

Neglect Of 'Theory' In Quantitative Research Method In International Relations

Posted on June 8, 2023

Introduction

Quantitative research has become central to International Relations because it allows scholars to compare cases, estimate relationships, evaluate uncertainty, and test claims using transparent evidence. Large datasets on war, trade, alliances, sanctions, democracy, public opinion, and international organizations have expanded what researchers can study. Yet numerical sophistication does not automatically produce meaningful knowledge. Statistical models require concepts, assumptions, and explanations that specify what the variables represent and why the relationships should exist.

The original essay draws on John Mearsheimer and Stephen Walt's argument that parts of quantitative International Relations have elevated hypothesis testing while giving insufficient attention to theory. Their criticism is not that statistics should be abandoned. It is that models become shallow when researchers select variables, estimate coefficients, and report significance without developing causal mechanisms or connecting findings to broader explanations of world politics. This essay examines that concern, considers the strengths and limits of the critique, and argues that theory and quantitative method should be treated as mutually dependent rather than competing activities. (Mearsheimer, 2013)

What Theory Does in International Relations

A theory is more than a prediction that one variable will be associated with another. It identifies important actors or structures, defines concepts, specifies relationships, describes causal mechanisms, and states the conditions under which an explanation should apply. A theory of alliance formation, for example, should explain why states perceive threats, how they evaluate partners, and when balancing, bandwagoning, neutrality, or nonalignment is expected.

Theory performs several tasks before data analysis begins. It tells researchers what should be measured, which cases are relevant, which variables may be confounders, and what evidence would contradict the argument. It also helps distinguish correlation from explanation. Two countries may trade heavily and avoid war, but theory is needed to determine whether commerce raises the cost of conflict, peaceful relations encourage trade, common institutions produce both outcomes, or selection processes make the association misleading.

Robert Cox's statement that theory is always "for someone and for some purpose" adds another dimension. Theories emerge from historical and political locations. They highlight some actors and problems while making others less visible. Recognizing perspective does not require abandoning evidence. It requires scholars to examine the assumptions and interests built into their categories.

The Rise of Simplistic Hypothesis Testing

Mearsheimer and Walt criticize a research style in which scholars generate narrow hypotheses, locate available data, and estimate statistical relationships without building or refining substantial theory. The resulting articles may be technically competent but disconnected from one another. A positive coefficient in one study and a negative coefficient in another do not cumulate into knowledge if the underlying concepts and mechanisms differ.

Publication incentives can encourage this pattern. Journals may favor novel findings, sophisticated estimators, and statistically significant results. Researchers face pressure to divide large questions into small testable claims. Data availability can determine the research question: scholars study what has been coded rather than what is theoretically important. Variables become substitutes for concepts, and the choice of estimator receives more attention than the explanation of political behavior.

Christopher Achen's criticism of "garbage-can regressions" addresses models that include many variables without a coherent causal structure. Adding controls may appear rigorous, but a control can introduce bias when it is caused by the treatment, selected by the outcome, or measured after the relevant process. Theory is necessary to decide what belongs in the model. (Achen, 2002)

Why Quantitative Evidence Still Matters

The critique should not be interpreted as a defense of theory that remains insulated from evidence. International Relations has a history of broad theoretical debates in which competing schools explain nearly every outcome after it occurs. Quantitative methods can force scholars to define expectations, compare systematic evidence, and acknowledge uncertainty. They can reveal patterns that selected historical cases obscure.

Statistical analysis is especially useful when the argument concerns frequency, average effects, variation across many countries, or changes over time. Research on democratic peace, economic sanctions, civil conflict, and international institutions has benefited from datasets that enable replication and comparison. Quantitative work can also identify cases that do not fit a theory and therefore require revision.

The problem is not hypothesis testing itself. A well-designed test connects an observable implication to a theory and evaluates evidence that could plausibly disconfirm it. The problem arises when the

hypothesis has no explanatory foundation or when statistical significance is treated as theoretical importance.

Concept Formation and Measurement

Theory guides measurement. Concepts such as power, democracy, threat, legitimacy, polarization, and state capacity are not directly observable. Researchers create indicators, but every indicator captures some dimensions and excludes others. Military expenditure may measure resources but not readiness, geography, alliance support, or political willingness. A democracy score may combine electoral competition, civil liberties, and institutional constraints in ways that affect the result.

Measurement error can weaken estimates or create misleading comparisons. Cross-national data may be less complete in states with low administrative capacity or active conflict. Official figures may reflect political incentives. Coding decisions can embed Western institutional assumptions or treat contested boundaries as settled facts. A theoretical discussion should explain why the selected measure represents the concept and where it may fail.

Construct validity is therefore not a preliminary technical issue. It is part of the substantive argument. If a study claims to explain state repression but measures only reported human-rights violations, variation may reflect monitoring access and media freedom as well as actual abuse.

Causal Mechanisms and Statistical Identification

A causal mechanism describes the process connecting an explanatory factor to an outcome. Statistical identification concerns whether the research design can distinguish that effect from alternative explanations. The two are related but not identical. A credible natural experiment may estimate an effect even when the mechanism is uncertain, while a persuasive mechanism can be paired with a weak design that does not establish causation.

Modern quantitative research increasingly uses matching, instrumental variables, regression discontinuity, difference-in-differences, panel models, and experiments. These tools can improve causal inference, but each depends on assumptions. An instrument must affect the outcome only through the proposed treatment. Parallel-trends assumptions must be plausible for difference-in-differences. Experiments may identify effects in a particular population while leaving external validity uncertain.

Theory helps justify these assumptions. It explains why an instrument should be exogenous, why comparison groups are appropriate, and why the treatment should operate similarly across contexts. Without theory, methodological sophistication can conceal rather than solve the inferential problem.

Levels of Analysis and Model Specification

International Relations research moves among individuals, leaders, organizations, states, dyads, regions, and the international system. A theory developed at one level cannot automatically be tested with data from another. Public attitudes may influence foreign policy, but state behavior is mediated by institutions and leaders. Systemic power distributions may shape opportunities while domestic politics determines choices. (Cox, 1981)

Data are often nested and dependent. Years within countries, countries within regions, and dyads sharing the same states are not independent observations. Multilevel models, clustered standard errors, network methods, and spatial models may be appropriate, but the choice should follow the theoretical structure. Treating every country-year as an independent unit can produce misleading certainty.

Temporal order also matters. A model that predicts conflict using variables measured after mobilization may incorporate consequences of the outcome. Theory should specify when causes operate and when observations should be taken.

Prediction Versus Explanation

Machine learning and large-scale data have strengthened prediction in political science. A model may forecast protest, conflict onset, or treaty violation without offering a simple causal explanation. Prediction is valuable for early warning and policy planning, but predictive performance should not be confused with theory confirmation.

A variable can improve prediction because it is a proxy for many processes, because it captures information close in time to the outcome, or because it reflects consequences already underway. Such a variable may be poor evidence for a causal argument. Conversely, a theoretically important cause can have limited predictive power when outcomes are rare or highly contingent.

The most productive research states its goal clearly. Explanatory studies should focus on causal estimands and mechanisms. Predictive studies should use out-of-sample validation, calibration, and transparent error analysis. Some projects can combine both, but they should not evaluate one goal with the standards of the other.

The Value of Mixed Methods

Case studies, archival research, interviews, process tracing, formal models, and quantitative analysis can address different parts of the same theoretical problem. A statistical pattern may establish scope, while detailed cases reveal mechanisms and sequencing. Case selection can investigate typical,

deviant, or influential observations rather than using anecdotes as decoration.

Mixed methods are not automatically superior. Combining weak methods does not create strong evidence. The components should be integrated around a common theoretical question. A case study should examine the same mechanism implied by the statistical model, and quantitative analysis should reflect insights from historical evidence.

International Relations is especially suited to this integration because major events are few, heterogeneous, and historically embedded. Wars, revolutions, and institutional transformations cannot always be understood through averages alone.

Counterarguments to the “Leaving Theory Behind” Thesis

Critics of Mearsheimer and Walt argue that their definition of theory privileges broad explanatory traditions and underestimates the theoretical value of middle-range research. A carefully specified mechanism concerning one institution or behavior may contribute more cumulative knowledge than a grand theory that explains too much. Empirical anomalies can also stimulate theory rather than waiting for theory to dictate every test.

Quantitative scholars frequently engage theory through identification strategies, measurement models, and formal derivation even when the article does not resemble traditional realist or liberal theory-building. The field’s methodological development has exposed assumptions that older theoretical debates left implicit. Greater attention to research design can be interpreted as an improvement rather than a retreat.

These objections are persuasive when the empirical work has a clear conceptual purpose. They do not defend mechanical significance testing, post hoc storytelling, or models assembled from available variables. The debate is therefore partly about what counts as theory and how much theoretical ambition each study should possess.

Building Cumulative Research

Cumulative knowledge requires more than repeated studies on related topics. Scholars need shared definitions, transparent data, preregistered or clearly specified analyses, replication, and serious engagement with contradictory results. Null findings and failed replications should contribute to theory revision rather than disappear through publication bias.

Researchers should specify scope conditions: the countries, periods, actors, and institutional settings in which the theory is expected to apply. Heterogeneous effects should be theorized rather than treated as inconvenient noise. A relationship that differs across regime types or regions may reveal

the mechanism more clearly than a single global average.

Open data and code improve verification, but transparency cannot replace judgment. A fully reproducible model can still measure the wrong concept or answer an unimportant question. Theoretical significance and methodological transparency are complementary standards.

A Productive Relationship Between Theory and Method

A strong quantitative IR project begins with a substantive puzzle, defines concepts, proposes a mechanism, derives observable implications, and identifies plausible alternatives. It then selects data and a research design suited to the claim. After estimation, the researcher returns to theory: Do the results support the mechanism, narrow the scope, reveal heterogeneity, or require revision?

Theory should not protect preferred claims from evidence, and method should not become an end in itself. Statistical tools are languages for expressing assumptions and uncertainty. Their value depends on the political explanation they serve.

Conclusion

The neglect of theory in quantitative International Relations is a genuine risk when hypothesis testing becomes disconnected from concept formation, causal mechanisms, historical context, and cumulative explanation. Yet the solution is not to retreat from quantitative evidence. Statistical methods can discipline theory, reveal broad patterns, test observable implications, and expose uncertainty. The most valuable scholarship integrates theoretical reasoning with measurement, identification, diagnostics, and substantive interpretation. Theory explains why evidence should matter; method determines how confidently the evidence can support the claim. International Relations needs both. (Dunne et al., 2013)

References

Achen, Christopher H. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science*, vol. 5, 2002, pp. 423-450.

Cox, Robert W. "Social Forces, States and World Orders: Beyond International Relations Theory." *Millennium*, vol. 10, no. 2, 1981, pp. 126-155.

Dunne, Tim, Lene Hansen, and Colin Wight. "The End of International Relations Theory?" *European Journal of International Relations*, vol. 19, no. 3, 2013, pp. 405-425.

Mearsheimer, John J., and Stephen M. Walt. "Leaving Theory Behind: Why Simplistic Hypothesis Testing Is Bad for International Relations." *European Journal of International Relations*, vol. 19, no. 3, 2013, pp. 427-457.

King, Gary, Robert O. Keohane, and Sidney Verba. *Designing Social Inquiry*. Princeton University Press, 1994.