Supplemental Appendix for

Top-Category Inflation in Ordered International Relations Outcomes

Benjamin E. Bagozzi, Minnie M. Joo, and Bumba Mukherjee

Contents

A.1. TiOP(C) Model: (Log-)Likelihood Function

Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\beta}', \boldsymbol{\mu}')'$for the TiOP model and $\widehat{\boldsymbol{\theta}} = (\boldsymbol{\gamma}', \boldsymbol{\beta}', \boldsymbol{\mu}', \rho_{\varepsilon u})'$for the TiOPC model. The likelihood of the TiOP model without correlated errors is:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \prod_{i=1}^{N} \prod_{j=0} [\Pr(s_i = 1) \Pr(\widetilde{y}_i = j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N} \prod_{j>0}^{J-1} [\Pr(s_i = 1) \Pr(\widetilde{y}_i = j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N} \prod_{j=J} [\Pr(s_i = 0) + \Pr(s_i = 1) \Pr(\widetilde{y}_i = J)]^{d_{ij}}
\end{aligned}$$

where J is the top-category of the ordered dependent variable y_i, and $d_{ij} = 1$ if outcome j is realized in i ($d_{ij} = 0$ otherwise). The TiOPC (correlated errors) model's likelihood is:

$$\begin{aligned}
\mathcal{L}(\widehat{\boldsymbol{\theta}}) = & \prod_{i=1}^{N} \prod_{j=0}^{0} [\Pr(s_i = 1, \widetilde{y}_i = j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N} \prod_{j>0}^{J-1} [\Pr(s_i = 1, \widetilde{y}_i = j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N} \prod_{j=J}^{J} [\Pr(s_i = 0) + \Pr(s_i = 1, \widetilde{y}_i = J)]^{d_{ij}}
\end{aligned}$$

The TiOP model's log-likelihood is $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=0}^{J} d_{ij} \ln[\Pr(y_i = j | \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta})]$ with outcome probabilities given by equation 4 in the main paper. The TiOPC model's log-likelihood is

$$\ell(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{N} \sum_{j=0}^{J} d_{ij} \ln[\Pr(y_i = j | \mathbf{x}_i, \mathbf{z}_i, \widehat{\boldsymbol{\theta}})]$$

with outcome probabilities defined in equation 5. These log-likelihood functions can be consistently and efficiently estimated using maximum likelihood which yields asymptotically normally distributed parameter estimates.

A.2. Analyzing *Escape Flexibility*

Recall that the main dependent variable in Baccini et al. is the 0-to-4 ordered *Escape Flexibility* measure, which is an additive index that corresponds to the number of the four escape flexibility provisions (SGs, STCs, AD, CVDs) in each PTA. *Escape Flexibility*= 0 when a given PTA does not include any flexibility provision and increases to its top-most ordered category of 4 when it includes all four (i.e., maximum) flexibility provisions. Further, as illustrated in Figure 1a in the main text, the ordered *Escape Flexibility* dependent variable's top-category likely incorporates an excessive share of observations since there are three times more observations in this category than any other three outcome categories. In the main text, we argued that the preponderant share of observations in the top-category outcome—"maximum flexibility provisions"—of the ordered *Escape Flexibility* dependent variable contains both non-inflated cases and inflated-cases. We assess this claim empirically below.

A.2.1. Inflated and Non-Inflated Top-Category Cases

We suggest in the main text that the non-inflated top-category cases of maximum flexibility provisions in *Escape Flexibility* are based on the theoretically-assumed ordered continuum. This is because the non-inflated observations are generated by PTA-member states that negotiate and adopt each of the four flexibility provisions step-by-step—that is, gradationally from the minimal provision of STCs to eventually all four flexibility provisions. Step-by-step adoption of flexibility provisions occurs when PTA-member countries sequentially negotiate and obtain each flexibility provision at a time. Doing so allows them to obtain temporary trade barriers to insure their economies against exogenous trade shocks while (also) genuinely signaling their commitment to revert back to implementing PTA-mandated reforms once the trade shocks dissipate. By contrast, the inflated set of top-category observations of maximum flexibility provisions in the ordered *Escape Flexibility* measure are not based on the theoretically-assumed ordered continuum. This is because the inflated top-category observations result from PTA-member countries negotiating and adopting all four flexibility provisions in concurrently. We argued in our paper that this process is distinct from the step-by-step pursuit and adoption of flexibility provisions as PTA-member states that seek and adopt all four (maximum) flexibility provisions in a concurrent manner do so to obtain immediate protection for special interests rather than for insurance against trade shocks.

The possibility that the top-category of the ordered *Escape Flexibility* dependent variable contains a mixture of non-inflated and inflated cases along the lines described above is borne out by the empirical evidence from Baccini et al.'s data. To see why, first note that Baccini et al.'s data contains information on the *Depth* of PTA agreements adopted by each PTA member-state. *Depth* in this context is defined and operationalized within PTAs by Baccini et al. as the degree of "tariff cuts and provisions concerning services, government procurement, investments, standards, intellectual property rights, and competition" (p. 766) accepted and implemented by PTA-member states. The *Depth* index theoretically[1] ranges from a minimum of 0 to a maximum of 48, and an index value of less than 6 in this range are, as noted by trade economists, PTA-member states that engage in trade diversion to protect domestic firms from import competition (Bhagwati et al., 1999; Bhagwati, 2008; Teh et al., 2009).[2]

As such, the information provided by the data on *Depth* allows us to identify, categorize, and list two types of PTA-member states: those that use PTAs for trade diversion that results in more import discrimination (denoted as "protectionist PTA members"), and those that use PTA to tie their hands to enact trade reforms by reducing tariff barriers ("reformist PTA members"). In fact, trade economists widely recognize that protectionist PTA members tend to use the entire menu of temporary PTA trade barriers—i.e., escape flexibility provisions—to protect their domestic firms as rapidly as possible from import-competition by non-PTA

[1]Empirically, the maximum *Depth* value observed in Baccini et al.'s data is 40.

[2]This is emphasized by Bhagwati (2008) who notes that, "while Article 24 requires that the external tariffs not be raised when the PTA is formed so as to not harm nonmembers, the fact is that they can be raised when the external (MFN) tariffs are bound at higher levels than the actual tariffs. In these cases, a member of the PTA is free to raise the external MFN tariffs up to the bound level" (p.53).

countries in particular (Bhagwati 2008; Teh et al. 2009; Prusa 2016).

Building on these claims, we thus argue in the main text that protectionist PTA members who, by default, exhibit low or negligible levels of *Depth* are the inflated cases in the top-category of *Escape Flexibility* who are not based on the theoretically-assumed ordered continuum. The main reason for this is because these PTA-members negotiate and adopt all four (i.e., maximum) flexibility provisions concurrently for immediate trade protection of domestic firms. This is distinct from reformist PTA members who exhibit higher levels of depth in PTAs which mandate larger tariff cuts and deviates further away from their pre-trade agreement positions (Baccini et al. 2015). These reformist PTA members constitute the non-inflated set in the top-category of *Escape Flexibility* that are based on the theoretically-assumed ordered continuum as these states incorporate maximum PTA escape flexibility clauses in a step-by-step manner to temporarily hedge against uncertainties stemming from PTA mandated trade reforms.

Next, we use the aforementioned list of protectionist and reformist PTA states to extract and analyze the proportion (in percentage terms) of observations in the top-category—that is, "maximum flexibility provisions"—of *Escape Flexibility* that fall under the category of protectionist PTA members,[3] and those that are in the category of reformist PTA members.[4] The results from this exercise are illustrated below in Figure A.1. This figure shows that about 35% of the top-category observations in *Escape Flexibility* are protectionist PTA members that constitute the inflated set, while the remaining 65% in the top-category of this ordered dependent variable constitute the non-inflated share of observations. Table Table A.1 provides examples of these inflated and non-inflated observations found within the top-category ("maximum flexibility provisions") of *Escape Flexibility* as drawn from Baccini et al.'s data. This Table also includes brief narratives that corroborate the aforementioned classification of these illustrative examples, reinforcing our contentions of top-category inflation

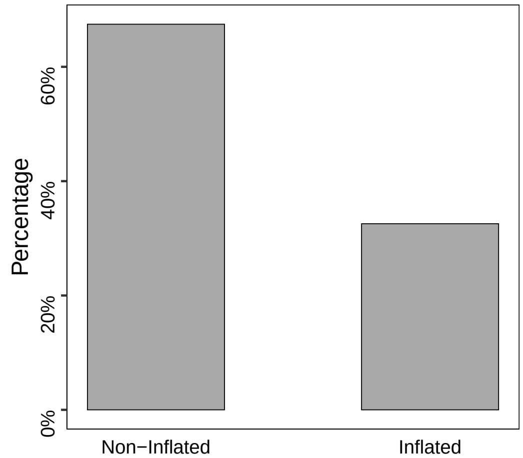
[3]Recall that these are the inflated set of top-category observations that exhbit low or negligible levels of *Depth* in the Baccini et al. data.

[4]These are the non-inflated set of top-category observations that exhibit sufficiently high levels of *Depth* in the Baccini et al. data.

Table A.1: Inflated and Non-inflated Cases from Baccini et al.

Non-Inflated Cases	Inflated Cases
Australia-Chile: The Australia-Chile Agreement aims to "remove most barriers to Australia's exports of goods, and provide economic integration for markets... liberalize and facilitate trade and investment... eliminate tariffs on the imports of most goods from the other party" (Parliament of Australia). And Indeed, the level of the PTA-mandated trade liberalization Depth is 38, which is extremely high. As such, Australia and Chile adopted the flexibility provisions in a sequential manner as needs arose: in 2006, Chile and Australia agreed to incorporate for the 2008 trade agreement. "safeguard duties applied in accordance with Article XIX of GATT 1994 and the Safeguards Agreement" as well as the "antidumping or countervailing duty". Then in 2007, the two countries added "compensation and suspension of concessions" (i.e., tariff concessions) under Article 21.2 in the Trade Agreement (Australian Government DFAT *a*). See: IMF (2015), Australian Government DFAT (*b*).	**South Asian Free Trade Area: (SAFTA):** SAFTA is a trade agreement between Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka which was concurrently negotiated in 2005 and adopted in 2006. SAFTA was a rather shallow trade agreement. The mean Level of Import Duties in SAFTA region in 2006 was approximately 53 per cent, which was among the highest in the world. Moreover, the level of required Depth in PTA is 0 on a 0-48 scale, which implies no trade liberalization requirement. However, SAFTA includes all four measures of escape flexibility—safeguards, anti-dumping duties, suspension of tariff cuts, and countervailing measures—and notes that all countries in SAFTA can have a "longer list of sensitive products exempted from liberalization commitments" and are "allowed smaller initial tariff reduction." See: Baysan, Panagariya and Pitigala (2006), Weerakoon (2010), United Nations LCD Portal.
Japan-Thailand Economic Partnership Agreement (JTEPA): In April 2007, Japan and Chile signed the STEPA free trade agreement, which aimed to "event eliminate the tariffs on over 90 per cent of their bilateral trade" (JUS Laws & Consult), and to "liberalize and facilitate trade in goods and services between the Parties" (EDIT *a*). The level of the PTA-mandated trade liberalization Depth is 31, which is also extremely high. Thus, Japan and Thailand sequentially negotiated and adopted escape clases in their agreement: "Countervailing duties" was discussed and adopted in 2005; "antidumping clauses" was negotiated and incorporated in 2006; and STCs and safeguards were introduced in 2007 as a part of the "Uruguay round agreement... to deal with the damage caused by import surges" (Urata 2007). Also see: Kawai and Wignaraja (2009).	**Asia Pacific Trade Agreement [APTA]** (amended Bangkok Agreement): APTA is a trade agreement between Bangladesh, China, India, Laos, Mongolia, South Korea, Laos, and Sri Lanka. The mean level of Import Duties among the APTA member-states in the amended Bangkok Agreement is a sizeable 29 per cent, and the level of PTA-mandated Depth is as low (5 out of 48). Yet, in 2005, APTA concurrently negotiated and implemented all four escape flexibility clauses—suspension of tariff cuts, safeguards, anti-dumping duties, and countervailing measure—in Chapter 4, Article 17 of the amended Bangkok agreement (University of Solo). See: Feridhanusetyawan (2005), UNESCAP.
Mexico Northern Triangle: In 2000, Mexico, El Salvador, Guatemala, and Honduras entered into an FTA whose Article 1.01 states that "the Parties establish a free trade area in accordance with the provisions of Article XXIV of GATT 1994 and Article V of GATS," while Article 1.02 emphasizes that the agreement's key objective is to "eliminate barriers to trade and facilitate the movement of originating goods and services between the Parties" (EDIT *b*). Indeed, the level of PTA-mandated trade liberalization Depth is 32, which is considerably high. Therefore, the four member-states negotiated and adopted the escape flexibility provisions in a step-by-step manner. The Mexico Northern Triangle FTA countries first "negotiated special safeguard provisions individually" in early 1999 which was initiated by Mexico (Pereira 2001, 396). They then negotiated and incorporated anti-dumping and countervailing duties in late 1999 (Villareal 2017, 59). Finally, STCs were implemented in the FTA in 2000 before the Northern Triangle Trade Agreement came into effect formally.	**Brazil-Mexico:** In 2002, Brazil and Mexico entered into the Brazil-Mexico Agreement of Economic Complementation No. 53 (PTA). The level of the PTA-mandated Depth is 6 out of 48, which is extremely low and implies non-negligible amount of PTA-determined trade reform commitments. However, Brazil and Mexico concurrently negotiated and adopted safeguard measures, including STCs, anti-dumping duties, and countervailing duties in July 2002 within the framework of the Latin American Integration Association (LAIA). See: Estevadeordal et al. (2004), Galindo et al. (2002), Estevadeordal et al. (2009).

Figure A.1: Non-inflated and Inflated Top-Category Cases %:Escape Flexibility



in this case.

A.2.2. TiOP(C) Inflation- and Outcome-Stage Covariates

In the TiOP(C) models' OP outcome stage in which *Escape Flexibility* is the dependent variable, we include all covariates in Baccini et al.'s main (Model 3, Table 3) OP specification. These variables include an additive index that measures each PTA's *Depth* of trade liberalization commitments, the *Depth*×*Regime Dummy* interaction term and its two constitutive components, the *Democratization* dummy for PTA member-countries that have recently democratized and the following control variables: logged measures of *GDP*, *Trade* flows and GDP per capita (*GDPpc*) for each PTA member-state, the *WTO* dummy for PTA member-states that are WTO members, and the number of member-states in each PTA (*No. Members*). Additionally, we include the following theoretically-identified covariates in the TiOP(C) models' probit inflation-stage equation: logged measures of *GDP* and GDP per capita (*GDPpc*) for each PTA member-state, the *Democratization* dummy, and the *Regime Dummy*. The operationalization of each variable listed above is described in the paper's main text. Summary statistics for these variables are provided below in Table A.2.

Next, we draw from extant theoretical claims on PTA-flexibility provisions to identify and

Table A.2: Summary Statistics for Variables in Baccini et al.

	Mean	Median	Minimum	Maximum	Standard Deviation
Escape Flexibility	2.682	3.000	0.000	4.000	1.454
Depth	8.151	4.000	0.000	40.000	10.098
Regime	1.554	5.000	−10.000	10.000	7.138
Regime Dummy	0.494	0.000	0.000	1.000	0.500
GDP	21.310	21.250	14.130	27.390	1.805
GDPpc	8.409	8.539	5.067	10.580	0.991
Trade	2.965	2.430	0.000	11.985	2.709
Democratization	0.233	0.000	0.000	1.000	0.423
WTO	0.498	0.000	0.000	1.000	0.500
No. Members	5.706	2.000	2.000	91.000	8.722

include the following four covariates in the TiOP(C) models' probit inflation-stage equation. First, we include log *GDPpc* in the inflation stage. This is because richer PTA-member countries are less likely to pursue maximum flexibility provisions for insurance as these states have sufficient resources to compensate domestic industries or adjust to trade shocks resulting from PTA-induced reforms via counter-cyclical fiscal policies (Baier and Bergstrand 2004; Teh et al. 2009). By contrast, Prusa (2016) and Bhagwati, Krishna and Panagariya (1999) argue that richer PTA-member states seek maximum PTA-escape clauses for import discrimination against non-PTA countries. Second, we add log *GDP* of PTA-member states as Kucik and Reinhardt (2008) suggest that economically larger member-states may pursue maximum import relief measures like AD in PTAs to "[shift] the terms of trade in their favor" (p.492) at the expense of smaller PTA-participating economies. Teh et al. (2009), however, claim that economically larger PTA-member countries typically have asset-specific import-competing industries that hinder adjustment to trade-shock induced growth contractions stemming from PTA-mandated trade reforms. Hence, these PTA-member states pursue all PTA-flexibility clauses to hedge against this outcome.

Third, *Regime Dummy* (coded as 1 for fully democratic PTA member-states) is included in the inflation-stage to evaluate two competing claims. First, democratically elected incumbents may face punishment from voters if they (i) do not protect citizens from the adverse effects of trade liberalization and (ii) use tariffs at a higher rate compared to NTBs or PTA-flexibility provisions to (Kono 2006, 374; Mansfield and Milner 2012, 2018). Thus, incumbents in PTA-member democracies will pursue maximum escape flexibility measures to signal to voters that they intend to shield them from PTA-induced terms-of-trade shocks, while avoiding tariffs (Maggi and Rodriguez-Clare 2007; Mansfield and Milner 2012). Second, democratic PTA member-state incumbents may seek substantial escape clauses to protect domestic industries from import competition as they are "tempted by the rents that accrue from furnishing protection" (Mansfield and Milner 2018, 374; Maggi and Rodriguez-Clare 2007). Notwithstanding these competing perspectives, greater political transparency of democratic PTA member-states facilitates bargaining among PTA-negotiating countries, thereby allowing them to obtain maximum flexibility provisions in PTAs (Mansfield and Milner 2012; Baccini et al. 2015). Fourth, we add the *Democratization* dummy in the TiOP(C)'s inflation-stage. This is because incumbents in recently democratized PTA-member countries "face high levels of uncertainty about the future state of the world" (Baccini et al. 2015, 771) as their economies are susceptible to import surges engendered by PTA trade reforms. They can also be punished by voters for using tariffs at a higher rate than escape clauses. Hence, recently democratized PTA-participating countries will bargain for maximum escape clauses in PTAs for insurance.

A.3. Evaluating *Repression Intensity*

Girod, Stewart and Walters' (2016, hereafter Girod et al.) evaluate the dynamics of repression, the resource curse, and anti-government protests across several countries. They focus on which "type" of governments frequently employ extreme repression against mass anti-government protests and whether this helps to quell dissent. To test their key hypothesis that "oil-rich autocracies are more likely to follow through on the threat to use force" (p.505) against mass protest campaigns relative to autocracies without significant oil wealth, Girod et al. estimate an OL model on a sample of 662 mass protest campaign-years (71 distinct country-protest campaigns) between 1945 and 2006. This particular OL model includes an *Oil Rents*×*Authoritarianism* interaction term and its separate constitutive components for testing their aforementioned hypothesis, but excludes any additional controls (see Table 2, Model 1 in Girod et al.). *Oil Rents* is measured as the logarithm of per capita oil rents in constant 2005 US dollars. *Authoritarianism* is operationalized as an inverted and re-scaled version of the Polity project's measures of executive constraints and competitive executive recruitment. Accordingly, larger and more positive values denote higher levels of authoritarianism (Girod et al., 508). Girod et al.'s dependent variable in their OL model, *Repression Intensity*, is a four-category ordered variable that ranges from none ("0") to extreme ("3") repression (Figure 1b). This dependent variable is drawn from the Nonviolent and Violent Campaigns and Outcomes (NAVCO 2.0) dataset (Chenoweth and Lewis 2013*b*). Specifically, *Repression Intensity* measures the "degree of state repression in response to campaign activity" when the "government uses the coercive apparatus of the state to quell opposition" (Chenoweth and Lewis 2013*a*, 13).

Girod et al.'s results statistically support their hypothesis mentioned above since they find and report that "autocracies with oil wealth one standard deviation above the mean are 12% more likely to use the highest intensity repression (86% likely) than autocracies whose oil wealth is one standard deviation below the mean (74% more likely)" (p.513). Notwithstanding this insightful result, Figure 1b reveals that a staggering 68% of all observations falls under the ordered *Repression Intensity* measure's top-category—"extreme repression"—therein implying that this dependent variable is top-category inflated. Further, note that *Repression Intensity*'s top-category is coded as "extreme repression" by NAVCO when governments respond to anti-government "campaign activity" with "mass violence," "torture," "intent to violently silence opposition," or "kill(ing)" (Chenoweth and Lewis 2013*a*, 13).[5] This operationalization suggests that there are sound theoretical reasons to suspect that the excessive observations in the "extreme repression" top-category includes two types of obser-

[5] *Without* specific mention that such violence is directed against campaign members.

vations summaried in the paper's main text: non-inflated cases which are effectively incumbents (in this case, autocrats) who *target* heavy-handed repression against anti-government campaign specific activities, and inflated cases which essentially incorporates observations of governments resorting to substantial yet indiscriminate repression against non-campaign specific activities. We turn to evaluate this claim empirically in the next subsection.

A.3.1. "Extreme Repression": Inflated and Non-Inflated Cases

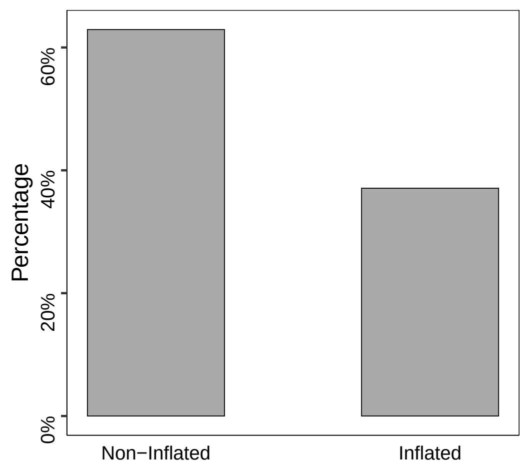
In the main paper, we employed extant claims about the differences between targeted and indiscriminate repression by Kalyvas (2006), Lyall (2009), and Dugan and Chenoweth (2012) to explain on theoretical grounds why extreme repression of non-campaign specific activities constitute the inflated set of top-category observations that are substantively distinct from extreme repression of anti-government campaign-specific activities that are non-inflated top-category observations. Using insights from the scholars mentioned above, we argued that the top-category of "extreme repression" in Girod et al.'s ordered *Repression Intensity* dependent variable includes both high levels of (i) *targeted repression* against anti-government campaign-specific (e.g., regime replacement or overthrowing government) activities and (ii) *indiscriminate repression* against non-campaign specific activities. To this end, we first suggest that substantial repression of anti-government campaign-specific activities transpires when governments target their coercion narrowly against specific opposition groups and their leaders as well as core activists. Substantial levels of targeted repression against anti-government campaign-specific activities are harder to execute as targeting anti-government activities requires fine-grained information and preparation. Importantly, observations of extreme repression against anti-government campaign-specific activities—which, by default, are targeted—are non-inflated top-category cases. This is because these observations are based on Girod et al.'s theoretically-determined ordered continuum as these scholars theorize that extreme repression are perpetrated by autocrats in oil-rich states against specifically anti-government opposition campaigns.

By contrast, repression of non-campaign specific activities, especially by autocrats, is indiscriminately applied against a wider swath of the population that can include sympathizers and even bystanders (Kalyvas 2006; Lyall 2009). It is also often a cheaper tactic that is simply used to demonstrate power (Kalyvas 2006, 146-148). Since heavy-handed repression against non-campaign specific activities is indiscriminate and thus not targeted towards specific groups, repression of this sort is *not* about extreme repression of anti-government campaign-specific activities. Accordingly, "extreme repression" observations in the ordered *Repression Intensity* dependent variable that results from autocrats repressing non-campaign specific activities constitute inflated top-category observations that are not based on Girod et al.'s theoretically-assumed ordered continuum that focuses on heavy-handed repression unleashed by oil-rich autocrats against specific anti-government opposition campaigns.

We carefully analyze the top-category of the ordered *Repression Intensity* measure in the Girod et al. data and other relevant information in their data to empirically parse out the latent inflated set of "extreme repression" observations from the non-inflated ones in the said top-category. This exercise is conducted in three steps. For the first step, we extracted and examined all the cases of campaign activities listed in the "Campaign" column in their data. We learn from this data that the campaign activities range from anti-government opposition campaigns carried out by well defined and easily identified domestic actors (e.g., ETA in Spain, EOKA in Cyprus) to highly diffuse and amorphous non-campaign specific activities such as broad-based random protests against poor governance or corruption. We then classified these campaigns into the following two categorized lists—those in which the incumbent (including autocrats) can easily identify the specific anti-government campaigns and their core activists as opposed to which the incumbent simply cannot or finds it extremely difficult to identify the opposition owing to their diffuse or amorphous nature.

For the second step, we employ the lists mentioned above to analyze the proportion of "extreme repression" (top-category) observations in Girod et al.'s ordered *Repression Intensity* dependent variable that are observed in the categorized context of anti-government opposition campaigns (the non-inflated set) such as regime replacement and those that occurred in the context of the non-campaign specific activities (the inflated set). For the third step, we illustrate in Figure A.2 the proportion of "extreme repression" (top-category) observations—expressed in percentage terms—in the aforementioned non-inflated set and inflated set of top-category observations. The figure reveals that about 37% percent of NAVCO's extreme repression episodes incorporated in the top-category of extreme repression of Girod et al's *Repression Intensity* measure are indeed inflated cases in which incumbents, including autocrats, resorted to heavy-handed and likely indiscriminate repression against a wide spectrum of citizens that were engaged in non-campaign specific activities.

Figure A.2: Non-inflated and Inflated Top-Category %:*Repression Intensity*



As such, these inflated cases are "NAs" in the top-category of *Repression Intensity* as they are not based on Girod et al.'s theoretically-assumed ordered continuum of autocratic repression of anti-government campaign-specific activities. The remaining non-inflated share (63%) of extreme repression top-category observations are, however, based on the theoretically-determined ordered scale which assumes that substantial repression of this sort is targeted against anti-government campaign-specific activities that can pose an existential threat to those in office. Illustrative examples of these (non-)inflated set of observations in the top-category ("extreme repression") of *Repression Intensity* drawn from the Girod et al. data appear in Table A.3. As for our earlier application above, this Table also includes brief narratives that corroborate our classifications of these examples, and our overall contentions of top-category inflation.

Table A.3: Inflated and Non-inflated Cases from Girod et al.

Non-Inflated Cases	Inflated Cases
Spain: Spain is coded to have used the highest level ("extreme repression") of repression against the Euskadi Ta Askatasuna's (ETA) anti-government campaign for several years in the dataset. However, Spain has only targeted ETA members in their efforts to repress the movement.	**Morocco:** Morocco is coded to have used the highest level ("extreme repression") of repression against the pro-independence movement. However, irrespective of the pro-independence campaign, the Moroccan government used indiscriminate violence, including chemical weapons, against civilian populations (Porges and Leuprecht 2016, 79). Scholars have also contended that there are "well-documented evidence" which suggests that the "Moroccan government committed crimes against humanity from the beginning of the occupation in 1975 and onwards, continuously violating the human rights of the Saharawi population living on the occupied territories" (Sántha, Lennartsson Hartmann and Klamberg 2010, 4).
Bolivia: Bolivia is coded to have employed the highest level of repression against the anti-government campaign directed at the Junta regime (1977-1981). However, during this time, only those who were involved in the specific campaign was targeted by the Bolivian government, including "generalized police violence against demonstrators and the arrest of the movement's leaders" (Zunes 2018, 60).	**Egypt:** In the Girod et al. dataset, Egypt is coded to have the highest level ("extreme repression") of repression against the Kifaya (Kefaya) that protested against Hosni Mubarak's rule. However, even before the anti-government campaign, as soon as he took office, "Hosni Mubarak reimposed a 'state of emergency law'" (Takriti 2020) that allowed the government to engage in "mass torture and arbitrary detention" of "tens of thousands of people" (Amnesty International 2020). As there were many scattered protests events throughout the country, the Egyptian government's repressive acts could have been attributed to Kifaya operations when it was not.
Bangladesh: Bangladesh is coded to have responded to the anti-government campaign in 1987-1990 against the Ershad government. Despite the employment of extreme repression, the Bangladesh government used targeted repression, including firing tear gasses at demonstrators as well as arresting, detaining, and executing demonstrators; especially, student leaders were most frequently arrested and harrassed (Immigration and Refugee Board of Canada 1991).	**Guatemala:** In the Girod et al. dataset, Guatemala is coded to have the highest level ("extreme repression") of a left-wing guerrilla anti-government campaign, for most of the years during the Guatemalan Civil War (1960-1996). However, Guatemalan military regime has been named and shamed by the international community for its use of indiscriminate violence against civilians, regardless of their involvement with the anti-government campaign, which has been labeled as 'acts of genocide' by the UN-sponsored Historical Clarification Commission (CEH) (Kubota 2017; Schwartz and Straus 2018, 223, 227). It is likely that the authoritarian government used extreme repression to not just suppress URNG but to silence its opponents.

Table A.3: Inflated and Non-inflated Cases from Girod et al. Continued

Non-Inflated Cases	Inflated Cases
United Kingdom (Northern Ireland): The UK is coded to have used the highest level ("extreme repression") of repression against the Irish Republican Army (IRA) for several years in the dataset. However, the UK has only targeted IRA members in their efforts to repress the movement.	**Serbia (FR Yugoslavia):** In the Girod et al. dataset, Serbia is coded to have the highest level ("extreme repression") of repression against an anti-government campaign during 1997-1999. However, irrespective of the anti-government campaign, the "Serbian police committed numerous serious abuses including extrajudicial killings, disappearances, torture, brutal beatings, and arbitrary arrests and detentions... severely restricted freedom of speech and of the press, and used overbearing police intimidation... infringed on freedom of worship by minority religions..." (US Department of State 1998). Moreover, during the same period, the government was also engaged in a conflict in Kosovo, where President Milosevic and his inner circle of political and military leaders committed war crimes and crimes against humanity, which may have also impacted the government's level of repression.
Cyprus: The United Kingdom is coded to have employed the highest level of repression against Ethniki Organosis Kyprios Agoniston (EOKA) during their anti-(British) government campaign during 1956-1958. While the British government has been accused of torture and other human rights abuses, the repression was generally targeted at EOKA members and suspects (French 2015).	**Indonesia:** In the Girod et al. dataset, Indonesia is coded to have the highest level ("extreme repression") of repression against the Timorese resistance for independence for all years between 1988-1999 except for 1992. Since the annexation, Indonesia has repressed the Timorese population regardless of their involvement in resistance: "indiscriminate nature and were apparently aimed at terrorizing and intimidating entire villages or communities perceived as hostile to the pro-integration cause" (United Nations General Assembly 1999). Moreover, in 1999, extreme repression has been also linked to the terror and intimidation caused by the Indonesian government and anti-Independence militias before and after the referendum for the independence of East Timor in 1999 rather than direct Timorese resistance (Political Economy Research Institute N.d.).

A.3.2. Inflation- and Outcome-Stage Covariates: Girod et al.

Unlike Girod et al.'s OL model, our TiOP(C) models account for the non-inflated and inflated share of observations in *Repression Intensity*'s top "extreme repression" category. The TiOP(C) models do so by jointly estimating two latent equations: the (i) probit inflation-stage equation that estimates the effect of covariates on the probability of governments resorting to extreme repression against anti-government campaign activity (*non-inflated* cases) versus non-campaign specific activity (*inflated* cases) and (ii) ordered probit (OP) outcome-stage equation that evaluates how another set of covariates influences each ordered category of *Repression Intensity*, conditional on the government's engagement in campaign-specific extreme repression. We thus test Girod et al.'s hypothesis stated earlier by estimating our TiOP(C) models on their data. To this end, we first replicate the authors' original analyses with an OP, rather than OL, model. We then proceed to estimate our TiOP(C) models alongside these OP estimates. For the TiOP(C) models' OP outcome-stage, we replicate Girod et al.'s OL specification described above. This entails our use of their ordered *Repression Intensity* measure as the dependent variable and the inclusion of all independent and control variables reported in Girod et al.'s main OL model that appeared in Model 1, Table 2 of their article.[6]

First, we include Girod et al.'s *Authoritarianism* variable in the probit inflation-stage. This follows from our earlier claim that autocrats have political incentives to *preemptively* unleash heavy-handed repressive tactics not to suppress campaign activities *per se* but to "publicize their brutality to deter opposition or energize supporters" (Guriev and Treisman 2019) and signal the regime's "coercive power domestically" (Davenport 2007, 1; Escribà-Folch and Wright 2015). Doing so may help autocrats prolong their survival in office. Hence,

[6]These variables are *Oil Rents*×*Authoritarianism* and each of the two separate constitutive components of this interaction term. Next, following the extant theoretical literature on state repression and the preceding discussion about the dual d.g.p. underlying the top-category of *Repression Intensity*, we identify and include the following covariates in the TiOP(C) models' probit inflation-stage. This inflation-stage equation estimates the probability with which governments perpetrate extreme repression in response to anti-government campaign activity versus non-campaign-specific activity.

we expect that the association between *Authoritarianism* and the inflated share of observations (non-campaign specific activity) in *Repression Intensity's* top-category will be positive. This implies that the influence of *Authoritarianism* on the non-inflated share of observations (campaign-specific activity) in the top-category of *Repression Intensity* will be negative in the TiOP(C) models' probit inflation-stage. Second, reports of violent repression are more likely to be miscoded as "mass violence" and hence coded in the top-category of "3" in *Repression Intensity* for countries that frequently experience civil conflicts, independent of a given protest campaign because non-state actors such as militias often commit violent atrocities against civilian who are not associated with anti-government campaign activity (Carey, Colaresi and Mitchell 2015; Raleigh 2016).

Given the noisy and contradictory information environment that prevails in civil conflicts, the sources used by NAVCO to operationalize *Repression Intensity* may mistakenly report militias' independent repressive actions as government-perpetrated repression in response to domestic resistance campaigns when in reality it is not so. Hence, we measure an observation's exposure to the number of *Civil Conflicts* using the number of internal armed conflict between the government and domestic opposition group(s), absent international interventions, from the UCDP dyadic dataset (Harbom et al. 2008). We anticipate that the estimate of this variable will be negative in the probit inflation-stage equation. Third, primary sources that cover government repression in low-information countries where media coverage is weak tend to inaccurately report state repression with respect to which groups were targeted and why. Since NAVCO uses these sources to code the top-category of "3" in *Repression Intensity*, these "extreme repression" events may have been miscoded as campaign-specific when they were not. Therefore, in the probit inflation-stage equation, we include the ordered *Domestic Media Salience* measure drawn from the NAVCO dataset, which measures the extent of domestic media coverage of the campaign of interest. We predict this covariate's estimate to be negative in the TIOP(C) models' inflation-stage. We also control for *Oil Rents* as states with greater oil wealth typically have greater capacity to violently repress anti-government campaign-specific activities. This variable is expected to be negatively associated with non-inflated campaign-specific extreme repressions. Summary statistics for these probit inflation-stage variables and the ordered outcome stage covariates drawn from Girod et al. appear in Table A.4.

Table A.4: Summary Statistics for Variables in Girod et al.

	Mean	Median	Minimum	Maximum	Standard Deviation
Repression Intensity	3.000	2.366	0.000	3.000	1.093
Authoritarianism	−0.700	2.000	−7.000	6.000	4.819
Oil Rents	2.909	2.793	0.000	8.539	2.618
Civil Conflict	0.628	0.000	0.000	5.000	0.875
Domestic Media Salience	1.497	2.000	0.000	2.000	0.759

A.3.3. TiOP(C) Results for ***Repression Intensity***

For the Girod et al. application, we first replicated the authors' original analyses with an OP, rather than OL, model. We then estimate a pair of matching TiOP and TiOPC models (denoted TiOP 1 and TiOPC 1) that include the (i) theoretically identified covariates mentioned above in the models' probit inflation-stage and (ii) all of Girod et al.'s OL variables listed above in the ordered outcome stage. For robustness checks, we add all of Girod et al.'s ordered outcome stage variables to the TiOP(C) model's inflation-stage (TiOP 2 and TiOPC 2).

We first discuss the TiOP(C) models' probit inflation-stage results that assess the effect of each inflation-stage variable on the probability of an observation being a *non-inflated* case. We also illustrate the marginal effect of these inflation-stage covariates in Figure A.3 to facilitate interpretation of the said results.[7] Note that the coefficient estimate *and* marginal effect of *Authoritarianism* on the predicted probability of observations being in the non-inflated group in the top-category ("extreme repression") of *Repression Intensity* is

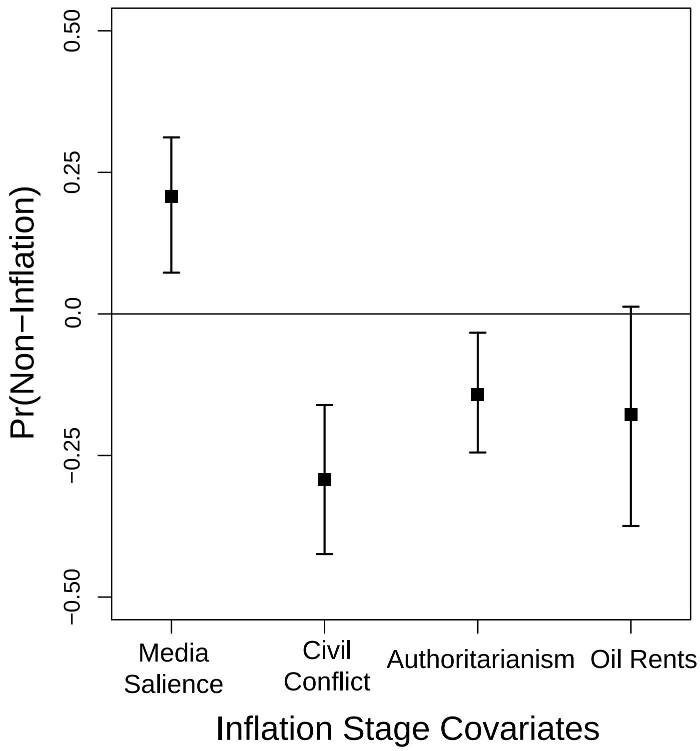
[7]Obtained from the coefficients of the theoretically-identified TiOP inflation-stage covariates and parametric bootstraps ($m = 1,000$ and all other remaining covariates are held at their mean or mode).

consistently negative and significant in the TiOP(C) models' probit inflation-stage (Table A.5; Figure A.3). This result corroborates our claim that events of "extreme repression" committed in autocracies are more likely to be miscoded as campaign-specific when, in fact, they are non-campaign-specific. We argue that this is an artifact of heightened preemptive government repression and draconian censorship that together mask the true nature of acute government repression in authoritarian countries, leading to top-category inflation in the *Repression Intensity* measure.

Table A.5: Replication Results For Girod et al. (Table 2, Model 1)

	Dependent Variable: Repression Intensity				
	OP	TIOP 1	TIOP 2	TIOPC 1	TIOPC 2
	Ordered Probit Stage				
Authoritarianism	-0.078^{***}	-0.118^{***}	-0.147^{***}	-0.128^{***}	-0.163^{***}
	(0.023)	(0.040)	(0.061)	(0.040)	(0.049)
Oil Rents	-0.051^{*}	-0.123^{***}	-0.229^{**}	-0.144^{***}	-0.192^{***}
	(0.028)	(0.047)	(0.061)	(0.047)	(0.056)
Oil Rents×Authoritarianism	0.026^{***}	0.039^{***}	0.037^{***}	0.040^{***}	0.043^{***}
	(0.006)	(0.009)	(0.012)	(0.009)	(0.011)
Cut1	-1.488^{***}	-1.079^{***}	-1.393^{***}	-1.380^{***}	-1.483^{***}
	(0.127)	(0.207)	(0.279)	(0.252)	(0.215)
Cut2	-1.321	-0.801^{***}	-1.110^{***}	-1.101^{***}	-1.194^{***}
	(0.122)	(0.210)	(0.293)	(0.258)	(0.219)
Cut3	-0.942^{***}	-0.052	-0.326	-0.317	-0.339
	(0.115)	(0.272)	(0.429)	(0.296)	(0.303)
	Inflation Stage				
Domestic Media Salience	.	0.834^{***}	0.760^{***}	0.971^{***}	0.830^{***}
		(0.190)	(0.237)	(0.218)	(0.229)
Civil Conflict	.	-0.872^{***}	-1.007^{***}	-0.898^{***}	-1.014^{***}
		(0.153)	(0.190)	(0.155)	(0.179)
Authoritarianism	.	.	-0.084^{**}	.	-0.083^{***}
			(0.035)		(0.031)
Oil Rents	.	.	-0.177^{*}	.	-0.133^{*}
			(0.099)		(0.070)
Intercept	.	-0.988^{***}	-0.133	-1.293^{***}	-0.533
		(0.309)	(0.549)	(0.357)	(0.494)
ρ	.	.		-0.445	-0.408
				(0.275)	(0.289)
Observations	406	406	406	406	406
Log-likelihood	-284.45	-232.65	-228.27	-231.67	-227.49
AIC	580.90	483.30	478.54	483.34	478.98
BIC	604.93	519.36	522.61	523.40	527.06
Note:					*p<0.1;**p<0.05; ***p<0.01

Figure A.3: TiOP Inflation Stage Marginal Effects



The estimate and marginal effect of *Domestic Media Salience* is positive and statistically significant across the TiOP(C) inflation-stage specifications (Table A.5, Figure A.3). This implies that state-perpetrated "extreme repression" events are indeed significantly more likely to be correctly attributed to campaign-specific protests when there is more media coverage of the campaign of interest. Conversely, in low information environments, there is a higher probability that extreme government repression is mis-associated with a protest campaign of interest when in fact there is *no* association, leading to top-category inflation. Next, the estimate and marginal effect of *Civil Conflicts* in the TiOP(C) inflation-stage is also negative and significant (Table A.5, Figure A.3), suggesting that the higher the number of civil conflicts in a country, the more likely that "extreme repression" will be misattributed to a campaign of interest, thereby leading to top-category inflation in *Repression Intensity*. This supports our theoretical claim that the noisy information environment during civil conflicts increases the likelihood of an arbitrary militia-perpetrated violence independent of government intentions being mis-ascribed as campaign-specific "extreme repression" by the government. Higher *Oil Rents* are also associated with more top-category inflation even though this statistical relationship is only significant at the $p < .10$ level. Furthermore, the inflation-stage results reported above remain robust when we include all the other Girod et al.'s ordered-outcome stage covariates (i.e., *Oil Rents*$\times$ *Authoritarianism*) to the probit inflation-stage of the TiOP and TiOPC models (see Table A.5).

We discussed and illustrated the marginal effect results (see Figure A.4) obtained from the TiOP(C) models estimated on Girod et al.'s data above. To determine which models to focus on for deriving and analyzing the substantive effects plots from the TiOP(C) models' ordered outcome stage, we review our model selection criteria for the OP, TiOP, and TiOPC models reported in Table A.5. We find that our AICs, BICs, and likelihood ratio tests all favor the TiOP(C) models over the OP model,[8] providing strong support for addressing top-category inflation via the TiOP(C) models. The AIC and BIC statistics for the TiOP and TiOPC models also consistently favor the TiOP models over the TiOPC models. This finding in favor of the TiOP (TiOP 1 and TiOP 2) models over the TiOPC (TiOPC 1 and TiOPC) models is further reinforced by (i) the ρ estimate reported in Table A.5 (statistically insignificant) and (ii) a likelihood ratio test, which prefers the TiOP model.[9] These results suggest that the TiOP models are a superior choice to the TiOPC models when modeling *Repression Intensity*, leading us to focus on the TiOP and OP model comparisons below.

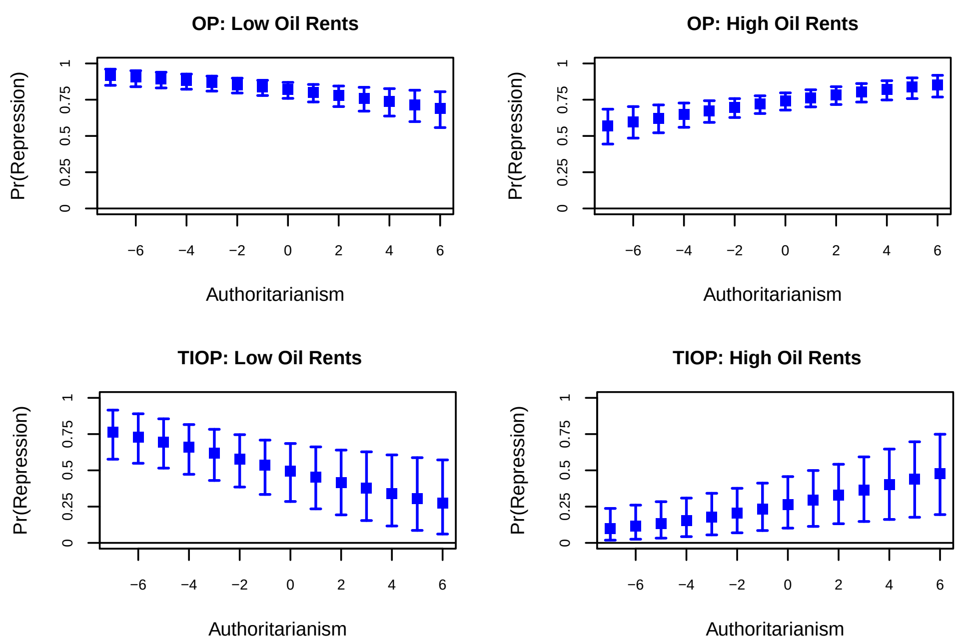
With these inflation-stage findings in mind, we now turn to the ordered outcome-stage results of the OP, TiOP, and TiOPC models in Table A.5. We focus on first reporting the ordered-outcome stage coefficient estimates from these models and then further assess these findings by plotting the relevant results below. To begin with, the coefficient estimate of *Authoritarianism* is negative and statistically significant, whereas the estimate of *Oil Rents*$\times$*Authoritarianism* is consistently positive and statistically significant. The estimate for *Oil Rents* is likewise statistically significant (and negative) across all three models,

[8]In each case, at least at the $p < 0.01$ level.

[9]At least the $p < 0.01$ level

but achieves a noticeably higher threshold of significance in the TiOP(C) models' ordered-outcome stage compared to the OP model. Additionally, the TiOP and TiOPC models' ordered-outcome stage results reported here remain robust when the probit inflation-stage of these models include the theoretically-identifies covariates listed earlier as well as the *Oil Rents*$\times$*Authoritarianism* interaction term (see Table A.5). While these results support Girod et al.'s findings, the TiOP(C) models' ordered-stage estimates are often 1.5-2 times greater than those obtained under an OP model. As elaborated in further detail below, this indicates that one's primary findings can at times become even *stronger* once top-category inflation is accounted for.

Figure A.4: Estimated Ordered Probit Stage Marginal Effects



Before presenting our substantive effect results, it is worth noting here that the AIC, BIC, and likelihood ratio tests all favor the TiOP(C) models over the OP model,[10] providing strong support for addressing top-category inflation via the TiOP(C) models. The AIC, BIC, and ρ test results consistently favor the TiOP models over comparably specified TiOPC models.[11]

[10]In each case, at least at the $p < 0.01$ level.
[11]At least at the $p < 0.01$ level.

These results suggest that the TiOP models are a superior choice to the TiOPC models when modeling *Repression Intensity*, leading us to focus on the TiOP and OP model comparisons. In fact, turning to our substantive effect comparisons, we specifically evaluate the OP and TiOP models' ordered-stage interaction effect of *Oil Rents*×*Authoritarianism* by calculating the predicted probability of "Extreme Repression" (*Repression Intensity's* category "3") at each level of *Authoritarianism* for both low and high values of *Oil Rents*. Mirroring Girod et al.'s evaluation of their Model 1 (Table 2, p. 513) substantive effects, "low oil rent" denotes one standard deviation (SD) below and "high oil rent" denotes one SD above *Oil Rent* variable's sample mean. We estimate these ordered outcome-stage effects and their 95% confidence intervals using parametric bootstraps with $m = 1,000$ (see Figure A.4).

While the direction of the estimated interaction effect across the OP and TiOP models in Figure A.4 is similar, we can observe within the TiOP model that—once we account for top-category inflation—the baseline probabilities of extreme repression for all values of *Authoritarianism* and *Oil Rents* are lower in the TiOP model compared to the OP model. More crucially, Figure 5b reveals that the TiOP model's interaction effects are notably larger in magnitude than those of the OP model. A shift from low-to-high *Oil Rents* in highly authoritarian countries (*Authoritarianism*= 6.) yields a 20% increase in the likelihood of "extreme repression" in the TiOP model, but just a 16% increase in the OP model (Figure 5b). An increase from -7 to 6 on *Authoritarianism* for a high (low) *Oil Dependence* state yields a 38% increase (49% decrease) in the probability of extreme repression in the TiOP model, but only a 28% increase (22% decrease) in the aforementioned probability from the OP model. Given the risks to human lives posed by extreme repression, these differences are substantively important. Thus, these findings further indicate that it is important to address top-category inflation especially since it may occasionally *strengthen* key effects as opposed to *diminishing* them.

A.4. Exploring *Disappearance* and *ICC Ratification*

A.4.1. Piazza and Walsh (2009): *Disappearance*

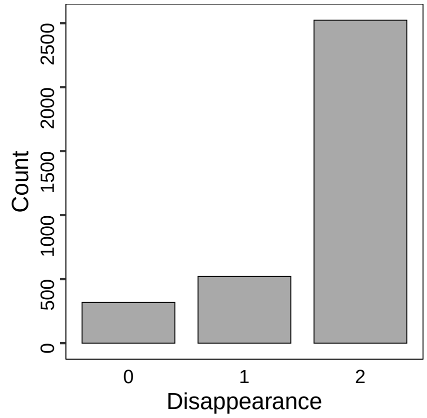
In their *ISQ* article, Piazza and Walsh (2009) suggest that governments are more likely to respond to transnational terror attacks by engaging, for example, in "disappearances" that directly violate the physical integrity rights of citizens. But in the absence of transnational terror attacks or when such attacks are negligible, governments are more prone to respect citizens' physical integrity rights and are thus less likely to engage in disappearances. They assess this claim by employing the ordered dependent variable, *Disappearance*. This dependent variable is drawn from the Cingranelli and Richards (CIRI) Human Rights data (Cingranelli and Richards 2004), which ranges from 0 to 2 on an ordinal scale. In this ordinal scale, 0 indicates the lowest level of (i.e., "least") respect for human rights by governments, and 2 denotes "maximum respect" or, in other words, "zero" abuses (Piazza and Walsh 2009, p.133).[12] They find robust support for their claims summarized above.

However, as indicated in the frequency distribution illustrated in Figure A.5, the top-category of "maximum respect" (coded as "2") has three times more observations than it has observations within the other two ordered categories (i.e., 0 and 1) combined. This means that the "maximum respect" top-category in *Disappearance* contains an excessive share of observations that likely includes both non-inflated and inflated cases that are generated from distinct theoretical processes. More specifically, it is plausible that the non-inflated cases in the top-category of *Disappearance* (i.e., "maximum" respect or "zero" abuses) are countries that are genuinely characterized by high respect for their citizens physical integrity rights owing to a variety of observable (e.g., rule of law) and latent (e.g., social norms that foster human rights protection) factors.

Note that these non-inflated cases are based on the ordered continuum conceptualized by CIRI—employed by Piazza and Walsh (2009) to test their theoretical claims—which posits

[12] CIRI primarily relies on the US State Department's annual country reports on Human Rights practices.

Figure A.5: Excessive Top-Category Observations from Piazza and Walsh (2009)



that countries in the top-category of maximum respect are indeed those that fully value and respect the physical integrity rights of their citizens. By contrast, the inflated cases in the top-category of *Disappearance* are countries that engage in repression (including disappearances) against citizens. Yet, CIRI may inadvertently misattribute or miscode these repressive states as countries that exhibit maximum respect for human rights in the *Disappearance* measure owing to a variety of factors such as poor information, censorship or even biased reporting by certain media outlets or other agencies (Simmons 2009; Poe, Carey and Vazquez 2001). Since these inflated observations are misattributed as observations of "maximum respect" for human rights in *Disappearance* when this is not the case in reality, they are thus not based on CIRI's conceptualized ordered continuum which assumes (as Piazza and Walsh (2009) do) that states in the maximum respect top-category are, in fact, those that genuinely value their citizens' physical integrity rights.

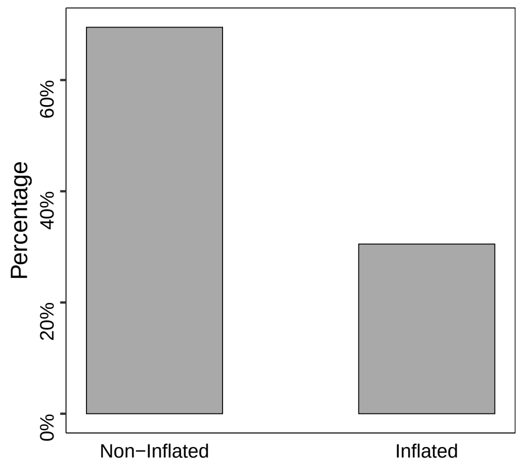
A careful empirical assessment of all the top-category "maximum respect" observations in Piazza and Walsh's *Disappearance* dependent variable reveals that there indeed, as described above, exists non-inflated and inflated cases in the said top-category. We assessed this in three steps. For the first step, we merged Piazza and Walsh's data with the Political Terror Scale (PTS) dataset (Wood and Gibney 2010; Gibney et al. 2022), which includes two variables—*PTS_A* and *PTS_S*—that each code the level of state-perpetrated human rights violations (also denoted as 'political terror') carried out in each country based on the Amnesty International Annual Reports and the US Department of State's annual Country Reports on Human Rights Practices, respectively. Specifically, both PTS variable ranges from 1 to 5 where 1 denotes the lowest level of human rights abuses while 5 denotes the maximum level of abuses.[13] Countries whose PTS score is 3 or above are those that often resort to substantial levels of state-perpetrated repression (Gibney et al., 2022). Countries that have a PTS score that lie between 1 and 3 are states that exhibit high respect for the physical integrity rights of their citizens. For our analysis, we take the average score of *PTS_A* and *PTS_S* and denote it as *PTS*.

For the second step, we employ the (averaged) PTS scale to classify the top-category maximum respect observations in *Disappearance* into the following two sets of states: those with a PTS score of (i) 3 or above as these countries routinely violate their citizens' physical integrity rights and (ii) less than 3 as this latter set of countries have strong observable recording of protecting their citizens' rights. Note that states observed in the top-category of Disappearance whose PTS score is 3 or above are the inflated cases. This is because these states have been likely misattributed by CIRI as units that have maximum respect for physical integrity rights of citizens even though the actual repressive behavior of these states suggest exactly the opposite. We noted above that these inflated cases are not based on CIRI's assumed ordered continuum which conceptualizes and operationalizes the top-category of disappearance for states that are assumed to be genuine guarantors of their citizens' physical integrity rights.

By contrast, states observed in the top-category of *Disappearance* whose PTS score is

[13] The PTS measures levels of political violence and terror that a country experiences in a particular year based on a 5-level "terror scale." The data used in compiling this index comes from three different sources: the yearly country reports of Amnesty International, the U.S. State Department Country Reports on Human Rights Practices (only for years after 2013), and Human Rights Watch's World Reports. More specifically, 1 indicates that there is a "secure rule of law, people are not imprisoned for their views, and torture is rare or exceptional. Political murders are extremely rare" whereas a 5 indicates that "terror has expanded to the whole population. The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals" (`https://www.politicalterrorscale.org/Data/Documentation.html`).

Figure A.6: Non-inflated and Inflated Cases %:*Disappearance*



less than 3 are the non-inflated cases as the maximum respect for physical integrity rights recorded for these states is genuinely based on their *de facto* behavior of protecting and preserving such rights of their citizens. These non-inflated cases that are genuine guarantors of their citizens' physical integrity rights are thus based on CIRI's conceptually assumed ordered continuum that produces the "maximum respect" top-category in the *Disappearance* measure that is used by Piazza and Walsh (2009).

For the third step, we compute and illustrate the proportion of inflated and non-inflated cases in the top-category of "maximum respect" in the ordered disappearance dependent variable employed in Piazza and Walsh's (2009) study. Figure A.6 reveals that 30% of the top-category observations in disappearance are indeed inflated cases whose PTS score is 3 or above. But the remaining 70% of the said measure's top-category observations are non-inflated cases whose PTS score is below 3. A few examples of these inflated and non-inflated set of observations in the top-category of the *Disappearance* variable which is drawn from the Piazza and Walsh's data and a brief narrative that corroborates the classification of these examples are provided below in Table A.6.

Table A.6: Inflated and Non-inflated Cases from Piazza and Walsh (2009)

Non-Inflated Cases	Inflated Cases
Italy: Italy is coded as the highest category of "2" in the *Disappearance* dependent variable (i.e., "zero" disappearances). The top-category of the *Disappearance* variable for Italy is less likely to have been miscoded due to the abundant availability of information, and there is no particular incentive for the US to overstate their human rights practices.	**Haiti:** Haiti, an US ally during the Cold War, is coded as "zero" incidents of *Disappearance* (i.e., highest category of 2) for most years between 1981 and 2001 according to CIRI. However, multiple sources, such as the AI and Human Rights Watch report that there were "periodic" and "systematic or widespread" instances of enforced disappearances in Haiti, especially during the 1970s and the 1980s (Human Rights Watch 2011; Amnesty International 2011).
Trinidad and Tobago: Trinidad and Tobago is coded as the highest category of "2" in the *Disappearance* dependent variable (i.e., "zero" disappearances). The top-category of the *Disappearance* variable for Trinidad and Tobago is less likely to have been miscoded as there is no particular incentive for the US to overstate their human rights practices.	**Egypt:** An US ally, Egypt, is coded as the top-category of 2 ("zero" disappearances) for all years between 1981 and 2003 except for 1999 in CIRI. Yet, according to the United Nations Commission on Human Rights (UNCHR), there were 19 cases of enforced disappearance between 1988 and 1994.
Belgium: Belgium is coded as the highest category of "2" in the *Disappearance* dependent variable (i.e., "zero" disappearances). The top-category of the *Disappearance* variable for Belgium is less likely to have been miscoded due to the abundant availability of information, and there is no particular incentive for the US to overstate their human rights practices.	**Israel:** Israel, an US ally, is coded as 2 for *Disappearance* in 1992, UNCHR also reports two cases of enforced disappearance for Israel in 1992.
Switzerland: Switzerland is coded as the highest category of "2" in the *Disappearance* dependent variable (i.e., "zero" disappearances). The top-category of the *Disappearance* variable for Switzerland is less likely to have been miscoded due to the abundant availability of information, and there is no particular incentive for the US to overstate their human rights practices.	**Nepal:** The US has been promoting democracy in Nepal, and Nepal is coded as 2 (top-category) for *Disappearance* in CIRI for most years between 1983-1996, several "disappearances" were reported in 1985 and 1996, according to AI (Amnesty International 2003).
Japan: Japan is coded as the highest category of "2" in the *Disappearance* dependent variable (i.e., "zero" disappearances). The top-category of the *Disappearance* variable for Japan is less likely to have been miscoded due to the abundant availability of information, and there is no particular incentive for the US to overstate their human rights practices.	**Nicaragua:** Nicaragua is another country that the US has been involved with democratization and is coded as 2 (highest category) for all years in the post-Cold War period (1992-2003) but there were two accounts of disappearances according to the UNCHR.

A.4.2. Brysk and Mehta (2014): *ICC Ratification*

An important field of research in which IR scholars have consistently employed ordered outcome measures is the study of international organizations (IOs). While previous and current research on IOs is too vast to discuss here, it is worth mentioning that scholars have analyzed not just binary but also ordinal measures to assess the extent to which member-states ratify conventions designed by IOs, the degree to which states comply with rules and regulations set by IOs in which they participate, the influence of IOs on a variety of economic policy outcomes, "quality" of election results, and the prospects for democracy (e.g., Simmons and Martin 2002; Hyde and Marinov 2014; Mukherjee and Singer 2010; Brysk and Mehta 2014). It is possible that the top-category of particularly ordered dependent variables used to assess the role or effect of IOs may be "inflated." To understand this in some depth, consider the study by Brysk and Mehta (2014) in the *Journal of Peace Research* that explores why particularly (but not only) democracies with greater levels of domestic gender equity tend to initiate and ratify several international human rights treaties.

Among the several human rights treaties that they examine, Brysk and Mehta (2014) assess in their paper's appendix the association between the level of gender equity within states and ratification of the "Rome Statue" that both established and empowered the International Criminal Court (ICC) to formally charge individuals and put them on trial in international courts. More specifically, once states formally ratify the Rome Statue in their national legislature, the ICC has the *de jure* capacity to directly intervene in states that have ratified the Rome Statute to both investigate as well as incarcerate those individuals in these states who are charged with the gravest crimes against the international community. These grave crimes include genocide, war crimes, crimes against humanity, and the crime of aggression. Brysk and Mehta (2014) theorize in their paper that countries, particularly democracies, that exhibit higher levels of gender equity are more likely to formally ratify the Rome Statue as greater sexual equality may lead to feminist socialization of the wider society to promote human rights values.

To evaluate the aforementioned claim, Brysk and Mehta (2014) construct the ordered *ICC* dependent variable which operationalizes whether participating governments have (i) *neither* signed nor ratified the Rome Statute, i.e. *no action* ("0"), (ii) have *signed* but not ratified the said Statute ("1"), or (iii) have signed and *ratified* the Statute ("2"). As shown in Figure A.7 the top-category of the *ICC* dependent variable labeled as "ratified" contains an excessive share of observations that likely include both non-inflated and inflated cases that are produced from distinct theoretical processes.

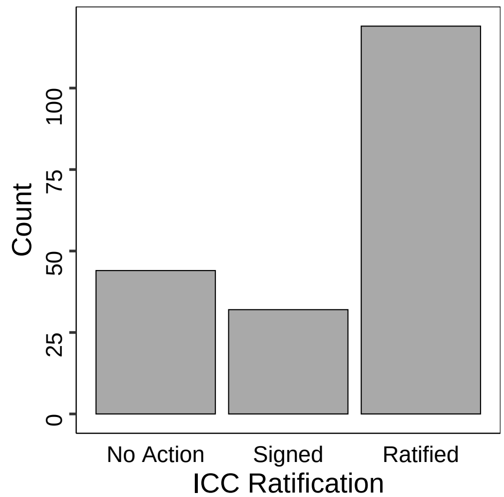
Note that the non-inflated top-category cases are likely states that overtly exhibit a good record or have a reputation for protecting human rights. It is thus plausible that states with a good human rights record ratified the Rome Statute (ICC) primarily because of their normative concerns regarding "grave" crimes (listed above) against the international community, higher levels of pro-human rights socialization, and genuine respect for international law. Accordingly, these non-inflated cases are based on Brysk and Mehta's (2014) theoretically-assumed ordered continuum which posits that states that ratify the Rome Statute are indeed those that have normative concerns against grave international crimes or those that exhibit higher levels of pro-human rights socialization.[14]

In sharp contrast, we posit that the inflated top-category cases are those countries that not only have a poor human rights record but have also engaged in repression targeted against citizens. It is possible that the inflated cases are countries—which are serial abusers of human rights—that ratify the Rome Statute to avoid "naming and shaming" (Hafner-Burton and Tsutsui 2005), prevent economic sanctions targeted against the regime, or to simply curry favor among advanced democracies to secure foreign aid or other material concessions. Hence, unlike the non-inflated top-category cases whose decision to ratify the Rome Statute are driven by normative concerns, the inflated cases are those that ratified the Rome Statute for "window-dressing". Consequently, these inflated cases in the top-category of *ICC* are *not* based on the theoretically-assumed ordered continuum as their ICC

[14]Brysk and Mehta (2014) emphasize that norms against grave international crimes and socialization in support of human rights are driven by higher levels of gender equality within countries.

ratification decision is *not* driven by an inherent revulsion against grave international crimes or genuine proclivity for protecting human rights.

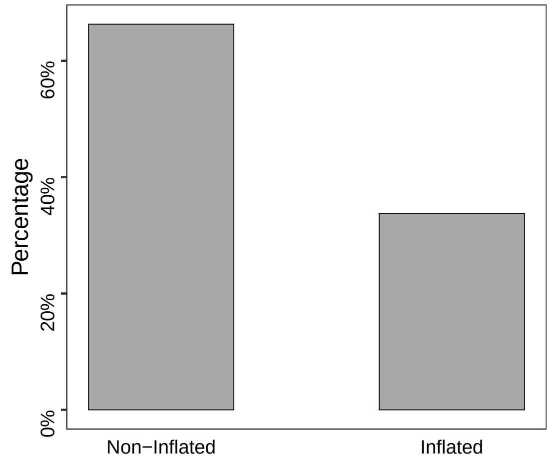
Figure A.7: Excessive Top-Category Observations from Brysk and Mehta (2014)



A concise yet careful empirical examination of *ICC's* top-category (i.e., "ratified") observations in Brysk and Mehta's data reveals that there are indeed non-inflated and inflated observations along the lines suggested above. To see this in more depth, recall the Political Terror Scale (PTS) that ranges from 1-to-5. Countries whose PTS score is high in that it is 3 or above on this scale—including those that exhibit high PTS scores 5 years after the ICC came into effect—are habitual human rights abusers who carry out extensive and systematic state-perpetrated human rights violations. By contrast, countries whose PTS score is below 3, again including those that exhibit low PTS scores 5 years after the ICC came into effect, are states that typically have an impeccable record in terms of respecting and protecting human rights.

We thus employ the aforementioned averaged PTS scale (i.e, mean of PTS_A and PTS_S) to classify the observations in the top-category of *ICC* in Brysk and Mehta's measure into the following two types of states in their data: serial human-rights abusers and those that demonstrate high *de facto* protection of (hence, respect for) human rights. Doing so reveals that about 34% of countries in the top-category of *ICC* "ratified" states in Brysk and Mehta's

Figure A.8: Non-inflated and Inflated Cases %: *ICC Ratification*



data are serial human rights abusers. Yet, despite being human rights abusers, these countries ratified the Rome Statute. This suggests, as illustrated in Figure A.8, that the said countries are likely inflated cases in the top-category of *ICC* who have ratified the Rome Statute to avoid international opprobrium or to perhaps obtain foreign aid. Accordingly, these inflated cases are not based on the theoretically-assumed ordered continuum of states that ratify human rights treaties out of genuine concern and respect for human rights.

Table A.3 provides numerous examples of these inflated cases in the top-category of ICC ratified in Brysk and Mehta's data. Unlike these inflated cases, the remaining 66% of the states observed in the *ICC*'s top-category of "ratified" are *de facto* human rights protectors. This latter set of top-category observations are thus non-inflated cases (see Figure A.8) that have ratified the Rome Statute owing to normative concerns against war crimes, higher levels of human rights socialization, and genuine respect for the physical integrity rights of citizens. A few examples of these top-category non-inflated cases in Brysk and Mehta's *ICC* ordered dependent variable are listed and described in Table A.7.

Table A.7: Inflated and Non-inflated Cases from Brysk and Mehta (2014)

Non-Inflated Cases	Inflated Cases
Netherlands: The Netherlands has had good record of human rights even before the ratification of the Rome Statute and to this date continues to have the best record of human rights according to the PTS index (i.e., "1").	**Afghanistan:** While Afghanistan has ratified the Rome Statue, it may have done so for their international reputation, as the ratification followed just two years after the war in Afghanistan and the 9/11 attacks. Indeed, the Human Rights Watch (HRW) has continuously called out Afghanistan for its violations of human rights (Human Rights Watch 2020). Moreover HRW "strongly doubts the Afghan government's capacity and willingness to bring alleged perpetrators to justice" (HRW 2020, no page). Moreover, Afghanistan has always scored 4-5 on the PTS index since it ratified the Rome Statute.
New Zealand: New Zealand has had good record of human rights even before the ratification of the Rome Statute and to this date continues to have the best record of human rights according to the PTS index (i.e., "1").	**Burundi:** Burundi ratified the Rome Statue against the backdrop of a ceasefire and the establishment of a transitional government after almost a decade of civil war in Burundi. Therefore, they may have been hoping to gain international legitimacy through ratifying the Rome Statue. However, Burundi continued to commit human rights violations and scored 3-5 on the PTS index since it ratified the Rome Statute. Moreover, in 2017, Burundi withdrew from the Rome statute claiming that the ICC targets African countries.
Finland: Finland has had good record of human rights even before the ratification of the Rome Statute and to this date continues to have the best record of human rights according to the PTS index (i.e., "1").	**Chad:** Although Chad has ratified the Rome Statute, Chad has generally scored very high in the PTS index following its ratification. In fact, Chad has also been non-compliant with the "cooperation requests issued by the Court regarding the arrest and surrender of Omar Hassan Ahmad Al-Bashir" (International Criminal Court 2013).
Luxembourg: Luxembourg has had good record of human rights even before the ratification of the Rome Statute and to this date continues to have the best record of human rights according to the PTS index (i.e., "1").	**The Democratic Republic of Congo:** The Rome Statute was ratified in the midst of the Second Congo War, and it is possible that the DRC ratified the Rome Statute to gain legitimacy in the international arena and punish the opponents in the aftermath of the war rather than out of genuine respect for international human rights norms. Indeed, the DRC has constantly scored the highest (i.e., worst) in the PTS index since it ratified the Rome Statute. The country "continue[s] to experience serious human rights violations, including mass killings in the context of armed conflict and inter-communal violence, a crackdown on dissent and ill-treatment of detainees" (Amnesty International 2022).

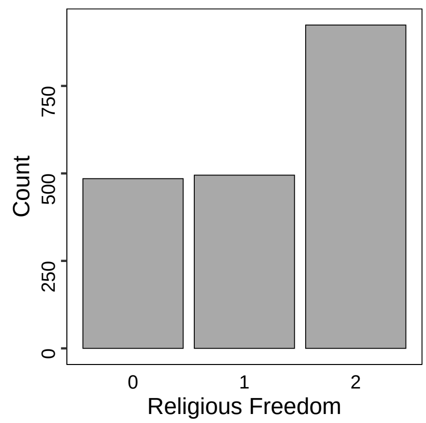
A.5. Other IR Top-Category Inflated Ordered DVs

We emphasized in the introduction of our paper that IR scholars have in recent years employed ordered dependent variables—that exhibit an excessive share of observations in their respective top-category—to evaluate a variety of outcomes that span distinct areas of research. These research areas in IR in which top-category inflated ordered dependent variables have been used for hypothesis testing include the following: government restrictions on religious freedom, interstate war outcomes, WTO trade concessions in the wake of trade dispute settlements analyzed by IPE scholars, survey responses that track political attitudes or perceptions of trust by respondents, and monitoring of election quality by international government and non-government organizations. Below, we briefly discuss examples of top-category inflated ordered dependent variables within each of these five research areas.

A.5.1. Lupu (2015): ***Religious Freedom***

In addition to the Piazza and Walsh (2009) article introduced in the previous section, consider the ordered dependent variable—government restrictions on *Religious Freedom* for its citizens—analyzed by Lupu (2015) in his study published in the *American Journal of Political Science*. This *Religious Freedom* measure is also drawn from the CIRI Human Rights dataset (Cingranelli and Richards 2004) and it is operationalized on a 0 to 2 ordered scale where 0 denotes "high restrictions" and 2 indicates "no restrictions." The distribution of this ordered dependent variable illustrated below in Figure A.9 clearly reveals that that its top-category of "no restrictions" incorporates an excessive share of observations relative to the other outcome categories. Similar to the top-category of the Piazza and Walsh (2009) measures, this top-category of *Religious Freedom* is potentially generated from distinct d.g.p.'s that leads to same top-category outcome: non-inflated cases of states that genuinely respect their citizens' religious freedom because of greater political accountability (this is based on the empirically-assumed ordered continuum) versus those (the inflated cases) that engage in "window-dressing" by demonstrating respect for their citizens religious freedom to avoid international sanctions but (in reality) persecute religious minority groups or informally clamp down on individual religious practices and beliefs (these cases are not based on the ordered continuum).

Figure A.9: Excessive Top-Category Observations from Lupu (2015)

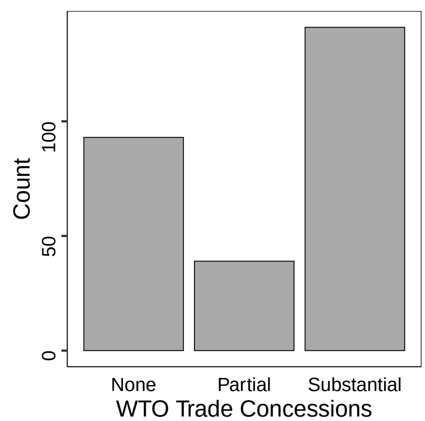


A.5.2. Bobik and Smith (2013): *WTO Trade Concessions*

The first empirical application in the text of our paper focused on the ordered *Escape Flexibility* provisions measure drawn from Baccini et al.'s (2015) Preferential Trade Agreements (PTAs), which are international trade agreements negotiated between states. A careful examination of this ordinal dependent variable revealed that the top-category of *Escape Flexibility* includes an inflated share of observations that results from a dual d.g.p. discussed in depth in our paper. Another key example of international trade agreements is Bobik and Smith's (2013) study in the *Review of International Organiations* that focuses on trade dispute settlement between states that are members of the World Trade Organization (WTO). These two authors evaluate the resolution of bilateral trade disputes within each pair of WTO member-states by employing a directed-dyad-year dataset of WTO member-states' trade dispute settlements. Each directed-dyad in their sample consists of a defendant-plaintiff "pair" that are WTO member-states engaged in a bilateral trade dispute and thus settlement of the said dispute. The dependent variable in their study is the degree of trade dispute settlement for each WTO member-state. It is operationalized as the level of trade concessions granted by the defendant WTO member-state to the plaintiff member-state or vice-versa depending on which state lost the trade dispute in the WTO's appellate body.

More specifically, Bobik and Smith (2013) operationalize such trade concessions—we label this as *WTO Trade Concessions*—on a 0 to 2 ordered scale where 0 denotes "No Concessions," 1 is "Partial Concessions," and 2 is "Substantial Concessions." The frequency distribution of this ordered dependent variable illustrated in Figure A.10 unambiguously reveals that the top-category of this measure has a preponderant share of observations relative to other ordered outcome categories.

Figure A.10: Excessive Top-Category Observations, Bobik and Smith (2013)



It is plausible that the inflated top-category ("Substantial Concessions") in the ordered dependent variable illustrated above is produced from two distinct theoretical processes that incorporate the same top-category outcome associated with two different types of WTO-member states : the non-inflated cases of WTO-member state defendants (or plaintiffs) whose domestic norms or robust legal system induced them to *voluntarily* comply with the WTO's ruling by providing substantial concessions and the inflated cases of WTO-member state defendants (or plaintiffs) who were compelled or *coerced* into providing such concessions because of concerns about retaliatory action imitated by the economically larger (i.e., more powerful) state involved in the bilateral trade dispute.

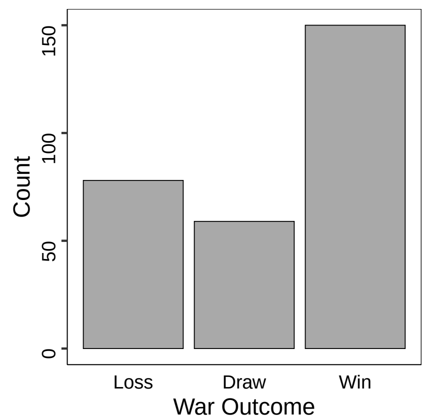
A.5.3. Carter et al. (2012): Interstate *War Outcome*

Ordered (and binary) measures of interstate conflict outcomes in directed-dyad-year or dyad-year datasets are frequently zero-inflated (eg., Senese 1997; Lemke and Reed 2001; Clark and Regan 2003; Pevehouse 2004; Xiang 2010). Yet, it is possible that ordinal outcome measures of interstate conflict, in particular, can exhibit top-category inflation. A good example of such top-category inflation is Carter, Bernhard, and Palmer's (2012) (hereafter Carter et al.) ordered measure of *War Outcome* analyzed in the *Journal of Conflict Resolution*. Specifically, the authors evaluate the relationship between social revolutions and interstate war outcomes. They employ directed-dyad-year data (1900-2001) to test their prediction in "Hypothesis 3" about the likelihood with which postrevolutionary states win interstate wars relative to nonrevolutionary states (Carter et al., 444). The authors code their 0 to 2 ordered *War Outcome* dependent variable which—given their directed-dyad research design—is coded "as 2 if state A won, 1 if the war ended as a draw, and 0 if state A lost" (Carter et al., 449) the war. As illustrated in Figure A.11, the top-category of 2 ("State A won") of *War Outcome* incorporates a preponderant share of observations. We contend that this top-category is potentially generated by the following distinct theoretical processes: non-inflated cases of states (i.e., State A) that have won the war because of their postrevolutionary status (as suggested and operationalized by Carter et al. in the measure's ordered continuum) and inflated cases of states that have won the war because of "structural" reasons such as *geographic advantages* that are unrelated to or distinct from their postrevolutionary status.

A.5.4. Getmansky et al. (2018): *Trust in Reintegration*

Ordered (and binary) survey response dependent variables that are drawn from survey experiments or standard survey-response questionnaires are increasingly used by IR scholars. Survey response outcome measures of this sort have been used to assess political or (foreign) policy attitudes of citizens, policy-making elites, and ex-combatants who served in rebel

Figure A.11: Excessive Top-Category Observations from Carter et al. (2012)

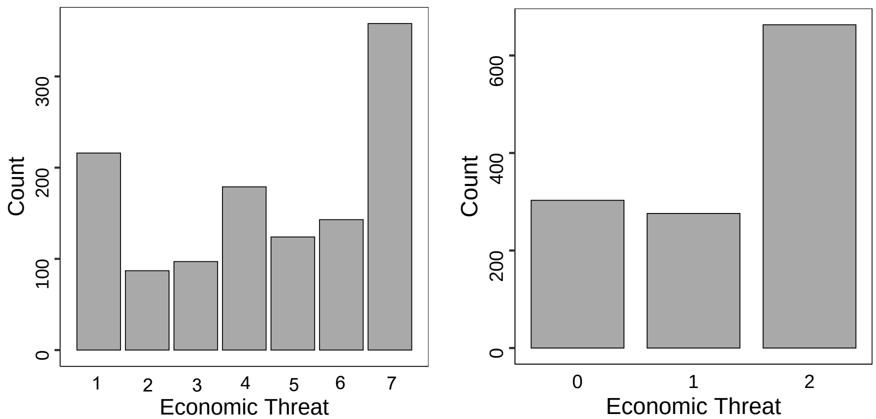


groups or militia organizations (e.g., Getmansky, Sınmazdemir and Zeitzoff 2018; Kreutz and Nussio 2019; Tomz and Weeks 2020). An analysis of these ordinal survey-response dependent variables reveals that some of these measures are top-category inflated. For instance, consider Getmansky, Sınmazdemir, and Zeitzoff's (2018) (hereafter Getmansky et al.) ordered survey-respondent measure on perceptions of *Economic Threat* from refugees published in the *Journal of Political Research*. This ordered measure was developed from a survey experiment that these authors conducted in Turkey to assess how individual perceptions about the adverse consequences of refugee inflows (operationalized through different primes) influence the attitude of survey respondents towards refugees. Their ordered dependent variable, *Economic Threat*, measures the degree to which survey "respondents view the refugees as an economic threat" (p.497). The ordinal *Economic Threat* variable ranges from 1 to 7, which increases from the minimum value of 1 that indicates "strong agreement" with the statement that refugees pose an economic threat to the maximum value of 7 which indicates "strong disagreement" with said statement (Getmansky et al. 2018, 497). As illustrated in Figure A.12a, the combination of the highest outcome categories in the aforementioned ordered survey response measure that focuses on disagreement with the economic threat statement—and likewise, the highest outcome category of "disagreement" in the 0 to 2 coarse version of this measure (Figure A.12b)—has an excessive share of observations.

It is possible that the preponderant share of observations in the highest outcome category

Figure A.12: Excessive Top-Category Observations, Getmansky et al. (2018)

(a) Economic Threat (b) Economic Threat (Coarse)



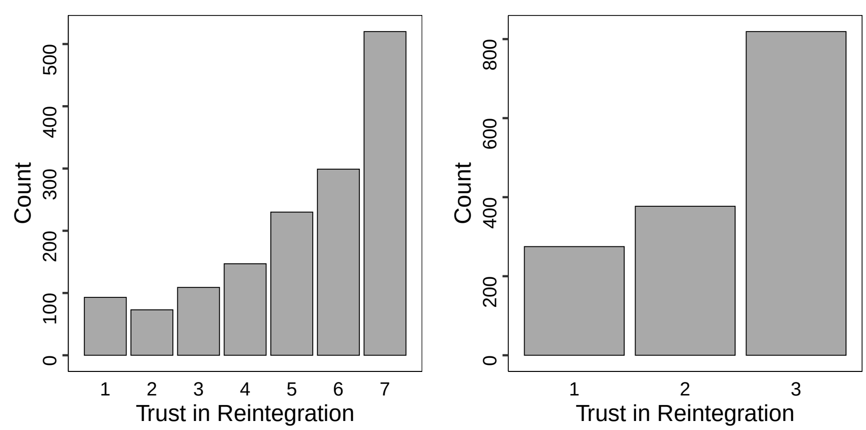
Note: The 7-category *Economic Threat* variable in Figure A.12a was transformed to a 0-1 scale in Getmansky et al.'s (2018) replication dataset.

(or categories) of disagreement include responses from the following two groups of survey respondents. The first group is the non-inflated set of *informed* respondents who opt for the disagreement outcome categories based on their knowledge or research—and hence informed evaluation—about the actual lack of economic threat from refugees. Note that this non-inflated set of survey respondents in the topcategory follow Getmansky et al.'s theoretially-assumed ordered continuum of respondents who opt for the highest outcome category given their knowledge about the economic costs of refugee inflows. The second group, however, are the inflated set of respondents that includes *uninformed* respondents who do not have much (if any) knowledge about the economic impact of refugee flows but who select the disagreement outcome category for social desirability reasons including their desire to avoid being perceived as bigoted or as an individual who lacks empathy for refugees. As scuh, this inflated set of uninformed respondents are

Next, we briefly discuss Kreutz and Nussio's (2019) innovative study in which they use data from a 2008 survey questionnaire that was administered to ex-combatants from several rebel groups in Colombia who had fought the Colombian government during the country's long drawn-out civil conflict. A key objective underlying this survey response analysis is to estimate the extent to which these ex-combatants trusted the government's intention to genuinely reintegrate them (via the reintegration scheme called "ACR") into the country's society, polity, and economy. To this end, they operationalize from their survey response data of ex-combatants an ordered dependent variable that captures their level of trust—which ranges from "no" to "much" (i.e., very high) trust—in the government's reintegration effort. Figures A.13a–A.13b illustrate the 1-7 fine-grained and 0 to 2 coarse versions of this ordered *Trust Reintegration* survey-response dependent variable,

Figure A.13: Excessive Top-Category Observations, Kreutz and Nussio (2019)

(a) Trust Reintegration (b) Trust Reintegration (Coarse)



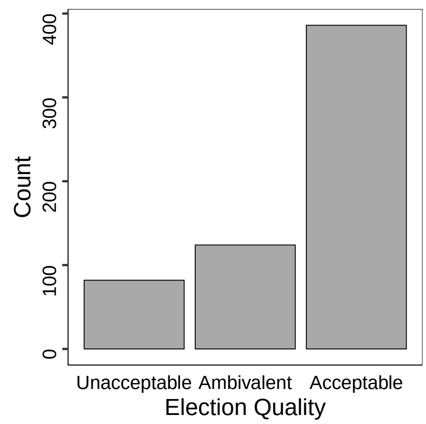
It is clear that the top-category of "much" (that is, very high trust) of the ordered *Trust Reintegration* in both these figures is inflated as it includes a substantially higher share of observations compared to the other ordered outcome categories. There are sound theoretical reasons to believe that this inflated top-category emerges from two distinct d.g.p.'s: ex-combatants who chose the very high trust category since they may have believed the sincerity of the government's reintegration effort (or because of their genuine desire for peace) versus those who opted for this top-category—as opposed to other lower categories—either because of social desirability reasons or because they were fearful of retaliatory action by government officials (in case their identity was compromised).

A.5.5. Kavakli and Kuhn (2020): *Election Quality*

Next, consider the recent study by Kavakli and Kuhn (2020) published in *International Organization*. Their study evaluates when election monitors from International Intergovernmental Organizations (IGOs) and Non-Governmental Organizations (NGOs) endorse national election results (as opposed to rejecting these results on grounds of electoral fraud) across developing states (1990-2004) where democratic institutions are fragile. The main hypothesis that Kavakli and Kuhn (2020) propose is that international election monitors from IGOs and NGOs are more likely to endorse elections in states where Islamic opposition parties operate in the context of Islamic terrorist activity (p.152). They test this hypothesis by employing a binary dependent variable (labeled *Acceptable*) and an ordered dependent variable (denoted as *Election Quality*) for robustness tests. Given our focus on the TiOP(C) model, we explore their ordered *Election Quality* dependent variable that is operationalized as per the following ordinal scale: the measure is coded "0" if the election monitors from IGOs and NGOs state that the election represents the will of the people, "0.5" if they are ambivalent and thus not possible to discern the overall veracity of the election outcome, and "1" if the election does not represent the will of the people.

The frequency distribution of the ordered *Election Quality* dependent variable illustrated in Figure A.14 reveals that this measure's top-category—namely, the election monitor's assessment that the election results in question are acceptable—is inflated as this category includes almost two-thirds of all the coded observations in the dependent variable. We suggest from a theoretical perspective that this inflated top-category results from two distinct d.g.p.'s. The first d.g.p. stems from the possibility that the election monitors from the international IGOs and NGOs likely obtained credible information about the validity of the election results and were thus able to accurately infer that the said results are *not* "questionable." The second d.g.p., which is also plausible, is that election monitors may have not obtained sufficient information about the election outcome and thus compensated for this poor information by simply exaggerating or inaccurately reporting the said outcome

Figure A.14: Excessive Top-Category Observations, Kavakli and Kuhn (2020)



as acceptable (also driven by organizational imperatives) even if that was not the case.

A.6. Monte Carlo Experiments

This section presents the results from six main Monte Carlo experiments. These experiments collectively assess the performance of the TiOP, TiOPC, and OP models under conditions where the highest-most category of a four category ordered dependent variable is believed to have (or not have) inflation due to a second, (correlated) data generating process (d.g.p.) that is itself distinct from the primary ordered d.g.p. of interest. The first of these experiments examines the performance of TiOP, TiOPC, and OP models when one's d.g.p. can be characterized as TiOP. Experiment 2 reevaluates our respective findings from Experiment 1 when the d.g.p. is instead TiOPC. Experiment 3 then compares the TiOP, TiOPC, and OP models under conditions where the true d.g.p. is instead OP. Finally, Experiments 4-6 reassess the performance of the TiOP, TiOPC, and OP models under TiOPC d.g.p.'s where the correlation between one's inflation stage and OP stage (ρ) is varied from 0.25 (Experiment 4) to 0.50 (Experiment 5) to 0.75 (Experiment 6).

For Monte Carlo Experiments 1-2 and 4-6, we set the degree of inflation in the highest-most category of the TiOP- and TiOPC-generated ordered dependent variable to the moderately high level of 85%. We set inflation to 0% for Experiment 3. Each Monte Carlo experiment uses simulated datasets with N's of 1,000 and with $sims = 1,000$. In each case, we draw an outcome (i.e., ordered probit) stage covariate $\mathbf{x}$ from $\mathbf{x} = (\mathbf{1}, \mathbf{x_1})'$ where $\mathbf{x_1}$ is the natural log of $Uniform[0, 100]$. For Experiments 1-2 and 4-6, we define two inflation stage covariates $\mathbf{z} = (\mathbf{1}, \mathbf{z_1}, \mathbf{z_2})'$. We assign our first inflation stage covariate $\mathbf{z_1} = \mathbf{x_1} = ln(Uniform[0, 100])$ to approximate real world conditions where a covariate exhibits effects on the propensity to provide both an ordered response and an inflated response. We then set our second inflation stage covariate to $\mathbf{z_2} = 1_{Uniform[0,1]>0.25}$. This second inflation stage covariate thereby approximates the types of imbalanced binary independent variables that are commonly used in analyses of international relations data.

For Experiment 2, we set the correlation parameter ρ within our corresponding TiOPC d.g.p. to correspond to the value of 0.50. This value precisely matches the value of ρ that is used in a number of similar Monte Carlo experiments examining either the zero-inflated ordered probit with correlated error (ZiOPC) d.g.p. or the middle-inflated ordered probit with correlated errors d.g.p. (Harris and Zhao, 2007; Bagozzi and Mukherjee, 2012). We then reconsider ρ across a wider range of values for the TiOPC d.g.p.'s that we employ in Experiments 4-6; in these cases, we vary ρ across $\{0.25, 0.50, 0.75\}$. Note that our TiOP d.g.p. in Experiment 1 can also be seen as an instance of a TiOPC d.g.p. where $\rho = 0$.

Across all experiments, and consistent with extant simulation studies of inflated limited dependent variable models (Harris and Zhao, 2007; Bagozzi and Mukherjee, 2012), draws of $\mathbf{x}$ and $\mathbf{z}$ were taken once at each N and then held fixed for each simulation therein. Parameter values were assigned as $(\beta_0, \beta_1)' = (4.5, -1)'$ for our ordered-stage predictors in Experiments 1-6; and as $(\gamma_1, \gamma_2, \gamma_3)' = (1, -2.5, -1)'$ for our inflation stage predictors (i.e., $\mathbf{z}$'s) in Experiments 1-2 and 4-6. These specifications produced a four category ordered dependent variable $\mathbf{y} = (0, 1, 2, 3)$ with either (i) an average rate of 85% inflation in the top-category, $y = 3$, in Experiments 1-2 and 4-6[15] or (ii) 0% inflation in the top-category of y (in the case of Experiment 3).

[15]I.e., 85% inflation as a share of all $\mathbf{y} = 3$ responses. This yields 65% (global) inflation as a share of all outcomes of $\mathbf{y}$; and places approximately 75% of our highest-most ordered outcome ($\mathbf{y} = 3$).

Like many limited dependent variable models, coefficient estimates for the TiOP, TiOPC, and OP models are not directly interpretable, and researchers employing these models are therefore likely to be most interested in estimated first differences in the predicted probability of each outcome category on y (hereafter "first differences"), and potential biases therein. For all experiments—and consistent with Harris and Zhao (2007) and Bagozzi et al. (2015)—our Monte Carlos accordingly calculate the estimated effect of the following two scenarios on the probability of observing each category of y: (i) an increase in $\mathbf{z_1} \equiv \mathbf{x_1}$ from its mean to one standard deviation above its mean or (ii) a $0 \rightarrow 1$ change in $\mathbf{z_2}$. We use parametric bootstraps to calculate all corresponding estimated effects and their associated 95% confidence intervals, using $m = 1,000$. When calculating a first difference for a particular variable, we hold our other variable value to its mean (in the case of $\mathbf{z_1} \equiv \mathbf{x_1}$) or mode (in the case of $\mathbf{z_2}$). The recovered first differences in predicted probabilities for a particular simulation run are then compared to the true first differences obtained during that same simulation run in order to obtain several of the quantities of interest discussed below.

The results for Monte Carlo Experiments 1-3 appear in Tables A.8-A.10. Within each table, we report the mean first difference estimates for our TiOP, TiOPC, and OP models; the root mean squared errors (RMSEs) for these first difference estimates, these first difference estimates' 95% empirical coverage probabilities (CPs),[16] and a number of model selection criteria. With respect to the latter items, we separately report the percentage of times that the correct model was identified by (i) the Akaike information criterion (AIC), (ii) the Bayesian information criterion (BIC), (iii) a likelihood ratio (LR) test; and (iv) a test of whether ρ is statistically significant at the $p < 0.05$ level (denoted as "ρ test").[17] Figures A.15-A.17 then present and compare the complete distributions of each estimated TiOP, TiOPC, and OP first difference quantity across all simulations for each of our first three

[16]The average proportion of times—out of $1,000$ simulations—that a true first difference value fell within the 95% confidence intervals of a first difference estimate.

[17]Instances of nonconvergence were relatively rare and were largely limited to the application of the TiOPC to an OP generated variable (i.e., Experiment 3). In these instances, we treated nonconvergence of the TiOPC as a failed ρ test.

Table A.8: Marginal Effects For OP, TiOP, and TiOPC Models under a TiOP DGP

			Change in $Pr(Y=0)$					Change in $Pr(Y=1)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	*Mean*	0.052	0.020	0.051	0.051	$x_1 \equiv z_1$	-0.031	0.015	-0.032	-0.033
	RMSE		(0.032)	(0.012)	(0.012)			(0.047)	(0.009)	(0.010)
	CP		(0.029)	(0.950)	(0.953)			(0.000)	(0.956)	(0.956)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	*Mean*	-0.067	.	-0.067	-0.066	z_2	-0.127	.	-0.127	-0.127
	RMSE		.	(0.009)	(0.015)			.	(0.015)	(0.016)
	CP		.	(0.941)	(0.916)			.	(0.947)	(0.953)
			Change in $Pr(Y=2)$					Change in $Pr(Y=3)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	*Mean*	-0.079	0.011	-0.077	-0.079	$x_1 \equiv z_1$	0.059	-0.047	0.058	0.059
	RMSE		(0.091)	(0.009)	(0.009)			(0.106)	(0.023)	(0.023)
	CP		(0.000)	(0.953)	(0.967)			(0.000)	(0.946)	(0.948)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	*Mean*	-0.142	.	-0.138	-0.139	z_2	0.336	.	0.332	0.333
	RMSE		.	(0.015)	(0.021)			.	(0.029)	(0.029)
	CP		.	(0.942)	(0.961)			.	(0.949)	(0.957)

	% Sims Preferring TiOP to OP	% Sims Preferring TiOPC to OP	% Sims Preferring TiOP to TiOPC
AIC	100	100	84
BIC	100	100	100
LR Test	100	100	95
ρ Test	.	.	94

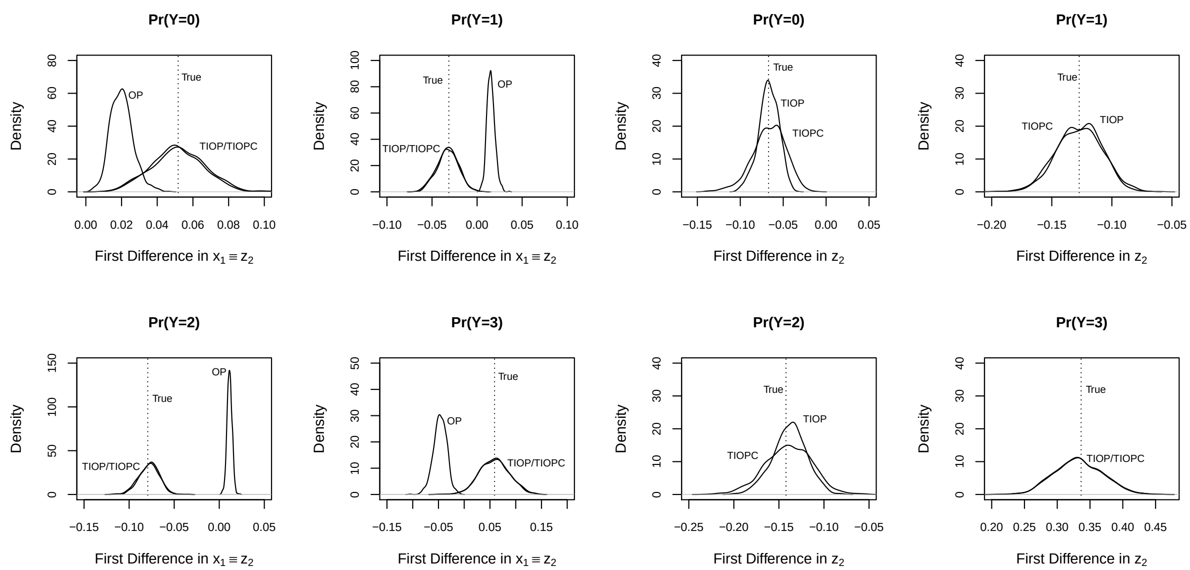
experiments.

We first discuss our result for Experiment 1. Recall that Experiment 1 evaluates the performance of the TiOP, TiOPC, and OP models when the true d.g.p. is TiOP. Table A.8 reveals very favorable results for TiOP and TiOPC models with respect to both accuracy and coverage. The TiOP model's first difference estimates pertaining to $x_1 \equiv z_1$ are highly similar to the TiOPC model's first difference estimates. Moreover, each (TiOP and TiOPC) first difference estimate is remarkably close to its true first difference value reported in Table A.8. This is not the case for the OP $x_1 \equiv z_1$ first difference estimates. For example, for $Pr(Y = 1)$, our OP first difference estimate for $x_1 \equiv z_1$ is in the opposite direction of the true first difference of $x_1 \equiv z_1$, with a mean first difference estimate of 0.015 relative to the true first difference value of -0.031.[18] By comparison, the average TiOP and TiOPC first difference estimates of $x_1 \equiv z_1$ on $Pr(Y = 1)$—of -0.032 and -0.033, respectively—are each

[18]This OP estimation of a first difference effect that is reversed in sign relative to the true first difference effect is also the case for the $Pr(Y = 2)$ and $Pr(Y = 3)$ outcomes in Table A.8.

Figure A.15: Distributions of First Differences Across All Simulations, Experiment 1

(a) First Differences for $x_1 \equiv z_1$ (b) First Differences for z_2



much closer to the true first difference effect of -0.031.

These observations are reinforced by each model's RMSEs for Experiment 1's $x_1 \equiv z_1$ first difference estimates in Table A.8, as well as in the distributions of first difference estimates depicted in Figure A.15a. For the former quantities, we find that the TiOP(C) RMSEs are consistently superior to those of the OP. Moreover, in Table A.8, the TiOP RMSEs for our $x_1 \equiv z_1$ effects are either comparable to, or superior to, the associated RMSE's for the TiOPC model's estimated first difference. In regards to Figure A.15a (i.e., the distributions of estimated first differences for $x_1 \equiv z_1$ across all simulations), one can likewise observe that the TiOP and TiOPC first differences are centered over the true first difference values for each outcome category considered. By contrast, the OP model's first difference distributions for $x_1 \equiv z_1$ fall above or below the true first difference effect in every instance—illustrating the biases that arise when an OP model is misapplied to a TiOP-generated dependent variable.

Looking beyond these estimated first differences and RMSEs, we can also note in Table A.8 that the 95% CPs for $x_1 \equiv z_1$ favor the TiOP model over the OP and TiOPC models in every instance. These CPs further indicate that the TiOP model's first difference estimates encompass each outcome category's true first difference effect in 94.6%-95.6% of our simulations, whereas the OP model's first difference estimates encompass these true values in only 0.0%-2.9% of our simulations. The TiOPC CPs for $x_1 \equiv z_1$ (of 94.8%-96.7%) are comparable to those of the TiOP model in Table A.8, and also consistently favor the TiOPC model over the OP model's aforementioned CPs (of 0.0%-2.9%). For our z_2 CPs, the TiOP and TiOPC models' 95% confidence intervals recover our true first difference values within the vast majority (94.1%-94.9% and 91.6%-96.1%, respectively) of all simulations. The TiOP(C) models also recover our z_2 first differences with a high degree of accuracy, as evidenced by (1) the relatively low RMSEs for these first difference estimates across all ordered outcomes in Table A.8 and (2) the overall distributions of our z_2 first difference estimates depicted in Figure A.15b.

Returning to Table A.8, our four model selection criteria (i.e., the *AIC*, *BIC*, *LR* test, and ρ test) select the correct model in virtually every instance—favoring the TiOP (and TiOPC) over the OP in 100% of our simulations, and favoring the TiOP over the TiOPC in 84%-to-100% of all simulations.[19] In sum, when the d.g.p. is TiOP, our OP first difference estimates exhibit substantial bias and poor empirical coverage. The TiOP and TiOPC models, on the other hand, exhibit impressive—and often comparable—degrees of accuracy and coverage. However, model selection criteria (correctly) suggest that in this case the TiOP should be preferred to the TiOPC—a conclusion that is reinforced by the slightly more precise TiOP estimates of each z_2 first difference quantity in Figure A.15b.

We next turn to our second Monte Carlo experiment. Recall that Experiment 2 reevaluates the performance of the TiOP, TiOPC, and OP models when the true d.g.p. is TiOPC, rather than TiOP. For this second experiment, we find in Table A.9 and Figures A.16a-A.16b that our TiOPC model performs admirably with respect to both empirical coverage and accuracy when the underlying d.g.p. is TiOPC. In terms of coverage, we can note that

[19] We find here that the *BIC* is more accurate in identifying the proper (i.e., the TiOP) model under a TiOP d.g.p. than is the AIC. The former correctly selects the TiOP over the TiOPC in 100% of our simulations, while the latter only correctly identifies the correct (TiOP) model in 84% of our simulations.

across all estimated first difference effects and ordered outcome categories of interest, the 95% confidence intervals to our TiOPC model's first difference estimates recover the true first difference values for each outcome category in 94% of all simulations, on average, in Experiment 2. This average 95% CP rate is slightly lower than the TiOPC CP rate obtained under the TiOP d.g.p. in Experiment 1, but is far superior to the average TiOP and OP 95% CPs in Experiment 2—whose averages come to 68% and 24%, respectively.

Table A.9: Marginal Effects For OP, TiOP, and TiOPC Models under a TiOPC DGP

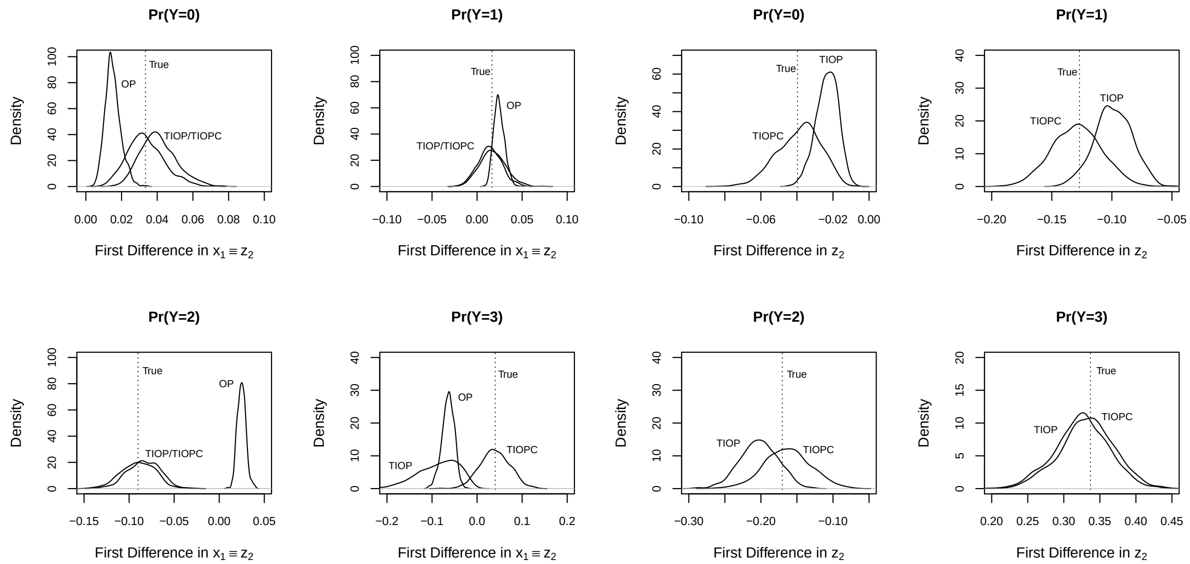
			Change in $Pr(Y=0)$					Change in $Pr(Y=1)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	0.033	0.015	0.041	0.041	$x_1 \equiv z_1$	0.016	0.024	0.014	0.017
	$RMSE$		(0.018)	(0.010)	(0.008)			(0.008)	(0.011)	(0.011)
	CP		(0.102)	(0.918)	(0.931)			(0.859)	(0.955)	(0.953)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	-0.040	.	-0.023	-0.039	z_2	-0.127	.	-0.099	-0.129
	$RMSE$		.	(0.016)	(0.010)			.	(0.029)	(0.017)
	CP		.	(0.361)	(0.944)			.	(0.579)	(0.969)
			Change in $Pr(Y=2)$					Change in $Pr(Y=3)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	-0.090	0.025	-0.084	-0.090	$x_1 \equiv z_1$	0.040	-0.064	-0.086	0.040
	$RMSE$		(0.115)	(0.015)	(0.015)			(0.105)	(0.126)	(0.026)
	CP		(0.000)	(0.919)	(0.956)			(0.000)	(0.000)	(0.924)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	-0.170	.	-0.203	-0.163	z_2	0.337	.	0.325	0.332
	$RMSE$		.	(0.035)	(0.026)			.	(0.031)	(0.030)
	CP		.	(0.770)	(0.927)			.	(0.915)	(0.935)

	% Sims Preferring TiOP to OP	% Sims Preferring TiOPC to OP	% Sims Preferring TiOPC to TiOP
AIC	100	100	82
BIC	100	100	42
LR Test	100	100	67
ρ Test	.	.	80

The RMSEs reported in Table A.9 likewise suggest that our TiOPC model—and to a lesser extent our TiOP model—each exhibit relatively high levels of accuracy when applied to a TiOPC-generated dependent variable. In this case, our obtained RMSEs consistently favor the TiOPC model over the TiOP model when one considers the estimated first difference effects for our exclusively-inflation stage covariate (i.e., z_2). We similarly find that our TiOPC RMSEs are comparable to—or superior to—those of the TiOP model for each of the estimated first differences associated with $x_1 \equiv z_1$. In three (two) out of our four ordered outcome categories of interest, we likewise find lower (i.e., superior) RMSEs values

Figure A.16: Distributions of First Differences Across All Simulations, Experiment 2

(a) First Differences for $x_1 \equiv z_1$ (b) First Differences for z_2



for our TiOPC (TiOP) model's first difference estimates in relation to the corresponding OP-estimated first differences. For the latter (OP) estimates, empirical coverage also remains remarkably poor with an average of only 24% of all simulations exhibiting empirical coverage at the 95% level.

These findings are reinforced by the distributions of estimated first differences in Figure A.16. Figures A.16a-A.16b together demonstrate that the TiOPC model is clearly superior to the OP and TiOP models in approximating our true $x_1 \equiv z_1$ first differences for all four outcome categories, and that the TiOP model similarly outperforms the OP model across at least three of our four ordered outcomes of interest. Within Figure A.16b, the TiOPC model is typically superior to the TiOP model in recovering all true first differences for z_1. That being said, we do find in these Experiment 2 Figures that the TiOP and TiOPC models are at times each slightly less accurate in recovering our true first difference effects in comparison to Figure A.15 (Experiment 1).

We next turn to Experiment 2's model selection criteria reported in Table A.9. Recall that these quantities report the percentage of simulations that correctly select the true model (in this case the TiOPC) relative to a comparison model. We find that each selection paradigm consistently identifies the TiOPC (and TiOP) model as superior to the OP model across all simulations considered. Our selection criteria further identify the TiOPC model as a superior choice to the TiOP model in 42%-to-82% of all simulations considered. Looking across these model selection criteria for both Experiment 1 and Experiment 2, the *LR* test and ρ test each appear to be the most effective in correctly selecting the proper TiOP or TiOPC model depending upon the underlying d.g.p. The *AIC* and *BIC* are highly effective as well, though the *AIC* (*BIC*) tends to marginally favor the TiOPC (TiOP) no matter the circumstance.[20]

Altogether, and returning to the full set of Table A.9 results discussed above, our Experiment 2 results indicate that when a d.g.p. is believed or suspected to be TiOPC, the TiOPC model is the optimal choice for estimation and inference, while the OP model is an especially poor choice. This conclusion is widely supported by our model selection statistics, and especially by the *LR* test and ρ test. We withhold further discussion of these TiOPC-d.g.p. findings until our evaluations our TiOPC-d.g.p. results anew under two additional levels of correlation between the inflation and ordered stages of a TiOPC-generated dependent variable in Experiments 4-6 below.

Experiments 1-2 indicate that when the underlying d.g.p. is either TiOP or TiOPC, the TiOP(C) models outperform the OP model in terms of accuracy and coverage. In Experiment 3, we evaluate how the TiOP(c) models perform when they are each incorrectly applied to an ordered dependent variable whose true d.g.p. is OP. We do so by applying our three models of interest to an ordered outcome variable that contains no inflation. To mirror real-world situations of potential TiOP(C) misspecification in such contexts, we continue to include our inflation stage covariates within the TiOP(C) specifications for this experiment. Turning to Table A.10 and Figure A.17, we can first notice in Experiment 3 that our OP first difference estimates and their 95% CP's have substantially improved relative to Experiments 1-2. For

[20]However, note that these latter observations are likely sensitive to N and k.

Table A.10: Marginal Effects For OP, TiOP, and TiOPC Models under an OP DGP

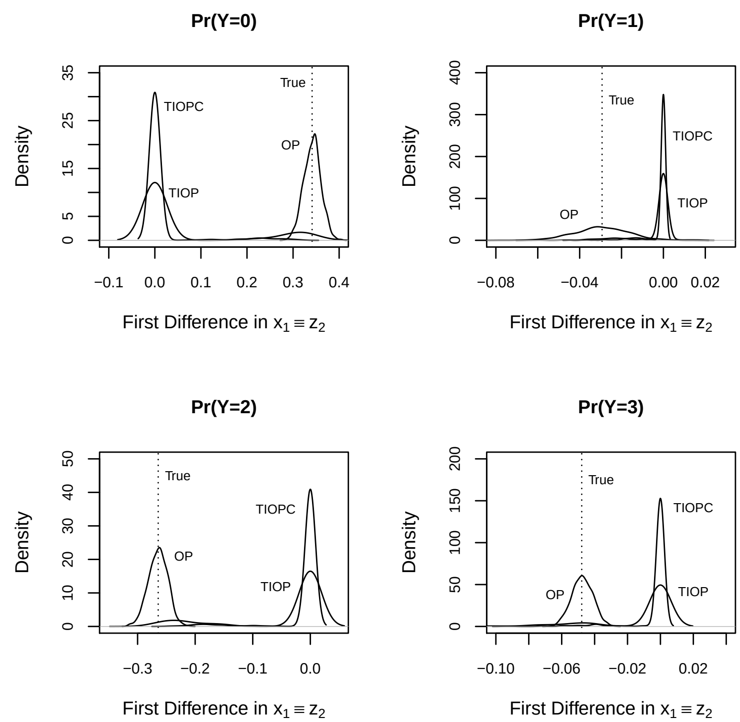
			Change in $Pr(Y=0)$					Change in $Pr(Y=1)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	0.341	0.342	0.056	0.056	$x_1 \equiv z_1$	-0.029	-0.030	-0.003	-0.001
	$RMSE$		(0.015)	(0.008)	(0.008)			(0.010)	(0.003)	(0.001)
	CP		(0.948)	(0.164)	(0.053)			(0.943)	(0.154)	(0.060)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	0.000	.	-0.001	-0.000	z_2	0.000	.	-0.003	-0.000
	$RMSE$		.	(0.002)	(0.001)			.	(0.004)	(0.001)
	CP		.	(0.183)	(0.057)			.	(0.182)	(0.057)
			Change in $Pr(Y=2)$					Change in $Pr(Y=3)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	-0.264	-0.265	-0.041	-0.010	$x_1 \equiv z_1$	-0.048	-0.047	-0.012	-0.002
	$RMSE$		(0.013)	(0.009)	(0.006)			(0.005)	(0.004)	(0.001)
	CP		(0.937)	(0.164)	(0.053)			(0.948)	(0.181)	(0.060)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	0.000	.	-0.003	-0.000	z_2	0.000	.	0.007	0.001
	$RMSE$		.	(0.004)	(0.001)			.	(0.010)	(0.004)
	CP		.	(0.183)	(0.057)			.	(0.181)	(0.059)

	% Sims Preferring OP to TiOP	% Sims Preferring OP to TiOPC	% Sims Preferring TiOP to TiOPC
AIC	86	98	98
BIC	100	100	100
LR Test	94	99	99
ρ Test	.	.	98

each and every ordered outcome category considered for Experiment 3, our OP-estimated first differences in the effect of $x_1 \equiv z_1$ fall within a thousandth of a decimal place of the true parameter value, and exhibit empirical CPs that range from 93.7% to 94.8%. The *AIC*, *BIC*, and *LR* test results reported in Table A.10 similarly indicate that these criteria identify the OP model as the correct model (relative to either the TiOP or TiOPC models) across 86% to 100% of all simulations—with the *BIC* proving to be the most consistent.

By comparison, the TiOP and TiOPC models' average estimated first differences for $x_1 \equiv z_1$ are fairly close to one another, but are each attenuated towards zero in relation to the true first difference values for each ordered outcome category of interest. This can be clearly seen in both Table A.10 and Figure A.17, wherein our TiOP(C) first difference distributions for $x_1 \equiv z_1$ are centered over zero (see Figure A.17). This bias in our TiOP(C)-estimated first differences is in large part a function of our inclusion of x_1 as z_1 within these models' inflation stages under Experiment 3. That is, the inclusion of x_1 as z_1 under an OP d.g.p. leads the TiOP(C) models to (accurately) estimate a value for γ_1 that is close to

Figure A.17: Comparison of First Differences for $x_1 \equiv z_1$ Under Experiment 3 (OP DGP)



zero; which then biases our TiOP(C)-first difference estimates of x_1 given our changing of x_1 and z_1 simultaneously during the TiOP(C) first difference calculations for this covariate's estimated effect.

As such, our comparisons in Table A.10 and Figure A.17 present an especially high bar for the TiOP(C) models in this instance. Even so, we find that the mean estimated TiOP(C) first differences in these cases do not differ from our true first differences in as severe a manner as was highlighted for the OP first difference estimates under a TiOP or TiOPC d.g.p. above (i.e., under Experiments 1-2).[21] The latter observation is reinforced by the RMSEs for the TiOP(C) models in Experiment 3 (Table A.10), which are generally (albeit slightly) superior to those reported for the OP model for each outcome category. On the other hand, the TiOP(C) models' estimated first differences perform noticeably worse

[21]For example, Experiment 3's TiOP- and TiOPC-estimated first difference effects exhibit the same sign as the true first difference effects, which was not always the case for the OP model's marginal effects in the previous experiments.

than those of the OP model in terms of empirical coverage. For example, the 95% CPs obtained for the aforementioned TiOP(C) models in Experiment 3 only recover the true first difference values in 5.3%-18.3% of all relevant simulations. Thus, the OP model should be strongly preferred in instances where one suspects the true d.g.p. to be OP—a conclusion that is consistently supported by our model selection criteria. Moreover, our 95% CPs and Figure A.17 each strongly suggest that the TiOP(C) models are far less accurate than the OP model in these contexts.

Our next set of Monte Carlo experiments endeavor to more rigorously evaluate the performance of the TiOP, TiOPC, and OP models under a TiOPC d.g.p. when we vary the level of correlation between the inflation stage and ordered outcome stage. In this manner, these additional evaluations parallel the ZIOP(C) and OP experiments conducted by Bagozzi et al. (2015). We specifically consider three realistic levels of correlation between one's inflation and TiOPC-outcome stages: $\rho = \{0.25, 0.50, 0.75\}$.[22]

We begin with Experiment 4, which considers a TiOPC d.g.p. with a relatively low level of correlation (i.e., with $\rho = 0.25$). The results from this experiment are presented in Table A.11 and Figure A.18. In Table A.11, we find that our 95% CPs favor the TiOPC model over the TiOP model in six out of eight possible instances. The superior performance of the TiOPC model in this case is reaffirmed by an aggregate comparison of each model's average CP value across all relevant first differences, wherein we find that the TiOPC model exhibits a CP average of 86% in comparison to the TiOP model's CP average of 73%. Relative to the case of $\rho = 0.50$ (i.e., Experiment 2), this TiOPC average is lower than the TiOPC model's CP performance under situations of moderate correlation, whereas the TiOP model's CP average is now noticeably higher than was the case when $\rho = 0.50$. The TiOPC (TiOP) model also outperforms the OP model's 95% CPs in Experiment 4 across three (two) of four possible outcome categories. This finding is further reinforced by the OP model's average 95% CP value of 33% for Experiment 4, which is far lower than the TiOP(C) CP averages

[22]Recall that the case of $\rho = 0.50$ was also considered in Experiment 2.

Table A.11: Marginal Effects For OP, TiOP, and TiOPC Models under a TiOPC DGP with 25% Inflation

			Change in $Pr(Y=0)$					Change in $Pr(Y=1)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	0.033	0.018	0.048	0.048	$x_1 \equiv z_1$	0.016	0.020	-0.011	-0.009
	$RMSE$		(0.015)	(0.016)	(0.014)			(0.005)	(0.027)	(0.026)
	CP		(0.358)	(0.789)	(0.858)			(0.977)	(0.419)	(0.491)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	-0.040	.	-0.043	-0.055	z_2	-0.127	.	-0.116	-0.126
	$RMSE$		.	(0.008)	(0.017)			.	(0.016)	(0.016)
	CP		.	(0.933)	(0.819)			.	(0.889)	(0.958)
			Change in $Pr(Y=2)$					Change in $Pr(Y=3)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	-0.090	0.017	-0.083	-0.086	$x_1 \equiv z_1$	0.040	-0.055	-0.075	0.050
	$RMSE$		(0.107)	(0.013)	(0.012)			(0.095)	(0.115)	(0.025)
	CP		(0.000)	(0.897)	(0.945)			(0.000)	(0.000)	(0.944)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	-0.170	.	-0.168	-0.149	z_2	0.337	.	0.327	0.330
	$RMSE$		.	(0.018)	(0.029)			.	(0.030)	(0.030)
	CP		.	(0.945)	(0.886)			.	(0.930)	(0.940)

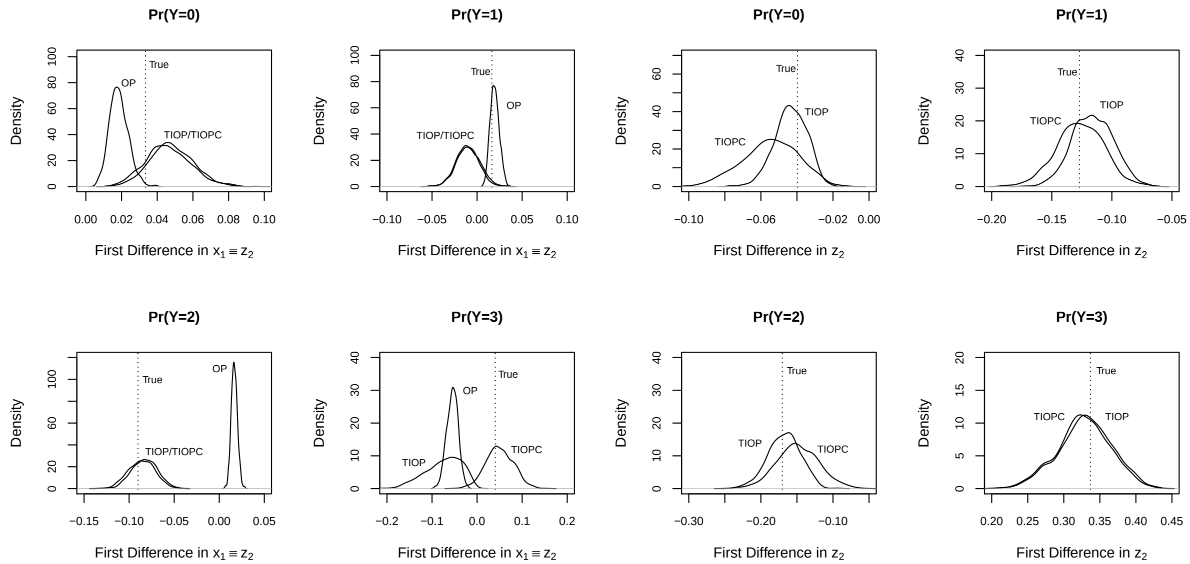
	% Sims Preferring TiOP to OP	% Sims Preferring TiOPC to OP	% Sims Preferring TiOPC to TiOP
AIC	100	100	39
BIC	100	100	7
LR Test	100	100	19
ρ Test	.	.	34

(of 86% and 73%) discussed above.

Table A.11 indicates that our TiOPC model exhibits comparable RMSE values—and hence accuracy in estimated first differences—to those of the TiOP model in most instances. However, at times we find that our TiOP and TiOPC RMSEs are effectively identical, whereas in other cases one model marginally outperforms the other in RMSEs. The same can be said for the OP model's RMSEs, which are inferior to the TiOPC (TiOP) model in three (one) of four possible instances—in one case almost by a factor of 10. Nevertheless, at least in the aggregate, each model—and especially the TiOP(C) models—tend to perform comparably in RMSE and thus accuracy when the level of correlation (ρ) in a TiOPC-generated dependent variable is low. Yet a more detailed comparison of estimated first differences in Figure A.18 suggests that the TiOP(C) models notably outperform the OP model in terms of (low) bias for this experiment. For the TiOPC model, this can be seen for at least three of our four outcome categories in Figure A.18a. One can reach similar

Figure A.18: Distributions of First Differences Across All Simulations, Experiment 4

(a) First Differences for $x_1 \equiv z_1$ (b) First Differences for z_2



conclusions for the TiOP model in two of these same four outcome categories in Figure A.18a. On the other hand, our estimated first differences for z_2 in Figure A.18b suggest that our TiOP estimates have improved in relation to those presented in Experiment 2. To this end, we find in many instances that the TiOP model offers comparable, and in some cases slightly superior, first differences for z_2 in relation to the TiOPC model, although both models tend to perform well in recovering the true first difference effects in this instance.

Under this particular experiment, the model selection criteria reported in Table A.11 are not always effective in selecting the correct (i.e., TiOPC) model when compared to the performance of these same criteria in Experiment 2.[23] For instance, our four model selection paradigms now properly select the TiOPC model as preferable to the TiOP model in only 7%-39% of all simulations. In this case, the BIC remains the most conservative in its favoring of the TiOPC model in only 7% of all simulations, whereas the ρ test and AIC perform best in correctly selecting the TiOPC model over the TiOP in 34%-39% of all simulations. Given the low ρ considered here, these results are anticipated, and naturally imply that our

[23]And also when compared to the performance of these selection criteria in Experiments 5-6 below.

model selection criteria tend to select the more parsimonious model (amongst the TiOP and TiOPC models) when ambiguity as to the true TiOPC d.g.p. arises. These points aside, we *do* continue to find that *all* four model selection criteria properly favor the TiOP(C) models over the OP model within 100% of simulations considered under Experiment 4. This reaffirms the above findings, in implying that no matter the level of correlation between one's inflation and outcome category, our model selection statistics will properly identify the TiOP(C) models as optimal to an OP model when a d.g.p. is TiOPC.

For a TiOPC d.g.p. with a low level of correlation between one's inflation process and ordered outcome process, the TiOPC model is hence preferred over the TiOP model—and especially over the OP model—in terms of empirical coverage. With respect to accuracy and bias, the differences between the TiOPC and TiOP become much more slight, although the TiOP(C) models do continue to often outperform the OP model on these dimensions. These broader conclusions are each reaffirmed by the model selection statistics considered, wherein these criteria consistently favor the TiOP(C) models over the OP model, but are unable to always identify the TiOPC model as preferable to the TiOP model.

Experiment 5 repeats the same TiOPC d.g.p. setup as presented in Experiment 2, where in this case $\rho = 0.50$. We report and reinterpret these results anew in Table A.12 and Figure A.19 for convenience. Here, we again find that our model selection criteria *always* favor the TiOP model over the OP model, and *always* favor the TiOPC model over the OP model. With regards to the performance of these criteria in correctly favoring the TiOPC model over the TiOP model, the AIC (84%), LR test (67%), and ρ test (80%) each favor the TiOPC model across a majority of relevant simulations. However, the BIC only correctly selects the TiOPC in 42% of all simulations.[24] Altogether, the performance of these model selection statistics suggests that under conditions of modest correlation between one's inflation stage and outcome stages, multiple model selection statistics should be considered, and the LR test and ρ test may be preferable to the AIC and BIC when there is disagreement between

[24]This is in keeping with our findings for the AIC and BIC above.

Table A.12: Marginal Effects For OP, TiOP, and TiOPC Models under a TiOPC DGP with 50% Inflation

			Change in $Pr(Y=0)$					Change in $Pr(Y=1)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	0.033	0.015	0.041	0.041	$x_1 \equiv z_1$	0.016	0.024	0.014	0.017
	$RMSE$		(0.018)	(0.010)	(0.008)			(0.008)	(0.011)	(0.011)
	CP		(0.102)	(0.918)	(0.931)			(0.859)	(0.955)	(0.953)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	-0.040	.	-0.023	-0.039	z_2	-0.127	.	-0.099	-0.129
	$RMSE$		.	(0.016)	(0.010)			.	(0.029)	(0.017)
	CP		.	(0.361)	(0.944)			.	(0.579)	(0.969)
			Change in $Pr(Y=2)$					Change in $Pr(Y=3)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	$Mean$	-0.090	0.025	-0.084	-0.090	$x_1 \equiv z_1$	0.040	-0.064	-0.086	0.040
	$RMSE$		(0.115)	(0.015)	(0.015)			(0.105)	(0.126)	(0.026)
	CP		(0.000)	(0.919)	(0.956)			(0.000)	(0.000)	(0.924)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	$Mean$	-0.170	.	-0.203	-0.163	z_2	0.337	.	0.325	0.332
	$RMSE$		.	(0.035)	(0.026)			.	(0.031)	(0.030)
	CP		.	(0.770)	(0.927)			.	(0.915)	(0.935)

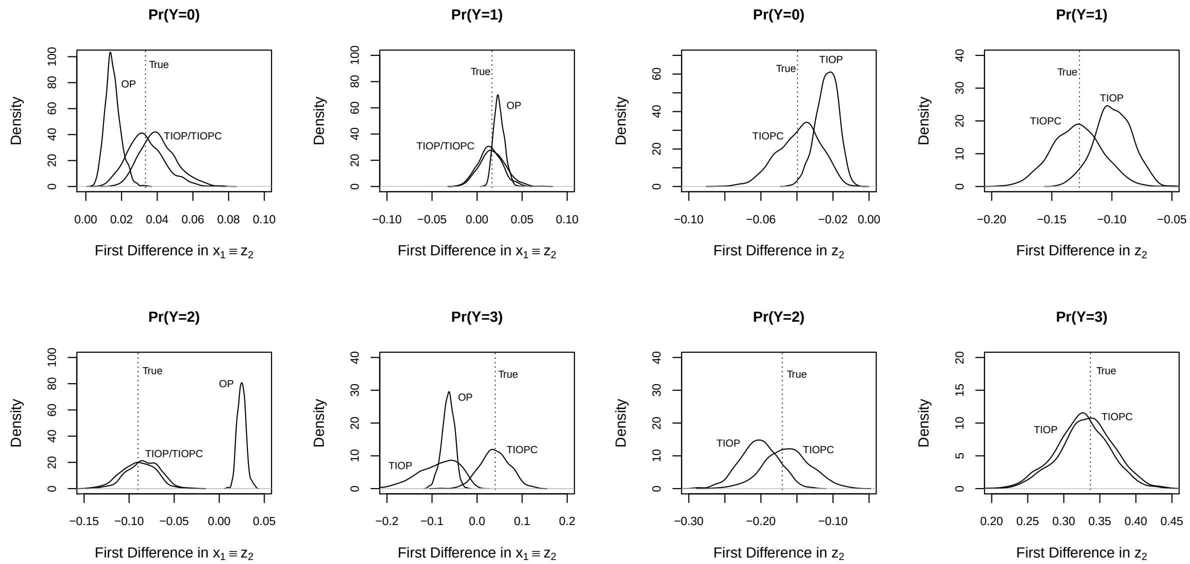
	% Sims Preferring TiOP to OP	% Sims Preferring TiOPC to OP	% Sims Preferring TiOPC to TiOP
AIC	100	100	82
BIC	100	100	42
LR Test	100	100	67
ρ Test	.	.	80

model selection criteria.

Turning to our first difference estimates, we find very favorable results for the TiOPC and TiOP models in terms of both accuracy and empirical coverage in Experiment 5. With regards to overall accuracy, and with the exception of our first difference estimates pertaining to the $Pr(Y = 1)$ outcome, we find for example that our TiOPC models' RMSEs are 2-10 times smaller (i.e., superior to) those of the OP model for $x_1 \equiv z_1$. The same can often be said for the TiOP model's RMSE's, though in this case the OP model offers modestly smaller RMSEs in relation to the TiOP model's RMSEs for two OP outcome categories: $Pr(Y = 1)$ and $Pr(Y = 4)$. For z_2, we find in Table A.12 that the TiOPC model in this case recovers more accurate first differences than its TiOP counterpart in every instance. These observations are reinforced by Figure A.19, wherein we find that the TiOPC model's first differences typically fall closer to the true first difference value, in comparison to the TiOP and OP first differences—and in some cases (e.g., for the effects of $x_1 \equiv z_1$ upon our

Figure A.19: Distributions of First Differences Across All Simulations, Experiment 5

(a) First Differences for $x_1 \equiv z_1$ (b) First Differences for z_2



highest-most outcome category) markedly so. Hence, in terms of both accuracy and bias, the TiOPC clearly outperforms our comparison models under the moderate levels of TiOPC d.g.p.-correlation considered for Experiment 5.

With respect to empirical coverage, the 95% confidence intervals associated with the TiOPC model's estimated first differences (across both $x_1 \equiv z_1$ and z_2) recover our true first differences in 94.2% of all Experiment 5 simulations, on average. By comparison, the 95% confidence intervals associated with the TiOP model's estimated first differences recover these true first difference effects across an average of 67.7% of all simulations in Experiment 5. For the OP model's equivalent 95% CP's associated with $x_1 \equiv z_1$ in Experiment 5, we find that the 95% confidence intervals to the OP's estimated first differences include the true first difference values in only 24.0% of all simulations, on average—a far lower rate than what was obtained in the TiOP(C) cases for this particular experiment. Altogether, these results reaffirm those of Experiment 2 in suggesting that the TiOP(C) models are far superior to the OP model in terms of accuracy and coverage under conditions of a TiOPC d.g.p. with a moderately-sized ρ. That being said, the choice between the TiOP and TiOPC models in these instances is sensitive to one's choice of model selection criteria, and likely bolstered by the consideration of multiple criteria.

Experiment 6 reassesses the performance of our TiOP, TiOPC, and OP models within circumstances where a researcher is faced with a TiOPC-distributed dependent variable that exhibits a high level of correlation ($\rho = 0.75$) between its associated inflation and ordered outcome stages. The results are reported in Table A.13 and Figure A.20. We can begin by noting in Table A.13 that our four key model selection criteria each consistently favor the TiOP(C) models over the OP model across 100% of our simulations. These same criteria also now properly select the TiOPC model over the TiOP model at noteworthy rates of accuracy that in each case fall between 93%-to-100% of all simulations, depending on the model selection criteria considered. Hence, we can conclude that at high levels of correlation between one's inflation and outcome stages, our four model selection criteria each perform quite well in identifying the correct model as being a TiOPC model.

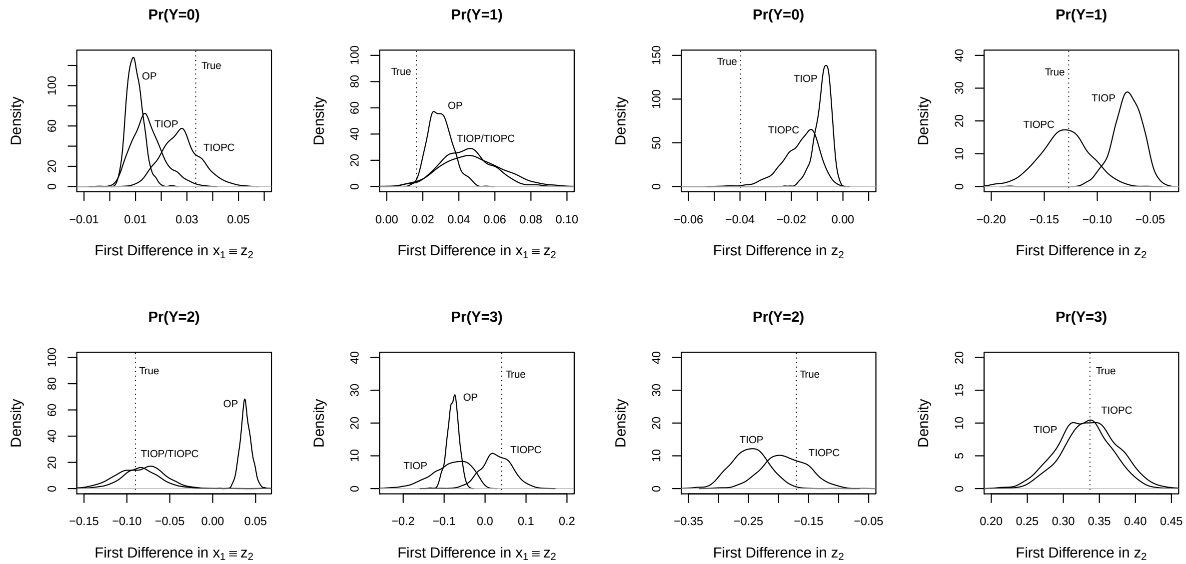
Table A.13: Marginal Effects For OP, TiOP, and TiOPC Models under a TiOPC DGP with 75% Correlation

			Change in $Pr(Y=0)$					Change in $Pr(Y=1)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	*Mean*	0.033	0.010	0.028	0.028	$x_1 \equiv z_1$	0.016	0.030	0.044	0.048
	RMSE		(0.024)	(0.008)	(0.019)			(0.013)	(0.028)	(0.031)
	CP		(0.003)	(0.881)	(0.296)			(0.592)	(0.478)	(0.641)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	*Mean*	-0.040	.	-0.008	-0.017	z_2	-0.127	.	-0.072	-0.131
	RMSE		.	(0.032)	(0.023)			.	(0.055)	(0.019)
	CP		.	(0.002)	(0.341)			.	(0.064)	(0.972)
			Change in $Pr(Y=2)$					Change in $Pr(Y=3)$		
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
$x_1 \equiv z_1$	*Mean*	-0.090	0.038	-0.078	-0.089	$x_1 \equiv z_1$	0.040	-0.078	-0.085	0.028
	RMSE		(0.129)	(0.021)	(0.021)			(0.119)	(0.125)	(0.031)
	CP		(0.000)	(0.879)	(0.942)			(0.000)	(0.001)	(0.890)
		True	OP	TiOP	TiOPC		True	OP	TiOP	TiOPC
z_2	*Mean*	-0.170	.	-0.248	-0.189	z_2	0.337	.	0.327	0.339
	RMSE		.	(0.078)	(0.034)			.	(0.031)	(0.030)
	CP		.	(0.266)	(0.918)			.	(0.937)	(0.945)

	% Sims Preferring TiOP to OP	% Sims Preferring TiOPC to OP	% Sims Preferring TiOPC to TiOP
AIC	100	100	100
BIC	100	100	93
LR Test	100	100	99
ρ Test	.	.	100

Figure A.20: Distributions of First Differences Across All Simulations, Experiment 6

(a) First Differences for $x_1 \equiv z_1$ (b) First Differences for z_2



Likewise, the TiOPC model consistently outperforms the OP and TiOP models in terms of both 95% empirical coverage and overall accuracy in Table A.13. Starting with the 95% empirical coverage probabilities in Table A.13, we find that our TiOPC model's estimated first differences' 95% confidence intervals recover our true first difference quantities of interest in 74% of all simulations, on average. By comparison, under conditions where one's TiOPC d.g.p. has an underlying ρ of 0.75, our 95% empirical CPs for the TiOP and OP models are on average 44% and 15% respectively—far lower than the aforementioned average CP value that was obtained for the TiOPC model under this particular experiment. Interpreting these CP findings alongside those for Experiments 4-5, we can conclude that the TiOPC model becomes increasingly optimal in relation to the TiOP and OP models in terms of empirical coverage when one's underlying TiOPC d.g.p. exhibits an increasingly high level of correlation in its inflation and ordered-outcome stages.

Turning next to the RMSEs reported in Table A.13, we find that our TiOPC model exhibits superior accuracy to the other two models considered across all estimated first differences and ordered outcome categories. The same can arguably also be said for TiOP model's RMSE values in comparison to the OP model's RMSEs. Unpacking these RMSE results in further detail, we can note that our Experiment 6 RMSEs consistently favor the TiOPC over the TiOP for each estimated first difference that is associated with z_2, but do not systematically favor either model in the case of our estimated first differences for $x_1 \equiv z_1$. However, for the latter first differences, the TiOP and TiOPC models' RMSEs generally outperform our OP model's RMSE for a majority of the first difference estimates considered—in several cases by a factor of 3-to-5.

Figure A.20 reinforces the findings discussed above. For instance—under Experiment 6's conditions of a high ρ—the TiOPC model's corresponding distributions of first differences generally encompass each true first difference effect more consistently than the TiOP model's comparable first difference distributions, which in turn tend to outperform those of the OP model. However, we can also note that all plotted distributions in Figures A.20a-A.20b appear to exhibit slightly higher bias than the distributions depicted for Experiments 4-5 (i.e., in comparison to our TiOPC d.g.p. experiments with $\rho = 0.25$ or $\rho = 0.5$). This indicates that—all else equal—an increasingly high ρ leads to heightened challenges in accurately recovering true parameters of interest, no matter whether one uses a TiOP or TiOPC model. This latter observation notwithstanding, we can conclude in this instance that for a TiOPC d.g.p. with a high level of correlation between one's inflation process and ordered outcome process, the TiOPC model is preferred over the TiOP—and especially over the OP model—in terms of both accuracy and empirical coverage.

References

Amnesty International. 2003. *Nepal Widespread "disappearances" in the context of armed conflict.* ASA 31/045/2003.

Amnesty International. 2011. *You Cannot Kill the Truth.* London, UK: Amnesty International.

Amnesty International. 2020. "Hosni Mubarak: A Living Legacy of Mass Torture and Arbitrary Detention." Available at: `https://www.amnesty.org/en/latest/news/2020/02/hosni-mubarak-legacy-of-mass-torture/`.

Amnesty International. 2022. "DEMOCRATIC REPUBLIC OF THE CONGO 2022." Available at: `https://www.amnesty.org/en/location/africa/east-africa-the-horn-and-great-lakes/democratic-republic-of-the-congo/report-democratic-republic-of-the-congo/`.

Australian Government DFAT. *a*. "Australia-Chile Free Trade Agreement." Available at: `https://www.dfat.gov.au/sites/default/files/Australia-Chile-Free-Trade-Agreement.pdf`.

Australian Government DFAT. *b*. "Australia-Chile FTA." Available at: `https://www.dfat.gov.au/trade/agreements/in-force/aclfta/Pages/australia-chile-fta`.

Baccini, Leonardo, Andreas Dür and Manfred Elsig. 2015. "The Politics of trade agreement design: revisiting the depth–flexibility Nexus." *International Studies Quarterly* 59(4):765–775.

Bagozzi, Benjamin E. and Bumba Mukherjee. 2012. "A Mixture Model for Middle Category Inflation in Ordered Survey Responses." *Political Analysis* 20:369–386.

Bagozzi, Benjamin E., Daniel W. Hill, Will H. Moore and Bumba Mukherjee. 2015. "Modeling Two Types of Peace: The Zero-inflated Ordered Probit (ZiOP) Model in Conflict Research." *Journal of Conflict Resolution* 59(4):728–752.

Baier, Scott L and Jeffrey H Bergstrand. 2004. "Economic determinants of free trade agreements." *Journal of International Economics* 64(1):29–63.

Baysan, Tercan, Arvind Panagariya and Nihal Pitigala. 2006. "Preferential Trading in South Asia." *World Bank Policy Research Working Paper* (3813).

Bhagwati, Jagdish. 2008. "Why PTAs Are a Pox on the World Trading System." *Termites in the Trading System* .

Bhagwati, Jagdish N, Pravin Krishna and Arvind Panagariya. 1999. *Trading blocs: Alternative approaches to analyzing preferential trade agreements.* Cambridge, MA: MIT Press.

Bobick, Talya and Alastair Smith. 2013. "The impact of leader turnover on the onset and the resolution of WTO disputes." *The Review of International Organizations* 8(4):423–445.

Brysk, Alison and Aashish Mehta. 2014. "Do rights at home boost rights abroad? Sexual equality and humanitarian foreign policy." *Journal of Peace Research* 51(1):97–110.

Carey, Sabine C, Michael P Colaresi and Neil J Mitchell. 2015. "Governments, informal links to militias, and accountability." *Journal of Conflict Resolution* 59(5):850–876.

Carter, Jeff, Michael Bernhard and Glenn Palmer. 2012. "Social revolution, the state, and war: How revolutions affect war-making capacity and interstate war outcomes." *Journal of Conflict Resolution* 56(3):439–466.

Chenoweth, Erica and Orion A Lewis. 2013*a*. "Nonviolent and Violent Campaigns and Outcomes (NAVCO) Data Project Version 2.0." Campaign-Year Data Codebook. University of Denver.

Chenoweth, Erica and Orion A Lewis. 2013*b*. "Unpacking nonviolent campaigns: Introducing the NAVCO 2.0 dataset." *Journal of Peace Research* 50(3):415–423.

Cingranelli, David and David Richards. 2004. *The Cingranelli-Richards (CIRI) Human Rights Dataset*.

Clark, David H and Patrick M Regan. 2003. "Opportunities to fight: A statistical technique for modeling unobservable phenomena." *Journal of Conflict Resolution* 47(1):94–115.

Dugan, Laura and Erica Chenoweth. 2012. "Moving beyond deterrence: The effectiveness of raising the expected utility of abstaining from terrorism in Israel." *American Sociological Review* 77(4):597–624.

EDIT. *a*. "Japan - Thailand EPA (2007)." Available at: `https://edit.wti.org/document/show/5db863d6-0758-4ba5-99bf-1fdac96e8112`.

EDIT. *b*. "Mexico - Northern Triangle FTA (2000)." Available at: `https://edit.wti.org/document/show/66d8409c-0249-43d9-9dc8-513b9075efa3`.

Escribà-Folch, Abel and Joseph Wright. 2015. *Foreign Pressure and the Politics of Autocratic Survival*. New York: Oxford University Press.

Estevadeordal, Antoni, Dani Rodrik, Alan M Taylor and Andres Velasco. 2004. *Integrating the Americas: FTAA and beyond*. Harvard University Press Cambridge, MA.

Estevadeordal, Antoni, Kati Suominen, Jeremy Harris and Matthew Shearer. 2009. *Bridging Regional Trade Agreements in the Americas: Special Report on Integration and Trade*. Inter-American Development Bank.

Feridhanusetyawan, Tubagus. 2005. "Preferential trade agreements in the Asia-Pacific Region.".

French, David. 2015. *Fighting EOKA: The British counter-insurgency campaign on Cyprus, 1955-1959*. Oxford University Press.

Galindo, Arturo, Eduardo Fernández-Arias, José Luis Machinea, Carolyn Robert, Suzanne Duryea, Marcos Sawaya Jank, Stephen Meardon, César Manuel Serra, Ernesto H Stein, José Ernesto López Córdova et al. 2002. "Beyond borders: the new regionalism in Latin America." *IDB Publications (Books)* .

Getmansky, Anna, Tolga Sınmazdemir and Thomas Zeitzoff. 2018. "Refugees, xenophobia, and domestic conflict: Evidence from a survey experiment in Turkey." *Journal of Peace Research* 55(4):491–507.

Gibney, Mark, Linda Cornett, Reed Wood, Peter Haschke, Daniel Arnon, Attilio Pisano, Gray Barrett and Baekkwan Park. 2022. "The political terror scale 1976–2021." Available at: http://www.politicalterrorscale.org/.

Girod, Desha M, Megan A Stewart and Meir R Walters. 2016. "Mass protests and the resource curse: The politics of demobilization in rentier autocracies." *Conflict Management and Peace Science* 35(5):503–522.

Guriev, Sergei and Daniel Treisman. 2019. "Informational autocrats." *Journal of Economic Perspectives* 33(4):100–127.

Hafner-Burton, Emilie M and Kiyoteru Tsutsui. 2005. "Human rights in a globalizing world: The paradox of empty promises." *American journal of sociology* 110(5):1373–1411.

Harbom, Lotta, Erik Melander and Peter Wallensteen. 2008. "Dyadic dimensions of armed conflict, 1946–2007." *Journal of peace research* 45(5):697–710.

Harris, Mark N. and Xueyan Zhao. 2007. "A Zero-Inflated Ordered Probit Model, with an Application to Modelling Tobacco Consumption." *Journal of Econometrics* 141:1073–1099.

Human Rights Watch. 2011. *Haiti's Rendezvous with History*. New York, NY: Human Rights Watch.

Human Rights Watch. 2020. "Q and A: The International Criminal Court and the United States." Available at: `https://www.hrw.org/news/2020/09/02/qa-international-criminal-court-and-united-states`.

Hyde, Susan D and Nikolay Marinov. 2014. "Information and self-enforcing democracy: The role of international election observation." *International Organization* 68(2):329–359.

IMF. 2015. "Trade Integration in Latin America and the Caribbean: Hype, Hope, and Reality." Regional Economic Outlook: Western Hemisphere, Chapter 4October. Washington, D.C.: International Monetary Fund.

Immigration and Refugee Board of Canada. 1991. "The Fall of President Ershad, 1 February 1991,." Available at: `https://www.refworld.org/docid/3ae6a8050.html` [accessed 5 December 2023].

International Criminal Court. 2013. "Decision on the Non-compliance of the Republic of Chad with the Cooperation Requests Issued by the Court Regarding the Arrest and Surrender of Omar Hassan Ahmad Al-Bashir." Available at: `https://www.icc-cpi.int/sites/default/files/CourtRecords/CR2013_02245.PDF`.

JUS Laws & Consult. . "JTEPA - Japan Thailand Economic Partnership Agreement." Available at: `https://www.juslaws.com/business-commercial/jtepa`.

Kalyvas, Stathis N. 2006. *The Logic of Violence in Civil War*. Cambridge, UK: Cambridge University Press.

Kawai, Masahiro and Ganeshan Wignaraja. 2009. "Multilateralizing regional trade arrangements in Asia." *Multilateralizing Regionalism: Challenges for the Global Trading System* pp. 495–553.

Kono, Daniel Y. 2006. "Optimal obfuscation: Democracy and trade policy transparency." *American Political Science Review* 100(3):369–384.

Kreutz, Joakim and Enzo Nussio. 2019. "Destroying trust in government: Effects of a broken pact among Colombian ex-combatants." *International Studies Quarterly* 63:1175–1188.

Kubota, Yuichi. 2017. "Explaining state violence in the Guatemalan civil war: Rebel threat and counterinsurgency." *Latin American Politics and Society* 59(3):48–71.

Kucik, Jeffrey and Eric Reinhardt. 2008. "Does flexibility promote cooperation? An application to the global trade regime." *International Organization* 62(3):477–505.

Lemke, Douglas and William Reed. 2001. "The relevance of politically relevant dyads." *Journal of Conflict Resolution* 45(1):126–144.

Lupu, Yonatan. 2015. "Legislative veto players and the effects of international human rights agreements." *American Journal of Political Science* 59(3):578–594.

Lyall, Jason. 2009. "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya." *Journal of Conflict Resolution* 53(3):331–362.

Maggi, Giovanni and Andres Rodriguez-Clare. 2007. "A political-economy theory of trade agreements." *American Economic Review* 97(4):1374–1406.

Mansfield, Edward D and Helen V Milner. 2012. *Votes, vetoes, and the political economy of international trade agreements*. Princeton, NJ: Princeton University Press.

Mansfield, Edward D and Helen V Milner. 2018. "The domestic politics of preferential trade agreements in hard times." *World Trade Review* 17(3):371–403.

Mukherjee, Bumba and David Andrew Singer. 2010. "International institutions and domestic compensation: The IMF and the politics of capital account liberalization." *American Journal of Political Science* 54(1):45–60.

Parliament of Australia. . "Chapter 3 Australia - Chile Free Trade Agreement." Available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Completed_Inquiries/jsct/4june2008/report1/chapter3.

Pereira, Guillermo. 2001. "Mexico-Central America Free Trade Agreement." *Law & Bus. Rev. Am.* 7:383.

Pevehouse, Jon C. 2004. "Interdependence theory and the measurement of international conflict." *The Journal of Politics* 66(1):247–266.

Piazza, James A and James Igoe Walsh. 2009. "Transnational terror and human rights." *International Studies Quarterly* 53(1):125–148.

Poe, Steven C, Sabine C Carey and Tanya C Vazquez. 2001. "How are these pictures different? A quantitative comparison of the US State Department and Amnesty International human rights reports, 1976-1995." *Hum. Rts. Q.* 23:650.

Political Economy Research Institute. N.d. "HEast Timor - Indonesia (1975 - 1999)." University of Massachusetts Amherst. Available at: https://peri.umass.edu/fileadmin/pdf/Easttimor.pdf.

Porges, Matthew and Christian Leuprecht. 2016. "The puzzle of nonviolence in Western Sahara." *Democracy and Security* 12(2):65–84.

Prusa, Thomas J. 2016. Antidumping provisions in preferential trade agreements. In *The World Trade System: Trends and Challenges*. Cambridge, MA: MIT Press pp. 117–148.

Raleigh, Clionadh. 2016. "Pragmatic and promiscuous: Explaining the rise of competitive political militias across Africa." *Journal of Conflict Resolution* 60(2):283–310.

Sántha, Hanga, Ylva Lennartsson Hartmann and Mark Klamberg. 2010. "Crimes against humanity in Western Sahara: The case against Morocco." *Juridisk Publikation* 2:175–199.

Schwartz, Rachel A and Scott Straus. 2018. "What drives violence against civilians in civil war? Evidence from Guatemala's conflict archives." *Journal of Peace Research* 55(2):222–235.

Senese, Paul D. 1997. "Between dispute and war: The effect of joint democracy on interstate conflict escalation." *The Journal of Politics* 59(1):1–27.

Simmons, Beth A. 2009. *Mobilizing for human rights: international law in domestic politics*. Cambridge University Press.

Simmons, Beth A and Lisa L Martin. 2002. International Organizations and Institutions. In *Handbook of international relations*, ed. Carlsnaes W, Risse T and Simmons BA. Thousand Oaks,CA: Sage Publications pp. 192–211.

Takriti, Danna. 2020. "Hosni Mubarak's death and despotic rule, briefly explained." Available at: https://www.vox.com/world/2020/2/25/21152870/hosni-mubarak-death-despotic-rule-explained.

Teh, Robert, Thomas J Prusa and Michele Budetta. 2009. Trade remedy provisions in regional trade. In *Regional rules in the global trading system*, ed. K. Suominen Estevadeordal, A. and R. Teh. Cambridge: Cambridge University Press.

Tomz, Michael R and Jessica LP Weeks. 2020. "Human rights and public support for war." *The Journal of Politics* 82(1):182–194.

UNESCAP. . "Asia-Pacific Trade Agreement (APTA)." Available at: https://www.unescap.org/apta/agreement-text.

United Nations General Assembly. 1999. "Human rights questions: human rights situations and reports of special rapporteurs and representatives." A/54/660. Fifty-fourth session, Agenda item 116(c).

United Nations LCD Portal. . "South Asia Free Trade Agreement (SAFTA)." Available at: https://www.un.org/ldcportal/content/south-asian-free-trade-area-safta.

University of Solo. . "Asia-Pacific Trade Agreement (APTA)." Available at: https://www.jus.uio.no/english/services/library/treaties/09/9-05/asia_pacific_trade_ag.html#treaty-header1-4.

Urata, Shujiro. 2007. "Japan's FTA Strategy and Free Trade Area of Asia Pacific (FTAAP)." Pacific Economic Cooperation Council. Available at: https://www.pecc.org/resources/trade-and-investment-1/2038-japans-fta-strategy-and-free-trade-area-of-asia-pacific-ftaap/file.

US Department of State. 1998. "Serbia-Montenegro Country Report on Human Rights Practices for 1998." Available at: https://1997-2001.state.gov/global/human_rights/1998_hrp_report/serbiamo.html.

Villareal, M. 2017. "Mexico's Free Trade Agreements." Congressional Research Service.

Weerakoon, Dushni. 2010. "Current Status and Prospects." *Promoting Economic Cooperation in South Asia* p. 71.

Wood, Reed M and Mark Gibney. 2010. "The Political Terror Scale (PTS): A re-introduction and a comparison to CIRI." *Hum. Rts. Q.* 32:367.

Xiang, Jun. 2010. "Relevance as a latent variable in dyadic analysis of conflict." *The Journal of Politics* 72(2):484–498.

Zunes, Stephen. 2018. "The Role of Civil Resistance in Bolivia's 1977-1982 Pro-Democracy Struggle." *Middle Atlantic Review of Latin American Studies* 2(1).