COLUMBIA | SIPA
Center for Development Economics and Policy

CDEP-CGEG WORKING PAPER SERIES

CDEP-CGEG WP No. 65

Within but Without: Involuntary Displacement and Economic Development

Alice Tianbo Zhang

October 2018

COLUMBIA | SIPA
Center on Global Economic Governance

Within but Without: Involuntary Displacement and Economic Development*

Alice Tianbo Zhang†

October 23, 2018

Job Market Paper

Click Here For Latest Version

Abstract

Between 1995 and 2010, China's Three Gorges Dam uprooted more than one million people, resulting in the largest involuntary displacement from dam construction in history. This paper provides the first evidence of the causal impact of dam-induced inundation on migration and labor market outcomes by combining micro-level census and satellite data. Using a novel identification strategy and remote sensing techniques to capture exogenous variations in flooding intensity, I find that inundation has imposed large and enduring costs on the local economy. Rising water levels in the dam's reservoir displaced between 1.5 and 1.9 million people. Most migrants in partially flooded counties relocated to other towns or villages within the same county. Flooded counties saw a steep and persistent decline of employment by 30 to 50 percent in the manufacturing sector and capital-intensive occupations, and the effects were more harmful to residents that did not move. In the long run, the decline in manufacturing was partly offset by reallocating workers to agriculture. Industrial firms in flooded counties responded by downsizing the workforce and reducing compensation. Overall, these findings highlight the need for policy evaluations to carefully weigh the broad benefits of infrastructure against the concentrated costs to local communities.

*I am deeply grateful to Geoffrey Heal, Wolfram Schlenker and Rodrigo R. Soares for invaluable guidance and support. I thank Douglas Almond, Eyal Frank, Jonas Hjort, Jaehyun Jung, Anouch Missirian, John Mutter, Suresh Naidu, Suanna Oh, Cristian Pop-Eleches, Miikka Rokkanen, Jeffrey Shrader, Christopher Small, Anna Tompsett, Miguel Urquiola, Eric Verhoogen, Jack Willis, Andrew Wilson, participants of the Sustainable Development Colloquium, Development Colloquium and Sustainable Development Research Symposium at Columbia, the Heartland Environmental and Resource Economics Workshop at UIUC, and the National Academy of Sciences Sackler Colloquium for helpful comments and suggestions. I thank Jeremiah Trinidad-Christensen and Eric Glass from the Digital Social Science Center at Columbia for generous help with acquiring historical GIS data. All errors are my own.

†Ph.D. Candidate in Sustainable Development, School of International and Public Affairs, Columbia University. 420 W. 118th St., New York, NY 10027. Email: alice.zhang@columbia.edu

1 Introduction

Large-scale infrastructure development projects often create winners and losers. Infrastructure, such as railroad, telecommunication network, and electrical grid, lies at the heart of economic development by lowering trade costs, creating jobs, raising labor productivity and spurring growth (Dinkelman 2011; Donaldson and Hornbeck 2016; Donaldson 2018; Lipscomb, Mobarak and Barham 2013; Roller and Waverman 2001). But while the benefits of such projects are spatially diffused over a large population, the costs are often concentrated in the surrounding communities (Duflo and Pande 2007). Dams, in particular, are hailed as a leading renewable energy source and justified on the grounds of flood control, electricity generation, irrigation, and river navigation. But they can also cause irreversible damage to the surrounding ecosystem and community by destroying natural habitats and uprooting millions of people. Worldwide, dams are a major source[1] of development-induced internal displacement, forcing an estimated 40 to 80 million people to move out of their homes over the last half-century (World Commission on Dams 2000). As the world is witnessing a boom in dam construction with at least 3,700 large dams[2] being planned or built, understanding their social and economic impacts will have important policy implications (Zarfl et al. 2015).

While the benefits of infrastructure projects have been extensively studied in the literature, the social costs and spatial distribution of economic impacts are largely unknown. In this paper, I provide the first micro-level evidence of the causal impact of dam-induced inundation on migration and labor market outcomes by using plausibly exogenous variation from the Three Gorges Project (TGP) in China as a natural experiment. Drawing on a comprehensive dataset that I have constructed, including individual- and firm-level census, satellite data on elevation and inundation extent, historical administrative GIS boundaries, archival government legislative documents and geo-referenced data on dams, I make four main contributions to improve our understanding of the social costs of energy infrastructure. First, I propose a novel identification strategy that uses the reservoir maximum design water level to capture exogenous variations in inundation intensity. Resi-

[1]The World Bank (WB) estimated that dam construction accounted for 63 percent of the total population displacement by WB-funded development projects between 1986 and 1993, or about 1.2 million people from 39 projects (World Bank 1996). The current scale of such development-induced displacement is around 10 to 15 million people annually, almost on par with disaster or conflict-induced displacement in magnitude (Cernea 1997; IDMC 2017, 2018).

[2]Zarfl et al. (2015) define large dams as those with generation capacities of more than 1 megawatt (MW).

dents living below the water level were forced to move out, while those living at higher elevations could stay. Using satellite data and remote sensing techniques, I calculate the fraction of county area inundated by the dam's reservoir and show that ex-ante elevation relative to the water level accurately predicts ex-post inundation intensity. Second, I document the magnitude and direction of the population displacement flow and provide an empirically-derived estimate to bound the government's official and often contested displacement count. Access to microdata also allows me to uncover new sources of variation in migration patterns and map out the bilateral migration matrix between all county pairs in China. Third, I estimate the short- and long-run impact of inundation on the local labor market, focusing specifically on the transitional costs of reallocating the labor force. I also separately examine the heterogeneous effects on migrants and nonmigrants. Fourth, I use establishment-level data to shed light on the mechanisms through which firms may have adjusted in response to inundation, population displacement, and plant relocation.

The Three Gorges Dam (TGD) is the largest hydroelectric power plant in the world with an installed capacity of 22,500 MW, equivalent to about 20 coal-fired power plants. It is also one of the largest man-made structures on the planet and is lauded as the crown jewel of infrastructure development by its proponents. But the project was marred by controversy throughout its 17-year construction period between 1994 and 2010. When fully operational, the dam raised water levels of the Yangtze River from 68 to 175 meters above sea level and inundated an area of more than 1000 km^2. The ensuing flooding led to the largest dam-induced involuntary displacement flow affecting at least 1.3 million people. The displaced were promised compensation for the loss of their home and income, but in most cases, it was either inadequate to restore their standard of living or never materialized (Duan and Steil 2003; Li and Rees 2000; Wilmsen, Webber and Duan 2011*a*,*b*). As a result, while the majority of migrants were within the homeland,[3] they were often left without stable housing or income source – their former way of life irreparably lost.

Several features of the TGP make it a particularly appealing setting. First, the project

[3]People that are displaced within their country of residence are often called Internally Displaced Persons (IDPs), defined as "persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized border" (UNHCR 1998). IDPs are among the most vulnerable populations in the world. Unlike refugees, who cross national borders and are protected by international law, IDPs are under the jurisdiction and protection of the very government that may have caused their plight.

provides an exogenous source of variation that can be used as a natural experiment to causally estimate the labor market impacts of inundation and involuntary displacement. Throughout the long deliberation and eventual approval process, decision-making was highly centralized with little participation by local stakeholders. The dam location was determined by an extensive geological survey in 1979, almost two decades before construction began. Second, the magnitude and duration of the relocation process were both unprecedented enough to warrant a closer examination of the short- and long-run impacts on migrants and nonmigrants. Third and most importantly, the government used a simple and transparent rule to classify residents as part of the displaced population. All residents living below the 175-meter flood line in the reservoir catchment zone were relocated at some point between 1995 and 2010, and those whose homes were at lower elevations were ordered to leave earlier.

A major challenge for identifying the causal impact of population displacement is that even in settings where the shocks are plausibly exogenous, such as civil conflicts and natural disasters, there may still be selection into who moves away and returns. In light of the government's relocation policy, I propose a novel identification strategy that uses a key engineering feature of dams, the Maximum Design Water Level (MDWL),[4] to assign treatment. Individuals living near the reservoir at elevations below the MDWL were inundated and forced to move out. In the ideal experiment, I would classify individuals to be treated by TGP using the exact longitude and latitude of their homes and a continuous measure of topography to compare those living just below and above the water level. But county is the lowest level of geography reported in the census for the place of origin. Therefore, I assign treatment at the county level using its minimum elevation and actual changes in the flooded area before and after TGD construction calculated from satellite imageries. A county is treated if it 1) has a minimum elevation below 175m, 2) has any changes in river area between the 1990s and 2010s, and 3) is upstream of the dam. To get a continuous flooding intensity measure, I interact this treatment dummy with the fraction of county area covered by the Yangtze River in the 2010s. I find that a county's ex-ante minimum elevation accurately predicts its ex-post changes in the inundated area.

The empirical specification uses a difference-in-differences model to compare changes in migration and labor market outcomes between inundated and nearby non-inundated

[4]The MDWL is defined as "the highest elevation of water determined as a result of safely passing the Inflow Design Flood" by U.S. Fish and Wildlife Service (2003), Part 361 Dam Safety regulations.

counties in the same province and with similar pre-TGP characteristics, relative to the baseline in 1990 or 1995. For the main analysis, I create a county-level panel between 1990 and 2010 by harmonizing 1990, 2000, 2005 and 2010 population census data. For geographic comparability, I construct crosswalk relationship files for all counties in China between the census years and adjust the data to hold 1990 geographic boundaries constant. The main sample consists of 2,460 counties with consistent county geography between 1990 and 2010. Moreover, access to microdata in 2000 and 2005 allows me to retrospectively construct a county-year panel of outmigration between 1995 and 2005 and map out the bilateral migration matrix between all *county* pairs in China. This would not be possible with aggregated county, prefecture or province level census commonly used in the literature. I show that information on the fine-scale migration patterns is crucial for validating the identification assumptions and informing the research design.

I begin by documenting the magnitude and direction of population displacement to bound the government's official and often contested displacement count. While the displacement shock was involuntary in nature, in most cases individuals could and were encouraged by the government to choose their destination and resettle on their own. I find that TGP-induced population displacement was unprecedented in magnitude and duration but spatially concentrated in the affected provinces, Chongqing and Hubei. The majority of migrants in flooded counties moved along the river to other towns or villages within the same county, and some moved to immediate neighboring counties that were also partially flooded. Peaks in outmigration flow coincided with key construction milestones in 1998 and 2003 when the water level in the reservoir was elevated. Since spatial spillover of the displaced population was confined within the treatment group, non-flooded counties in the same province with similar pre-trend in outcomes would constitute the proper control group. I estimate that between 1996 and 2005, inundated counties had 998,556 more within- and across-county migrants than control counties relative to the baseline in 1995. This implies that between 1995 and 2010, the total number of people directly displaced by TGP-induced flooding was likely to be between 1.5 and 1.9 million, considerably higher than the 1.4 million reported by the government.

In addition to uprooting millions of residents and redistributing them across flooded counties, inundation has wide-ranging impacts on the local economy by reallocating the labor force across sectors. Accompanied by the loss of land and physical capital, inundated counties witnessed a large and persistent decline of employment in the manufacturing sector and capital-intensive occupations by 30 to 50 percent of the mean. Overall employment in flooded counties exhibited a U-shaped pattern over time, such that the decline in manufacturing was partly offset by reallocating workers to agriculture in the long run. There was little change in service sector employment, providing evidence against a process of "creative destruction," through which the destruction of land and physical capital would spur a structural transformation in the local economy by creating new technologies. To shed light on the transitional costs of reallocating the labor force, I show that flooded counties experienced a significant increase in the share of unemployed migrants in the short run. Displaced migrants had few financial assets to buffer the loss of job and house and mostly relied on pension, family support and other informal sources of income.

To further examine the heterogeneous labor market impacts on migrants and nonmigrants, I show that TGP-induced inundation has affected not only those displaced but also the local population at large. Results from a subsample analysis on nonmigrants are robust and essentially identical to the full sample. This suggests the closure of firms and factories during the relocation process led to a "reverse structural transformation" in flooded counties, in which the decline of manufacturing forced local residents to seek jobs in agriculture and other less capital-intensive industries. The impact of the TGP was also more detrimental and longer-lasting for residents that did not move, possibly due to their more limited ability to find employment in other sectors and occupations. Finally, to shed light on firm adjustments, I use establishment-level data to show that firms in inundated counties responded by employing a variety of cost-cutting measures.

In a falsification test, I provide further evidence for the validity of the identifying assumptions by exploring differential changes in migration and labor market outcomes in placebo counties. To generate a distribution of estimated coefficients, I construct an ensemble of placebo samples, which include counties within varying distance cutoffs of the Yangtze River in upstream provinces outside of the reservoir's catchment zone. Considering upstream rather than downstream provinces ensures that the results are comparable to the main analysis and that these counties do not directly benefit from the dam, such as through flood control. In the absence of TGP-induced flooding, placebo counties did not experience any change in migration flow within or across counties. There was also no comparable change in overall, primary or tertiary sector employment. The only notable difference was that counties near the Yangtze River had a higher share of manufacturing employment. This could reflect secular changes in industry composition as China transitioned from a predominantly agricultural and centrally-planned economy in the 1980s to a more market-based economy since the 1990s. Most importantly, across all placebo samples, changes in overall and manufacturing employment were either marginally significant or slightly positive, indicating that the main results provide a lower bound for the negative labor market impacts of the TGP.

This paper contributes to three strands of literature. The first is on the social and economic impact of infrastructure development projects. Most previous studies have focused on the benefits of infrastructure, such as telecommunication (Roller and Waverman 2001), transportation (Donaldson and Hornbeck 2016; Donaldson 2018), and electrification (Dinkelman 2011; Lipscomb, Mobarak and Barham 2013). But the social costs and spatial distribution of economic impacts are largely unknown. Dams represent a particularly interesting case where the equity and efficiency tradeoffs can be more closely examined. A seminal paper by Duflo and Pande (2007) shows that dam construction in India benefits districts downstream at the expense of those upstream and in the vicinity of the dam.

Building upon these earlier results, this paper is different along several key dimensions. First, I examine the impact of one major natural experiment from the largest hydroelectric dam in the world, whereas Duflo and Pande (2007) explore the construction of over 2,000 mid-sized irrigation dams over a thirty-year period. The different settings necessitate a departure in identification strategies, where I use exogenous variations induced by the government's relocation policy, whereas they use variations induced by differences in river gradient. Second, Duflo and Pande (2007) conduct analysis at the aggregated district level, whereas I use individual- and firm-level data for county-level analysis. There were 466 districts in India in 1991, compared to 2,460 counties in China in my sample. Access to micro-level census and satellite data allows me to precisely measure the inundation intensity for each county and consider detailed population displacement patterns, both of which are not possible with district-level data. Lastly, the impact of irrigation dams is very different from that of multipurpose mega-dams as the TGD. While it is reasonable in their setting to assume that labor and capital are immobile when considering agricultural production and poverty outcomes, in the case of the TGD, the displacement of labor and capital have substantial and long-lasting impacts on the local economy. As such, this paper presents the first micro-level evidence of the social costs of dam-induced inundation.

The research design of this paper relates to the broader literature using exogenous shocks as push factors for migration outcomes. Previous studies have employed a variety of natural experiments for identification, including civil conflicts, political regime shifts, policy shocks, and natural disasters (Bazzi et al. 2016; Boustan, Kahn and Rhode 2012; De Silva et al. 2010; Deryugina, Kawano and Levitt 2018; Friedberg 2001; Hornbeck 2012; Hornbeck and Naidu 2014; McIntosh 2008; Ruiz and Vargas-Silva 2015; Sarvimäki, Uusitalo and Jäntti 2010). Dam-induced flooding represents a unique setting, where the displacement is permanent and there are no never-takers or selective return. As a result, the estimates in this paper can be generalized to shed light on the potential effect of irreversible flooding in other settings, such as sea-level rise caused by climate change.

Finally, the paper is related to previous studies that examine the distributional labor market impacts of exogenous shocks, particularly adjustment costs of reallocating workers across sectors and industries. I provide new evidence to show that in addition to trade exposure (Autor, Dorn and Hanson 2013; Autor et al. 2014) and environmental regulation (Greenstone 2002; Walker 2013), infrastructure projects, though typically viewed as beneficial to economic development, can impose large and long-lasting costs on the local economy. This paper also offers a new angle to examine distributional consequences beyond demographic and socioeconomic dimensions – the spatial distribution of costs and benefits is an important consideration when evaluating government policies.

The remainder of the paper is organized as follows: Section 2 highlights the quasi-experimental nature of the TGP and describes the government's relocation policy. Section 3 documents the data sources and measurement. Section 4 discusses the research design and provides evidence validating the identification assumptions. Section 5 presents the main results on population displacement, local labor market adjustments in the short and long run, and heterogeneous impacts on migrants and nonmigrants. In Sections 6 and 7, I show the results are robust to falsification exercises and explore potential mechanisms by looking at firm-level adjustments. Section 8 concludes.

2 Historical Background

2.1 Approval and Siting of the Three Gorges Dam

The Three Gorges Dam (hereafter TGD or TGP for Three Gorges Project) is the largest hydroelectric power plant and one of the most controversial infrastructure development projects in the world. Situated along the longest river in Asia, the Yangtze, TGD has a total installed capacity of 22,500 MW, about 8 percent of the installed hydropower capacity and

2 percent of the total installed electricity generating capacity in China as of 2013 (CTG 2013*b*; EIA 2015). Similar to other large-scale development projects in China, throughout the long deliberation and eventual approval process by the National People's Congress (NPC) in 1992, decision-making was highly centralized with little participation by local stakeholders. The idea of building a dam at the Three Gorges section of the Yangtze river was first proposed by Sun Yat-Sen in 1919. After China regained independence in 1949, subsequent leaders Mao Zedong and Deng Xiaoping reopened the case to examine its feasibility. An extensive geological survey that lasted 24 years was conducted to compare two possible dam regions, and the final location for TGD was selected at Sandouping in 1979, due to its solid granite riverbed that was earthquake and seepage safe (CTG 2010*b*; Highland 2008). The siting and approval process highlights TGP's exogenous nature.

Hailed by the government and proponents as an engineering marvel, the project was justified on the grounds of flood control, power generation, and river navigation (State Council 1992). Historically, the middle and lower reaches of the Yangtze River downstream of TGD have experienced devastating floods that affect millions of residents and stifle economic activity in one of the main industrial hubs of China (CTG 2010*a*; Yu et al. 2009). By regulating water flow, reducing its speed, and raising the water level in the reservoir, the dam promises to decrease the frequency of downstream flooding from once every decade to once every century and allow larger vessels to navigate the Three Gorges section of the Yangtze, which is famous for its treacherous waters. In addition, TGD heralds an era of dominance by hydropower as the leading renewable energy source in China, accounting for around 20 percent of total electricity generation in 2015 (IHA 2016). It generates power that fuels the rising electricity demand in a large area encompassing Central, East and South China, and contributes to the nation's emission reduction targets. As of 2017, TGD is estimated to have abated 319 million tonnes of coal and 858 million tonnes of carbon dioxide, which is about half of the annual electricity sector emissions in the United States (EIA 2018; Xinhua 2017).

After 17 years of construction and an official price tag of $36 billion,[5] however, much uncertainty remains about whether TGD is the crown jewel of infrastructure development projects in China or an impending disaster threatening the livelihoods of millions. Crit-

[5]The National Audit Office (2013) released its final project cost calculations for the TGP in 2013. The total cost is 248.5 billion CNY (36 billion USD) and much higher than initial projections and interim estimates, which are around $23 billion.

ics have pointed to the myriad environmental, socioeconomic and political issues associated with the project (Fearnside 1988; Gleick 2009; Webber 2012). Among the many cited environmental impacts are biodiversity loss of endangered fish species, habitat destruction of flora and fauna in the region, irreversible changes to the hydrology, water quality, sediment regime and geology of the Yangtze River, and landslide and seismic activities triggered by the rapid changes in water levels in the reservoir (Stone 2008; Tullos 2009).

2.2 Inundation and Involuntary Displacement

The most controversial and detrimental aspect of the TGP is the involuntary displacement of residents from villages and towns inundated by the reservoir. When fully operational in 2010, the dam's reservoir reached a water level of 175 meters above sea level and covered an area[6] of 1045 km^2, of which 600 km^2 is land. The inundated land spans 20 counties or municipal districts, 277 towns, 1,680 villages, and 26,000 hectares of farmland (State Council 2000, 2006; Xinhua 2000*a*). It is also home to 1,599 industrial and mining firms, of which 98 percent are small enterprises with less than 300 workers (CTG 2002; NBS 2006). While the affected area is geographically confined, the ensuing forced migration flow is unprecedented in magnitude because river valleys of the Yangtze River and its tributaries are densely populated.

There is a lack of definitive statistics on the actual number of forced migrants. To the best of my knowledge, the government did not release a detailed record for the magnitude, origin and destination of the displacement flow throughout the 17-year construction period. Estimates vary considerably, depending on the source and time frame considered (Duan and Steil 2003; Li and Rees 2000; Wilmsen, Webber and Duan 2011*b*). In 1992, the NPC approved 725,500 migrants to be relocated (State Council 1992). At the end of the official resettlement program in 2010, the Three Gorges Project Construction Committee (TGPCC) estimated that 1.4 million people moved between 1995 and 2010 (NetEase 2010; Xinhua 2006). But recent projections add another 100,000 migrants as TGD-induced environmental hazards, such as landslides, may render the surrounding areas uninhabitable (Wee 2012).

The government used a simple and transparent rule to classify residents as part of the displaced population. All residents below the 175m flood line were relocated at some

[6]This is the government's *ex-ante* estimate of the inundated area using the maximum water level and elevation map. As I will show in section 3.1.2, this *ex-ante* estimate is very close to my *ex-post* calculation using satellite data, which shows an actual inudated area of around 1035 km^2.

point between 1995 and 2010, and those living at lower elevations were ordered to leave earlier. In Figure 1, I document the official relocation program timeline using public information on the websites of TGPCC and China Three Gorges Corporation (CTG), news articles, and government legislative documents. From the initial proposal in 1992 to the final completion in 2012, the TGP progressed in three phases that marked the four waves of population displacement. Relocation was concurrent to dam construction and often preceded major project milestones to clear the way for the rising water level. In Phase 1 (1992-1997), the official relocation program was kicked off in 1995 with several hundred migrants moving out to pilot receiving sites. Relocation began in earnest in 1996 to prepare for the diversion of Yangtze River a year later in 1997. In Phase 2 (1998-2003), relocation was in full force with all residents living below 135m ordered to move by 2003, when the reservoir was filled for the first time. In Phase 3 (2004-2010), water level gradually increased to 156m in 2006 and eventually reached the target level of 175m in 2010, when the official relocation program ended.

[Figure 1 about here.]

State Council (1993, 2001) legislations stipulate that migrants should be relocated to nearby areas through a combination of government- and self-directed efforts. Residents are classified into three types of migrants based on destination: within the same county, across counties in the same province, and across provinces. According to historical accounts, almost 90 percent of migrants remained in the affected provinces in Chongqing and Hubei by moving to higher elevations in the same county or to another nearby county (NetEase 2010; Xinhua 2000*b*). The displaced were promised compensation for the loss of their home and income, but in most cases it was either inadequate to restore their standard of living or never materialized (Duan and Steil 2003; Li and Rees 2000; Wilmsen, Webber and Duan 2011*a*,*b*). About 45 percent of the initial budget of 90 billion CNY or $6 billion was designated for population resettlement (CTG 2013*a*; Xinhua 2003). But the compensation scheme was vaguely described in the legislation and poorly implemented by city and county officials. The process was marred by corruption and lack of transparency and effective oversight.[7] As a result, while the majority of migrants were within the homeland, they were often left without stable housing or income source - their former way of

[7] For instance, the National Audit Office (2013) found that $40 million of the resettlement fund was misappropriated.

life irreparably lost.

3 Data Construction

I construct a comprehensive dataset to study the social costs of the TGP, including satellite data on elevation and inundation extent, individual-, establishment- and county-level census, historical administrative boundary shapefiles, hardcopy township-level census tabulations, archival government legislations, and geo-referenced data on dams. The goal is to calculate the inundated area, document the forced migration flow, and estimate the causal impact of inundation and displacement on economic development in the surrounding communities.

3.1 Satellite Data

3.1.1 Minimum Elevation

I use Digital Elevation Map (DEM) data from the NASA Shutter Radar Topographic Mission (SRTM). The data was originally acquired in 2000 with a resolution of 3 arc-seconds or 90 meters. I use the Version 4 resampled data to 250 meters available for the entire globe from the Consortium for Spatial Information (CGIAR-CSI).[8] SRTM is considered the highest-resolution topographic data and a "major breakthrough in digital mapping of the world" by filling data voids (NASA, 2017; CGIAR-CSI, 2017). I describe in detail the procedures for calculating the minimum county elevation based on raw raster SRTM data in the Appendix.

Figure 2 shows the minimum elevation for counties along the Yangtze River, with the color gradient representing water levels during the four waves of TGD-induced forced migration: 1-135m (Waves 1 & 2: 1995-2003), 135-156m (Wave 3: 2004-2006) and 157-175m (Wave 4: 2007-2010). As evident in the figure, most counties along the Yangtze River have minimum elevation below 156m, implying the majority of affected residents have moved before 2006. There is also a gradual decline in elevation from the western Tibetan plateau to the eastern coastal seaboard, with most counties to the east of TGD lying below 175m. Since only counties upstream of the dam are inundated, I differentiate between counties upstream and downstream using geo-coordinates of their centroids.

[Figure 2 about here.]

[8] Available here: `http://srtm.csi.cgiar.org/`

3.1.2 Inundation Map

I create an intensity of treatment measure by using Landsat 5 TM satellite data and remote sensing techniques to calculate the fraction of county area inundated by TGD's reservoir.[9] I select 12 spatially contiguous scenes, each covering an area of approximately 106 miles by 115 miles, that together span the reservoir of TGD in Hubei and Chongqing provinces. I pre-process each scene and then mosaic them together to get a raster image of only the Yangtze River for the time periods 1990s (pre-TGD) and 2010s (post-TGD). In the Appendix, I list the individual Landsat scenes and describe the procedure for creating the inundation map.

Figure 3 shows changes in Yangtze River extent between the 1990s and 2010s near the TGD, overlaid with county boundaries in 2000.[10] The river channel upstream of TGD has clearly expanded due to elevated water levels in the reservoir. Using the change detection algorithm in ENVI, I calculate that the section of the Yangtze River in Hubei and Chongqing covered an area of about 713 km^2 before TGD construction in 1990s. The area expanded to 1035 km^2 after construction was completed in 2010s, a change of about 322 km^2. This is likely a conservative estimate of the actual inundation area, because tributaries and smaller streams connected to the main channel of Yangtze River are excluded.

[Figure 3 about here.]

3.2 Administrative Data

3.2.1 Individual-Level Population Census

The research design is made possible by access to individual-level census data that contains detailed migration information. I use microdata from the 1 percent random sample of the 2000 Population Census and the 15 percent sample of the 2005 1 percent Population Survey to construct a county-level panel of outmigrants between 1995 and 2005.[11] The

[9] Available through the U.S. Geological Survey (USGS) Global Visualization Viewer here: `https://glovis.usgs.gov/`

[10] I check that there were few county boundary changes in this region between 1990 and 2010 that might be endogenous to dam construction. The most important change was the creation of Chongqing, formerly a sub-provincial city in Sichuan, as a direct-controlled municipality in 1997. This is mainly a change in administrative jurisdiction with the county boundaries largely intact. A map of boundary changes is shown in Appendix Figure A2.

[11] The 2000 data is the first decennial census that asks retrospective questions on previous place of permanent residence and year of migration. To the best of my knowledge, individual-level data of the 2010 census

main advantage of these microdata is that it allows me to map out the bilateral migration matrix between all *county* pairs in China and account for the spatial patterns of migration in a rigorous manner. This would not be possible with aggregated county, prefecture or province level census commonly used in the literature, because they only report broad origin types such as "within county," "across county" and "across province" that mask much of the variation in intra-provincial migration flow. As I show in Section 4, information on the fine-scale migration patterns is crucial for validating the identification assumptions and informing the research design.

I calculate origin-based outmigration flow using variables in the census that retrospectively record the origin and destination counties, migration reason, and migration year up to five years prior to the census. The corresponding census questions are listed in the Appendix. An individual is considered an outmigrant if 1) he/she has moved within the past 5 years; 2) the move is to another town in the same county, another county in the same province or another province; 3) the place of origin is not missing.

To fully account for the multi-dimensional spatial patterns of migration, I calculate the number of migrants from any sending county to their neighbors as follows: 1) From the individual-level census, I first calculate the number of outmigrants between all sending and receiving county pairs by the year of migration; 2) I then classify the pairwise neighbor relationship between all 2,876 counties in China by constructing a file of length 8,271,376 with origin county, residence county and dummy variables indicating whether two counties are first degree, second degree, third degree, and up to the tenth degree neighbors; 3) I merge these two files and collapse by county of origin and year of migration to get a county-year panel for the number of out-migrants from each origin county to their neighboring county of degree i.

Figure 4 demonstrates the nearest neighbor classification for one county, Zigui, where the TGD is located. Using this procedure, I am able to assign neighbor relationships to more than 92 percent of origin-destination pairs in the census. I exclude 25,274 migrants who are unmatched due to missing or misreported origin information. I also exclude Hongkong, Taiwai, Macau, Nansha Island, Xisha Island, and Zhongsha Island that are not part of mainland China.

[Figure 4 about here.]

has not been released to researchers. As a result, I can only construct origin-based migration spells between 1995 and 2005.

While access to microdata allows me to calculate the migration flow between all origin-destination county pairs and account for the multi-dimensional spatial patterns of migration in a rigorous manner, several caveats should be noted. First, the earliest migration spell that can be constructed from the data is 1995. Individuals who have moved prior to November 1, 1995 are considered nonmigrants, because the census does not report their place of origin. Since the official resettlement program started in 1995, excluding those who moved before 1995 will not understate the magnitude of TGD-induced migration flow. Second, the census only observes single migration spells and provides limited information on step migration. Moreover, while return migrants are recorded in the 2000 census, they are considered nonmigrants in the 2005 census. In the Appendix, I discuss these issues and provide a detailed procedure to adjust for the systematic undercounting of return migrants in the 2005 census.

3.2.2 County-Level Population Census

In addition to microdata in 2000 and 2005, I harmonize census data in 1990 and 2010 to create a county-level panel between 1990 and 2010 for the main analysis. Historical county-level data in 1982 was used to check for pre-trends, but it contains a very limited number of variables and is excluded from the main analysis. To check for pre-treatment characteristics of people living near the Yangtze River, I collect hardcopy 1990 census data with township identifiers from archival records and manually geocode the location of towns in Chongqing and Hubei. To the best of my knowledge, microdata in 1990 does not contain corresponding township identifier information.

Two main challenges underlie the data harmonization process. First, county names, codes and boundaries have changed considerably over time.[12] Second, the census questionnaires have been modified over time, such that categories are added or subtracted for existing variables and new variables are created. For geographic comparability, I manually harmonize and construct crosswalk relationship files for all counties in China for the 1990, 2000, 2005 and 2010 census. I adjust the data to hold 1990 geographic boundaries constant for backward consistency. The main sample consists of 2,460 counties with consistent county geography between 1990 and 2010. For variable consistency, I identify a

[12] In the Appendix, I identify five main types of changes in county geography: 1) GB code only; 2) Name only; 3) Both GB code and name; 4) Redistrict, including one county splitting into multiple smaller counties, multiple counties aggregating into one larger county, and other idiosyncratic redrawing of boundaries; 5) New District. Figure A1 shows the number of counties in each census year and those that have experienced any changes in the above categories.

comparable set of questions across census years and recode them similar to the IPUMS-International data harmonization procedure. In the Appendix, I list and describe the harmonized industry and occupation categories, which are of particular interest for the labor market analysis.

4 Empirical Strategy

4.1 Treatment Assignment

A major challenge for identifying the causal impact of population displacement is that even in settings where the shocks are plausibly exogenous, such as civil conflicts and natural disasters, there may still be selection into who moves away and returns. In light of the government's relocation policy, I propose a novel identification strategy that uses a key engineering feature of dams, the Maximum Design Water Level (MDWL),[13] to assign treatment. Individuals living near the reservoir at elevations below the MDWL will be inundated and forced to move out. In the ideal experiment, I would classify individuals to be treated by TGP using the exact longitude and latitude of their homes and a continuous measure of topography to compare those living just below and above the water level. But county is the lowest level of geography reported in the census for the place of origin. Therefore, I assign treatment at the county level using its minimum elevation and actual changes in the flooded area before and after TGD construction calculated from satellite imageries. As I will show, a county's ex-ante minimum elevation accurately predicts its ex-post changes in inundated area.

A county i in province j is considered treated by the TGD if it satisfies all three criteria:

C1. County minimum elevation ≤ 175 meters

C2. Change in flooded area between 1990s and 2010s $\neq 0$

C3. Upstream of the TGD

C1 is an extensive margin measure for whether any part of the county is below the reservoir water level. C2 complements C1 by ensuring that counties are not misclassified as treated due to rugged terrain. For instance, a valley lying in between mountains will not actually be inundated by the TGD. C3 reflects the fact that only counties upstream of the

[13]The MDWL is defined as "the highest elevation of water determined as a result of safely passing the Inflow Design Flood" by U.S. Fish and Wildlife Service (2003), Part 361 Dam Safety regulations.

dam are inundated. C2 in itself is not sufficient because while sections of the Yangtze beyond the reservoir of TGD also exhibit very small changes in river area, their minimum elevation is above 175m and therefore not affected. To get a continuous flooding intensity measure, I interact this treatment dummy with the fraction of county area covered by the Yangtze River in the 2010s. Figure 5 displays the 25 treated counties using these criteria, overlaid with the proportion of county area inundated in 2010. I verify that the treatment assignment is consistent with official documentation and captures all counties with any reported displaced population (Xinhua 2000*b*).[14]

[Figure 5 about here.]

4.2 Model Specification

The empirical specification uses a difference-in-differences model to compare changes in migration and labor market outcomes between inundated and nearby non-inundated counties in the same province and with similar pre-TGP characteristics, relative to the baseline in 1990 or 1995. As I will show, spatial spillover of the displaced population is confined within the treatment group. As a result, nearby non-inundated counties in the same province with similar pre-trend in outcomes constitute the proper control group.

For each outcome variable Y of county i in province j and year t, I estimate the following model on the fraction of county area flooded in 2010, county fixed effects and province-by-year fixed effects:

$$Y_{it} = \alpha_i + \beta_t Inundated_i + \gamma_{jt} + \varepsilon_{it} \tag{1}$$

The coefficient of interest is β_t, which varies by year and represents the average differences between inundated and nearby non-inundated counties in each time period relative to the base year, which is 1990 for labor market analysis and 1995 for migration analysis. As documented earlier, 1995 serves as a reasonable base year for analyzing migration outcomes, because water level in the reservoir was still at its pre-TGD level and the relocation program wasn't in full swing until 1996.

[14] The list of inundated counties is available through the Chongqing Resettlement Affairs Bureau of Three Gorges here [in Chinese]: `http://www.cq.xinhuanet.com/sanxia/yimin/ymgk.htm`. One county, Wuxi, is excluded from the treated because it doesn't show a change in flooded area. While most official inundation maps include Wuxi, it only had 520 TGD-induced migrants, considerably smaller than the rest of the inundated counties. In addition, these migrants moved toward the end of the construction process in 2008, which is beyond my sample period. Government report (CTG, 2009) available here [in Chinese]: `http://root.tgic.cn/xwzx/news.php?mnewsid=39443`.

A comprehensive set of control variables are included to account for possible differences between inundated and non-inundated counties. County fixed effects α_i absorb time-invariant county characteristics, such as distance to the Yangtze River, terrain, climate, longitude, and latitude. Province-by-year fixed effects γ_{jt} account for changes that happen smoothly over time in each province that may be correlated with the construction of TGD, such as fertility control policies and secular trends in industrialization and urbanization. Robust standard errors are clustered at the county level to allow for heteroskedasticity and serial correlation within counties over time.

4.3 Identification Assumptions

Identification of the causal impact of the TGP hinges on the following assumptions:

1. Inundation is not endogenous to unobserved characteristics of the county
2. Non-inundated counties in the same province serve as the proper control group

Assumption 1 requires that assignment to inundation is plausibly exogenous and not correlated with other unobserved county characteristics. As documented in Section 2, inundation and displacement are largely determined by the elevation of individual's homes relative to the targeted water level of 175 meters above sea level in the reservoir. Figure 6 illustrates variations in the extent of TGD-induced inundation for counties within 30 km of the Yangtze River.[15] There is a strong negative relationship between the change in flooded areas from the 1990s (pre-TGD) to 2010s (post-TGD) and the minimum elevation for a given county: counties with lower minimum elevations experience the greatest change in flooded area and there is a clear cutoff at 175 meters. In addition, Figure 7 shows that for treated counties, there is a strong positive correlation between the fraction of county area flooded in the 2010s and the fraction of county area with elevation below 175m. This shows the ex-ante elevation of individuals' homes relative to 175m is a strong predictor of the likelihood their homes are flooded and they are forced to move out ex-post.

[Figure 6 about here.]

[Figure 7 about here.]

[15] For this figure, the sample is restricted to counties within 30 kilometers of the Yangtze River for readability. Including the full sample merely adds a long right tail of counties with very little change in flooded area.

Table 1 reports baseline differences in 1990 county characteristics between inundated (N = 24) and nearby non-inundated (N = 103) counties in Chongqing and Hubei.[16] Treated counties and their neighboring control counties are very similar across a comprehensive set of variables on demographics, educational attainment, sectoral and occupational employment, and migration reason. The only notable difference is that residents in treated counties are about two years older and have a slightly higher literacy rate. But the magnitude of such difference in literacy rate is small and not reflected in other educational attainment or labor market outcomes.

[Table 1 about here.]

To examine possible selection of migrants within treated counties, I use township-level 1990 census data from hardcopy tabulations and estimate differences between people living closer and farther from the Yangtze River. While this historical dataset contains a very limited number of variables, it is the only data source with township identifiers. Table 2 shows that people living in towns closer to the river in treated counties are slightly different and tend to be more urban, compared to those farther away. To the extent that people with urban Hukou tend to have higher socioeconomic status than those with rural Hukou, the displaced population may be positively selected within flooded counties. This will bias the negative impact of forced displacement downward.

[Table 2 about here.]

Assumption 2 has a few testable implications. First, before the construction of TGD, inundated and non-inundated counties in the same province should have similar pre-trends in the outcome variables. I test this by harmonizing and combining historical county-level decennial census data in 1982 and 1990. To account for changes in administrative boundaries, I adjust the historical census to have consistent 1990 geographic boundaries by manually harmonizing and creating a cross-walk file for all counties in China.[17] As shown in Table 3, inundated and nearby non-inundated counties show similar pre-trends across a range of demographic and socioeconomic variables, including the share of males, sex ratio, middle school and college graduates, employment overall, and employment in agriculture. Unfortunately, retrospective questions on migration history were not asked in the

[16] One treated county, Dadukou, is excluded because there is no observation in 1990.

[17] In Section 3 and the Appendix, I discuss details of the administrative boundary harmonization process.

census until 2000. As a result, I am not able to include migration variables in the pre-trend test.

Second, identification requires that non-inundated counties are not themselves treated, such that there should be little spillover effect of people moving from the treated to control counties. I test this directly by estimating the migration flow from treated counties to their neighbors of degree 1, 2, 3, and so forth. In Table 5, I find little evidence for spillover effects to non-inundated counties: The majority of the displaced moved within the county, some moved to another nearby *treated* neighbor and few moved to an *untreated* neighbor or to another province. Since spatial spillover of the displaced population is confined within the treatment group, nearby non-inundated counties in the same province with similar pre-trend in outcomes constitute the proper control group.

[Table 3 about here.]

Third, the empirical strategy assumes that in the absence of TGP, inundated and non-inundated counties in the same province and with similar pre-TGP outcomes would have evolved similarly over time. In other words, there are no unobserved changes at the county level that coincide with the timing of TGP and also affect migration and labor market outcomes. I conducted a comprehensive database search of government legislative documents and policy memos during the period of 1990-2010 and did not find other programs that may confound the effects of TGP.[18] In addition, I identified 39 dams and reservoirs in Hubei and Chongqing provinces from the Global Dam Tracker (GDAT) database (Zhang, Urpelainen and Schlenker 2018). Among these, only one small reservoir, Xinqiao, was situated upstream and near the reservoir of TGD and its construction was completed in 1991. This provides further evidence that there are no other major dam projects underway during the same period that may confound the effects of TGD.

5 Main Results

I begin by documenting the effect of inundation on outmigration patterns in treated counties between 1995 and 2005 by type of destination and reason of move. I then estimate the short- and long-run impact of inundation on the local labor market, focusing specifically

[18] I used the Policies and Laws of China (PLOC) database on Wanfang Data, which is an affiliate of the Chinese Ministry of Science & Technology. Available at: `http://c.g.wanfangdata.com.cn.ezproxy.cul.columbia.edu/Claw.aspx`.

on the transitional costs of reallocating the labor force. I also separately examine the heterogeneous effects on migrants and nonmigrants.

5.1 Population Displacement

Figure 8a shows the average number of outmigrants for each year between 1995 and 2005 in treated and control counties. Figure 8b separates this into the fraction of outmigrants by three destination types: within-county, across-county in the same province, and across-province. Throughout the analysis, I restrict the sample to two provinces affected by TGD, Chongqing and Hubei, and counties that have migrant observations for each year between 1995 and 2005. Treated and control counties had similar numbers of outmigrants in 1995 and 1996 before water level in the reservoir was first elevated from 66m to 88m when the Yangtze River was diverted in 1998. This provides justification for using 1995 as the baseline for estimating migration flow changes in inundated counties. TGD-induced population displacement started to pick up in 1997 and the gap in outmigration flow between treated and control counties increased gradually.

[Figure 8 about here.]

Several patterns emerge when I examine the direction and reason of migration in Figure 9. First, most TGD-induced migrants moved within the same county, followed by across counties in the same province. Second, among across-county migrants, the outmigration flow was almost entirely directed to another *treated* first or second degree neighbor. Third, inundated counties had more migrants moving out due to house demolition and the temporal patterns were consistent with that of within-county migration. Fourth, across-province migration was mostly driven by people moving due to work or business. As such, there was a relative decline in the share of migrants moving across provinces in inundated counties.

[Figure 9 about here.]

Table 4 reports the migration flow in more detail by three destination types: within-county, across-county in the same province or across-province, and shows estimated differences in the log fraction of outmigrants between inundated and non-inundated counties relative to the omitted base year in 1995. Counties that had 1 percent of their land area inundated witnessed an average 0.2 percentage point increase in within-county migration for each year between 1996 and 2005. Peaks in within-county migration flow clearly match the TGP construction timeline. The largest number of migrants moved to another neighborhood or town in the same county in 1998 and 2003, which corresponded to the beginning of Phase 2 of the TGP when the resettlement program was in full swing and the filling of the reservoir to its interim level of 135m respectively. The increase in across-county migrant share was milder at about 0.18 percentage points per year and the magnitude got slightly larger over time. This is in agreement with the government's resettlement policy, which stipulates that across-county moves should only be considered if migrants couldn't be relocated to the same county. There was also a relative decline in the share of across-province migration in treated counties, because the majority of TGD-induced migrants remained in the affected provinces. In terms of magnitude, I estimate that between 1996 and 2005, inundated counties had 998,556 more within- and across-county migrants than control counties relative to the baseline in 1995. This implies that between 1995 and 2010, the total number of people directly displaced by TGP-induced flooding was likely to be between 1.5 and 1.9 million, considerably higher than official estimates of 1.4 million.

[Table 4 about here.]

To account for the finer spatial patterns of across-county moves, I examine the migration flow from inundated counties to their neighbors of up to the sixth degree.[19] I also break this down to see if the neighbors themselves are inundated. Interestingly, the estimated differences in across-county migration in inundated counties were almost entirely driven by outmigration to treated first degree neighbors shown in Figure 10 and Table 5. To provide further evidence that the estimated increase in outmigration is most likely due to the construction of TGD, in Figure 11 and Table 6 I find that inundated counties had on average 0.2 percentage point more house demolition-related migrants per year and the temporal patterns closely mirror that of within-county migration flows. In contrast, inundated counties had less migrants who moved for work or business related reasons. There is also no statistically significant change in migration due to other reasons, including job change, entry-level hire, dependents, study/training, marriage or join relatives. This provides further evidence that the increase in migration flow in treated counties is directly

[19]Results for higher order neighbors are generally very small and not statistically significant due to the very small number of observations between treated counties and these higher order neighbors.

due to the construction of TGD.

[Figure 10 about here.]

[Table 5 about here.]

[Figure 11 about here.]

[Table 6 about here.]

5.2 Local Labor Market - Sectoral and Occupational Employment

In addition to population displacement, TGD has wide-ranging impacts on the local economy by flooding farmland, transportation infrastructure, and firms. Having established that TGD has led to a significant increase in the number of migrants within the affected provinces, in this section I explore the subsequent local labor market impacts of inundation and focus particularly on changes in employment overall and by industry and occupation.

In Figure 12, I harmonize individual-level census data in 1990, 2000, 2005 and county-level census in 2010 and aggregate them into the county level to estimate changes in overall employment by the share of county area inundated, relative to the baseline in 1990. The sample includes all residents of a given county at census time, irrespective of their migration status. From the pre-trend results earlier in Table 3, the share of population employed has been changing similarly between inundated and non-inundated counties prior to the construction of TGD in 1982–1990. Starting from 1990, however, overall employment in inundated counties followed a U-shaped pattern, in which it first declined significantly between 1990 and 2000 but then slowly reverted back to the baseline level by 2010. For an average treated county with 2 percent land area flooded, the share of working-age population employed declined by 1.2 percentage points in 2000 and 0.7 percentage points in 2005. In the long run, there was a gradual convergence in employment between inundated and non-inundated counties and by 2010, employment had reverted back to the baseline level. This U-shaped pattern suggests that in the short run, the TGP may have reduced employment by flooding farmland, destroying physical capital, and forcing firms to relocate. Over time, this detrimental employment effect may have been partly offset by the creation of new job opportunities and a sectoral reallocation of the labor force across industries and occupations.

[Figure 12 about here.]

To explore further the extent of this labor market reallocation, I examine sectoral and occupational employment in Figure 13. I harmonize and classify 20 industries and 6 occupations across the 1990, 2000, 2005 and 2010 census into primary, secondary and tertiary sectors.[20] Due to the loss of land and physical capital, there was a large and persistent decline in the share of manufacturing employment by about 2-3 percentage points in inundated counties. Accompanied by this large and significant loss of manufacturing jobs, there is suggestive evidence for a moderate gain in the share of agricultural employment. However, the coefficients are smaller at around 1.5 percentage points and not significant at the 5 percent level. In addition, there is no significant change in agriculture or mining employment share when I separately examine industry-specific employment in the primary sector. The results also show no change in the share of tertiary sector employment, providing evidence against a process of "creative destruction," through which the destruction of land and physical capital spurs a structural transformation in the local economy by creating new technologies. On the contrary, the closure of firms and factories during the TGP resettlement process appears to have forced local residents to seek job in agriculture and other low-skill sectors and resulted in a "reverse structural transformation" that reallocates labor from the secondary to the primary sector.

[Figure 13 about here.]

In Figure 14, I take a more detailed look at the heterogeneous employment effects across occupations. Consistent with the sectoral analysis, individuals who are employed as "Production and Transportation" workers are the most adversely affected. Counties with 1 percent area flooded experienced a large and persistent decline in the share of production and transportation workers by about 2-3 percentage points. While there was a slight increase in the share of service and administrative workers, the gains were an order of magnitude smaller at around 0.5-0.9 percentage points and far from enough to offset the loss of manufacturing and transportation related jobs. Overall, these results suggest TGD-induced inundation has not only led to a spatial relocation of the local population within flooded counties, but also contributed to a significant loss of manufacturing employment and a broader sectoral reallocation of the labor force.

[20] Please see Appendix Tables A1 and A2 for a detailed breakdown of the industries, occupations and their corresponding sectors.

[Figure 14 about here.]

5.3 Labor Market Impact on Migrants

In this section, I examine the transitional costs of reallocating the labor force by focusing on those displaced. I first show that migrants from flooded counties are comparable in observable characteristics to those from non-flooded nearby counties. A potential threat to identifying the causal impact of forced displacement is that the composition of migrants may be changing differently in inundated and non-inundated counties. To examine the composition effect, I estimate equation 1 with pre-determined characteristics of migrants as the dependent variables to compare the differences between inundated and non-inundated counties relative to 1995. In particular, I consider the share of migrants moving in each year between 1996 and 2005, who are 1) of Han ethnicity; 2) rural Hukou holders 35 and above; 3) literate and 18 and above. I also consider 4) the average years of schooling for those 18 and above and 5) age at first marriage for those between 35 and 60.

This exercise relies on the assumption that "pre-determined" characteristics of migrants are unlikely to be changed by the construction of TGD. For educational attainment, since the average years of schooling in treated counties is only around 9 years, it is reasonable to assume that most people would have completed schooling by age 18. For Hukou status and age at first marriage, an important consideration is that TGD may affect the marriage market outcomes in inundated counties. Therefore, I restrict the sample to those 35 and above because 90 percent of individuals first got married before 27 in the 2000 census and 99 percent did so before age 34. This implies that marriage-related Hukou status change is likely to be completed by age 35. For age at first marriage, I exclude those who report an age beyond 60 due to possible reporting errors. There are only 1,815 such cases in the 2000 census and more than 93 percent of individuals are nonmigrants.

Table 7 shows that there was little change in the composition of pre-determined demographics among migrants in inundated counties. Migrants from inundated and non-inundated counties had similar educational attainment, literacy, Hukou status and ethnicity. There is a slight increase in the age at first marriage, but it is small in magnitude and only statistically significant in 1998 and 2003. These results provide suggestive evidence that migrants from flooded counties are comparable in observable characteristics to those from non-flooded nearby counties.

[Table 7 about here.]

To understand the direct impact of inundation and displacement on migrants in flooded counties, I create a retrospective panel using micro-level census in 2000 and 2005 to compare the labor market outcomes for migrants that moved in each year between 1996 and 2005, relative to those who moved in 1995. Since the outcome variables are measured in the census years, this exercise is essentially capturing the differential effects of displacement on migrants who moved in different waves of the TGP relocation process. It is important to note that in the absence of longitudinal data, there may still be unobserved heterogeneity in individual characteristics, such that migrants in flooded counties may be a selected group. To the extent captured by educational attainment and hukou status, Table 7 shows that the composition of TGD-induced migrants does not appear to be systematically different compared to that in nearby control counties. In addition, township-level analysis in Table 2 shows that people living in towns closer to the Yangtze River in flooded counties tend to be more urban. Since people with urban Hukou tend to have higher socioeconomic status than those with rural Hukou, selection of migrants in flooded counties, if any, is likely positive. This will bias the negative impact of forced displacement downward.

With these caveats in mind, Table 8 reports changes in the employment status of migrants in flooded and non-flooded counties. The dependent variables are the log fraction of migrants unemployed overall and by job searching status. Flooded counties experienced a significant increase in the share of unemployed migrants and those that moved in 2003 had the highest unemployment rate. Since 2003 was when water level in the reservoir was elevated to 135m, this provides suggestive evidence that those displaced by the TGD lost their jobs and remained unemployed in the short run. In terms of labor force participation status, flooded counties had significantly more migrants not in the labor force, either due to retirement or discouragement from searching.

[Table 8 about here.]

Displaced migrants had little financial assets to buffer the loss of job and house. As shown in Table 9, the unemployed largely relied on pension, family support and other informal sources of income. This is consistent with historical accounts, which documented that for most displaced families, compensation only accounted for 30 percent of the total cost of constructing a new home (Duan and Steil 2003). Survey of TGP-migrants had revealed that the majority of migrants considered the compensation received as inadequate and did little to fight for their interest (Xi and Hwang 2011). As a result, many migrant families were left with little physical or financial capital and no sustainable sources of income.

[Table 9 about here.]

5.4 Labor Market Impact on Nonmigrants

TGP-induced inundation has affected not only those displaced but also local residents at large. In Table 10, I conduct a subsample analysis wth only nonmigrants to see if the results are driven entirely by those displaced. Columns 1, 3, 5 and 7 replicate the main results for the full sample on changes in overall and sectoral employment by the share of county area flooded, relative to the baseline in 1990. Columns 2, 4, 6 and 8 show results for the subset of individuals who never moved. The subsample analysis is only extended to 2005, because individual-level data is not available in 2010.

[Table 10 about here.]

The estimated coefficients for nonmigrants are robust and essentially identical to the full sample, suggesting the impacts of inundation have reached beyond those displaced. In particular, there is a large and persistent decline in manufacturing employment by 2-3 percentage points for nonmigrants. This represents roughly one third of the mean county-level employment in that sector. A comparison of columns 1 and 2 also shows that the overall negative employment effects may be larger and more persistent for nonmovers. Overall, these results provide evidence that the closure of firms and factories during the relocation process has led to a "reverse structural transformation" in flooded counties, in which the decline of manufacturing has forced local residents to seek jobs in agriculture and other less capital-intensive industries.

In Table 11, I explore the channels through which nonmovers may have adjusted by working in other sectors or occupations. While the results for most occupation categories are similar between the full and nonmigrant-only sample, nonmigrants in flooded counties didn't experience the same gain in commercial and service jobs as the full sample of all residents. Therefore, TGP-induced inundation not only imposes large and persistent costs on the local labor market, but the effects are also more detrimental and longer lasting for non-movers, due to their more limited ability to find employment in other sectors and occupations.

[Table 11 about here.]

6 Robustness

6.1 Falsification Test

Identification of the causal impact of TGP hinges on the assumptions that non-inundated counties in Chongqing and Hubei serve as the proper control group and that in the absence of dam construction, treated and control counties would have evolved similarly over time. To provide additional evidence for the validity of the identifying assumptions, I conduct a falsification exercise to explore differential changes in migration and labor market outcomes in placebo counties.

In particular, I construct a placebo sample that includes counties near the Yangtze River upstream and outside of the catchment zone of TGD's reservoir.[21] Considering upstream rather than downstream provinces ensures that the results are comparable to the main analysis and that these counties do not directly benefit from the dam, such as through flood control. Counties in the placebo sample are considered "pseudo-flooded" if they are within 50km of the Yangtze River. To get a distribution of estimated coefficients, I also use distance cutoffs between 0 and 50 in 5km increments to generate 11 placebo samples. The number of "pseudo-flooded" counties varies from 42 to 77 depending on the distance cutoff, and 264 counties are in the placebo sample in total.

Figure 15 shows that placebo counties near the Yangtze River evolved similarly in demographics, educational attainment, overall and agricultural employment between 1982 and 1990 compared to counties farther away.[22] In the absence of TGD-induced flooding, placebo counties near the Yangtze River did not experience any changes in outmigration flow within or across counties. There is a slight increase in the share of across-province migrants since 2003, but the coefficients from most placebo samples are not significant. There is also no change in the share of migrants moving due to house demolitions.

[21] Appendix Figure A4 shows that placebo counties are in the Sichuan and Yunnan Provinces, which are neighbors to the treatment provinces, Chongqing and Hubei. I restrict the placebo sample to counties upstream of the TGD to ensure comparability with the main analysis. Two upstream provinces that the Yangtze River runs through, Tibet and Qinghai, have no migration observations during the sample period and are excluded.

[22] The only statistically significant difference is that placebo counties closer to the Yangtze River have about 1 percent more middle school graduates. However, this slight difference in education attainment does not translate into differences in overall or agricultural employment.

[Figure 15 about here.]

[Figure 16 about here.]

In addition, placebo counties experienced no comparable changes in overall, primary or tertiary sector employment as shown in Figure 17. The only notable change is that counties near the Yangtze River had a higher share of manufacturing employment. This could reflect secular changes in industry composition as China transitioned from a predominantly agricultural and centrally-planned economy in the 1980s[23] to a more market-based economy since the 1990s. Overall, placebo counties do not exhibit the same outmigration patterns or labor market reallocation as treated counties. Most importantly, across all 11 placebo samples, the change in overall and sectoral employment is either not statistically significant or slightly positive, suggesting that the estimated coefficients from flooded counties provide a lower bound for the negative labor market impacts of TGD. The falsification exercise lends further credence to the identifying assumption that inundated counties would have evolved similarly to non-inundated counties had the TGD not been built.

[Figure 17 about here.]

7 Mechanism

7.1 Firm Adjustment

How did firms respond to TGP-induced inundation and relocation? In this section, I use establishment-level data to examine firm adjustments in wage and employment in inundated counties. The data comes from the Annual Survey of Industrial Enterprises (ASIE) collected by the National Bureau of Statistics of China (NBS), and contains data on "above-scale" firms in the mining, manufacturing and utility industries. Manufacturing firms account for around 90 percent of the observations in each year. Here the "above-scale" size threshold includes all state-owned enterprises and non-state-owned enterprises with annual sales above 5 million CNY.[24] Together, firms contained in the database account for

[23] From the 1982 census, on average 76 percent of the county population was employed in agriculture, forestry and fishing.

[24] At the exchange rate of 1 USD to 6.8 CNY, this is about 733,000 USD.

around 95 percent of total industrial output in China. While the NBS has started collecting annual firm-level data as early as 1992, sample definition and firm identifiers have changed substantially prior to 1998. Previous studies have also noted coverage issues and measurement errors of post-2008 data. Therefore, I collapse the data into a county-level panel and restrict the analysis to 1998-2007.

This exercise relies on the assumptions that 1998 serve as the proper baseline and that in the absence of TGD, firms in treated and control counties would have evolved similarly. Although historical record on firms' relocation is scant, my reading of the government legislative documents suggests that most firms moved after 1998. In June 1999, the State Council issued the first set of guidelines on firm resettlement. Subsequently in July 1999 and September 2000, the Chongqing municipal government issued its official legislative decree to guide the firm relocation and reconstruction process. Contemporaneous studies and news articles published in that period also corroborate late 1990s and early 2000s as the time when most firms started to move. For instance, out of the 1599 firms that were estimated to be situated below 175m, only 208 firms moved as of 1997 (Zhu, 1999). In addition, the majority (1379 or 88 percent) of affected firms were in Chongqing, which announced its firm resettlement guidelines in 2000. Given that the reservoir was first filled in 2003, it is reasonable that the majority of firms started the relocation process in the early 2000s.

In Table 13, I show that industry size and firm composition evolved similarly in inundated and non-inundated counties between 1999 and 2007, relative to 1998. In particular, there is no statistically significant change in the number of firms overall and those that started prior to the TGD. There is also little change in the composition of firms by operation status or size. Given that flooded counties had around 1,600 firms according to historical accounts, these results may seem counterintuitive. But since most inundated firms are small enterprises with less than 300 employees, they are most likely to be excluded from the sample.

[Table 12 about here.]

Firms in inundated counties appear to have responded by employing a variety of cost-cutting measures while maintaining output. Table 13 shows that inundated counties experience a 4-8 percent reduction in the number of workers and a 3-8 percent reduction in wages per year. Annual benefit cuts are more pronounced at around 5-11 percent for overall operations and 6-12 percent for the firm's primary operation. Accompanied by this labor downsizing, firms in flooded counties also witness a persistent reduction in fixed assets in Table 14. Sales revenue drops significantly in 1999-2000 and 2004-2007 by around 5 percent per year. At the same time, cost reduction in sales is larger at around 6 percent per year. Output remains largely unchanged, except for a temporary decline in 1999. This provides suggestive evidence that firms may have been more resilient to relocation compared to residents. Overall, these results are consistent with the main finding that the destruction of land and physical capital has led to a large and persistent reduction in manufacturing employment.

[Table 13 about here.]

[Table 14 about here.]

8 Conclusion

This paper uses the Three Gorges Project (TGP) in China as a natural experiment to provide the first evidence of the causal impact of dam-induced inundation on population displacement and labor market reallocation. Drawing on a comprehensive dataset combining micro-level census and satellite data, I make four main contributions to improve our understanding of the social costs of energy infrastructure. First, I propose a new identification strategy that uses the reservoir maximum design water level to capture exogenous variations in inundation intensity. Residents living below the water level were forced to move out, while those living at higher elevations could stay. Using satellite data and remote sensing techniques, I calculate the fraction of county area inundated by the dam's reservoir and show that ex-ante elevation relative to the water level accurately predicts ex-post inundation intensity.

Second, I document the magnitude and direction of the displacement flow and provide an empirically-derived estimate to bound the government's official and often contested displacement count. Access to microdata also allows me to uncover new sources of variation in migration patterns and map out the bilateral migration matrix between all county pairs in China. Third, I estimate the short- and long-run impact of inundation on the local labor market, focusing specifically on the transitional costs of reallocating the labor force. I also separately examine the heterogeneous effects on migrants and nonmigrants. Fourth, I use establishment-level data to shed light on the mechanisms through which firms may have adjusted in response to inundation, population displacement, and plant relocation.

Inundation has imposed large and long-lasting costs on the local economy. I estimate that between 1995 and 2010, the total number of people directly displaced by TGP-induced flooding was likely to be between 1.5 and 1.9 million, considerably higher than the 1.4 million reported by the government. The displaced were promised compensation for the loss of their home and income, but in most cases, it was either inadequate to restore their standard of living or never materialized. As a result, while the majority of migrants stayed within the affected provinces, they were often left without stable housing or income source – their former way of life irreparably lost.

In addition to uprooting millions of residents and redistributing them across flooded counties, inundation has led to a reallocation of the labor force across sectors. Due to the destruction of land and physical capital, flooded counties saw a steep and persistent decline of employment in the manufacturing sector and capital-intensive occupations by about 30-50 percent. The closure of firms and factories also led to a "reverse structural transformation," in which the decline of manufacturing forced local residents to seek jobs in agriculture and other less capital-intensive industries in the long run.

The stakes are high for understanding the social costs and distributional impacts of infrastructure development projects. In 2015, World Bank Group President Jim Yong Kim acknowledged shortcomings in the bank's population resettlement policy: "We took a hard look at ourselves on resettlement and what we found caused me deep concern" (World Bank 2015). As the World Bank and other multilateral development banks are making a major push to fund mega-dams and other large energy infrastructure projects in developing countries (World Bank 2013), accurate assessment of their social and economic impact is vital for effective policymaking. To the extent that large-scale infrastructure projects are integral to economic development and dams, in particular, are necessary for combating climate change, institutions need to provide fair and adequate compensation for communities that bear the burdens of development.

This paper has highlighted the need for impact assessments to carefully weigh the projected economic benefits of infrastructure against the potential costs to local communities. In the case of TGP, the feasibility study prepared by the Canadian International Development Agency to endorse the dam has been widely critiqued, due to conflict of interest and inadequate evaluation of ecological and socioeconomic risks (Fearnside 1994; Probe

International 2009). Indeed, my correspondence with a World Bank economist, who was involved in supervising TGP's socioeconomic impact in the 1980s, revealed that the Canadian feasibility study only approved filling the reservoir to 160m. Raising the water level to 175m would reduce room for flood storage, improve power generation only marginally, but would double the population displaced from around 720,000 to 1,300,000. This suggests that dams may do more harm than good when they exceed a certain size.

Beyond dams, lessons learned from the TGP can be helpful for coastal communities to cope with sea-level rise from climate change and irreversible habitat destruction from natural disasters. The analysis here focuses on economic impacts and labor market adjustments, leaving many important dimensions of impact unexplored. Whether permanent inundation triggers ecological hazards, perturbs local ecosystems, disrupts cultural and social networks, and intensifies civil conflicts are open questions left for future research.

References

Autor, David H, David Dorn, and Gordon H Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *Am. Econ. Rev.*, 103(6): 2121–2168.

Autor, David H, David Dorn, Gordon H Hanson, and Jae Song. 2014. "Trade Adjustment: Worker-Level Evidence." *Q. J. Econ.*, 129(4): 1799–1860.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong. 2016. "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia." *Am. Econ. Rev.*, 106(9): 2658–2698.

Boustan, Leah Platt, Matthew E Kahn, and Paul W Rhode. 2012. "Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century." *Am. Econ. Rev.*, 102(3): 238–244.

Cernea, Michael. 1997. "The risks and reconstruction model for resettling displaced populations." *World Dev.*, 25(10): 1569–1587.

CTG. 2002. "Three Gorges Project Q&A." *http://www.ctg.com.cn/sxjt/_300821/gczds/_303331/index.html*, Accessed: 2018-10-12.

CTG. 2010*a*. "Flood Control." *http://www.ctgpc.com/en/_304051/_304529/_304717/_304736/index.html*, Accessed: 2018-8-30.

CTG. 2010*b*. "Selection of the Dam Site of the Three Gorges Project." *http://www.ctgpc.com/en/_304051/_304549/611219/index.html*, Accessed: 2018-8-30.

CTG. 2013*a*. "Project Investment." *http://www.ctg.com.cn/sxjt/_300821/gcgk/_301322/index.html*, Accessed: 2018-8-31.

CTG. 2013*b*. "Three Gorges Project - Overview." *http://www.ctg.com.cn/sxjt/_300821/gcgk/_301302/index.html*, Accessed: 2018-8-30.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns." *Am. Econ. J. Appl. Econ.*, 10(2): 202–233.

De Silva, Dakshina G, Robert P McComb, Young-Kyu Moh, Anita R Schiller, and Andres J Vargas. 2010. "The Effect of Migration on Wages: Evidence from a Natural Experiment." *Am. Econ. Rev.*, 100(2): 321–326.

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *Am. Econ. Rev.*, 101(7): 3078–3108.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *Am. Econ. Rev.*, 108(4-5): 899–934.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *Q. J. Econ.*, 131(2): 799–858.

Duan, Yuefang, and Shawn Steil. 2003. "China Three Gorges Project Resettlement: Policy, Planning and Implementation." *J. Refug. Stud.*, 16(4): 422–443.

Duflo, Esther, and Rohini Pande. 2007. "Dams." *Q. J. Econ.*, 122(2): 601–646.

EIA. 2015. "China." *https://www.eia.gov/beta/international/analysis.php?iso=CHN*, Accessed: 2018-8-30.

EIA. 2018. "Monthly Energy Review August 2018."

Fearnside, Philip M. 1988. "China's Three Gorges Dam: "Fatal" project or step toward modernization?" *World Dev.*, 16(5): 615–630.

Fearnside, Philip M. 1994. "The canadian feasibility study of the three gorges dam proposed for china's yangzi river: A grave embarrassment to the impact assessment profession." *Impact Assessment*, 12(1): 21–57.

Friedberg, Rachel M. 2001. "The Impact of Mass Migration on the Israeli Labor Market." *Q. J. Econ.*, 116(4): 1373–1408.

Gleick, Peter H. 2009. "Three Gorges Dam Project, Yangtze River, China." In *The World's Water 2008-2009: The Biennial Report on Freshwater Resources*. 139–150. Island Press Washington DC.

Greenstone, Michael. 2002. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." *J. Polit. Econ.*, 110(6): 1175–1219.

Highland, Lynn M. 2008. "Geographical Overview of the Three Gorges Dam and Reservoir, China - Geologic Hazards and Environmental Impacts." USGS.

Hornbeck, Richard. 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *Am. Econ. Rev.*, 102(4): 1477–1507.

Hornbeck, Richard, and Suresh Naidu. 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *Am. Econ. Rev.*, 104(3): 963–990.

IDMC. 2017. "Dams and Internal Displacement." Internal Displacement Monitoring Centre.

IDMC. 2018. "Global Report on Internal Displacement." Internal Displacement Monitoring Centre.

IHA. 2016. "Hydropower Status Report." International Hydropower Association.

Li, Heming, and Philip Rees. 2000. "Population displacement in the Three Gorges reservoir area of the Yangtze River, central China: relocation policies and migrant views." *Int. J. Popul. Geogr.*, 6(6): 439–462.

Lipscomb, Molly, A Mushfiq Mobarak, and Tania Barham. 2013. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." *Am. Econ. J. Appl. Econ.*, 5(2): 200–231.

McIntosh, Molly Fifer. 2008. "Measuring the Labor Market Impacts of Hurricane Katrina Migration: Evidence from Houston, Texas." *Am. Econ. Rev.*, 98(2): 54–57.

National Audit Office. 2013. "Financial Audit Report for the Three Gorges Project." The Central People's Government of the People's Republic of China.

NBS. 2006. "Enterprise Size Classification." National Bureau of Statistics of China.

NetEase. 2010. "Three Gorges Project resettlement program completed: 1.4 million people displaced." `http://news.163.com/10/0918/21/6GT5OPKU0001124J.html`, Accessed: 2018-9-1.

Probe International. 2009. *Damming the Three Gorges: What Dam-Builders Don't Want You To Know.* . Second Edition ed., Earthscan Publications Limited.

Roller, Lars-Hendrik, and Leonard Waverman. 2001. "Telecommunications Infrastructure and Economic Development: A Simultaneous Approach." *Am. Econ. Rev.*, 91(4): 909–923.

Ruiz, Isabel, and Carlos Vargas-Silva. 2015. "The Labor Market Impacts of Forced Migration." *Am. Econ. Rev.*, 105(5): 581–586.

Sarvimäki, Matti, Roope Uusitalo, and Markus Jäntti. 2010. "The Long-Term Effects of Forced Migration."

State Council. 1992. "Resolution on the Construction of the Yangtze River Three Gorges Project."

State Council. 1993. "Yangtze River Three Gorges Project Population Resettlement Guidelines."

State Council. 2000. "Three Gorges Project Population Resettlement Status Report." http://www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/2000/1025/Document/327802/327802.htm, Accessed: 2018-8-31.

State Council. 2001. "Yangtze River Three Gorges Project Population Resettlement Revised Guidelines."

State Council. 2006. "Three Gorges Reservoir Inundation Overview." http://www.gov.cn/ztzl/2006-01/02/content_145309.htm, Accessed: 2018-8-31.

Stone, Richard. 2008. "China's environmental challenges. Three Gorges Dam: into the unknown." *Science*, 321(5889): 628–632.

Tullos, Desiree. 2009. "Assessing the influence of Environmental Impact Assessments on science and policy: an analysis of the Three Gorges Project." *J. Environ. Manage.*, 90 Suppl 3: S208–23.

UNHCR. 1998. "Guiding Principles on Internal Displacement." E/CN.4/1998/53/Add.2, United Nations High Commissioner for Refugees.

U.S. Fish and Wildlife Service. 2003. "Chapter 2 Dam Safety Program Description, Definitions, and Standards." 418.

Walker, W Reed. 2013. "The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce*." *Q. J. Econ.*, 128(4): 1787–1835.

Webber, Michael. 2012. "The Political Economy of the Three Gorges Project." *Geographical Research*, 50(2): 154–165.

Wee, Sui-Lee. 2012. "Thousands being moved from China's Three Gorges - again." *Reuters*.

Wilmsen, Brooke, Michael Webber, and Yuefang Duan. 2011*a*. "Development for Whom? Rural to Urban Resettlement at the Three Gorges Dam, China." *Asian Stud. Rev.*, 35(1): 21–42.

Wilmsen, Brooke, Michael Webber, and Yuefang Duan. 2011*b*. "Involuntary Rural Resettlement: Resources, Strategies, and Outcomes at the Three Gorges Dam, China." *J. Environ. Dev.*, 20(4): 355–380.

World Bank. 1996. "Resettlement and development: the Bankwide review of projects involving involuntary resettlement 1986-1993." The World Bank 12971.

World Bank. 2013. "IDA support to transformational projects with regional impact." The World Bank 75927.

World Bank. 2015. "World Bank Acknowledges Shortcomings in Resettlement Projects, Announces Action Plan to Fix Problems."

World Commission on Dams. 2000. *Dams and Development: A New Framework for Decision-making : the Report of the World Commission on Dams.* Earthscan.

Xi, Juan, and Sean-Shong Hwang. 2011. "Relocation Stress, Coping, and Sense of Control Among Resettlers Resulting from China's Three Gorges Dam Project." *Soc. Indic. Res.*, 104(3): 507–522.

Xinhua. 2000*a*. "Inundated Area." http://www.cq.xinhuanet.com/sanxia/yimin/ymgk.htm, Accessed: 2018-8-31.

Xinhua. 2000*b*. "Resettlement Planning." http://www.cq.xinhuanet.com/sanxia/yimin/ymgh.htm, Accessed: 2018-8-31.

Xinhua. 2003. "Three Gorges Project Investment Analysis." *http://www.cq.xinhuanet.com/sanxia/gongcheng/011.htm*, Accessed: 2018-8-31.

Xinhua. 2006. "Total number of Three Gorges migrants will reach 1.4 million." *http://news.sina.com.cn/c/2006-10-01/212910153892s.shtml*, Accessed: 2018-9-1.

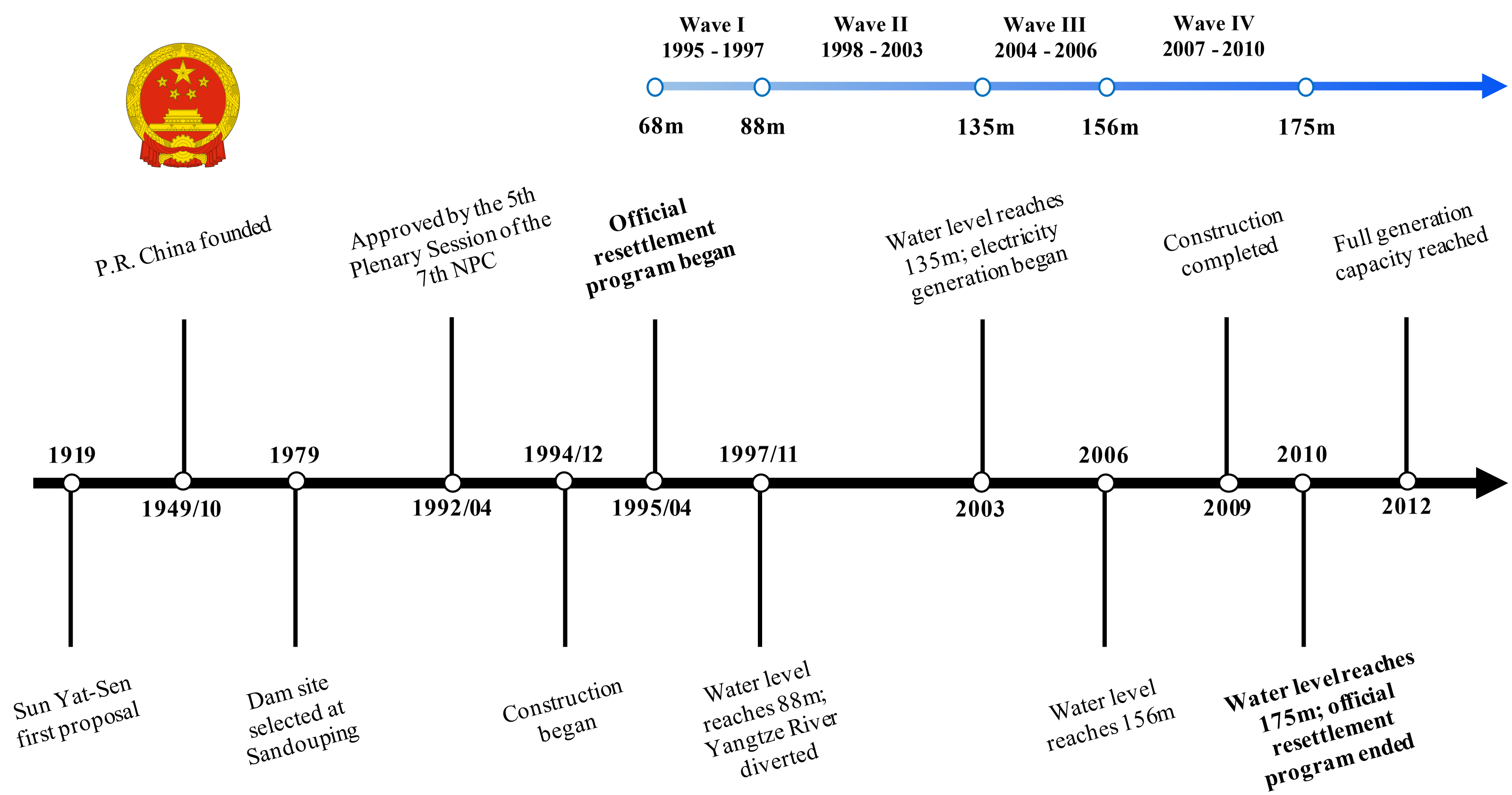
Xinhua. 2017. "Three Gorges Project generates 1 trillion kWh of electricity." *http://www.xinhuanet.com/english/2017-03/01/c_136094226.htm*, Accessed: 2018-8-30.

Yu, Fengling, Zhongyuan Chen, Xianyou Ren, and Guifang Yang. 2009. "Analysis of historical floods on the Yangtze River, China: Characteristics and explanations." *Geomorphology*, 113(3): 210–216.

Zarfl, Christiane, Alexander E Lumsdon, Jürgen Berlekamp, Laura Tydecks, and Klement Tockner. 2015. "A global boom in hydropower dam construction." *Aquat. Sci.*, 77(1): 161–170.

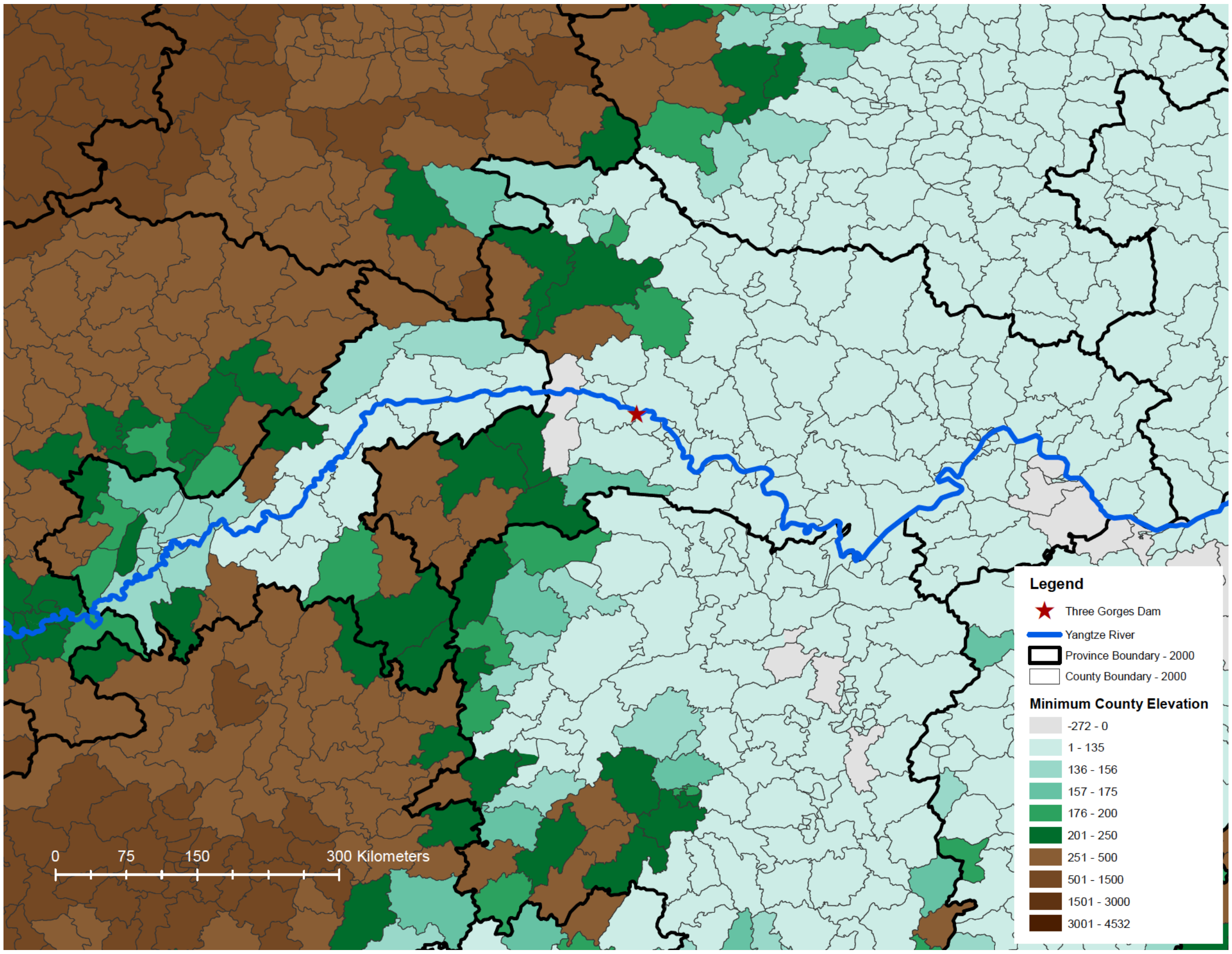
Zhang, Alice Tianbo, Johannes Urpelainen, and Wolfram Schlenker. 2018. "Power of the River: Introducing the Global Dam Tracker (GDAT)." Center on Global Energy Policy. Forthcoming.

Figure 1: Timeline of the Three Gorges Project



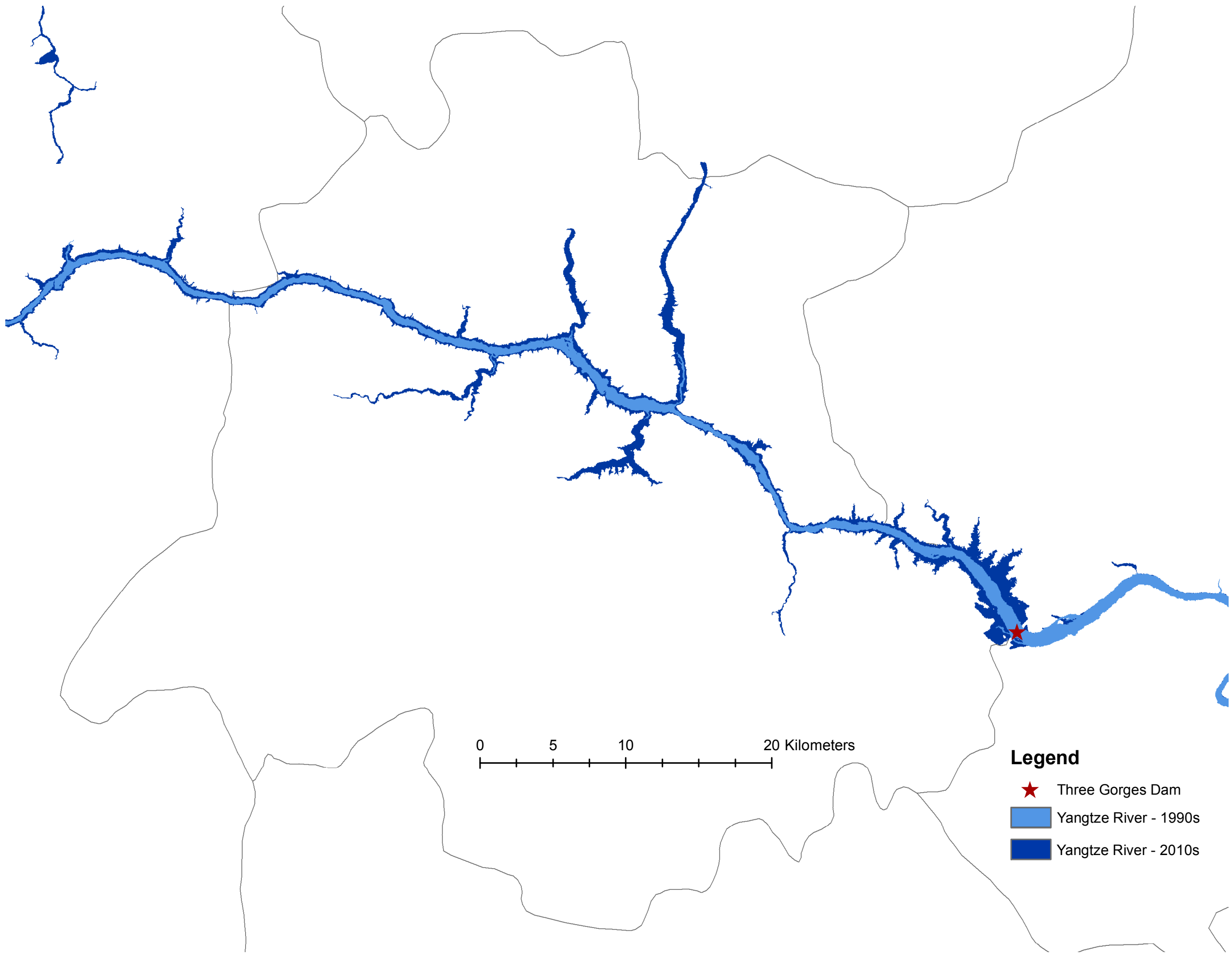
Notes: This figure documents the timeline of TGP using government legislative records, news articles, websites, and academic publications. Dam construction progressed in three phases that marked the four waves of forced relocation: Phase 1 (1992-1997), Phase 2 (1998-2003) and Phase 3 (2004-2010). Population displacement often preceded major construction milestones to clear the way for the rising water level.

Figure 2: Map of County Minimum Elevation



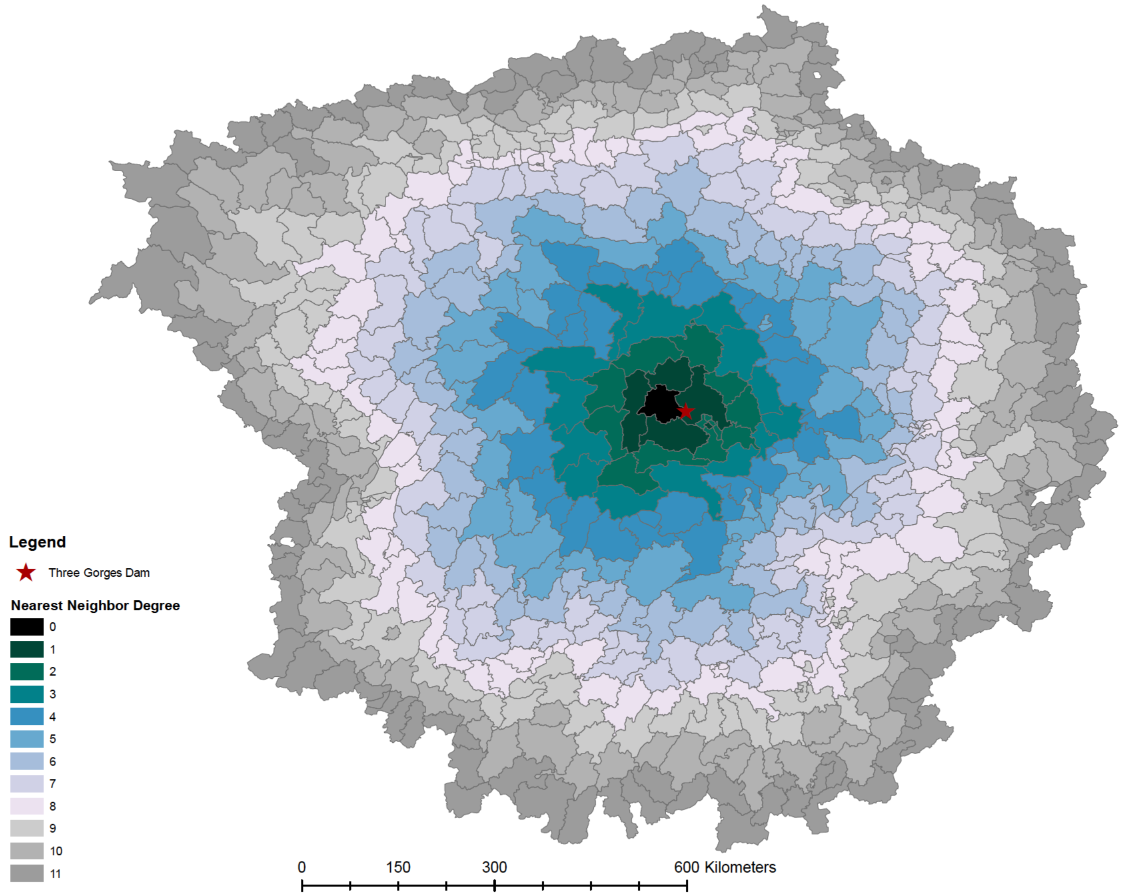
Notes: This map shows the minimum elevation for counties along the Yangtze River. The color gradient represents water levels during the four waves of TGD-induced forced migration: 1-135m (Waves 1 & 2: 1995-2003), 136-156m (Wave 3: 2004-2006) and 157-175m (Wave 4: 2007-2010). Most counties along the Yangtze River have minimum elevation below 156m. Calculations are done in ArcGIS using Digital Elevation Model data from the NASA SRTM.

Figure 3: Change in Yangtze River Extent between 1990s and 2010s near the TGD



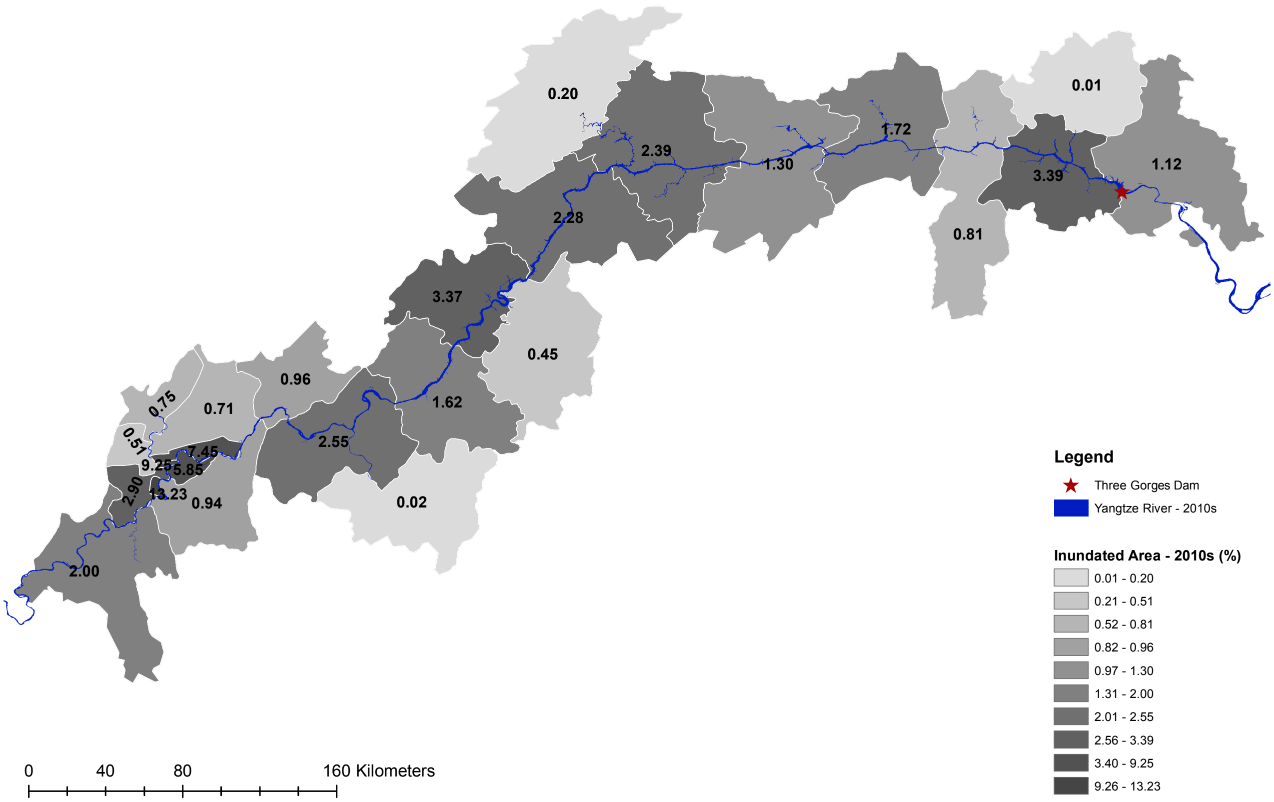
Notes: This map overlays the Yangtze River extent before and after TGD construction and county boundaries in 2000. It is zoomed into Zigui County, where the TGD is located. Lighter blue represents Yangtze River in the 1990s and darker blue that in the 2010s. There is a clear expansion of the river channel upstream of TGD: the dam elevated water levels in the reservoir to 175 meters and resulted in an increase of 322 km^2 in flooded area. Calculations are done in ENVI using remote sensing techniques and Landsat 5 TM satellite data from the USGS.

Figure 4: Map of Nearest Neighbor Classification



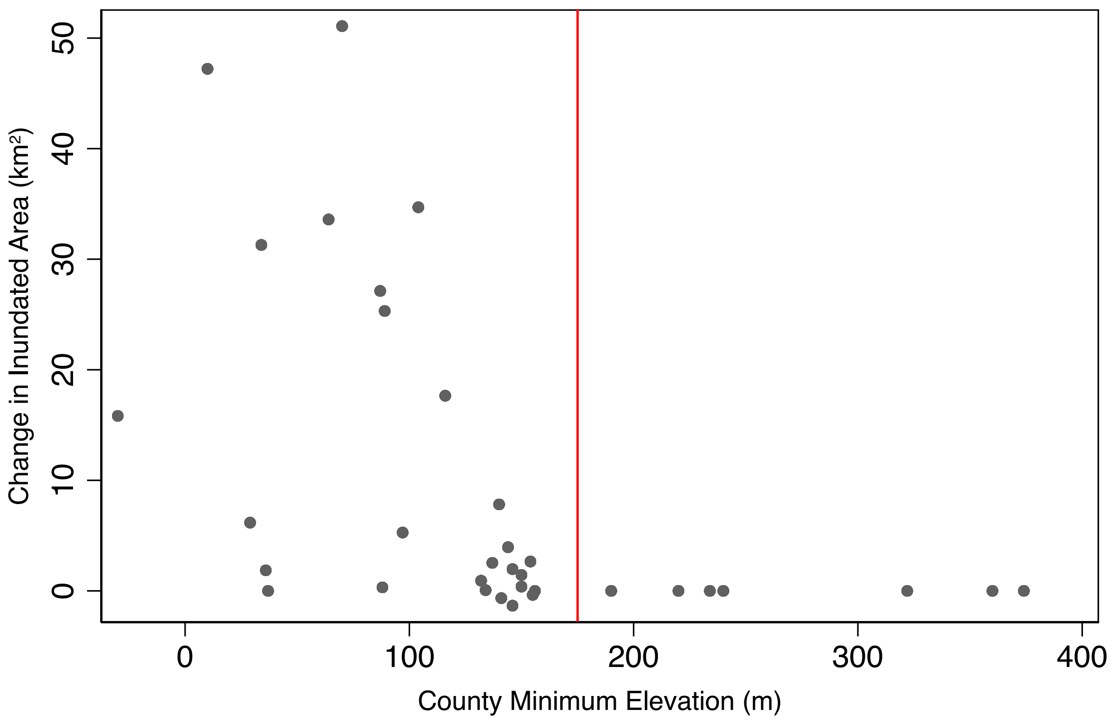
Notes: This map demonstrates the nearest neighbors of Zigui County, where the TGD is located, from the first up to the eleventh degree neighbors. Using the same procedure, I classify the pairwise neighbor relationship between all counties in China and create a file with origin county, residence county, and dummy variable indicating whether the counties are first degree, second degree, third degree, and up to the tenth degree neighbors. This allows me to construct the bilateral migration matrix between all county pairs and uncover new sources of variation in intra-provincial migration patterns. In particular, it enables me to examine potential TGD-induced outmigration to neighboring non-flooded counties.

Figure 5: Map of Inundated Counties



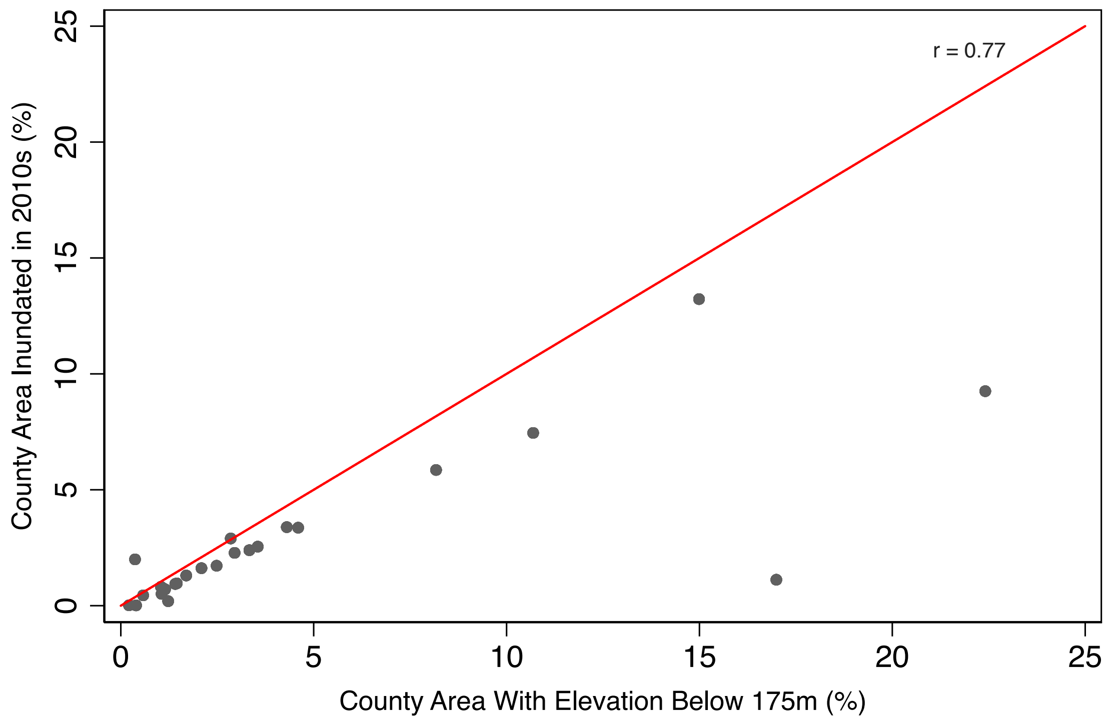
Notes: This map shows the 25 counties partially inundated by the TGD. The color gradients and numbers represent the percent county area inundated. Yangtze River extent in the 2010s (post-TGD) is shown. Inundation area and river extent are derived from Landsat 5 TM satellite imageries using remote sensing techniques in ENVI.

Figure 6: Change in Inundated Area between the 1990s and 2010s vs. Minimum Elevation



Notes: This scatterplot shows the change in inundated area between the 1990s and 2010s vs. minimum elevation for treated and nearby untreated counties. The red vertical line represents 175 meters, which is the maximum design water level in the reservoir of TGD. Change in inundated area is constructed from Landsat 5 TM satellite data using remote sensing techniques in ENVI. Minimum county elevation is calculated from the SRTM DEM satellite data. The sample is restricted to counties within 30 kilometers of the Yangtze River for readability. Including the full sample merely adds a long right tail of counties with little change in flooded area.

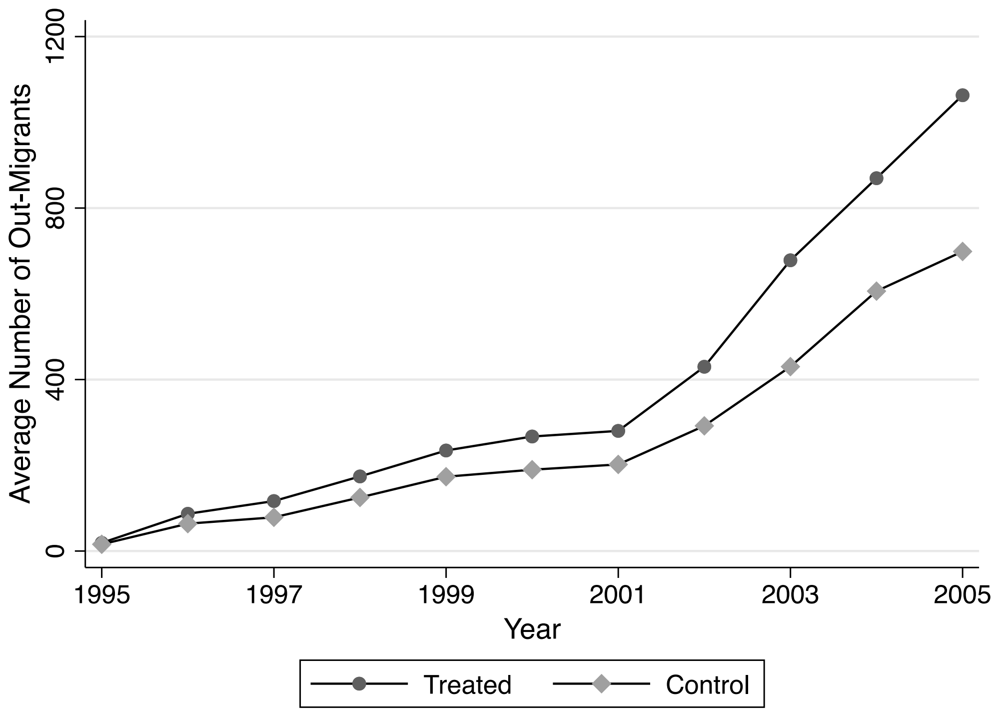
Figure 7: County Area Inundated in the 2010s vs. County Area Below 175m



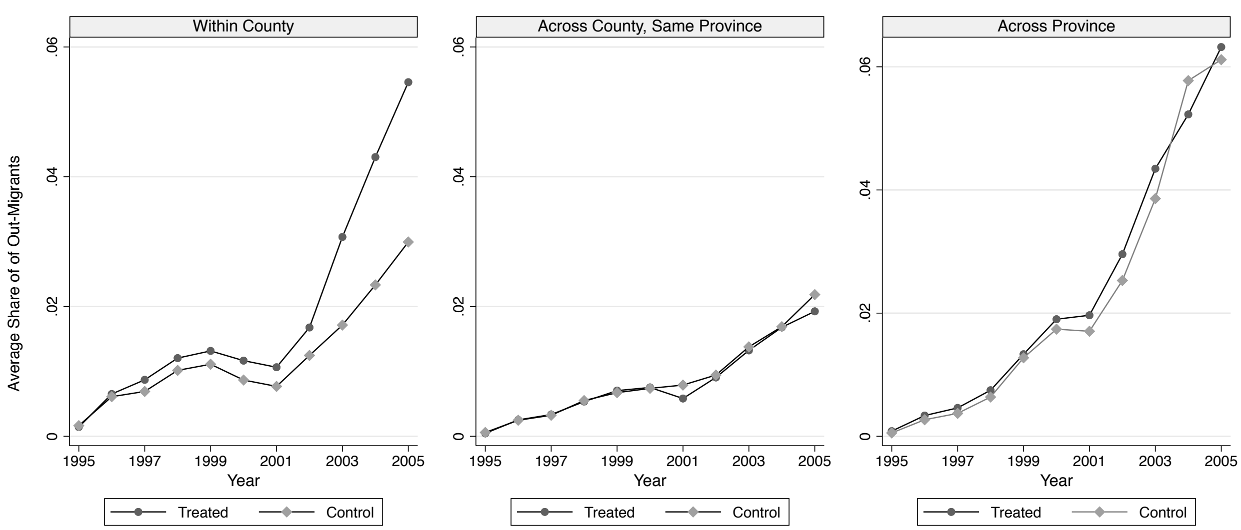
Notes: This scatterplot shows the fraction of county area inundated in the 2010s vs. the fraction of county area below 175m in treated counties. Inundated area is constructed from Landsat 5 TM satellite data using remote sensing techniques in ENVI. Area of county elevation below 175m is calculated from the SRTM DEM satellite data.

Figure 8: Out-Migration Patterns in Treated vs. Control Counties

(a) Average Number of Out-Migrants

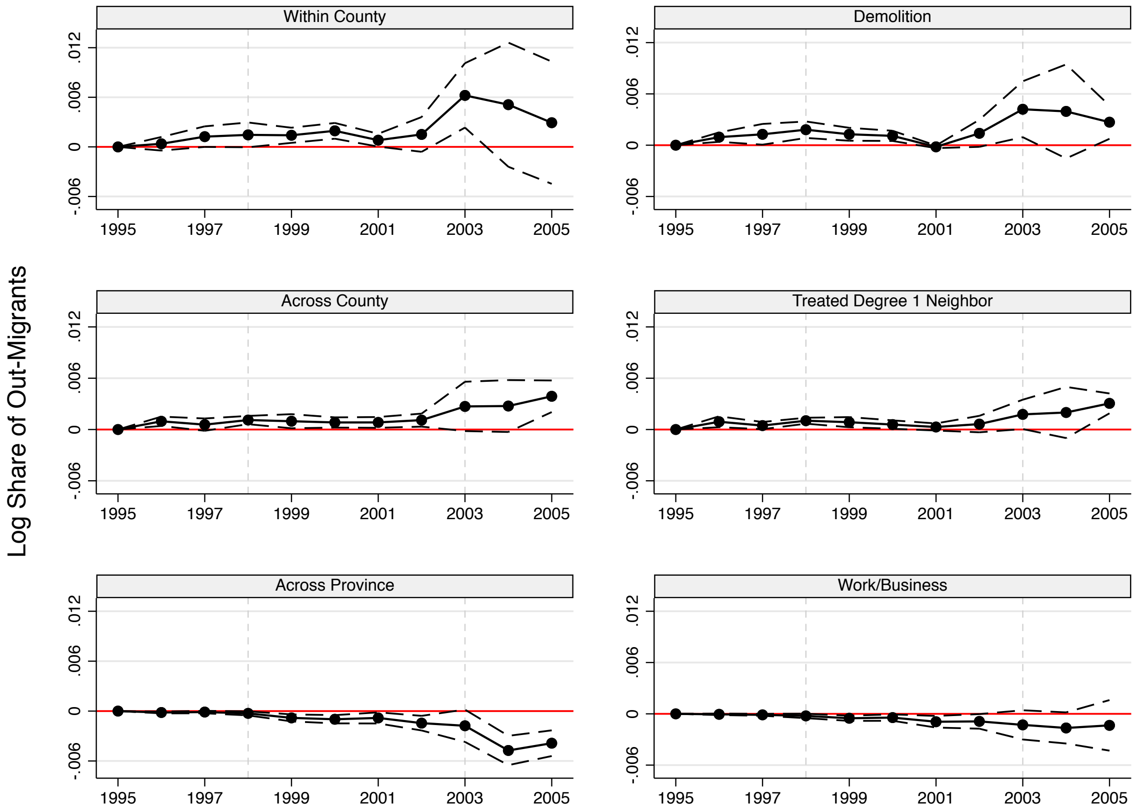


(b) Average Share of Out-Migrants by Destination Type



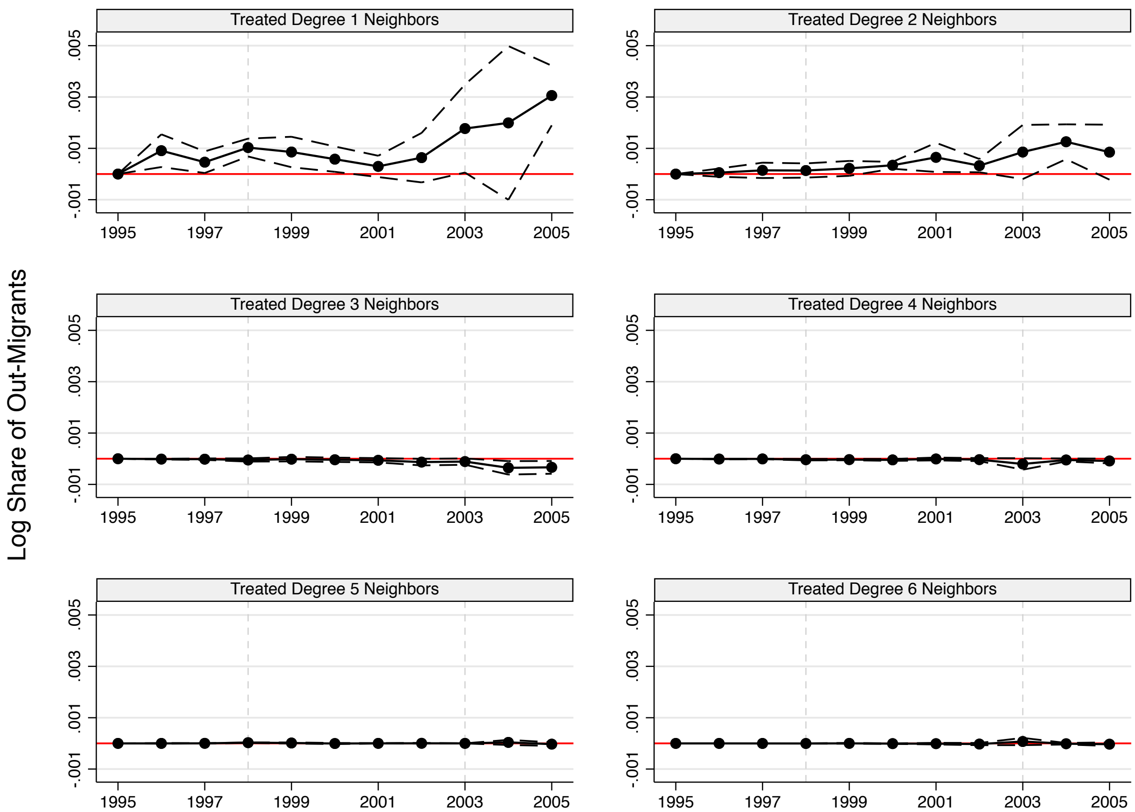
Notes: These graphs show the outmigration patterns in inundated (treated) and nearby non-inundated (control) counties between 1995 and 2005. Panel (a) shows the average number of outmigrants each year. Panel (b) shows the average share of outmigrants each year by three destination types: within county, across county in the same province, and across province. Migration flow is calculated retrospectively using individual-level census data in 2000 and 2005. See text for details.

Figure 9: Estimated Difference in Out-Migration Flow from Inundated Counties by Destination Type



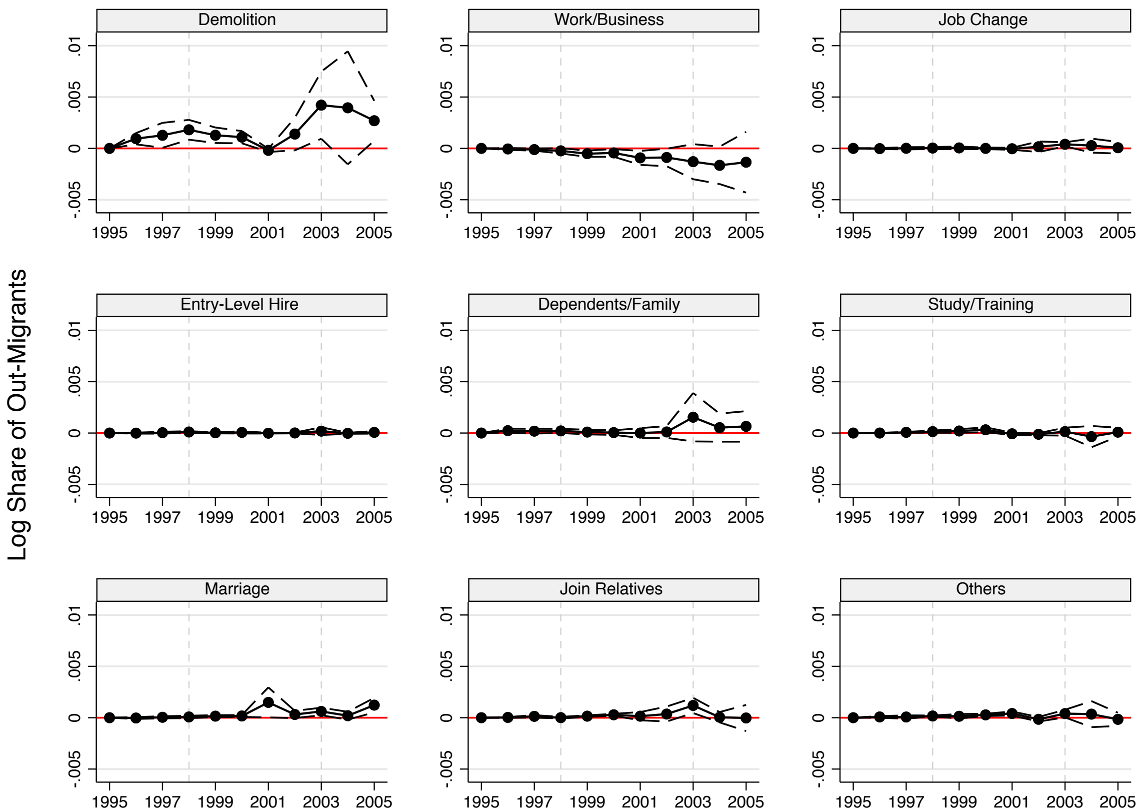
Notes: These graphs show the estimated differences in outmigration flow between inundated and non-inundated counties relative to the omitted base year in 1995. In the first column, the dependent variables are the log fraction of outmigrants in three destination types reported in the census: within county, across county in the same province and across province. In the second column, I show that within-county moves are mostly driven by house demolition, whereas across-province moves are driven by work or business. Migration across counties are mostly directed to another treated first degree neighbor. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The black dashed lines represent 95 percent confidence intervals. The gray vertical dashed lines at 1998 and 2003 show the years when the water level of Yangtze was elevated due to the construction of TGD. Migration flow is calculated retrospectively using individual-level census data in 2000 and 2005.

Figure 10: Estimated Difference in Out-Migration Flow from Inundated Counties to Treated Neighbors



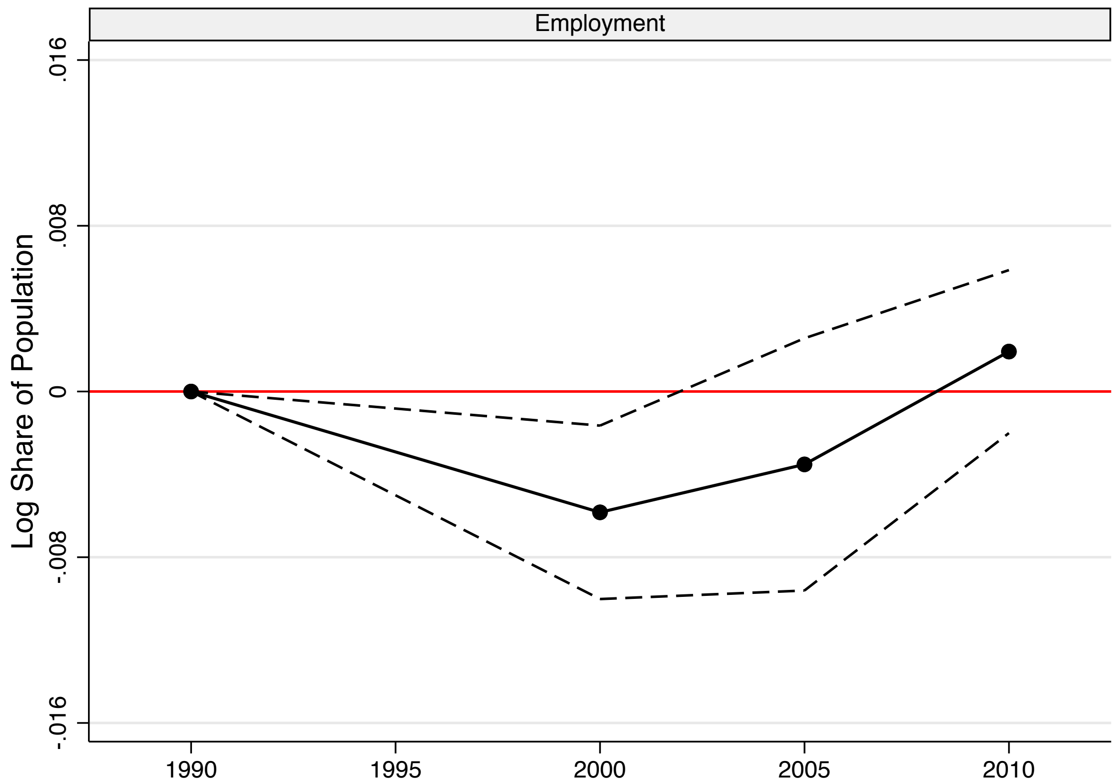
Notes: These graphs show the estimated differences in outmigration flow between inundated and non-inundated counties relative to the omitted base year in 1995. The dependent variables are the log fraction of outmigrants to treated neighboring counties up to the 6th degree. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The black dashed lines represent 95 percent confidence intervals. The gray vertical dashed lines at 1998 and 2003 show the years when the water level of Yangtze was elevated due to the construction of TGD. Migration flow is calculated retrospectively using individual-level census data in 2000 and 2005.

Figure 11: Estimated Difference in Out-Migration Flow from Inundated Counties by Migration Reason



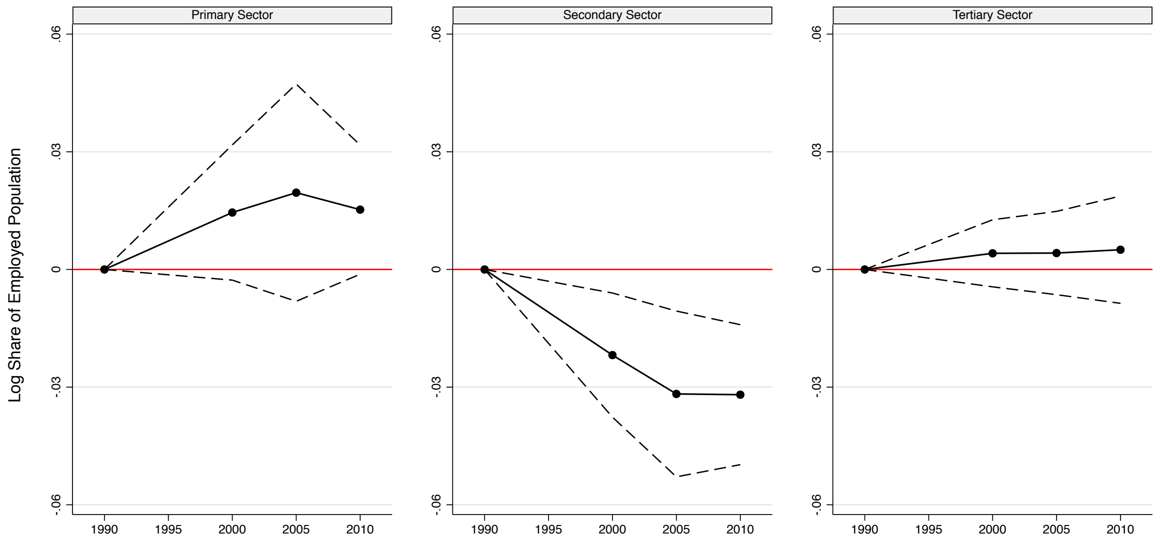
Notes: These graphs show the estimated differences in outmigration flow between inundated and non-inundated counties relative to the omitted base year in 1995. The dependent variables are the log fraction of outmigrants by reported reason of migration. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The black dashed lines represent 95 percent confidence intervals. The gray vertical dashed lines at 1998 and 2003 show the years when the water level of Yangtze was elevated due to the construction of TGD. Migration flow is calculated retrospectively using individual-level census data in 2000 and 2005.

Figure 12: Estimated Difference in Overall Employment by Inundation Share



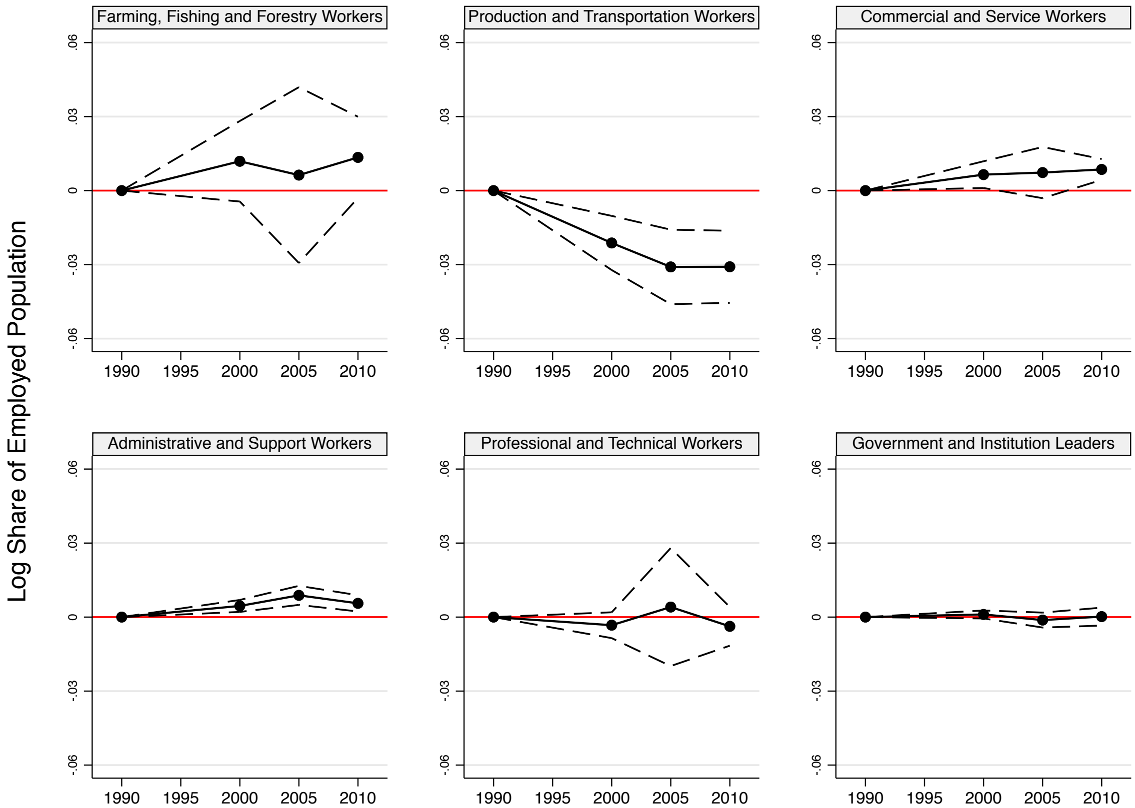
Notes: This graph shows the estimated differences in overall employment between inundated and non-inundated counties relative to the omitted base year in 1990. The dependent variable is the log fraction of population 15 years old and above who are employed. Using equation 1, it is regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The black dashed lines represent 95 percent confidence intervals.

Figure 13: Estimated Difference in Sectoral Employment by Inundation Share



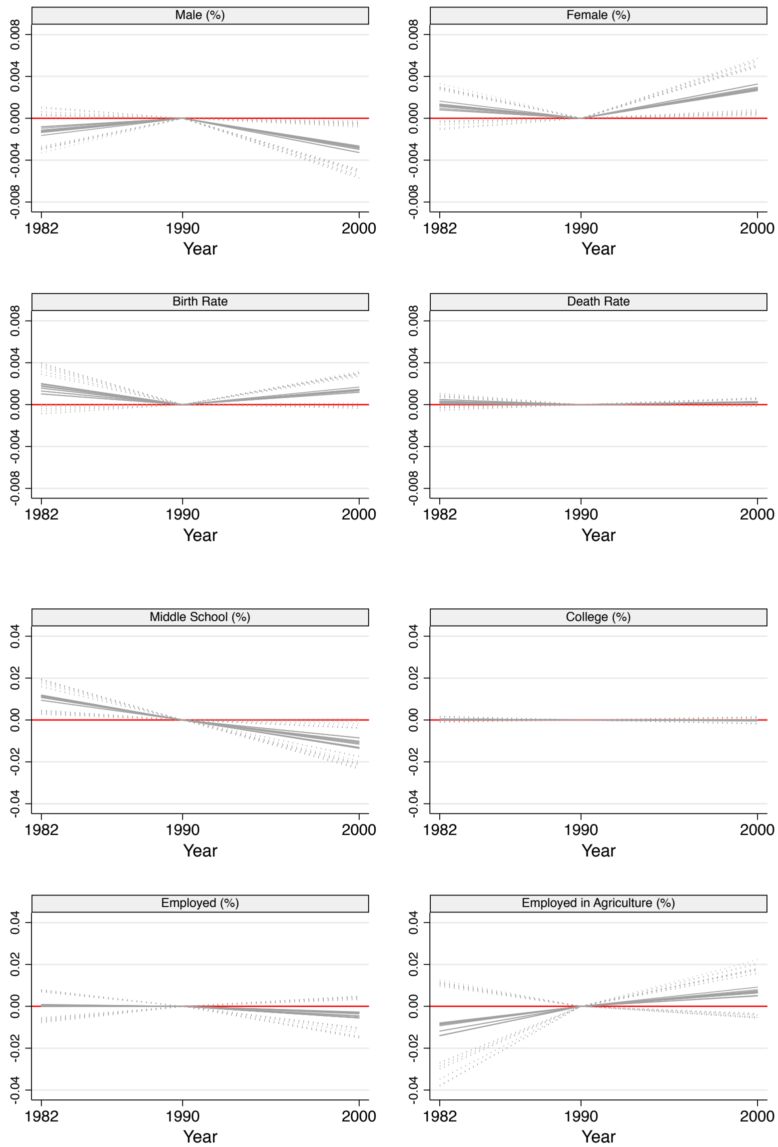
Notes: These graphs show the estimated differences in sectoral employment between inundated and non-inundated counties relative to the omitted base year in 1990. The dependent variables are the log fraction of population 15 years old and above, who are employed in the primary, secondary and tertiary sectors. Using equation 1, it is regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The black dashed lines represent 95 percent confidence intervals.

Figure 14: Estimated Difference in Occupational Employment by Inundation Share



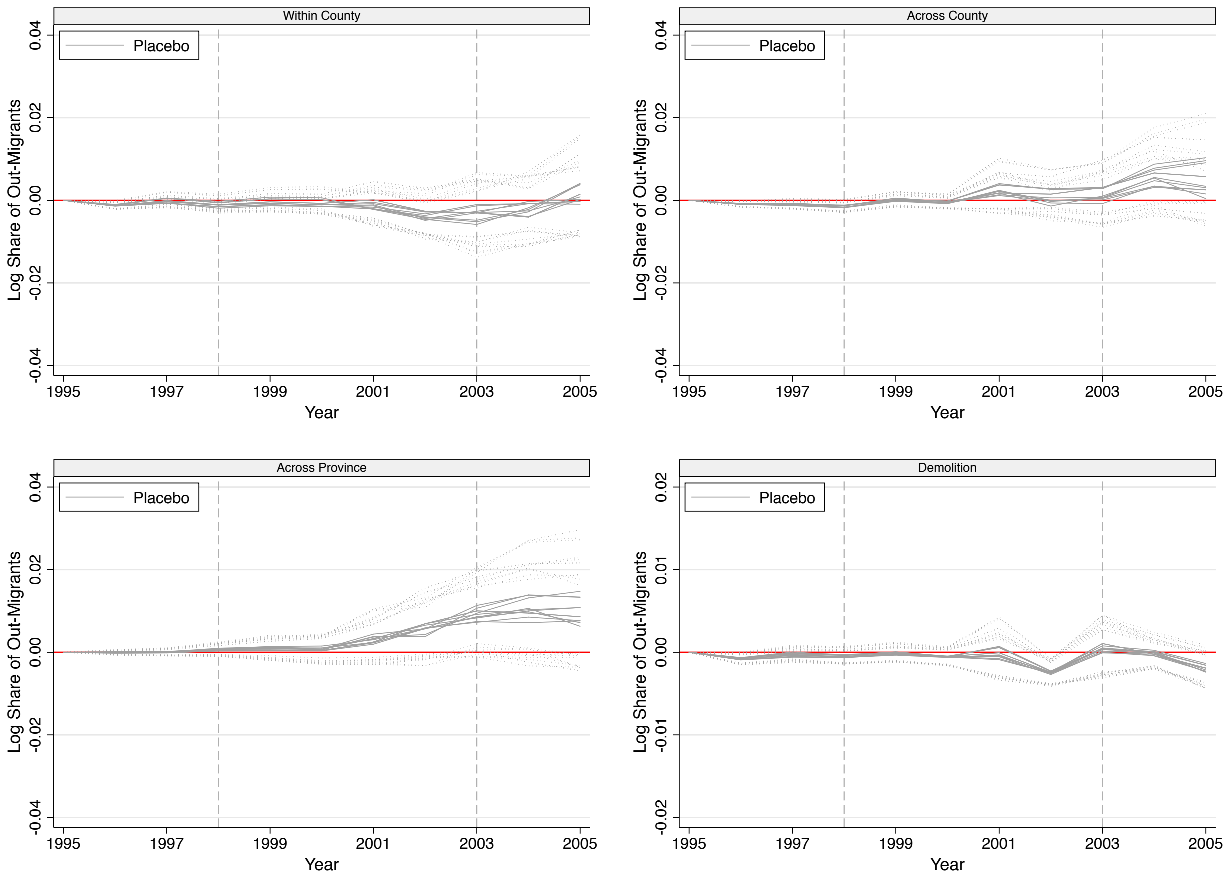
Notes: These graphs show the estimated differences in occupational employment between inundated and non-inundated counties relative to the omitted base year in 1990. The dependent variables are the log fraction of population 15 years old and above who are employed in 6 harmonized occupations. Using equation 1, it is regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The black dashed lines represent 95 percent confidence intervals.

Figure 15: Estimated Difference in Baseline Characteristics of Placebo Counties



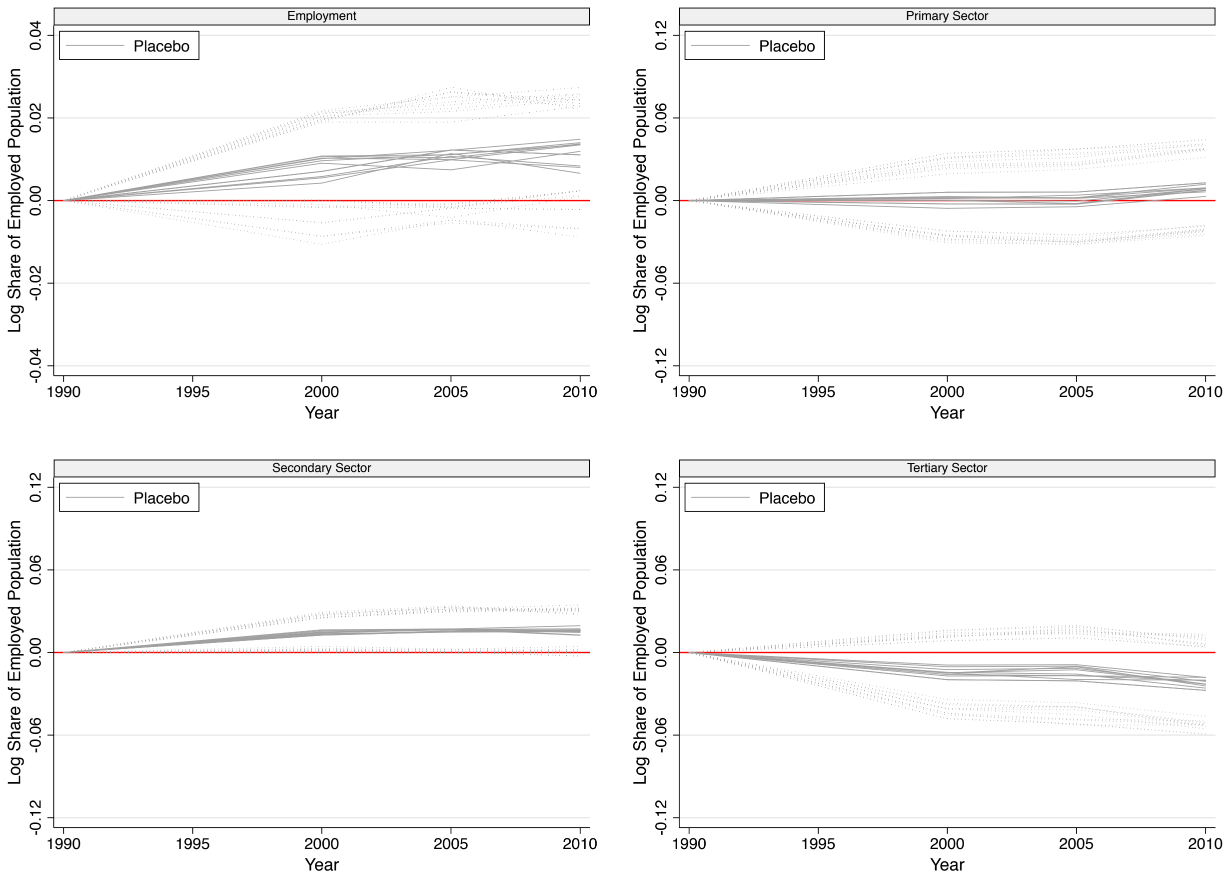
Notes: These graphs show the estimated differences in baseline characteristics between placebo counties near the Yangtze River and those farther away relative to the omitted base year in 1990. Counties are considered 'pseudo-flooded' if they are within 50 km of the Yangtze River in provinces upstream of the TGD. I construct 11 placebo samples by varying the distance cutoffs between 0 and 50 in 5km increments. The dependent variables are fraction of males and females, birth and death rates, fraction of middle school and college graduates, and fraction of those employed and employed in agriculture. Using equation 1, these are regressed on a dummy for whether the county lies within the distance cutoff interacted with a year dummy, county and state-by-year fixed effects. To account for the underlying population age structure, I also include the share of population below 14 and above 65 in 1982 interacted with the year dummy as controls. Robust standard errors are clustered by county. The dashed lines show 95 percent confidence intervals.

Figure 16: Estimated Difference in Out-Migration Flow from Placebo Counties



Notes: These graphs show the estimated differences in outmigration flow between placebo counties near the Yangtze River and those farther away relative to the omitted base year in 1995. Counties are considered 'pseudo-flooded' if they are within 50 km of the Yangtze River in provinces upstream of the TGD. I construct 11 placebo samples by varying the distance cutoffs between 0 and 50 in 5km increments. The dependent variables are the log fraction of out-migrants in three destination types reported in the census: within county, across county and across province, and those who moved due to house demolition. Using equation 1, these are regressed on a dummy for whether the county lies within the distance cutoff interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The dashed lines represent 95 percent confidence intervals. The gray vertical dashed lines at 1998 and 2003 show the years when the water level of Yangtze was elevated due to the construction of TGD.

Figure 17: Estimated Difference in Overall and Sectoral Employment of Placebo Counties



Notes: These graphs show the estimated differences in overall and sectoral employment between placebo counties near the Yangtze River and those farther away relative to the omitted base year in 1990. Counties are considered 'pseudo-flooded' if they are within 50 km of the Yangtze River in provinces upstream of the TGD. I construct 11 placebo samples by varying the distance cutoffs between 0 and 50 in 5km increments. The dependent variables are the log fraction of population 15 years old and above, who are employed and those employed in the primary, secondary and tertiary sectors. Using equation 1, these are regressed on a dummy for whether the county lies within the distance cutoff interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. The dashed lines represent 95 percent confidence intervals.

Table 1: Baseline Difference in Average 1990 County Characteristics by Treatment Status

	Treated Mean (N = 24)	Control Mean (N = 103)	P-value of Difference
Demographics			
Total Population	7412.000	6150.301	0.131
Average age	30.966	28.653	0.000
Female	0.481	0.484	0.446
Han ethnicity	0.954	0.928	0.545
Rural hukou	0.705	0.794	0.192
Nonmigrant	0.949	0.971	0.148
Married	0.659	0.679	0.156
Educational Attainment			
Literate	0.767	0.699	0.000
Primary school	0.391	0.370	0.309
Middle school	0.247	0.227	0.151
High school	0.094	0.086	0.648
University	0.024	0.008	0.268
Employment by Sector			
Employed	0.796	0.816	0.287
Primary sector	0.673	0.731	0.419
Secondary sector	0.177	0.114	0.131
Tertiary sector	0.150	0.155	0.907
Employment by Occupation			
Agricultural workers	0.664	0.707	0.554
Production and transportation workers	0.181	0.138	0.277
Commercial and service workers	0.062	0.061	0.981
Admin and support workers	0.020	0.019	0.841
Professional and technical workers	0.058	0.055	0.801
Government and institution leaders	0.014	0.019	0.375
Migration Reason			
Work and business	0.010	0.007	0.289
Study and training	0.021	0.005	0.263
Marriage	0.006	0.005	0.530
Other	0.002	0.003	0.390

Notes: This table compares average baseline characteristics between inundated and non-inundated counties in 1990 before the TGD is approved. The sample is restricted to the two affected provinces, Chongqing and Hubei. Across a comprehensive set of variables on demographics, educational attainment, sectoral and occupational employment, and migration reason, treated counties and their untreated neighbors are very similar.

Table 2: Baseline Difference by Distance to the Yangtze River

		Demographics (Fraction)		
	Total Pop	Male Pop	Sex Ratio	Rural Hukou
Distance to Yangtze (km)	-133.9420***	0.0001***	0.0006***	0.0028***
	(38.4113)	(0.0000)	(0.0002)	(0.0006)
Townships	139	139	139	139
R^2	0.06	0.06	0.06	0.08
μ_{y}	24222.14	0.52	1.07	0.76
σ_{y}	13826.72	0.01	0.06	0.25

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table examines possible selection of migrants within treated counties using hardcopy township-level census tabulations in 1990. The dependent variables are total population, the fraction of males, sex ratio (males/females) and the fraction of rural Hukou holders. These are regressed on the distance of each township to the Yangtze River using a simple OLS model. Towns closer to the river tend to be more urban and densely populated, suggesting that migrants may be positively selected within the flooded counties. This will bias the negative effects of forced displacement downward.

Table 3: Estimated Changes in County Characteristics 1982–1990

	Dependent var: County characteristics (fraction)					
	Male	Sex Ratio	Middle School	College	Employed	Employed in Ag
Panel A: Within Province, No Controls						
Pre-TGD changes 1982–1990	-0.0002	-0.0009	-0.0003	-0.0023	0.0034*	0.0096
	(0.0002)	(0.0009)	(0.0016)	(0.0015)	(0.0018)	(0.0067)
R^2	0.80	0.80	0.95	0.83	0.77	0.98
Panel B: Age Structure Controls						
Pre-TGD changes 1982–1990	-0.0005	-0.0020	-0.0003	-0.0011	0.0016	0.0082
	(0.0003)	(0.0013)	(0.0017)	(0.0012)	(0.0015)	(0.0062)
R^2	0.86	0.86	0.97	0.91	0.86	0.98
Counties	105	105	105	105	105	105
μ_{y}	0.52	1.08	0.23	0.01	0.81	0.78
σ_{y}	0.01	0.04	0.08	0.02	0.05	0.17

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated changes in baseline characteristics between 1982 and 1990 in inundated and nearby non-inundated counties in the same province. The dependent variables are the fraction of males, sex ratio (male/female), fraction of middle school and college graduates, fraction of working age population employed, and fraction of employed population in agriculture. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Panel A shows pre-trend results without any controls. To account for the underlying population age structure, Panel B includes the share of population below 14 and above 65 in 1982 interacted with the year dummy as controls. Robust standard errors are clustered by county. Analysis uses harmonized county-level decennial census data for 1982 and 1990. I manually create crosswalk relationship files for all counties to keep administrative boundaries 1990-consistent.

Table 4: Effect of TGD on Out-Migration Flow by Destination Type

	Dependent var: Log fraction of out-migrants for each destination type		
	Within County	Across County, Same Province	Across Province
1996	0.0004	0.0010***	-0.0002***
	(0.0004)	(0.0003)	(0.0001)
1997	0.0012*	0.0006	-0.0001
	(0.0006)	(0.0004)	(0.0001)
1998	0.0015*	0.0011***	-0.0003**
	(0.0008)	(0.0002)	(0.0001)
1999	0.0014***	0.0010**	-0.0008***
	(0.0005)	(0.0004)	(0.0002)
2000	0.0019***	0.0008***	-0.0010***
	(0.0005)	(0.0003)	(0.0003)
2001	0.0008**	0.0008**	-0.0008**
	(0.0004)	(0.0003)	(0.0003)
2002	0.0015	0.0011***	-0.0014***
	(0.0011)	(0.0004)	(0.0004)
2003	0.0062***	0.0027*	-0.0018*
	(0.0020)	(0.0015)	(0.0010)
2004	0.0051	0.0027*	-0.0047***
	(0.0038)	(0.0015)	(0.0009)
2005	0.0029	0.0039***	-0.0039***
	(0.0037)	(0.0009)	(0.0008)
Observations	1496	1496	1496
Counties	136	136	136
R^2	0.52	0.54	0.74

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in outmigration flow between inundated and non-inundated counties relative to the omitted base year in 1995. The dependent variables are the log fraction of outmigrants in three destination types reported in the census: within county, across county in the same province and across province. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 5: Effect of TGD on Out-Migration Flow to Degree N_i Neighbor Counties

	Dependent var: Log fraction of out-migrants to neighbor degree N_i					
	N1	N2	N3	N4	N5	N6
Panel A: All Neighbors						
1996	0.0009**	0.0001	-0.0000*	-0.0000	-0.0000	-0.0000
	(0.0003)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1997	0.0004	0.0002	0.0000	-0.0000	-0.0000	0.0000
	(0.0002)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1998	0.0010***	0.0002	-0.0001	-0.0000	0.0000**	-0.0000
	(0.0002)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1999	0.0007*	0.0003	-0.0000	-0.0001**	-0.0000	-0.0000**
	(0.0003)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
2000	0.0005*	0.0003***	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
2001	0.0002	0.0009*	-0.0002*	-0.0000	-0.0001	-0.0000
	(0.0002)	(0.0003)	(0.0001)	(0.0000)	(0.0001)	(0.0000)
2002	0.0005	0.0004**	0.0001	-0.0001*	-0.0000	0.0000
	(0.0005)	(0.0001)	(0.0002)	(0.0001)	(0.0000)	(0.0001)
2003	0.0015	0.0011	0.0002	-0.0003*	0.0003	0.0001
	(0.0009)	(0.0006)	(0.0002)	(0.0001)	(0.0004)	(0.0001)
2004	0.0015	0.0014***	-0.0003*	-0.0002	-0.0001	-0.0001
	(0.0015)	(0.0004)	(0.0002)	(0.0001)	(0.0001)	(0.0000)
2005	0.0027***	0.0009	0.0002	0.0000	-0.0001	-0.0000
	(0.0006)	(0.0006)	(0.0002)	(0.0002)	(0.0000)	(0.0001)
Observations	1496	1496	1496	1496	1496	1496
Counties	136	136	136	136	136	136
R^2	0.44	0.39	0.36	0.28	0.19	0.16

	Dependent var: Log fraction of out-migrants to neighbor degree N_i					
	N1	N2	N3	N4	N5	N6
Panel B: Treated Neighbors						
1996	0.0009***	0.0001	-0.0000*	-0.0000**	-0.0000	-0.0000
	(0.0003)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1997	0.0005**	0.0001	-0.0000	-0.0000	0.0000	-0.0000
	(0.0002)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1998	0.0010***	0.0001	-0.0000	-0.0000***	0.0000***	-0.0000
	(0.0002)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1999	0.0009***	0.0002	-0.0000	-0.0000**	0.0000	0.0000
	(0.0003)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
2000	0.0006**	0.0003***	-0.0000	-0.0000**	-0.0000	-0.0000*
	(0.0002)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
2001	0.0003	0.0007**	-0.0001	-0.0000	0.0000	-0.0000
	(0.0002)	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
2002	0.0006	0.0003**	-0.0001*	-0.0000	0.0000	-0.0000
	(0.0005)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
2003	0.0018**	0.0009	-0.0001*	-0.0002*	0.0000	0.0001
	(0.0009)	(0.0005)	(0.0001)	(0.0001)	(0.0000)	(0.0001)
2004	0.0020	0.0013***	-0.0004***	-0.0000	0.0000	-0.0000
	(0.0015)	(0.0003)	(0.0001)	(0.0000)	(0.0001)	(0.0000)
2005	0.0031***	0.0009	-0.0003***	-0.0001*	-0.0000	-0.0000
	(0.0006)	(0.0005)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Observations	1496	1496	1496	1496	1496	1496
Counties	136	136	136	136	136	136
R^2	0.65	0.58	0.41	0.30	0.16	0.17

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in outmigration flow between inundated and non-inundated counties relative to the omitted base year in 1995. The dependent variables are the log fraction of outmigrants to all neighbors (Panel A) and treated neighbors only (Panel B) up to the 6th degree. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 6: Effect of TGD on Out-Migration Flow by Migration Reason

	Dependent var: Log fraction of out-migrants for each migration reason			
	Work/Business	Job Change	Entry-level Hire	Dependents/Family
Panel A: Work-related out-migration				
1996	-0.0001	-0.0000	-0.0000	0.0002*
	(0.0000)	(0.0000)	(0.0000)	(0.0001)
1997	-0.0001*	0.0000	0.0000	0.0002
	(0.0001)	(0.0001)	(0.0000)	(0.0001)
1998	-0.0002	0.0000	0.0001**	0.0002*
	(0.0001)	(0.0001)	(0.0000)	(0.0001)
1999	-0.0005**	0.0000	0.0000	0.0001
	(0.0002)	(0.0001)	(0.0000)	(0.0001)
2000	-0.0004*	0.0000	0.0001	0.0000
	(0.0002)	(0.0000)	(0.0001)	(0.0001)
2001	-0.0009**	-0.0000	-0.0000	0.0000
	(0.0003)	(0.0000)	(0.0000)	(0.0002)
2002	-0.0009*	0.0002	0.0000	0.0001
	(0.0004)	(0.0003)	(0.0000)	(0.0003)
2003	-0.0013	0.0004***	0.0002	0.0015
	(0.0009)	(0.0001)	(0.0002)	(0.0012)
2004	-0.0016	0.0003	-0.0000	0.0005
	(0.0009)	(0.0003)	(0.0000)	(0.0007)
2005	-0.0013	0.0001	0.0001	0.0006
	(0.0015)	(0.0003)	(0.0001)	(0.0008)
Observations	1496	1496	1496	1496
Counties	136	136	136	136
R^2	0.77	0.22	0.25	0.48

	Dependent var: Log fraction of out-migrants for each migration reason				
	Study/Training	Demolition	Marriage	Join Relatives	Others
Panel B: Non-work related out-migration					
1996	0.0000	0.0009***	-0.0000	0.0000	0.0001*
	(0.0000)	(0.0003)	(0.0000)	(0.0000)	(0.0000)
1997	0.0001***	0.0013**	0.0000	0.0001**	0.0001
	(0.0000)	(0.0006)	(0.0001)	(0.0001)	(0.0001)
1998	0.0001**	0.0018***	0.0001	0.0000	0.0002***
	(0.0001)	(0.0005)	(0.0001)	(0.0000)	(0.0000)
1999	0.0002**	0.0013***	0.0002***	0.0002***	0.0001
	(0.0001)	(0.0004)	(0.0000)	(0.0000)	(0.0001)
2000	0.0003***	0.0011***	0.0002***	0.0003***	0.0003***
	(0.0001)	(0.0003)	(0.0000)	(0.0000)	(0.0001)
2001	-0.0001	-0.0002**	0.0015**	0.0002	0.0004***
	(0.0001)	(0.0001)	(0.0007)	(0.0002)	(0.0001)
2002	-0.0001**	0.0014*	0.0003*	0.0004	-0.0002
	(0.0001)	(0.0008)	(0.0002)	(0.0004)	(0.0001)
2003	0.0001	0.0042**	0.0006***	0.0012***	0.0004**
	(0.0002)	(0.0017)	(0.0002)	(0.0004)	(0.0002)
2004	-0.0003	0.0040	0.0002	0.0000	0.0003
	(0.0005)	(0.0028)	(0.0002)	(0.0003)	(0.0006)
2005	0.0001	0.0027***	0.0012***	-0.0000	-0.0002
	(0.0002)	(0.0010)	(0.0004)	(0.0006)	(0.0003)
Observations	1496	1496	1496	1496	1496
Counties	136	136	136	136	136
R^2	0.22	0.52	0.38	0.48	0.36

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in outmigration flow between inundated and non-inundated counties relative to the omitted base year in 1995. The dependent variables are the log fraction of outmigrants by reported reason of migration. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 7: Estimated Difference in Migrant Composition by Inundation Share

	Dependent var: Pre-determined individual demographics (fraction)				
	Han	Rural (35+)	Literate (18+)	Edu Years (18+)	Age 1st Marriage (35-60)
1996	-0.001	0.012	0.004	-0.023	0.194
	(0.002)	(0.011)	(0.005)	(0.081)	(0.193)
1997	-0.003	-0.001	0.002	0.009	0.146
	(0.002)	(0.010)	(0.005)	(0.079)	(0.154)
1998	-0.003	-0.001	0.000	0.029	0.302**
	(0.002)	(0.010)	(0.005)	(0.075)	(0.151)
1999	-0.003*	0.003	-0.001	-0.029	0.221
	(0.002)	(0.009)	(0.003)	(0.047)	(0.161)
2000	-0.005**	0.005	0.004	0.014	0.124
	(0.002)	(0.009)	(0.005)	(0.072)	(0.148)
2001	-0.001	0.013	0.005	-0.010	0.355
	(0.002)	(0.015)	(0.005)	(0.115)	(0.224)
2002	-0.004	0.008	-0.001	0.026	0.157
	(0.003)	(0.016)	(0.007)	(0.057)	(0.165)
2003	-0.004	-0.013	0.007	0.057	0.361**
	(0.003)	(0.010)	(0.005)	(0.051)	(0.176)
2004	-0.003	-0.007	0.006	0.115*	0.170
	(0.003)	(0.011)	(0.005)	(0.066)	(0.152)
2005	-0.002	-0.017	0.004	0.077	0.219
	(0.002)	(0.011)	(0.005)	(0.084)	(0.142)
Observations	1496	1414	1496	1496	1392
Counties	136	136	136	136	136
R^2	0.92	0.56	0.15	0.56	0.28

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Notes: This table shows the estimated differences in migrant composition between inundated and noninundated counties, relative to migrants that moved in 1995. The dependent variables are characteristics of migrants that are unlikely to be affected by the construction of TGD. Specifically, I consider the share of migrants moving in each year between 1996 and 2005 that are 1) of Han ethnicity; 2) rural Hukou holders 35 and above; 3) literate and 18 and above. I also consider 4) the average years of schooling for those 18 and above and 5) age at first marriage for those between 35 and 60. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 8: Estimated Difference in Migrants' Employment Status by Inundation Share

	Overall	By Status - Unemployed and Job Searching		By Status - Not in Labor Force			
	Unemployed	Recent Graduate	Lost Job	Retired	Homemaking	Disabled	Other
1996	0.0005***	0.0000**	0.0001***	0.0002***	-0.0000	0.0000	0.0000
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
1997	0.0008**	0.0001***	0.0001***	0.0003	0.0001	0.0000	0.0001***
	(0.0004)	(0.0000)	(0.0000)	(0.0002)	(0.0001)	(0.0000)	(0.0000)
1998	0.0010***	0.0001***	0.0002***	0.0003**	0.0001**	0.0000	0.0000*
	(0.0004)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
1999	0.0009***	0.0000	0.0002***	0.0003**	0.0000	0.0000	0.0001***
	(0.0002)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
2000	0.0010***	0.0001***	0.0002***	0.0003**	-0.0000	0.0000	0.0001**
	(0.0003)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
2001	-0.0000	-0.0000	0.0000	0.0001	-0.0000	-0.0000	-0.0001
	(0.0002)	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0000)	(0.0001)
2002	0.0006	-0.0000	0.0000	0.0007	-0.0001	0.0001	-0.0002
	(0.0004)	(0.0000)	(0.0001)	(0.0006)	(0.0002)	(0.0001)	(0.0001)
2003	0.0035***	-0.0000	0.0002	0.0021***	-0.0003	-0.0001	0.0012**
	(0.0011)	(0.0000)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0006)
2004	0.0022	-0.0000	0.0001	0.0018*	0.0001	-0.0000	0.0002
	(0.0019)	(0.0000)	(0.0001)	(0.0009)	(0.0002)	(0.0001)	(0.0004)
2005	0.0025	0.0000	0.0001	0.0019***	-0.0003	-0.0002	0.0006*
	(0.0017)	(0.0001)	(0.0002)	(0.0006)	(0.0003)	(0.0001)	(0.0003)
Observations	1496	1496	1496	1496	1496	1496	1496
Counties	136	136	136	136	136	136	136
R^2	0.49	0.17	0.27	0.49	0.44	0.24	0.41

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in employment status of migrants that moved in each year between 1996 and 2005 in inundated and non-inundated counties, relative to migrants that moved in 1995. The dependent variables are the log fraction of unemployed migrants overall and by job search status, including recent graduates, recent unemployed, retired, homemaker, disabled, and others not in the labor force. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 9: Estimated Difference in Unemployed Migrants' Income Source by Inundation Share

	Dependent var: Income source of the unemployed (log fraction)					
	Pension	Allowance/Welfare	Family Support	Property Income	Insurance	Others
1996	0.0002***	0.0000	0.0002***	0.0000	0.0000	0.0000
	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
1997	0.0003	0.0000***	0.0004*	0.0000	-0.0000	0.0001***
	(0.0002)	(0.0000)	(0.0002)	(0.0000)	(0.0000)	(0.0000)
1998	0.0003**	0.0001*	0.0005***	-0.0000	-0.0000	0.0001*
	(0.0001)	(0.0001)	(0.0002)	(0.0000)	(0.0000)	(0.0000)
1999	0.0003**	0.0000	0.0005***	0.0000	-0.0000	0.0001*
	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0001)
2000	0.0003**	0.0001***	0.0005***	0.0000	0.0000	0.0002***
	(0.0001)	(0.0000)	(0.0002)	(0.0000)	(.)	(0.0001)
2001	0.0001	0.0000	-0.0001	-0.0000*	-0.0000	0.0000
	(0.0001)	(0.0000)	(0.0002)	(0.0000)	(0.0000)	(0.0000)
2002	0.0007	-0.0000*	-0.0001	-0.0000	0.0000	-0.0001**
	(0.0006)	(0.0000)	(0.0003)	(0.0000)	(.)	(0.0000)
2003	0.0026***	0.0002	0.0008	-0.0000	0.0000	0.0000
	(0.0003)	(0.0001)	(0.0010)	(0.0000)	(0.0000)	(0.0001)
2004	0.0019*	0.0000	0.0004	-0.0000	-0.0000	0.0001
	(0.0010)	(0.0001)	(0.0010)	(0.0000)	(0.0000)	(0.0001)
2005	0.0019***	0.0001	0.0002	0.0002	0.0001	0.0001
	(0.0007)	(0.0001)	(0.0008)	(0.0002)	(0.0001)	(0.0002)
Observations	1496	1496	1496	1496	1496	1496
Counties	136	136	136	136	136	136
R^2	0.50	0.20	0.40	0.16	0.12	0.18

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in the income source of unemployed migrants that moved in each year between 1996 and 2005 in inundated and non-inundated counties, relative to unemployed migrants that moved in 1995. The dependent variables are the log fraction of unemployed migrants relying on pension, welfare, family support, property income, and insurance for the main source of income. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 10: Estimated Difference in Overall and Sectoral Employment by Inundation Share

	Employment Overall		Primary Sector		Secondary Sector		Tertiary Sector	
	(1) All	(2) Non-Migrant	(3) All	(4) Non-Migrant	(5) All	(6) Non-Migrant	(7) All	(8) Non-Migrant
2000	-0.006***	-0.007***	0.015*	0.018*	-0.022***	-0.022***	0.004	0.002
	(0.002)	(0.002)	(0.009)	(0.010)	(0.008)	(0.008)	(0.004)	(0.004)
2005	-0.004	-0.007*	0.020	0.021	-0.032***	-0.032***	0.004	0.004
	(0.003)	(0.004)	(0.014)	(0.015)	(0.011)	(0.011)	(0.005)	(0.005)
2010	0.002		0.015*		-0.032***		0.005	
	(0.002)		(0.008)		(0.009)		(0.007)	
Observations	508	381	508	381	508	381	508	381
Counties	127	127	127	127	127	127	127	127
R^2	0.84	0.85	0.88	0.87	0.73	0.76	0.85	0.83
μ_y	0.55	0.56	0.48	0.52	0.09	0.09	0.21	0.17
σ_y	0.06	0.06	0.19	0.19	0.10	0.11	0.14	0.13

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in overall and sectoral employment between inundated and non-inundated counties relative to the omitted base year in 1990. It compares results from the full sample (columns 1, 3, 5, 7) with that of a subsample with only nonmigrants (columns 2, 4, 6, 8). The dependent variables are the log fraction of population 15 years old and above who are employed overall and in the primary, secondary and tertiary sectors. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 11: Estimated Difference in Occupational Employment by Inundation Share

	Farming, Fishing and Forestry		Production and Transportation	
	All	Non-Migrant	All	Non-Migrant
Panel A: Primary and Secondary Sectors				
2000	0.012	0.015	-0.021***	-0.021***
	(0.008)	(0.009)	(0.006)	(0.006)
2005	0.006	0.008	-0.031***	-0.026***
	(0.018)	(0.019)	(0.008)	(0.008)
2010	0.013		-0.031***	
	(0.008)		(0.007)	
R^2	0.87	0.86	0.72	0.76
	Commercial and Service		Administrative and Support	
	All	Non-Migrant	All	Non-Migrant
Panel B: Tertiary Sector				
2000	0.006**	0.003	0.005***	0.004***
	(0.003)	(0.003)	(0.001)	(0.001)
2005	0.007	-0.000	0.009***	0.011**
	(0.005)	(0.004)	(0.002)	(0.004)
2010	0.009***		0.006***	
	(0.002)		(0.002)	
R^2	0.88	0.84	0.82	0.79
Observations	508	381	508	381
Counties	127	127	127	127

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the estimated differences in occupational employment between inundated and non-inundated counties relative to the omitted base year in 1990. It compares results from the full sample with that of a subsample with only nonmigrants. The dependent variables are the log fraction of population 15 years old and above who are employed in 4 harmonized occupations. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county.

Table 12: Estimated Difference in Industry Size and Firm Composition by Inundation Share

	Industry Size		Firm Operation Status			Firm Size		
	All Firms	Firms Started Before 1995	In Operation	Under Construction	Ceased Operation	Large	Medium	Small
1999	-0.001	-0.007	0.004	-0.038	-0.034	-0.028*	-0.025	0.002
	(0.006)	(0.006)	(0.006)	(0.024)	(0.025)	(0.015)	(0.018)	(0.006)
2000	0.015	0.001	0.020**	-0.038	-0.024	-0.002	-0.020	0.018**
	(0.009)	(0.010)	(0.009)	(0.024)	(0.022)	(0.013)	(0.015)	(0.008)
2001	0.014	0.009	0.017	-0.036	-0.013	0.014	0.008	0.011
	(0.013)	(0.017)	(0.012)	(0.026)	(0.026)	(0.013)	(0.041)	(0.012)
2002	0.016	0.010	0.022	-0.039	-0.026	0.013	0.008	0.014
	(0.015)	(0.015)	(0.015)	(0.028)	(0.022)	(0.018)	(0.044)	(0.012)
2003	0.008	0.014	0.010	-0.032	0.027	-0.025	-0.013	0.010
	(0.020)	(0.016)	(0.021)	(0.029)	(0.049)	(0.044)	(0.037)	(0.019)
2004	-0.008	-0.001	-0.002	-0.037	-0.027	0.000	0.000	0.000
	(0.027)	(0.020)	(0.028)	(0.025)	(0.023)	(.)	(.)	(.)
2005	-0.010	-0.011	-0.006	-0.039	-0.025	-0.037	-0.047	-0.004
	(0.042)	(0.029)	(0.044)	(0.024)	(0.023)	(0.055)	(0.047)	(0.041)
2006	-0.027	-0.021	-0.019	-0.038	-0.045*	-0.044	-0.043	-0.022
	(0.045)	(0.034)	(0.046)	(0.024)	(0.027)	(0.059)	(0.047)	(0.044)
2007	-0.040	-0.020	-0.035	-0.035	-0.035	-0.051	-0.058	-0.035
	(0.050)	(0.036)	(0.051)	(0.026)	(0.024)	(0.069)	(0.059)	(0.047)
Observations	1380	1380	1380	1380	1380	1242	1242	1242
Counties	138	138	138	138	138	138	138	138
R^2	0.89	0.86	0.88	0.26	0.24	0.74	0.80	0.88

Notes: This table shows the estimated differences in industry size and firm composition between inundated and non-inundated counties relative to the omitted base year in 1998. The dependent variables are the log number of industrial enterprises in total, those that started before 1995, have an operation status of in operation, under construction, ceased operation, and have a size of large, medium and small. Firm size categories are based on the number of employees (N) and sales revenue (R, in millions of CNY). In the sample, the average cutoffs for "Large" firms are around $N \geq 1000$, $R \geq 400$; for "Medium" firms are $300 \leq N \leq 1000$, $50 \leq R \leq 400$; and for "Small" firms are $100 \leq N \leq 300$, $5 \leq R \leq 50$. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. Analysis uses establishment-level data from the Annual Survey of Industrial Enterprises.

Table 13: Estimated Difference in Employment and Payroll of Firms by Inundation Share

		Overall		Primary Operation	
	Total Employee	Wage	Benefit	Wage	Benefit
1999	-0.079***	-0.078***	-0.116***	-0.082***	-0.121***
	(0.015)	(0.016)	(0.021)	(0.015)	(0.020)
2000	-0.048***	-0.035*	-0.054**	-0.047***	-0.071***
	(0.016)	(0.019)	(0.026)	(0.014)	(0.026)
2001	-0.046***	-0.025	-0.062***	-0.028	-0.069***
	(0.014)	(0.017)	(0.019)	(0.019)	(0.021)
2002	-0.040**	-0.025	-0.059**	-0.030	-0.067***
	(0.017)	(0.019)	(0.023)	(0.021)	(0.024)
2003	0.000	-0.006	-0.048*	-0.011	-0.059**
	(.)	(0.023)	(0.028)	(0.024)	(0.030)
2004	-0.041***	-0.025	-0.052**	-0.031*	-0.083***
	(0.014)	(0.017)	(0.026)	(0.017)	(0.022)
2005	-0.046**	-0.034*	-0.078***	-0.051***	-0.093***
	(0.018)	(0.018)	(0.027)	(0.019)	(0.026)
2006	-0.047***	-0.033*	-0.069***	-0.041**	-0.073***
	(0.016)	(0.017)	(0.020)	(0.018)	(0.020)
2007	-0.057***	-0.044**	-0.060**	-0.051***	-0.074***
	(0.015)	(0.019)	(0.025)	(0.020)	(0.026)
Observations	1242	1380	1379	1380	1379
Counties	138	138	138	138	138
R^2	0.74	0.65	0.60	0.60	0.60

Notes: This table shows the estimated differences in employment and payroll of industrial firms in inundated and non-inundated counties relative to the omitted base year in 1998. The dependent variables are the log number of employees, and log wages and benefits for firms' overall and primary operations. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. Analysis uses establishment-level data from the Annual Survey of Industrial Enterprises.

Table 14: Estimated Difference in Asset, Sales, Output and Profit of Firms by Inundation Share

	Asset		Sales		Output		Profit	
	Fixed	Current	Revenue	Cost	Total	Sales	Total	Operating
1999	-0.072***	-0.052**	-0.070***	-0.076***	-0.037***	-0.036***	0.128	0.214
	(0.018)	(0.023)	(0.021)	(0.024)	(0.013)	(0.013)	(0.086)	(0.776)
2000	-0.059**	-0.047*	-0.052*	-0.055*	-0.014	-0.017	0.145*	0.287
	(0.027)	(0.028)	(0.028)	(0.031)	(0.019)	(0.018)	(0.074)	(0.786)
2001	-0.049**	-0.041	-0.042	-0.047	0.000	-0.004	0.139**	0.264
	(0.024)	(0.028)	(0.030)	(0.033)	(0.020)	(0.022)	(0.068)	(0.793)
2002	-0.059	-0.025	-0.048	-0.054	-0.013	-0.014	0.156**	0.264
	(0.049)	(0.032)	(0.032)	(0.035)	(0.021)	(0.021)	(0.076)	(0.792)
2003	-0.001	0.012	-0.020	-0.020	0.001	-0.002	0.136*	0.206
	(0.026)	(0.022)	(0.019)	(0.019)	(0.018)	(0.020)	(0.078)	(0.791)
2004	-0.077**	-0.018	-0.055**	-0.059**	0.000	0.000	0.109	0.214
	(0.035)	(0.022)	(0.026)	(0.028)	(.)	(.)	(0.081)	(0.790)
2005	-0.088**	-0.018	-0.054*	-0.058*	-0.020	-0.020	0.074	0.174
	(0.040)	(0.028)	(0.028)	(0.031)	(0.021)	(0.021)	(0.075)	(0.795)
2006	-0.088**	-0.024	-0.051**	-0.054*	-0.020	-0.019	0.063	0.191
	(0.038)	(0.025)	(0.025)	(0.028)	(0.019)	(0.020)	(0.078)	(0.796)
2007	-0.065***	-0.011	-0.062**	-0.067**	-0.025	-0.025	0.110	0.243
	(0.022)	(0.022)	(0.026)	(0.028)	(0.019)	(0.019)	(0.089)	(0.791)
Observations	1380	1380	1380	1380	1242	1242	1154	1069
Counties	138	138	138	138	138	138	138	138
R^2	0.56	0.62	0.61	0.63	0.77	0.76	0.61	0.54

Notes: This table shows the estimated differences in asset, sales, output and profit of industrial firms in inundated and non-inundated counties relative to the omitted base year in 1998. The dependent variables are log fixed and current asset, log sales revenue and cost, log overall and sales output, and log overall and operating profit. Using equation 1, these are regressed on the fraction of county area inundated interacted with a year dummy, county and state-by-year fixed effects. Robust standard errors are clustered by county. Analysis uses establishment-level data from the Annual Survey of Industrial Enterprises.

Within but Without: Involuntary Displacement and Economic Development

Online Appendix

Alice Tianbo Zhang*

October 2018

Job Market Paper

Click Here For Latest Version

Contents

*Ph.D. Candidate in Sustainable Development, School of International and Public Affairs, Columbia University. 420 W. 118th St., New York, NY 10027. Email: alice.zhang@columbia.edu

A1 Harmonize Historical Census

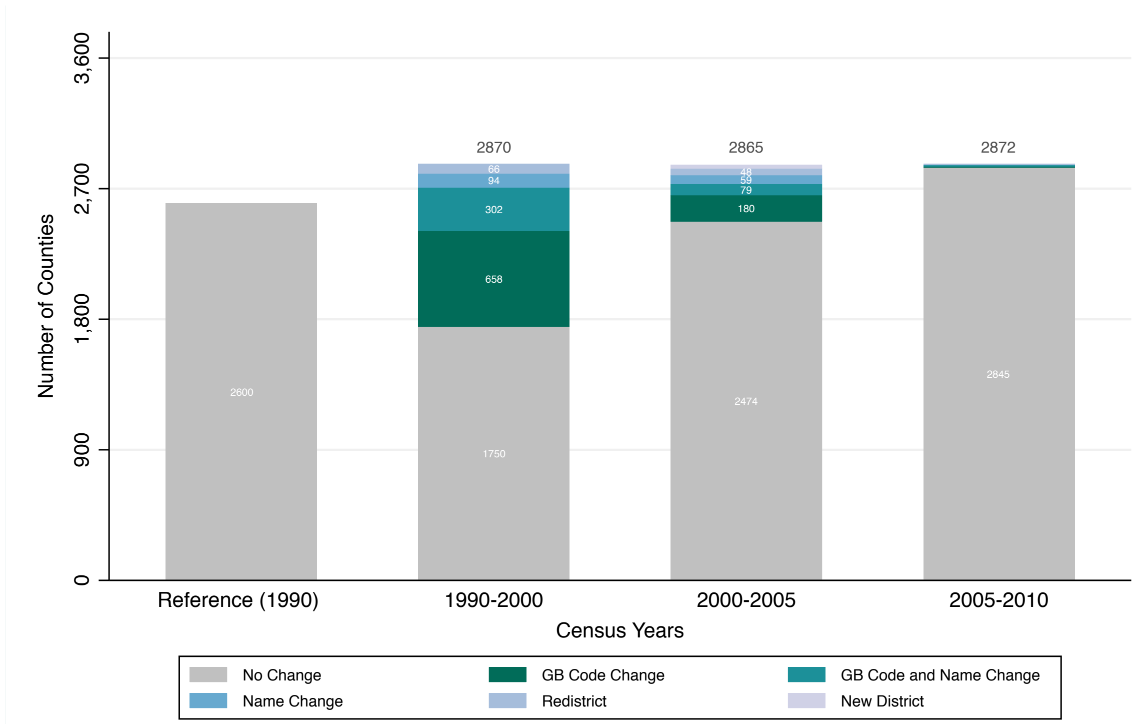
This section describes the data cleaning and harmonization procedures I use to obtain a consistent county-level panel using historical census. Table A1 shows the list of datasets used in the paper and their sources. Two main challenges underlie the data harmonization process. First, county names, codes and boundaries have changed considerably over time. Second, the census questionnaires have been modified over time, such that categories are added or subtracted for existing variables and new variables are created.

A1.1 County Boundary Crosswalk

For geographic comparability in the main analysis sample, I construct crosswalk relationship files for all counties in China for the 1990, 2000, 2005 and 2010 census. To the best of my knowledge, no such crosswalk files are publicly available through either the government or previous research. I create the crosswalk file to have consistent 1990 county boundary definitions by first eliminating duplicate records in the individual-level census and then performing an one-to-one merge between 1990 and 2000 county administrative division codes (GB code). In cases where the GB codes have changed, I use an one-to-one matching algorithm based on the county Chinese name. For the remaining cases unmatched by either GB code or county name, I manually verify the boundary changes and redistricting history for each county using publicly available records. Based on historical administrative division files published by the Ministry of Civil Affairs, I then assign the correct GB code and name to redistricted counties. I separately perform this crosswalk process between two consecutive census years: 1990-2000, 2000-2005, 2005-2010, and then merge them together by 1990 county GB code.

I identify five main types of changes in county geography: 1) GB code only; 2) Name only; 3) Both GB code and name; 4) Redistrict, including one county splitting into multiple smaller counties, multiple counties aggregating into one larger county, and other idiosyncratic redrawing of boundaries; 5) New District. Figure A1 shows the number of counties in each census year and those that have experienced any changes. The most significant change occurred between 1990 and 2000, when almost 40% of counties underwent one of the above changes and the total number of counties increased from 2,600 in 1990 to 2,870 in 2000. As shown in Figure A2, most counties in the two provinces affected by the Three Gorges Dam (TGD), Chongqing and Hubei, have consistent boundaries between 1990 and 2010. One notable exception is Chongqing, which became a direct-controlled municipality and separated from Sichuan province in 1997. As a result, between 1990 and 2000, the former Chongqing City was split into smaller sub-districts. For the main sample, I exclude counties that are newly created, have idiosyncratic redrawn boundaries or are aggregated since 2000. For backward consistency, I adjust the data to hold the 1990 geographic boundaries constant by summing up values for counties that were split in later years. The remaining sample consists of 2,460 counties with consistent county geography between 1990 and 2010.

Figure A1: County Administrative Crosswalk - Summary of Changes



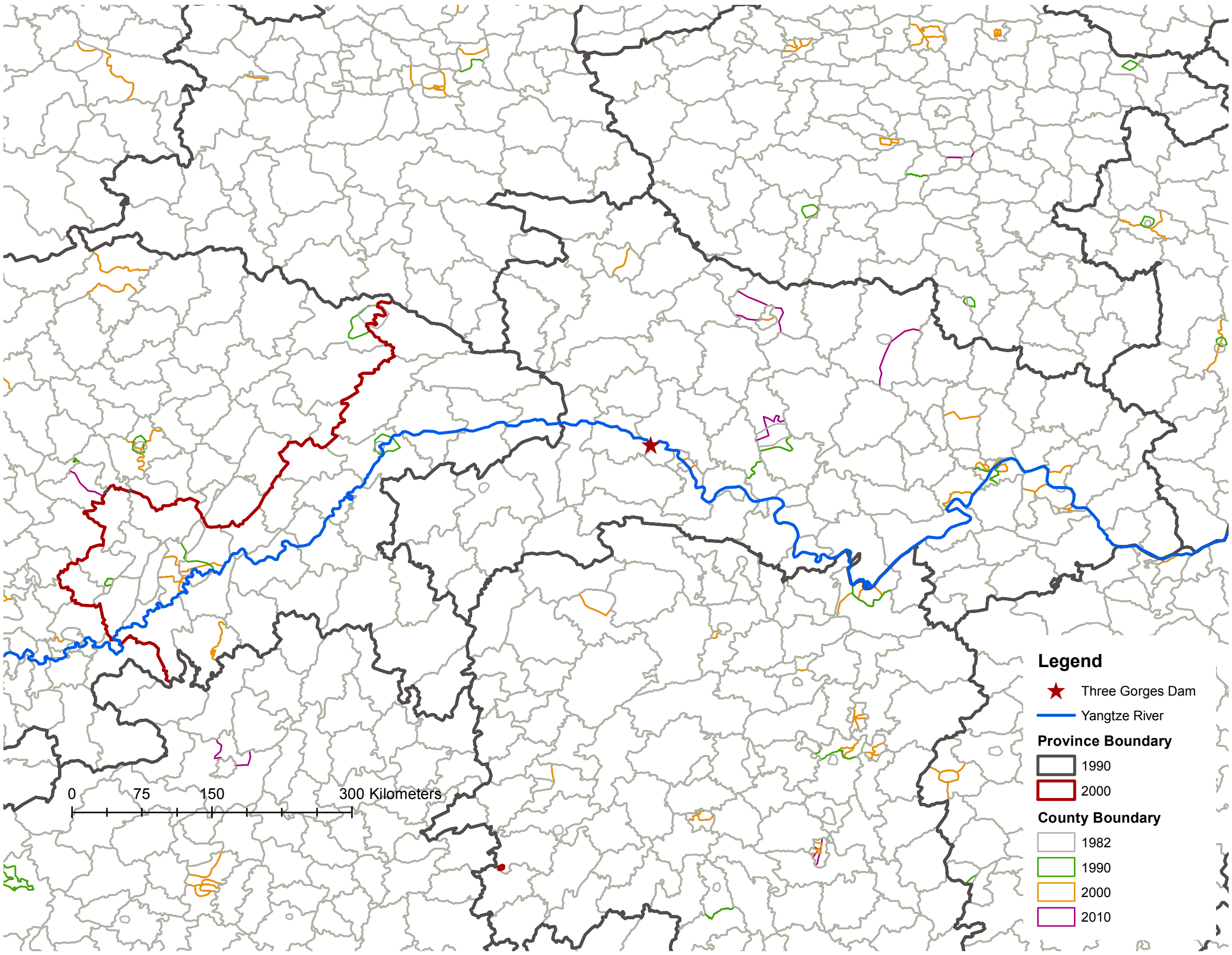
Notes: This figure shows the number of counties by the type of geography changes between two consecutive census years: 1990-2000, 2000-2005, 2005-2010. Census data is adjusted to hold the 1990 county boundaries constant. See text for details.

Table A1: List of Datasets and Sources

Type	Dataset Name	Year	Unit of Obs. / Resolution	Source
Satellite	Landsat 5 Thematic Mapper Level-1 Data Products	1990s, 2010s	30 meters	USGS
	Shuttle Radar Topographic Mission Digital Elevation Model	2000	250 meters	NASA
Population	China Decennial Census	1990, 2000	Individual	National Bureau of Statistics (NBS)
	China 1% Population Survey	2005	Individual	NBS
	Historical China County Population Census Data	1982, 1990, 2000	County	UMich China Data Center (UMCDC)
	China 2010 County Population Census Data	2010	County	UMCDC
	China 2000 Township Population Census Data	2000	Township	UMCDC
	Tabulation of the 1990 Population Census	1990	Township	NBS
Industry	Annual Survey of Industrial Enterprises (ASIE)	1998-2007	Firm	NBS
Auxillary	Archival Government Legislations	Various	Document	Wanfang Data
	Global Dam Tracker (GDAT)	Various	Dam	Zhang et al. (2018)

Notes: This table shows the datasets used in the paper and their corresponding sources. Please see data section for details.

Figure A2: Map of Administrative Boundary Changes



Notes: This map illustrates province and county administrative boundary changes between 1982 and 2010. The map is zoomed to the two provinces affected by the TGD - Chongqing and Hubei. County boundaries for 1982, 1990, 2000 and 2010 are overlaid with province boundaries in 1990 and 2000. There is no change in province boundaries in this region between 2000 and 2010. The most notable change on the map is the creation of Chongqing (marked by the maroon bold line in 2000), formerly a sub-provincial city in Sichuan, as a direct-controlled municipality in 1997. This is largely a change in administrative jurisdiction with the county boundaries intact. Other minor changes involve the splitting of large counties into smaller ones, such as the division of Chongqing city into districts. Overall, I find few boundary changes that may be potentially endogenous to the construction of TGD. Historical county boundary GIS data comes from the China Data Center, University of Michigan. Census data is adjusted to hold the 1990 county boundaries constant. See text for details.

A1.2 Industry and Occupation Crosswalk

For variable consistency, I identify a comparable set of variables across the census years and recode them using a process similar to the IPUMS-International data harmonization procedure[1]. While most variables in the census have comparable categories over time, industry of employment information is reported differently over time. In particular, the 1990 and 2000 census contain finer level three-digit industry group information, whereas the 2005 and 2010 census only have two-digit industry division or top-level industry section information respectively. To accommodate the lowest common denominator, I first perform a many-to-one merge of reported industry codes in the individual-level data with two or three-digit Chinese Industrial Classification (CIC) codes published in 1984, 1994 or 2002 corresponding to a given census year. I create a crosswalk to account for CIC code changes across the years. I then aggregate these codes into 20 top-level industry sections comparable with the International Standard Industrial Classification of All Economic Activities (ISIC) system[2]. Table A2 shows the 20 harmonized industries used in the sample and their corresponding sectors. I use a similar crosswalk procedure to create 6 harmonized occupation groups shown in Table A3.

[1]As explained by IPUMS-I here: `https://international.ipums.org/international/harmonization.shtml`.

[2]The ISIC classifies economic activity into four hierarchical levels: Section, Division (two-digit), Group (three-digit) and Class (four-digit). For ISIC Rev.4, industries are aggregated into 21 sections as detailed here: `https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf`.

Table A2: Harmonized Industries and Corresponding Sectors

Sector	CIC Industry Section	Industry Description
Primary	A	Agriculture, Forestry, Husbandry and Fishery
	B	Mining
Secondary	C	Manufacturing
Tertiary	D	Electricity, Gas and Water
	E	Construction
	F	Transportation and Communication
	G	Information technology and Computer Systems
	H	Wholesale and Retail Trade
	I	Hotel and Restaurant
	J	Financial Service and Insurance
	K	Real Estate
	L	Rental and Business Services
	M	Professional, Scientific and Technical Activities
	N	Water and Environmental Management
	O	Residential and Other Services
	P	Education
	Q	Health and Social Work
	R	Art, Culture and Sports
	S	Public Administration and Defense
	T	Other Industry

Notes: This table shows the 20 harmonized industries of employment across the 1990, 2000, 2005 and 2010 census and their corresponding sectors.

Table A3: Harmonized Occupations and Corresponding Sectors

Sector	Occupation Description
Primary	Farming, Fishing and Forestry Workers
Secondary	Production and Transportation Workers
Tertiary	Commercial and Service Workers Administrative and Support Workers Professional and Technical Workers Executives of Government, Civil and Private Institutions

Notes: This table shows the 6 harmonized occupation of employment across the 1990, 2000, 2005 and 2010 census and their corresponding sectors.

A2 Create Treatment and Placebo Samples

A2.1 Inundation Map

I create an intensity of treatment measure by using Landsat 5 TM satellite data to calculate the fraction of county area inundated by the reservoir of TGD. In this section, I describe the procedures for creating the inundation maps.

First, I select 9 spatially contiguous scenes for the 1990s (pre-TGD) and 2010s (post-TGD) available through the U.S. Geological Survey Global Visualization Viewer (USGS GloVis). Each scene covers an area of approximately 106 miles by 115 miles and together they span the reservoir of TGD in Hubei and Chongqing provinces. Table A4 lists the individual scenes, their date of acquisition, the scene center location designated by the Worldwide Reference System (WRS) Path and Row numbers, and the percent of cloud cover. Since Landsat 5 TM has a 16-day revisit time, individual scenes from different paths are usually not acquired on the same day. To facilitate inter-temporal comparison, I select scenes with minimum cloud cover (cloud coverage < 10%) and from the dry season in Spring between April and May when possible (rainy season is from June - September) in the provinces under study. These restrictions ensure that atmospheric conditions are relatively stable and the estimated river area are not confounded by rainfall shocks. Second, I apply atmospheric corrections to each scene in ENVI, classify pixels based on the spectral profile of different land cover types, including substrate, water and vegetation, and extract pixels for the Yangtze River. Third, I mosaic the 9 scenes together to get a raster image of only the Yangtze River for the two time periods: 1990s and 2010s.

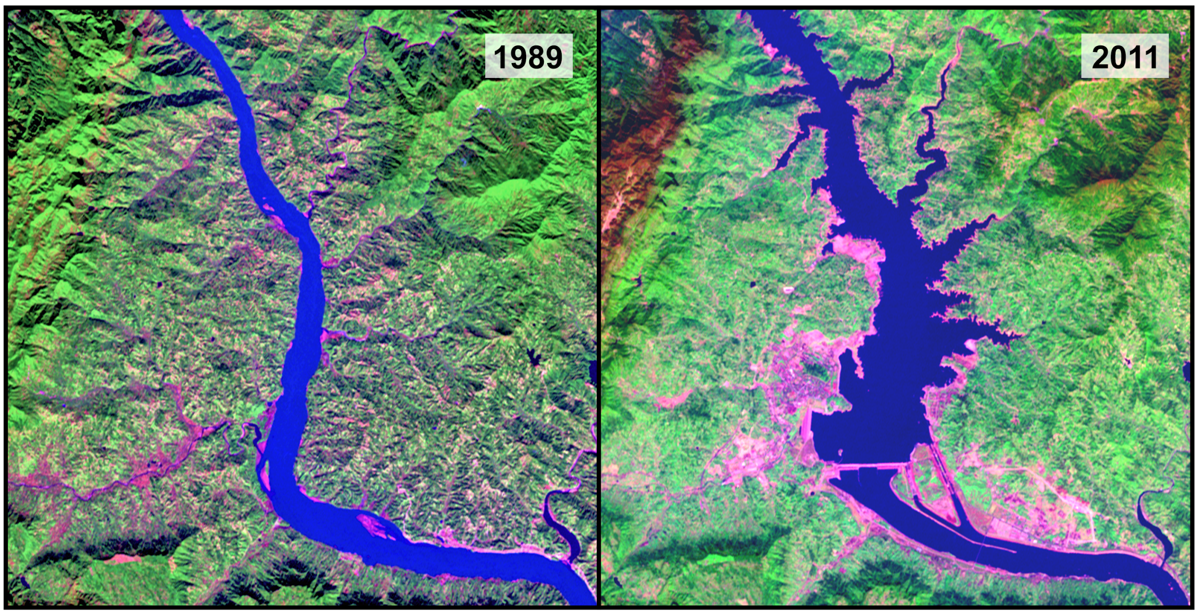
Figure A3 illustrates two raw satellite images acquired before and after the construction of TGD. The left image shows the region around the TGD in 1989 prior to construction, while the right image shows the same region in 2011 when construction was completed and electricity generation had begun. The structure of the dam standing across the river can be clearly seen in the right image. There is a notable expansion of the river channel upstream of TGD due to the filling of the reservoir. Using the change detection algorithm in ENVI, I calculate that the section of the Yangtze River in Hubei and Chongqing covered an area of about 713 km^2 before the TGD in 1990s and 1035 km^2 after construction was completed in 2010s, a change of about 322 km^2. This is likely a conservative estimate of the actual inundation area, because tributaries and smaller streams connected to the main channel of Yangtze River are excluded.

Table A4: Landsat Scene List

Time Period	Date	WRS Path/Row	Cloud Cover (%)
Pre-TGD	9/13/1989	125/38	2
	9/13/1989	125/39	0
	5/18/1990	126/38	0
	5/18/1990	126/39	0
	5/9/1990	127/38	0
	5/9/1990	127/39	0
	5/9/1990	127/40	0
	5/24/1993	128/39	0
	5/24/1993	128/40	0
Post-TGD	4/19/2011	125/38	0
	4/19/2011	125/39	0
	4/26/2011	126/38	1
	4/26/2011	126/39	0
	5/19/2011	127/38	0
	5/19/2011	127/39	0
	5/19/2011	127/40	0
	5/23/2010	128/39	0
	5/23/2010	128/40	0

Notes: This table lists the individual Landsat 5 TM scenes used to calculate the area inundated by the reservoir of TGD. For each scene, it shows the date of acquisition, the scene center location designated by the Worldwide Reference System (WRS) Path and Row numbers, and the percent of cloud cover. I select 9 scenes for the 1990s and 2010s each covering an area of approximately 106 miles by 115 miles. For each time period, I pre-process each scene and then mosaic them together to create a raster image for the Yangtze River. See text for details.

Figure A3: Yangtze River Extent Before (1989) vs. After (2011)



Notes: This figure shows two raw satellite scenes acquired by Landsat 5 TM before and after the construction of TGD. The left image shows the region around the TGD in 1989 prior to construction, while the right image shows the same region in 2011 when construction was completed and electricity generation had begun. There is a clear expansion of the river channel due to the filling of the reservoir. I calculate an additional area of 322 km^2 was flooded due to the rising water levels in the reservoir. The structure of the dam standing across the river can also be clearly seen in the right image.

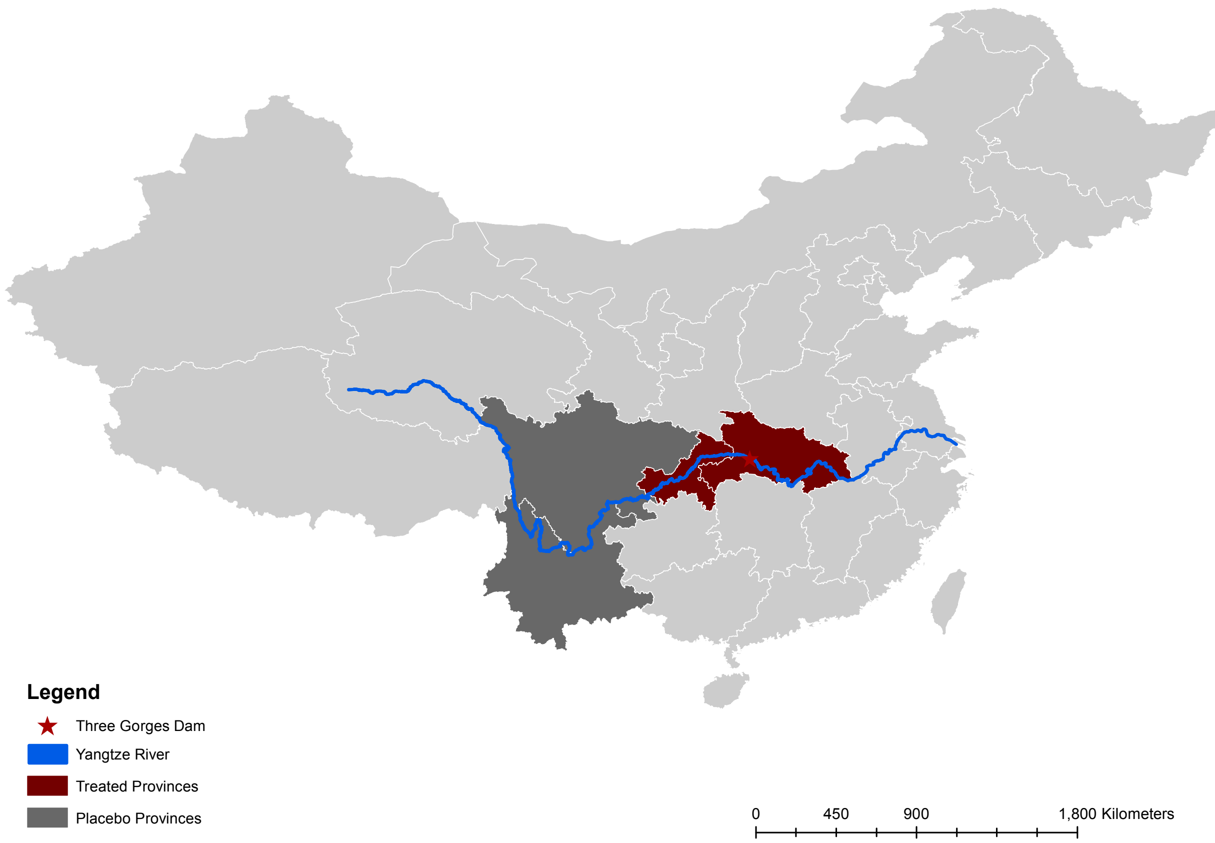
A2.2 Placebo Sample

To conduct a falsification exercise, I generate a sample of placebo counties that are near the Yangtze River outside of the catchment zone of TGD's reservoir. Using surface water occurrence data from the Global Surface Water Explorer (GSWE)[3], I create a shapefile that spans the entire Yangtze River from the Tibetan Plateau to the East China Sea. Since existing boundary maps of the Yangtze River have very coarse resolution and do not match the satellite data, I geo-rectify the shapefile to ensure that it matches the river pixel locations in GSWE. For each county, I then calculate its distance to the Yangtze River.

To facilitate comparisons with the treatment sample, I restrict the placebo sample to counties upstream of the TGD and outside of the reservoir catchment area. As shown in Figure A4, the placebo provinces are Sichuan and Yunnan, which are neighbors to the treatment provinces, Chongqing and Hubei. Counties in the placebo sample are considered 'pseudo-flooded' if they are within 50km of the Yangtze river. I also use different distance cutoffs between 0 and 50 in 5km increments to generate 11 placebo samples. Together, results from these placebo samples test the validity of the identifying assumption that in the absence of the TGD and inundation, flooded counties in the treatment provinces would have evolved similarly to non-flooded counties.

[3]The dataset is created using full-resolution images from the Landsat 5, 7 and 8 satellites. Data and documentation available here: `https://global-surface-water.appspot.com/`.

Figure A4: Map of Treatment and Placebo Samples



Notes: This figure shows provinces in the treatment and placebo samples along the Yangtze River. The placebo provinces are Sichuan and Yunnan and they are neighbors to the treatment provinces, Chongqing and Hubei. Counties in the placebo provinces are considered 'pseudo-flooded' if their distance to the Yangtze are within 50km. I use different cutoffs between 0 and 50 in 5km increments to generate 11 placebo samples. Shapefile for the Yangtze River is geo-rectified using satellite data from the Global Surface Water Explorer. See text for details.

A3 Construct Migration Flow

To construct an annual panel of outmigrants between 1995 and 2005, I use individual-level data from the 1 percent sample of the 2000 Population Census and the 15 percent sample of the 2005 1% Population Survey. In this section, I describe in detail how I calculate the migration flow.

A3.1 Data Description

The National Bureau of Statistics (NBS) of China has conducted decennial Population Census in 1953, 1964, 1982, 1990, 2000 and 2010. The Population Census covers all individuals who have Chinese nationality and reside in China at the census reference time. I use a random 1 percent sample of the 2000 Population Census with more than 11.8 million individual observations. The sample is unweighted and each observation carries an expansion factor of 100. During inter-census years, the NBS has conducted 1% Population Surveys in 1995, 2005 and 2015. For the 2005 survey, households were interviewed using a stratified multi-stage cluster sampling design and a questionnaire similar to the census long form[4]. The full sample consists of 16.9 million individuals from 77,417 census enumeration districts, representing 1.32 percent of the total population and covering all counties in China in 2005. I use a random 15 percent sample with more than 2.5 million individual observations, representing 0.2 percent of the total population. The NBS provides weight for each observation to account for the underlying survey sampling design.

I calculate origin-based outmigration flow using variables in the census that retrospectively record the county of origin, year of migration and reason of migration for each individual. The corresponding questions from the 2000 and 2005 data are listed below. In particular, for the 2000 census, an individual is considered an outmigrant from origin county c if R9 ≥ 3 and R10 is not missing. Similarly, for the 2005 population survey, an individual is considered an outmigrant from origin county c if $2 \leq$ R6 ≤ 3 and $2 \leq$ R8 ≤ 9.

<u>2000 Population Census</u>

- Year of migration (R9): "When did you move to the current town/township/neighborhood?"

    1. From the date born (skip to R14)
    2. Before October 31, 1995 (skip to R14)

[4]Details on the 2005 1% Population Survey sampling design is available on the NBS website: `http://www.stats.gov.cn/tjsj/ndsj/renkou/2005/renkou.htm`. Accessed: April 2, 2018.

3. Between November 1st, 1995 and December 31st, 1995
4. 1996
5. 1997
6. 1998
7. 1999
8. 2000

- County of origin (R10): “What is your previous place of permanent residence?"
  1. Within the current county, city or district
  2. Outside the current county, city or district, please specify: province, prefecture and county (city/district)
- Reason of migration (R12): “What is the reason of moving?"
  1. Work/business
  2. Job change
  3. Entry-level hire
  4. Study/training
  5. Demolition
  6. Marriage
  7. Dependents/family for work
  8. Join relatives
  9. Others

2005 Population Survey

- Year of migration (R8): “When did you leave the place of Hukou registration?"
  1. Never left (skip to R11)
  2. Less than 6 months
  3. 6 months to 1 year
  4. 1 to 2 years
  5. 2 to 3 years

6. 3 to 4 years
7. 4 to 5 years
8. 5 to 6 years
9. More than 6 years

- County of origin (R6): "Where is your place of Hukou registration?"

  1. This town (township/neighborhood)
  2. Another town (township/neighborhood) in this county (city/district)
  3. Another county (city/district), please specify province, prefecture and county (city/district)

- Reason of migration (R9): "What is the reason of leaving the place of Hukou registration?"

  1. Work/business
  2. Job change
  3. Entry-level hire
  4. Study/training
  5. Demolition
  6. Marriage
  7. Dependents/family for work
  8. Join relatives
  9. Affiliated Hukou attached to non-family or collective households
  10. Temporal business trip
  11. Others

A3.2 Panel Construction

While access to microdata allows me to calculate the migration flow between all origin-destination county pairs and account for the multi-dimensional spatial patterns of migration in a rigorous manner, several caveats should be noted. First, the earliest migration spell that can be constructed from the data is 1995. Individuals who have moved prior to November 1st, 1995 are considered non-migrants, because the census does not report their place of origin. In addition, 2000 is the first decennial census that asks retrospective questions on previous place of permanent residence and year of migration. To the best of my knowledge, individual-level data of the 2010 census has not been released to researchers. As a result, I can only construct origin-based migration spells between 1995 and 2005.

Second, the question on county of origin is asked slightly differently in 2000 and 2005. The 2000 census asks information on the previous county of permanent residence, whereas the 2005 survey asks information on the county of Hukou. These two places are the same for individuals who move directly from the county of Hukou registration to the current destination county, i.e. their previous county of permanent residence is the same as the county of Hukou. This can be a concern for individuals that have moved more than once in the past 5 years and have another permanent address distinct from their Hukou registration. Since the Hukou system is still fairly restrictive between 1995 and 2000, it is unlikely that individuals would move and change their permanent address more than once within a five-year period. It is therefore reasonable to assume that for the majority of the population, their county of permanent residence corresponds to the county of Hukou.

Third, the census only observes single migration spells and provides limited information on step or return migration. Consider first the case of step migration. Suppose an individual leaves her origin county of Hukou O_H for destination county D_1 in 1998 and then transit to another destination county D_2 in 2000. In both the 2000 and 2005 data, I would only observe the last destination D_2 and miss any intermediate destinations. While I would still correctly calculate origin-based migration flow for the county of Hukou or permanent residence O_H, the 2000 data will incorrectly attribute the year of departure from origin, because it asks for the year of arrival at destination. On the other hand, the 2005 data will correctly attribute the year of departure.

Now consider the case of return migration. Suppose an individual leaves her origin county of Hukou O_H for destination county D_1 in 1998 and then return to O_H in 2000. In the 2000 data, if the duration of stay at D_1 is longer than 6 months, the census would correctly capture this return migration flow. However, the 2005 data makes no distinction between people who have never left O_H and those who have left for more than 6 months and returned at the time of survey like in the 2000 data. As such, the 2005 data omits outmigration flows for people that have returned to their county of Hukou. Similarly, individuals who have moved and updated their Hukou registration within the last five years will be considered as non-migrants in the 2005 data. Such conversions are rare because the Hukou system poses considerable administrative and legal constraints on making frequent changes.

With these caveats in mind, I adjust the migration flow in 2001-2005 to account for the systematic undercounting of return migrants in the 2005 data. Since the two datasets have one year of overlap in 2000, comparing the number of outmigrants in 2000 measured in both datasets allows me to estimate the degree of undercounting. More specifically, I compute the adjustment factor according to equation 1, in which $outMig_i^{2000}$ and $outMig_i^{2005}$ are the number of migrants from province i in 2000 measured in the 2000 and 2005 datasets respectively. I then calculate the average across all 31 provinces in China and multiply the raw migrant counts in 2001-2005 by 3.97 to obtain comparable migration flows between 1995 and 2005. The main assumption here is that the fraction of return migrants is constant each year between 2001 and 2005. From the 2000 census, I find that around 60% of all migrants to a given county are returning to their county of Hukou registration in 1995. Taking this as the steady-state share of return migrants each year, the adjustment factor of 3.97 provides a reasonable estimate for the return flow.

$$AdjustmentFactor = \frac{1}{p}\sum_{i=1}^{p}\frac{outMig_i^{2000}}{outMig_i^{2005}} = 3.97 \tag{1}$$

Figure A5 shows the distribution of province-specific adjustment factors, which range from 1.3 to 5.8. The red vertical line is at the provincial average of 3.97. The systematic undercounting of return migrants can be seen in the left panel of Figure A6, where we observe a sharp drop in the number of migrants between 2001 and 2005. After adjustment, we get a more consistent migration flow in the right panel. For robustness, I check that province specific and provincial average adjustment factors give essentially the same migration flow results.

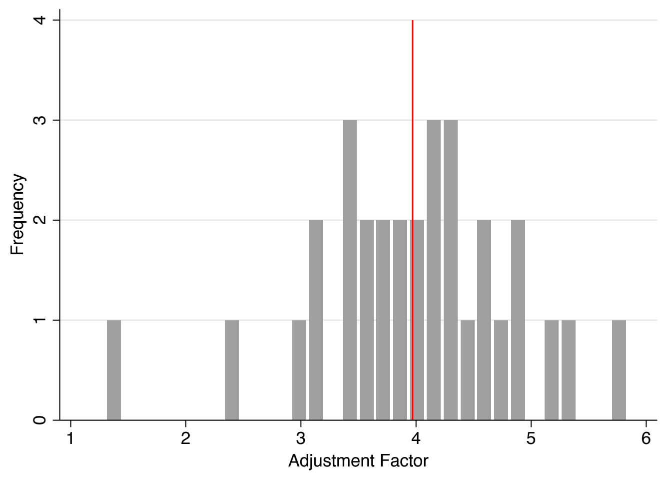


Figure A5: Histogram of Province-Specific Adjustment Factor

Notes: This histogram shows the distribution of province-specific adjustment factors to account for the systematic undercounting of return migrants in the 2005 survey data. The red vertical line is at the provincial average of 3.97. I use this number to scale the 2005 data to obtain consistent migration flows relative to the 2000 census.

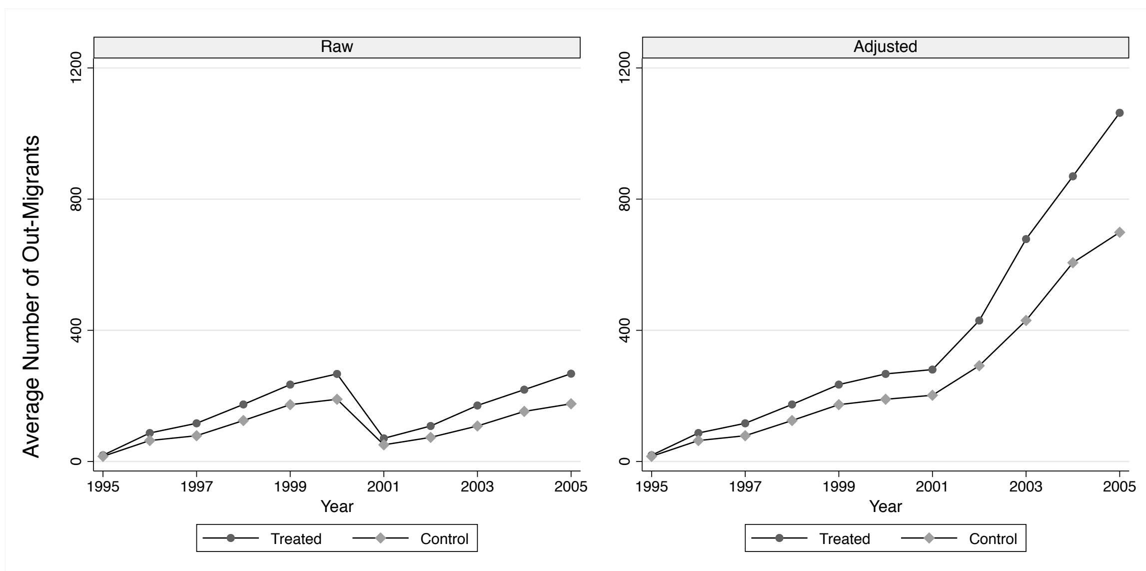


Figure A6: Comparison of Raw vs. Adjusted Migration Flow 1995-2005

Notes: These figures show the average number of outmigrants in treated and control counties between 1995-2005 using data from the 2000 census and 2005 population survey. The left panel shows the raw time series, where we observe a sharp drop in the number of migrants in 2001-2005. This highlights the systematic undercounting of return migrants in the 2005 data. The right panel shows that after adjustment, the migration flow is consistently measured during the sample period.