WWII Events Reduce Population in One Shanghai District While Increasing It in Others*

Impact Analysis of Japanese Occupation on Population Shifts Across Shanghai's Districts During WWII

Sandy Yu

December 3, 2024

This paper evaluates the impact of three historic events on the population dynamics of Shanghai during WWII: the Japanese occupation of the Chinese-administered area, France's surrender in WWII, and the Japanese occupation of the International Settlement. Data from the Virtual Shanghai research platform and historic publications are analyzed using a difference-in-difference model to estimate the effects of these events. The results indicate that the Japanese occupation of the Chinese-administered area led to an overall increase in Shanghai's population across all districts. In contrast, France's surrender during WWII had a negative impact on population numbers in all districts except the French Concession. Similarly, the Japanese occupation of the International Settlement caused population declines in all districts except the International Settlement itself. These findings underscore the complex demographic shifts resulting from geopolitical events and reveal how wartime occupation reshaped urban populations in uneven ways, reflecting resilience and displacement patterns.

Table of contents

*Code and data are available at: https://github.com/Jingying-yu/Shanghai_population_change

1 Introduction

The 1930s-1940s were marked by extreme volatility and chaos. World War II devastated regions across Europe and Asia. On the Western front, the Allies clashed with Nazi Germany, while in the East, Japanese forces invaded China. This paper examines the impact of key WWII-related events on the population dynamics of Shanghai.

The focus of this study is to evaluate how three significant historic events—the occupation of the Chinese-administered District, France's surrender in WWII, and the occupation of the International Settlement—affected population changes within Shanghai.

The Japanese occupation of the Chinese-administered area led to an overall increase in Shanghai's population across all districts, a counterintuitive result likely driven by an influx of refugees fleeing agricultural disruptions in rural areas for industrial and economic opportunities in the city. In contrast, France's surrender during WWII had a negative impact on population numbers in all districts except the French Concession, which retained relative stability due to its continued governance under French authority, albeit under Japanese surveillance. Similarly, the occupation of the International Settlement caused population declines in all districts except for the International Settlement itself, which remained one of the safest and most resource-secured areas in Shanghai despite the occupation. These findings reflect the complex interplay between governance, security, and economic opportunity in shaping population dynamics during wartime.

Understanding the demographic shifts caused by major historic events provides insight into the broader socio-economic and political factors at play during times of war. By examining these population changes, this paper highlights how external forces and governance structures shaped the resilience and migration patterns of urban populations during WWII. This research also underscores the importance of context-specific analyses when studying the impacts of conflict on urban demographics.

Analyses and findings in this paper are structured into several sections: Section 2 – Data, Section 3 – Model, Section 4 – Results, and Section 5 – Discussion. The Data section examines all variables and attributes kept for analysis, followed by an explanation of their measurement and purpose. The Model section defines the generalized linear model used for further analysis, explain its components, and presents model justifications. The Result section focuses on visualizing and presenting the model results through data presented in Data section. The Discussion section further evaluate the interpretations behind the model results presented in the previous section, and touches on any expanded topics as well as weaknesses and next steps.

2 Data

Data is cleaned and analyzed using the open-source statistical programming language R (R Core Team 2023) and supporting packages including tidyverse (Wickham et al. 2019), arrow (Richardson et al. 2024), readr (Wickham, Hester, et al. 2023), knitr (Xie 2023), kableExtra (Zhu 2023), modelsummary (Arel-Bundock 2023), rstanarm (Goodrich et al. 2022), dplyr (Wickham et al. 2023), and broom.mixed (Bolker et al. 2023). These tools facilitate data manipulation, visualization, modeling, and reporting. Detailed descriptions of each dataset are provided in the subsections below.

2.1 Data Sources

The primary dataset is sourced from the appendix of the research publication (Zou (1980)), which includes selected excerpts from historical government census records. Due to the age of the original census documents, they are no longer available online. To address gaps in the primary dataset, complementary data was obtained from the research platform Virtual Shanghai (Henriot and Contributors (2024)), providing crucial insights to fill missing areas.

2.2 Historical Background

Between 1936 and 1942, China was in a state of political and social upheaval, marked by increasing Japanese aggression and internal instability. Shanghai, a critical urban center, was divided into three districts: the Chinese-administered district, the International Settlement (primarily governed by the U.K. and U.S.), and the French Concession (under French governance). The International Settlement and French Concession operated with significant autonomy compared to the Chinese government's limited authority over its district. Each district occupied distinct proportions of Shanghai's urban landscape.

Reference Timeline:

1. 1937-08-13: Japanese armed forces entered Shanghai

2. 1937-11-12: Japanese armed forces claims occupation of Shanghai –> ends Chinese district

3. 1940-06-22: France surrendered to Germany –> French Concession stopped taking refugees

4. 1942-01: Japanese armed forces claims authority over International Settlement (which was mainly under the governance of U.K and U.S prior to this date)

5. 1945-08-15: Japan surrendered in WWII

6. 1945-10: most Japanese armed forces withdrew from Shanghai

2.3 Measurement & Methodology

The majority of the data used in this analysis were collected by police officers or hired local authorities within each district of Shanghai, as detailed in the source (Zou (1980), page 2).

This paper focuses on population data from 1936 to 1942, a timeframe determined by both historical relevance and data availability. The analysis centers on the impact of three major historical events, beginning with the Japanese occupation of the Chinese-administered district on November 12, 1937. Given that Japanese forces first entered Shanghai on August 13, 1937, the year 1936 is chosen as the baseline or "no-treatment" year for comparison. The analysis concludes with the year 1942, as district-specific population data after this year is unavailable. The lack of records for the period between 1943 and 1949 coincides with the political transition to the People's Republic of China, making 1942 the final year for which reliable district-based population data exists.

The survey process faced several challenges, primarily due to population volatility caused by wartime conditions. Population shifts were frequent and unpredictable as military conflicts unfolded. Additionally, survey efforts encountered varying degrees of difficulty across districts due to logistical and political constraints. Resistance from the local population, often rooted in political risks, further complicated data collection efforts. These challenges underline the complexities of accurately recording population dynamics during a period of significant unrest and upheaval.

2.4 Variables of Interest

This paper primarily evaluates three historic events between 1936 - 1942:

1. The Japanese occupation of Shanghai's Chinese-administrated District – 1937-11-12

2. France surrendered to Germany – 1940-06-22

3. Japanese armed forces claims authority over International Settlement – 1942-01

The first and third events should have very straight-forward consequences on population. The shifting over governing power to a foreign, hostile country would intuitive cause local population to panic and evacuate. In this paper, we initially assumed that Japanese occupation would decrease population in the corresponding area.

The effect of the second event is less direct. In 1940, European battlefield is still filled with WWII soldiers. When France surrendered to Germany in June, Japan – an ally of Germany, accepted France's new status as mutual German allies. The Japanese forces chose to stay out of the French Concession District in Shanghai. (User (n.d.)) The French Concession remained un-conquered until the Japanese withdrew from Shanghai in 1945.

2.4.1 Outcome variable

The outcome variable in this analysis is the *population* from 1936 to 1942. Table 1 contains population data for each district within the time-frame of interest.

District-level population data is missing for the years 1938, 1939, 1940 (partially), and 1941 in the government records. For 1938, 1939, and 1940, the population counts for all districts are unavailable due to undocumented reasons. To estimate these values, the population for the International Settlement and French Concession is deduced by adding the refugee population in their camps to their respective 1937 population counts. The population of the Chinese-administered district is then estimated by subtracting the total number of refugees in the two foreign districts from its 1937 population.

The missing population data for 1941 is estimated by calibrating the pre-established 1942 population dataset (zhengfu (n.d.)) against the primary data. A multiplier is derived to adjust the 1941 data to align as closely as possible with the available records, providing an approximate but consistent estimation.

Table 1: Explanatory models of flight time based on wing width and wing length

Year	Chinese-administered Population	International Settlement Population	French Concession Population
1936	2155717	1180969	477629
1937	2155717	1218630	477629
1938	2074693	1272552	504731
1939	2098331	1257703	495942
1940	1479726	1233394	477629
1941	957339	446692	835523
1942	1049403	1585673	854380

2.4.2 Predictor variables

Time Variable: Year

The observations used for modeling and analysis in this paper range from 1936 to 1942. No observations prior to 1936 are included, as population growth patterns before this period differ significantly from those influenced by the events of interest, namely the Japanese occupation of Shanghai's districts. Similarly, no data beyond 1942 is utilized due to data constraints, as annual district-level population records are unavailable after this year. The foreign districts were dissolved in 1949 following the rise to power of a new political regime and the establishment of the People's Republic of China.

2.4.2.1 Binary Indicators

Column 3 and 4 of Table 2 contains district identity indicators used in the analysis model. Variable *district_is* and *district_fc* are two binary variable used to indicate the district identity of an observation. There are 3 possible district identity in Shanghai during 1930s-40s: Chinese-administrated District, International Settlement, and French Concession.

The Chinese-administrated District is governed by local Chinese authority. Land type in this area is primary rural, most of the local population are farmers. Percentage of foreign (out-of-province) immigrants or refugees is low. (Zou (1980)) In this analysis paper, Chinese-administrated District takes value of *district_is = 0* and *district_fc = 0* (it is neither International Settlement nor French Concession).

The International Settlement district is a generalized term for section of lands that allowed the occupation of foreign (non-Chinese) residence. Foreigners in Shanghai were not allow to reside in areas outside of the bounds of this district. Governance power of foreign population within the district – including the policing, taxing, etc, are transferred to foreign authorities. This district takes up most of the central-city in Shanghai during the 1930s-40s, with foreign investment attracting many business opportunities and subsequently job opportunities. Due to the complicated political authorities within the district, the part of Shanghai remains initially untouched by the Japanese forces as the entered Shanghai's Chinese-administrated District. International Settlement district takes the value of *district_is = 1* and *district_fc = 0* in this paper.

French Concession district is a land district given solely to the French. French residency are given the authorization to self-govern within this area. Due to political reasons on the Western battlefield, this district remains untouched by the Japanese forces all the way until they withdrew from Shanghai. The French Concession district takes value of *district_is = 0* and *district_fc = 1* for all analysis in the paper.

Column 5 to 7 in Table 2 are historic event indicators. Variables *cd_occupied*, *french_surrender*, and *is_occupied* are binary variables that mark the whether or not a major historic event that may have impacted Shanghai's population during the observation period (1936-1942) has occurred yet. 1. The Japanese armed forces declared occupation of the Chinese-administrated district on 1937-11-12. With expectation that Shanghai's population will internally shift out of the Chinese-administrated district, the binary variable *cd_occupied* is used to measure the impact of this event on the population change in Shanghai. 2. As WWII impacted the Eastern globe, on the primary battlefield (Western Battlefield), France surrendered to Germany on 22 June 1940. Since Japan sided with Germany in WWII, France's whiteflagging to Germany led to Shanghai's Japanese forces promising not to invade the French Concession on the condition that the district stop taking refugees and support Japanese governance. The paper uses the binary variable *french_surrender* to evaluate the impact this event has on the overall population in Shanghai. 3. In January of 1942, the Japanese forces occupied the International Settlement. Local authority handed over its

governing power to the invading forces. This event's impact on Shanghai's population is measuring using the binary variable *is_occupied*.

Table 2: Explanatory models of flight time based on wing width and wing length

Year	Population	district_is	district_fc	cd_occupied	french_surrender	is_occupied
1936	477629	0	1	0	0	0
1937	2155717	0	0	1	0	0
1938	1272552	1	0	1	0	0
1940	477629	0	1	1	1	0
1941	957339	0	0	1	1	0
1942	1585673	1	0	1	1	1

3 Model

The goal of our modelling strategy is to evaluate the impact of 2 historical events in Shanghai during WWII. Here we briefly describe the Difference-in-Difference analysis model used to investigate the impact of the Japanese forces taking over the Chinese district and International Settlement in November 1937 and December 1941 on population shift between different Districts in Shanghai.

Background details and diagnostics are included in Appendix Section A.3.

3.1 Model set-up

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_{i,t}, \sigma)$$

$$\mu_i = \alpha + \beta_k$$
$$\alpha \sim \text{Normal}(0, 2.5)$$

$$\beta_k \sim \text{Normal}(0, 2.5) \quad \text{for } k = 1, ..., 12$$

$$\sigma \sim \text{Exponential}(1)$$

Where:

- y_i: Observed population in district i at year t.

- α: The intercept for when all variables take value of 0.
- β_1: coefficient for temporal observation *year*. Controls for the effect of time (on an annual basis) on Shanghai's population.
- β_2: coefficient for the dummy variable *district_is*. Take on value of 1 if the district of the observation is International Settlement districts (1 if *district type* is IS, 0 otherwise). The coefficient measures the effect of district identity on population of Shanghai within a given year.
- β_3: *district_fc*: Dummy variable for French Concession districts (1 if i is FC, 0 otherwise). The coefficient measures the effect of district identity on population of Shanghai within a given year.
- β_4: the coefficient for *cd_occupied* indicator. This indicator takes a value of 1 after the Chinese-administered district was occupied by Japanese forces in year 1937 (1 if $t \geq 1937$, 0 otherwise). The coefficient measures the effect of this indicator on population of Shanghai within a given year.
- β_5: coefficient for indicator *french_surrender* which takes on a value of *1* after the French Surrendering to Germany in 1940, leading to the French Concession rejecting refugees in Shanghai (1 if $t \geq 1940$, 0 otherwise). The coefficient measures the effect of this indicator on population of Shanghai within a given year.
- β_6: the coefficient for the indicator *is_occupied*. This indicator marks the Japanese occupation of the International Settlement in 1942 (1 if $t \geq 1942$, 0 otherwise). The coefficient measures the effect of this indicator on population of Shanghai within a given year.
- β_i *where* $i = 7, ..., 12$: coefficients for interaction terms below. Each term ajusts the impact of a historical events on a district-specific basis.

Interaction Terms:

- *district_is* $\times$ *cd_occupied*: Captures the effect of the 1937 occupation on the International Settlement in comparison to the Chinese-administered district (base district).
- *district_fc* $\times$ *cd_occupied*: Captures the effect of the 1937 occupation on the French Concession in comparison to the base district.
- *district_is* $\times$ *french_surrender*: Effect of the 1940 refugee rejection on the International Settlement.
- *district_fc* $\times$ *french_surrender*: Effect of the 1940 refugee rejection on the French Concession.
- *district_is* $\times$ *is_occupied*: Effect of the 1942 occupation on the International Settlement.
- *district_fc* $\times$ *is_occupied*: Effect of the 1942 occupation on the French Concession.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

The Difference-in-Difference model is chosen to evaluate the impact of historical events by comparing the observed outcomes, population changes, to an imaginary counter-factual scenario where these events did not occur. The year variable captures the temporal trends, allowing the model to account for changes in population over time. The district variable identifies the three distinct regions in Shanghai (Chinese-administered district, International Settlement, and French Concession), each subject to unique governance and historical impacts. These inclusions enable the model to control for factors beyond the primary variables of interest (historical events).

The interaction terms in the model are essential because they capture how the impact of historical events (e.g., the 1937 occupation of Chinese-administered districts, the 1940 refugee rejection in the French Concession, and the 1942 occupation of the International Settlement) varies across different district types. By including interactions between event indicators and district types, the model accounts for district-specific responses to each event, controlling for the fact that the effect of an event on population changes may not be uniform across all districts. For instance, the 1937 occupation might lead to population declines in Chinese-administered districts but population increases in the International Settlement due to refugee influx. These terms ensure that the model can differentiate and estimate the unique effects of each event within each district type, thereby improving precision and interpretability.

A negative relationship is expected between year and population because the period under study (1936–1942) was marked by military conflict and foreign occupation, creating an unstable living environment. Specifically: - Chinese-administered District: A population decline is expected due to the Japanese occupation in 1937. - French Concession: A population decrease is anticipated following France's surrender in 1940, which led to refugee rejection. - International Settlement: A population decline is expected after its occupation by Japanese forces in 1942.

4 Results

This section presents the population trends in Shanghai's districts between 1936 and 1942, the estimated effects of district type and key historical events on population, and the district-specific impacts of these events as captured by interaction terms. The analysis highlights distinct population shifts driven by both district characteristics and the varying effects of war-related events.

4.1 Population Trend

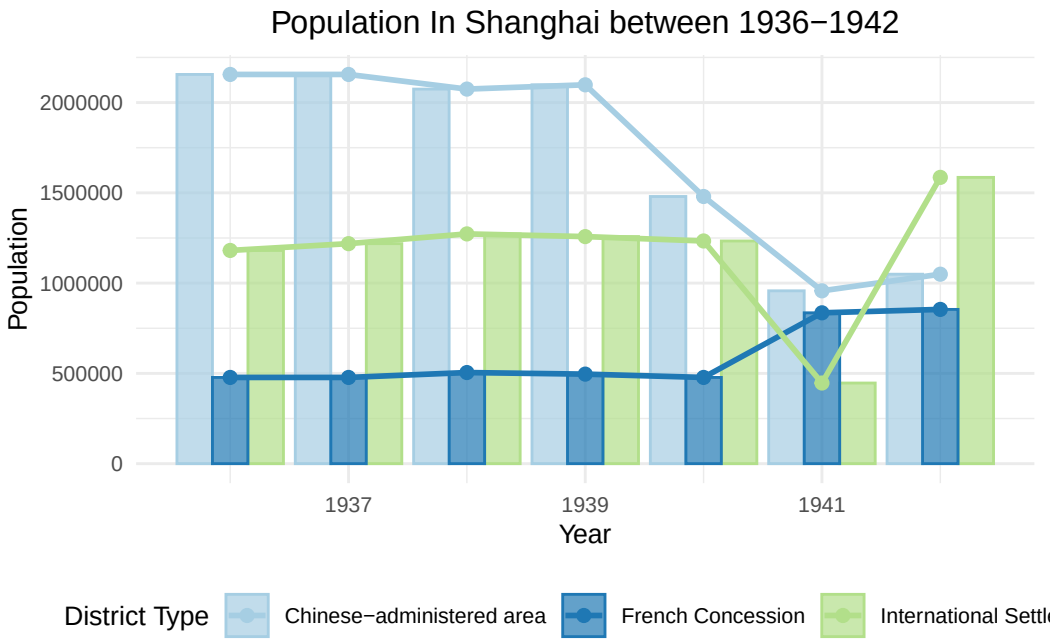


Figure 1: Population in Shanghai between 1936 - 1942

Figure 1 presents distinct population trends from 1936 to 1942 across the three districts in Shanghai: the Chinese-administered area, the French Concession, and the International Settlement. In the Chinese-administered area, the population remained stable until 1937, after which a decline began in 1940, continuing through 1942. The International Settlement experienced consistent population growth throughout the period, with the largest increases occurring after 1938. The French Concession maintained population stability until 1940, followed by a slight increase in 1941, which plateaued in 1942.

4.2 Model results

Model results are summarized in Table 5 in Appendix (Section A).

4.2.1 District Impact on Population

Table 3: Explanatory models of flight time based on wing width and wing length

Parameters	Estimated Effect	Standard Error
(Intercept)	117928165.58	189545438.85
year	-59841.02	98298.59
district_is	-943281.63	381299.41
district_fc	-1642126.91	350808.44

Table 3 estimates highlight the influence of district type on population. The baseline population, Chinese-administrated district population, is represented by the intercept, is estimated at 11,792,816.58, with a standard error of 18,954,543.85. This estimate is not a realistic estimate since the year variable is not included. This intercept is population at year = 0, which does not actually exist. Temporal impact on population is estimated using the *year* variable, observing a decline of 59,841.02 across all districts, with a standard error of 98,298.59. The International Settlement exhibits a population difference of -943,281.63 (standard error: 381,299.41) compared to the Chinese-administered area, while the French Concession shows a difference of -1,642,126.91 (standard error: 350,808.44).

4.2.2 Historical Events

Table 4: Explanatory models of flight time based on wing width and wing length

Parameters	Estimated Effect	Standard Error
cd_occupied	89350.01	350665.6
french_surrender	-735223.28	340852.1
is_occupied	-88982.32	361506.9

The effects of three historical events on population are also quantified in Table 4. The Japanese occupation of the Chinese-administered area in 1937 (cd_occupied) is associated with an increase of 89,350.01 (standard error: 350,665.57). France's surrender in 1940 (french_surrender) results in a population decrease of 735,223.28 (standard error: 340,852.1). The Japanese occupation of the International Settlement in 1942 (is_occupied) corresponds to a population decrease of 88,982.32 (standard error: 361,506.9).

4.2.3 Historic Events' Impact on Each District

The interaction terms, which capture district-specific responses to historical events, provide additional insight. Figure 2 contains the relevant section in the model summary.

The Japanese occupation of the Chinese-administered area increases the population in the International Settlement by 89,430.69 (standard error: 414,726.8) and in the French Concession by 38,396.87 (standard error: 410,271.2). France's surrender increases the population in the International Settlement by 470,617.65 (standard error: 352,215.9) and in the French Concession by 1,044,470.54 (standard error: 352,469.9). The Japanese occupation of the International Settlement itself is associated with an increase of 934,059.1 (standard error: 476,934.2) in the International Settlement's population. Interaction terms differ from main effects by isolating the impact of historical events within each district type, providing a detailed analysis of population dynamics during this period.

District-based population trend after the occurrence of events are shown in Figure 2, Figure 3, and Figure 1.

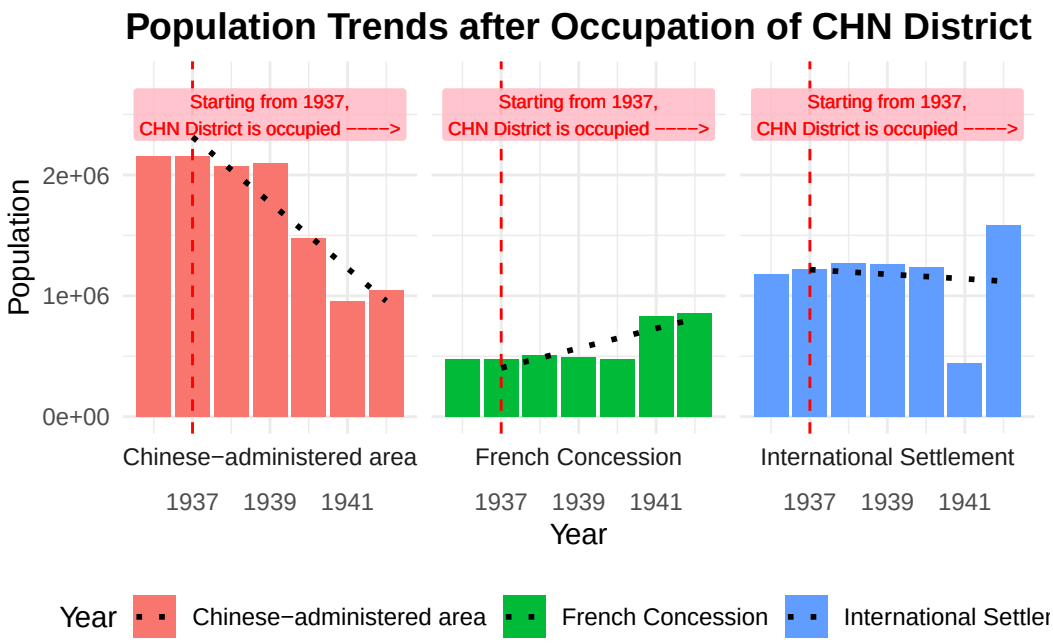


Figure 2: Trend of Population after Chinese-administrated District became occupied in 1937

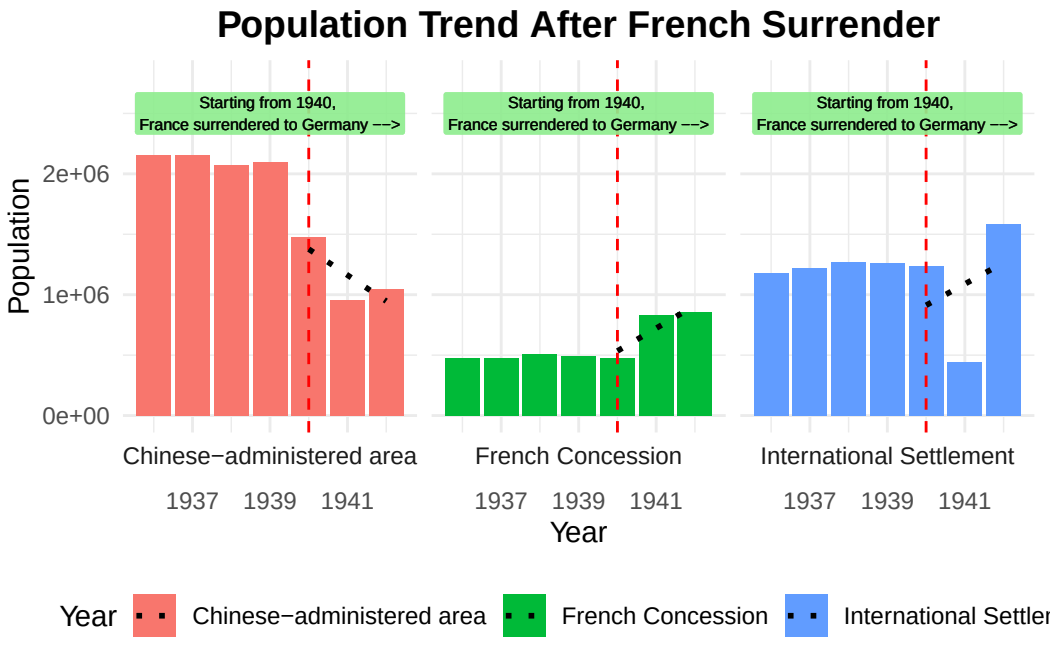


Figure 3: Visualization of Population trend after French surrender in 1940

5 Discussion

The results highlight the varying population shifts in Shanghai's three districts during key historical events in WWII. The Japanese occupation of the Chinese-administered district led to an overall increase in Shanghai's population across all districts. In contrast, France's surrender during WWII negatively impacted population numbers in all districts except the French Concession. Similarly, the Japanese occupation of the International Settlement caused population declines in all districts except the International Settlement itself.

5.1 Occupation in Chinese District

Contrary to expectations, the Japanese occupation of the Chinese-administered district did not result in population declines. Instead, this district experienced an increase in population, likely driven by refugees fleeing rural provinces. The devastation of agricultural economies due to ongoing warfare forced farmers to seek alternative livelihoods in urban centers. Shanghai, being an economic hub with abundant industrial job opportunities, became a natural destination for displaced populations (Zou (1980), page 14).

Industrialization during the 1930s and 1940s was concentrated in cities like Shanghai, while most provinces remained predominantly rural. Disrupted farming activities caused widespread

unemployment, prompting many to migrate to Shanghai, where income was not tied to land productivity. The high population density in Shanghai further fueled its economic growth, offering a range of job opportunities, especially in manufacturing industries located in the foreign settlements.

However, access to jobs was not uniform. Historical records suggest that certain industries were dominated by migrants from specific provinces due to the strong influence of regional associations. For example, hotel businesses were often monopolized by individuals from Hunan, while the footbath service industry was controlled by migrants from Northern Jiangsu. This social network-based hiring created barriers to entry for outsiders, shaping the employment landscape in Shanghai.

5.2 French Surrendering to Germany on Shanghai Population

The French Concession displayed unique population dynamics during WWII. Following France's surrender to Germany on June 22, 1940, the concession ceased accepting refugees. This sudden shift in refugee policy aligns with geopolitical developments, as France temporarily allied with Germany during this period. The Japanese armed forces allowed the French Concession to maintain governance autonomy, likely due to this alliance.

Despite retaining governance, the French Concession operated under the watchful eye of Japanese forces, who occupied the rest of Shanghai by 1942. Compliance with Japanese regulations, including bans on anti-Japanese propaganda, was mandatory. The political alignment with Germany thus shielded the French Concession from direct occupation but altered its role as a haven for refugees.

5.3 Occupation in International Settlement

The Japanese occupation of the International Settlement resulted in a population pattern contrary to initial expectations. While occupation typically drives residents away, the settlement remained one of the safest and most expensive areas in Shanghai. With the French Concession no longer accepting refugees, the International Settlement became the primary refuge for displaced individuals, stabilizing its population despite foreign control. Its status as an elite residential district further insulated it from severe population declines.

5.4 Instability after Japanese Occupation ends in WWII: Chinese Civil War

The end of Japanese occupation in 1945 did not bring stability to Shanghai. Instead, the city became embroiled in the Chinese Civil War, as the two dominant political parties—Kuomintang and the Chinese Communist Party—vied for control. Unlike Western political parties, these factions operated as militarized entities with distinct ideologies and armed forces.

The civil war, which engulfed much of China, ended in 1949 with the Kuomintang retreating to Taiwan, leaving behind a landscape marked by prolonged instability.

5.5 Weaknesses and next steps

The study acknowledges several limitations. First, data recorded during wartime is incomplete, with many population figures missing. Additionally, population counts were based on households, with the number of individuals per household estimated from historical averages. This approach may not accurately capture the actual population. Furthermore, under-reporting or partial reporting for economic or convenience reasons may have compromised data accuracy.

Future research should focus on improving data reconstruction methods, incorporating archival records, and leveraging advanced modeling techniques to address these limitations. Expanding the analysis to include other affected regions in China could also provide a broader understanding of the population dynamics during WWII.

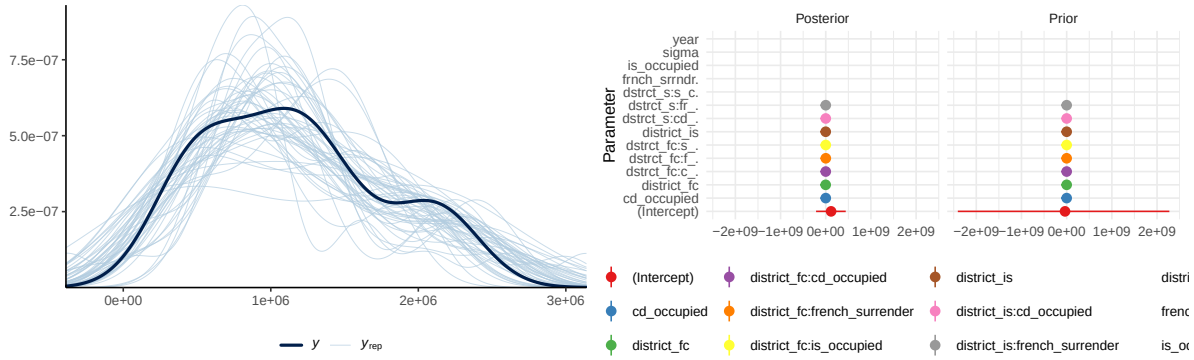
Appendix

A Model details

A.1 Posterior predictive check

In Figure 4a we implement a posterior predictive check. The posterior predictive check evaluates how well the model-generated predictions (y_{rep}) match the observed data (y). The overlay of the predictive distribution with the observed data indicates that the model captures the general trends and variability in the data. The similarity between the observed data and the model's predictions suggests that the model provides a good fit for the observed population data.

In Figure 4b we compare the posterior with the prior. The comparison between the posterior and prior distributions shows how the data influences the model parameters. The posterior distributions are narrower and more distinct compared to the priors, indicating that the data strongly informed the parameter estimates. Parameters like *cd_occupied* and *district_fc* show clear shifts in their posterior means, signifying their significant role in explaining population changes. Conversely, parameters with less distinct posteriors might indicate weaker evidence from the data.



(a) Posterior prediction check

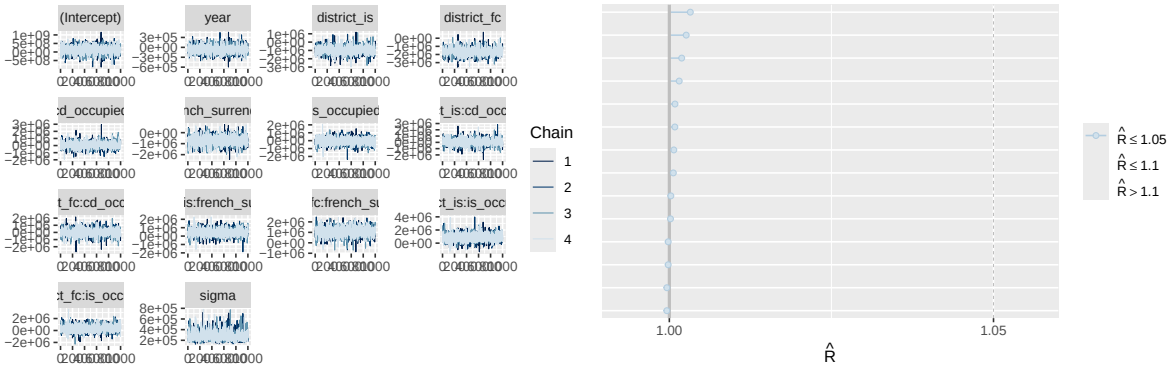
(b) Comparing the posterior with the prior

Figure 4: Examining how the model fits, and is affected by, the data

A.2 Diagnostics

Figure 5a is a trace plot. The trace plot displays the MCMC sampling for each parameter across the chains. The chains appear to mix well, with no clear trends or patterns, indicating that the algorithm explores the parameter space efficiently. This suggests good convergence, as the chains reach their stationary distribution without being stuck in local modes.

Figure 5b is a Rhat plot. The Rhat plot evaluates the convergence of the chains by comparing the within-chain and between-chain variance. All Rhat values are close to 1.0, indicating that the chains have converged. This suggests that the parameter estimates are reliable, and the sampling process provides consistent results across chains.



(a) Trace plot

(b) Rhat plot

Figure 5: Checking the convergence of the MCMC algorithm

A.3 Model Summary

Table 5: Complete Model Summary for DiD model used in analysis

	Population model
(Intercept)	117 928 165.58
	(189 545 438.85)
year	−59 841.02
	(98 298.59)
district_is	−943 281.63
	(381 299.41)
district_fc	−1 642 126.91
	(350 808.44)
cd_occupied	89 350.01
	(350 665.57)
french_surrender	−735 223.28
	(340 852.10)
is_occupied	−88 982.32
	(361 506.88)
district_is × cd_occupied	89 430.69
	(414 726.84)
district_fc × cd_occupied	38 396.87
	(410 271.18)
district_is × french_surrender	470 617.65
	(352 215.89)
district_fc × french_surrender	1 044 470.54
	(352 469.86)
district_is × is_occupied	934 059.10
	(476 934.16)
district_fc × is_occupied	377 436.02
	(466 616.20)
Num.Obs.	21
R2	0.837
Log.Lik.	−289.106
RMSE	162 825.22

B Idealized Survey Methodology

This appendix presents an idealized design for a population sampling survey, assuming that a data-recorder has unlimited time and authority to collect data within the historical time-frame of 1936–1942. The proposed methodology outlines a systematic approach to gathering detailed population data from Shanghai's districts during pivotal wartime events, with considerations for the practical constraints and technology available during that period. While acknowledging the inherent challenges of conducting such a survey, this design aims to provide a comprehensive framework for understanding population dynamics within the Chinese-administered areas, International Settlement, and French Concession.

B.1 Objective of the Survey

The purpose of this survey is to evaluate the relationship between key historical events and population shifts within Shanghai's districts between 1936 and 1942. Specifically, the study aims to document the population changes in three major districts—Chinese-administered areas, International Settlement, and French Concession—during pivotal events such as the 1937 Japanese occupation of Chinese districts, the 1940 rejection of refugees in the French Concession, and the 1942 occupation of the International Settlement.

B.2 Target Population

The survey targets residents and local representatives from the three major districts of Shanghai. These include long-term residents who were present before and during the historical events under study, as well as recent migrants and refugees arriving in these districts as a result of wartime displacement. The goal is to achieve representative coverage across the three districts to capture the diversity in population trends and movement patterns.

B.3 Sampling Design

The sampling design employs stratified random sampling to ensure representation across the three major districts. Each district is treated as a separate stratum: Chinese-administered areas, International Settlement, and French Concession. A fixed number of households per district are selected based on district size, population density, and anticipated population shifts. Special attention is given to refugee populations within the International Settlement and French Concession to capture the impact of displacement. Enumerators use household lists and local community organizations, such as neighborhood associations and refugee aid committees, to randomly select respondents within strata.

B.4 Recruitment Strategy

The recruitment strategy involves direct engagement with local leaders, community groups, and refugee aid organizations. Enumerators collaborate with influential figures in each district to gain trust and permission to conduct the survey. Recruitment booths are set up in public spaces such as markets and parks to engage with potential respondents. Additionally, refugee aid centers are utilized to access displaced populations. Recruitment materials are printed in both Chinese and English to accommodate the linguistic diversity of Shanghai's residents.

B.5 Survey Instrument

The survey instrument consists of a structured questionnaire covering demographic information, migration history, wartime impact, and population and housing details. The demographic section collects age, gender, occupation, family size, and household composition. Migration history includes the origin, year of migration to the current district, and reasons for relocation. Wartime impact focuses on displacement, property loss, and wartime experiences. Lastly, the population and housing section gathers data on the number of residents, housing type, and living conditions. Open-ended questions are also included to capture respondents' personal narratives about wartime displacement and resettlement.

B.6 Data Collection Procedures

Data collection occurs between 1936 and 1942, with enumerators visiting households, public spaces, and refugee shelters. Enumerators are trained to conduct interviews in Chinese, Shanghainese, or English based on respondent preferences. Responses are recorded on pre-printed forms, which are later transcribed and stored securely. To ensure accuracy, enumerators cross-check entries at the end of each survey day. Each enumerator is assigned a specific district to maintain consistency in data collection practices.

B.7 Limitations and Challenges

Some areas, particularly those under military occupation, may not be accessible for data collection, leading to incomplete coverage. Additionally, respondents may underreport family size or migration history due to privacy concerns or fear of government reprisal. Refugee populations are often transient and may not remain in one location long enough for follow-up interviews. Despite these limitations, the survey employs robust sampling and recruitment strategies to mitigate potential biases.

B.8 Ethical Considerations

Informed consent is obtained by informing respondents about the purpose of the survey and how their data will be used before they agree to participate. Personal identifying information is anonymized to ensure respondent privacy and security. Participation is entirely voluntary, and respondents can opt out at any time without consequences. The survey complies with ethical standards to respect the rights and dignity of all participants.

References

Arel-Bundock, Vincent. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models and Data*. https://vincentarelbundock.github.io/modelsummary/.

Bolker, Ben, Davis Vaughan, John Mount, Kristoffer Magnusson, et al. 2023. *Broom.mixed: Tidying Methods for Mixed Models*. https://broom.tidymodels.org/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Henriot, Christian, and Contributors. 2024. "Virtual Shanghai: A Research and Resource Platform on the History of Shanghai (1843-1949)." https://www.virtualshanghai.net.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

User, Bilibili. n.d. " ." https://www.bilibili.com/video/BV1D64y1J7ku?vd_source=960969a3da000a1a0ec76d3c506d1101.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, et al. 2023. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org/.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

zhengfu, Shanghai tebie shi. n.d. "Population of the Urban and Rural Districts of the Shanghai Municipality by Sex and Age Group (1941-1943)." Virtual Shanghai; Dataset accessed via Virtual Shanghai. https://www.virtualshanghai.net/Data/Tables?ID=46.

Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://cran.r-project.org/package=kableExtra.

Zou, Yi. 1980. *Jiu Shanghai Ren Kou Bian Qian Di Yan Jiu*. Di 1 ban. Shanghai: Shanghai ren min chu ban she.