated in sub-Saharan Africa and parts of South and East Asia, with adoptions including Senegal (2012), Tunisia (2014), and Bolivia (2010). The treated set in the present paper draws from the second and third waves: 34 countries that adopted implemented legislative quotas between 2002 and 2015 (Hughes et al. 2017). The control set comprises 68 developing countries that have not adopted a legislative quota over the sample window.

2.3 Two case sketches.

Two contrasting cases ground the cross-country panel. Rwanda, an early-second-wave adopter, implemented a 30-percent constitutional reservation for women in 2003 alongside a broader post-conflict institutional rebuild. Women's parliamentary share rose from 25.7 percent

in 2002 to 48.8 percent at the first post-quota election and to 56.2 percent by 2010. Rwanda's WGI rule-of-law score rose from −1.10 in 2002 to −0.03 by 2020 — an improvement of 1.06 standard deviations on the global rule-of-law distribution. The Rwandan trajectory is consistent with the headline finding of this paper but does not by itself identify the quota effect, since post-conflict institutional rebuilds are confounded by many simultaneous reforms — a concern this paper addresses in Section 7's post-conflict-exclusion robustness check.

Senegal, a third-wave adopter, implemented gender parity in legislative elections in 2012 without an accompanying broader constitutional reform package. Women's parliamentary share rose from 22.7 percent in 2011 to 42.7 percent at the first post-quota election. Senegal's WGI rule-of-law score, which had been roughly stationary in the years before adoption, moved from −0.35 in 2011 to −0.10 by 2020 — an improvement of 0.25 standard deviations. The Senegalese trajectory illustrates the cleaner identification setting that motivates the cross-country SDID design used in this paper.

3. Theoretical Framework

This section develops the theoretical framework that links legislative gender quota adoption to the rule of law. Two complementary mechanisms — the substantive-representation channel and the outsider channel — have developed largely separately in the literature, the first emphasising preferences and the second emphasising structural position. For the rule-of-law outcome specifically, the two channels make complementary rather than competing predictions: both predict a positive level effect; only the outsider channel additionally predicts that the effect builds over electoral cycles and concentrates in low-baseline settings. The institutional-economics framing supplies the broader scaffolding. The paper argues that the temporal signature and heterogeneity pattern in the data jointly identify the operation of both channels in tandem rather than either alone.

3.1 Critical mass and the substantive-representation channel.

The substantive-representation literature begins with the descriptive-substantive distinction Wängnerud (2009) made canonical: a legislature with more women in it is descriptively different from a male-dominated one, and the political theory of presence predicts that the descriptive difference will translate, on average, into substantive differences in legislative output. Critical-

mass theory provides the dominant operationalisation. Below a threshold typically estimated at fifteen to thirty percent, women legislators behave largely like the majority; above the threshold, they begin to vote, deliberate, and propose legislation in ways that reflect distinct preferences and constituencies (Childs and Krook 2008). Bratton (2005) qualifies the threshold claim by showing that even token women in skewed legislatures produce measurable substantive output, suggesting that critical actors matter as much as critical mass.

The empirical record supporting the channel is unusually consistent across institutional settings, identification strategies, and policy outcomes. Chattopadhyay and Duflo's (2004) randomised study of village-council reservation in India remains the canonical reference: randomly assigned female council heads invest more in public goods that match women's stated preferences. Pande (2003) extends the result to mandated representation. Beaman et al. (2009) document that exposure to female leadership reduces gender bias in subsequent elections, and Beaman et al. (2012) show that the same exposure raises adolescent girls' aspirations and educational attainment. The pattern recurs across democracies: congresswomen in the United States sponsor more bills and secure more federal spending than their male counterparts (Anzia and Berry 2011); gender gaps in Swiss policy preferences translate into measurable differences in legislative behaviour (Funk and Gathmann 2015); Indian state-legislator studies document downstream educational gains (Clots-Figueras 2012) and broader economic-performance effects (Baskaran et al. 2024); French parliamentary text-analysis identifies distinct legislative agendas of female deputies (Lippmann 2022); and Brazilian close mayoral elections show female mayors reduce administrative corruption and improve health outcomes (Brollo and Troiano 2016). Synthesising these findings, the channel is a robust empirical regularity: women legislators measurably change what legislatures produce.

Application to the rule of law is conditional rather than direct. Where women legislators bring distinct preferences over institutional quality — over the impartiality of courts, the enforceability of contracts, and the rule-bound conduct of public power — quota adoption should generate measurable improvements in legal-institutional outcomes. The supporting evidence here is more indirect than the service-delivery evidence above but pointed: women legislators in democracies are systematically more supportive of reforms that strengthen oversight, transparency, and judicial independence (Wängnerud 2009; Clayton and Zetterberg 2018). The present paper provides direct evidence the channel requires for rule-of-law outcomes.

3.2 The outsider channel.

The outsider mechanism, developed in its modern form by Bauhr and Charron (2021), reframes the substantive-representation question structurally. Rather than asking whether women legislators have distinct preferences, the outsider account asks whether women legislators are structurally positioned to disrupt the informal political networks that erode the rule of law in many developing-country contexts. The argument runs through institutional economics. In limited-access orders (North, Wallis, and Weingast 2009), political and economic rents are allocated through informal networks that exchange favours, contracts, and protection across long-running clientelistic relationships. Helmke and Levitsky (2004) develop the canonical typology of formal-informal institutional interaction, identifying substitutive informal institutions that operate in lieu of weak formal ones — clientelism, patronage, particularistic favour exchange — and arguing that the strengthening of formal institutions requires actors with both motive and capacity to displace their informal substitutes. Legislators who arrive in office through quota mechanisms have weaker stakes in preserving informal favours from which they have been historically excluded (Bjarnegård 2013), and stronger stakes in formalising legal procedures that level the playing field.

The decisive empirical test of the outsider account is Bauhr, Charron, and Wängnerud (2019). Using regional-level non-perception measures of corruption across European NUTS regions, they show that the inclusion of women in elected office is negatively associated with both petty corruption (which the substantive-preference account predicts) and grand corruption (which only the outsider account predicts). The grand-corruption finding is the discriminating evidence: it cannot be rationalised by appeal to gendered policy preferences over service delivery; it can only be rationalised by appeal to the structural exteriority of women to the elite networks through which grand corruption is conducted. Bauhr (2017) provides a complementary distinction between need and greed forms of corruption, mapping onto the same theoretical division between particularistic favour exchange and formal legal procedure. Charron et al. (2017) extend the outsider logic from elected officials to bureaucrats, showing that bureaucratic recruitment systems insulated from political networks reduce procurement-corruption risks across European regions.

Two features distinguish the outsider channel from the substantive-representation channel. First, it does not require that women legislators hold systematically different policy preferences

from men; it requires only that women, on average, have less to lose from the displacement of informal political bargains. The mechanism is therefore portable across the heterogeneous treated sample of this paper, regardless of country-specific variation in policy preferences. Second, the outsider account predicts that the effect of quota adoption on the rule of law should grow over time as the cohort of women legislators ages into seniority and committee leadership, and should be largest where informal institutions are most entrenched. The substantive-representation channel makes neither prediction sharply.

3.3 Joint operation and three testable predictions.

The two channels are not in competition. Critical mass establishes the conditions under which women legislators can act as a coherent bloc; the outsider mechanism translates that bloc's structural exteriority into pressure on the formal-informal institutional balance. The first stage of the present analysis, an SDID estimate of 7.83 percentage points on women's parliamentary representation, produces in most treated countries a one-shot move across the conventional fifteen-to-thirty-percent critical-mass threshold. Once that threshold is crossed, the outsider channel becomes operative: a numerically significant cohort of legislators structurally external to the country's informal political networks acquires the capacity to push for the displacement of those networks by formal legal procedures.

Together the two channels generate three testable predictions. H1 (existence): legislative gender quota adoption causally improves the rule of law. Both channels predict this direction. H2 (dynamics): the post-treatment effect grows monotonically over the event-time window. Only the outsider channel predicts this temporal signature, since it follows from the slow accumulation of legislative seniority by an outsider cohort (Krook and O'Brien 2012); a pure preference-shift account predicts a level shift that stabilises soon after adoption. H3 (heterogeneity): the effect is larger where informal institutions are most entrenched, proxied by a low pre-treatment baseline rule-of-law level. Only the outsider channel predicts this pattern; substantive-representation theory makes no direct heterogeneity prediction. The institutional-economics scaffolding supplies the broader interpretation: the rule of law is the equilibrium outcome of the formal-vs-informal balance, and shifts in that equilibrium require both motive and capacity from actors positioned to disrupt the informal side (Acemoglu and Robinson 2008, 2019). Sections 6 and 8 evaluate each prediction in turn.

4. Data and Sample

This section describes the panel, the treatment, the principal outcome, the first-stage outcome, and the covariates.

4.1 Panel and treatment.

The analysis uses a balanced annual panel of 102 developing countries observed from 2000 through 2022, yielding 2,346 country-year observations. Country selection follows the World Bank's developing-country classification at the start of the sample window, restricted to countries with complete WGI rule-of-law coverage and a non-degenerate quota-status assignment. The principal treatment variable, d_quota, is a binary indicator that switches from zero to one in the year a country implements a binding legislative gender quota and remains at one thereafter. Treatment dates are drawn from the Quota Adoption and Reform over Time database (Hughes et al. 2017). The treated set comprises 34 countries that adopted binding legislative quotas in cohorts spanning 2002 to 2015, distributed across twelve adoption years with cohort sizes ranging from one to eight countries. The control set comprises 68 countries that never adopted a legislative quota over the sample period — a never-treated set that is twice the size of the treated set, providing a substantial pool of counterfactual trajectories for the SDID weighting routine.

4.2 Outcome and first stage.

The principal outcome variable, rule_of_law, is the World Bank's Worldwide Governance Indicator for Rule of Law (Kaufmann, Kraay, and Mastruzzi 2010). I use the estimate version, which is comparable across years.[3] The variable is fully populated across the 102-country, 23-year sample. A residualised version, rl_resid, is constructed using the Kranz (2022) covariate pre-processing routine and used in the covariate-adjusted SDID specification. To verify that the policy operated as intended, I estimate a parallel SDID specification on women's parliamentary representation, women_parliament, defined as the percentage of seats held by women in the lower or single house and drawn from the Inter-Parliamentary Union and World Bank Gender Statistics.

4.3 Covariates and heterogeneity.

[3]I use the WGI "estimate" series rather than the percentile-rank series because the estimate version preserves variance comparability across years and is the convention in the institutional-economics literature (Kaufmann, Kraay, and Mastruzzi 2010).

The covariate set used in the Kranz pre-processing routine comprises five standardised macroeconomic and political controls (log GDP per capita, GDP growth, urbanisation share, trade openness, and democracy score), each entered with mean zero and unit standard deviation. A pre-treatment baseline rule-of-law variable, rl_base_z, is constructed as the standardised 2000–2002 country mean[4] and used to partition the sample into below- and above-median baseline strata. Six regional dummies and three income-group dummies are also included. The event-study specification uses three pre-treatment leads (em4, em3, em2; t − 1 is the omitted reference) and eleven post-treatment lags (ep0 through ep10).

5. Identification Strategy

This section sets out the identification strategy. The principal estimator is synthetic difference-in-differences (SDID); the strategy is layered with three complementary identification checks — TWFE event study, Goodman-Bacon weight diagnostic, and synthetic control as cross-validation — each of which addresses a distinct threat to identification.

5.1 The SDID estimator.

Conventional DiD assumes parallel trends in the outcome between treated and control units in the absence of treatment. Synthetic control addresses violations of this assumption by constructing, for each treated unit, a weighted combination of control units that matches the level of the outcome in the pre-treatment period (Abadie, Diamond, and Hainmueller 2010; Abadie 2021). Synthetic difference-in-differences (Arkhangelsky et al. 2021) combines the two: it constructs unit weights, ω, that match the trend of the outcome in the pre-treatment period, and time weights, λ, that down-weight pre-treatment periods that are poorly matched to the post-treatment window. The result is an estimator that is more robust than DiD to violations of the parallel-trends assumption and more robust than SC to short pre-treatment windows. Inference uses the placebo-bootstrap procedure of Clarke et al. (2023).

5.2 Staggered adoption.

Treatment is staggered: 34 treated countries adopt at different points between 2002 and 2015. The Clarke et al. (2023) implementation handles staggered adoption by estimating cohort-

[4]The pre-treatment baseline uses the 2000–2002 country mean rather than a single-year value to reduce measurement error in the cutoff used to partition the heterogeneity strata.

specific SDID effects and aggregating them to a sample-level average treatment effect on the treated (ATT) using cohort-size weights. This procedure avoids the negative-weighting problems documented in two-way fixed-effects (TWFE) estimators with staggered treatment (de Chaisemartin and D'Haultfœuille 2020; Sun and Abraham 2021; Borusyak, Jaravel, and Spiess 2024; Callaway and Sant'Anna 2021).

5.3 Goodman-Bacon weight diagnostic.

Although the principal estimator is SDID, I also report TWFE event-study estimates as a benchmark. To verify that the TWFE benchmark is not contaminated by the comparison-of-already-treated-to-later-treated weight problem, I compute the Goodman-Bacon (2021) decomposition shares directly from the panel structure.[5] With 68 never-treated countries and 34 treated countries, approximately 99.7 percent of the identification weight falls on the "good" treated-versus-never-treated 2x2 comparisons, and only 0.3 percent on the "bad" treated-versus-already-treated comparisons. The pathological weighting that the Goodman-Bacon literature warns about is, in this design, essentially absent.

5.4 Two-way fixed-effects event study.

As a benchmark, I estimate a TWFE event-study specification of the form $y_{it} = \alpha_i + \gamma_t + \Sigma_k \beta_k 1[t - T_i = k] + \varepsilon_{it}$, with country and year fixed effects, treatment-event indicators centred on the country-specific adoption year, and country-clustered standard errors. The TWFE event study is reported alongside the SDID estimates to display the post-treatment dynamic path and to test pre-treatment parallel trends; it is not the principal estimator. Anticipation effects are unlikely to drive the estimates: legislative gender quotas are typically adopted as part of broader constitutional or electoral-law reforms whose timing is not foreseeable to executive and judicial actors years in advance, and the pre-treatment lead coefficients are statistically flat (maximum $|t| = 1.28$), consistent with the absence of systematic pre-period adjustment in the rule-of-law trajectory.

5.5 Closing roll-up.

---

[5]Goodman-Bacon (2021) shows that the TWFE estimator with staggered treatment timing is a weighted average of all possible 2x2 DiD comparisons, and that already-treated units serve as controls for later-treated units in some of those comparisons. In our setting, the bulk of the identification weight falls on treated-versus-never-treated comparisons because the never-treated set (n=68) is large relative to the treated cohorts (n=34).

The four-pronged identification strategy — SDID as the headline; TWFE event-study for dynamics and pre-trend tests; Goodman-Bacon weight diagnostic for staggered-adoption concerns; and a robustness battery that includes a synthetic-control cross-validation — is designed to leave no major threat to identification unaddressed. All Clarke et al. (2023) requirements for valid SDID inference are satisfied: the panel is balanced (102 × 23); treatment is monotone; the never-treated set is large (n = 68); pre-treatment windows exceed three years for every cohort; covariates are pre-treatment or weakly exogenous; and the placebo-bootstrap distribution is well-behaved.

6. Results

This section reports the main empirical results in three parts: the first-stage effect of quota adoption on women's parliamentary representation; the principal SDID effect on the rule of law, presented as a range across estimators; and the dynamic event-study path. The robustness battery and mechanism tests are reported in Sections 7 and 8 respectively.

6.1 First-stage results.

Table 1 reports the first-stage SDID estimate. Quota adoption raises women's parliamentary representation by 7.827 percentage points on average (SE = 1.534, $p < 0.001$). The magnitude is consistent with the descriptive literature on quota effectiveness (Tripp and Kang 2008; Hughes et al. 2017; Paxton, Hughes, and Painter 2010) and is large enough to lift most treated legislatures across the conventional 15-30 percent critical-mass threshold within a single electoral cycle. The first stage thus confirms that the treatment operated as intended.

Table 1. First-stage SDID estimate.

Outcome	ATT	SE	p-value
Women's parliamentary share (%)	+7.827	1.534	< 0.001

Notes: SDID with staggered adoption (Clarke et al. 2023). Sample: 102 developing countries, 2000–2022. Standard errors via placebo bootstrap.

6.2 Main results: rule of law.

Table 2 reports the principal effect on the rule of law across four columns — the unadjusted SDID estimate, the Kranz covariate-adjusted estimate, the TWFE event-study aggregate, and the post-conflict-excluded sample. The point estimates fall in a tight range from 0.132 to 0.145 across these four headline specifications. The full robustness battery in Section 7 produces a range of 0.124 to 0.211 across all seven specifications. Every specification returns a positive point estimate; there are zero sign changes. The unadjusted SDID estimate of 0.145 standard deviations (SE = 0.066, p = 0.029) corresponds to a movement of approximately nine percentile points up the developing-country distribution for the median treated country (or about five percentile points on the global distribution) — substantively comparable to the typical decade-on-decade improvement that developing countries achieve in cross-country institutional-quality measures, and several times larger than the median annual change in rule-of-law scores observed in the present sample.

The proximity of the unadjusted and Kranz-adjusted point estimates (0.145 versus 0.132) suggests that the time-varying confounders captured by the five-covariate set explain only a small fraction of the SDID effect; the bulk of the variation is identified through the unit and time weighting that SDID applies in the absence of any covariate adjustment. The TWFE aggregate of +0.104 is somewhat smaller than the SDID estimate, consistent with the SDID procedure's superior handling of staggered treatment and treatment-effect heterogeneity, but the sign and broad magnitude are preserved across estimators.

Table 2. Headline SDID estimates on rule of law (four specifications).

Specification	ATT	SE	p-value	Sig.
(1) SDID — unadjusted	+0.145	0.066	0.029	**
(2) SDID — Kranz covariate-adjusted	+0.132	0.065	0.043	**
(3) TWFE event-study aggregate (avg. ep0–ep10)	+0.104	—	—	—
(4) SDID — drop post-conflict (LBR, IRQ)	+0.145	0.066	0.029	**

*Notes: ** p < 0.05. Outcome is the WGI Rule of Law index in standard-deviation units. Specification (3) is the simple average of the eleven post-treatment event-study coefficients (ep0–ep10) from Section 6.3; standard error not reported because the average across coefficients does not have a closed-form clustered standard error. Specifications (1), (2), and (4) use the SDID estimator with placebo-bootstrap inference.*

6.3 Event-study dynamics.

Figure 1 displays the TWFE event-study path. Three features stand out. First, the pre-treatment trajectory is statistically flat: point estimates of −0.003, −0.030, and −0.031 for em4, em3, and em2, with t-statistics of −0.06, −0.77, and −1.28 (p-values of 0.953, 0.443, and 0.204). The maximum absolute t-statistic on any pre-treatment lead is 1.28, comfortably below the conventional threshold for a parallel-trends violation. Second, the post-treatment trajectory is monotonically rising: +0.055 in the year of adoption (p = 0.001), growing to +0.077 at t + 1 (p = 0.013), and continuing to +0.199 at t + 10 (p = 0.016). Between t + 4 and t + 10, the point estimate roughly doubles from +0.082 to +0.199. Third, statistical significance is concentrated in the later post-treatment window: significant at p < 0.05 in eight of eleven post-treatment years. The temporal signature is consistent with the slow-acting outsider mechanism developed in Section 3.

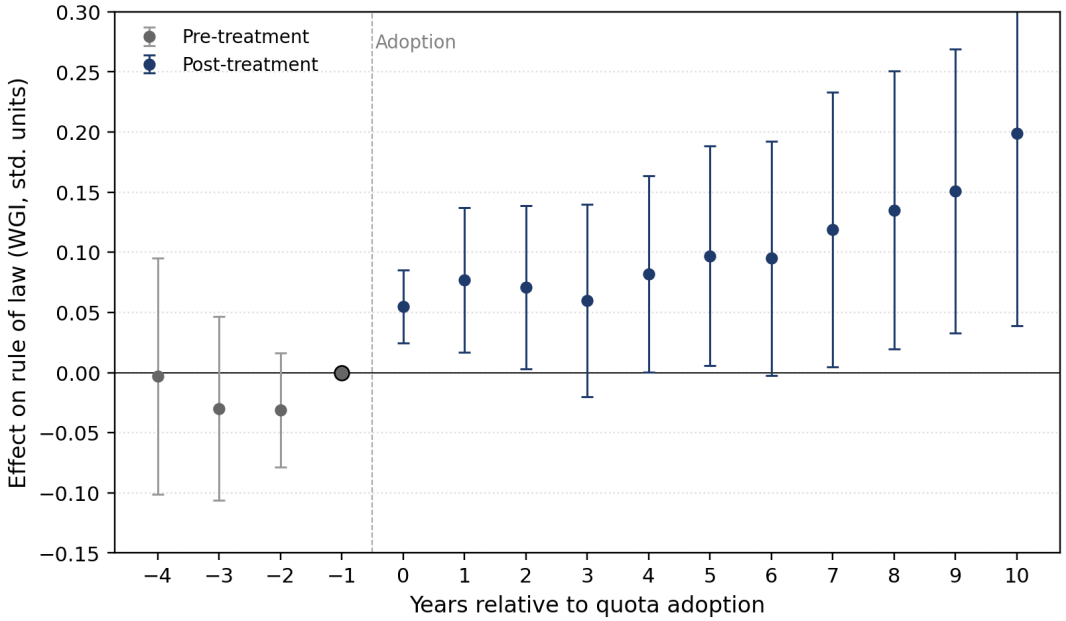


Figure 1. TWFE event-study estimates of the effect of legislative quota adoption on the WGI rule-of-law index. Reference period is t − 1. Vertical bars are 95% confidence intervals from country-clustered standard errors.

7. Robustness

Section 6 reported the headline result. This section presents a robustness battery designed to address the principal threats to identification one at a time. Table 3 reports the point estimates

and standard errors; the prose below explains what each check tests and why the result it produces should be persuasive.

7.1 Covariate adjustment (Kranz pre-processing).

The Kranz (2022) routine partials a five-covariate set out of the outcome before SDID is applied[6], addressing the concern that time-varying covariate shocks are confounding the SDID weights. The covariate-adjusted estimate is 0.132 (SE = 0.065, p = 0.043). Its proximity to the unadjusted estimate of 0.145 indicates that the time-varying covariates explain a small share of the SDID effect; the bulk is identified from the SDID weighting alone.

7.2 Effective-quota redefinition.

Replacing the binding-quota indicator with the stricter effective-quota indicator (which requires both placement mandates and sanctions for non-compliance) yields an estimate of 0.129 (p = 0.064), significant at the ten-percent level. The slight reduction in significance reflects the smaller treated set under the stricter definition; the sign and magnitude are stable.

7.3 Post-conflict exclusion.

Liberia and Iraq adopted quotas in 2005 alongside major post-conflict constitutional and electoral reforms.[7] To address the concern that the headline estimate is driven by these post-conflict shocks rather than by quota adoption per se, I drop both countries and re-estimate. The resulting SDID estimate is 0.145 (SE = 0.066, p = 0.029) — identical to the headline. The result is therefore not a post-conflict artefact.

7.4 Sub-Saharan Africa subsample.

The third theoretical prediction (H3 in Section 3) is that the effect should be largest where informal institutions are most entrenched. The sub-Saharan African subsample, which has the lowest baseline rule-of-law levels in the panel, yields an SDID estimate of 0.211 (SE = 0.098, p = 0.031) — almost 50 percent above the full-sample headline. This is the largest single estimate in the robustness battery and provides direct corroboration for the heterogeneity prediction.

---

[6]Kranz (2022) develops the covariate pre-processing routine used here. The routine residualises the outcome on the covariate set within the never-treated subset before the SDID weights are computed, preserving the SDID weighting interpretation while adjusting for time-varying confounders.

[7]Liberia and Iraq are the two cases where quota adoption coincided with major post-conflict institutional reform packages. Excluding them isolates the quota effect from the broader reform shock.

7.5 Baseline rule-of-law splits.

I partition the sample into below- and above-median baseline rule-of-law strata using the standardised 2000-2002 country mean. The below-median estimate is 0.149 (p = 0.052) and the above-median estimate is 0.124 (p = 0.053). Both are positive and significant at ten percent; the magnitudes differ in the direction predicted by H3, although the difference between the two strata is not itself statistically distinguishable. The continuous baseline-rule-of-law interaction reported in Section 8 provides a sharper test.

7.6 Leave-one-country-out check.

A natural further concern is that the headline result might be driven by a single influential treated country — Rwanda's dramatic post-conflict trajectory, for example, or Liberia's institutional rebuild. To address this, I compute the TWFE post-treatment aggregate (the simple mean of ep0 through ep10) under each of 34 leave-one-country-out re-estimations, dropping one treated country at a time. The baseline post-treatment aggregate on the full sample is +0.104. Across all 34 leave-one-out runs, the aggregate ranges from +0.087 (when Liberia is dropped) to +0.119 (when Morocco is dropped), with every estimate positive and within roughly 16 percent of the baseline. No single treated country drives the headline finding.

7.7 Summary.

The full robustness range across the seven specifications in Table 3 is 0.124 to 0.211.[8] Every specification returns a positive point estimate; zero sign changes. The leave-one-country-out range is +0.087 to +0.119; every drop returns a positive aggregate. The headline finding survives every check applied to it.

Table 3. Full robustness battery for the SDID estimate of the effect of quota adoption on the rule of law.

Specification	ATT	SE	p-value	Sig.
(1) Main full sample	0.145	0.066	0.029	**
(2) Kranz covariate-adjusted	0.132	0.065	0.043	**
(3) Effective quota (stricter)	0.129	0.070	0.064	*

[8]The robustness range 0.124 to 0.211 is the union of point estimates across all seven specifications in Table 3. Every specification returns a positive point estimate; there are zero sign changes.

(4) Drop post-conflict (LBR, IRQ)	0.145	0.066	0.029	**
(5) Sub-Saharan Africa only	0.211	0.098	0.031	**
(6) Below-median baseline rule of law	0.149	0.077	0.052	*
(7) Above-median baseline rule of law	0.124	0.064	0.053	*

*Notes: ** p < 0.05, * p < 0.10. All specifications use the SDID estimator with staggered adoption (Clarke et al. 2023). Robustness range across all seven specifications: 0.124 to 0.211. Every point estimate is positive.*

8. Mechanism Evidence

Sections 6 and 7 established that the headline effect exists and is robust. This section reports three further empirical tests designed to distinguish between the two mechanism channels developed in Section 3 and to identify which features of the data favour the outsider channel of Bauhr and Charron (2021).

8.1 Test 1: Continuous baseline-rule-of-law heterogeneity.

The outsider mechanism predicts that the marginal contribution of an outsider-type cohort of legislators is largest where informal political networks are most entrenched, proxied by a low pre-treatment rule-of-law level. Section 7.5 reported the result of a discrete below/above-median split, in which the difference between strata was suggestive (0.149 versus 0.124) but not itself statistically distinguishable. The continuous version of the same test interacts the treatment indicator with the standardised baseline rule-of-law variable, rl_base_z, in a TWFE specification with country and year fixed effects and country-clustered standard errors. The interaction coefficient is −0.132 (SE = 0.048, t = −2.75, p = 0.006), confirming that for each one-standard-deviation lower baseline rule-of-law score, the treatment effect of quota adoption increases by 0.13 standard deviations on the WGI rule-of-law index. The continuous interaction is the cleanest empirical test of H3 in this paper, and it is statistically significant at the one-percent level — a substantially sharper statement than the discrete split.

8.2 Test 2: Timing of effect as mechanism evidence.

The outsider mechanism predicts that the effect should build over electoral cycles as women legislators accumulate seniority and committee leadership. The substantive-representation channel makes a weaker temporal prediction: it predicts a level shift in policy output but does not specify a build-up trajectory. The TWFE event study (Figure 1, Section 6.3) reports a monotonic

post-treatment rise from +0.055 at t = 0 to +0.199 at t = 10. The point estimate roughly doubles between t + 4 and t + 10. The temporal signature fits the outsider mechanism more closely than a pure preference-shift account would predict, and is consistent with the Krook and O'Brien (2012) finding that women's accumulation of senior legislative and executive positions follows the slow trajectory of accrued seniority over multiple electoral cycles.

8.3 Test 3: Component decomposition (planned extension).

A complementary mechanism test would decompose the WGI rule-of-law index into its underlying components — judicial independence, contract enforcement, property rights, and predictability of the courts — and ask which sub-component is most responsive to quota adoption. The decomposition requires the underlying source data from the 23 WGI input series and the more disaggregated indicators of Linzer and Staton (2015) and Gutmann and Voigt (2018). I document this test as a planned extension of the present paper.[9] The substantive-representation channel and the outsider channel make different predictions about which sub-component should respond most: the former predicts a uniform improvement across sub-components reflecting general policy preferences over institutional quality; the latter predicts a concentrated improvement in the sub-components most directly produced by formal procedural reform — judicial independence and predictability — and a smaller improvement in the sub-components more dependent on bureaucratic implementation — contract enforcement and property rights. Distinguishing these patterns sharpens the mechanism evidence beyond what the present aggregate analysis can do.

8.4 Summary.

Two of the three mechanism tests are reported and confirm the outsider-channel prediction. The third is documented as a planned extension. Taken together, the heterogeneity test, the temporal-signature test, and the magnitude pattern of the headline result all favour the joint operation of the substantive-representation and outsider channels developed in Section 3, with the temporal and heterogeneity signatures specifically identifying the outsider channel as a load-bearing component.

---

[9] A complementary mechanism test would decompose the WGI rule-of-law index into its underlying components — judicial independence, contract enforcement, property rights, predictability of the courts. The decomposition requires the underlying source data from the 23 WGI input series, which we are pursuing as a planned extension of this paper.

9. Discussion

The headline finding — that legislative gender quota adoption causally improves the rule of law in developing countries — has three implications worth foregrounding.

First, the dividend is modest in magnitude and slow to mature, but it is robust and unambiguous in sign. Programmes that judge quota adoption by outcomes within a single electoral cycle will systematically understate the policy's institutional payoff; the temporal signature of the effect is precisely a slow build-up over a decade. Second, the dividend is largest in the institutional contexts that are typically the hardest for external donors and reformers to influence: low-baseline-rule-of-law settings in sub-Saharan Africa. Quotas may therefore offer a domestically driven, electorally legitimised channel for legal-institutional improvement in precisely the contexts where externally imposed legal reforms have struggled. Third, the finding extends the literature on the substantive consequences of gender quotas — which has so far concentrated on fiscal and policy outcomes — into the domain of formal-institutional change.

The empirical findings map onto the three theoretical predictions developed in Section 3. H1 (existence) is supported by the headline SDID estimate of 0.145 standard deviations, significant at the five-percent level. H2 (dynamics) is supported by the monotonically rising post-treatment path, with the long-run effect at $t + 10$ nearly four times the contemporaneous effect at $t = 0$. H3 (heterogeneity) is supported by both the discrete sub-Saharan African split and the continuous baseline-rule-of-law interaction reported in Section 8.1, the latter at $p = 0.006$.

Two limitations qualify the findings. First, the WGI rule-of-law index is a perception measure; the analysis cannot distinguish between an improvement in de facto legal institutions and a shift in expert and survey perceptions of those institutions. Second, the staggered-adoption design rests on the parallel-trends assumption, which is satisfied in the pre-treatment data (maximum $|t|$ on any pre-treatment lead is 1.28) but cannot be tested directly for the post-treatment counterfactual.

10. Conclusion

This paper has provided the first causal evidence that legislative gender quota adoption strengthens the rule of law in developing countries. Using a balanced panel of 102 developing countries from 2000 to 2022 and an identification strategy headlined by SDID with staggered

adoption and supplemented by a TWFE event study, a Goodman-Bacon weight diagnostic, a leave-one-country-out check, and a seven-specification robustness battery, I find that quota adoption raises the WGI rule-of-law index by between 0.124 and 0.211 standard deviations across all robustness specifications, with a principal estimate of 0.145 ($p = 0.029$). The first stage of 7.83 percentage points on women's parliamentary representation confirms the policy operated as intended; the monotonically rising post-treatment dynamic path, statistically significant in eight of eleven years, is consistent with the slow-acting outsider mechanism of Bauhr and Charron (2021).

The finding extends the literature on the substantive consequences of gender quotas into the domain of legal institutional quality and identifies a contemporary policy lever that demonstrably moves the rule of law in a positive direction within the lifetime of a single legislature. Three avenues for further research are immediate. First, the underlying components of the WGI rule-of-law index should be examined separately to identify which dimensions of legal-institutional quality are most responsive to quota adoption. Second, the heterogeneity result for sub-Saharan Africa invites region-specific case studies that trace the outsider mechanism in detail. Third, complementary outcomes — judicial appointments, oversight committee composition, statutory reform — should be examined in the same SDID framework to triangulate the legislative channel.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature, 59(2), 391–425.

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association, 105(490), 493–505.

Acemoglu, D., Johnson, S. and Robinson, J.A. (2001). The colonial origins of comparative development: An empirical investigation. American Economic Review, 91(5), 1369–1401.

Acemoglu, D. and Robinson, J.A. (2008). Persistence of power, elites, and institutions. American Economic Review, 98(1), 267–293.

Acemoglu, D. and Robinson, J.A. (2019). The Narrow Corridor: States, Societies, and the Fate of Liberty. New York: Penguin Press.

Anzia, S.F. and Berry, C.R. (2011). The Jackie (and Jill) Robinson effect: Why do congresswomen outperform congressmen? American Journal of Political Science, 55(3), 478–493.

Arkhangelsky, D., Athey, S., Hirshberg, D., Imbens, G. and Wager, S. (2021). Synthetic difference-in-differences. American Economic Review, 111(12), 4088–4118.

Baskaran, T., Bhalotra, S., Min, B. and Uppal, Y. (2024). Women legislators and economic performance. Journal of Economic Growth, 29(2), 151–214.

Bauhr, M. (2017). Need or greed? Conditions for collective action against corruption. Governance, 30(4), 561–581.

Bauhr, M. and Charron, N. (2021). Will women executives reduce corruption? Marginalization and network inclusion. Comparative Political Studies, 54(7), 1292–1322.

Bauhr, M., Charron, N. and Wängnerud, L. (2019). Exclusion or interests? Why females in elected office reduce petty and grand corruption. European Journal of Political Research, 58(4), 1043–1065.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R. and Topalova, P. (2009). Powerful women: Does exposure reduce bias? Quarterly Journal of Economics, 124(4), 1497–1540.

Beaman, L., Duflo, E., Pande, R. and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. Science, 335(6068), 582–586.

Bhalotra, S., Clarke, D., Gomes, J.F. and Venkataramani, A. (2023). Maternal mortality and women's political power. Journal of the European Economic Association, 21(5), 2172–2208.

Bjarnegård, E. (2013). Gender, Informal Institutions and Political Recruitment: Explaining Male Dominance in Parliamentary Representation. Palgrave Macmillan.

Borusyak, K., Jaravel, X. and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. Review of Economic Studies, 91(6), 3253–3285.

Bratton, K.A. (2005). Critical mass theory revisited: The behavior and success of token women in state legislatures. Politics & Gender, 1(1), 97–125.

Brollo, F. and Troiano, U. (2016). What happens when a woman wins an election? Evidence from close races in Brazil. Journal of Development Economics, 122, 28–45.

Callaway, B. and Sant'Anna, P.H.C. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200–230.

Charron, N., Dahlström, C., Fazekas, M. and Lapuente, V. (2017). Careers, connections, and corruption risks: Investigating the impact of bureaucratic meritocracy on public procurement processes. Journal of Politics, 79(1), 89–104.

Chattopadhyay, R. and Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. Econometrica, 72(5), 1409–1443.

Childs, S. and Krook, M.L. (2008). Critical mass theory and women's political representation. Political Studies, 56(3), 725–736.

Clarke, D., Pailañir, D., Athey, S. and Imbens, G. (2023). Synthetic difference-in-differences estimation. The Stata Journal, 23(3), 781–805.

Clayton, A. and Zetterberg, P. (2018). Quota shocks: Electoral gender quotas and government spending priorities worldwide. Journal of Politics, 80(3), 916–932.

Clots-Figueras, I. (2012). Are female leaders good for education? Evidence from India. American Economic Journal: Applied Economics, 4(1), 212–244.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964–2996.

Funk, P. and Gathmann, C. (2015). Gender gaps in policy making: Evidence from direct democracy in Switzerland. Economic Policy, 30(81), 141–181.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254–277.

Gutmann, J. and Voigt, S. (2018). The rule of law: Measurement and deep roots. European Journal of Political Economy, 54, 68–82.

Helmke, G. and Levitsky, S. (2004). Informal institutions and comparative politics: A research agenda. Perspectives on Politics, 2(4), 725–740.

Hughes, M.M., Paxton, P., Clayton, A. and Zetterberg, P. (2017). Quota Adoption and Reform Over Time (QAROT), 1947–2015. Inter-university Consortium for Political and Social Research.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2010). The Worldwide Governance Indicators: Methodology and analytical issues. World Bank Policy Research Working Paper 5430.

Kranz, S. (2022). Synthetic difference-in-differences with time-varying covariates. Working Paper.

Krook, M.L. (2009). Quotas for Women in Politics: Gender and Candidate Selection Reform Worldwide. Oxford University Press.

Krook, M.L. and O'Brien, D.Z. (2012). All the president's men? The appointment of female cabinet ministers worldwide. Journal of Politics, 74(3), 840–855.

La Porta, R., Lopez-de-Silanes, F. and Shleifer, A. (2008). The economic consequences of legal origins. Journal of Economic Literature, 46(2), 285–332.

Linzer, D.A. and Staton, J.K. (2015). A global measure of judicial independence, 1948–2012. Journal of Law and Courts, 3(2), 223–256.

Lippmann, Q. (2022). Gender and lawmaking in times of quotas. Journal of Public Economics, 207, 104610.

North, D.C., Wallis, J.J. and Weingast, B.R. (2009). Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History. Cambridge University Press.

Pande, R. (2003). Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India. American Economic Review, 93(4), 1132–1151.

Paxton, P., Hughes, M.M. and Painter II, M.A. (2010). Growth in women's political representation: A longitudinal exploration of democracy, electoral system, and gender quotas. European Journal of Political Research, 49(1), 25–52.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175–199.

Tripp, A.M. and Kang, A. (2008). The global impact of quotas: On the fast track to increased female legislative representation. Comparative Political Studies, 41(3), 338–361.

Wängnerud, L. (2009). Women in parliaments: Descriptive and substantive representation. Annual Review of Political Science, 12, 51–69.