Digging the “Ethnic Violence in China” Database The Effects of Inter-Ethnic Inequality and Natural Resources Exploitation in Xinjiang

Author(s): Xun Cao, Haiyan Duan, Chuyu Liu, James A. Piazza and Yingjie Wei
Source: China Review, Vol. 18, No. 2, SPECIAL THEMED SECTION: Frontiers and Ethnic Groups in China (MAY 2018), pp. 121-154
Published by: Chinese University Press
Stable URL: http://www.jstor.org/stable/26435650
Accessed: 30-06-2018 16:08 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms
Digging the “Ethnic Violence in China” Database: The Effects of Inter-Ethnic Inequality and Natural Resources Exploitation in Xinjiang*

Xun Cao, Haiyan Duan, Chuyu Liu, James A. Piazza, and Yingjie Wei

Abstract

Unrest in the Xinjiang region of China currently poses the most imminent threat to the internal security of China and to central government control over peripheral regions. Instability in Xinjiang, furthermore, has ramifications for the wider security environment in Central Asia as the conflict becomes linked with jihadist groups in other security hotspots, like Pakistan and Syria. However, our understanding of important potential factors affecting political instability in Xinjiang is limited by the lack of systematically collected event data of ethnic

Xun Cao is Associate Professor of Political Science at the Pennsylvania State University, USA.

Haiyan Duan is Assistant Professor in the School of Political Science and Public Administration, East China University of Political Science and Law.

Chuyu Liu is a Ph.D. candidate in the Department of Political Science, Pennsylvania State University, USA.

James A. Piazza is Liberal Arts Professor of Political Science at the Pennsylvania State University, USA.

Yingjie Wei is Assistant Professor in the School of International and Public Affairs, Shanghai Jiao Tong University, China. For the event data and replication files for the empirical analysis in the paper, correspondence should be sent to weiyj@sjtu.edu.cn.

*We want to thank the journal reviewers and editors for their helpful comments to improve the article. All errors are ours.
violence. In this article, we introduce the first effort to fill this gap in data collection, that is, the Ethnic Violence in China (EVC) Database: the Xinjiang Region. This is a geocoded database of yearly incidents of ethnic violence at the county level in Xinjiang from 1990 to 2005. Using the EVC database, we demonstrate some initial results modeling ethnic violence in Xinjiang. We find that ethnic violence is positively associated with interethnic inequality; resources such as oil and cotton, on the other hand, are unrelated to the likelihood of ethnic violence.

China is a prominent country on the world stage, and the status of its domestic stability has regional and global implications. A key factor affecting the internal security—and the political stability—of China is persistent political violence in the Xinjiang Autonomous Region in western China. This resource-rich but underdeveloped province is home to the Muslim Uyghur ethnic minority group. In the past several decades, Xinjiang has seen a large in-migration of Han Chinese and has been the site of acute ethnic violence, often in the form of Uyghur riots and antigovernment terrorism perpetrated by Uyghur separatists.\footnote{Former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.} Both the Chinese government and international communities, furthermore, worry about growing transnational ties between Uyghur militants in Xinjiang and al-Qaeda-affiliated terrorists in South and Central Asia and the Middle East. During a recent visit to China, former US National Security Advisor Susan Rice alleged that Chinese Uyghurs from Xinjiang had traveled to Iraq and Syria to fight in the ranks of the Islamic State (ISIS) terrorist insurgency.\footnote{In a 2004 speech, former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.\footnote{Former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.} The Chinese central government has implemented various policies to address violent unrest in Xinjiang, ranging from traditional policing and counterterrorism tactics such as those featured in the various “Strike Hard” campaigns to a recent move to send 200,000 civil servants to villages in Xinjiang to better provide local public goods. These strategies, however, have not successfully quelled the unrest or dampened tensions between ethnic minorities and the government.}

Unrest in this region currently poses the most imminent threat to internal security and to Chinese central government control over peripheral regions.\footnote{Former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.\footnote{Former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.} Though no published study to date has provided comprehensive figures, we recorded 213 ethnic violent events between 1990 and 2005 in the Ethnic Violence in China (EVC) database. In a 2004 speech, former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.\footnote{Former Chinese President Hu Jintao identified Uyghur “separatism, extremism and terrorism” as the paramount domestic security question facing contemporary China.} The Chinese central government has implemented various policies to address violent unrest in Xinjiang, ranging from traditional policing and counterterrorism tactics such as those featured in the various “Strike Hard” campaigns to a recent move to send 200,000 civil servants to villages in Xinjiang to better provide local public goods. These strategies, however, have not successfully quelled the unrest or dampened tensions between ethnic minorities and the government.}
The conflict in Xinjiang has security implications globally. Jihadist groups in South Asia and the Middle East, including the Iraq- and Syria-based ISIS movement, have issued statements in support of Xinjiang’s independence from China and integration into a global Islamic Caliphate, while Uyghur militants from China and Central Asia have joined and fought within a host of terrorist movements such as the Taliban and the al-Qaeda core. The East Turkestan Islamic Movement (ETIM), based in Pakistan, appears on the US State Department’s Foreign Terrorist Organizations list and is believed to have been implicated in attacks within China and in third countries. Understanding what drives ethnic violence in Xinjiang helps depict the global jihadist picture.

With regard to contributions to academic literatures on ethnic violence and civil conflicts, Xinjiang is an ideal case that allows us to examine what we suspect is a key process in fostering an environment in which political instability and ethnic conflict are likely to occur in a large number of developing countries: liberal economic reform coupled with continued authoritarian rule. Though China embarked on free market economic reforms starting in the early 1980s and has applied those policies in Xinjiang as an economic development strategy and lure for Han migration to the region, it has engaged in little political reform. This situation of liberal economic policies with illiberal rule in Xinjiang can be identified in countries in the Middle East (e.g., Jordan and Gulf states like Qatar and Bahrain) and in some Southeast Asian countries like Singapore and Vietnam. Studying violence and conflict in Xinjiang helps us to better understand civil order in such pluralistic societies.

The literature on conflict in Xinjiang speculates on several causal factors for ethnic violence in the region: economic inequality between ethnic minority groups and Han Chinese; a legacy of suppression of local cultural and religious life by Chinese authorities; political repression and underrepresentation of ethnic minority groups in provincial affairs and political life; excessive use of force by government agents during policing and counterterrorism efforts; and growing Islamist sentiment, some of which is fueled by radical Muslim movements in the neighboring Central Asian republics and in Pakistan and Afghanistan. However, scholarly and expert understanding of the contours and causes of ethnic violence in Xinjiang has been limited by one key factor: existing work on Xinjiang is often qualitative and has not systematically evaluated patterns and precipitants of violence using data and other forms of analysis, leaving us
with little understanding of how and why ethnic violence occurs and what can be done about it.\(^9\)

In this article, we introduce the first effort to fill this gap in data collection—the EVC Database: the Xinjiang Region, 1990–2005. This is a geocoded database of yearly incidents of ethnic violence (e.g., riots, terrorist incidents, acts of interethnic violence) occurring at the county level in Xinjiang from 1990 to 2005. The specific collection and coding technique is described in detail in the next section. Incidents of ethnic violence, such as riots, and political violence, such as acts of terrorism credited to extremists, are included in this event data set as well as information such as location and number of casualties, to measure the “intensity” of the event.\(^{10}\) Moreover, using the EVC event data, we demonstrate some initial results modeling ethnic violence in Xinjiang as a function of factors such as horizontal inequality between the Han majority and ethnic minorities, resources such as oil and cotton, income per capita, population density, local ethnic composition, and government fiscal spending. We find that ethnic violence is positively associated with horizontal inequality; resources such as oil and cotton, however, are unrelated to the likelihood of ethnic violence.

1. Collection and Coding Ethnic Violence for Xinjiang, China

In this section we discuss the construction of the EVC database. Our data collection and analysis is focused on the county-year level: Xinjiang has 15 prefectures that are further divided into 100 counties. We have coded violent events at the county level from 1990 to 2005. Various types of events are included such as bombing, assassination, riot, arson, and armed attacks. We also have coded variables on the time, location, perpetrators and targets, casualties, and event types for each event and also include detailed notes on the event.

For the purposes of the study, we consider ethnic violence to be the intentional execution of violent acts, perpetrated by individuals or groups of ethnic minority status, with political motivations. Acts of ethnic violence include but are not limited to (1) acts of political terrorism—for these we use the operational definition of terrorism provided by the Global Terrorism Database (GTD) hosted by the University of Maryland START center; (2) acts of political insurgency—for these we use the operational definition provided by the Peace Research Institute Oslo (PRIO); (3) riots; (4) violent street demonstrations and protests; and (5) assassinations.
Note that this conceptualization associates intentionality with the committing of the act itself, irrespective of the outcome it produces. We accordingly define violent attempts also as violence in our database. For instance, bombing and assassination attempts that were prevented by the police, often via violent means (e.g., arrests followed by siege and gun fights), are counted as violent events. Moreover, because our focus is on predictors of violence rooted in the Xinjiang crisis, we only include violent events that at least involved one ethnic minority. In other words, we do not include violent events that occur between members of the Han majority or events occurring between Han civilians and the government.

a. Variables in the EVC Xinjiang Database

The EVC database is an event data set. We first include variables that measure the timing and duration of an event. *Year original* is the exact description of time from the original document. Second, we use three variables, day, month, and year, to further describe the date(s)/month(s)/year(s) of an event onset, if available. For example, a value of 1 represents the first day of a month/January for the day/month variable; 1990 represents year 1990. Finally, the variable *duration* counts the number of days that an event lasted.

We also coded variables measuring the location of an event. Most events do not have location information below the county level; therefore, the most disaggregated spatial precision for our event data is county. The variable *location* is the name(s) of the place(s) where an event occurred from the original source: note that there is only one report that mentioned multiple social unrest events in multiple counties, beginning since 5 March 1992 and lasting 4 days; the five counties are Khotan, Kashgar, Quca, Cocak, and Bortala. County is the name(s) of the county(ies) where an event occurred. County precision describes the level of precision when we coded the event at the county level. There are two values for this variable: “confirmed”—either the original document provides county name(s) or we are able to locate county name(s) by our archival or Internet search; “missing”—the original document does not provide county name(s) and we are unable to locate county name(s) by our archival or Internet search.

Among the total 213 events coded between 1990 and 2005, there are 35 events for which despite our best efforts, the county location is
missing. For instance, Bovingdon reports that on 8 November 1999 that “Turghun Aqniyaz, a policeman, was killed by violent terrorists on his way home.”\textsuperscript{14} We only know that this event happened in the Hetian prefecture, and we are unable to find information on the exact county location of the event.\textsuperscript{15} Finally, the prefecture variable gives the name(s) of the prefecture(s) where an event occurred.

Furthermore, we provide variables on perpetrators and targets of an event. Perpetrator provides information on the perpetrator of an event—Uyghur, others (i.e., members of non-Han and non-Uyghur ethnic groups), mixed (i.e., Uyghur and other non-Han ethnic groups), and missing (i.e., whether information regarding the perpetrator side is missing). Similarly, the target variable depicts the nature of target of an event following the same coding rule: Uyghur, Han, others, mixed, government, and missing. Finally, target govt shows whether government agencies (e.g., police station) were directly targeted: we identify three categories—government, civilians, and government and civilians.

We also include variables that describe event types. Type includes events types such as bombing, assassination, riot, arson, gunfight, and attack on local governments. Death is the number of people killed in the event, if available; injury is the number of people injured in the event, if available; number involved is the number of people involved in the event, if available.

Finally, the variables source1, source2, source3, source4, and source5 list the data sources: 58 events were reported in more than one source. Note that different sources were used to collect event data. This is partly motivated by the advice to address the common concern of reporting bias when it comes to event data collection, which is to triangulate different sources.\textsuperscript{16} These sources can be categorized into three broad categories: (1) government documents (Xinjiang Public Security Gazette, government white papers, and county gazettes of Xinjiang), (2) online event data sets and news search engines (GTD, Minorities at Risk [MAR], and WiseNews), and (3) secondary data from existing scholarly work.\textsuperscript{17}

We acknowledge the limitation of gathering information on ethnic violence in Xinjiang and in China in general. For instance, local officials and state media might seek to hide these incidents from public view.\textsuperscript{18} While social media have increased our access to these incidents in the recent years, there is no certainty in our ability to know whether our event data has captured most, if not all, incidents that occurred during 1990–2005. However, even though it is impossible to completely rule out
media bias, the fact that in addition to media reports, we also heavily rely on government documents—some were written for circulation within the government (e.g., government white papers)—helps to address the concern of potential reporting biases.

In addition to the aforementioned variables, for each event, we also include a note, which is a detailed description on an event from the original source(s). We provide this so that any researcher can check our coding of variables and also draw information that is not included in our coded variables. For example, one note includes an event on 4 August 2008 reported by both WiseNews and GTD:

On Monday, at an unknown time, in Kashgar, Xinjiang, China, two unidentified assailants drove a truck into a group of policemen jogging near the barracks. The attack was claimed by the Eastern Turkistan Islamic Movement. There were no reported casualties in this incident.

b. Data Collected, 1990–2005

We have finished coding ethnic violent events between 1990 and 2005 in Xinjiang. According to our data, there were 213 ethnic violent events in the region. Figure 1 shows the spatial distribution of the cumulated number of events between 1990 and 2005 for each county, with a darker shading indicating a greater number of violent events. The Yecheng county of the Kashgar prefecture (the darkest polygon, southwest of the region) experienced 17 ethnic violent events during this 16-year period, closely followed by the Kashgar county of the same prefecture (14 events) and Yining county of the Ili prefecture (14 events). On the other hand, a number of counties to the east and southeast of Xinjiang, such as Yiwu, Barkol Kazakh, Ruoqiang, and Qiemo, experienced no violence at all. Figure 2 further displays the temporal distributions of violent events. The top panel shows the provincial annual sums; the highest number of events was in 1997. The bottom panel presents the patterns at the prefecture level. Violent events tend to be geographically concentrated in four prefectures: Akesu, Hetian, Kashgar, and Yili.
Figure 1: Distribution of Ethnic Violent Events in Xinjiang, 1990–2005.

Figure 2: Annual Violent Events in Xinjiang, Provincial and Prefectural Trends.

(a) Provincial Trends
2. An Application of the EVC Database

In this section, we demonstrate one application of the EVC database by modeling ethnic violence at the county-year level in Xinjiang. We believe that such a unit of analysis is crucial because we expect local factors to be of vital importance in determining patterns of violent events in the region. Regarding the dependent variable, we create a binary variable that takes the value of 1 in county-years where one or more ethnic violent events occurred, and 0 otherwise. This binarization process loses some information because there are county-years that experienced more than one event. However, multiple counts within one county-year are likely to be realizations of the same conflict process. Because of missing values from the right-hand-side variables, our regression analysis covers 1996–2005.

We have collected many potential predictor variables. This enables us to run logistic regressions to evaluate factors that might affect chances of ethnic violence. However, the exercise here is meant to demonstrate what we can learn by using the EVC event data rather than exhausting all choices of potential explanatory variables. We focus our theoretical
discussion on the effects of interethnic inequalities and natural resources. We control for the effects of many relevant local factors on violence such as mosque density, local government public spending, local state capacity, and ethnic fractionalization and polarization.

a. Interethnic Inequalities

Interethnic inequalities are often considered a salient factor driving grievances not only in Xinjiang, but also in other parts of the world. Indeed, the grievances thesis has a long tradition and posits that civil conflicts are often associated with ethnic or religious hatred, political repression, political exclusion, and economic inequalities. Conflict therefore is explained as a result of widely felt grievances among the relatively disadvantaged in the society. Inequality has often been the focus of this tradition including theories of ethnic conflict and structural inequality and relative deprivation. Many studies of inequality and conflict indeed relate to the theory of relative deprivation, which posits that while absolute poverty may lead to apathy and inactivity, comparisons with those in the same society who do better may inspire radical action and even violence.

Recent studies have pointed out the fact that to properly test the grievances argument, one needs inequality measures based on differences between collective actors. Moreover, given the fact that armed conflicts and ethnic violence are often local events, it is also important to measure inequalities at the subnational level. Part of the ongoing effort in the study of civil conflicts and local violence is to come up with better measures using geocoded data newly available to researchers such as Demographic and Health Surveys (DHS), spatial distribution of ethnic groups and GDP data, and census data.

In the case of Xinjiang, socioeconomic inequality between the ethnic minority groups and the Han populations is perhaps the most frequently mentioned source of political violence in the region by scholars. For instance, economic, social, health and educational outcomes for Uyghurs and Han in Xinjiang exhibit stark differences. Schuster finds that the infant mortality rate for Han Chinese is 13 per 1,000, but 102 per 1,000 for Uyghurs. According to Hasmath, relative to minorities, Han Chinese have a tendency to settle in wealthier urban areas, where Uyghurs tend to constitute the majority in rural areas or the poorer urban areas of southern Xinjiang. Cao demonstrates that Xinjiang experienced an increasing urban-rural income disparity in the 1990s.
that are characterized by the worst urban-rural income gaps are overwhelmingly concentrated in southwest Xinjiang (26 of 27 counties), where ethnic minorities usually account for more than 90 percent of the total population. In addition, these counties lag far behind the average level of per capita GDP in Xinjiang. In sum, we expect,

Hypothesis 1: Counties in Xinjiang with higher levels of interethnic inequality experience higher levels of ethnic violence.

b. Natural Resources

Xinjiang is China’s most crucial energy producing region. Around 60 percent of the gross regional product is produced by oil and gas extraction and ancillary industries, and the Chinese government has made large-scale investments in pipelines connecting Xinjiang with coastal cities like Shanghai. Though oil and gas are Xinjiang’s premier heavy industries, the central government has also promoted industrial-scale cotton production. Many believe that the significant orientation of the regional economy toward natural resource production has significantly contributed to conflict in Xinjiang. This is due to several reasons. First, the oil, gas, and cotton boom in Xinjiang has attracted large numbers of Han migrants, exacerbating ethnic conflict. Second, the majority Uyghur population of Xinjiang expects local resource revenues to be spent on Uyghur priorities. Large-scale investment in oil, coal, and other natural resource extraction by Han-dominated government firms is perceived as unjustly exploiting the wealth of the local, non-Han population, fostering sharp resentments. The benefits of government investment in the oil and gas industries have also been more heavily reaped in urban areas where Uyghurs are less dominant. Finally, natural resource exploitation has generated environmental costs that sharpen conflict. Only 4.3 percent of Xinjiang’s territory is inhabitable. The in-migration of Han, governmental agricultural policies, and oil and gas extraction have put a significant strain on water and arable land resources. The combination of the migration-fueled population boom and the despoliation of the environment has served to deepen grievances in the region.

These conclusions are often supported in the wider literature linking resource wealth and resource extraction with increased risks of internal political conflict and violence, including recent empirical work showing that when coupled with ethnic minority grievances and political
disenfranchisement, oil greatly increases the chances of armed conflict.\textsuperscript{36} Indeed, the natural resource–conflict nexus is one of the three branches of the resource curse literature.\textsuperscript{37} In theory, there are multiple causal mechanisms connecting natural resources to chances of violence. For instance, some argue that rebels from an ethnically marginalized region could be motivated by the prospect of establishing an independent state, so that local resource revenues would not be shared with the rest of the country.\textsuperscript{38} In sum, this leads us to expect,

Hypothesis 2: Counties in Xinjiang with more natural resource wealth experience higher levels of ethnic violence.

c. Coding Horizontal Inequality and Resources

Educational attainment is widely used to measure interethnic inequality. For example, Fjelde and Østby use education years provided by the DHS to calculate the relative wealth of the poorest and richest group at the subnational level;\textsuperscript{39} using census data, Barron, Kaiser, and Pradhan construct a district-level horizontal inequality index based on the ratio of group-level average education indictors.\textsuperscript{40} We use education attainments rather than direct measures on wealth also because the Chinese census data that we use do not report information on wealth.

To measure the horizontal inequality, we use the 1 percent sample of the Chinese National Population Census of 1990.\textsuperscript{41} It is reasonable to use horizontal inequality measures from 1990 to explain conflicts in 1996–2005.\textsuperscript{42} First, since we measure the value of inequality in 1990, it makes this indicator exogenous to the following conflicts between 1996 and 2005. Second, there is a widely shared assumption that the temporary changes in horizontal inequalities are relatively slow.\textsuperscript{43} The 1 percent sample of 1990 census allows us to construct prefecture level inequality measures along ethnicity lines using information on individual-level educational attainments: the census data do not report county information, so we have to measure horizontal inequality at the higher, prefecture, level.\textsuperscript{44}

We limit our sample to adults who were above 18 years old in 1990. There are five categories of educational attainments: less than primary completed, primary completed, lower secondary school completed, high school completed, and university completed. Our horizontal inequality measure is the prefecture-level difference between Han and the largest minority group in terms of the percentage of individuals who at least
completed the lower secondary school. The choice of lower secondary school completion as the threshold for education attainment fits the Chinese context. Despite China’s “nine years” compulsory education system that requires a student to finish lower secondary school, the implementation of this national policy has been far from being ideal, especially in periphery regions like Xinjiang. Therefore, measures using lower secondary school completion as the threshold for education attainment better serve as proxies for wealth and income.

For the choices of ethnic minority groups, we compare the Han majority to the largest ethnic minority group in a prefecture. Despite the media’s focus on the Uyghur population when it comes to violence in Xinjiang, ethnic violent events are not limited to this ethnic group. Among the 213 events between 1990 and 2005, there are 166 events (77 percent) that we are certain did involve the Uyghur ethnic group. Moreover, in 5 among 15 prefectures in Xinjiang, the largest ethnic minority group is not Uyghur. In prefectures such as Tacheng, Changji, Shihezi, and Aletai, Uyghurs are less than 5 percent of the local population. If we want to understand all ethnic violence, we should focus on the differences between Han and the largest ethnic minority group.

Many studies have found a robust association between natural resources, especially oil wealth, and the prevalence of civil violence. The redistribution of natural resource bounties often exacerbates the grievances of local minorities since they are usually excluded from sharing these benefits. We have collected county-level data regarding the geographic distribution of oil fields in Xinjiang. These data are taken from the General Chronicles of Xinjiang: Oil Industry (1999). We code this variable as 1 when there was at least one oil field in a county-year and 0 otherwise. Finally, in addition to oil wealth, cotton is of great importance to the local economy in Xinjiang. It is also an economic sector where we see high level of wealth extraction by the government. Therefore, we include a variable for county-year level cotton production (in tons) per capita.

d. Control Variables

Violence is more likely to happen in areas with higher population intensity, because population pressure exacerbates resource scarcity and worsens intergroup competition. Xinjiang 50 Years provides detailed annual statistics on county-level socioeconomic variables, including population density data. Chances of ethnic conflict also depend on the
size of local minority groups because large groups often are equipped with more resources for mobilization. We therefore include the county-year-level largest ethnic minority group size, measured as a percentage of total county population. Past studies have shown that income matters greatly for conflicts. For example, Buhaug et al. find strong evidence for a negative correlation between per capita income and political violence. To control for the impact of low income and poverty, we include county-year GDP per capita.

To control for ethnic fractionalization and polarization, we use county ethnic composition data from the *Statistical Yearbook of Xinjiang 1990–2005*. We construct the fractionalization index following Fearon and Laitin: \[ \text{fractionalization} = \sum_{i=1}^{N} \pi_i (1 - \pi_i); \] and polarization indicator following Montalvo and Reynal-Querol: \[ \text{polarization} = 4 \sum_{i=1}^{N} \pi_i^2 (1 - \pi_i). \] In both cases, \( \pi_i \) is the percentage of people who belong to ethnic group \( i \) in a given county-year, and \( N \) is the number of ethnic groups in that county-year. Intuitively, ethnic fractionalization captures the probability that two randomly selected individuals do not belong to the same group. The polarization index, on the other hand, aims to capture how far the distribution of ethnic groups is from the bipolar distribution, which represents the highest level of polarization.

Recent studies on religious institutions reveal their potential pacifying effects. For instance, religious institutions often assist local population by proving public goods. Public goods and emergency aid provided by local religious institutions help to address grievances and to prevent desperate citizens from using extreme actions against the government and other ethnic groups. Moreover, local public goods can improve living conditions of the population. We collect data on the numbers of mosques for each county. We use mosque density, standardizing the number of mosques by the 1990 county-level non-Han population. The mosque data are from 81 Xinjiang official county gazettes published between the late 1980s and the early 2000s. Since our regression analysis covers 1996–2005, we use mosque density of 1996 when the number of mosques in 1996 is available; for counties without 1996 mosque counts, we use the value from the most recent year between 1985 and 1995. This strategy allows us to minimize missing data: if we use only the numbers of mosques in 1996, we would lose a large number of observations.

We also include a number of geographic variables: the distance to the provincial capital, the distance to the prefectural capital, and a dummy variable indicates whether the county is a border county. These variables
often serve as proxies for state capacity. For instance, the reach of the state declines as we move away from the center and into peripheral regions. Given the large size of Xinjiang, the distance to the provincial capital, Ürümqi, may not fully capture state reach: this is the reason why we also include the distance to the prefectural capital. Furthermore, the Xinjiang Production and Construction Corps (XPCC) is a unique organization that combines functions of government, military and production. It has a hierarchical structure composed of 14 divisions and 175 regiments and these administrative domains of XPCC spread across the whole area of Xinjiang. XPCC can serve as an instrument of government control in the region. We include a dummy variable to indicate whether there was a XPCC administrative unit in a given county: note this is a time-invariant variable.

National border has been identified as a determinant of civil conflict. State boundaries can offer porous exits for insurgents to find sanctuaries in neighboring countries. Moreover, in the case of Xinjiang, border counties are often more exposed to the transnational diffusion of radical Islam movement that overwhelmed many central Asian states such as Afghanistan and Pakistan. We therefore create a dummy variable for border counties.

To further control for the roles of local governments, we also calculate government expenditure as a percentage of GDP. Finally, spatial dependence between the units of observations often poses a challenge to the analysis of conflict. For a binary dependent variable, it is often computationally challenging to address this issue. Following Pierskalla and Hollenbach, we construct a temporally lagged spatial lag of the dependent variable by dividing the number of neighboring counties with violent events at year $t-1$ by the total number of neighboring counties. To account for temporal dependence, we follow Carter and Signorino and add cubic polynomial approximation ($t$, $t^2$, $t^3$): $t$ is the number of years since last violent event. We have also included year fixed effects in all model specifications to take into account common exogenous shocks. We use logit models and all standard errors are clustered by county. To weaken the problem of reverse causality, all time-variant independent variables are lagged by one year; this does not affect the time-invariant variables included in the analysis though (distance to the prefecture’s capital, distance to Ürümqi, border county, mosques per 1,000 non-Han population, XPCC, and horizontal inequality). Descriptive and correlation statistics are presented in Tables 1 and 2.
Table 1: Descriptive Statistics.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflict event</td>
<td>957</td>
<td>0.090</td>
<td>0.286</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Spatial lag</td>
<td>979</td>
<td>0.084</td>
<td>0.162</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Largest minority group percentage</td>
<td>940</td>
<td>49.612</td>
<td>34.655</td>
<td>0.150</td>
<td>99.453</td>
</tr>
<tr>
<td>GDP per capita</td>
<td>848</td>
<td>0.309</td>
<td>0.425</td>
<td>0.052</td>
<td>5.780</td>
</tr>
<tr>
<td>Distance to the prefecture’s capital</td>
<td>1,001</td>
<td>96.593</td>
<td>93.590</td>
<td>0</td>
<td>406</td>
</tr>
<tr>
<td>Distance to Ürümqi</td>
<td>1,012</td>
<td>540.826</td>
<td>338.210</td>
<td>0</td>
<td>1,241</td>
</tr>
<tr>
<td>Border county</td>
<td>1,122</td>
<td>0.304</td>
<td>0.460</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Population density</td>
<td>932</td>
<td>588.821</td>
<td>2,156.332</td>
<td>1.339</td>
<td>26,000</td>
</tr>
<tr>
<td>Ethnic fractionalization</td>
<td>951</td>
<td>0.388</td>
<td>0.165</td>
<td>0.011</td>
<td>0.754</td>
</tr>
<tr>
<td>Ethnic polarization</td>
<td>951</td>
<td>0.550</td>
<td>0.213</td>
<td>0.022</td>
<td>0.970</td>
</tr>
<tr>
<td>Mosques per 1,000 non-Han</td>
<td>712</td>
<td>1.635</td>
<td>1.116</td>
<td>0.128</td>
<td>5.398</td>
</tr>
<tr>
<td>XPCC</td>
<td>1,122</td>
<td>0.098</td>
<td>0.298</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Horizontal inequality</td>
<td>1,056</td>
<td>29.871</td>
<td>21.330</td>
<td>−8.659</td>
<td>62.850</td>
</tr>
<tr>
<td>Oil</td>
<td>1,122</td>
<td>0.137</td>
<td>0.344</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Cotton production per capita</td>
<td>804</td>
<td>0.063</td>
<td>0.100</td>
<td>0.000</td>
<td>0.869</td>
</tr>
<tr>
<td>Government expenditure /GDP</td>
<td>850</td>
<td>0.184</td>
<td>0.149</td>
<td>0.020</td>
<td>1.051</td>
</tr>
</tbody>
</table>

Note: XPCC = Xinjiang Production and Construction Corps.

Table 2: Correlation Statistics.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conflict event</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Spatial lag</td>
<td>.28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Minority group percentage</td>
<td>.08</td>
<td>.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. GDP per capita</td>
<td>−.06</td>
<td>−.15</td>
<td>−.28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Distance to the prefecture’s capital</td>
<td>−.09</td>
<td>−.03</td>
<td>.02</td>
<td>−.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Distance to Ürümqi</td>
<td>.09</td>
<td>.12</td>
<td>.63</td>
<td>−.35</td>
<td>.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Border county</td>
<td>−.10</td>
<td>−.10</td>
<td>−.18</td>
<td>−.07</td>
<td>.00</td>
<td>.15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Population density</td>
<td>.20</td>
<td>.00</td>
<td>.09</td>
<td>−.04</td>
<td>−.17</td>
<td>.19</td>
<td>−.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Ethnic fractionalization</td>
<td>−.06</td>
<td>−.08</td>
<td>−.64</td>
<td>.15</td>
<td>−.11</td>
<td>−.49</td>
<td>.21</td>
<td>−.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Ethnic polarization</td>
<td>−.03</td>
<td>−.03</td>
<td>−.46</td>
<td>.18</td>
<td>−.05</td>
<td>−.41</td>
<td>.14</td>
<td>.00</td>
<td>.79</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Mosques per 1,000 non-Han</td>
<td>.03</td>
<td>.09</td>
<td>.75</td>
<td>−.30</td>
<td>−.11</td>
<td>.77</td>
<td>−.10</td>
<td>−.06</td>
<td>−.50</td>
<td>−.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. XPCC</td>
<td>−.06</td>
<td>−.04</td>
<td>−.20</td>
<td>.00</td>
<td>−.23</td>
<td>−.16</td>
<td>.00</td>
<td>.02</td>
<td>−.06</td>
<td>−.04</td>
<td>−.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Horizontal inequality</td>
<td>.14</td>
<td>.19</td>
<td>.40</td>
<td>−.31</td>
<td>.04</td>
<td>.57</td>
<td>−.20</td>
<td>.16</td>
<td>−.49</td>
<td>−.43</td>
<td>.54</td>
<td>−.19</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Oil</td>
<td>−.04</td>
<td>−.01</td>
<td>.07</td>
<td>.18</td>
<td>.22</td>
<td>−.15</td>
<td>−.08</td>
<td>−.08</td>
<td>−.04</td>
<td>.01</td>
<td>−.07</td>
<td>−.13</td>
<td>.10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>15. Cotton production per capita</td>
<td>−.01</td>
<td>.00</td>
<td>.03</td>
<td>.10</td>
<td>.19</td>
<td>−.11</td>
<td>−.31</td>
<td>−.11</td>
<td>−.10</td>
<td>.06</td>
<td>.04</td>
<td>.02</td>
<td>.07</td>
<td>.13</td>
<td>1</td>
</tr>
<tr>
<td>16. Government expenditure /GDP</td>
<td>−.12</td>
<td>−.07</td>
<td>.18</td>
<td>−.29</td>
<td>.33</td>
<td>.49</td>
<td>.22</td>
<td>−.10</td>
<td>−.19</td>
<td>−.23</td>
<td>.34</td>
<td>−.19</td>
<td>.22</td>
<td>−.10</td>
<td>−.18</td>
</tr>
</tbody>
</table>

Note: XPCC = Xinjiang Production and Construction Corps.
e. Empirical Results

Table 3 presents the empirical results. Our theoretical focus is on the effects associated with interethnic education differences and natural resources. Across all four model specifications, horizontal inequality, which is constructed as the difference between Han and the largest minority group in the percentage of individuals who at least completed lower secondary school, has a positive effect on the chances of ethnic violence. On the other hand, the oil variable is negatively associated with violence; but the statistical significance level of this association varies by model specifications—the only statistically significant effect is found in the last model specification when we include government expenditure as a percentage of GDP in the regression analysis. County-level cotton production per capita also has no effect on violence. However, we need to be cautious in interpreting these nonresults: we use the location of oil fields and county-level cotton production to capture resource extraction; but the location of oil fields and total cotton production might not reflect how jobs and resources are distributed in these extractive industries and especially to ethnic minority groups, if at all.68

Table 3: Binary Logistic Regression Models Predicting Ethnic Violence in Xinjiang

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial lag</td>
<td>1.448*</td>
<td>1.540*</td>
<td>1.586*</td>
<td>1.368</td>
</tr>
<tr>
<td></td>
<td>(0.842)</td>
<td>(0.909)</td>
<td>(0.926)</td>
<td>(0.917)</td>
</tr>
<tr>
<td>Largest minority</td>
<td>0.017</td>
<td>0.011</td>
<td>0.021</td>
<td>0.025</td>
</tr>
<tr>
<td>percentage</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.013)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>GDP per capita</td>
<td>1.545***</td>
<td>1.202*</td>
<td>1.251**</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>(0.504)</td>
<td>(0.601)</td>
<td>(0.611)</td>
<td>(0.708)</td>
</tr>
<tr>
<td>Distance to the</td>
<td>0.002</td>
<td>−0.0001</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>prefecture’s capital</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Distance to Ürümqi</td>
<td>−0.002</td>
<td>−0.001</td>
<td>−0.002</td>
<td>−0.001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Border county</td>
<td>0.931*</td>
<td>0.839*</td>
<td>1.141**</td>
<td>1.101**</td>
</tr>
<tr>
<td></td>
<td>(0.526)</td>
<td>(0.468)</td>
<td>(0.520)</td>
<td>(0.491)</td>
</tr>
<tr>
<td>Population density</td>
<td>0.572***</td>
<td>0.565***</td>
<td>0.603***</td>
<td>0.451***</td>
</tr>
<tr>
<td></td>
<td>(0.173)</td>
<td>(0.165)</td>
<td>(0.168)</td>
<td>(0.164)</td>
</tr>
<tr>
<td>Ethnic fractionalization</td>
<td>0.381</td>
<td>0.780</td>
<td>1.282</td>
<td>1.733</td>
</tr>
<tr>
<td></td>
<td>(2.473)</td>
<td>(2.329)</td>
<td>(2.374)</td>
<td>(2.246)</td>
</tr>
<tr>
<td>Ethnic polarization</td>
<td>−1.167</td>
<td>−1.316</td>
<td>−1.450</td>
<td>−1.708</td>
</tr>
<tr>
<td></td>
<td>(1.758)</td>
<td>(1.646)</td>
<td>(1.682)</td>
<td>(1.468)</td>
</tr>
<tr>
<td>Mosques per 1,000 non-Han</td>
<td>−0.354</td>
<td>−0.370</td>
<td>−0.465</td>
<td>−0.466</td>
</tr>
<tr>
<td></td>
<td>(0.341)</td>
<td>(0.304)</td>
<td>(0.319)</td>
<td>(0.319)</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>XPCC</td>
<td>−1.635</td>
<td>−1.046</td>
<td>−1.233</td>
<td>−1.018</td>
</tr>
<tr>
<td></td>
<td>(1.093)</td>
<td>(0.941)</td>
<td>(0.942)</td>
<td>(0.958)</td>
</tr>
<tr>
<td>Government expenditure/GDP</td>
<td></td>
<td>−0.067</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal inequality</td>
<td>0.071**</td>
<td>0.049*</td>
<td>0.065**</td>
<td>0.072**</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.027)</td>
<td>(0.030)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Oil</td>
<td>−0.350</td>
<td>−0.903</td>
<td>−1.108*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.471)</td>
<td>(0.557)</td>
<td>(0.559)</td>
<td></td>
</tr>
<tr>
<td>Cotton production per capita</td>
<td>2.838</td>
<td>2.945</td>
<td>1.778</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.005)</td>
<td>(2.044)</td>
<td>(2.009)</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>−0.451</td>
<td>−0.281</td>
<td>−0.264</td>
<td>−0.246</td>
</tr>
<tr>
<td></td>
<td>(0.301)</td>
<td>(0.325)</td>
<td>(0.327)</td>
<td>(0.333)</td>
</tr>
<tr>
<td>t²</td>
<td>0.058</td>
<td>0.024</td>
<td>0.023</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.060)</td>
<td>(0.060)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>t³</td>
<td>−0.002</td>
<td>−0.0004</td>
<td>−0.0004</td>
<td>−0.0002</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Constant</td>
<td>−12.518***</td>
<td>−9.870***</td>
<td>−11.050***</td>
<td>−8.954***</td>
</tr>
<tr>
<td></td>
<td>(2.271)</td>
<td>(2.351)</td>
<td>(2.543)</td>
<td>(2.465)</td>
</tr>
<tr>
<td>Observations</td>
<td>682</td>
<td>580</td>
<td>580</td>
<td>580</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−120.628</td>
<td>−108.010</td>
<td>−106.909</td>
<td>−105.081</td>
</tr>
<tr>
<td>Akaike information criterion</td>
<td>295.255</td>
<td>268.021</td>
<td>267.818</td>
<td>266.162</td>
</tr>
</tbody>
</table>

Note: XPCC = Xinjiang Production and Construction Corps. *p < .1. **p < .05. ***p < .01.

To get a sense of the substantive effects, based on the third model specification in Table 3, we simulated the change in probability of a county-year experiencing ethnic violence given a one-standard-deviation increase from the mean of an independent variable, when holding all other variables at their mean levels. The rope ladders in Figure 3 are 95 percent confidence intervals of the simulated probability changes. Horizontal inequality stands out in terms of the substantive effect: holding all other variables at their mean levels, one-standard-deviation increase from the mean of the horizontal inequality variable results in an increase in the probability of ethnic violence for a county-year by about 0.03. This is a substantively important effect because ethnic violent events do not occur often in the region; indeed, between 1996 and 2005, the mean of our binary dependent variable is only 0.08.
Other variables that affect the chances of ethnic violence include GDP per capita, population density, and border county status. Interestingly, we find that GDP per capital increases the chances of violence, which contradicts some recent studies on political violence. Many believe that economic prosperity reduces chances for violence because it alleviates ethnic competition by offering more resources; higher per capita income also presents higher opportunity cost for violence; economic development may also mitigate the level of grievances among marginalized groups. This positive effect of GDP per capita agrees with other recent studies though. For instance, in the context of the collapse of Soviet Union, Beissinger finds that the level of urbanization (a proxy of economic development) increases the frequency of protests over ethnonationalist issues because economic development often fosters the very condition for local minorities to form nationalist networks, which increase the chances of violence.

Furthermore, we find that population density increases ethnic violence, confirming previous findings from earlier conflict studies. Finally, border counties are also associated with higher chances of conflicts; this is also similar to recent studies that has identified national border as a determinant of civil conflict.
3. Conclusion and Discussion

In this article, we introduce the EVC Database: the Xinjiang Region, 1990–2005. This is a geocoded database of yearly incidents of ethnic violence occurring at the county level in Xinjiang from 1990 to 2005. Moreover, using the EVC event data, we further demonstrate some initial results explaining ethnic violence in Xinjiang. We find that ethnic violence is positively associated with horizontal inequality, which is measured by education attainment difference between Han and the largest ethnic minority group in a prefecture. This suggests that grievances as a function of interethnic inequalities need to be addressed to reduce ethnic violence in the region. Recent literature has also suggested that resource wealth and resource extraction increase risks of internal political conflict and violence. However, we do not find evidence supporting such a resource curse argument; neither oil nor cotton production is related to chances of ethnic violence in Xinjiang.

The results here are preliminary because there are other potential factors not controlled in the analysis. For instance, what is the role of marketization on violence and conflicts? Do free-market economic reforms—such as those implemented by China in Xinjiang—have an impact on political stability and the prospects for violence and armed conflict? These are unanswered questions not only in the study of Xinjiang unrest, but also in the broader literature of conflict studies.73

Indeed, one of the most transformative politico-economic phenomena in the world in the past few decades has been the process of economic liberalization (marketization), especially in the developing world. Yet, we know little about how marketization affects domestic political stability in many countries of the world, especially those with histories of domestic conflict. Roland Paris, discussing the effectiveness of immediate democratization and marketization as a common strategy for postconflict peace building, argues that such political and economic liberalization is inherently tumultuous, and thereby undermines the prospects for stable peace.74 Fearon and Laitin also point to the role of economic modernization in enhancing the positive effect of ethnic diversity on civil war. They argue that this is because more modernization should imply more discrimination and more nationalist contention in culturally divided countries.75 However, other scholars have argued the opposite, associating economic liberalization with increased efficiency and welfare, and therefore lower chances of conflicts. For instance, recent
research suggests that states that trade frequently experience fewer onsets of civil wars,\textsuperscript{76} and lower chances of mass protests and political violence.\textsuperscript{77}

As far as we know, no existing research has studied the distributive effects of market liberalization in an ethnically diverse society and how it affects chances of ethnic violence and political stability. Aggregated, country-level analyses often have difficulties to reveal nuanced differences in the ways that various aspects of marketization affect a society. This is especially relevant when we look at localized violent events instead of large-scale civil wars. Another complication and empirical challenge is that democratization and marketization are often intertwined so that it is really difficult to separate the effects. Regarding both empirical challenges, the Xinjiang case provides a unique testing ground to study the effects of marketization on ethnic and political violence: our disaggregated event data approach allows us to better specify local precipitants of ethnic violence; the fact that China has engaged in substantial economic reform without substantial political reform allows us to examine effects of economic factors on political violence while holding political factors constant.

The aforementioned case of testing the marketization-violence nexus using the EVC data is just one example of what we can potentially learn by using this event data set. Accompanying this article, we will make the database and replication code publicly available. One temporal limitation of the current event data is that they cover only 1990 to 2005.\textsuperscript{78} There are questions concerning more recent events that cannot be answered using our data. For example, have the dynamics of interethnic conflict changed since the 7.5 Ürümqi riots in 2009?\textsuperscript{79} Therefore, we plan to extend event coding to cover the 2006–2016 period.\textsuperscript{80} Our hope is that this data set, as an important first step, will help scholars in fields such as civil armed conflicts, ethnic violence, and Chinese politics to not only better study and understand this important region of China, but also use Xinjiang as a test case for theories that can be generalized to other regions of the world.

Appendix: Data Sources for the EVC

1. Chinese Government Documents
   a. Xinjiang Public Security Gazette (XPSG: 新疆通誌：公安誌):
      This was published by the Xinjiang Uyghur Autonomous
Regional Government Compilation Committee in 2004. It provides detailed information on ethnic conflicts in Xinjiang during the 1990s. The most valuable information in this book is presented in the fifth and sixth sections of the appendix (pp. 840–862). The title of the fifth section, which covers protests, riots, and (armed) fights, is “Events of All Types.” In the sixth section, the gazette records important violent incidents from 1996 to 2000. XPSG offers valuable local sources for ethnic conflicts. This book is based on internally circulated government statistics, which enhances its reliability and credibility.

b The unpublished draft of Xinjiang Public Security Gazette (unpublished XPSG 新疆公誌 草稿油印本): This is an unpublished draft of Xinjiang public security gazette discussed above. The draft provides a rare opportunity to access to original, internally circulated sources on violent events in Xinjiang. Main criminal activities and violent events are recorded annually in this draft. Different from the final published version, the draft covers episodes from 1949 to 1995. Moreover, the draft also presents each case thoroughly by articulating the time, locations, and detailed descriptions. This source is highly reliable and authoritative owing to the nature of internal circulation. This draft gives us firsthand information regarding ethnic conflicts in Xinjiang after 1949. Same as the published version, the draft is informative about the location of each violent event, which makes the county-level coding feasible.

c Government white paper: The Information Office of the State Council of China made a white paper, “East Turkistan Terrorist Forces Cannot Get Away with Impunity,” publicly available via People’s Daily on 21 January 2002.81 This white paper includes events such as explosions, assassinations, and attempted assassinations, attacks on police and government institutions, crimes of poison and arson, training and arms manufacturing bases, and plotting and organizing riots and creating terror. Whereas the white paper claims over 200 terrorist incidents between 1990 and 2001, it reports details on only about 39 episodes. Moreover, many scholars doubt the reliability of this report.82 However, this white paper is still the most important official document that reveals valuable statistics about ethnic tensions in Xinjiang.
during the 1990s. According to Clarke, the “report is the sole source of material regarding incidents of assassinations and attacks involving poison and arson.”

d County gazettes of Xinjiang (新疆縣誌) http://www.fudan.edu.cn/englishnew/library/library.html. These gazettes record most detailed socioeconomic information regarding counties in Xinjiang. Strikingly, only a few of them mention ethnic violence, and these records tend to include only the most salient cases such as the 1997 Ghulja incident.

2 Online Databases and News Archives and Search Engines

a The Global Terrorism Database (GTD, http://www.start.umd.edu/gtd/): This is a global database covering both international and domestic terrorist attacks. It is widely used in empirical research on terrorism. GTD records 39 events associated with Uyghur participants in Xinjiang from 1996 to 2011. Within these 39 events, however, 23 of them are listed repetitively: there was a bombing occurred on 20 July 1996, which is recorded 23 times in the GTD. We do not include other 22 records in our dataset. For each case, GTD provides information on timing, location, group involved, fatalities, target type, attack type, and weapon type.

b Minorities at Risk (MAR, http://www.cidcm.umd.edu/mar/chronology.asp?groupId=71003): The MAR is a university-based research project that monitors the status and conflicts of politically active communal groups in all countries with a current population of at least 500,000. The project is designed to provide information in a standardized format that aids comparative research and contributes to the understanding of conflicts involving relevant groups. Although MAR’s quantitative data do not disaggregate the geographic distribution of ethnic violent events in Xinjiang, the chronology of Turkmen in Xinjiang (Turkic minorities including Uyghurs, Kazakhs and Tajiks) is useful. After scrutinizing this chronology, we recorded 44 events.

c WiseNews (http://libguides.ust.hk/wisenews): WiseNews is a Hong Kong based online newspaper data search engine. It covers over 600 key newspapers, magazines, and websites from mainland China, Hong Kong, Macau, Taiwan, and the United States.
WiseNews provides digital archives of 259 newspapers based in Mainland China from 2000 to date. Among them, 125 are general-interest newspapers. Geographically, these newspapers cover 29 out of 31 provinces. In terms of government affiliation, ownership, and the type of readership, the WiseNews sample represents the whole newspaper industry in Mainland China. We use the relevant strings of Chinese characters (e.g., 新疆 [Xinjiang] and 暴力 [violence]) to search for the mass media coverage of political violence in the area. After ruling out overlapped events already reported by other sources, we found 5 new events between 2000 and 2011.

3 Secondary Datasets from Existing Scholarly Work

a Bovingdon: This book offers very comprehensive quantitative evidence on the amount and frequency of political violence in Xinjiang between 1949 and 2005. This book carefully balances between Chinese official statistics and messages from transnational Uyghur organizations and human rights organizations. The appendix of this book, titled “Organized Protests and Violent Events in Xinjiang, 1949–2005,” provides the year, duration, location, type, number of people involved, and comments (which provides detailed descriptions of coded events) for mass protests and violent events. This appendix records 158 political violent events and organized protests from 1949 to 2005. Within these episodes, 142 had clear ethnic component. However, these events are recorded case by case without disaggregating at the county level. Our dataset reorganizes the structure of this appendix by geocoding each case at the county level.

b Ma: This is a highly valuable internally circulated source regarding ethnic conflicts in Xinjiang. According to Bovingdon, “Ma’s importance as both a researcher and a policy adviser on Xinjiang cannot be overestimated.” The research that he and his assistants conducted for the book was commissioned by top party leaders in Beijing and written in consultation with high officials in Xinjiang. This book offers detailed information on 28 ethnic violent events in Xinjiang from the 1950s to the 1990s. Furthermore, the book highlights the evolution of ethnic violence in Xinjiang during the 1990s.
Notes

1. One should note the difference between terrorism and separatism, especially given the fact that they have sometimes been used interchangeably by the government and media in China. Terrorism can be defined, for example, by the UN as criminal acts intended or calculated to provoke a state of terror in the general public, a group of persons or particular persons for political purposes. The defining characteristics are often, first, no distinction between civilian and military targets and, second, the victim not being the target. Separatists, on the other hand, do not have to engage in terrorist activities. Separatism is often considered the advocacy or practice of separation of a certain group of people from a larger body on the basis of ethnicity and religion.


5. Some believe that at least some of the civil servants were associated with conducting intelligence and surveillance work; see James Leibold and Adrian Zenz, “Beijing’s Eyes and Ears Grow Sharper in Xinjiang,” Foreign Affairs, 23 December 2016.


However, see Ji Yeon Hong and Wenhui Yang, “Oilfields, Mosques, and Violence: Is There a Resource Curse in Xinjiang,” *British Journal of Political Science* (Forthcoming), for how natural resources and local mosques affect violence in the region; Chuyu Liu, “Local Public Goods Expenditure and Ethnic Conflict: Evidence from China” (working paper, 2017), for how local public goods spending decreases ethnic violence in Xinjiang; and Xun Cao, Haiyan Duan, Chuyu Liu, and Yingjie Wei, “Religious Institutions and the Impact of Inter-ethnic Inequality on Conflict: The Case of Xinjiang, China,” *International Studies Quarterly* (Forthcoming), for how local religious institutions mediate the effect of horizontal inequality on ethnic violence. Liu and Cao et al. use the EVC data introduced in this article for their main empirical analyses.

A “county-year” version of panel data will accompany this database to best capture the granularity and local nature of the events. Such a unit of analysis is crucial, we argue, as we expect local factors to be of importance in determining patterns of violent events in the region.

We also assume that arms manufacturing is a step toward violent attacks; therefore weapons production captured by the government is counted as violence. There were a total of six such events between 1990 and 2005.

Based on our definition, Han violence on minorities is part of our data. However, we found no record of Han violence on minorities. We suspect that there are two reasons. First, as Côté argues, the Han Chinese political mobilization was generally rare and limited in Xinjiang. See Isabelle Côté, “Political Mobilization of a Regional Minority: Han Chinese Settlers in Xinjiang,” *Ethnic and Racial Studies*, Vol. 34, No. 11 (2011), pp. 1855–1873. Han Chinese tend to perceive themselves as the ethnic group in power; thus, they are more likely to press the government to address their grievances. Second, it is possible that there are reporting biases that the media and the government tend to treat Han violence on minorities as general crime rather than politically charged ethnic violence.


Ibid.

In the regression analysis, we assign the events with only prefecture location to the prefecture capital city as the county-level location. However, excluding these events from the regression analysis does not change our main results.


See Bovingdon, *Uyghurs*; Dazheng Ma, *National Interests the Highest* (Ürümqi: Xinjiang People’s Publishing House, 2002). None of the sources that we use
are considered to be classified. More details about each data source are provided in the appendix.


22 See Gurr, Why Men Rebel. Others, on the other hand, disagree and view mobilization as the key factor affecting chances of civil unrest because...
grievance is always present in every society while conflict is not. See Charles Tilly, *From Mobilization to Revolution* (New York: McGraw-Hill, 1978). While grievances might provide motivation, there is still often the collective action and mobilization issue.


24 Fuller and Lipman, “Islam in Xinjiang.”


39 Fjelde and Østby, “Socioeconomic Inequality and Communal Conflict.”

40 Barron, Kaiser, and Pradhan, “Understanding Variations in Local Conflict.”


42 For the pre-1996 period, the only publicly available Chinese census sample data are from 1982 and 1990. We chose not to use the 1982 data in the main text because 1990 is much closer to the period covered by our study (1996–2005).

The 2000 census data became publicly available in June 2017. The data were downloaded from the IPUMS-International (https://international.ipums.org/international/, accessed 4 June 2017). This enables us to conduct a robustness check to see if we could replicate the main findings of this article when using a new horizontal inequality measure based on both the 1990 and 2000 census data. An online appendix to this article (https://sites.psu.edu/xuncao/files/2016/01/EVC-research-note_sep_2017_online_app-uoglyf.pdf) discusses the details of the variable construction and presents the results of this robustness check: briefly, our main findings remain unchanged.


Data are from the county statistics of China data online: http://chinadata-online.org/member/county/ (accessed 9 November 2016).


Qiao Yuanrao, Xinjiang 50 Years (Beijing: China Statistics Press, 2005).

Data are also from Qiao, Xinjiang 50 Years.


Qiao, Xinjiang 50 Years. We adjust GDP per capita based on 1990 Xinjiang price index to account for inflation.


The overwhelming majority of the non-Han population is Muslim, while few Han Chinese affiliate with Islam.

The Spatial Religion Explorer from China Data Online also provides mosque data (http://chinadataonline.org/religionexplorer). However, the Explorer data quality varies greatly from county to county. For example, for counties in southern Xinjiang where the Muslim population concentrates, few mosques are reported, which contradicts what we found from county gazettes.

For example, if the mosques data are available for a county only in 1985, 1990, and 1993, we use the value from 1993.

This is a crude measure for the density of local religious institutions. For instance, the data from county gazettes do not allow us to differentiate mosques attended by different Muslim ethnic groups such as Uyghur mosques versus Kazak mosques. More importantly, there is the potential risk that the spatial distribution of mosques changed during the investigation period of 1996–2005. However, Li shows that since September 1990, the government has enforced a regulation on religious activities that strongly restricts, if not completely prohibits, the construction of new mosques in the region. As a result, the change in the number of mosques since late 1990 has been small. For instance, there were 22,900 mosques in Xinjiang in 1995, only about 600 (or 2.3%) more than in 1991. Xiaoxia Li, “Changes in the Number of Mosques and Regulatory Policies in Xinjiang,” Sociology of Ethnicity, Vol. 164 (2014), pp. 40–48.

To calculate the distance between each county and the provincial prefectural capital, we used Google Maps to find the latitude and longitude of each county office building in Xinjiang.
The list was downloaded from http://baike.baidu.com/view/38528.htm on 21 September 2016.


The government spending variable is listed as sum of expenditures (Zhi Chu He Ji) from the National Prefecture and County Finance Statistics Compendium (Quanguo Di Shi Xian Caizheng Tongji Ziliao), 1994–2005, published by the Ministry of Finance.


Empirically, however, it is extremely difficult to measure the amount of natural resource wealth distribution to local ethnic minorities. Until recently, only a small fraction of oil revenue was kept by local governments while the rest went to central-state-owned oil companies and the central government. For example, focusing on the city of Korla, whose economy is built around oil extraction, Cliff shows that SOEs in the oil sector such as the local Tarim Oilfield Company predominantly employ Han Chinese. Moreover, even for the small fraction of oil revenue that was kept local, no data exist regarding what proportion was used to provide services and create jobs for the local population. Tom Cliff, Oil and Water: Being Han in Xinjiang (Chicago: University of Chicago Press, 2016).

For year fixed effects, it was hard to interpret the results if we used the mean level, which is 0.1, for the simulations in Figure 3. We therefore chose to use the latest year for our investigation period, 2005, for the simulations. Using other years changed the magnitudes of predicted probabilities, but the rope ladder shapes did not change much. Simulation figures using other years are available on request from the authors.


Note that we find that GDP per capita increases the chances of violence. In the Chinese context, wealth is highly correlated with marketization. See Gang Fan, Wang Xiaolu, and Zhu Hengpeng, NERI Index of Marketization of China’s Provinces 2011 Report (Beijing: Economic Science Press, National Economic Research Institute, China Research Foundation, 2011). This result seems to suggest possible connections between marketization and ethnic violence.

Fearon and Laitin, “Ethnicity, Insurgency, and Civil War.”

De Soysa, “Paradise Is a Bazaar?”


There are a few reasons why we collected data only for 1990–2005. First, the collection of the data is very time-consuming: For example, one of the most challenging tasks is to find county-level locations of the events because many events from news reports do not provide precise location information such as county names for the events; we had to conduct online searches and consult secondary sources to find county names. Second, in addition to online event data sets, news search engines, and secondary data from existing scholarly work, we heavily rely on government documents. These government documents help to address the concern of potential reporting biases by the mass media. See David Ortiz, Daniel Myers, Eugene Walls, and Maria-Elena Diaz, “Where Do We Stand with Newspaper Data?” *Mobilization: An International Quarterly*, Vol. 10, No. 3 (2005), pp. 397–419. Moreover, many government documents are based on police sources and administrative archives, which significantly improve the scope of coverage. See John D. McCarthy, Clark McPhail, and Jackie Smith, “Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991,” *American Sociological Review*, Vol. 61, No. 3 (1996), pp. 478–499. However, one trade-off of using government documents is the fact that most of the government documents are available only till the early 2000s—this is the second reason why, given the resources we had, we decided to code the 1990–2005 period first.

Our empirical analysis shows that horizontal inequality is associated with violence during the 1996–2005 period. Without additional analysis covering the post-2005 period, we cannot answer questions like this. However, our main finding seems to be consistent with a few recent studies focusing on the post-“7.5” period, such as Hillman and Tuttle, *Ethnic Conflict and Protest*. Our main finding that interethnic hostility in Xinjiang was driven by local grievances is also consistent with recent studies on social exclusion (Tom Cliff, “Lucrative Chaos: Inter-ethnic Conflict as a Function of the

We will first be working with some of the same sources that we used for the 1990–2005 period. Moreover, for more recent periods such as the late 2000s, more online data sources become available: We can turn to English databases such as LexisNexis—this is the approach employed by the AidData project to code China’s aid projects in developing countries. We can use keyword searches on public search engines such as Google and Baidu to code events—the same method has been successfully applied by recent research on sensitive topics in Chinese politics such as in Jeremy L. Wallace and Jessica Chen Weiss, “The Political Geography of Nationalist Protest in China: Cities and the 2012 Anti-Japanese Protests,” *The China Quarterly*, Vol. 222 (2015), pp. 403–429. Finally, in addition to general search engines, there are a number of websites that have closely followed the Xinjiang issue over recent years (e.g., www.ifeng.com, www.people.com.cn/). For example, a keyword search at http://news.qq.com/a/20140302/005359.htm provides detailed reports (in Chinese) of 16 violent events in Xinjiang from mid-2011 to early 2014.


Clarke, “China’s War on Terror,” pp. 282–283; and Bovingdon, *Uyghurs*.

Clarke, “China’s War on Terror,” pp. 271–301.

We also searched popular English news online archives such as LexisNexis. However, we found no additional events. Therefore, we do not list them in the main text of this proposal.


Bovingdon, *Uyghurs*, p. 174: “Bibliographic references for each of these reported events available on request. As discussed in chapter 4, all but a handful of these events are described in multiple sources. Wherever possible, I compared the Chinese and foreign sources when preparing the descriptions.”

Ma, *National Interests*.

Bovingdon, *Uyghurs*. 