The economic consequences of UN peacekeeping operations

Word Count:

Abstract:

Keywords:

**Introduction**

Scholars of peacebuilding and conflict management are largely interested in ***causal*** research questions. Do United Nations (UN) peacekeeping operations (PKOs) produce stable post-conflict environments? Can foreign aid serve as an instrument to promote peace and economic development? Questions such as these lead to widespread usage of causal language when interpreting results in the conflict management literature. A review of prominent pieces in the recent PKO literature reveals frequent use of causally suggestive terms such as “effect”, “impact”, “increase/decrease”, “leads to”, and “results in” (**Hultman et al. 2013, 2014, Beardsley and Gleditsch 2015, Bove and Ruggeri 2016, Ruggeri et al. 2017, Beardsley et al. 2018, Bove and Ruggeri 2018, Bara 2020**)**.** However, causally loaded language ***begs*** a causal interpretation. Absent a clearly-defined causal model, an identification strategy that allows a causal interpretation of an otherwise descriptive finding, and (if appropriate) a sensitivity analysis to assess the problem of unobserved confounding, researchers should refrain from using causally-loaded terminology when describing their results.

Instead, the extent of model coefficients in ***most*** PKO analyses should be interpreted ***descriptively***, rather than ***causally***. That is, an interpretation of an imagined coefficient taking the value of 1.13 should make no assumptions concerning the causal or predictive nature between and in its interpretation. “A 1-unit increase in is ***associated*** with a 1.13-unit increase in ” is an appropriate descriptive interpretation of linear regression output, absent the appropriate steps required to elevate a descriptive finding to a causal finding. Such an interpretation simply describes the degree to which two variables ***move*** together. For scholars, policymakers, and the public, such an interpretation is rather uninteresting. In truth, hardly anybody cares whether two variables move together. To the extent that scholars do care, this is motivated by an underlying suggestion that the descriptive finding may reflect a causal relationship. Within the context of PKO research, we wish to know if PKOs **work**. That is, do PKOs ***cause*** a change in some outcome that is considered favorable (termination of violence, development, etc.).

Some scholars may naturally object that this critique is a moot point. One could argue that conflict management scholars should not focus on making causal claims when such scholars are largely limited to the use of observational data. This critique is rooted in some truth. Indeed, because of this limitation, a growing inter-disciplinary movement (encompassing political science, economics, sociology, psychology, etc.) that allows scholars to make causal inferences with observational data has surprisingly flown under the radar in conflict management research. I argue that conflict management scholars ***must*** embrace the tools of causal inference to further progress in this important domain of research. The benefits of a reorientation towards explicitly causal research are threefold. First, causal research allows scholars to marry causal theory with causal methods, closing gaps between theory, methods, and results. Second, the normative goal driving conflict management research is the creation of scientific output that informs policymaker decision-making. In this manner, causal inference allows scholars to directly answer a very reasonable question from policymakers, “does this conflict management program *work*?”. Lastly, norms in causal inference concerning the use of sensitivity of results is likewise necessary in conflict management research. Prior to the suggestion of the adoption of costly (in economic and human terms) conflict management programs, scholars should do their due diligence and ensure that their estimates are not highly sensitive to unmeasured confounding effects.

While many scholars may naturally feel skeptical of a shift towards overtly causal research, it is worth remembering that the aim of most existing political science research is causal inquiry (**Gerring 2012, Samii 2016**). The suggestions that will be made in this paper are not practices far removed from the research that conflict management scholars currently engage in. Rather, I argue for the adoption of certain tools and research strategies to overcome some of the unnecessary limitations that scholars place on their findings and allow scholars to make causal inferences. The remainder of this paper is structured as follows. First, I outline the potential outcomes framework of causal inference and use this framework to describe why causal inference with observational data has conventionally been considered an impossibility. Second, I outline three steps that researchers should follow when attempting to make causal inferences from observational data. This three-step strategy is formally implemented in an analysis of the effects of UN PKOs on economic development. I find that, contrary to expectations, UN PKO deployments lead to a reduction in economic development that only furthers over time. While the sign on the effect of UN PKO withdrawals is also negative, the estimated effects are not significant and are sensitive to personal modeling decisions.

**Causal inference with observational data**

To effectively understand causality and subsequent issues with inferring causality, the potential outcomes framework (POF) developed by **Rubin (2005)** denotes a useful starting point. Under the POF, each observed unit (in the context of conflict management research, a country-year observation, for example) has one observed outcome. The GDP per capita (an observed outcome) of Rwanda in 1994 is fixed in history. A scholar may wish to know if the United Nations Assistance Mission for Rwanda (UNAMIR), authorized in 1993, had an economic effect in Rwanda. According to the data from **Fariss et al. (2021)**, Rwanda’s GDP per capita in 1994 fell roughly 24% from its value the year prior. On the one hand, UNAMIR (the treatment) may have caused a decrease in Rwanda’s economic development. Alternatively, Rwanda’s economic development could have been ***even worse*** had UNAMIR not been authorized. Under the POF, the causal effect of UNAMIR is the difference in economic development between the observed Rwanda that received treatment and the counterfactual ***exact same*** Rwanda that only differed in its treatment status. That is, the counterfactual Rwanda that never received UNAMIR.

Of course, the impracticality of estimating a causal effect of UNAMIR under the POF is glaring. The counterfactual of Rwanda with no UNAMIR is not real and can never be observed. UNAMIR ended in 1996, but Rwanda in 1997 is fundamentally different than Rwanda in 1994, so we cannot compare the two to isolate a causal effect of UNAMIR. Barring a time machine, we cannot observe the exact same unit twice and simply manipulate treatment status. This issue is known as “the fundamental problem of causal inference”. Instead, we could compare Rwanda in 1994 to a ***very*** similar country that did not have a UN PKO. However, any difference in outcome between the two may be attributed to unique factors related to the two cases. Because of this, researchers are typically interested in estimating ***average*** causal effects of some population. However, this strategy does not resolve an infamous issue for making causal inferences, the pesky confounding effect.

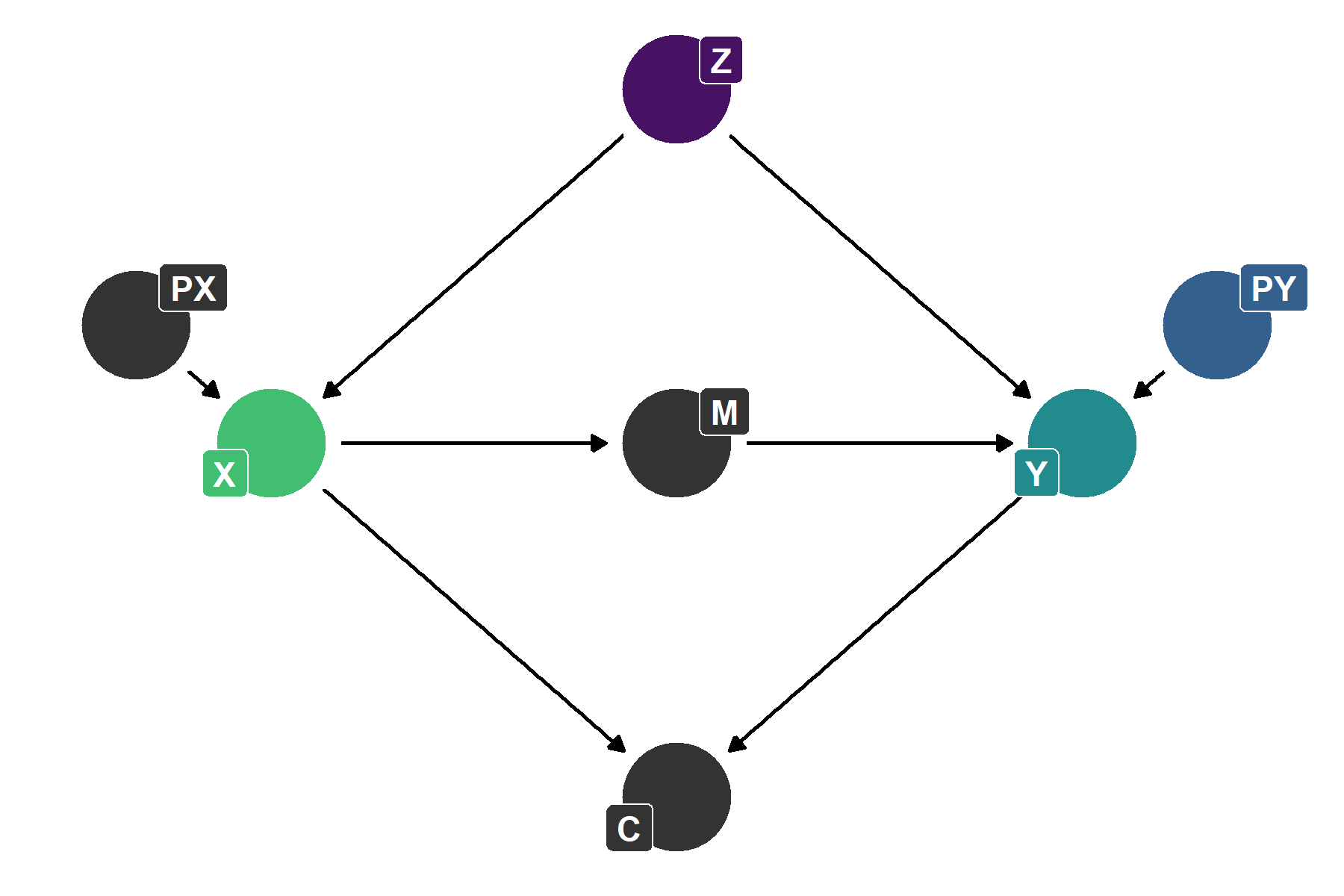
A confounder is any variable that ***causes*** some change in both treatment and outcome. Continuing the example of the economic effects of UN PKOs, conflict duration represents a very clear confounding effect. Increased conflict duration may incentivize the onset of a PKO deployment ***and*** damage a country’s development. If we were to ignore this confounder, we could not be sure that the effect of PKOs on economic development is actually ***caused*** by PKOs or if the relationship is spurious and is driven by conflict duration. When the time machine is avaliable, adjusting for conflict duration is not an issue since the exact same country-year is avaliable to make a comparison to. However, when the time machine is in the shop, we have to compare observed outcomes with approximate counterfactuals (a country very similar to Rwanda in 1994, for example). When doing this, scholars should be sure that Rwanda in 1994 and a similar non-treated comparison have nearly identical conflict durations, lest the confounding influence of conflict duration bias the causal effect. However, this approach is ultimately unsatisfying because ***conflict duration is not the only confounding effect***. Other factors such as peace agreements, conflict intensity, government military capacity, etc. all have the potential to impact both treatment and outcome. Researchers can do their best to identify ***every*** confounder possible (and they should!), but it is fundamentally impossible to know if ***all*** confounders have been accounted for.

For decades, the conventional method for getting around the issue of unobserved confounding was the randomized controlled trial (RCT). The RCT can practically eliminate confounding bias by randomizing access to treatment. As a result, estimates of an RCT can be interpreted causally. If access to treatment is assigned by pure chance (the flip of a coin, the toss of a die), nothing else can cause a change in the chances of treatment. As a result, no confounders, whether they are known or unknown, can exist since a confounder causally affects ***both*** treatment and outcome. In a RCT, the treated and control groups should be well-balanced meaning that, on average, the two groups systematically differ ***only*** in treatment status. Unfortunatley, conflict management scholars do not have the luxury of executing RCTs. The inability of conflict management scholars to use RCTs is quite apparent. First, policymakers have a vested interest in making sure that expensive conflict management programs ***work***. The scholarly community would have a difficult time convincing policymakers to allow them to ***randomly*** assign peacekeeping operations or billion-dollar foreign aid packages. Second, assuming scholars commanded a great amount of political clout and could convince policymakers to let researchers randomly assign various conflict management programs, randomization would hardly be an ethical criterion for potentially life-saving assistance.

Given practical limitations and ethical considerations, the RCT has been applied sparingly in conflict management research. Due to legitimate concerns of unobserved confounding, many have assumed that making causal inferences with observational data was not a possibility. However, numerous innovations over the past couple of decades driven by a growing community are making this very feat possible. While not ***easy***, making causal inferences with observational data is ***possible***. Researchers must pay increasing detail to both their theoretical assumptions and research designs to make causal inference possible in conflict management settings. In the following section, I outline a three-step process that can guide researchers seeking to make causal inferences with their observational data.

**Step 1: Develop a causal model**

Most conflict management scholars organically develop causal theories. However, these causal theories are rarely accompanied by a formal visual representation of said causal assumptions. Scholars make it very clear that they expect to cause some change in . However, most other causal assumptions regarding the relationships between, , and covariates are not made transparent. Despite this, formally outlining causal assumptions and the relationship between *all*relevant variables in a research design is critically important for the pursuit of causal estimates. Perhaps the most popular tool for making causal assumptions clear is the *directed acyclic graph*(DAG). DAGs are incredibly powerful tools that allow researchers to visualize complex causal relationships between variables. The benefits of DAGs are twofold. First, DAGs promote increased scrutiny in the theory-building process of research given that theoretical assumptions regarding the relationships between variables are made explicitly clear. Second, DAGs execute behind-the-scenes math that explicitly inform researchers which variables *should*and *should not*be adjusting for to make causal inferences. While it is beyond the scope of this paper to provide a satisfactory introduction to DAGs and appropriate adjustment (see Rohrer 2018 or Cinelli et al. 2022), Figure 1 demonstrates an intentionally simplistic DAG where is assumed to causally impact .



**Figure 1. A simple DAG**

Previous political science scholars have critiqued conventional control variable culture, arguing that researchers are too eager to throw “everything” into a single (or series of) “garbage can” models (Achen 2005, Schrodt 2014). Such criticisms are likewise shared by scholars familiar with causal inference (Clarke et al. 2018, Rohrer 2018, Cinelli et al. 2022). While the list of the “good” and the “bad” controls is exhaustive, a good rule of thumb is to adjust for confounders and predictors (parents/ancestors) of the outcome (Cinelli et al. 2022). Table 1provides a summary of common nodes found within DAGs along with information concerning whether adjusting for these nodes improves or harms causal inferences. From the DAG in Figure 1, it is easy to spot the variables that should be adjusted for. However, in practice, DAGs can get much more complicated, and the human eye will be unable to identify which variables are confounders, mediators, etc. In such cases, it is advisable that researchers rely on software such as the DAGitty website or the DAGitty R package.

**Table 1. Common nodes in a DAG**

| **Name** | **Description** | **Adjustment Status** |
| --- | --- | --- |
| Confounder (Z) | A node that causes change in the treatment and outcome: X ← Z → Y. | Adjusting for Z assists in producing an unbiased causal estimate. |
| Mediator (M) | A node that is caused by the treatment and influences the outcome: X → M → Y. | Adjusting for M creates bias by preventing the total causal effect from being estimated. |
| Collider (C) | A node whose status is changed by both the treatment and outcome: X → C ← Y. | Adjusting for C creates a spurious association between X and Y. |
| Parent of Outcome (PY) | A node that causes change in the outcome: PY → Y. | Adjusting for PY reduces variation in Y and may help improve the precision of a causal estimate. |
| Parent of Treatment (PX) | A node that causes change in the treatment: PX → X. | Adjusting for PX reduces variation in X and may damage the precision of a causal estimate. |

Given that DAGs reflect assumptions from their creators, and do not provide any quantitative output themselves, researchers may be tempted to speed their way through Step 1. However, it should be emphasized that a poor causal model can render subsequent causal estimates incorrect or highly sensitive to unobserved confounding. Without careful attention to the theoretical processes that connect to , researchers run the risk of failing to identify confounders and incorrectly specifying causal relationships. A highly technical and complex research design cannot overcome a bad theoretical framework and causal model. As a result, researchers should take their time and carefully consider the validity of their causal models.

**Step 2: Develop an identification strategy**

Following the development of a causal model, a researcher must develop an identification strategy that generates a statistical estimate and justifies the interpretation of this estimate as a causal effect. The term “identification strategy” derives from the inherent difficulty in ***identifying*** causal effects due to the fundamental problem of causal inference (Keele 2015). In an identification strategy, researchers should be explicit concerning the assumptions required by a particular strategy to estimate causal effects and the degree to which these assumptions are satisfied.

As referenced previously, the RCT is often cited as an identification “gold standard” given that it permits researchers to randomly assign treatment, ensuring that no observed or unobserved factor is biasing the treatment effect. However, RCTs are often not practical or ethical in most conflict management settings. Further, RCTs face their own unique struggles concerning external validity, attrition bias, selection biases, and more (**Keele 2015, Deaton and Cartwright 2018**). Barring the RCT, scholars can explore other identification strategies that mimic random assignment, such as instrumental variables (IV) or regression discontinuity designs (RDD). In instrumental variables settings, researchers are tasked with discovering an “instrument”, some variable that, meeting strict assumptions, approximates the aspects of treatment that do not suffer from confounding. Given the restrictive assumptions required to identify a valid instrument, accompanied by the intense debate surrounding seemingly “valid” instruments (**Mellon 2021**), the instrumental variables approach may not always be a practical option for many researchers. If a scholar can identify a discontinuity, a clear cut-off criterion in which some units receive treatment and others do not, a regression discontinuity design may reflect a fruitful next step given that treatment assignment near the discontinuity can be considered effectively randomized. However, discontinuities are rare, especially in the conflict management literature where clear and observed criteria for the application of policies such as PKOs, foreign aid, mediation, etc. are not apparent or perhaps even do not exist.

Inevitably, most conflict management scholars find their projects ill-suited for random assignment. As a result, the bulk of causal research in conflict management will likely be developed under the “selection on observables” identification strategy in which a researcher acknowledges that treatment is not randomly assigned, and observations are imbalanced along a set of observed covariates. However, if the appropriate covariates are adjusted for (through methods such as regression adjustment, matching, or inverse probability weighting), treatment can be considered random ***conditional*** on the observed covariates. Of course, selection on observables requires the satisfaction of the ***ignorability*** assumption that all relevant confounding effects have been adjusted for. Fundamentally, this assumption is ***impossible*** to directly satisfy. A scholar can develop a DAG and identify ten confounders and still not be certain that a hidden eleventh confounder has not been identified and accounted for. As a result of this conundrum, sensitivity analyses (elaborated on in Step 3) have been developed to assess the plausibility and consequences of unobserved confounders.

Before discussing the final step of the causal analysis process, it is important to review identification assumptions that are applicable to ***all*** identification strategies. First, the positivity assumption (also known as common support) requires that, for each combination of covariates, there exists treated and control units. The logic for this assumption is intuitive. A similar non-treated unit must be available to serve as a counterfactual for a treated unit. For example, in the evaluation of the effect of PKOs on economic development, a researcher may adjust for conflict duration, conflict intensity, peace agreement presence, government military capacity, etc. Along these four dimensions, treated and non-treated units should be represented in each possible combination of the four covariates for the positivity assumption to be satisfied.

Second, the so-called stable unit treatment value assumption (SUTVA) encompasses two criteria, the consistency assumption, and the non-interference assumption. The consistency assumption requires that treatment be uniformly applied to all units. The non-interference assumption requires that the treatment status of a treated unit does not affect the outcome of a non-treated unit. If non-interference is violated, non-treated units cannot serve as pure controls and counterfactuals of treated units. Unfortunately, both components of SUTVA are likely to be regularly violated in conflict management research. Treatments are often measured in a manner that simplifies the underlying variation within the treatment itself (such as dummy variables measuring UN PKOs). This issue can be somewhat resolved by the adoption of a method that allows for non-binary outcomes. However, for treatments measured as counts, ordinal scales, or nominal values, extant tools may be limited. Further, conflict management programs are often designed to promote peace in the broader region for which they are applied, rather than the specific unit (city, unit, state, etc.) that treatment was officially administered to. As a result, ***formally*** non-treated units located close to treated units may have their outcomes ***informally*** affected by the treatment status of nearby units. Unfortunatley, there is no settled and robust mechanism to overcome non-interference violations, justifying increased academic attention to the issue (**Keele 2015**).

**Step 3: Execute a sensitivity analysis**

Under selection on observables, researchers are tasked with satisfying the non-verifiable ignorability assumption that all confounders have been accounted for. However, no statistic or amount of careful theoretical deliberation can confirm whether *all*confounders have been identified. Instead, researchers can rely on sensitivity analyses to quantify the sensitivity of a causal estimate to various specifications of hypothetical unknown confounders.Such an approach will provide researchers a quantifiable measure of sensitivity concerning a causal estimate. Researchers then rely on expert domain knowledge and theory to assess the plausibility of an unobserved confounder existing that could significantly alter a causal estimate. Importantly, these sorts of sensitivity analyses do not *satisfy* the ignorability assumption. Instead, they provide much needed contextual evidence concerning the robustness of results. In the following section, I develop a research design to examine the impact of UN PKOs on economic development, incorporating each of these three steps into the development of the research design. Prior to this, I offer a brief review of the PKO-development literature.

**United Nations peacekeeping and economic development**

Many scholars have documented the phenomena of “peacekeeping economies”, where higher levels of economic activity in a host country are directly attributed to the deployment of an international peacekeeping force (**Jennings 2015, 2018, Rolandsen 2015, Jennings and** Bøås **2015**). Following a PKO deployment, there are numerous channels through which a PKO may contribute to economic growth and development. PKOs can assist in the peace agreement process and police and monitor armed and criminal actors in the post-conflict context, leading to a safer environment in which economic activity can recover. A consistently stable environment may also incentivize increased domestic and foreign business activity and investment as the rule of law and property protection is fostered. Additionally, the presence of a large foreign force artificially stimulates demand for a variety of goods and services such as housing, food, transportation, retail, etc. PKOs can assist in the development of civil infrastructure projects such as schools and hospitals and can directly or indirectly (through INGOs) provide immediate relief through humanitarian aid.

Alternatively, it may also be the case that economic benefits of peacekeeping economies are not sustainable and are PKO dependent. For example, when a PKO withdraws, the artificial demand for a variety of goods and services from the PKO personnel subsequently go away. If the core issues of the prior conflict were not resolved and an enduring peace was not established, the stabilization that fostered economic activity provided by a PKO may likewise erode following a PKO withdrawal. Despite sound theoretical reasoning for a causal effect of UN PKO deployments and withdrawals on economic development, prior quasi-experimental evidence of such effects is mixed (**Caruso et al. 2017, Bove and Elia 2018, Beber et al. 2019, Mvukiyehe and Samii 2021, Bove et al. 2022**).

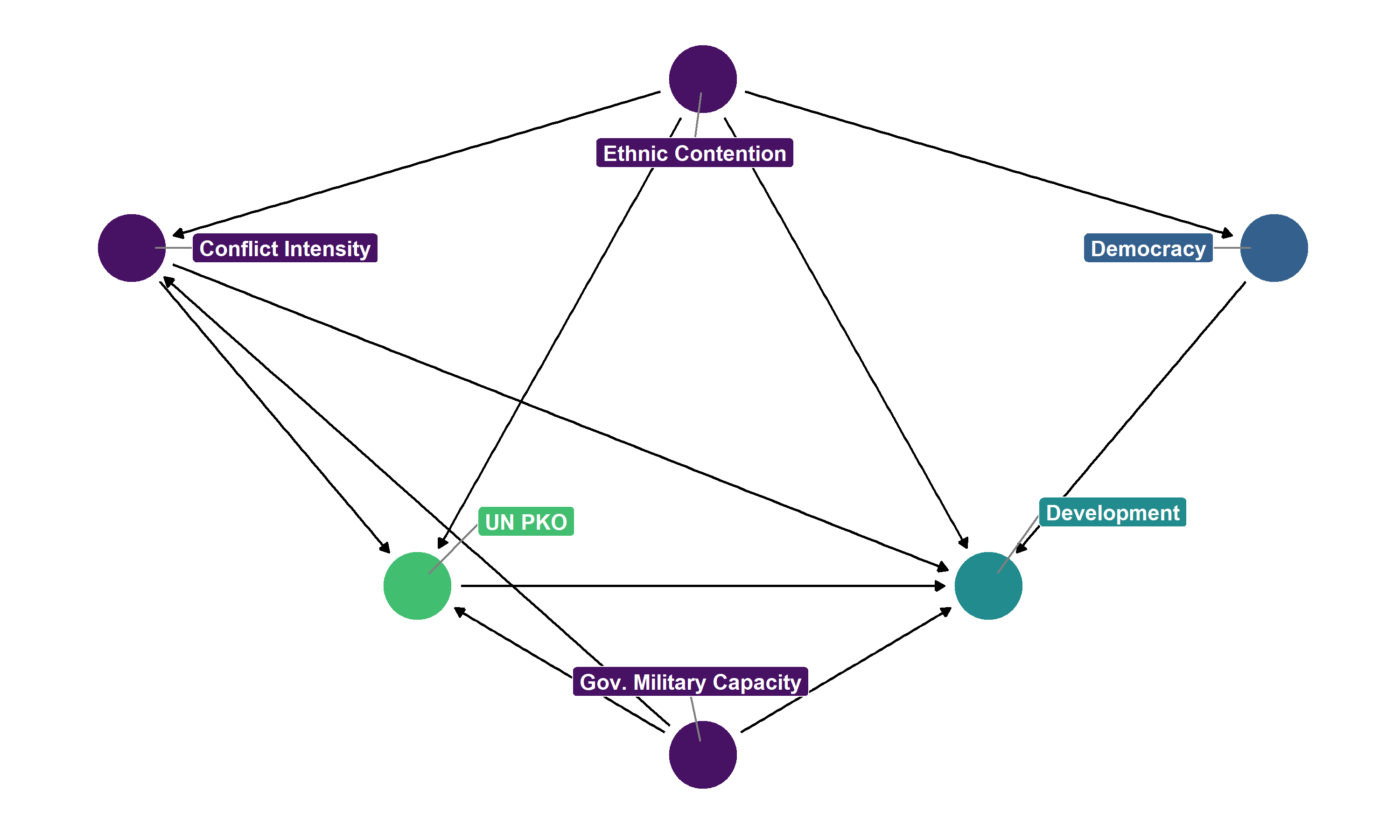
**Research** design

This paper seeks to estimate the causal effect of UN PKOs on economic development in conflict and post-conflict countries. Prior to defining and measuring treatment and outcome, I elaborate on the qualification for conflict and post-conflict cases. Cases enter the analysis when identified by the Uppsala Conflict Data Program (UCDP)/International Peace Research Institute in Oslo (PRIO) Armed Conflict Dataset (**Gleditsch et al. 2002, Davies et al. 2022**). To qualify as a conflict-state, a country must experience at least 25 battle-related deaths per year. A post-conflict state is subsequently any state that has experienced prior conflict and whose battle-related death count has fallen below the threshold. To avoid artificial cases of civil war caused by bloody military coups that resulted in at least 25 deaths, I recode instances of military coups (defined by Powell and Thyne 2011) as non-civil war. Finally, temporary lulls in conflict may result in the battle deaths threshold dropping below 25 despite conflict remaining ongoing. A consequence of leaving this unaddressed is an incorrectly inflated number of short post-conflict “peace” spells that represent short-term lulls in conflict intensity. To correct for this, I recode three-year (or less) spells of “peace” as conflict-level cases.

Information on the location of UN peacekeeping operations is acquired from the Geocoded Peacekeeping Operations (Geo-PKO) Dataset v. 2.0 (**Cil et al. 2020**) which covers UN PKOs from 1994-2020. UN PKOs are measured using a binary variable due to practical issues with implementing a non-binary treatment given the methods employed for this analysis. UN PKOs feature a great degree of diversity in scope, so it is advisable that future studies embrace alternative methods that account for the variability ***within-treatment***. I restrict the measurement of PKOs towards those that are not designed to mitigate ***interstate conflict***. Missions designed to manage conflict between states are often clustered along borders and the extent to which such PKOs will impact the ***overall*** economic performance of a host nation is questionable. Still, for robustness purposes, I estimate the effect of UN PKOs deployments and withdrawals using the restricted (intrastate conflict management) and non-restricted (intra- and interstate conflict management) criteria. To measure economic development, I follow standard convention and use GDP per capita (log-transformed). Data on GDP per capita is acquired from **Fariss et al. (2022)**.

**Causal model**

Step 1 of a causal research design requires a formal representation of a causal model that a researcher molds their analysis around. Figure 2reports a DAG that formally states my causal assumptions regarding the relationship between UN PKOs and economic development along with common or theoretically-relevant covariates. The green node denotes treatment while the teal node denotes outcome. Purple nodes are confounding effects, and the blue node is a parent of the outcome. As a result, this DAG suggests that government military capacity, conflict intensity, and ethnic contention are necessary to adjust for to eliminate confounding bias from these effects. Figure 2also implies that it may improve the precision of the causal estimate if democracy is adjusted for. Many readers may observe the DAG and immediately raise objections. This is a good response, and it highlights the partial value of DAGs as tools to make theoretical assumptions transparent. Below, I clarify my assumptions illustrated in this DAG.



**Figure 2: Causal model of UN PKO-development link**

First, some may contend that the DAG is fairly sparse. After all, how can the study of UN PKOs and development be simplified to six nodes? First, it is important to note that not *all*variables need to be included in a DAG (given that parents of outcome and colliders should not be adjusted for, it makes little sense to include them in a final DAG). Second, some may argue that I am leaving out an important confounder. This is a completely valid point that motivates the use of sensitivity analysis which is executed in this research design. If indeed a relevant confounder is being omitted, sensitivity analysis will quantify the strength required for an omitted unknown confounder to impact the causal estimate.

Beyond the inclusion/exclusion of certain variables, this DAG also warrants that discussion concerning the rationale behind the assumed causal mechanisms. First, many authors have argued that the UN has a general aversion to entering armed combat with military forces of the state (**Gilligan and Stedman 2003, Fjelde et al. 2018, Phayal and Prins 2019**). Such confrontation may be more likely when government military forces are more militarily capable of engaging potentially hostile forces. Stronger government forces may also be emboldened to deny UN PKO personnel from entering certain regions of the country they are deployed to. Further, increased government military capacity may also lead to higher levels of development since a stronger state has the capacity to ensure order and stability while leveraging its increased coercive capacity to extract taxes and reinvest taxes into the economy. One may contend that the causal arrow could equally flow in the opposite direction. As economic development increases, the government can reinvest increased tax revenue in its own military capacity. Thus, a feedback loop exists. DAGs must be acyclic, so the arrow between government military capacity and development cannot bidirectional. While I do not deny the existence of this feedback loop, it is important to know that the long-term relationship between government military capacity and development is not necessarily contemporaneous. It is much more likely that government military capacity precedes development given the reasons listed above. As such, I expect government military capacity within a time unit (such as a year) to ***cause*** some change in development within the same time frame. Government military capacity is measured using the log-transformed number of military personnel per capita acquired from the Correlates of War National Military Capabilities data set v.6.0 (**Singer et al. 1972, Singer 1987**).

Second, increased conflict activity may motivate the authorization of a UN PKO. Given the UN’s willingness to engage in some of the worst humanitarian crises in the post-Cold War era, a valid assumption can be made that levels of conflict intensity cause some change in the likelihood of a UN PKO being deployed. Also, increased conflict intensity features major destruction to human lives, civilian infrastructure, and various other forms of human and economic capital. Like the concern of directionality with the first confounder, one may also argue that a UN PKO impacts conflict intensity. Again, I would not contest this point. However, it is reasonable to assume that **conflict intensityt** leads to **UN PKOt** which subsequently leads to **conflict intensityt+1**. Importantly, lead and lag effects of are separate nodes from t itself. As a result, I assume that conflict intensity causes some change in the prospects of a UN PKO being authorized while also acknowledging that UN PKOs cause changes in levels of conflict intensity. To measure conflict intensity, I use a log-transformed value of the count of all battle-related and one-sided violence-related deaths. Aggregating all forms of violence into a single measure is favorable given that it allows for counts of violence in both the conflict and post-conflict context. Data on conflict intensity is acquired from UCDP’s Georeferenced Events Dataset v. 22.1 (**Sundberg and Melander 2013, Davies et al. 2022**).

Third, ethnic contention likely impacts the chances of a UN PKO being authorized given the theoretical and associational link between ethnic conflicts and various measures of violence (**Mattes and Savun 2009, Mason et al. 2011**). Causally, this should imply that UN PKOs should be less likely to deploy PKO forces to an ethnic conflict. Of course, ethnic diversity and ethnic contention are not the same concept. Nor does ethnic diversity imply ethnic contention. Still, the underlying theoretical logic is that UN PKOs are deployed with the expectation to succeed and being bogged down in a quagmire of ethnic conflict is a possibility UN decision-makers consider. That is, the underlying level of ethnic contention in a country should impact the decisions of relevant policymakers at the UN. Ethnic contention is a confounding effect because it may have negative associations with poor developmental prospects and growth rates (**Easterly and Savine 1997, Collier 2000, Alesina et al. 2003**). Rather than promoting general welfare through public infrastructure, ethnic contention may foster competition over public goods which become prioritized for certain ethnic groups at the expense of others. Such competition can result in poor government and economic management. I measure ethnic contention using the Variety of Democracy’s (V-Dem) exclusion by social group index. This index is a combined metric of other indices that measure the distribution of power, civil liberties equality, access to public services, access to state jobs, and access to state businesses by social group. A social group as defined by V-Dem encompasses any identity based on “caste, ethnicity, language, race, region, religion, migration status, or some combination thereof” (**Coppedge et al. 2022, p. 212**).

Lastly, I adjust for the level of democracy to improve causal estimates. Democracy is not considered a confounding effect in this analysis given the lack of theoretical or empirical material suggesting that the regime type of a country in civil war motivates the decision for the UN to intervene. However, because I do not assume that democracy is a confounding effect, rather, a predictor of outcome, adjusting for democracy serves to lessen the variation in development and generate a more precise estimate. One could criticize this assumption as failing to account for the unclear causal direction between democracy and development. However, current literature suggests that democracy has an independent effect on economic growth and development (**Acemoglu et al. 2019**). Democracy is measured using V-Dem’s electoral democracy index (**Teorell et al. 2019, Coppedge et al. 2022**). I opt for an index measure of democracy rather than a binary indicator to increase balance between treated and non-treated units. Embracing an arbitrary cut-off point for democracy is problematic for causal inference because within-binary classifications host a great degree of variability (a democracy with an electoral democracy score of 0.6 is different than a democracy with a score of 0.9). Using an index over a binary allows for the construction of more accurate counterfactuals.

**Identification strategy**

Because UN PKOs are not randomly assigned, the effect of UN PKOs on economic development cannot be identified through an RCT. Further, a natural experiment is difficult to implement given limitations on satisfying associated assumptions. For example, valid instruments are difficult to justify in the PKO literature (**Gilligan and Sergenti 2008**) and seemingly valid instruments are subject to criticism over time (**Mellon 2021**). Discontinuities are non-existent for PKO deployments and withdrawals since the UN does not adhere to a strict set of criteria concerning which countries receive and do not receive PKOs. With these options exhausted, this design operates under the selection on observables identification strategy. Because it is impossible to know the entire set of covariates to adjust for, I execute a sensitivity analysis (elaborated on in greater detail in the following section), to estimate the issue of unobserved confounding.

To estimate the effect of PKOs on economic development, this analysis employs the novel panel data matching design developed by **Imai et al. (2021)**. Matching designs are lauded for their ability to create data sets evenly balanced along a specified set of covariates. In doing so, similar observations are compared to each other that differ in treatment status. The average difference in outcome between similar treated and non-treated units represents an average treatment effect on the treated (assuming the ignorability assumption is satisfied). Due to its ease of implementation and the intuitive logic of matching methods, matching designs are popular under the selection on observables identification strategy.

In the canonical setting, matching methods are applied to evaluate a scenario with a single treatment for various cross-sectional units at a given time. However, applying conventional matching techniques for settings when this is not the case introduces several complications. Within panel data, cross-sectional units can receive multiple treatments over time, units can reverse their treatment status over time, and treatment itself can be applied at different time intervals for different cross-sectional units. Acknowledging these deviations from the canonical setting, **Imai et al. (2021)** introduced a novel matching method that allows for matching analysis to be executed on panel data. Rather than a unit-time observation being matched with another similar unit-time observation, control observations are selected based on an identical pre-treatment history for a specified timespan (a matched set). For example, if a researcher specified a four-period lag, units would only be matched if their pre-treatment history of treatment is identical four-periods prior to treatment. Following this, a refinement method (nearest-neighbor matching, propensity score matching, inverse probability weighting, etc.) is applied that balances covariates between treated and non-treated units. Finally, a difference-in-differences (DID) estimator is applied which allows for the estimation of post-treatment non-contemporaneous treatment effects. This feature is particularly helpful given that there are reasons to suspect that UN PKO deployments and withdrawals create economic effects beyond the immediate time of their deployment/withdrawal.

Per the suggestions of **Imai et al. (2021)**, I iteratively experiment with different pre-treatment thresholds to determine matched sets. Researchers using this method face a clear tradeoff between statistical power and the validity of matched sets. On the one hand, defining a loose criterion for a matched set (for example, units are matched if their treatment status the year prior to treatment is the same) allows the user to keep many observations. However, as the number of pre-treatment years increases as a criterion for the generation of a matched set, the number of possible observations decreases. While this can dramatically reduce ***N***, such steps ensure that matched sets feature observations that are more similar to each other. Given the short temporal coverage of the data used in this analysis, stricter pre-treatment history criteria reduce the number of countries being analyzed to an impractically low number. As a result, unique matched sets are created using a 1 to 3 year lag. Three separate lag criteria are utilized to assess the robustness of results to alternative lag specifications, a concern that is necessary to address given the comparatively small spatial coverage of treatment where the exclusion of treated units due to a lack of a match have the capacity to notably shift the results.

Following the specification of matched sets, I examine how well different refinement methods balance covariates between treated and non-treated units. An effective refinement method should be one that minimizes the differences between treated and non-treated units on a specified set of covariates. In the context of this project, the goal of refinement is to create a balanced sample where countries that receive UN PKOs and those that do not are, on average, similar with respect to government military capacity, ethnic contention, conflict intensity, and democracy. **Figure 3** reports the standardized mean difference between these covariates between treated and non-treated units across three refinement methods. The first two refinement methods reflect nearest-neighbor matching using the Mahalanobis distance with a single match and up to five matches. Second, inverse probability weighting (IPW) is selected. Like matching, IPW generates weights to balance the data set along specified confounders. Unlike matching, IPW does this in a manner that does not omit any observations. IPW does so by following a two-step process. First, a numeric value for the propensity of receiving treatment predicted by specified covariates is generated for each observation, estimated with logistic regression for this analysis. Next, these observations are weighted by the inverse of their treatment propensity. Observations are weighted more heavily when their propensity to receive treatment differs largely from their ***actual*** exposure to treatment. Observations that experience the treatment in accordance with the predicted propensity to receive treatment are less heavily weighted. Rather than relying on distance to another observation to be weighted, IPW does not require any data points to be dropped as weights are generated agnostic of other observations.

Chart, line chart

Description automatically generated

**Figure 3: Covariate balance between treated and non-treated units**

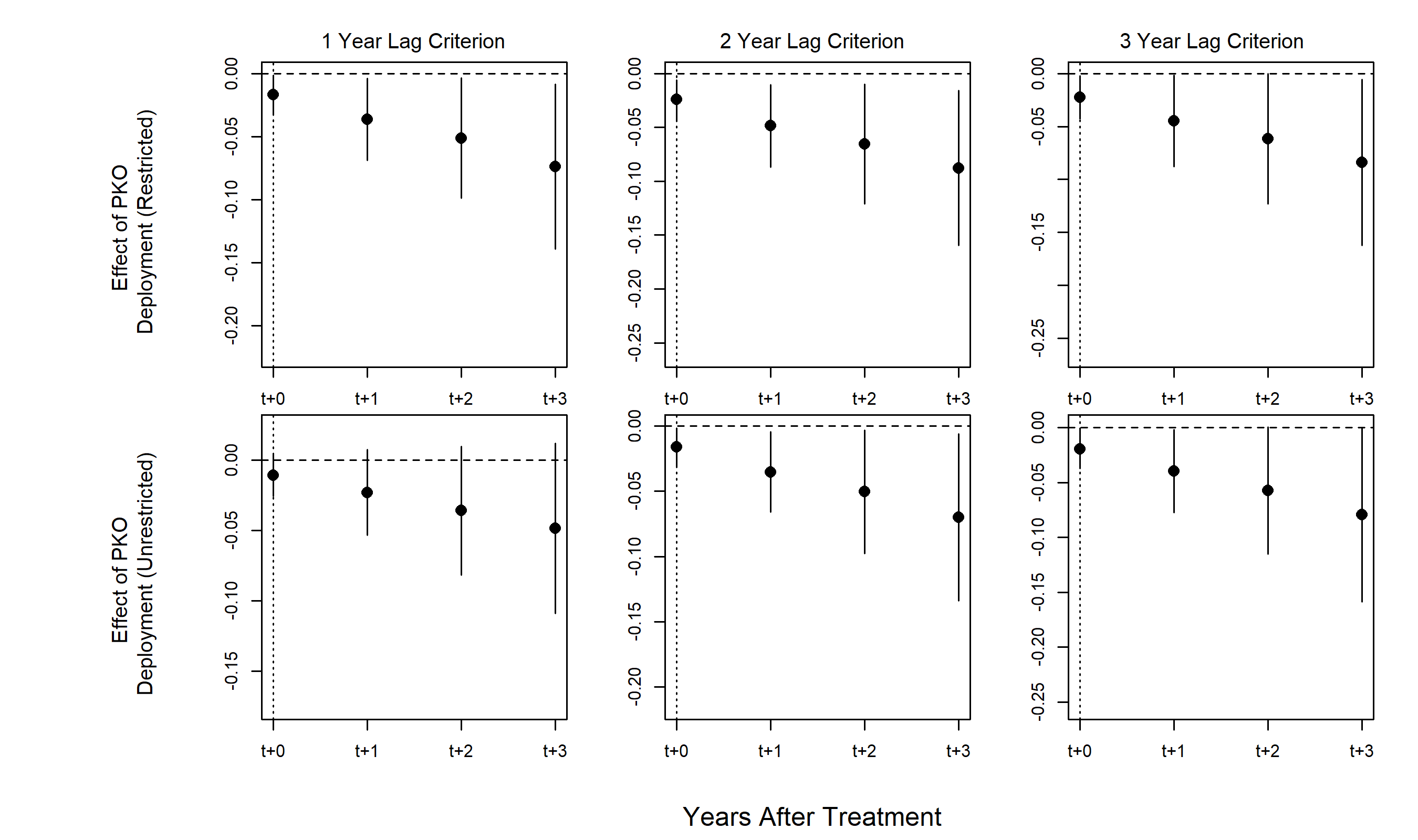
In Figure 3, the x-axis reflects the standardized mean difference between covariates for the treated and control units prior to the execution of the specified refinement method. The y-axis reflects the standardized mean difference between covariates for the treated and control units following the execution of the specified refinement method. Each dot in the plot reflects a specific covariate. Dots falling below the 45-degree dashed line suggest an improvement in covariate balance post-refinement. Dots falling above the dashed line imply that the select refinement method decreased covariate balance. To reach an ideal level of covariate balance, dots should cluster at the lower end of the y-axis. As Figure 3 demonstrates, IPW, across each lag criteria, clearly preforms the best of the three refinement methods at minimizing differences between treated and non-treated units. As a result, IPW is the selected refinement method employed for the remainder of the analysis.

**Sensitivity analysis**

A skeptical observer may rightly question whether four variables are sufficient to satisfy the ignorability assumption. With the assistance of sensitivity analysis, it is possible to assess the extent to which the omission of confounders biases estimated results. While the variety of sensitivity analyses is large, I follow the approach of Cinelli and Hazlett (2020). In this approach, a researcher specifies a benchmark covariate that the researcher believes should have a large substantive effect on the outcome. Hypothetical unobserved confounders are then constructed with varying levels of strength relative to the specified benchmark covariate (one times as large as the benchmark covariate, twice as large, three times as large, etc.). Following this, researchers can numerically and graphically assess the extent to which an unobserved confounder of varying strength has the capacity to reduce or flip a causal estimate. Unfortunately, this method is not directly applicable for panel data matching. Given the novelty of the panel data matching approach, this is not surprising. Instead of estimating the sensitivity of the estimates generated from the panel data matching analysis, I utilize sensitivity analysis to demonstrate the sensitivity of an estimate using standard inverse probability weighting on a model agnostic to the panel nature of the data. While this approach is not applicable for the results of the panel data matching analysis *in particular*, it is insightful for quantitatively assessing the sensitivity of estimates using a popular approach under the selection on observables identification strategy.

**Results**

Figure 4reports the contemporaneous and up to three year lead estimated effect of UN PKOs using a variety of lag specifications for the matched set criterion. The top row reports these estimated effects using the restricted non-interstate PKO operationalization while the bottom row reports estimated effects of PKOs on economic development with the unrestricted (all PKOs) measure. 90% confidence intervals are reported around point estimates. Across different matching criteria and measures of UN PKOs, the estimated effect of UN PKOs on economic development is surprisingly *negative*and *decreases* over time following a UN PKO. Both the contemporaneous and lead effects of UN PKOs are largely statistically significant (or hover around statistical significance) at the *p<0.1* level. Although, the estimated effect of UN PKOs on economic development with the unrestricted PKO measure and a 1-year lag criteria for generating a matched set reports 90% confidence intervals that overlap with zero. Given the estimated effects are generally not very sensitive to matched set criterion or the PKO measure selected, I consider the estimates from the 3-year lag criterion model with the restricted PKO measure as the most representative of the true effect of UN PKOs on economic development. As discussed earlier, a longer pre-treatment specification creates matched sets that are more similar with respect to treatment history (ensuring more similar counterfactuals) and the restricted PKO measure captures the types of PKOs that are considered to be the most relevant for managing intrastate conflict. The results from this model suggest that, for the countries that received a UN PKO, UN PKOs lead to an immediate 2.2% decrease in GDP per capita and an 8.4% decrease in GDP per capita three years following a PKO deployment.

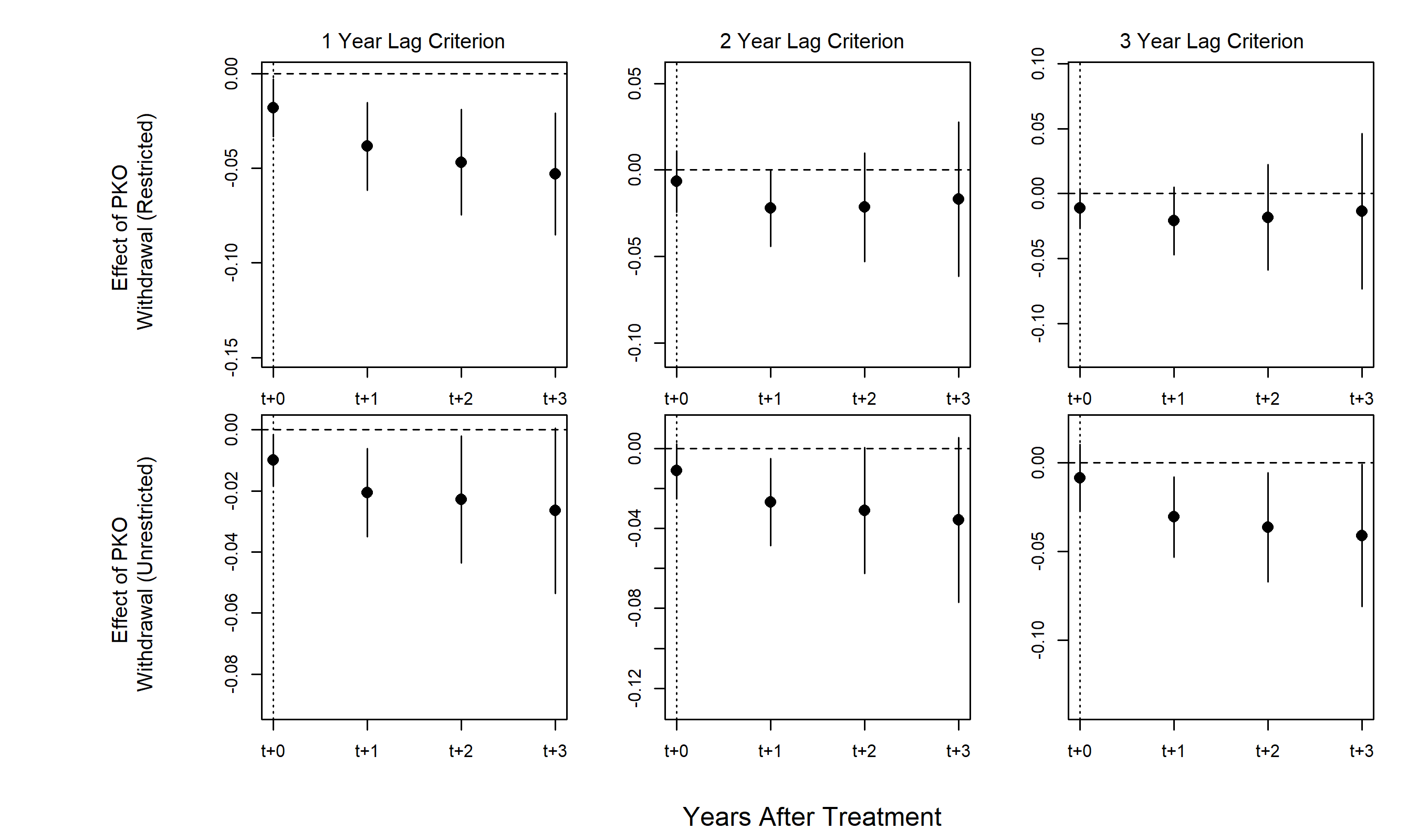


**Figure 4: Estimated economic effect of UN PKO Deployments**

Initially, these results appear to be very counterintuitive. Not only is the contemporaneous estimated effect negative, something that could potentially be explained by a delayed economic effect of PKOs, but the estimated lead effects are ***increasingly*** negative. While these results stand in contrast to works in the quasi-experimental literature on PKOs and development (**Caruso et al. 2017, Beber et al. 2019, Bove et al. 2022**), they also affirm other quasi-experimental studies that find counterintuitive null or negative effects (**Bove and Elia 2018, Mvukiyehe and Samii 2021**). Nonetheless, the results remain surprising given that, while PKOs may generate negative economic effects (such as PKO-dependent economies that are sensitive to PKO ***withdrawals***), immediate and longstanding declines in development attributed to PKO deployments represents a puzzle for both the theoretical and descriptive research on PKOs and development. Further, any selection bias explanation is countered by the implementation of the inverse probability weights and the construction of matched sets.

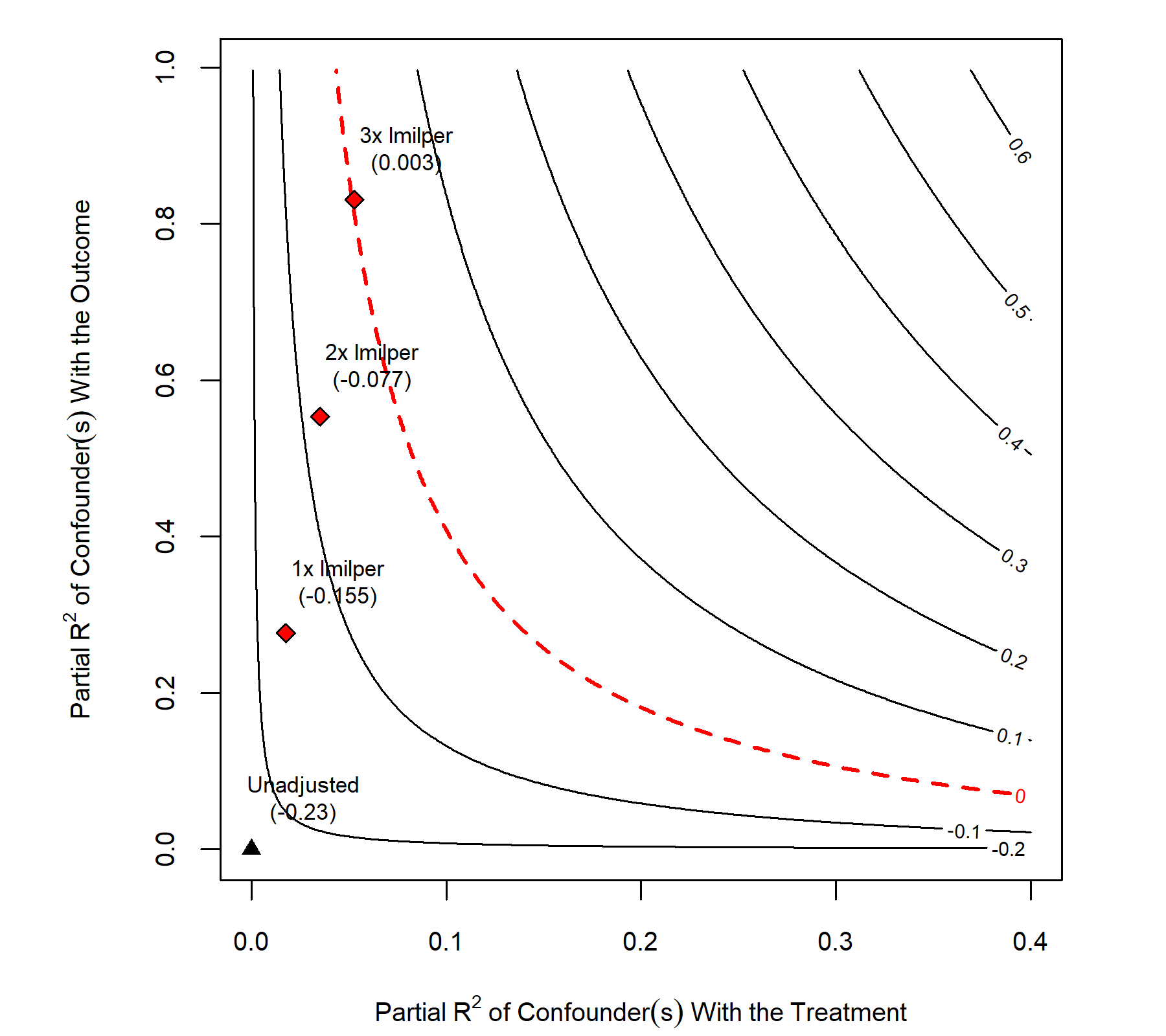
One solution to this puzzle may lie in the exploration of heterogenous treatment effects. For example, the average effect may be masquerading important distinctions in the effect of PKOs stratified by differing contexts. It may be the case that UN PKO deployments cause an increase in economic development ***only in certain contexts***. Likewise, UN PKO deployments may lead to increases in development ***generally***. However, a specific circumstance or set of circumstances may lead to dramatic negative declines in development, which may drag the average effect towards the negative. A second concern that should be addressed to explain this finding is the lack of consistency for “treatment” itself. As explained earlier, a binary measure for UN PKOs is not optimal given the wide variability concerning scope of mission, military personnel count, police personnel count, etc. Employing a continuous or count measure of treatment may be more appropriate and could lead to the estimation of results that are generally more aligned with existing theoretical and descriptive research.

**Figure 5** reports the estimated effect of PKO withdrawals on economic development among countries who experienced a (or several) PKO withdrawals. Unlike the PKO deployment results, estimated effects for PKO withdrawals are sensitive to both treatment history matching criteria ***and*** the restrictiveness of the PKO measure. Naturally, this suggests some caution in the interpretation of these results. Given that the restricted measure of UN PKOs improves consistency of treatment, I choose to focus on these results as a more accurate representation of the economic effect of UN PKOs. However, disparities in estimated effects due to matched set criteria pose an interesting interpretative question. While decreasing the number of years for the matched set criterion increases ***N***, it decreases the similarity between cases that are being matched, thereby creating fewer representative counterfactuals. Indeed, understanding the tradeoffs between the two is important given that the results suggest opposite outcomes (albeit the confidence intervals for the two- and three-year lag criterion overlap with zero). Nonetheless, while valid counterfactuals may have been omitted from the analysis with a stricter matched set criterion, invalid counterfactuals were almost certainly eliminated as well. Given that the endeavor of estimating causal effects should be a conservative one (false negatives are less impactful than false positives), I argue that the estimates generated from the models employing the stricter matched set criterion are more representative of the economic effect of PKO withdrawals than the alternative estimates. While it is tempting to note a potential recovery effect that the models using stricter matching criteria and the restricted PKO measure report, the confidence intervals overlap with zero, suggesting increased uncertainty of the estimates as the time post-withdrawal increases.



**Figure 5: Estimated economic effect of UN PKO withdrawals**

In any observational study, it is important to make an effort to assess the sensitivity of estimates to unobserved confounding. Figure 6reports the results of a sensitivity analysis designed to assess the sensitivity of IPW-generated estimates to unobserved confounding. Again, it is important to note that this sensitivity analysis *does not*assess the sensitivity of the estimates using the method developed by Imai et al. (2021). As noted in the research design, extant sensitivity analyses are not currently supported for the panel data matching design as implemented by Imai et al. (2021). Instead, the analysis below examines the sensitivity of estimates generated from a linear regression model using inverse probability weights, agnostic to the panel structure of the data. While not optimal, this strategy reflects the “next-best” approach in an honest effort to proxy an assessment of the ignorability assumption.



**Figure 6: Sensitivity analysis**

Because the size of a government’s military capability likely has a great degree of influence on the presence of a UN PKO (the UN may be hesitant to engage/stay in a conflict where the government is more capable) and the prospects for political stability and development, logged military personnel per capita (“lmilper”) is selected as a benchmark covariate. Simply put, this means that unobserved confounders are interpreted in reference to the size of the relationship between economic development and military personnel per capita. The coefficient for the PKO variable using inverse probability weighting is -0.23 (the unadjusted black triangle in **Figure 6**). If an unobserved confounder with a relationship with GDP per capita 1x as strong as the relationship between military personnel per capita and GDP per capita was specified in the IPW model, the PKO coefficient would increase to -0.155. If an unobserved confounder was 2x as strong, the sign on the coefficient for PKO would move closer to zero (the dotted red line). Finally, if an unobserved confounder was incorporated that was 3x as strong, the sign on the PKO coefficient would flip. Although, the effect is so small that it is practically a null effect.

Naturally, researchers must ask themselves, “is an unobserved confounder 3x as strong as the military personnel-GDP per capita effect likely to exist”? Theoretically, this analysis can be expanded to assess the estimate from an effect 4, 5, 6x, etc. as strong as “lmilper”, but these are all hypothetical unobserved confounders. Useful interpretation of these results is given only when researchers assess the plausibility of various hypothetical unobserved confounders. In this case, the effect of PKOs does not approach zero until a hypothetical confounder 3x as large as the military personnel-GDP per capita effect is incorporated into the model. Given the strong theoretical links between government military capacity and PKOs/GDP per capita, such a massive confounder going unobserved in this literature seems unlikely. This does not ensure, however, that unobserved confounding is not a serious issue. Rather, this analysis stresses the need for further theorizing so that researchers can identify and adjust for the entire plausible range of confounders in pursuit of estimating causal effects.

**Conclusion**

Traditionally, research on peacekeeping has been devoted to understanding its potentially pacifying effects. However, an emerging theoretical and empirical literature has likewise extended the study of peacekeeping to examine the potential economic effects of peacekeeping operations. Existing quantitative research is not clear, however, on whether UN PKOs have a substantive effect, negative or positive, on economic development. Employing a quasi-experimental design using matching and weighting techniques extended to a panel data framework, the analysis in this paper compliments findings from other scholars employing quasi-experimental designs who generally fail to find a consistent, positive effect of UN PKOs on economic development. Indeed, PKO deployments were found to have a negative impact in host country economies.

While this paper sought to make causal inferences concerning the economic effects of UN PKOs, a number of concerns place asterisks around the findings of this paper. Primarily, issues surrounding the violation of SUTVA warrant great caution in the causal interpretation of these results. Researchers can assuage the consistency assumption by incorporating fluid measures of PKO presence/activity rather than utilizing a binary. However, satisfying the non-interference assumption remains a serious and large task. Such issues are not limited to the analysis or specific topic of this paper. They remain present for the entirety of the conflict management literature. Following SUTVA concerns, researchers should consider alternative strategies to estimate causal effects in the peacekeeping literature. For example, scholars can execute sub-national analyses to ensure the generation of more comparable counterfactuals and more consistent treatments. Scholars may also embrace alternative methods, such as the synthetic control method, that may prove fruitful for analyzing the economic impacts of ***ac*** UN PKOs and bolstering qualitative analysis.

Overall, I argue the primary contribution of this paper stems from its illustrative guideline for causal research, rather than serving as a novel entry in PKO-development literature. A number of critiques could (and should) be made concerning the causal model and identification strategy developed in this work. Making causal inferences is difficult and, if scholars seek to identify causal effects, causal models and identification strategies should be critically examined and thoughtfully criticized. However, despite this difficulty, this endeavor is worthwhile if scholars are interested in answering the questions the public and policymakers are concerned with. It is my primary hope that this paper serves as a reference and a source of motivation for future causal research in the conflict management literature.

**References**

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2019. “Democracy Does Cause Growth.” *Journal of Political Economy* 127(1): 47–100.

Achen, Christopher H. 2005. “Let’s Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong.” *Conflict Management and Peace Science* 22: 327–39.

Alesina, Alberto et al. 2003. “Fractionalization.” *Journal of Economic Growth* 8: 155–94.

Bara, Corinne. 2020. “Shifting Targets: The Effect of Peacekeeping on Postwar Violence.” *European Journal of International Relations* 26(4): 979–1003.

Beardsley, Kyle, David E. Cunningham, and Peter B. White. 2018. “Mediation, Peacekeeping, and the Severity of Civil War.” *Journal of Conflict Resolution* 63(7): 1682–1709.

Beardsley, Kyle, and Kristian Skrede Gleditsch. 2015. “Peacekeeping as Conflict Containment.” *International Studies Review* 17(1): 67–89.

Beber, Bernd, Michael J. Gilligan, Jenny Guardado, and Sabrina Karim. 2019. “The Promise and Peril of Peacekeeping Economies.” *International Studies Quarterly* 63(2): 364–79.

Bove, Vincenzo, and Leandro Elia. 2018. “Economic Development in Peacekeeping Host Countries.” *CESifo Economic Studies* 64(4): 712–28.

Bove, Vincenzo, and Andrea Ruggeri. 2016. “Kinds of Blue: Diversity in UN Peacekeeping Missions and Civilian Protection.” *British Journal of Political Science* 46(3): 681–700.

———. 2018. “Peacekeeping Effectiveness and Blue Helmets’ Distance from Locals.” *Journal of Conflict Resolution* 63(7): 1630–55.

Bove, Vincenzo, Jessica Di Salvatore, and Leandro Elia. 2022. “UN Peacekeeping and Households’ Well‐Being in Civil Wars.” *American Journal of Political Science* 66(2): 402–17.

Caruso, Raul, Prabin Khadka, Ilaria Petrarca, and Roberto Ricciuti. 2017. “The Economic Impact of Peacekeeping: Evidence from South Sudan.” *Defence and Peace Economics* 28(2): 250–70.

Cil, Deniz, Hanne Fjelde, Lisa Hultman, and Desirée Nilsson. 2020. “Mapping Blue Helmets: Introducing the Geocoded Peacekeeping Operations (Geo-PKO) Dataset.” *Journal of Peace Research* 57(2): 360–70.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2022. “A Crash Course in Good and Bad Controls.” *Sociological Methods & Research* 0(0): 1–34.

Cinelli, Carlos, and Chad Hazlett. 2020. “Making Sense of Sensitivity: Extending Omitted Variable Bias.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82(1): 39–67.

Clarke, Kevin A., Brenton Kenkel, and Miguel R. Rueda. 2018. “Omitted Variables, Countervailing Effects, and the Possibility of Overadjustment.” *Political Science Research and Methods* 6(2): 343–54.

Collier, Paul. 2000. “Ethnicity, Politics and Economic Performance.” *Economics and Politics* 12(3): 225–45.

Coppedge, Michael et al. 2022. “V-Dem Codebook V12.” *Varieties of Democracy (V-Dem) Project*. <https://www.v-dem.net/data/the-v-dem-dataset/>.

Davies, Shawn, Therése Pettersson, and Magnus Öberg. 2022. “Organized Violence 1989–2021 and Drone Warfare.” *Journal of Peace Research* 59(4): 593–610.

Deaton, Angus, and Nancy Cartwright. 2018. “Understanding and Misunderstanding Randomized Controlled Trials.” *Social Science & Medicine* 210: 2–21.

Easterly, William, and Ross Levine. 1997. “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *The Quarterly Journal of Economics* 111(4): 1203–50.

Fariss, Christopher J., Therese Anders, Jonathan N. Markowitz, and Miriam Barnum. 2022. “New Estimates of Over 500 Years of Historic GDP and Population Data.” *Journal of Conflict Resolution* 66(3): 553–91.

Fjelde, Hanne, Lisa Hultman, and Desirée Nilsson. 2018. “Protection Through Presence: UN Peacekeeping and the Costs of Targeting Civilians.” *International Organization* 73(1): 103–31.

Gerring, John. 2012. “Mere Description.” *British Journal of Political Science* 42(4): 721–46.

Gilligan, Michael J., and Ernest J. Sergenti. 2008. “Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference.” *Quarterly Journal of Political Science* 3(2): 89–122.

Gilligan, Michael J., and John Stedman. 2003. “Where Do the Peacekeepers Go?” *International Studies Review* 5(4): 37–54.

Gleditsch, Nils Petter et al. 2002. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research* 39(5): 615–37.

Hultman, Lisa, Jacob Kathman, and Megan Shannon. 2013. “United Nations Peacekeeping and Civilian Protection in Civil War.” *American Journal of Political Science* 0(0): 1–17.

———. 2014. “Beyond Keeping Peace: United Nations Effectiveness in the Midst of Fighting.” *American Political Science Review* 108(4): 737–53.

Imai, Kosuke, In Song Kim, and Erik H. Wang. 2021. “Matching Methods for Causal Inference with Time‐Series Cross‐Sectional Data.” *American Journal of Political Science* 0(0): 1–19.

Jennings, Kathleen M. 2015. “Life in a ‘Peace-Kept’ City: Encounters With the Peacekeeping Economy.” *Journal of Intervention and Statebuilding* 9(3): 296–315.

———. 2018. “Peacekeeping as Enterprise: Transaction, Consumption, and the Political Economy of Peace and Peacekeeping.” *Civil Wars* 20(2): 238–61.

Jennings, Kathleen M., and Morten Bøås. 2015. “Transactions and Interactions: Everyday Life in the Peacekeeping Economy.” *Journal of Intervention and Statebuilding* 9(3): 281–95.

Keele, Luke. 2015. “The Statistics of Causal Inference: A View from Political Methodology.” *Political Analysis* 23(3): 313–35.

Mason, T. David, Mehmet Gurses, Patrick T. Brandt, and Jason Michael Quinn. 2011. “When Civil Wars Recur: Conditions for Durable Peace after Civil Wars: When Civil Wars Recur.” *International Studies Perspectives* 12(2): 171–89.

Mattes, Michaela, and Burcu Savun. 2009. “Fostering Peace After Civil War: Commitment Problems and Agreement Design.” *International Studies Quarterly* 53(3): 737–59.

Mellon, Jonathan. 2021. “Rain, Rain, Go Away: 195 Potential Exclusion-Restriction Violations For Studies Using Weather as an Instrumental Variable.” *SocArXiv*. <https://osf.io/preprints/socarxiv/9qj4f>.

Mvukiyehe, Eric, and Cyrus Samii. 2021. “Peacekeeping and Development in Fragile States: Micro-Level Evidence from Liberia.” *Journal of Peace Research* 58(3): 368–83.

Phayal, Anup, and Brandon C. Prins. 2019. “Deploying to Protect: The Effect of Military Peacekeeping Deployments on Violence Against Civilians.” *International Peacekeeping* 27(2): 311–36.

Powell, Robert, and Clayton Thyne. 2011. “Global Instances of Coups From 1950 to 2010: A New Dataset.” *Journal of Peace Research* 48(2): 249–59.

Rohrer, Julia M. 2018. “Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data.” *Advances in Methods and Practices in Psychological Science* 1(1): 27–42.

Rolandsen, Øystein H. 2015. “Small and Far Between: Peacekeeping Economies in South Sudan.” *Journal of Intervention and Statebuilding* 9(3): 353–71.

Rubin, Donald B. 2005. “Causal Inference Using Potential Outcomes: Design, Modeling, Decisions.” *American Statistical Association* 100(469): 322–31.

Ruggeri, Andrea, Han Dorussen, and Theodora-Ismene Gizelis. 2017. “Winning the Peace Locally: UN Peacekeeping and Local Conflict.” *International Organization* 71(1): 163–85.

Samii, Cyrus. 2016. “Causal Empiricism in Quantitative Research.” *The Journal of Politics* 78(3): 941–55.

Schrodt, Philip A. 2014. “Seven Deadly Sins of Contemporary Quantitative Political Analysis.” *Journal of Peace Research* 51(2): 287–300.

Singer, J. David. 1987. “Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985.” *International Interactions* 14: 115–32.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. “Capability Distribution, Uncertainty, and Major Power War, 1820-1965.” In *Peace, War, and Numbers*, Beverly Hills: Sage, 19–48. <https://correlatesofwar.org/data-sets/national-material-capabilities/>.

Sundberg, Ralph, and Erik Melander. 2013. “Introducing the UCDP Georeferenced Event Dataset.” *Journal of Peace Research* 50(4): 523–32.

Teorell, Jan, Michael Coppedge, Staffan Lindberg, and Svend-Erik Skaaning. 2019. “Measuring Polyarchy Across the Globe, 1900–2017.” *Studies in Comparative International Development* 54(1): 71–95.