

Yu, Lamont Bo; Tran, Trang My; Lee, Wang-Sheng

Working Paper

Revitalising the Silk Road: Evidence from Railway Infrastructure Investments in Northwest China

IZA Discussion Papers, No. 15454

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Yu, Lamont Bo; Tran, Trang My; Lee, Wang-Sheng (2022) : Revitalising the Silk Road: Evidence from Railway Infrastructure Investments in Northwest China, IZA Discussion Papers, No. 15454, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/263670>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 15454

**Revitalising the Silk Road:
Evidence from Railway Infrastructure
Investments in Northwest China**

Lamont Bo Yu
Trang My Tran
Wang-Sheng Lee

JULY 2022

DISCUSSION PAPER SERIES

IZA DP No. 15454

Revitalising the Silk Road: Evidence from Railway Infrastructure Investments in Northwest China

Lamont Bo Yu

City University of Macau

Trang My Tran

Monash University

Wang-Sheng Lee

Monash University and IZA

JULY 2022

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Revitalising the Silk Road: Evidence from Railway Infrastructure Investments in Northwest China*

China's Belt and Road Initiative was introduced in 2013 to revitalise the Silk Road and promote economic development and integration. This paper investigates the economic effects of the opening of the only high-speed rail (HSR) line in northwest China which connects China's northwestern provinces along this Silk Road land route. We use a recently developed machine-learning extended nightlight data series from 2000 to 2019 and employ the ridge augmented synthetic control method (Ben-Michael et al., 2021) to assess the effects of the HSR line connection on economic activity along this Silk Road land route. We further propose an algorithm that helps automate the donor pool selection process while ensuring optimal pre-treatment fitness. Our results show that there are winners and losers from the opening of the Lanzhou–Urumqi HSR line. While there is some indication of the role that HSR can help play in making progress towards breaking through the Hu Huanyong Line, a geographical demarcation in China that is of vast economic significance, not all counties benefited from the opening of the HSR line.

JEL Classification: O22, R11, R58

Keywords: high-speed railway, augmented synthetic control, Hu Huanyong Line

Corresponding author:

Wang-Sheng Lee
Centre for Development Economics and Sustainability
Monash Business School
Monash University
Caulfield East, VIC 3145
Australia
E-mail: wang.lee@monash.edu

* We thank Shihe Fu, Zibing Huang, Ming Lu, Russel Smyth, Yanrui Wu, Yucheng Wang and participants at the 2022 Chinese Economists Society annual conference, the 3rd China Regional Economics Forum, 2021 China Youth Economists Society annual conference, and the RIEM at Southwestern University of Finance and Economics, for helpful comments. This paper has previously circulated under the title 'Revitalising the Silk Road: The Economic Effects of High-Speed Rail in Northwest China'. All authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1 Introduction

In 2014, Chinese Premier Li Keqiang put forward some questions about the ‘Hu Huanyong Line’ (Hu Line, hereafter), an imaginary line that divides China into two parts.¹ To the east of this line is just over one-third of the nation’s land, housing almost 94 percent of the country’s population—more than 1.2 billion people. To the west, around 6 percent of citizens share the vast and varied terrain. Premier Li Keqiang stated that the nation needed to ‘break through’ the Hu Line to achieve modernisation in its central and western regions. The underdevelopment of provinces in Western China compared to provinces to the east of the Hu Line has been a long-standing issue recognised by the Chinese government.

Place-based policies aim to narrow development gaps within regions of a country and involve directing public resources to disadvantaged geographic areas, infrastructure investments, block grants, regulations, subsidies and tax exemptions (Glaeser and Gottlieb, 2008). Such policies are widespread in the United States and throughout the world. In the United States, a prominent example is the Tennessee Valley Authority, an ambitious federal programme conceived as part of the New Deal that consisted of massive public infrastructure investments in roads and electricity. It was designed to modernise a region that, at the time, was one of most underdeveloped in the country (Kline and Moretti, 2014a). Place-based policies are also popular in Asia and Europe. In China, an example is China’s Great Western Development (GWD) Programme, which was an initiative of the Chinese central government in 2000 to target 12 provinces in Western China to accelerate their industrialisation (Jia et al., 2020). In India, the Indian Government launched the Prime Minister’s Village Road Programme in 2000, premised on the idea that poor road connectivity is the biggest hurdle in faster rural development (Asher and Novosad,

¹The line is named after the Chinese demographer Hu Huanyong, who, in 1935, used a hand-drawn map of the line to illustrate a demarcation that has vast demographic, environmental, and political significance (Hu, 1935). Issues regarding the Hu Line have been dubbed ‘The Premier’s Question’ following Premier Li’s remarks. Chen et al. (2016) show using contemporary data that the Hu Line is still valid today in describing China’s demographics.

2020). In Europe, Italy has long provided regional transfers that single out distressed regions, especially in the south, for special infrastructure investments. Sweden, France, and Germany have similar programmes (Kline and Moretti, 2014b).

There have been substantial investments in Western China in recent years. In 2013, China launched the Belt and Road Initiative (BRI), which seeks to improve connectivity between northwest China, Central Asia, the Middle East, Africa and Europe (see Figure A1 for the BRI land trade route).² Central Asia is particularly affected by a series of major BRI infrastructure investments which are strategic transit routes for China, and which will simultaneously improve transport links within this remote region and between it and the rest of the world (Bird et al., 2020). In part, the BRI is about improving the physical infrastructure through land corridors that roughly equate to the old Silk Road.

Since the opening of the first line in 2008, HSR networks in China have expanded dramatically, meeting the need for long-distance transportation in the country. The route of the BRI within China extends from Xi'an to Lanzhou to Urumqi, and the Lanzhou–Urumqi high speed railway (HSR) is the only HSR route in northwest China that supports the land corridors of the BRI. The Lanzhou–Urumqi HSR line is an important path connecting the central and western provinces. It can play a key role in promoting economic integration between the eastern and western parts of China and serve as a vital gateway to Central Asia. In this paper, we estimate the causal effects of the opening of the Lanzhou–Urumqi HSR line in northwest China in 2014 on regional economic development along the Silk Road.

Previous research has examined the effect of HSR on economic development in China. However, these studies have generally focused on evaluating the effects of China's HSR network as a whole (e.g., Huang et al., 2019; Long et al., 2018; Qin, 2017; Wang et al., 2020)

²The BRI connects Central and Northern China principally through three corridors: the new Eurasian Land Bridge (a Northern route, connecting Xi'an, Lanzhou, and Urumqi in China with Astana and Moscow); the China Central-Asia West-Asia corridor (Xi'an, Lanzhou and Urumqi in China with Almaty, Tashkent, dividing to Aktau and to Ashgabat, Tehran, Ankara); and the China Pakistan corridor (Xi'an, Lanzhou and Urumqi to Urumqi, Almaty, Islamabad, and Gwadar) (Bird et al., 2020).

or a single line to the eastern side of the Hu Line (Liang et al., 2020).³ Given that the development of Western China is one possible way of breaking through the Hu Line, it is important to examine the impacts that the Lanzhou-Urumqi HSR line has had on local economies to the west of the Hu Line. For the counties along the Silk Road, to what extent does being connected to the HSR line affect economic growth? The advantage of focusing on a single HSR line rather than all HSR lines in the network is that it is more clear-cut to develop an appropriate causal model that attempts to accurately capture specific local effects of the quasi-experiment.⁴

There are two alternative ways in which HSR connection may affect regional economic development. From a core node perspective, HSR may create a decentralisation effect that significantly improves regional market access and promote service clustering by expanding the urban workforce in core cities. The labour force can be drawn from connected core counties within a commuting range, in turn helping to boost economic development over space as a result of additional income flowing into rural counties (Wang et al., 2020; Zheng and Kahn, 2013; Zheng et al., 2019). In contrast, from the ‘home market’ centralisation perspective, market heterogeneity may create an agglomeration effect in which resources are diverted from disadvantaged cities to the core city (Baum-Snow et al., 2017; Glaeser and Gottlieb, 2008). The introduction of a HSR line could enable the central megalopolis nodes within a region to further attract high-quality resources and production inputs from less-developed peripheral areas, thereby strengthening disparities between the core and peripheral areas and increasing regional economic differences (Faber, 2014; Qin, 2017). In particular, if the reduced trade costs were to facilitate temporal population

³Liang et al. (2020) examined the effect of a single HSR line in Southern China, the Guangdong–Guangxi–Guizhou (GGG) HSR, on regional economic growth using a combined propensity score and difference-in-differences method. As an extensive railway network linked the GGG cities multiple times, and because the demographic and geographic characteristics of cities in the GGG corridor are very different (see Figures A5a and A5b), their results cannot be directly applied to our context.

⁴Our narrower focus is somewhat similar to the approach in Ke et al. (2017), who examined how China’s recently completed ‘four vertical’ HSR projects influenced local economic growth in targeted city nodes. Their treatment group comprised of 21 HSR cities along the four HSR lines of interest, and estimated treatment effects were provided for each city separately.

flows of low-skilled labour which reduces their wage levels in the home market, the decreased wage costs could further result in higher outputs in labour-intensive industries (Kong et al., 2021).

Our paper contributes to the place-based policy literature by estimating the short- and medium-term effects of HSR on economic development in northwest China using a synthetic control method (SCM). More specifically, we utilise the recently developed ridge augmented synthetic control method (ASCM) by Ben-Michael et al. (2021) to analyse the change in machine-learning smoothed nightlight as a result of HSR opening for the counties of northwest China along the Silk Road land route. ASCM can help reduce bias when pre-treatment fit using the standard SCM is unattainable. It achieves this by allowing negative weights for donor counties, which can be used for extrapolation in the pre-treatment period. To avoid over-fitting, a ridge-type adjustment punishes excessive extrapolation.

In addition, we provide new evidence on the asymmetric impacts of having a HSR line connected to rural areas. Past studies suggest that there are positive effects of cities being connected to HSR in terms of promoting knowledge spillovers, wage premiums and increasing market potential (Dong et al., 2020; Lin, 2017; Zheng et al., 2019).

A noteworthy feature of our empirical approach is that our counterfactuals are constructed from a large donor pool of counties (accounting for 1,504 potential donor Chinese counties). This possibly allows the data-driven SCM to construct the best possible synthetic control. However, given that the risk of over fitting may increase with the size of the donor pool, we further validate our estimates by testing variations in the size of the donor pool and extending an exercise conducted by Abadie et al. (2015). We propose an algorithm that computes pre-treatment pairwise standard deviations across treated and control units, helping us systematically restrict the size of the donor pool, while ensuring an optimal fit prior to the treatment date.

As a basic requirement, SCM requires a sufficiently long pre-treatment period of out-

comes to create a counterfactual for the treated county, based on a composite of the control counties. We use the most recent machine-learning smoothed and integrated time-series nightlight data to measure economic activity along the Lanzhou-Urumqi HSR line (Chen et al., 2021). In order to create this data series, Chen et al. (2021) used a machine-learning approach to smoothly join two sets of satellites observations over the period 2000-2019, the Defense Meteorological Satellite Programme (DMSP) Operational Linescan System and the Suomi National Polar-orbiting Partnership Visible Infrared Imaging Radiometer Suite (VIIRS).⁵

Our identification strategy relies on the timing of the completion of the Lanzhou-Urumqi HSR line. We used the opening date of the HSR line as a source of exogenous variation to examine the economic effects of the opening of the new line. We also take advantage of the fact that the Lanzhou-Urumqi HSR line is entirely located in northwest China, unlike the central and coastal areas of China, which have seen new HSR lines connected repeatedly over our sample period (2000-2019). The natural geographical isolation of the Lanzhou-Urumqi HSR line helps ensure that the treatment effect of the HSR line is unlikely to be confounded by the connection of other nearby HSR lines.⁶

As a preview of our results, we find that the rural Lanzhou-Urumqi HSR line led to limited economic growth for counties between the two ends of the HSR line. We also find little evidence that Lanzhou-Urumqi HSR improved the economic attributes (growth in land transactions, fixed investment growth, and employment) of the Chinese counties along the land corridor of the BRI. However, at the same time, higher levels of growth were seen in counties within Lanzhou. Moreover, we see that following the initial opening of the Lanzhou-Urumqi line in late 2014, an extension of its tracks in mid-2017 al-

⁵DMSP and VIIRS nighttime light data are two widely used nightlight datasets. However, the difference in their spatial resolutions and sensor design means that a cross-sensor calibration of the datasets is required for analysing a long-term urbanisation process. This is discussed in more detail in Section 2.1.

⁶For example, a county that is connected to an east-west HSR line may experience a population boom and an increase in direct investment (i.e., a centralisation effect). These factors, which raise aggregate demand, may in turn increase the likelihood of the prosperous county becoming connected to another HSR line (e.g., north-south), which introduces challenges in terms of the empirical estimation, particularly with respect to treatment timing and intensity.

lowed Lanzhou to be fully connected to central China via rail. The positive effect we find for Lanzhou following its connection to the rest of the HSR network is a sign that economic disparities on both sides of the Hu Line might gradually dissipate. Growth in a major city in Western China can be beneficial as it can facilitate bridging Western China to the more developed eastern provinces.

Overall, we find that HSR construction in northwest China has resulted in both winners and losers in terms of economic growth. Economic growth may not always follow from the extension of HSR tracks. Some counties become vibrant following a few years after HSR connection, while others may be unable to leverage any benefits. There could be various explanations for this effect heterogeneity including differences in industrial investment attractiveness, sectoral agglomerations and population mobility. At the same time, the potential for urban agglomeration in rural counties in Western China, which are located at high altitudes and surrounded by deserts, appears to be rather limited. We therefore suggest prudence when contemplating constructing HSR lines in rural counties, especially when the construction and operational costs are high (the estimated construction cost for the Lanzhou-Urumqi HSR was \$22.56 billion USD).

The remainder of the paper is organised as follows. Section 2 introduces the data and counties' attributes. Section 3 discusses the details of our empirical strategy. Section 4 discusses the results and robustness checks. Finally, Section 5 concludes the paper.

2 Data and County Characteristics

2.1 Nightlight data

In this study, we utilise nightlight data as a proxy for GDP to investigate changes in economic activity and urbanisation in northwest China. Satellite nighttime light data used as a proxy variable of regional economic growth can help alleviate the potential

issue of inaccurate regional GDP measurements in China.⁷ It possibly more objectively reflects the impact of HSR opening on regional economic level along the line.

We used two sets of widely applied nightlight data over the period 2000-2019. The first dataset was sourced from DMSP, which provided yearly nightlight data for the period 2000-2012. The second dataset was sourced from VIIRS, which reported monthly and yearly nighttime light from April 2012 to December 2019. However, a practical integration issue is that DMSP and VIIRS recorded nightlight measurements via different satellites over the years. The DMSP nightlight data series has two disadvantages. First, there is inconsistency in the data recording because the DMSP satellites changed from F14 in 2000 to F18 in 2010. Therefore, some modelling is required to account for differences in the sources of nightlight to create an extended time series over the period 2000-2019. Second, DMSP has less predictive power with respect to GDP in less populated rural areas and the predictive power is especially poor for fine-granular units in rural areas (Gibson et al., 2021). In contrast, the newer but shorter VIIRS data series is more accurate and depicts relatively stronger correlations with economic activities.⁸

To address inconsistencies in the source of nightlight data and the inaccuracy of DMSP, we used Chen et al. (2021)'s machine-learning simulated nightlight data.⁹ To combine the DMSP and VIIRS data and create an extended nightlight time series, they proceeded as follows. Chen et al. (2021) first utilised a machine-learning approach to predict VIIRS-like nightlight data for the period 2000 to 2012. In their model, they used the 2013 enhanced vegetation index adjusted nightlight index (EANTLI) and 2013 VIIRS nightlight as labelled features to train their predictive model. Once the model had been built, they re-

⁷The use of nightlight data in this fashion follows from the pioneering work of Henderson et al. (2012) who applied two decades of multinational panel data to discuss the applicability of satellite nighttime light data to provide reasonably accurate estimates of subnational GDP.

⁸By focusing on larger aggregated county cities in this paper, we acknowledge the limitation suggested by Gibson et al. (2021) that neither DMSP nor VIIRS might be appropriate for a higher resolution analysis in the less populous and developing northwestern provinces.

⁹Some recent work on the economic impacts of HSR in China using nightlight data (e.g.: Long et al., 2018; Wang et al., 2020; Zheng et al., 2019) are based on nightlight data over a more restricted period 2004 to 2013, likely because of this data inconsistency.

tained the 2000-2012 EANTLI data as an input to develop the VIIRS-like 2000-2012 night-light data. The predicted 2000-2012 VIIRS-like nightlight data series was then smoothly appended with the actual 2013-2019 VIIRS data to generate an extended 2000-2019 time series.¹⁰ We obtained [Chen et al. \(2021\)](#)'s extended global nightlight data and used GIS software containing information on China's administrative county boundaries to create yearly nightlight data measurements by county. Average county nightlight per pixel was calculated using an arithmetic mean over each county city.

Figure 1 shows the change in nightlight intensity over East Asia between 2014 and 2019. For China, it can be seen that the most salient changes in nightlight intensity took place in the eastern and central parts. In general, apart from a few counties in Lanzhou, there was much less change in counties alongside the Lanzhou-Urumqi HSR line.

2.2 Socioeconomic and geographical variables

We collected the names of the HSR stations and their opening dates from the official HSR website (<http://www.gaotie.cn>). Entering the HSR station name into the Baidu Map allows the application to decode its longitude and latitude. In order to obtain a county-level data set that included the location of the HSR station together with detailed administrative information of the county, we spatially match the HSR stations with the county-by-prefecture vector layers from Baidu Map.

Table 1 lists the connection dates of the 16 connected counties along the Lanzhou-Urumqi line. For the purpose of comparison, we aggregate the counties along the Lanzhou-Urumqi HSR line into 3 groups: all counties within Lanzhou prefecture ("Within Lanzhou"), all HSR counties connected between Lanzhou ("Between L-U HSR"), and a major county within Urumqi prefecture ("Within Urumqi").

We also utilise the 2013 Annual Survey of Industrial Manufacturing Firms (the most

¹⁰Appendix Table A2 describes the machine-learning data smoothing process using an auto-encoder model in [Chen et al. \(2021\)](#). Data over the period 2000-2019 is available at <https://doi.org/10.7910/DVN/YGIVCD> in *tif* format with a spatial resolution of 15 arcsec (around 500 metres).

recent data available to us) to determine the exact locations of each firm via a search of firm name in the Baidu Map API. Figure 2 plots the firm locations, revealing a sparse distribution of manufacturing firms and the low industrial diversity across the northwestern counties. This corresponds to the demographic and nightlight distribution in China and suggests that counties in northwest China in general are much less developed than the prosperous eastern counties. In order to more closely assess the economic conditions, we also collect all the types of land transactions for the northwestern provinces in China from the official land transaction website (<https://www.landchina.com/>) and spatially match these transactions with its specific county vector layer.

To provide context regarding population density in the counties around the Lanzhou-Urumqi HSR line, we use data from the geo-coded China 2000-2020 population census. These gridded population estimates are obtained from NASA's Administrative Unit Center Points with Adjusted Population Estimates (V. 4.11). Figure A4 displays the changes in population densities from 2010 to 2020. It is evident that the most densely populated cities are located closer to Lanzhou, and it does not appear that there were any large changes in population density by county between 2015 and 2020.

There are additional meteorological and geographical variables that can provide further context regarding the economic environment of counties along the Lanzhou-Urumqi line. In particular, we calculate the number of days per year that had maximum temperatures below 5°C, averaged across 2000-2018. This serves as a proxy for the habitability of counties and also demonstrates the uniqueness of China's northwestern provinces. Other relevant meteorological variables include coldness and altitude (Chen et al., 2018). These data are derived from the daily surface meteorological dataset (V3.0) from the China Meteorological Data Service Centre (<http://data.cma.cn/>).¹¹ In addition, we downloaded geographical variables (such as altitude) from the ASTER Global Digital Elevation Model

¹¹This service provides daily temperature recordings of 839 ground-based monitoring stations. Outliers and missing weather values from station recordings are manually adjusted and corrected by the China National Meteorological Information Centre, and an inverse distance weighting interpolation method is used to generate daily county temperature measurements.

V3 (NASA Earthdata archives) and clipped and compiled the downloaded raster segments to a new raster that reflected the altitude and surface slopes of Chinese counties. Panel (a) of Figure A5 plots the average altitude of the counties. Panel (b) plots a measure of coldness measured using the number of days the outdoor temperature is less than 5°C. Combined, both graphs provide visual evidence that the northwestern provinces are less habitable than areas to the east of the Hu Line, and why the Hu Line serves as a natural demographic demarcation that divides China’s territories into two parts.

To assess the dynamic properties of each connected county and core-peripheral relationship for Lanzhou, we collected yearly county demographic and industrial characteristics from the *China County Statistical Yearbooks* from 2000 to 2019. Table 2 compares the mean values of some pre-treatment (2010-2014) and post-treatment (2015-2019) characteristics for three HSR groups connected by the Lanzhou-Urumqi HSR line (within Lanzhou, within Urumqi, and between Lanzhou and Urumqi), and another group containing all other neighbouring counties around Lanzhou (“Peripheral Lanzhou”, illustrated in the Figure 6). The descriptive evidence in Table 2 suggests that counties within Lanzhou seem to have experienced the largest change in nightlight intensity compared to the other three groups.

Having assembled a comprehensive dataset that comprises demographic, industrial, geographical and meteorological variables, in the next section, we provide causal estimates of the economic effects for the opening of the Lanzhou-Urumqi HSR line on a diverse set of counties along the HSR line.

3 Empirical Strategy

3.1 Augmented Synthetic Control Method

In their seminal article on SCM, [Abadie and Gardeazabal \(2003\)](#) introduced a data-driven method to evaluate causal effects in policy evaluation analysis. This method uses

a weighted average of the selected control units to construct a synthetic control unit that balances the treated unit's pre-treatment outcomes as closely as possible. Selected control units are chosen from a donor pool and assigned non-negative optimal weights, where the weights are determined by a solution to a minimisation problem, and the total weights are constrained to sum to one.

In this study, we are interested in assessing the effects of the opening of the Lanzhou-Urumqi HSR line on counties that are along the line. The ideal solution is to identify counties in China that are not close to a HSR line but which are very similar to the counties of interest across a range of different characteristics. The representation of the average effect in our case is given by:

$$\tau_{it} = \mathbb{E}[Y_{it}(1) - Y_{it}(0)], \quad (1)$$

where $Y_{it}(1)$ and $Y_{it}(0)$ denote nightlight intensity per pixel for the treated and control counties respectively. τ summarises the average treatment effect on treated unit i by calculating the difference between treated and synthetic units that have not been exposed to treatment in the post period. In the classical SCM approach, the synthetic estimator for unit i with a vector of non-negative, convex combination weights, $\gamma = (\gamma_2, \gamma_3, \dots, \gamma_{J+1})'$ is such that

$$\hat{\tau}_{it}(1) = Y_{it}(1) - \sum_{j=2}^{J+1} \gamma_j Y_{jt}. \quad (2)$$

However, in our case, the existence of a perfect synthetic control is unlikely to exist because the outcomes of our treated units (i.e. average nightlight density of HSR connected counties) are likely outside of the convex hull of the values of the predictors for the donor pool (the unconnected counties). In other words, our selected donor control units do not always provide a good match for the treated unit. A likely consequence of this is that the trajectory of the synthetic unit does not reproduce a similar trajectory as that of the

treated county prior to the treatment date. This results in an unreliable synthetic control estimator due to a poor pre-treatment fit (Abadie, 2019). There is no guarantee that the use of SCM will lead to the creation of a suitable synthetic control. In some cases, a good fit will be impossible to achieve, and Abadie et al. (2015) suggest only using SCM if there is a good pre-treatment fit. To maximise our chance of obtaining a good pre-treatment fit in our empirical analysis of northwestern counties in China, we employ a modified version of SCM–augmented SCM by Ben-Michael et al. (2021).¹²

ASCM relaxes the SCM assumption that weights must be non-negative and extrapolates away from the convex hull. The extrapolation allows for an enhancement of the pre-treatment fit. However, this can lead to an over-fitting bias. Therefore, to control the amount of extrapolation, ASCM proposes to use ridge regression to directly penalise over excessive extrapolations through the optimal choice of the ridge penalty hyperparameter.¹³ The ASCM estimator for the first unit in post-treatment period is given by:

$$\hat{Y}_{1T}^{aug}(0) = \sum_{W_i=0} \hat{\gamma}_i^{scm} Y_{iT} + \left(\hat{m}_{1T} - \sum_{W_i=0} \hat{\gamma}_i^{scm} \hat{m}_{iT} \right). \quad (3)$$

The key purpose of ASCM is that it corrects for imbalance bias using the term \hat{m}_{1T} . Equation 3 converts to a standard SCM estimator when \hat{m}_{1T} is a constant term. Through ridge augmentation that adjusts for $\hat{m}(X_i)$, the ASCM estimator is presented as:

$$\hat{Y}_{1T}^{aug}(0) = \sum_{W_i=0} \hat{\gamma}_i^{scm} Y_{iT} + \left(\mathbf{X}_1 - \sum_{W_i=0} \hat{\gamma}_i^{scm} \mathbf{X}_i \right) \cdot \hat{\eta}^{ridge}, \quad (4)$$

where $\hat{\eta}^{ridge}$ is a ridge regression coefficient of post-treatment outcomes on the pre-treatment outcome, X_i . Therefore, $\hat{Y}_{1T}^{aug}(0)$ can be viewed as a penalised synthetic control

¹²For a recent application of ASCM, see Cole et al. (2020) who examined the change in air pollution following Wuhan’s 2020 COV-19 lockdown. They found that ASCM provided a good pre-treatment fit matching when using 30 capital cities as the donor pool.

¹³When achieving excellent pre-treatment fit with original SCM is possible, SCM and ASCM will both converge. Otherwise, a certain amount of extrapolation involving negative weights will be beneficial for improvement in balance.

estimator (Ben-Michael et al., 2021). It allows for a certain amount of extrapolation to improve pre-treatment fit while at the same time including a penalty parameter to help prevent over-fitting.

In this study, several factors determine the control counties selection process. There are 2,560 counties listed in the *China County Statistical Yearbook*. We first exclude all other HSR-connected counties in China that were connected prior to the opening date of the Lanzhou-Urumqi HSR line. We also excluded counties connected to HSR between January 2015 and November 2019 because a HSR intervention for these later-connected counties would confound the estimation of post-treatment effects. Moreover, we exclude Hong Kong, Macau and Taiwan from the list of donor counties arising from their distinct administrative systems. These areas are typically not considered in the analysis of Chinese counties. Finally, we identify and further exclude counties that experienced administrative (border) changes (around 150 counties). Altogether, these exclusions will ensure that the intervention affected only our treated counties and not any counties in our donor pool. The screening results in 1,504 counties left for which data is available for the period 2000 - 2019.¹⁴

Appendix Figure A3 plots the location of our control counties. It shows that the majority of control counties are located neither in the developed coastal areas nor in the connected central parts of China. Moreover, Figure 2 shows that the control counties generally had lower industrial densities. This similarity to counties in northwest China will be helpful for balancing the characteristics of the treated and control counties. In our analysis, the pre-treatment starting year is 2000 which is the earliest year available for the machine-learning smoothed data. Given that our treatment date was the end of 2014, we have a sufficient pre-treatment period to develop our model using ACSM.

¹⁴We are aware of one other study that uses a large number of control units in a SCM approach (Zou, 2018). In this study, 2,429 counties without military bases were used in an SCM approach to evaluate local economic impacts of contractions in US military personnel in counties that hosted military bases. Utilising a large size of the donor pool is not a common practice in SCM research because there are usually few control units available to choose from, such as in a cross-country study.

4 Effects of the Lanzhou-Urumqi HSR

4.1 Baseline results

To quantify the economic effects of the opening of the Lanzhou-Urumqi HSR line opening, we calculate the average amount of nightlight per pixel for each county and apply the ASCM approach separately for each county.¹⁵ This provides us with an estimate of the average treatment effect on each treated county. We first present baseline results that are based on an ASCM approach that incorporates covariates and lagged outcomes for selected years (2004, 2009 and 2014).

Figures 4 and 5 plot the estimated economic effects for each county. The estimated effects are based on the difference between the actual and counterfactual trajectories, conditional on their demographic and socio-economic attributes listed in Table 2. The treatment date is represented by the vertical dash line. Based on the assessment of the pre-treatment fit for the focal counties, these figures suggest that for most counties, the combination of our donor counties closely reproduces the outcome path prior to treatment.¹⁶ From the synthetic control group matching exercise in Figures 4 and 5, a close pre-policy fit for some of the northwestern counties is not attainable for two reasons. First, the remaining donor counties do not closely match the unique characteristics of some of the treated counties in northwest China. In contrast, the eleven focal counties (sub-figures (a) - (k)) that we are able to obtain good pre-treatment fit for are mostly located in the downstream area of the Lanzhou-Urumqi line. They are therefore more likely to share similar attributes with other parts of China, which makes it easier to find matches. Second, even though ASCM relaxes the restriction of having positive weights to allow for extrapolation outside the support area of the donor pool, pre-treatment fit is still not always guaranteed.

¹⁵In China's administrative hierarchy, a province is larger and ranked administratively higher than a prefecture city, while a prefecture is larger than a county city. An autonomous region or urban district under a prefecture is administratively equivalent to a county city; we use these terms interchangeably.

¹⁶Figure A6 and Figure A7 display the corresponding trajectories of nightlight intensity for the focal counties and their counterfactuals.

To make inferences regarding the statistical significance of the results, we calculate standard errors based on the Jackknife+ approach of [Barber et al. \(2021\)](#).¹⁷ For this inference exercise, we restrict ourselves to using a uniform donor pool and a single model specification (the specification used in columns (1) - (4) of [Table 3](#)) for all counties, avoiding any concerns with specification searching and cherry-picking of results, which are concerns highlighted by [Donohue et al. \(2018\)](#) and [Ferman et al. \(2020\)](#).

[Botosaru and Ferman \(2019\)](#) show that the existence of a good balance in terms of both covariates and pre-treatment outcomes implies tighter bounds on the bias of the SCM estimator relative to when a good balance is achieved in terms of only the pre-treatment outcomes. [Ferman et al. \(2020\)](#) suggest focusing on a specification which only includes all pre-treatment periods because this is not subject to arbitrary decisions regarding which pre-treatment outcome lags to include as predictors. More recently, [Kaul et al. \(2021\)](#) note that if covariates are to be included, then one should not use a specification that includes all pre-treatment lags. They show that if all lagged outcomes are included as separate predictors, this will make the covariates irrelevant in the estimation of weights. Furthermore, they suggest that for applied researchers, it should not be the first-best solution to simply optimize the pretreatment fit by using all pretreatment values of the outcome as separate economic predictors. Therefore, our results are estimated using a model that includes covariates and the pretreatment outcome average as an additional economic predictor, the approach that works best in their Monte Carlo study (see [Table 3](#)).¹⁸

Overall, the baseline results suggest that the opening of the Lanzhou-Urumqi HSR line has resulted in asymmetric effects in economic activities. There are winners and

¹⁷One necessary condition of the Jackknife-type resampling process is that it requires a smooth function of outcome variables. As an informal test, we conduct a leave-one-out test that drops one observation from our sample (excluding the treated counties) each time to obtain its corresponding median. We repeat this exercise 27,208 times based on the size of our sample. The summary statistics shows the inter-quartile range is 0.000002 which increases our confidence in applying the Jackknife+ approach. A recent application of the Jackknife approach for computing standard errors for ridge ASCM estimators can be seen in [Bennett \(2021\)](#).

¹⁸As a robustness check, we tested a specification using covariates and different years of lagged outcomes (using lagged outcomes for 2004, 2009 and 2014). The results obtained using this alternative specification are very similar. These results are available upon request.

losers from the opening of the HSR line in this relatively rural part of China. The ASCM estimates show that the opening of the Lanzhou-Urumqi HSR did not lead to statistically significant economic impacts on counties within Lanzhou in the first two years following treatment. However, it is possible that the effect may take time to evolve. This is in line with the argument that HSR towns can take years to become vibrant (Dong et al., 2021). Notably, we observe pronounced and significant nightlight effects for Gaolan county in 2018, three years after the opening of the HSR line. Average nightlight intensity in Gaolan increased by 0.65 per pixel, which is about four times higher than the average level of nightlight intensity in the pre-treatment period. This increment in nightlight is also three times as large as the increment found in Gaolan's neighbouring counties, Yongdeng and Yuzhong.

On the other hand, we find that most HSR counties in our sample outside of Lanzhou either experienced some decline in growth or experienced negligible growth in terms of nightlight intensity in the post-treatment period. Exceptions to this general pattern of results occur for Ganzhou county (a popular tourist destination in the urban district of Zhangye prefecture) and Urumqi county (a major county within Urumqi prefecture), where we find that the opening of the Lanzhou-Urumqi HSR line led to relatively large statistically significant impacts by the end of post-treatment period. There is also an interesting pattern observed in the results in terms of the direction of the effect for some counties. In Suzhou, one year after the HSR line opening, there was a positive estimated effect whereas Minhe experienced a negative effect. However, towards the end of our observed sample period, the average treatment effect for both counties were reversed, which could indicate some inter-county flow of resources. In our analysis, we find that counties within Lanzhou experienced the largest growth in nightlight intensity, while the counties connected between Lanzhou and Urumqi HSR line generally experienced little growth. A possible explanation is that a centralisation effect (sometimes also referred to as a 'siphon effect') is occurring and Lanzhou, as the core city, might be draining the most suitable

workers and economic resources from the less developed peripheral counties. To explore if such a core-peripheral relationship documented in the literature (e.g. [Baum-Snow et al., 2017](#)) is taking place here, we identify all the peripheral counties around Lanzhou and estimate average treatment effects for them using ASCM in the next section.

4.2 Asymmetrical economic impacts of HSR

Figure 6 plots the geographical distribution of the eight peripheral counties around Lanzhou. The estimated results for these counties are shown in Figure 7. The figures indicate that Huzhu, Yongjin, Tianzhu and Lintao counties experienced a decline in nightlight intensity, implying that some drain in economic activities has taken place. In Anding, Jingyuan, Jingtai and Huining counties, although good pre-treatment fits are not attainable, the general pattern of the results is suggestive of no economic growth occurring over the post-treatment period.

To contrast the effects for the different counties over time, we present the average of the estimated results by group in Table 3. Columns (1) - (4) report the corresponding treatment effects for each of the four groups for each post-treatment year. There are several interesting patterns emerging from this table. We find economic growth within Lanzhou to be the largest across the four groups. However, the HSR line did not bring any economic prosperity for the group of counties that lay between the two ends of the HSR line. In addition, any HSR benefits took a few years to emerge for the counties in the Urumqi group.

It appears to be the case that only counties within Lanzhou experienced positive growth following the opening of the HSR line, while the counties located in-between the HSR line did not. As transportation infrastructure and agglomeration effects may cause productive capital and skilled labour to move from rural regions to cities over time, the asymmetric economic effects we find in the Lanzhou-Urumqi HSE line are not unexpected. [Asher and Novosad \(2020\)](#) show that the primary impact of new roads built

in India after 2000 under the Prime Minister's Village Road Programme was to increase the access of workers to non-agricultural jobs. However, even after four years of road construction, India's regional road programme has a limited impact on enhancing consumption or promoting industrialisation in remote areas. In related work on highways in China, [Baum-Snow et al. \(2020\)](#) investigate the effects of an extensive modern national highway network constructed in China between the late 1990s and 2010 on local economic outcomes. They find a distinct pattern of winners and losers, with economic output and population increasing in regional primates at the expense of hinterland prefectures. In addition, [Banerjee et al. \(2020\)](#) estimate the effect of access to transportation networks on regional economic outcomes in China over a twenty-year period (1986-2003) of rapid income growth. They conclude that proximity to a highway or railroad was beneficial for an average Chinese county over that period and per capita GDP was higher in places closer to the line, although the effect is not large.

Lanzhou prefecture is a core center in northwest China, and the question is whether its growth resulted in spillover effects that helped boost the economic activities of the region and its neighbors, or whether there was more of a centralisation effect taking place. In this latter case, resources will be diverted from other counties in the region to the core city. To better assess the mechanisms underlying the changes in each of the four county groups, we explore a set of economic indicators that sheds light on the potential channels underlying our results in the next section.

4.3 Why HSR effects vary within and across prefectures?

The finding that Lanzhou has recently experienced a spurt in economic growth is consistent with results from the Milken Institutes Best-Performing Cities China Index. The index highlights that Lanzhou outperformed other cities based on data up to 2018 in attracting foreign direct investments, creating jobs and coordinating regional develop-

ment.¹⁹ Table 4 presents some characteristics that are similar to the ones used in the Milken Index. In comparing the pre-post differences in the Milken-type variables for the four county groups, it is evident that counties within Lanzhou appear to have grown at the expense of the other three county groups.

In addition, we also utilise land transaction data to examine changes in land transactions for each of the four county groups over time. The results are provided in Figure A8, revealing that land transactions for commercial service, manufacturing and residential services increased for counties within Lanzhou relative to other counties along the HSR line and other peripheral counties.

Our results show that three counties within Lanzhou prefecture (Gaolan, Yongdeng and Yuzhong) as well as a neighboring county Minhe (in Haidong prefecture) and several other tourist areas further away from Lanzhou (Minle, Ganzhou and Urumqi county) experienced increases in nightlight intensity following the opening of the HSR line. Several factors may help explain the heterogeneity of observed HSR effects across the different prefectures and within each prefecture. We start by analysing the characteristics of counties across each prefecture. Lanzhou is the capital city of Gansu province and a key transportation hub, bridging the northwest of China to the central and coastal provinces. In comparison with prefectures to its northwest, Lanzhou prefecture is overall relatively more industrialised and abundant in human capital. For example, it has the highest number of universities along the Lanzhou-Urumqi route (including the prestigious Lanzhou University, ranked among the top 30 universities in China). Lanzhou has geographical advantages relative to the far northwest prefectures (Xining and Haidong) and it can be expected to derive larger benefits from fixed investments in transport infrastructure.²⁰

Among our estimates for counties within Lanzhou, we find that the effects are

¹⁹Please see <https://sme.asia/lanzhou-tops-best-performing-cities-china-index/> and <https://milkeninstitute.org/report/best-performing-cities-china-2020> for the Milken's 2020 ranking of city performance.

²⁰According to Figure A4, rural Xining and Haidong prefectures in Qinghai province are located to the northwest of Lanzhou. Their peripheral prefectures are undeveloped and much less populous than those of neighbouring Lanzhou prefecture.

largest toward the end of 2018. This might correspond to the connection to the new Lanzhou–Baoji HSR line that reinforced the existing HSR effect in Lanzhou. Figure 3 highlights that the Lanzhou–Baoji HSR line which opened in mid-2017 helped to complete the connection of the existing Lanzhou-Urumqi line to the rest of China’s HSR network. This connection of Lanzhou to central China makes it much more accessible to other major cities and facilitates a link between East and West on both sides of the Hu Line.

In addition, in 2019, Lanzhou New Area (located 30 kilometers away from its main city and split off from Yongdeng and Gaolan counties) was formally established as a county-level division of Lanzhou.²¹ The establishment of Lanzhou New Area is the culmination of multiple strategies both from Chinese Central Government and Gansu Provincial Government. As early as 2010, the Lanzhou municipal government had announced its goal of building Lanzhou as an important transportation hub in the western region. The ambition of the municipality began to expand when China’s central government was preparing for the “New Urbanisation Programme (2014-2020)” and the “Reconstruction Programme of Old Industrial Bases (2013-2022)” in 2012. The municipality declared that it would be creating the city as an industrial and trading platform for China’s westward strategy, and additionally that the city would serve as a national transport hub rather than merely on the regional level (Guo, 2018). Recognising that linking Lanzhou New Area with the original main city area is a crucial factor for its development, an inner-city railway line called the “Lanzhou-Zhongchuan Airport intercity railway” was constructed and started to operate from 30 September 2015. This railway line (61 kilometres or 38 miles long) connected the Lanzhou HSR Station with Lanzhou Zhongchuan International Airport, and includes a station in the Lanzhou New Area.

The positive results in several counties around Lanzhou in 2018 and the fact that Lan-

²¹Lanzhou’s main urban area is constrained in size by its geographical location as it is in a valley surrounded by mountains, with the Yellow River crossing through the city. The topography of Lanzhou’s main city hampers its ability to easily expand in size, explaining why a new area was chosen to facilitate Lanzhou’s development.

zhou was found by the Milken Institute in 2020 to be the city in the whole of China that has grown the most is a promising indication that disparities on both sides of the Hu Line might gradually dissipate as a result of these infrastructure investments.

There are also within-prefecture heterogeneity effects to consider. There are several possible reasons why we find the estimated effects for Gaolan county within Lanzhou to be the largest among the effects estimated. First, as mentioned above, Lanzhou New Area includes part of Gaolan. Increased economic activity we capture in Gaolan could reflect economic activity in Lanzhou New Area. In our analysis, we do not formally demarcate Lanzhou New Area as the official boundary was only created after 2019, which is the end of our sample period. Second, Gaolan county lies in between Yongdeng and Yuzhong counties and is closer to the HSR station stop in Lanzhou. It enjoys a proximity advantage over its neighbours which can be helpful in boosting local economic growth (see Table 1). Third, among these three counties, Gaolan successfully attracted more firms after it was connected to the HSR line.

With respect to Minhe within Haidong prefecture, the results are more mixed. Being connected to the HSR line led to a reduction in nightlight intensity for Minhe in the first three years before a reversal occurred and it became more vibrant in the fourth and fifth years. There could be two reasons for this reversal. First, there might be a centralisation effect occurring. Minhe county is located next to Lanzhou prefecture, and the falling trade costs between peripheral Minhe and metropolitan Lanzhou led to a decline in output from Minhe (Faber, 2014). The positive effect after 2018 likely arose when positive spillovers from Lanzhou helped to dominate the centralisation effect. Second, Minhe is more oriented towards primary industries, and it can take time for positive benefits to emerge in an agricultural county following the initial HSR connection.

This hypothesis for the observed reverse effect in Minhe is consistent with a prediction made in a recent work by Dong et al. (2021) who suggest that conditional on location and the size of existing nearby urban markets, it takes around four years before a HSR new

town begins to experience steady economic growth. In a similar vein, Wang et al. (2020) argue that HSR does not always promote economic growth in the short term. However, it could possibly accelerate economic growth over the longer term as sectoral transitions from secondary to knowledge-based tertiary industries take place (Dong et al., 2020; Vickerman, 2015). This hypothesis is possibly applicable to Ganzhou within Zhangye prefecture which shows an increasing trend in the size of its positive effect. Its firm density and tertiary sector employment share increases substantially following its connection to the HSR network. There are also no developed counties near Ganzhou to hamper its growth prospects, as Ganzhou is about six hours away by HSR train from the Lanzhou HSR station. In addition, it is home to attractive tourist sites such as Zhangye National Geopark, which is known for its colourful rock formations and was named a UNESCO World Heritage Site in 2009. The park is now a destination for many Chinese and international tourists, and the HSR line from Lanzhou has helped increase its accessibility and popularity.

Other cases in our sample which experience positive effects are Minle county and Urumqi county. The finding of a significant effect for Minle may be attributable to the fact that it is next to Ganzhou which is a tourist destination. The significant effect for Urumqi county only emerges in 2019, the end of our sample period. The delay in Urumqi benefitting from the HSR line might be due to the lack of agglomeration of industries. As it is relatively isolated at the western end of the HSR line, it does not enjoy positive spillovers from economic activity in any peripheral counties.

We also observe negative effects in Ledu district, Linze and Suzhou. Linze and Suzhou are located somewhere in the middle of the HSR line and could be recipients of a centralisation effect occurring that benefits Lanzhou. It is a little more surprising to find negative significant effects for Ledu district given that it is contiguous to Lanzhou, albeit the effects are small (around -0.08 in 2019). This result could represent a small local centralisation effect occurring, whereby local resources are being diverted to within Lanzhou.

4.4 Robustness checks

In this section, we conduct two further robustness checks to determine whether our ridge ASCM estimates are sensitive to the use of alternative donor pools and to make further inferences. [Abadie \(2019\)](#) notes that units in the donor pool should be chosen judiciously to provide a reasonable control for the treated unit. Including unsuitable units in the donor pool (with large differences in observed or unobserved characteristics relative to the treated unit) is a recipe for bias. While there are advantages in using a large reservoir of control units to enable the data-driven ASCM approach to optimise, choose and weight the most appropriate controls, there is a concern that this could create bias in the ASCM estimator.

Therefore, our first robustness check is to restrict the number of control units in the donor pool. Exploring all possible different sets of control units is not practically feasible because of the large donor pool and there will be millions of possible combinations. To address this issue, we design a simple objective criterion to choose the most appropriate control units in the first step. Our approach extends the idea of [Abadie et al. \(2015\)](#) from a related context. In conducting their placebo runs, they use only control units with a mean squared prediction error (MSPE) value that is within a certain multiple of the pre-treatment prediction error observed for the treated unit. Our criterion is also related to the approach proposed in [Zubizarreta \(2015\)](#). In an effort to design optimal weights to balance treated and control groups in an observational setting, he proposes constraining the absolute difference in means of the weighted pre-intervention outcomes to be less than a user-specified threshold.

Specifically, we repeatedly compute the pairwise standard deviation of the difference in outcome values for each treated unit i paired against a control unit. This exercise is conducted for each time period t across all periods prior to the treatment. We then sort the computed standard deviation vector and seek a group of minimum values before

treatment year T such that

$$\operatorname{argmin} SD_i = \sqrt{\frac{\sum_t^{T-1} \left(Y_{it}(1) - Y_{it}(0) - \mathbb{E}[Y_{it}(1) - Y_{it}(0)] \right)^2}{T - t}}, \quad \forall t < T. \quad (5)$$

The purpose of this algorithm is to minimise the amount of variation or dispersion in the difference in outcome values prior to treatment. An advantage of doing so is that it helps the algorithm focus on the control units that are most similar and greatly increases the likelihood of finding a good pre-treatment fit. In our analysis, we select the top 100, top 200, top 300 and top 500 counties with the lowest pre-treatment standard deviation values. This allows us to check the sensitivity of our main ASCM estimates based on the larger donor pool. Another advantage of our proposed algorithm for implementing SCM estimators with large donor pools is that it automates the donor pool selection process. This process efficiently looks for control units which are similar to the treated unit from a large donor pool, reducing bias and considerably cutting down on computational time.

Figure 8 shows the results for the eight counties where there are larger mean effects. There is evidence that the post-treatment point estimates across the alternative restricted donor pools for Gaolan and Yuzhong, the two counties where we have the largest estimated effects, move within a reasonable range that overlaps with the confidence interval of our larger donor pool. On the other hand, there are larger fluctuations in results for Ganzhou, Urumqi and Yongdeng counties. For Yongdeng, although the pre-treatment fit is good, the post-treatment impacts can be either positive or negative depending on the composition of the control group. This suggests that the results for Yongdeng are quite unstable and should be interpreted with caution. Overall, this algorithm helps to mimic the parallel trend assumption underlying difference-in-differences estimators.²²

²²In our balancing checks, we find that the top 100-500 counties often share similar economic attributes (e.g. industrial composition, air pollution and fixed investments) suggesting the practical utility of the algorithm. However, our proposed method also has some limitations. The algorithm is more likely to find perfect fits for the pre-treatment period in some cases when sufficiently long pre-treatment data are available, but this does not necessarily result in less biased estimates. This risk of over-fitting has been

The second robustness check we conduct is the leave-one-out placebo test using the original SCM approach suggested by [Abadie et al. \(2015\)](#). This is an alternative means of conducting statistical inference to gauge whether an effect is statistically significant by conducting placebo runs. [Figure 9](#) reports the results for our focus counties. Based on the top 200 counties with the lowest standard deviations as the donor pool, as seen previously in [Figure 8](#), we find that the results for Gaolan where the largest impact occurs remains robust.²³ The results for Yuzhong are slightly weaker but still suggestive that the impacts are positive and significant over a wider confidence interval. The results of the placebo test for Ganzhou, Urumqi and Yongdeng counties are consistent with the results from [Figure 8](#), suggesting that their small estimated impacts are not unique in the data. It is possible that non-treatment counties have equally sized or larger impacts estimated.

5 Conclusion

The Lanzhou-Urumqi HSR line lies along the land corridor of the BRI and is an important gateway for China to Central Asia, the Middle East, Africa and Europe under the BRI initiative. As the only HSR line in northwest China to the left of the Hu Line, it can also play an important role in promoting economic integration between the eastern and western parts of China. This paper estimates the causal effects of the opening of the Lanzhou-Urumqi HSR line in northwest China in 2014 on regional economic development along the Silk Road. We employ the ridge augmented SCM approach to assess the effect of the HSR line connection on the economic activity along this HSR route for the connected counties. To circumvent any measurement issues pertaining to regional GDP, we use a newly extended satellite nightlight data series from 2000 to 2019 (developed using machine-learning techniques) to proxy for economic activity.

highlighted by [Abadie \(2019\)](#).

²³There are heavy computation burdens when using a large donor pool (1,504 counties) as in our case. Therefore, we rely on [Equation 5](#) to search for optimal combinations and restrict the donor pool to the top 200 control counties.

The results from the Lanzhou-Urumqi HSR line show that HSR may help stimulate some local economies, in particular, counties which have a higher initial industrial endowment. In line with theory on regional economic development, there can be winners and losers from investments in economic infrastructure in a region such as the opening of the Lanzhou-Urumqi HSR line. Whether counties in the region benefit will depend on the relative strength of any spillover benefits, or whether a centralisation effect that enhances a core city dominates. Consistent with the Milken Institute's 2020 ranking of Lanzhou as the fastest growing city in the whole of China in recent years, our positive and robust results for Gaolan and Yuzhong counties within Lanzhou are an indication of a possible role that HSR can play in stimulating China's future regional growth. At the same time, the growth in Lanzhou appears to have come at the expense of some other counties in the region, especially counties in the middle of the HSR line between Lanzhou and Urumqi. Our evidence using a five-year follow up period since the opening of the Lanzhou-Urumqi HSR line suggests that the potential for urban agglomeration in rural counties in Western China, which are located at high altitudes and surrounded by deserts, appears to be rather limited.

As the construction costs for Lanzhou-Urumqi HSR were high (\$22.56 billion USD), it will take time for full returns to the investment to be recouped. We leave it to future research to examine the longer run effects and whether the HSR line can permanently alter the geographical distribution of economic activity in northwest China. At the moment, it remains more of an open question whether such a huge investment in transport infrastructure can help to gradually dissipate the disparities on both sides of the Hu Line and help sustain economic growth along China's ancient Silk Road.

References

- Abadie, Alberto (2019), "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of Economic Literature*.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2015), "Comparative politics and the synthetic control method." *American Journal of Political Science*, 59, 495–510.
- Abadie, Alberto and Javier Gardeazabal (2003), "The economic costs of conflict: A case study of the Basque country." *American Economic Review*, 93, 113–132.
- Asher, Sam and Paul Novosad (2020), "Rural roads and local economic development." *American Economic Review*, 110, 797–823.
- Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2020), "On the road: Access to transportation infrastructure and economic growth in China." *Journal of Development Economics*, 145, 102442.
- Barber, Rina Foygel, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani (2021), "Predictive inference with the Jackknife+." *The Annals of Statistics*, 49, 486–507.
- Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang (2017), "Roads, railroads, and decentralization of Chinese cities." *Review of Economics and Statistics*, 99, 435–448.
- Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt (2020), "Does investment in national highways help or hurt hinterland city growth?" *Journal of Urban Economics*, 115, 103124.
- Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2021), "The augmented synthetic control method." *Journal of the American Statistical Association*, 1–34.
- Bennett, Magdalena (2021), "All things equal? Heterogeneity in policy effectiveness against Covid-19 spread in Chile." *World Development*, 137, 105208.
- Bird, Julia, Mathilde Lebrand, and Anthony J Venables (2020), "The belt and road initiative: Reshaping economic geography in central Asia?" *Journal of Development Economics*, 144, 102441.
- Botosaru, Irene and Bruno Ferman (2019), "On the role of covariates in the synthetic control method." *The Econometrics Journal*, 22, 117–130.
- Chen, Mingxing, Yinghua Gong, Yang Li, Dadao Lu, and Hua Zhang (2016), "Population distribution and urbanization on both sides of the Hu Huanyong Line: Answering the Premier's question." *Journal of Geographical Sciences*, 26, 1593–1610.
- Chen, Yvonne Jie, Pei Li, and Yi Lu (2018), "Career concerns and multitasking local bureaucrats: Evidence of a target-based performance evaluation system in China." *Journal of Development Economics*, 133, 84–101.
- Chen, Zuoqi, Bailang Yu, Chengshu Yang, Yuyu Zhou, Shenjun Yao, Xingjian Qian, Congxiao Wang, Bin Wu, and Jianping Wu (2021), "An extended time series (2000–2019) of global NPP-VIIRS-like nighttime light data from a cross-sensor calibration." *Earth System Science Data*, 13, 889–906.

- Cole, Matthew A, Robert JR Elliott, and Bowen Liu (2020), "The impact of the Wuhan Covid-19 lockdown on air pollution and health: A machine learning and augmented synthetic control approach." *Environmental and Resource Economics*, 76, 553–580.
- Dong, Lei, Rui Du, Matthew Kahn, Carlo Ratti, and Siqi Zheng (2021), "Ghost cities versus boom towns: Do China's HSR new towns thrive?" *Regional Science and Urban Economics*, 103682.
- Dong, Xiaofang, Siqi Zheng, and Matthew E Kahn (2020), "The role of transportation speed in facilitating high skilled teamwork across cities." *Journal of Urban Economics*, 115, 103212.
- Donohue, John J, Abhay Aneja, and Kyle D Weber (2018), "Right-to-carry laws and violent crime: A comprehensive assessment using panel data, the lasso, and a state-level synthetic controls analysis (No. 23510)." *National Bureau of Economic Research*.
- Faber, Benjamin (2014), "Trade integration, market size, and industrialization: Evidence from China's national trunk highway system." *Review of Economic Studies*, 81, 1046–1070.
- Ferman, Bruno, Cristine Pinto, and Vitor Possebom (2020), "Cherry picking with synthetic controls." *Journal of Policy Analysis and Management*, 39, 510–532.
- Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li (2021), "Which night lights data should we use in economics, and where?" *Journal of Development Economics*, 149, 102602.
- Glaeser, Edward L and Joshua D Gottlieb (2008), "The economics of place-making policies." *Brookings Papers on Economic Activity*, 39, 155–253.
- Guo, Jie (2018), "Industrial restructuring and urban spaces transformation in Lanzhou city, with focus on its transition periods." *Gebhardt Hans (Ed.), Urban Governance, Spatial Planning and Economic Development in the 21th Century China*. Berlin: LIT Verlag.
- Henderson, J Vernon, Adam Storeygard, and David N Weil (2012), "Measuring economic growth from outer space." *American Economic Review*, 102, 994–1028.
- Hu, Huanyong (1935), "Distribution of China's population: Accompanying charts and density map (in Chinese)." *Acta Geographica Sinica*, 2 (2), 33–748.
- Huang, Ying et al. (2019), "The correlation between HSR construction and economic development—empirical study of Chinese cities." *Transportation Research Part A: Policy and Practice*, 126, 24–36.
- Jia, Junxue, Guangrong Ma, Cong Qin, and Liyan Wang (2020), "Place-based policies, state-led industrialisation, and regional development: Evidence from China's Great Western Development Programme." *European Economic Review*, 123, 103398.
- Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler (2021), "Standard synthetic control methods: The case of using all preintervention outcomes together with covariates." *Journal of Business & Economic Statistics*, 1–15.
- Ke, Xiao, Haiqiang Chen, Yongmiao Hong, and Cheng Hsiao (2017), "Do China's high-speed-rail projects promote local economy? New evidence from a panel data approach." *China Economic Review*, 44, 203–226.

- Kline, Patrick and Enrico Moretti (2014a), "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics*, 129, 275–331.
- Kline, Patrick and Enrico Moretti (2014b), "People, places, and public policy: Some simple welfare economics of local economic development programs." *Annu. Rev. Econ.*, 6, 629–662.
- Kong, Dongmin, Lihua Liu, and Zhiqing Yang (2021), "High-speed rails and rural-urban migrants' wages." *Economic Modelling*, 94, 1030–1042.
- Liang, Yutian, Keyang Zhou, Xun Li, Zhengke Zhou, Wei Sun, and Jiaqi Zeng (2020), "Effectiveness of high-speed railway on regional economic growth for less developed areas." *Journal of Transport Geography*, 82, 102621.
- Lin, Yatang (2017), "Travel costs and urban specialization patterns: Evidence from China's high speed railway system." *Journal of Urban Economics*, 98, 98–123.
- Long, Fenjie, Longfei Zheng, and Zhida Song (2018), "High-speed rail and urban expansion: An empirical study using a time series of nighttime light satellite data in China." *Journal of Transport Geography*, 72, 106–118.
- Qin, Yu (2017), "'No county left behind?' The distributional impact of high-speed rail upgrades in China." *Journal of Economic Geography*, 17, 489–520.
- Vickerman, Roger (2015), "High-speed rail and regional development: the case of intermediate stations." *Journal of Transport Geography*, 42, 157–165.
- Wang, Chunyang, Weidong Meng, and Xinshuo Hou (2020), "The impact of high-speed rails on urban economy: An investigation using night lighting data of Chinese cities." *Research in Transportation Economics*, 80, 100819.
- Zheng, Longfei, Fenjie Long, Zheng Chang, and Jingsong Ye (2019), "Ghost town or city of hope? The spatial spillover effects of high-speed railway stations in China." *Transport Policy*, 81, 230–241.
- Zheng, Siqi and Matthew E Kahn (2013), "China's bullet trains facilitate market integration and mitigate the cost of megacity growth." *Proceedings of the National Academy of Sciences*, 110, E1248–E1253.
- Zou, Ben (2018), "The local economic impacts of military personnel." *Journal of Labor Economics*, 36, 589–621.
- Zubizarreta, José R (2015), "Stable weights that balance covariates for estimation with incomplete outcome data." *Journal of the American Statistical Association*, 110, 910–922.

Table 1: Characteristics of counties along the Lanzhou–Urumqi HSR line (the northwestern Silk Road land route)

HSR Connected Counties	Prefecture, Province	Distance (km)	Connection Date	Group
Yongdeng County	Lanzhou, Gansu	112	26/12/2014	Counties within Lanzhou
Gaolan County	Lanzhou, Gansu	51	26/12/2014	
Yuzhong County	Lanzhou, Gansu	60	26/12/2014	
Ganzhou District	Zhangye, Gansu	3.1	26/12/2014	Between the Lanzhou–Urumqi HSR
Minle County	Zhangye, Gansu	7	26/12/2014	
Linze County	Zhangye, Gansu	8.7	26/12/2014	
Gaotai County	Zhangye, Gansu	11.6	26/12/2014	
Guazhou District	Jiuquan, Gansu	10	26/12/2014	
Yumen County	Jiuquan, Gansu	3	26/12/2014	
Suzhou District	Jiuquan, Gansu	6.7	26/12/2014	
Datong Autonomous Region	Xining, Qinghai	6.6	26/12/2014	
Ledu District	Haidong, Qinghai	3.6	26/12/2014	
Minhe Autonomous Region	Haidong, Qinghai	2.9	26/12/2014	
Menyuan Autonomous Region	Haibei, Qinghai	10	26/12/2014	
Shanshan County	Tulufan, Xinjiang	6.6	16/11/2014	County within Urumqi
Urumqi County	Urumqi, Xinjiang	42.6	16/11/2014	

Note: Distance standards for county government location to its nearest HSR station. We divide all HSR connected counties into three groups, including all HSR connected counties within Lanzhou prefecture, all HSR counties between Lanzhou and Urumqi prefectures, and a major county within Urumqi prefecture. Data for other counties within Urumqi prefecture were either not recorded in the China County Statistical Yearbooks or contained many missing values in the Provincial Statistical Yearbooks. We therefore did not include these counties in our analysis.

Table 2: Characteristics of four focal groups before and after the Lanzhou-Urumqi HSR connection

ASCM covariates	(1)		(2)		(3)		(4)	
	Within Lanzhou		Peripheral Lanzhou		Between L-U HSR		Within Urumqi	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Nightlight Intensity (per pixel)	0.156	0.282	0.0536	0.0649	0.0776	0.0991	0.114	0.164
Panel A: Industrial Characteristics								
Industrial firm density (no. /km ²)	0.0103	0.0102	0.00577	0.00624	0.00700	0.00774	0.00125	0.00141
Fixed investment ratio (100%)	1.205	1.129	1.435	1.507	1.125	1.270	2.131	2.452
Secondary ind. value added share (100%)	0.522	0.432	0.401	0.336	0.489	0.426	0.290	0.238
GDP per capita (CNY/ppl.)	19,410	25,543	14,335	16,683	33,026	36,372	30,471	37,416
Panel B: Human Resources								
Population density (ppl./km ²)	0.00960	0.00953	0.00986	0.00993	0.00769	0.00784	0.00158	0.00147
High Sch. Stu. (per 10k Hukou ppl.)	558.8	460.7	722.7	600.1	620.8	584.6	271.9	313.9
Primary Sch. Stu. (per 10k Hukou ppl.)	532.6	504.0	646.1	602.6	735.3	703.1	610.8	657.8
Secondary employee share (per 10K ppl.)	670.2	695.8	610.1	715.4	891.4	1,072	203.8	229.9
Tertiary employee share (per 10K ppl.)	1,473	1,540	1,070	1,240	1,263	1,547	1,058	1,136
Public expense per capita (CNY/ppl.)	4,294	5,989	5,636	7,483	6,327	8,480	14,964	20,002
Number of Counties	3		8		12		1	

Note: This table compares mean values of some pre-treatment (2010-2014) and post-treatment (2015-2019) characteristics used as covariates in the Augmented SCM for the four focal groups. In particular, these groups are all HSR connected counties within Lanzhou prefecture ('Within Lanzhou'), peripheral counties around Lanzhou ('Peripheral Lanzhou'), all HSR counties connected between Lanzhou ('Between L-U HSR'), and a major county within Urumqi prefecture ('Within Urumqi'). GDP and public expenditure *per capita*, secondary and tertiary employment share is calculated by GDP, total public expenditure, secondary and tertiary employees over total number of Hukou registrations. 'Population density' is total Hukou registration divided by size of county administrative areas. 'Fixed investment ratio' and 'Secondary ind. added value share' are a county's fixed investment and secondary industrial added value divided by its GDP, respectively. 'Primary and high Sch. Stu.' are the primary and high school students weighted by total Hukou registration.

Table 3: Average treatment effects on the treated using ridge ASCM for the opening of the Lanzhou-Urumqi HSR

	(1)	(2)	(3)	(4)
Year\Group	Within Lanzhou	Periperal	Between L-U	Within Urumqi
2015	-0.0035	-0.0061	0.0007	-0.0020
2016	0.0263	-0.0169	0.0041	-0.0187
2017	0.0664	-0.0351	-0.0257	-0.0440
2018	0.3586	-0.0475	-0.0195	-0.0097
2019	0.3057	-0.0510	0.0078	0.1341
Covariates	Yes	Yes	Yes	Yes
Avg outcome	Yes	Yes	Yes	Yes

Note: This table reports the average estimates for four focal county groups: 1. all HSR connected counties within Lanzhou prefecture ('Within Lanzhou'), 2. peripheral counties around Lanzhou ('Peripheral Lanzhou'), 3. all HSR counties connected between Lanzhou ('Between L-U HSR'), and 4. a major county within Urumqi prefecture ('Within Urumqi'). Please refer to Figures 4, 5 and 7 for the detailed results of the individual counties, including statistical significance. The donor pool comprises of 1,504 counties and excluded all counties connected to the HSR up to 2019 and counties not recorded in the China County Statistical Yearbooks, and also excluded any counties that experienced administrative border changes. 'Covariates' are the predictors listed in the Table 2. 'Avg outcome' refers to average outcomes from 2000 to 2013 as an additional covariate suggested by Kaul et al. (2021). We also estimated models which included various multiple combinations of lagged outcomes. The resulting estimates remain very similar.

Table 4: Characteristics using Milken variables before and after the Lanzhou-Urumqi HSR connection

Milken-type variables	(1)		(2)		(3)		(4)	
	Within Lanzhou		Peripheral Lanzhou		Between L-U HSR		Within Urumqi	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Fix investment growth (%)	0.206	0.215	0.584	0.396	0.413	0.297	0.680	0.343
Employment growth in the secondary industry (%)	-0.0332	-0.0170	0.751	0.243	2.500	0.747	0.696	0.449
Employment growth in the tertiary industry (%)	-0.0340	0.00158	1.319	0.230	0.821	0.246	0.184	0.0776
Hukou registration growth (%)	-0.0207	-0.008	-0.004	0.0015	0.0092	0.0033	-0.063	-0.044
Number of Counties	3		8		12		1	
ATT (across 2015 - 2019)	0.1252		-0.0378		-0.0083		0.0028	

Note: This table calculates variables used in the Milken report that compares the growing potential for Chinese cities. “Pre” refers to average values between 2010-2014 and “Post” refers to average values between 2015-2019.

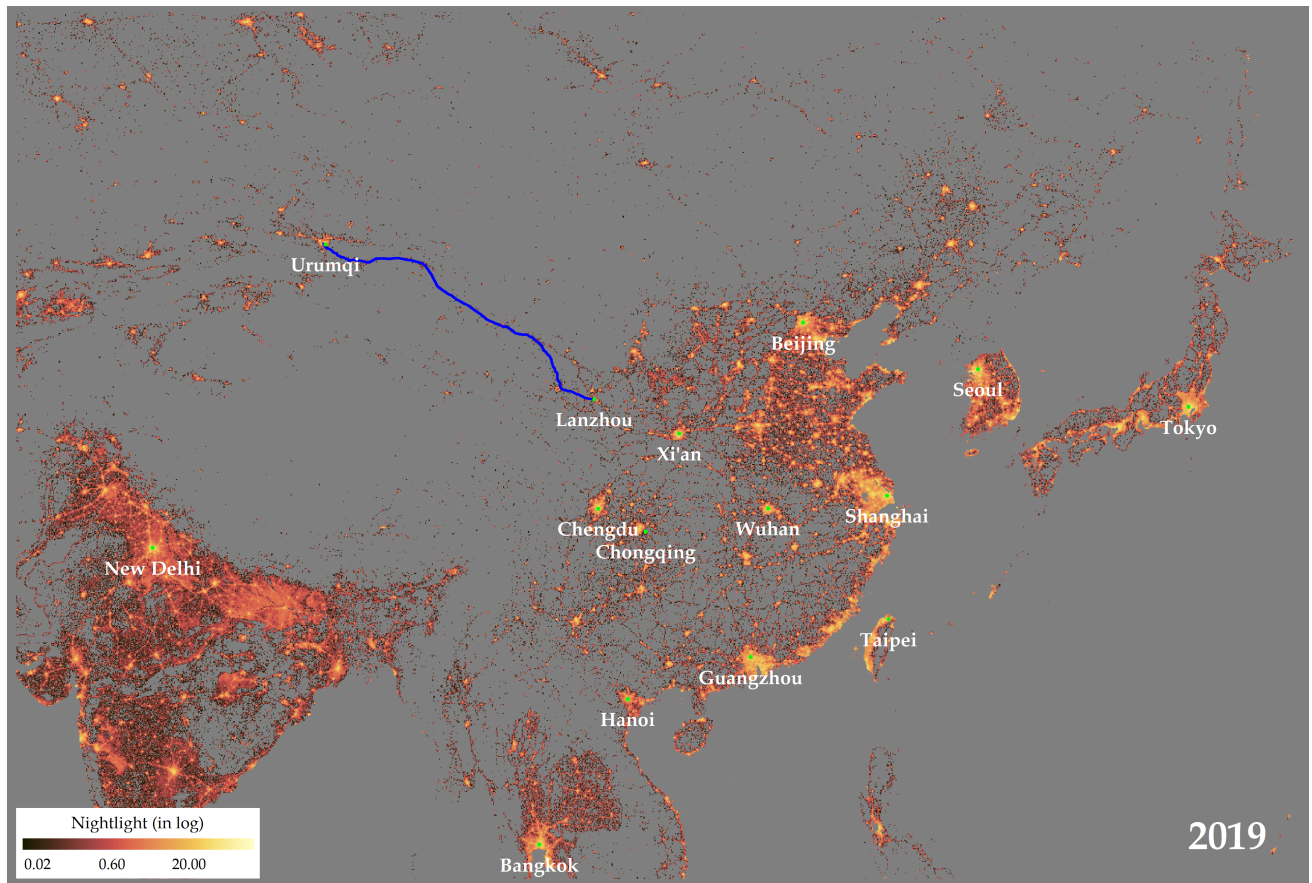
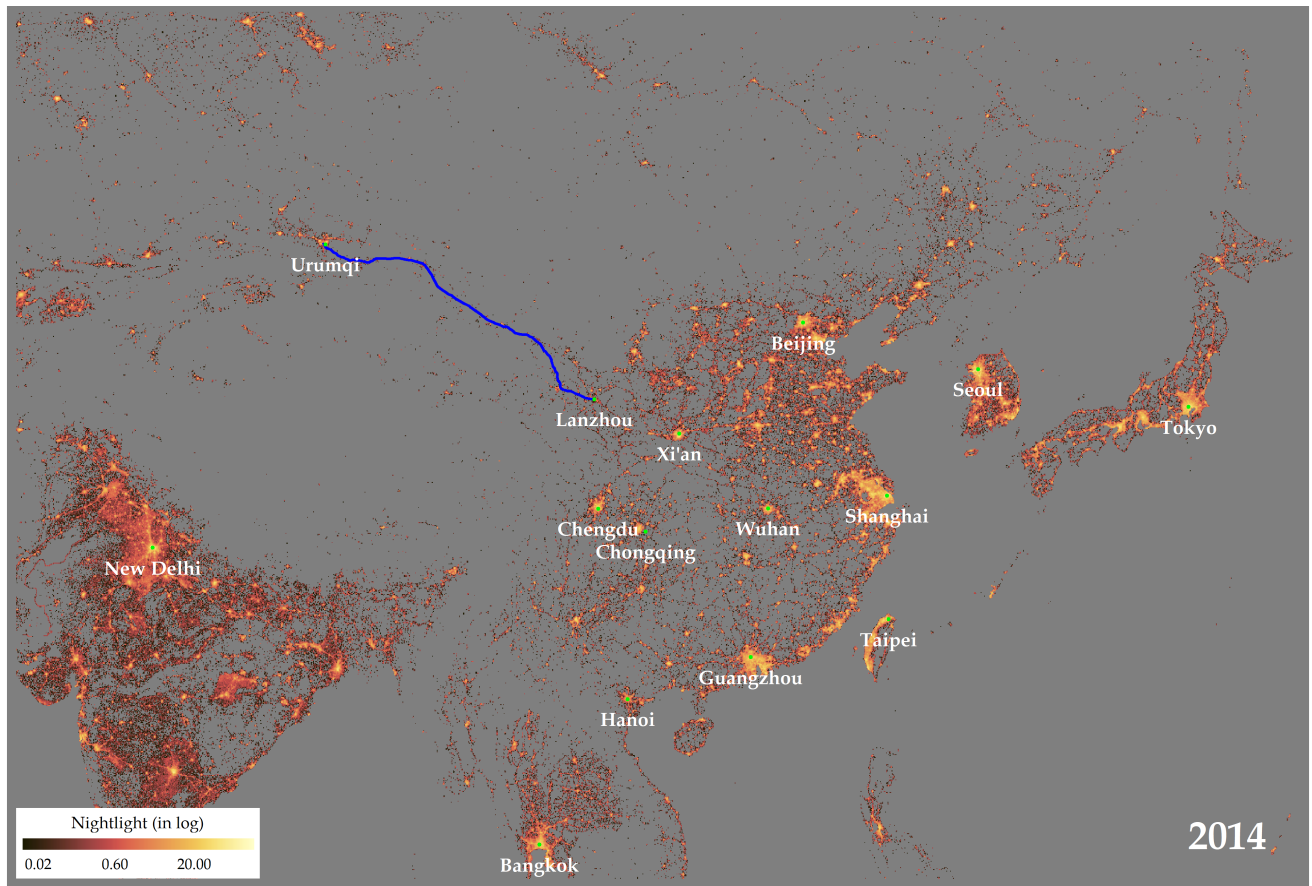


Figure 1: Nightlight intensity in 2014 and 2019

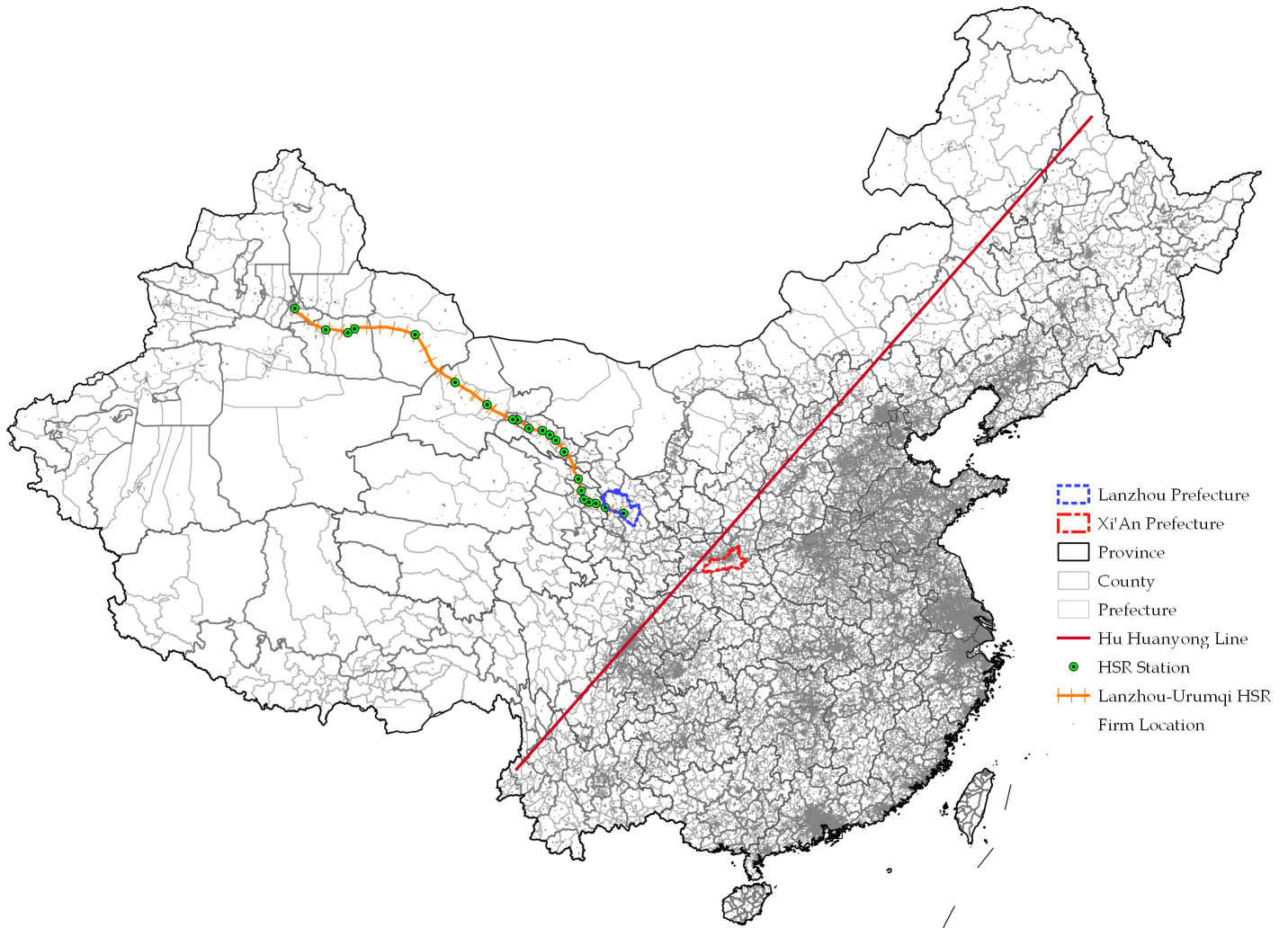


Figure 2: Location of Chinese manufacturing firms (2013) and the Hu Huanyong Line

Note: The stacked firm locations are from the 2013 Annual Survey of Industrial Manufacturing Firms. Their exact coordinates are taken using the Baidu Map API.

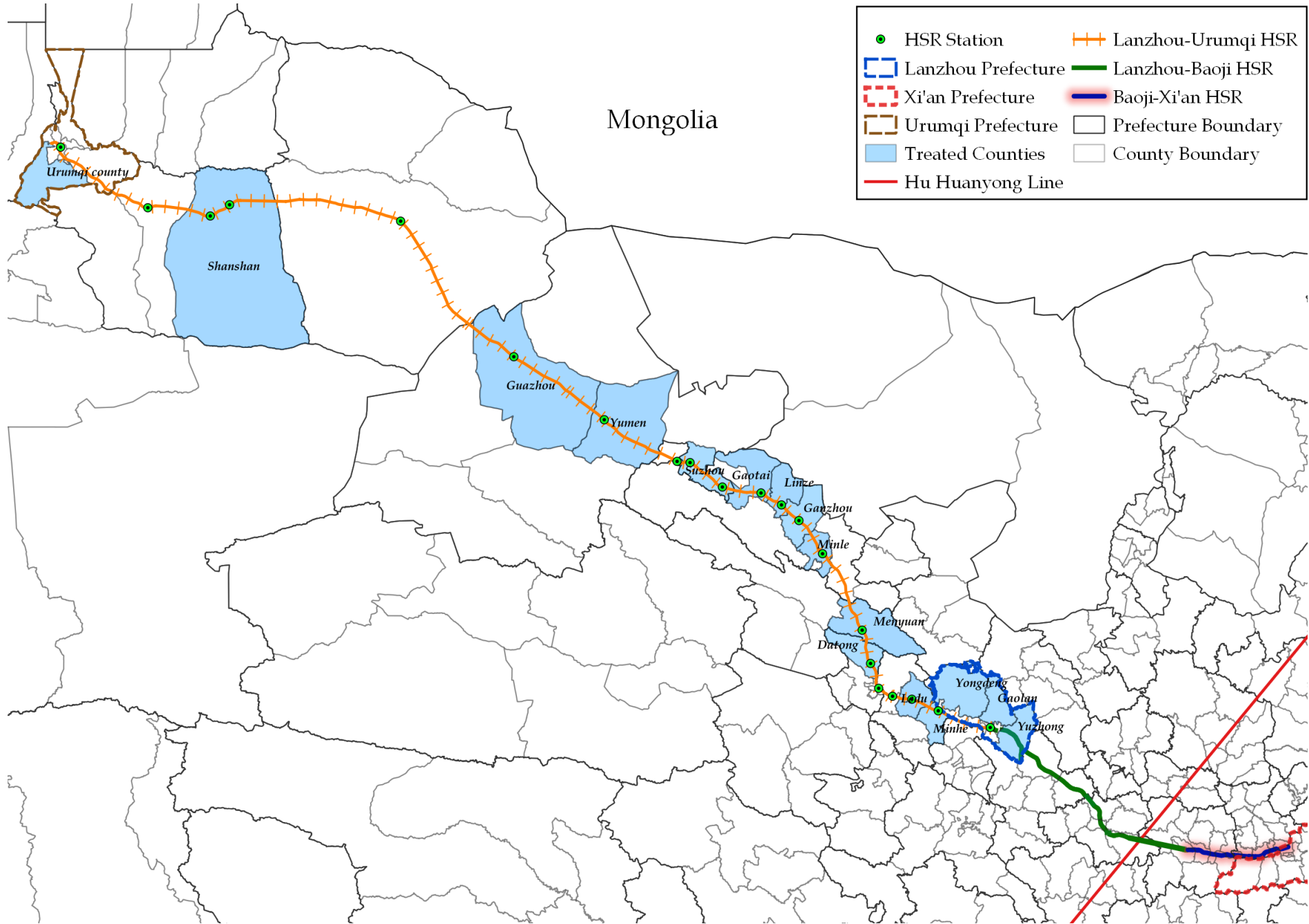


Figure 3: Location of the 16 treated counties (Source: *China County Statistical Yearbook*)

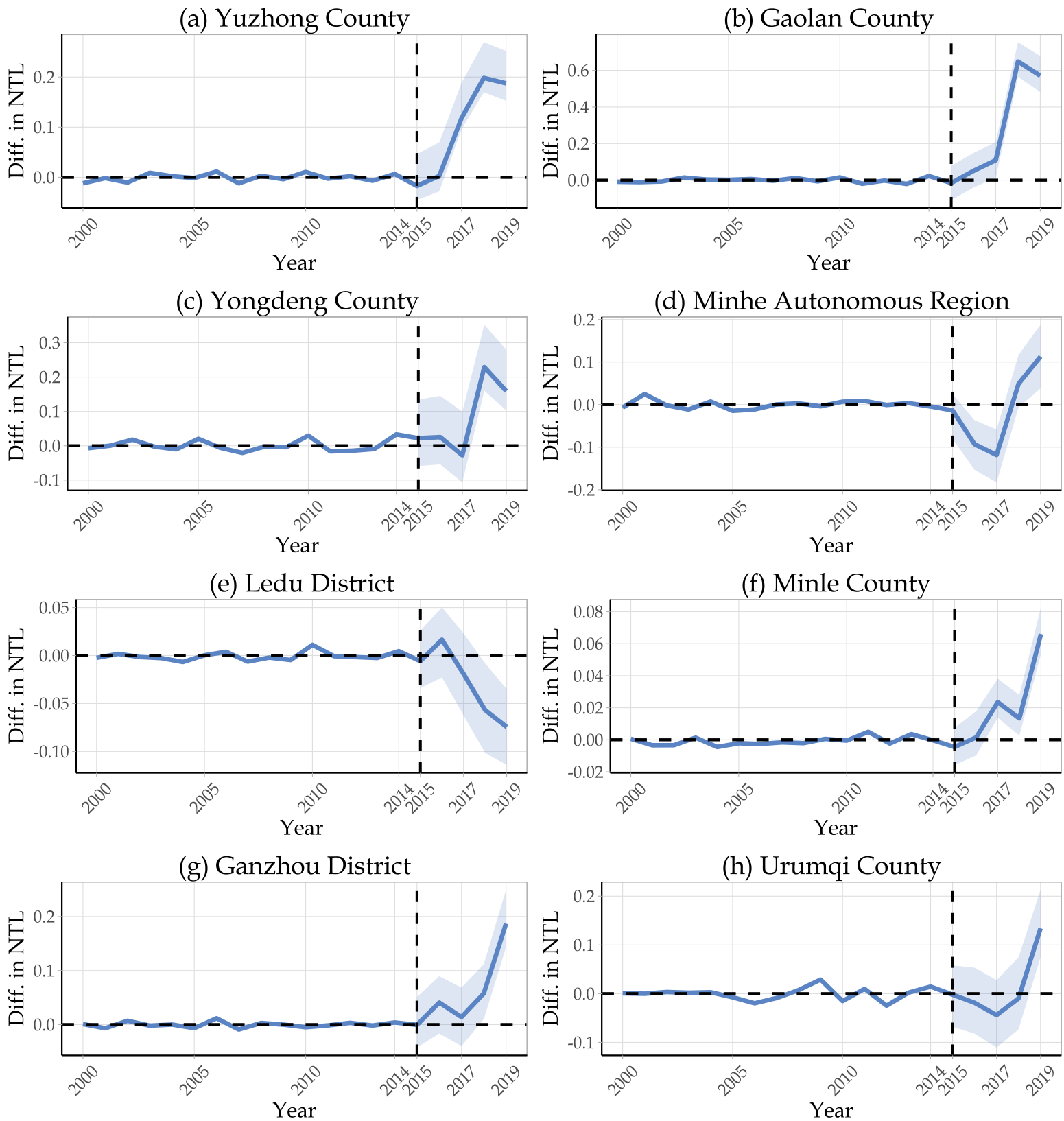


Figure 4: Estimated differences in nightlight intensity between treated and synthetic counties

Note: This figure plots point estimates for counties connected by the Lanzhou–Urumqi HSR. Shaded areas are 95% confidence intervals based on the Jackknife+ approach in [Barber et al. \(2021\)](#). The treatment date for the treated counties are either November or December 2014 and the treatment effect starts from 2015 represented by the vertical dashed line. Appendix Figure A6 shows the fit between treated and synthetic counties prior to the treatment.

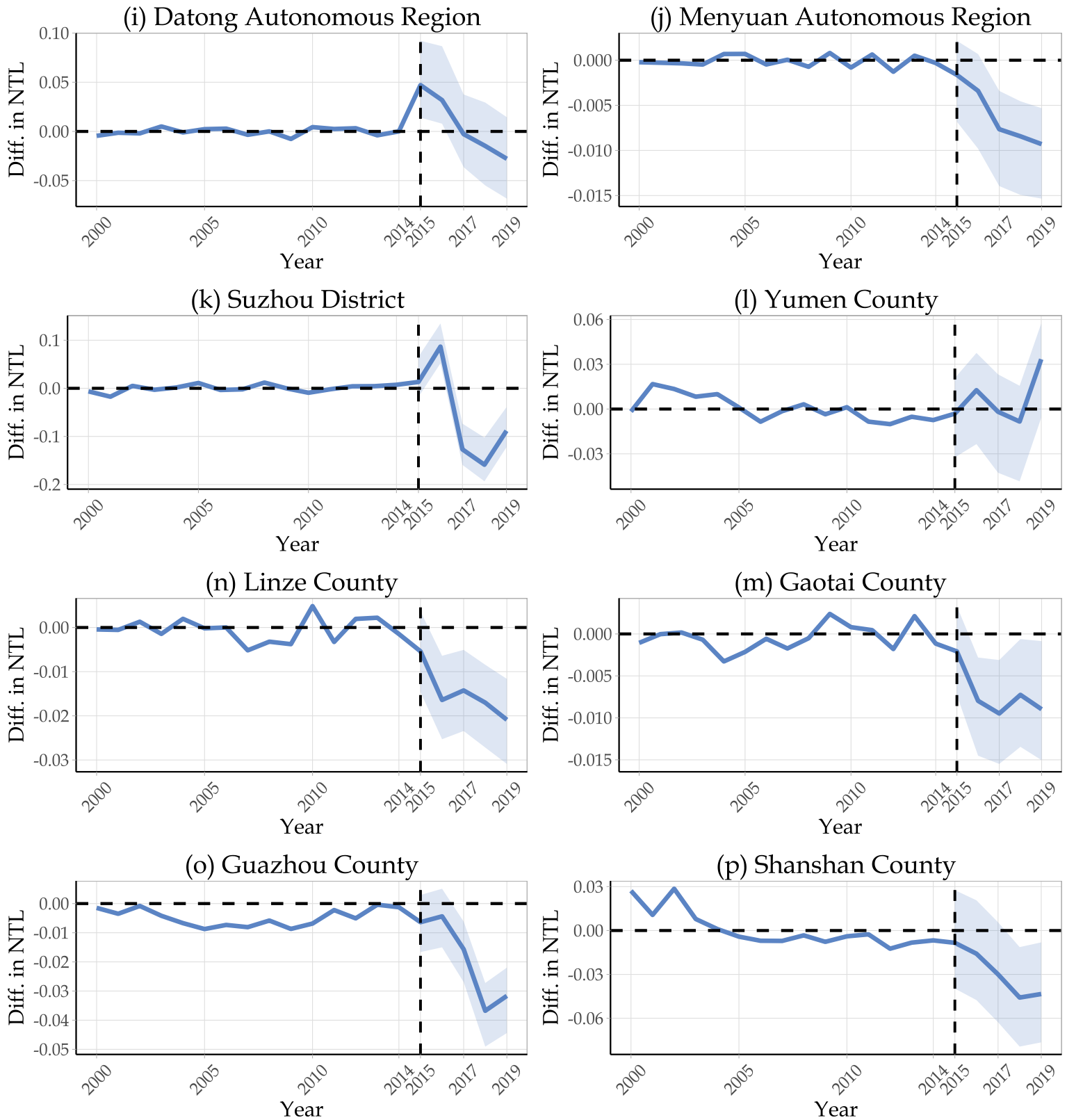


Figure 5: Estimated differences in nightlight intensity per pixel between treated and synthetic counties (continued)

Note: This figure plots point estimates for counties connected by the Lanzhou–Urumqi HSR. Shaded areas are 95% confidence intervals based on the Jackknife+ approach in Barber et al. (2021). The treatment date for the treated counties are either November or December 2014 and the treatment effect starts from 2015 represented by the vertical dashed line. Appendix Figure A7 shows the fit between treated and synthetic counties prior to treatment.

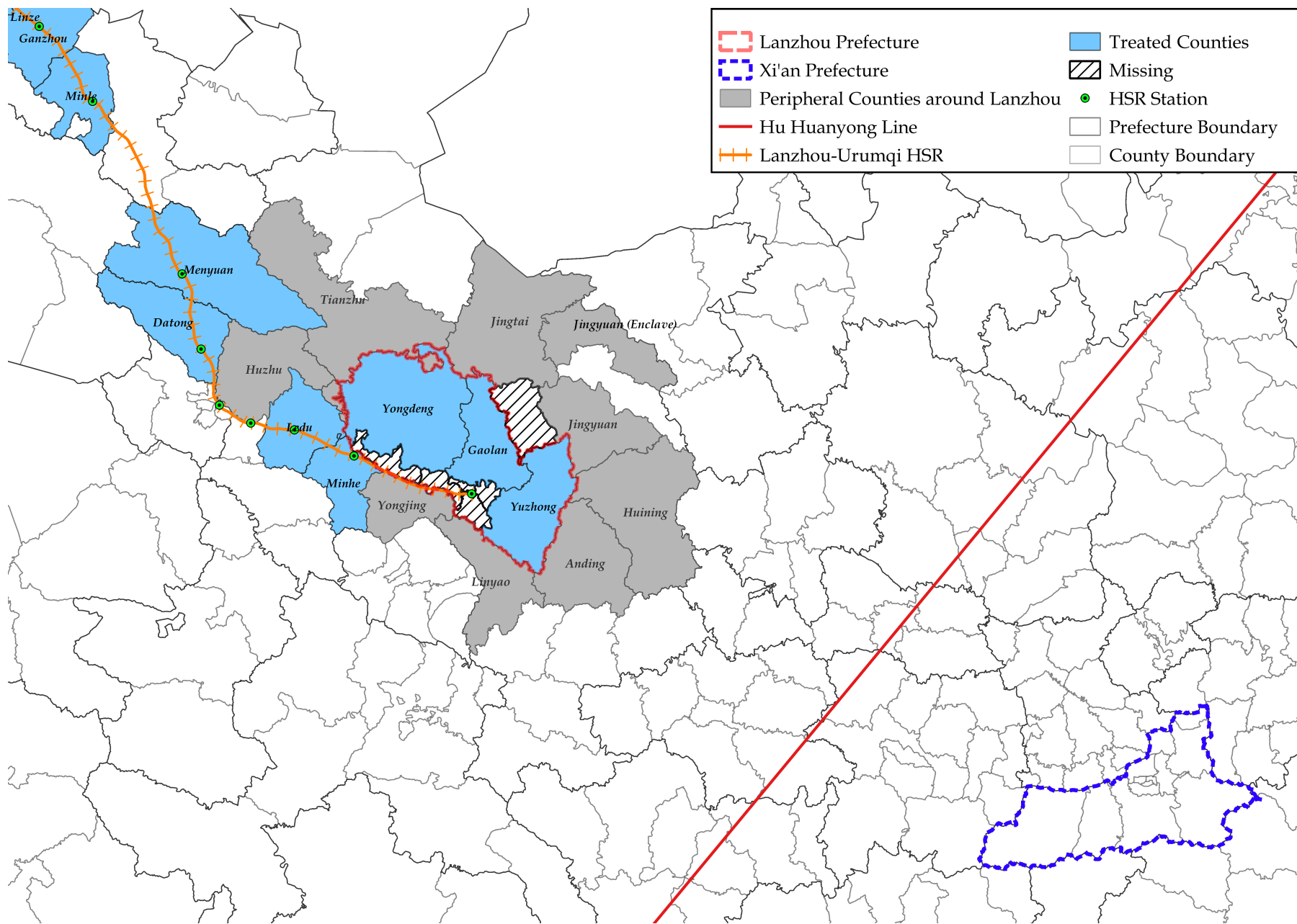


Figure 6: Location of peripheral counties around Lanzhou prefecture

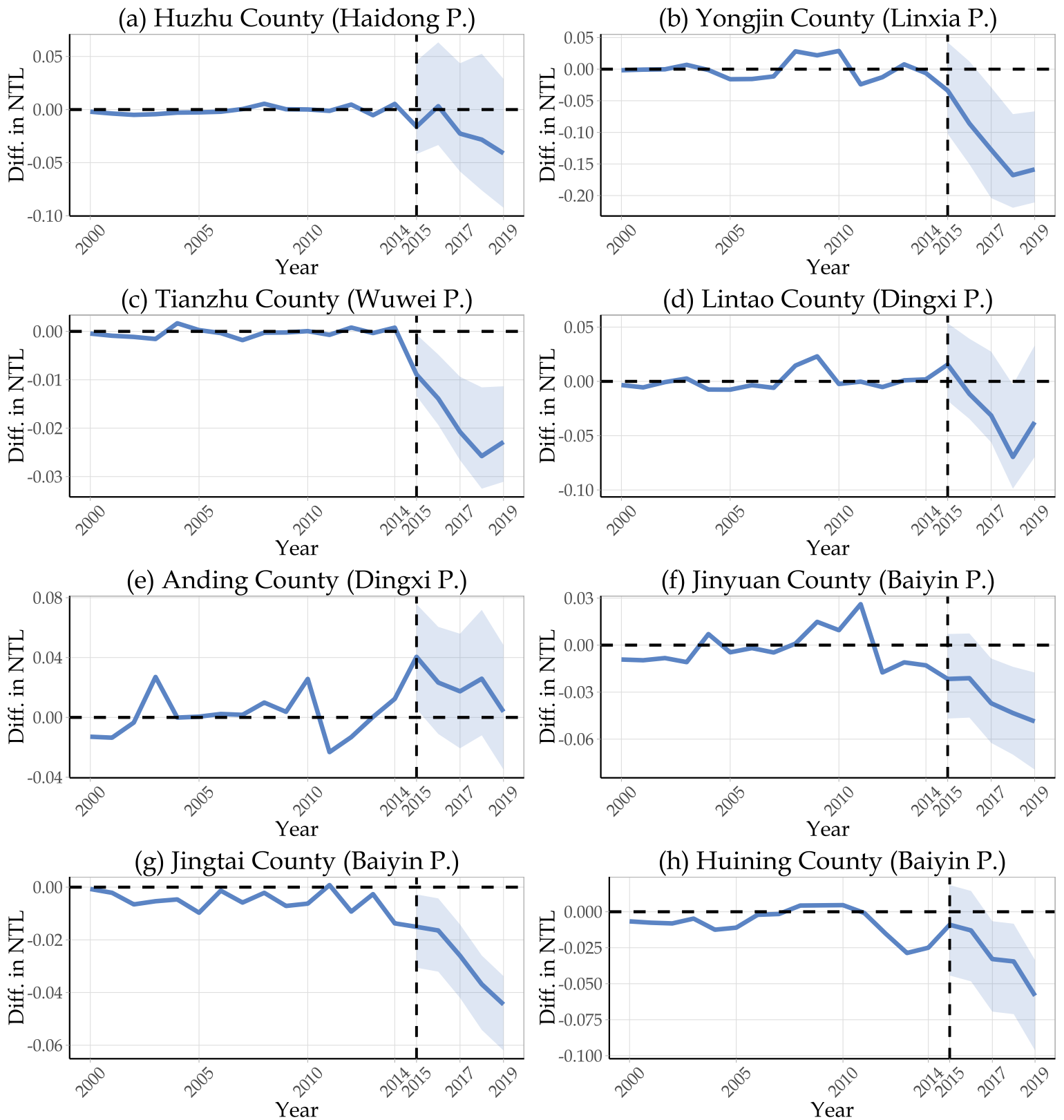


Figure 7: Estimated differences in nightlight intensity for peripheral counties around Lanzhou prefecture

Note: This figure plots point estimates for eight peripheral counties around Lanzhou prefecture including Huzhu county (within Haidong prefecture), Yongjin county (within Linxia prefecture), Tianzhu county (within Wuwei prefecture), Jinyuan and its enclave and Jingtai and Huining counties (within Baiying prefecture). Shaded areas are 95% confidence intervals based on the Jackknife+ approach in Barber et al. (2021).

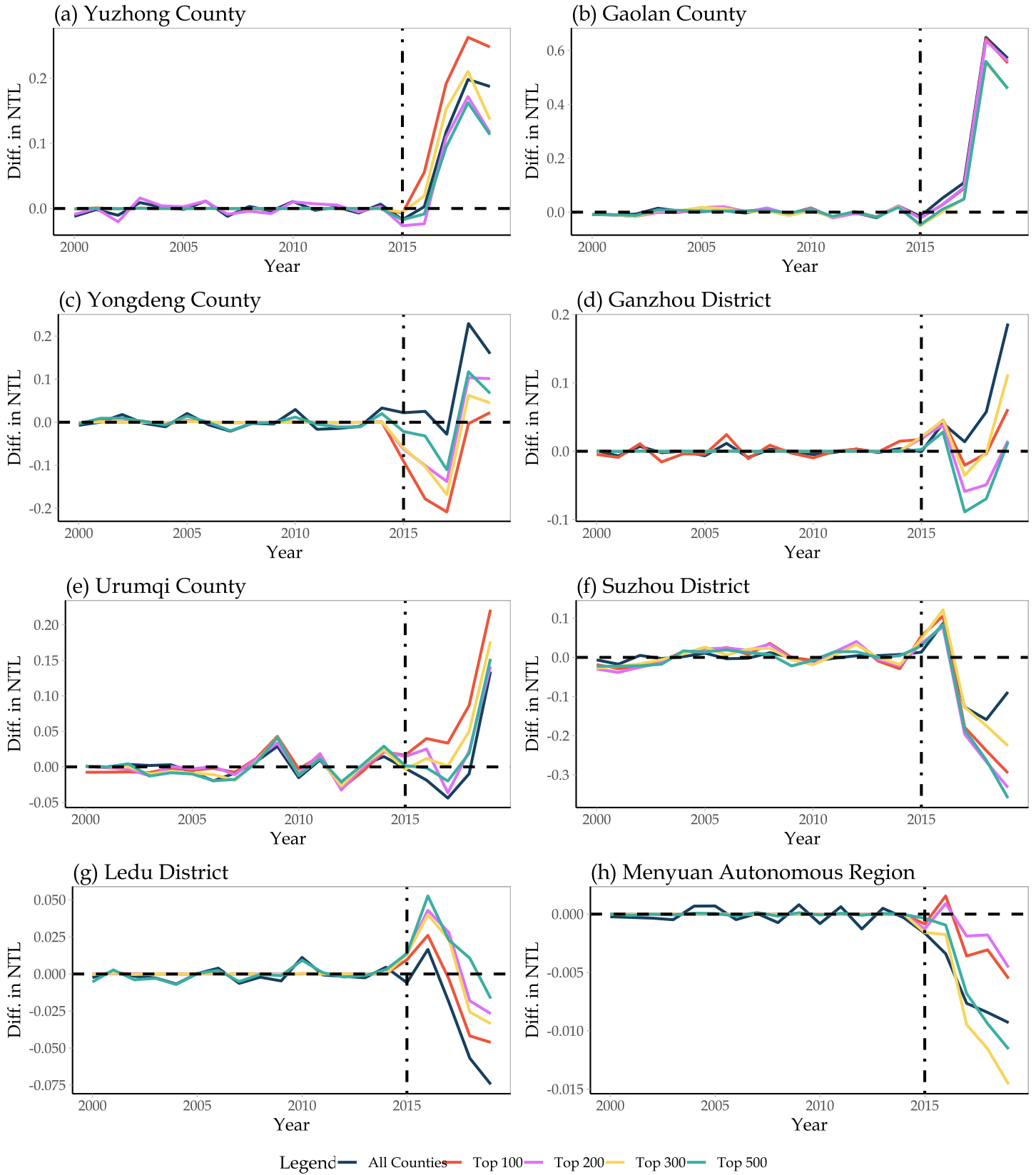


Figure 8: Placebo tests based on sorting standard deviation

Note: 'Top 100-500' represent groups of counties with the lowest standard deviation to reflect the sensitivity of the baseline ASCM estimates. For example, 'Top 100' means the first 100 counties ranked by descending order of standard deviations as the composite donor pool.

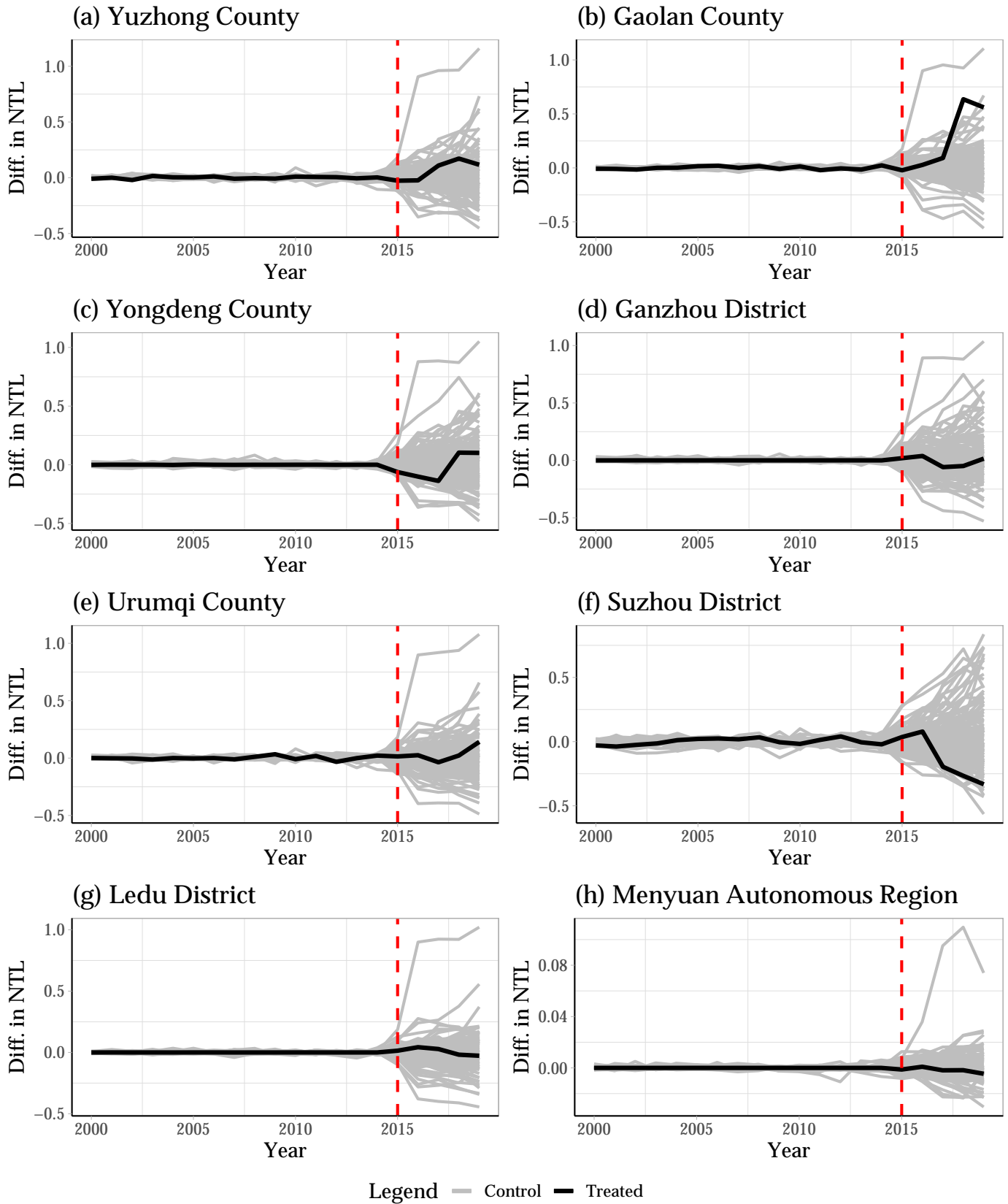


Figure 9: Leave-one-out placebo tests (ASCM)

Note: The leave-one-out placebo tests follow the approach proposed by [Abadie et al. \(2015\)](#). We used the first 200 counties ranked by descending order of standard deviations as the donor pool.

Online Appendix (Not for Publication)

This page is intentionally left blank.

Appendix

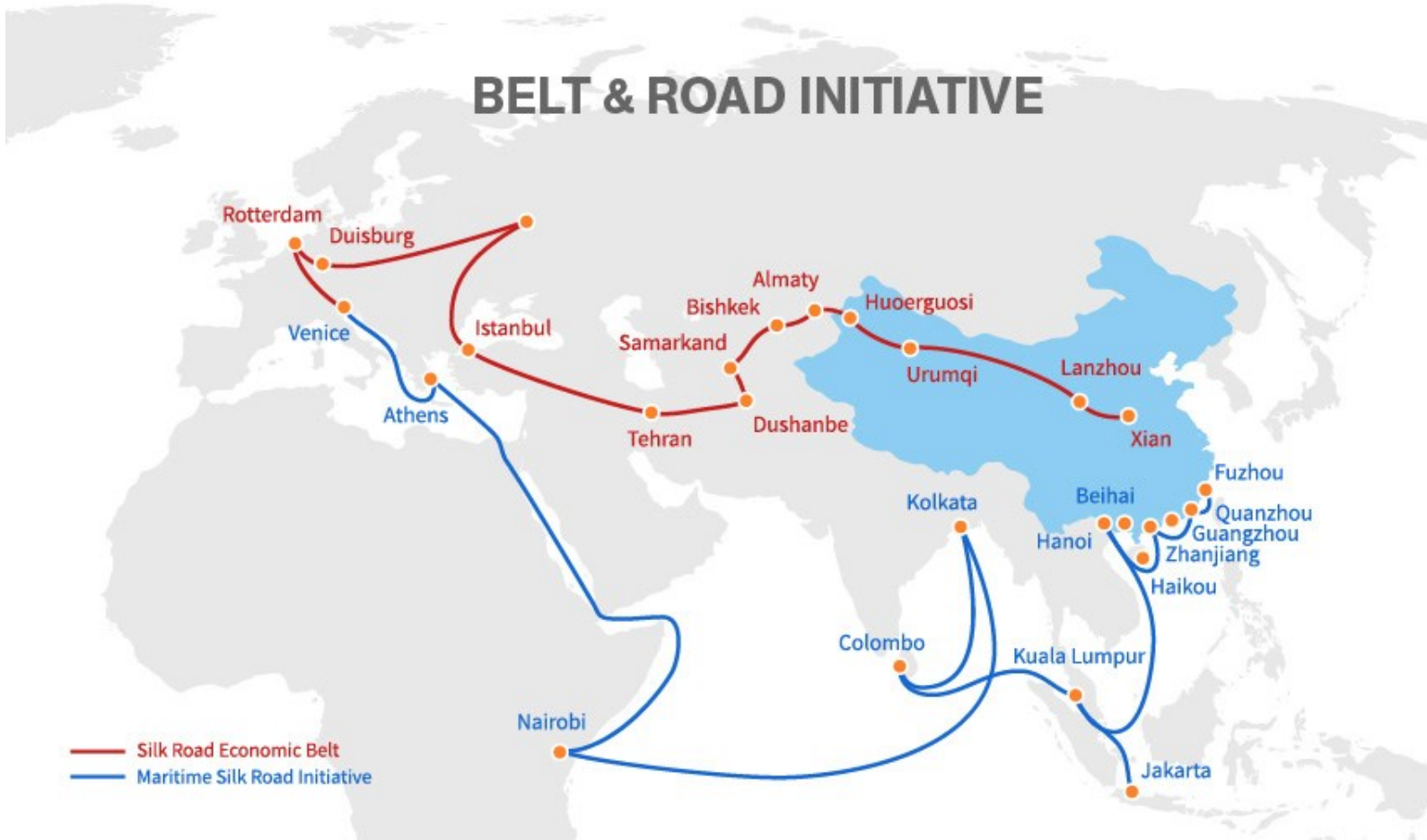


Figure A1: 'Belt and Road Initiative' Trade Route

Note: This figure shows the 'Belt and Road Initiative' that comprises several overland and maritime trade routes (Source: Silk Road Briefing retrieved from <https://www.silkroadbriefing.com/the-belt-and-road-initiative.html>).

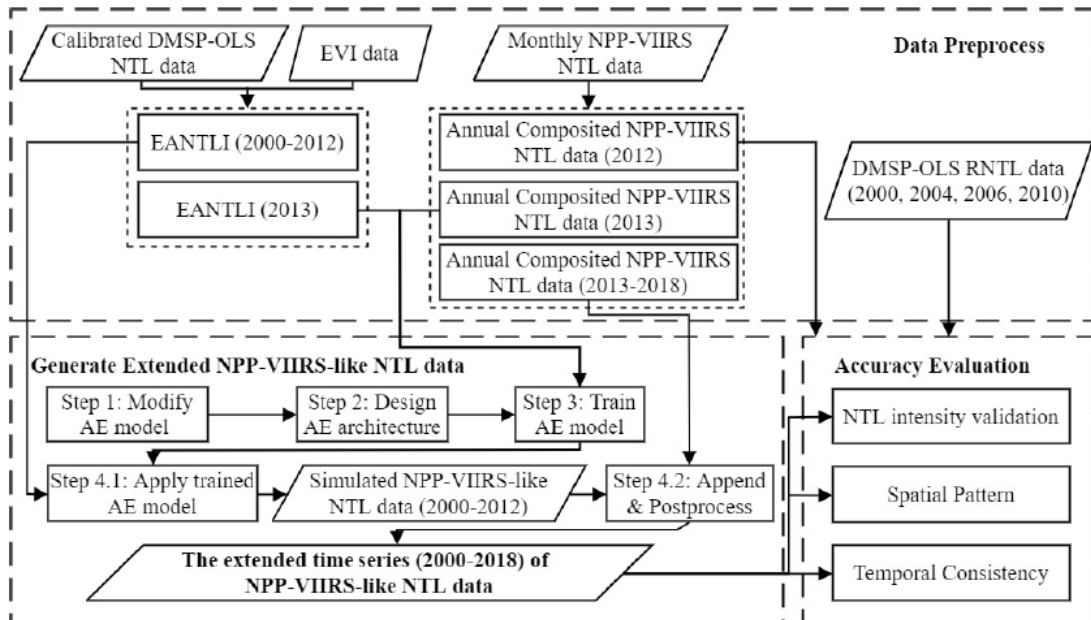


Figure A2: Process for joining nightlight datasets via machine-learning predictions (source: [Chen et al., 2021](#))

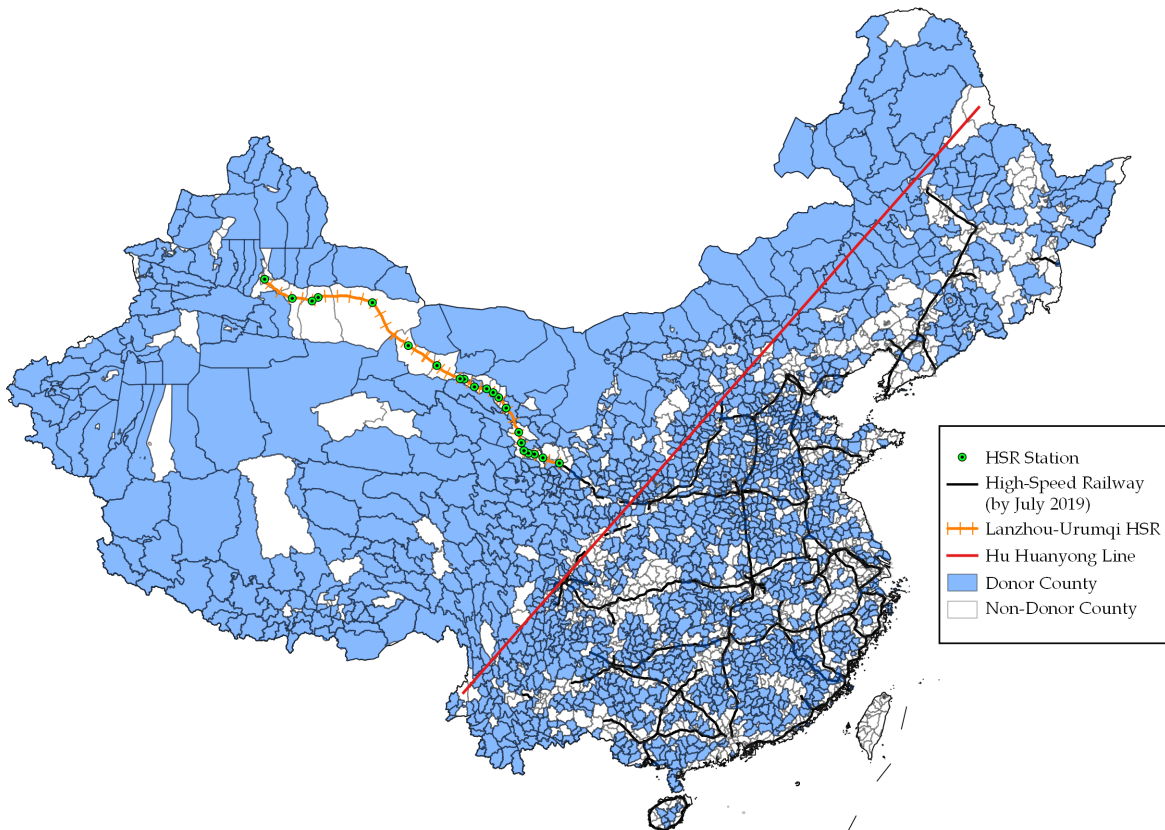


Figure A3: Geographical distribution of the donor pool

Note: This figure plots the distribution of the donor counties. We excluded counties connected to HSR between Jan 2015 and Nov 2019 and further excluded counties that experienced administrative (border) changes. The resulting donor pool comprises of 1,504 counties. Unlike metropolitan cities that have been connected multiple times, the Lanzhou-Urumqi HSR is the only HSR constructed to the west of the Hu Line, helping connect China’s northwestern and eastern provinces.

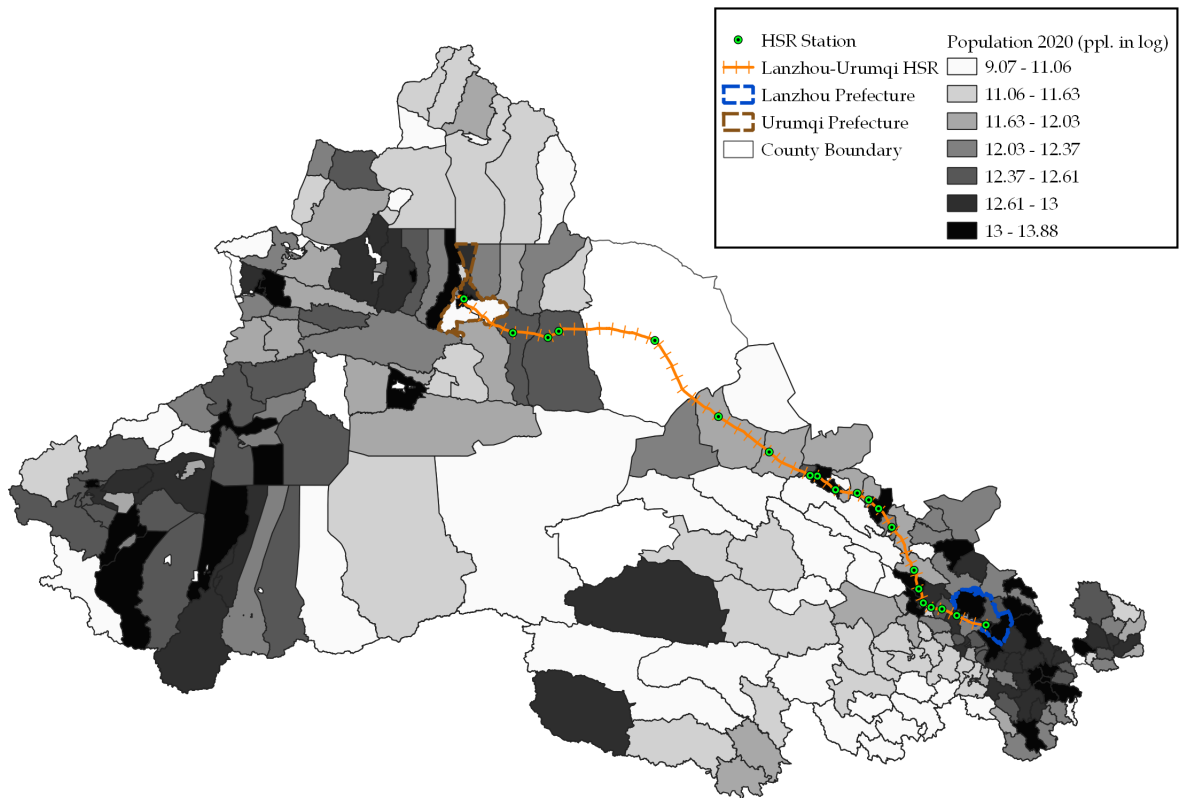
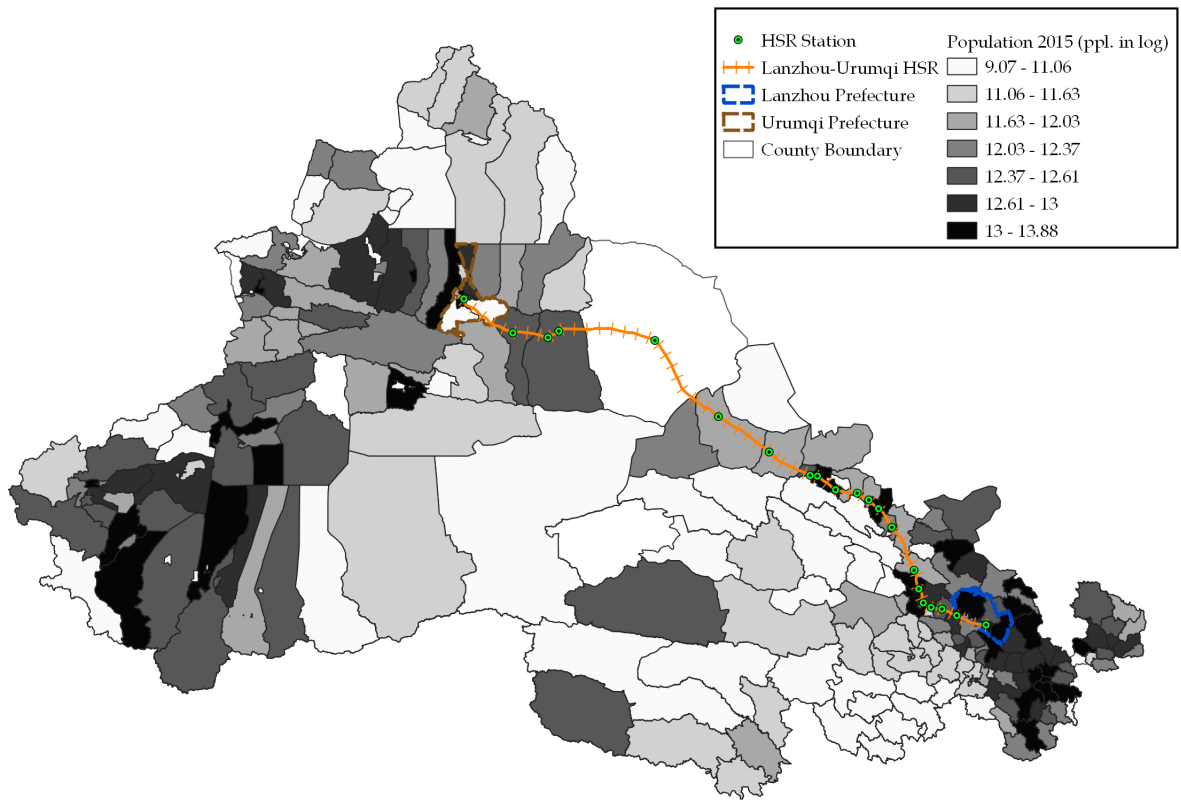
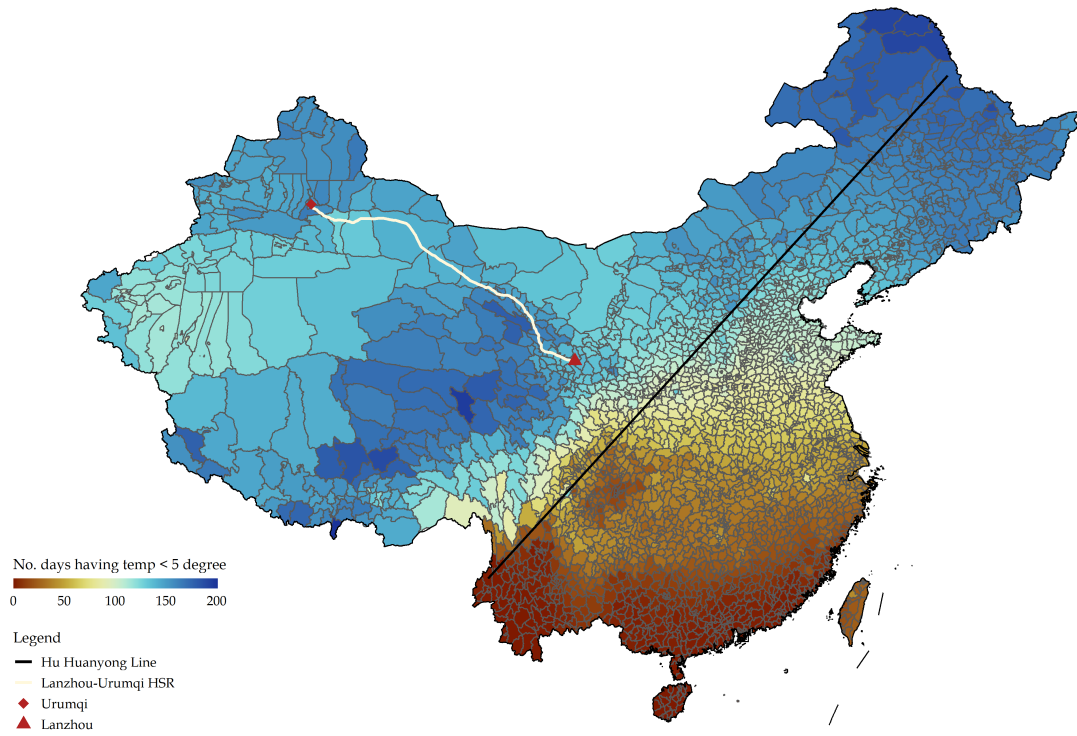
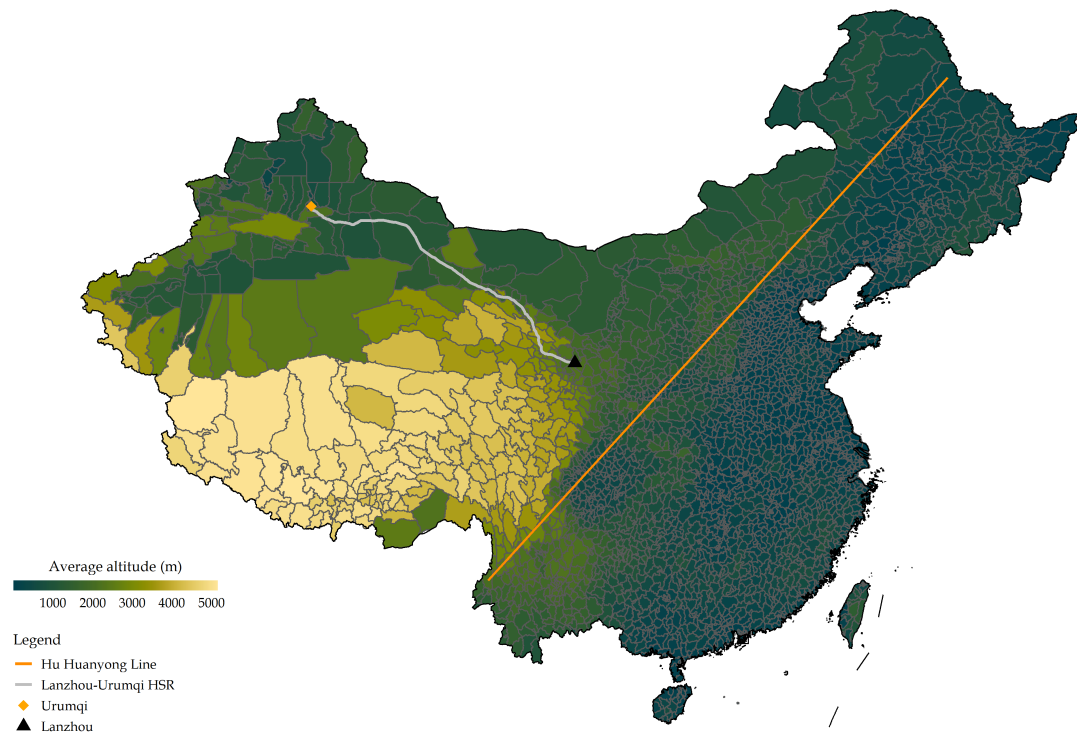


Figure A4: Population density (2015 and 2020)



(a) Coldness (no. of days $\leq 5^{\circ}\text{C}$)



(b) Average altitude (metres)

Figure A5: Habitability in China

Note: This figure shows habitability for residents in China. Panel (a) plots average altitude for each county in China. Panel (b) represents the average number of days below 5°C across 2000 to 2018.

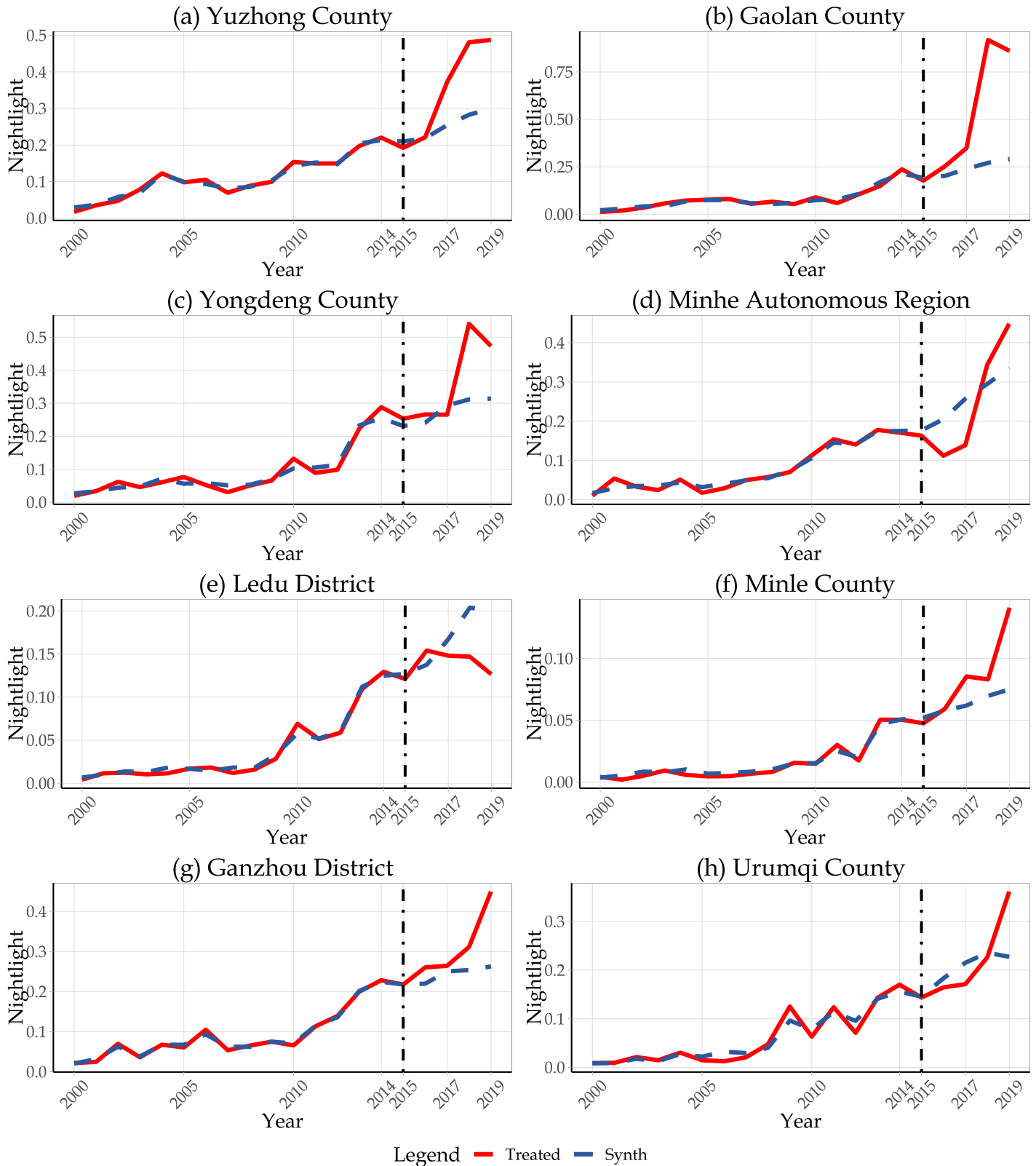


Figure A6: ASCM results for treated and synthetic counties along the Lanzhou-Urumqi HSR

Note: This figure plots the trends for observed nightlight per pixel on treated counties compared with those of synthetic counterfactuals using ridge ASCM by including the covariates in Table 2 and the averaged outcome from 2000-2013 as a predictor. Results are robust to inclusions of more years of lagged outcomes or exclusion of the averaged outcome.

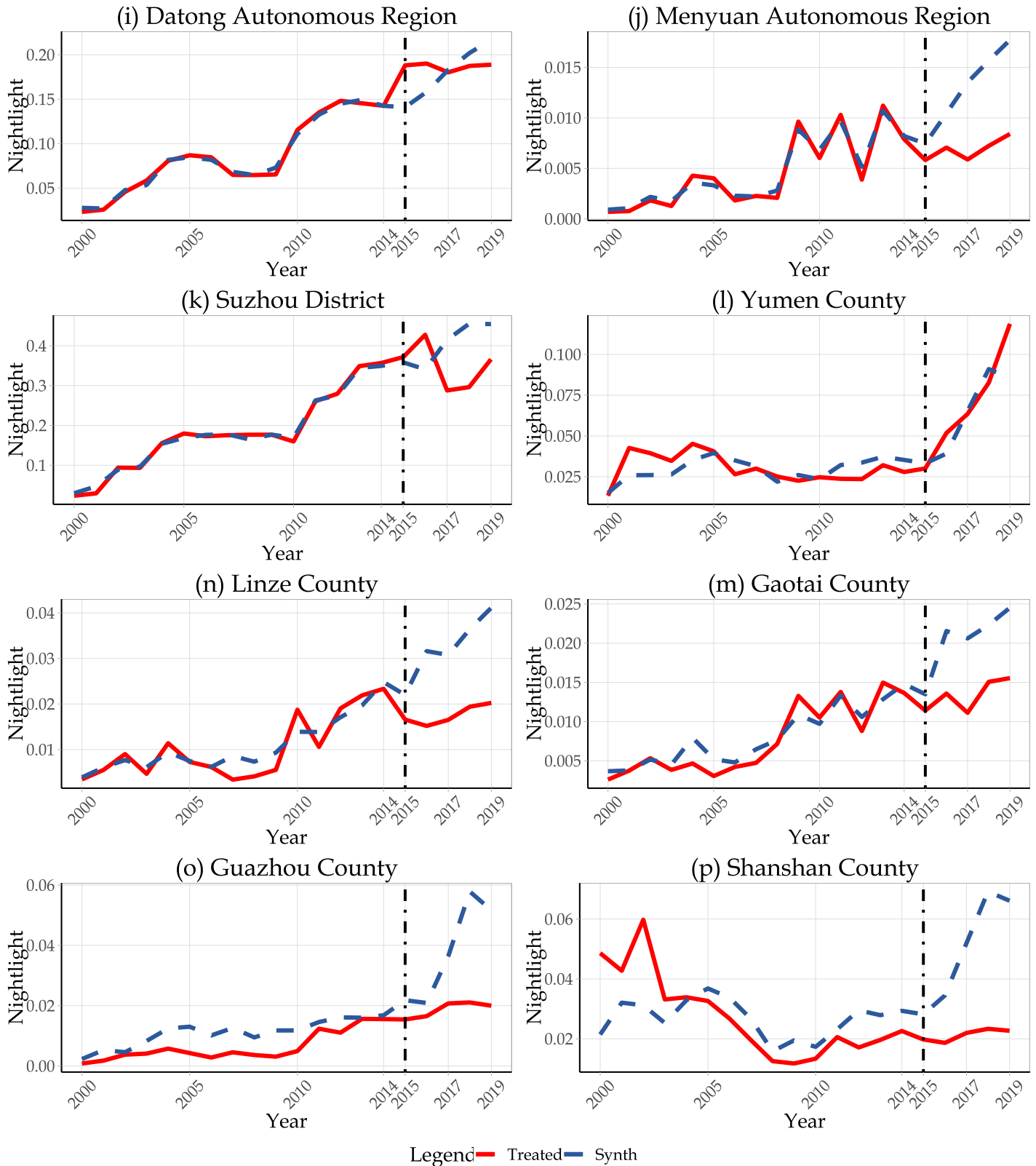


Figure A7: ASCM results for treated and synthetic counties along the Lanzhou-Urumqi HSR (continued)

Note: This figure plots trend of observed nightlight per pixel on treated counties against that of synthetic counterfactuals using Ridge ASCM by including covariates in Table 2 and the averaged outcome from 2000-2013 as a predictor. Results are robust to inclusions of more years of lagged outcomes or exclusion of the averaged outcome.

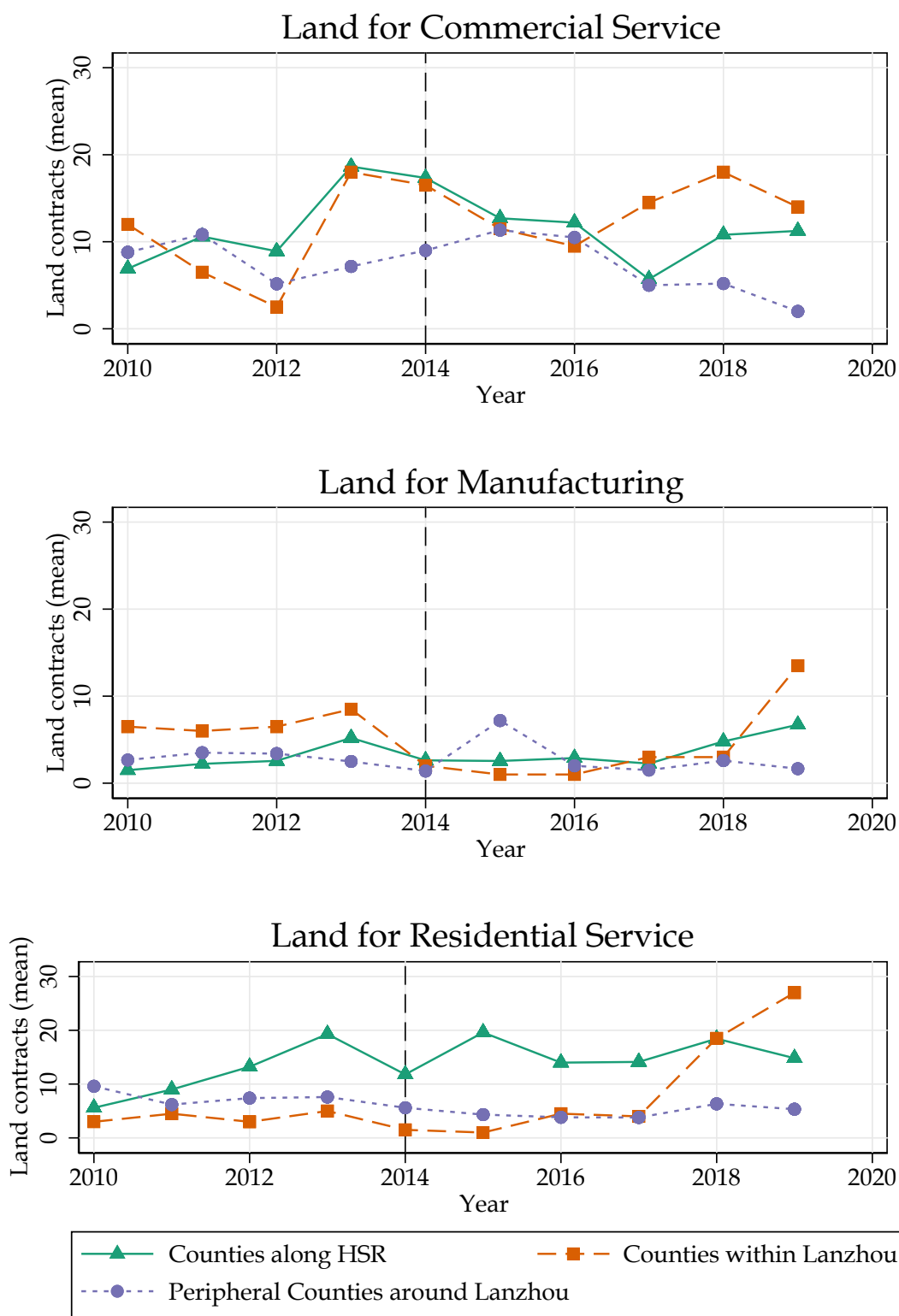


Figure A8: Land transactions along the Lanzhou-Urumqi HSR

Note: This figure compares trends of changes on averaged land transactions for three groups of interest. The first group contains three counties within Lanzhou; the second group involves the fourteen Lanzhou-Urumqi HSR connected counties along the northwestern Silk Road land route; the third group comprises eight peripheral counties around Lanzhou prefecture. The land transaction data for each county are collapsed by group and year.