Draft Version: Jun.26, 2017

Experiment to find the right thing or experimenting the right things? Evidence Revisited on China's "Opening-up" City Pilots

Shaowei Chen and Ciqi Mei

School of Public Policy and Management
Tsinghua University
Beijing, China

Comments and suggestions are welcome

Please correspond to:
Dr. Ciqi Mei
Associate Professor
School of Public Policy and Management
Tsinghua University
Beijing, China
cmei@tsinghua.edu.cn

(This is a preliminary draft. Please do not cite without authors' consent!)

Experiment to find the right thing or experimenting the right things? Evidence Revisited on China's "Opening-up" City Pilots

Shaowei Chen and Ciqi Mei
School of Public Policy and Management
Tsinghua University, Beijing

The puzzle of "China must have done something right" (Rawski, 1999; Woo, 1999; Guo and Hickey, 2010; Rothstein, 2015) presents a special edition for policy scientists. That is, how could so many major policies changes have been brought or "smuggled", -to use the Lindblom's famous term-, into China in the absence of major *ex ante* institutional transformation? Just to name a few, China has managed to decollectivize agricultural sector without privatizing farmland (Chung, 2000), to open up to foreign investors without property rights protection, to free the market price in the presence of dual-track system (Naughton, 1995). Even more interestingly, it seems that series of policy changes in China not only ended up with anticipated policy outcomes but also brought about, in an incremental fashion, unintended but desirable institutional transformation like property right protection, marketization and stabilized central-local relations, etc. From the perspective of endogenous institutional change, the most intriguing part of China puzzle is therefore whether China has found, or at least evidenced, a new way to bring about institutional transformation through incremental policy changes?

Against the "preoccupation of with optimal program design"[1] embodied by the reform practice in former Soviet Union, the China scholarship has long noticed the incrementalism in China's institutional transformation, specifically focusing on a series of feedback loops of reform, i.e. "how reforms begets further reform," as put by Barry Naughton (1995, p.320). Echoing the large literature on endogenous institutional change, such scholarship is especially keen to the role of self-reinforcing sequencing (Pierson, 2000; Grief and Laitin, 2004) in pushing forward institutional reconfiguration[2]. To substantiate this sequencing argument, several scholars have paid attention to how local experimentation has initiated virtuous feedback loops that eventually lead to scale-up of sound policy programs at the national level and break the path for institutional transformation (Naughton, 1995; Rawski, 1999; Heilmann, 2008a, 2008b; Xu, 2011; Ang, 2016). The underlying logic of such experimental logic is rather straightforward: for pressing policy issues, various local policy pilots experimenting different policy alternatives could identify novel policy solution which could be learned from by or forced upon other localities (Heilmann, 2008a; Zhu, 2014); informal institution, which provides the needed flexibility for policy experiment, would eventually adapt itself to accommodate then local experiments and now nationwide policy practice (Tsai, 2006; Ang, 2016).

This paper takes issues with such experiment-centered proposition on China's policy change. Our challenges stems from two angles. For one, the presumptive relation between success of an early policy experiment and its later diffusion is suspicious. While the experiment-based logic maintains

[1] Rawski,1999

[2] For a thorough literature review on endogenous institutional change related to China study, see Kellee Tsai (2006)

that the role of decentralized policy experiments is to seek novel solutions to pressing policy issues, an underlying assumption is that "positive results" of a policy experiment (Heilmann, 2008b, p.10) is a necessary, even if not sufficient, condition to for its later nationwide adoption. However, innovation literature has long warned about a pro-innovation bias, i.e. the assumption of "innovation benefits its adopters" (Abrahamson, 1991). A few policy literature has also evidenced that policy innovations, even failed in early experiments might nevertheless diffuse because of factors other than policy results (Park and Berry, 2013). In fact, as policy experiments are sometimes conducted under institutional uncertainty, the criteria to assess policy experimentation could be volatile by itself (Hall, 2010). In short, to start a self-reinforcing sequence leading to policy diffusion and institutional transformation, positive result of policy experiment might be neither sufficient nor necessary. For the other, the experiment-centered proposition has relatively downplayed the role of policy actors in the process of policy making and institutional transformation. While its proponents clearly point out the importance of patrons at the top for a policy pilot to scale up, evidence of success of policy experimentation is considered by them as more crucial. Such downplay of policy actors is abnormal for an authoritarian government where top-decision makers could forcefully put forward its policy agenda with selective evidence. Previous study has already challenges the latitude allowed for local government in policy experimentation and shows the authoritarian center's control on experimental variables (Mei and Liu, 2014). A closer look on how top decision makers have interpreted and utilized evidence of experimentresult is therefore needed to better understand the process from local experiment to nationwide policy change.

In this paper, we choose China's opening-up city policy since late 1970s as a case to show that the success of early opening-up pilot is not a necessary condition for its later scale-up at the national level; the presumptive causal relation between local policy experiment and policy change was false. Instead, local policy pilot, its later scale-up and policy change could primarily be outcomes of paramount leaders' construction of new type of means-end relation. As opening-up policy is cited as a prominent example for China's experiment-centered incrementalism, we use it as "the most likely case" to illustrate why the role of policy experimentation should not be overestimated when interpreting China's reform experience. Specifically, to examine the result of early opening up experiments, we employ the Synthetic Control Method (SCM) to compare economic performance of opening-pilots and their counterfactual controls. Empirically, the comparison doesn't find a significant advantage for opening-up pilots to their counterparts. With the help of historical archival research, we are able to show that policy environment has shifted by the top decision makers and hence favored the opening-up policy which eventually led its diffusion to a much larger scale since 1984.

We proceeds as follows. In the next section, we link the literature of policy experiments, policy diffusion and institutional transformation to discuss how local policy experimentation can and cannot contribute to policy changes. We then critique the current reform discourse on how opening-up policy is introduced into China's economy through city pilots and consecutive scale-up. In section 3, we briefly introduce the SCM method and show our empirical evidence on performance discrepancy (the absence thereof) between coastal cites piloted Opening-up polices and those not. Result of robustness check is also presented. Historical archival research in section 4 recapitulate the policy process of opening-up from 1978 to 1984 and shows paramount leaders' decision and

interaction, rather than the evidence of pilots' success, has determined the diffusion of opening-up policy. We conclude by discussing the implication of our study on the understanding of China's transformation and the function and limitation of policy experimentation.

1. Local policy experimentation and Policy changes

What could local policy experimentation do in the process of policy change? The aforementioned experiment-centered proposition has stressed on the prospect that policy pilots with positive result engendered a self-reinforcing sequencing which leads to policy change. Specifically, the self-reinforcing sequencing goes in two different ways. For one, as policy diffusion literature has long discovered, policy programs innovated in certain locality might go across the *geographic* boundaries through voluntary learning; as a result, the more localities adopting a policy innovation, the more likely others will follow suit because of pressures from competition or peers (Shipan and Volden, 2008). For the other, while local policy experimentation requires institutional flexibility to survive, replication of successful local policy experiments might reciprocally facilitate institutional transformation in need (Tsai, 2006); in other words, successful local policy experimentation might cross the *institutional* barriers too. In China, we do observe, for example, while the private sector has grown rapidly under the wings of informal institution protecting property rights (e.g. "red hat"), formal institution like the Property Right Law enacted in 2007 follows the growing-up of the private sector (Nee, 2015).

Criticism on such proposition could come from two strands of literature, i.e. the one on policy innovation and the one on institutional transformation. While most policy innovation research assumes implicitly that policy later diffused should be successful in where is innovated, many still work to find out other mechanism for policy innovation to diffused. Shipan and Volden (2008) identified four different mechanisms of diffusion, i.e. learning, competition, imitation and coercion. Among these four mechanisms, learning is more based upon the "proved success" (Berry and Baybeck, 2005) of a policy alternative, whereas policy results are less important for other mechanism. Besides the actual results of policy pilots, fruitful empirical research have identified many other factors accounting for policy diffusion. Zhu and Zhang (2016) even find out, for a same policy innovation with early successful pilots in China, different mechanism exists to account for its diffusion when policy environments changed. In short, as policy diffusion more likely than not starts from a successful local policy innovation (Volden, 2006), the policy innovation literature in no way support the idea that policy pilots with positive outcome could by themselves initiate a self-reinforcing sequencing leading to sound policy changes.

Furthermore, some other policy innovation literature eschewed from the idea that success of policy pilots is even relevant for policy diffusion. Policy programs born to the so-called New Public Management (NPM) Movement have attracted quite some criticism of this type. Christian Hood, in his 1991 review of NPM movement, has already warned the danger of "commodity cult." That is, while many policy programs adopted in NPM movement is borrowed from private sector, they could be borrowed without even thinking about why they have worked in private sector and whether they would work in public management. In this line, Park and Perry (2013) show that performance management programs were adopted by American governments even when its failure or at least

nonsuccess has been observed. Moynihan (2006) also observes that "managing for result" might be adopted for symbolic reasons in American states. Similar symbolic innovation/diffusion are also pervasive in China where pressure to conform is usually higher in its authoritarian setting (Teets, 2015). Beyond the public sector, Abrahamson (1991) was among the first to account for the irrelevance between results of innovation and its diffusion. According to him, when imitation process impels the diffusion or rejection, fads and fashions could better explain innovation diffusion.

To link to the literature on institutional transformation, we could better understand why fads and fashions, instead of policy results, affect policy change in a more fundamental way. It's worth noting that all policy programs have to function in certain institutional context. To explain institutional transformation or policy change separately is already hard; it is even hard to explain both at the same time. As Peter Hall mentioned, "Institutional stability is a fundamental issue for analyses of institutional change" (2010, 207). Specifically for policy changes, one thing crucial that could be provided by institutional stability is a shared belief on what means are available in the choice set and what ends for policies are, respectively. However, as Lindblom has figured out, means and ends could both be uncertain and subjective to policy actors' choice (1959, p83) in the absence of such shared belief or institutional stability. Uncertainty in the means and ends of policies apparently has direct impact on policy actors. Without a well-defined choice set, try-and-error type of experimentation would be pointless if not impossible; without a common-accepted goal, assessment on results of policy experimentation would be totally a subjective matter. In this scenario, following fashions and fads reduces uncertainty for the good of policy actors. When policy misbehaviors under uncertainty could at times cost one's career or even personal life in China, to follow fashions and fads, instead of looking into policy results, is even more crucial. A pair of examples regarding China's agricultural policy is quite telling. During the Great Leap Forward, local leaders have followed suit with each other to over-report crops yield while knowing its catastrophic consequences for local residents (Kung and Chen, 2012). In contrast, knowing decollectivization helped agricultural outputs, local leaders chose to "wait and see" when pioneers' early success was observed but consensus among top leaders was yet to come (Chung, 2000). In both cases, the ostensible policy consequences, good or bad for local residents, did not really matter for policy actors to follow suit.

That said, we in no way suggest that results of policy pilots don't matter at all. According to Hall's classic conception of policy change of three orders, policy changes might happen at first-order for policy settings, second-order for policy instruments and third-order for policy paradigm. When means and ends are well defined, policy pilots could certainly serve as a tool for "normal policy making" (Hall, 1993) to bring about first and second order changes. In this sense, Heilmann's model of "experimentation under hierarchy" is suitable to explain how local policy experiments have initiated second-order policy change, i.e. "experimenting units try out a variety of methods and processes to find imaginative solutions to predefined task" (Heilmann, 2008b, p.3). And in a more conservative version, Xu (2011) points out the role of policy experimentation in first-order policy change, i.e. to reduce technical noise by comparing different configurations of policy instruments.

In contrast, results of policy experimentation should matter less for the third order policy change, or policy change during paradigms shift. Result of a local policy practice is local by nature in the

sense that certain idiosyncratic local institutional settings have enabled and tolerated such experiment; otherwise the practice should have been institutionalized in other places (Berman, 2012). During the paradigm shift period when previous common beliefs on means and ends were vehemently challenged, opportunities appear for the previously *local* practice, if interpreted as appropriate means for a newly accepted ends, to migrate into other localities. In this process, interpretation of a local policy experiment in order to reconstruct an acceptable means-end relation is more important than the policy experiment's outcome engendered in certain locality. However, as interpretation is a subjective matter, competition for the right to build a new means-end narrative would be political by nature.

To sum up, the experiment-centered account explaining China's drastic policy change during the reform era could be wrong. Both policy innovation/diffusion literature and policy change literature suggest that positive results of local policy experimentation are unlikely the cause of a self-reinforcing sequencing which eventually leads to the policy change observed in China. Instead, successful reconstruction of means-end narratives has simultaneously facilitated diffusion of local policy experiment and institutional transformation to accommodate its being scaled-up.

2. Opening-up City Pilots in Coastal China

This paper chooses opening-up as the case to study the role of policy pilots in policy changes. Apparently, the concept of openness of China's economy is more inclusive than opening-up of cities. However, China's openness today does see its origin in a few piloting cities. Shenzhen, for example, now the fourth largest city in China and famous globally for its manufacturing and high tech industry, used to be a small costal market town adjacent to Kowloon of Hong Kong with a population of 30,000. Dramatic changes was brought in after it was chosen in May 1980 as one of the four first-batch Special Economic Zones (SEZs) together with Zhuhai, Shantou, and Xiamen. Until today, Shenzhen is still viewed as the flagship and role model of China's Opening-up. Multiple policy programs relating with foreign trade and foreign investment were first tried here before they migrate into the inner land.

A typical narrative of China's opening-up usually started from how the success of SEZs led its scale-up to the national level. As mentioned above, 4 SEZs was chosen in May 1980[3]; In May 1984, 14 coastal cities from north to south were designated as "coastal open city."[4] In less than 10 months, other 11 cities[5] was added to the pilot list as State Council decided to open up Yangtz Delta, Pearl Delta and South Fujian Delta in Feb. 1985. In Early 1988, East Liaoning Peninsula and Shandong Peninsula was opened up and 5 more cities[6] was included. In April 1988, Hainan Island became a separate province and the first and only SEZ at the provincial level. In 1992, following Deng's unexpected South Tour aiming to pick up the reform momentum after the post-Tiananmen retrenchment, the central government went one step further to open up more inner land cities, first

[3] Hainan province was designated as the fifth SEZ in 1988.

[4] Dalian, Qinhuangdao, Tianjin, Yantan, Qingdao, Lianyungang, Nantong, Shanghai, Ningbo, Wenzhou, Fuzhou, Guangzhou, Zhanjiang, Beihai

[5] Suzhou, Wuxi, Changzhou, Jiaxing, Huzhou, Quanzhou, Zhangzhou, Foshan, Jiangmen, Zhongshan

[6] Dandong, Yingkou, Weifang, Weihai, Rizhao.

those along Yangtze River, then provincial capital cities, and then cities along the landline borders. Official documents has dated 1993 as the year of the emergence of China's "multi-directional, multi-level and wide-ranging opening up."

Such linear depiction on the emergence of China's opening-up policy has left many important questions unanswered. First, how could the pilot city to even start its opening-up experiments? A typical experiment-centered answer would be that opening-up policy program was one of many policy alternatives which were experimented. But how could opening-up city emerge in the choice set in the first place? Second, did the performance of opening-up pilots really matter for its later scale-up? While some pilots like Shenzhen have indeed presented a glaring grade sheet in economic performance, some others were less impressive. For example, Hainan as the largest SE has not outperformed the national average and was often headlined for overheating and inefficient investment in the 1990s. Third, probably more importantly, when the old means-end narratives still dominated and the positive results accompanied by the new means-end relation was yet to come, how could proponents of opening-up weather through the attacks from its opponents? While a retrospective account could easily identify benefit of opening-up for China's development, such means-end relation was not so obvious in the policy environment when opening-up was introduced. Negative consequences observed in SEZs, e.g. rampant smuggling, deteriorating public safety, less obedient youth, etc., have provided abundant ammunitions for the conservatives sticking to the old means-end narratives. So, before enormous positive results in economic performance could establish a new means-end narratives, how could proponents of opening-up defend themselves?

We answer these questions in the next two sections. By comparing the economic performance of opening-up pilots and non-pilots with SCM, we show that performance gaps between pilots and its counterfactual control did not prove advantage of opening-up in a convincing way. Historical review in section 4 shows the process how change in policy environment has helped engender the opening-up pilots, how different political groups have competed to interpret pilots' result and how the new means-end narratives were constructed.

3. Comparing the Economic Performance of Opening-up Pilots and Non Pilots

Estimating economic effects of the opening-up policy on the experimental pilot cities is difficult for two reasons. Firstly, all the pilot cities are located in the coastal regions that has a geographically higher potential for economic growth. Direct comparison with the non-pilot cities (most of which are inner land cities) are likely to produce severely upward-biased estimations. Secondly, those pilot cities were deliberately chosen by the central government. As stated in the official document: "Those cities were the richest regions in China…with highly developed transportation system as well as the most advanced technology and education level" (cite…). This also indicates a high risk of selection bias in identifying casual effects. What's more, since the selection mechanism are not known with complete certainty, we also confront the challenge of omitted variable bias/hidden bias (Guo & Fraser, 2014).

A natural solution is to rely on longitudinal data and adopt the widely used difference-in-difference (DID) approach to control for time-invariant unobservable variables. Nevertheless, exploiting the

DID approach requires a "parallel assumption", i.e., the average outcomes of treated and control units would have followed parallel paths in the absence of treatment (Xu, 2015), which is extremely difficult to be satisfied in our case. Furthermore, the omitted unobservable variables could be time-varying, which is out of the scope of DID.

This paper adopts synthetic control methods (SCM), a newly-developed comparative case studies approach, to overcome the identification problems outlined above. The basic idea of SCM is to construct a "synthetic control unit" as a counterfactual for the treated unit by reweighting the control units (Abadie, Diamond & Hainmueller, 2010, 2015; Xu, 2015). This approach not only makes the comparison between the treated and control units transparent, but also extends the fixed-effects DID model by allowing the effects of unobservable variables vary with time. Below we describe this approach in detail.

3.1 ***The Method: SCM***

SCM constructs a counterfactual state of treated units as a weighted average of mathematically selected control units. The central premise of SCM is that once we create a synthetic unit that matches the treated unit in respect of both the covariates and outcomes in pretreatment period, we construct an unbiased counterfactual of treated unit in the post-treatment period. To simplify the exposition, we introduce the method as if only one city were chosen as the pilot city.

Following Abadie et al. (2010), suppose that we observe the data of $J+1$ cities during the periods $t = 1 \ldots, T$. Without loss of generality, suppose also that the first city is chosen as the pilot city (i.e., treated by the opening-up policy) at time period $T_0 \in (1, T)$. Let Y_{It}^N be the GDP (outcome variable) for city i at time t in the absence of treatment, and Y_{It}^I be the GDP observed for treated city at the post-treatment periods $t \in (T_0 + 1, T]$. The observed GDP for city i at time t is

$$Y_{it} = Y_{1t}^N + \alpha_{it} D_i$$

Since only the first city was chosen as the pilot city, we have that

$$D_i = \begin{cases} 1, if \ \ i = 1 \ \text{and} \ \ t > T_0 \\ 0, \qquad \text{otherswise} \end{cases}$$

The treatment effects we aim to estimate is

$$\alpha_{1t} = Y_{1t}^I - Y_{It}^N = Y_{1t} - Y_{It}^N, t \in (T_0 + 1, T]$$

Suppose that Y_{It}^N, the post-treatment periods counterfactual we need to construct is given by a factor model

$$Y_{1t}^N = \partial_t + \theta_t \mathbb{Z}_i + \lambda_t \mu_i + \varepsilon_{it}$$

where ∂_t is an unknown common factor with constant factor loadings across cities, $\mathbb{Z}_i$ is a $(r \times$

1) vector of observed covariates (not affected by the opening-up policy), θ_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is a $(F \times 1)$ factor loadings, and the error term ε_{it} are unobserved transitory shocks at city level with zero mean.

Consider a $(J \times 1)$ vector of weights $\mathcal{W} = (w_2, \dots w_{J+1})$ such that $w_j \geq 0$ for $j = 2, \dots, J+1$ and $w_2 + \cdots w_{J+1} = 1$. Then we have

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \partial_t + \theta_t \sum_{j=2}^{J+1} w_j \mathbb{Z}_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}$$

Suppose that there exists $\mathcal{W}^* = (w_2^*, \dots, w_j^*)$ such that

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \dots, \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \text{and } \sum_{j=2}^{J+1} w_j^* \mathbb{Z}_j = \mathbb{Z}_1$$

Abadie et al. (2010) proved that, if $\sum_{j=2}^{J+1} \lambda_t' \lambda_t$ is nonsingular, under standard conditions, $Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ will be close to zero if the number of pre-intervention periods is large relative to the scale of transitory shocks. This suggests that $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ will be an unbiased counterfactual of Y_{1t}^N in the post-treatment period, even with the existence of time-varying unobserved confounders.

The optimal weights vector $\mathcal{W}^*$ is given by

$$\mathcal{W}^* = arg\min \sum_{m=1}^{k} v_m (\mathcal{X}_{1m} - \mathcal{X}_{0m} \mathcal{W})^2$$

where $\mathcal{X}_{1m}$ is the a $(m \times 1)$ vector of predictors for the treated city, $\mathcal{X}_{0m}$ is a $(m \times J)$ vector that contains the same variables for untreated cities. $\mathcal{V} = (v_1, \dots, v_m)$ is a $(m \times m)$ vector of weights on predictors given by

$$v_m = arg \min_{v \in \mathcal{V}} ((\mathrm{Y}_1 - \mathrm{Y}_0 \mathcal{W}^*(\mathcal{V}))' (\mathrm{Y}_1 - \mathrm{Y}_0 \mathcal{W}^*(\mathcal{V}))$$

With the technique described above, we estimate the casual effects of opening-up policy by constructing synthetic counterfactuals for each of the pilot cities. This approach offers a transparent way to choose suitable comparison cities and construct unbiased post-treatment counterfactuals with time-varying unobserved variables (given that the treated and control cities match well during the pre-treatment periods), thus effectively conquer the identification difficulties mentioned above.

The SCM does not allow assessment on the validity of inference by calculating frequentist *p*-values

used in large sample regression settings (Billmeier & Nannicini, 2013; Fremeth, Holburn, & Richter, 2016). A more common way of doing inference in SCM is placebo tests, one approach for testing whether estimated results in the post-treatment period are spurious (Bertrand et al. 2004). The basic idea is that to replicate the SCM analysis using a hypothetical treatment on untreated units should not generate meaningful treatment effects.

In this paper, we mainly adopt the "across-unit" placebo tests by randomly assign the hypothetical treatment to the non-pilot cities. This allows us to assess whether the treatment effects estimated by SCM for the pilot cities is large relative to the hypothetical treatment effects estimated for non-pilot cities. To further evaluate the treatment effects of pilot cities relative to effects obtained from the placebo runs, we calculate the ratios of post/pre-opening-up policy RMSPE (root mean square prediction error). This creates a distribution of placebo effects against which we can assess the treatment effects of pilot cities. This process allows us to compute *p*-values based on permutation inference in the population (Rosenbaum, 2002a, b).

3.2 **Data and Sample**

The main source used in this research is the "Comprehensive Statistical Data and Materials on 50 Years of New China", a series of official statistical yearbook published by the China's National Bureau of Statistics and several provincial governments. They contain data for the city level economic development that covers the 1949-1998 period (most of the data in the yearbooks starts with 1952). Data for outcome variable and predictors are collected beginning with 1952, and ending in 1991, the last year before the South Tour of Deng Xiaoping.

Since we concern the economic impact of opening-up policy on pilot cities, we choose GDP as the outcome variable. We include all the economic-achievement-related variables that are completely recorded in the yearbooks as predictors, including population, the second industry output, fixed assets investment, fiscal revenue, total retail sales of social consumption and GDP in the previous years.

Our treated units contain the 11 first batch costal open cities (Dalian and Beihai are excluded due to missing data; the GDP of Shanghai is always in top of China, thus can't not be synthesized by other cities) and the 10 cities in the open costal economic areas. As these cities were designated in 1984 and 1985 respectively, the treatment period is 1984 for the costal open cities and 1985 for the open costal economic areas. Our donor pool includes 51 cities whose data are available in the statistical yearbooks while all the cities that are treated with similar opening-up policy during the post-treatment periods are excluded (Abadie et al. 2015).

3.3 **Empirical Findings**

3.3.1 Results for Costal Open cities (1984)

Figure 1 displays the GDP trajectory of 11 costal open cities. For all the cities, their synthetic counterparts exactly reproduce the GDP for pre-treatment periods (1952-1991). The predictor balance results (see SI) confirms that the synthetic control matches the treated cities well in terms of the initial GDP and other predictors. Good match in the pre-treatments ensures that the

counterfactuals of treated cities in the post-treatment periods are well constructed.

[Figure 1 about here]

Our results show that effects of opening-up policy vary among the treated cities. Only 4 cities of 11 (Yantai, Guangzhou, Zhanjiang, Ningbo) have outperformed their synthetic counterparts in the post-treatment period, while others display non-significant positive discrepancy comparing with the synthetics. Cities like Tianjin, Qingdao and Nantong even registered lower economic performance after piloting opening-up policy. These results indicate that effects of opening-up policy are not as good as the reformers have argued.

[Figure 2 about here]

We further implement the "across-units" placebo tests to check if the positive effects displayed in the four cities are spurious. We perform hypothetical SCM analysis to all the control cities, and test if the gaps found in the four treated cities is relative larger than the gaps generated among cities in donor pool. Figure 2 depicts the gaps of both the treated (presented with bold red line) and control cities. Graphical results show relative large gaps of the treated cities in comparison with the placebo gaps. We also run the permutation inference based on the distribution of the placebo effects. The results indicate that the treatment effects of four cities are significantly non-zero (p-value=0.06, see table 1). Both tests confirm the significance of the positive treatment effects for the four cities. Placebo tests for the remaining cities are also performed, which gives additional evidence that the treatment effects for the other 7 cities are not significant (details could be found in SI).

[Table 1 about here]

In sum, the SCM evidence shows that only 4 out of the 11 costal open cities are significantly affected by the opening-up policy experiment administrated in 1984.

3.3.2. Results for Open Costal Economic Areas (1985)

Figure 3 demonstrates the baseline SCM results for the 10 cities in the open costal economic areas determined by the Chinese government in 1985. The SCM constructs synthetic units that perfectly trace the economic growth path of the treated units in the pre-intervention periods (years before 1985). The predictors balance results (see SI) also indicates that SCM offers the appropriate counterfactuals. Graphical results show that 6 cities (Huzhou, Suzhou, Foshan, Jiangmen, Jiaxing, and Zhongshan) have produced significantly better economic outcomes compared to their synthetic counterparts, while the other 4 cities remain unaffected by the opening-up policy. However, the placebo tests indicate that the seemly significant treatment effects could be spurious. Here we show the placebo test results for the 6 cities that outperform synthetic units. Figure 4 displays the gaps of the 6 cities vis-à-vis the placebo ones. Only gaps of two cities (Foshan and Suzhou) in the post-treatment periods seem to be large comparing with that of donor pool units. Permutation test confirm this finding in Table 1 as only performance gaps observed in the case of Foshan and Suzhou are significantly large (p-value=0.02 and 0.06 respectively). In sum, results of placebo test show that

only two cities are positively affected by the opening-up policy while the treatment effects of the other eight cities are insignificant.

[Figure 3 about here]

[Figure 4 about here]

3.3.3 Robustness Checks

Our baseline results show that the treatment effects of opening-up policy is significant in only a small portion of the experimental pilot cities (6 out of 21). That is, success of the opening-up policy experiment is hard to tell. In this part, we implement several robustness checks to further confirm our results.

We first use the traditional difference-in-difference (DID) to estimate the average treatment effects (ATE) of opening-up policy experiment. Although the two batches of pilot cities are designated around the similar time (May, 1984 and February, 1985, respectively), we still consider them as two different treatment and conduct the DID analysis for them separately.

For the 11 costal open cities, the balance checks show that the both the GDP and predictors match badly in the pre-treatment period. All the indicators of the treated cities are much higher than the control units. The standard DID analysis find that the ATE is positive but small and insignificant (0.091, with *p*-value=0.964). We further combine the DID analysis with Kernel matching technique to reduce the difference in observed confounders that might be associated with the outcome dynamics (Abadie, 2005). Nevertheless, matching still resulted in bad balance on the pre-treatment predictors. At the same time, the treatment effects estimated are relative larger, but still insignificant (24.738, with *p*-value=0.325).

Similar analysis for the 10 cities in open costal economic areas is also conducted. Different for above, the balance check finds that the treated cities and control units matches well in terms of both GDP and covariates in the pre-treatment periods. Standard DID analysis results in a significant negative treatment effects (-4.616, *p*-value=0.000). The gap remains negative (but insignificant) using matched DID estimation (-13.788, *p*-value=0.504).

Evidences from DID analysis do not lend support to the view that opening-up policy experiment induced a significantly positive impact on the pilot cities. (More details of the DID analysis could be found in the Appendix X in the Supporting Information.)

Secondly, we also consider confounding effects of other ongoing policy experiments. One major challenge to our results is that the treatment effects identified could be mixed with the impact of the Economic System Reforms (ESR). ESR is the other bundle of reform policies that is believed to contribute positively to the China's remarkable economic achievements. Since the ESR experimentation also started and scaled up around 1980s, we do need to check the potential bias that it could bring to our estimations. Good news is that SCM offers explicit weights combination of the

control cities in the construction of synthetic units. This enables us to estimate the extent and direction of the potential bias brought by ESR.

We divide the treated cities into two groups, i.e. Group 1: those included in both in opening-up and ESR pilot list; Group 2: those only included in opening-up pilot list but not in ESR list. We then calculate for each treated city the proportion of ESR pilot cities in the composition of synthetic counterpart. The basic idea is as follows. For Group 1 cities, if ESR pilots constitute a large portion of a synthetic unit, our estimation is likely unbiased; if ESR pilots constitute only a small portion, our estimation is biased upward. For Group 2 cities, if ESR pilots constitute a small portion of a synthetic unit, our estimation is likely unbiased; otherwise it's biased downward. Table 2 shows the composition of the treated cities' synthetic counterparts.

[Table 2 about here]

Result shows that for cities in Group 1, the ESR city weights are mostly less than 0.5, which means that our estimation is likely biased upward for these cities, i.e., the actual treatment effects of opening-up policy is smaller than we have estimated. This further supports our argument that the opening-up policy is not as successful as it looks.

For Group 2, the ESR city weights are quite small (less than 0.1 in most cases), which suggests that our estimation is likely to be unbiased. To be cautious, we rerun the SCM for cities with ESR weights larger than 0.1 (Nantong and Jiaxing) by leaving the ESR cities out from the donor pool. The results for these cities are consistent with our baseline findings (See Appendix S in SI). This indicates that our findings is not significantly affected by donor cities that are included in ESR program.

4. Opening-up policy 1978-1984: A historical account

Empirical results above show that early opening-up pilots have not shown clear advantage in economic performance over other non-pilot cities. A direct implication is that the later scale-up of opening-up pilots is not caused by or contingent upon the success of early pilots. This section reviews the incipience of opening-up pilots in Shenzhen, the dispute between reformists and conservatives in interpreting the result and the build-up of the new means-end narratives. Against the experiment-centered proposition, the historical account shows that policy actors' intentional maneuver matters more in changing the policy paradigm and promoting local policy experiments.

4.1 Incipience of Opening-up Pilots in Shenzhen

Different accounts exist for why Shenzhen was chosen to build up the first opening-up pilot. An often cited account has followed closely the problem-solving logic, i.e. observing many Guangdong residents fleeing to Hong Kong from Shenzhen since late 1950s, policy makers at the top and at front both recognized the urgency to change and then choose opening-up pilot as the solution. In Nov. 1977, Deng Xiaoping commented that "(Fleeing to Hongkong) is due to problems in our policy and could not be solved by the army". Xi Zhongxun, then first party secretary of Guangdong, echoed Deng in 1978 and said "these people (fled) just migrate to another place, which is internal

contradiction instead of enemy-friend contradiction; once our economy improves, they'll come back to us."[7]

However, problem with this account is that to feel the urgency to change doesn't automatically lead to change. Wu Nansheng, then part secretary of Guangdong and later the first party secretary of Shenzhen who was local to Shenzhen, clearly disagreed with this account.

> *"For a local official with my ability, is it possible that I didn't know the situation (fleeing to Hong Kong)? I only knew after Deng said it (is due to problems of our policy)? Was I that stupid?... Was the whole Guangdong leadership that incapable? "*[8]

Wu's challenge is reasonable. The problem of "fleeing to Hong Kong" was not new to decision makers. Why was the opening-up piloted in 1978? While the reform discourse in China consistenly favored more the reformists group gripping power since late 1978 after the 3rd plenum of 11th Party Congress, PRC's economic opening-up to the western countries dated earlier before 1978. Two years after its reclaiming seat of China in United Nations in 1971, PRC launched an ambitious "Four-Three Program" to import equipment sets worth of 4.3 billion from major western countries, which was later completed in 1982 and partly contributed to the economic growth observed during the reform era. Furthermore, a more immediate event before the opening-up pilot, was the first ever official visit of a PRC economic delegate to western countries in May 1978 led by then vice premier Gu Mu. In his memoir, Gu Mu made it clear opening-up has already been put on the agenda before his trip. His later trip reports (Gu Mu Report) to paramount leaders basically reiterated the urgency to opening up. An added value of this trip was, however, the discovery that most western countries are quite eager to provide help to China's opening up, which helped produce an optimistic view to outside environment for policy change.

While Gu Mu report did not propose to pilot SEZ, it clearly shifted the policy environment for opening up policy in general, whose impact continued to ferment till the 3rd plenum of 11th Party Congress in the end of 1978. The idea to have some special region in Guangdong to develop export-oriented industry has actually appeared before but failed[9]. However, after policy environment shifted, this idea appeared again in a more substantial way. Wu Nansheng remembered, which is also confirmed in other memoirs, it was Luo Xinquan, one of the many Hong Kong businessmen who swarmed into Guangdong in late 1978 and early 1979, that first raised this idea of "Free harbor" to him. Wu then proposed this idea to Xi Zhongxun who in turn seek support from the paramount leaders in April 1979[10]. With the overall support and optimism toward a more open economy, the idea to have four SEZs met few serious challenges. Party central quickly approved in July 1979 the proposal to pilot special economic zone in Shenzhen, Zhuhai and later Shantou, Xiamen[11]. To explore the technical details of SEZs, a government delegate led by Jiang Zemin, then deputy

[7] http://www.thepaper.cn/newsDetail_forward_1260984

[8] "Experiencing the decision making process for SEZ," dictated by Wu Nansheng, recorded by Xiao Donglian and Yang Jisheng. *Yanhuang Chunqiu*, 2015, Issue 5. P. 6-12

[9] Ibid.

[10] Ibid. Memior of Gu Mu.

[11] Central [1979] No. 50

director of National Import & Export Commission, was dispatched in Sept. 1980 to six different countries to learn from their experiences in developing export-oriented industrial park and free trade zones. More details for SEZ was concretized in July 1981.[12] In short, against the commonly accepted problem-solving logic and try-and-error experimental logic, opening-up city pilots appeared in China because the shift of policy environment encouraged a more aggressive way to develop a more open economy. Special economic zone, despite its name given by Deng being new, was not really a brand new idea. However, only after the shift of policy environment was it possible.

4.2 Dispute over the results of SEZs: Negative Lessons or Positive Experiences?

In a little over 2 years since early 1979, SEZs won great popularity among the top leaders. However, as early pilots started to operate, disagreement emerged between the reformist and the conservative. How to interpret the results of early opening up pilots became a serious political issue. At the peak of dispute, Chen Yun, another paramount leader in Deng's time usually considered more conservative than Deng, commented in late 1981 that "the imminent task for SEZ now is to learn its own lessons."[13]

Chen's concern over SEZs was not unfounded. For one, the negative consequences have manifested itself at the incipience of opening-up. While reformists were willing to tolerate these negative consequences and consider them as necessary cost for the benefit, it was legitimate for conservatives to attribute all of them to the opening-up experiment itself, especially when the economic benefits of SEZ has yet to manifest. Among the negative consequences, smuggling attracted most of the criticism, together with "capitalist elements" introduced to SEZs. Hu Qiaomu, a leading figure in communist theorists was recorded to comment after a visit to Shenzhen, "Shenzhen is all white (meaning capitalist) except for the national flag (is red)".[14] For the other, probably more important, the positive results of SEZ in economic development were not convincing at that time. Gu Mu's memoir suggested that reformists also recognized problems in SEZ. Among the problems, the most controversial one was that SEZ has profiteered on its policy advantage over other inland areas in China, which was clearly against the original goal of SEZ of developing export-oriented economy. Good performance of SEZ, if any, could result from resource reallocation among different provinces. As for export-oriented economy, the total exportation of four SEZs only amounted to 400 million dollars in 1984 and importation outnumbered exportation by 1 billion dollars[15].

Facing serious challenges on the dispute, Deng's attitude was ambivalent. In the end of 1981, the central government convened a work meeting of provincial leaders mainly to counter the smuggling issues. Leaders of Guangdong and Fujian, where 4 SEZs were located, were criticized and warned for the problems spotted. In early 1982, Deng went to Shenzhen for his winter break. When reached out by Guangdong provincial leaders for consultation, Deng's reply was "(I) come to rest, not to listen."[16]

---

[12] Central [1981] No. 27

[13] Selected Works of Chen, volume 3

[14] "Hua Guofeng on Opening Up", dictated by Hua Guofeng recorded by Zhang Gensheng Yanhuang Chunqiu, 2011.

[15] Memoir of Gu Mu, 2014

[16] Fn.8

This dispute over the interpretation of opening-up pilot performance lasted about a year. With the incessant efforts of reformists, Chen Yun's attitudes softened in Oct. 1982 when he commented a ninth version report on SEZs prepared by Guangdong provincial government that, "(We) should continue SEZs and should summarize lessons to better SEZs."[17] Softening of conservatives' opposition clearly gave a break for the development opening-up pilots. In response, reformist also slowed down the scale up of SEZs, although several localities were still enthusiastic to obtain SEZ status. For example, Hainan didn't become SEZ until 1988.

4.3. The 1984 Turn: SEZ is good and we are right

The turning point didn't come too late. Deng, who chose to be silent in 1982, spoke out in the spring 1984. In his spring tour to South again, Deng this time wrote inscriptions for Zhuhai and Shenzhen, respectively, "Zhuhai SEZ is good" and "Development and experience of Shenzhen testifies that our decision to build SEZ is right." Quickly after Deng's SEZ inscription, scale-up of opening-up policy was put on agenda. In Mar. 26, leaders of coastal city were convened in Beijing to discuss having more opening-up pilots. In May 4, 14 new opening-up pilots were announced[18]. Quickly after that, more coastal cities were added to the list of pilots or granted similar latitude in developing a more open local economy.

The 1984 turn came as a surprise to many, even to Gu Mu who admitted that "situation evolved faster than I expected."[19] Admittedly, Deng's judgment was partly backed up by the fact that opening-up pilots had won more ground especially for its positive spillover to the national economy. Besides that, however, the interaction between Deng as the reformist leader and Chen as conservative leader is especially crucial in the process. In general, these two leaders have worked tacitly to change the nature of political struggle of CPC from a death/life game to a win-win one (Saich, 2001). In the case of opening up, while Chen Yun clearly favored a more conservative route signified by his late1981 comments on SEZ, he didn't objective categorically the idea of open-up championed by Deng. As Deng's chose to be silent facing the challenge in 1982, Chen's attitude softened. In this 1984 round, after the SEZ inscription and returning to Beijing, Deng convened a meeting on Feb. 24 at his home to announce his decision to promote more opening-up pilots. Interestingly, two of close allies of Chen were convened among the audience, Yao Yilin and Song Ping who later on Mar. 14 went to Chen Yun's home, as requested by Deng, to convey Deng's decision.[20] Before the politburo was convened to pass Deng's scale-up decision in May, Gu Mu, as the implementer of Deng, went to Chen Yun's place again to debrief the details for scale-up in Apr. 23[21]. The consensus was finally reached.

5. Conclusion

[17] Memoir of Gu Mu, 2014; p. 365; also fn. 8
[18] Central [1984] No.13.
[19] Gu Mu Memoir, p. 372
[20] Annals of Deng Xiaoping, p.962; Annals of Chen Yun, p. 401.
[21] Gu Mu Memoir, p. 377.

What we've found in this paper is twofold. On one hand, we counter the argument that positive result of previous policy pilots itself could initiate the self-reinforcing sequencing leading to policy changes at the level of paradigmatic shift. The performance gap between those Chinese coastal cities serving as early pilots of opening-up policy and those not was at most hard to tell. Results of SCM shows pilot cities didn't submit an overall better, if not worse, performance sheet in economic growth than their counterparts synthesized based upon their pre-pilot parameters. This result is robust using DID methods and even after we control the possible economy boosting effect of economic system reform. Second, instead of the experiment-centered logic, we propose a political account to explain why policy pilots brought about drastic policy changes observed in China. While paramount leaders chose to change the course, policy pilots emerged out of local spontaneity or center's promotion. In this process, political competition has emerged to interpret the results of policy pilots. With different consequences presented, paramount leaders' subjective construction of means-end narratives determined at the same time the fate of policy pilots and the direction of policy change. Our historical review shows how a new consensus out of Deng and Chen's interaction eventually led to the scale-up of opening-up pilots and eventual a more open economy in China.

We might be wrong for the performance gap between opening-up pilots and non-pilot. An easy challenge is that the hard-to-tell performance gap was not because pilot cities have performed badly but due to the facts that other cities have also performed so well that performance discrepancy diminishes. The good performance of other cities could be caused by the positive spillover of pilot cities. For example, the center allowing SEZs to try market-oriented reform measures may encourage other cities to boldly try similar or even more aggressive liberal policy in promoting their own economy; therefore, opening-up pilots have positive impact on economic growth. We admit it is possible as all cities in our dataset, pilots and non-pilots, have grew rapidly during reform era. However, if such reasoning is correct, it actually supports our argument that outcome of early policy pilots is not really important. If policy pilots serves primarily as the signal for the launch of more marketization reform and could hence bring about impressive economic growth, then the primary goal of policy pilots was not to be successful but to smuggle changes in.

Further, by challenging the success of early opening-up pilots, we in no way argue that opening-up is not good for China. As for today, benefit of a more open economy for China is a simple fact, period. What we are really challenging is the experimental logic used to interpret how an open economy in China was brought about. The role of local policy experimentation should not be exaggerated, especially for its capacity to bring about paradigmatic shift in policy environment. While the assessment of pilots' result is a subjective matter under different means-end narratives, to rely on policy experiment for policy change is to put the cart before the horse. Specifically for the success of opening-up policy, it is more likely a coincidence that reformist has chosen the right thing to experiment in the first place. Piloting itself, although helping fine details of policy program, might end up with catastrophic consequences. Above all, piloting or experimentation point (*Shidian*) was not a new practice in China (Heilmann, 2008a) and its conducive effect was not guaranteed.

Finally, we do have a little more ambition to redirect the scholarship in explaining China's success. Debates always exist whether China's experience converges to other countries or China has found a new and unique way to prosperity (Rawski, 1999; Woo, 1999; Acemoglu and Robinson, 2012;

Fukuyama, 2013). A polemic paradox is as follows. It is much easier to reject the convergence argument because of every country including China is unique and China's success until today has been robust. However, the uniqueness argument must be substantiated to the extent that it could have predicted rather than described the success of China. It is in the sense that we oppose the experiment-centered proposition.

Table 1 Permutation Inference for Cities with Significant Positive Gaps

Permutation Tests of Post/Pre RMSPE Ratio (Costal Open Cities, 1984)						
Cities	GUANGZHOU	NINGBO	YANTAI	ZHANGJIANG		
Ratio of Post-Pre RMSPE	33.44	29.01	34..09	31.64		
Rank, Highest to Lowest	3/51	3/51	3/51	3/51		
P-value, one tail test P (pool>treated)	0.06	0.06	0.06	0.06		
Permutation Tests of Post/Pre RMSPE Ratio (Open Costal Economic Areas, 1985)						
Cities	SUZHOU	FOSHAN	HUZHOU	JIANGMEN	JIAXING	ZHONGSHAN
Ratio of Post-Pre RMSPE	26.99	86.44	4.55	19.14	7.99	13.59
Rank, Highest to Lowest	3/51	1/51	37/51	8/51	27/51	15/51
P-value, one tail test P (pool>treated)	0.06	0.02	0.73	0.16	0.53	0.29

Note: The "Ratio of Post-Pre" equals the absolute value of the ratio of the rooted average Treated-Synthetic control square deviation in 1985-1992 divided the rooted average Treated-Synthetic control square deviation in the pre-period (1978-1983 for WENZHOU and NANTONG). We also calculate the same ratio for each city in the donor pool and construct a distribution of the 51 ratio statistics. The "rank" entry shows were the Treated City ranks in the distribution of 66 values (top to bottom) the p-value is a test of the probability that a random draw from the donor pool takes a higher than the treated city value.

Table 2: Weights of ESR Cities in Synthetic Combination

	Treated Cities Included in ESR Experimental Cities					
City	Guangzhou	Zhanjiang	Ningbo	Yantai	Foshan	Suzhou
ESR city weights	0.344	0.007	0.657	0.393	0.193	0.522
City	Qinhuangdao	Fuzhou	Wenzhou	Tianjin	Qingdao	
ESR city weights	0.028	0.621	0.097	1	1	
City	Changzhou	Huzhou	Jiangmen	Quanzhou	Wuxi	
ESR city weights	0.397	0.370	0.064	0	0.548	
	Treated Cities not Included in ESR Experimental Cities					
City	Lianyungang	Nantong	Zhangzhou	Zhongshan	Jiaxing	
ESR city weights	0.051	0.229	0.045	0.001	0.115	

Figure 1 Graphical Results for Costal Open Cities

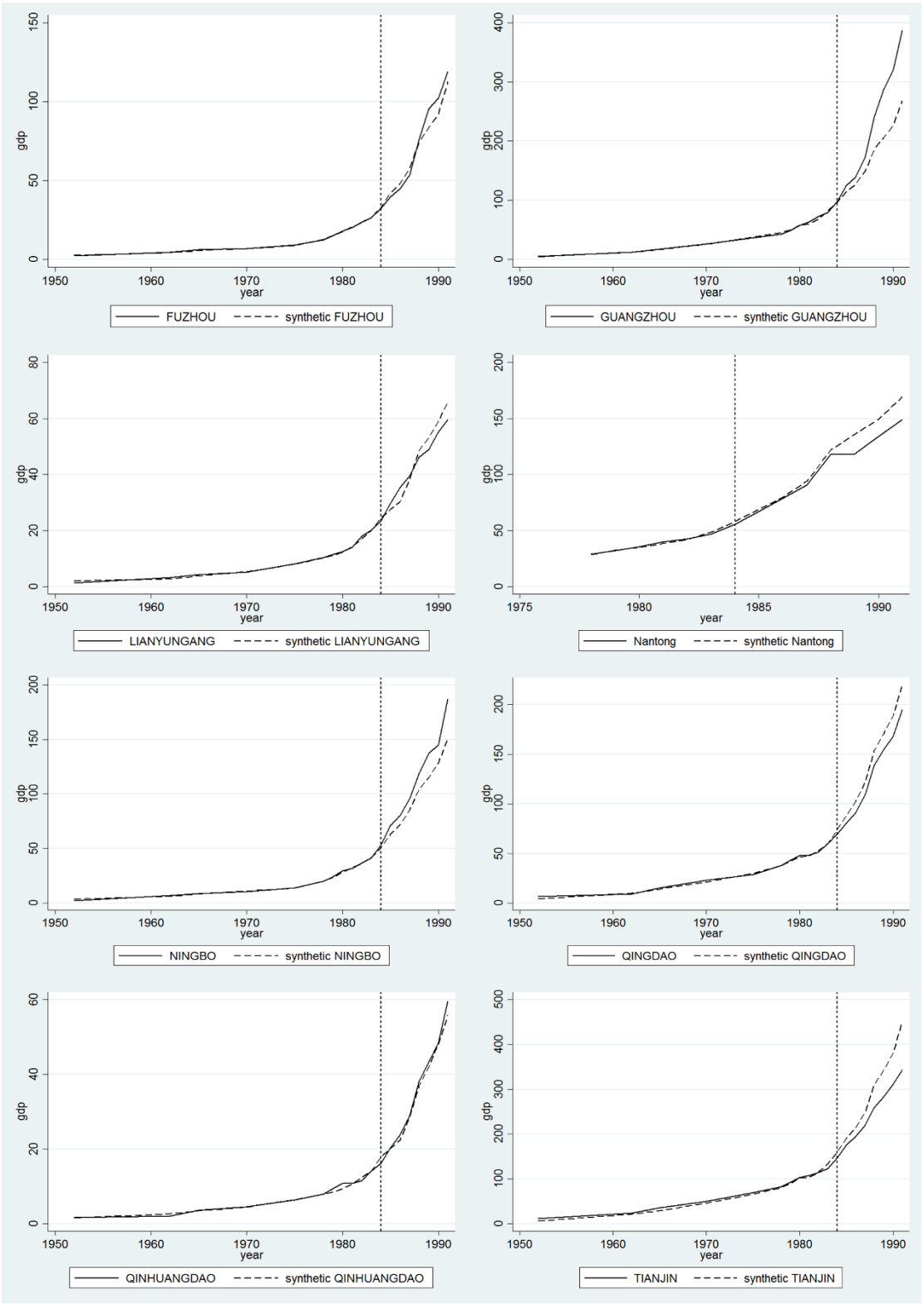

gdp
year
FUZHOU
synthetic FUZHOU
GUANGZHOU
synthetic GUANGZHOU
LIANYUNGANG
synthetic LIANYUNGANG
Nantong
synthetic Nantong
NINGBO
synthetic NINGBO
QINGDAO
synthetic QINGDAO
QINHUANGDAO
synthetic QINHUANGDAO
TIANJIN
synthetic TIANJIN

Figure 1 Graphical Results for Costal Open Cities (continued)

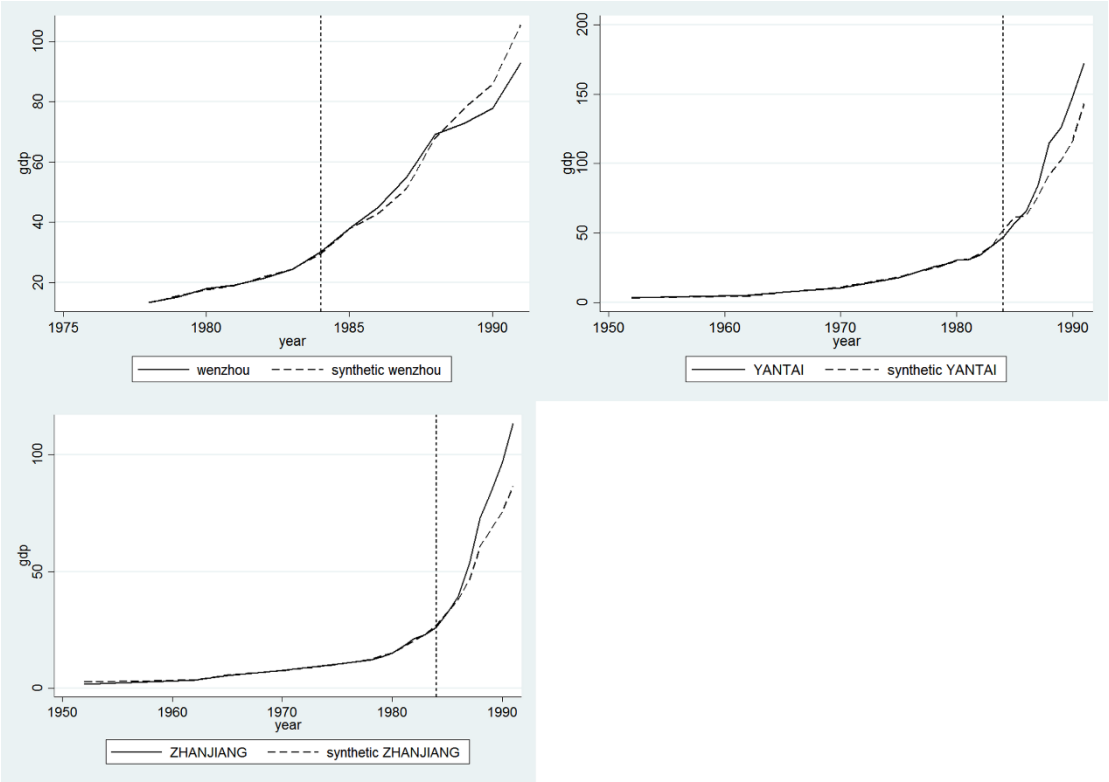


Figure 2 Placebo Tests for Costal Open Cities

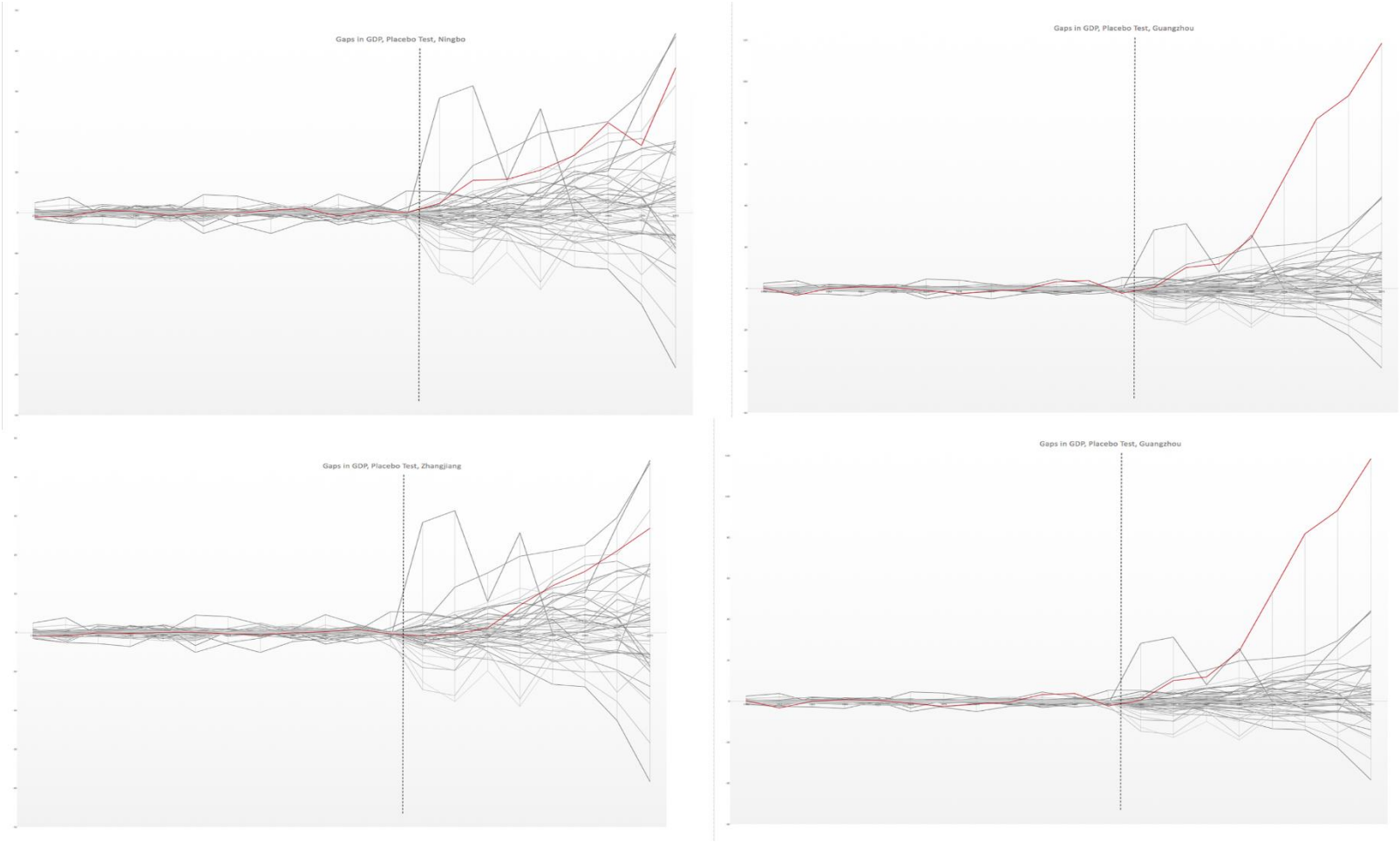


Figure 3 Graphical Results for Open Costal Economic Areas

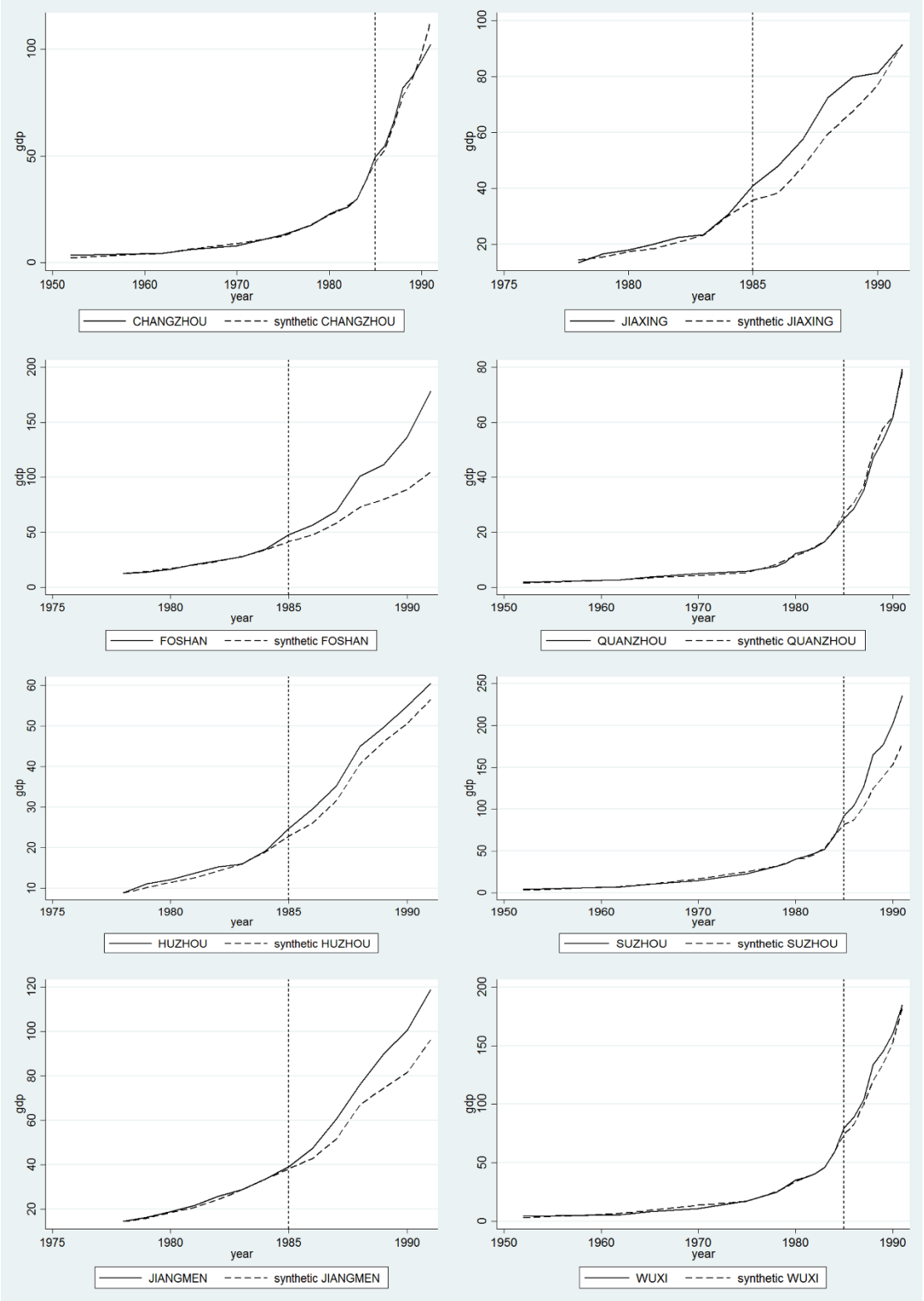


Figure 3 Graphical Results for Open Costal Economic Areas (continued)

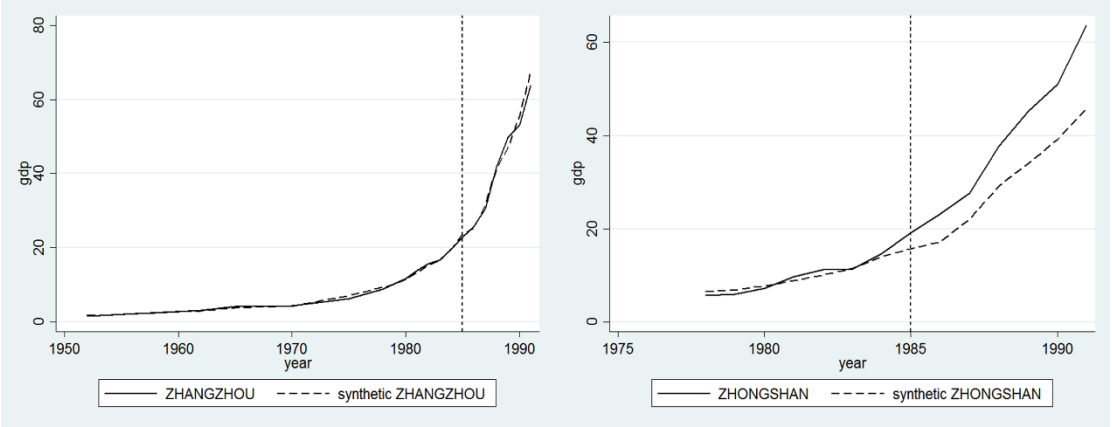
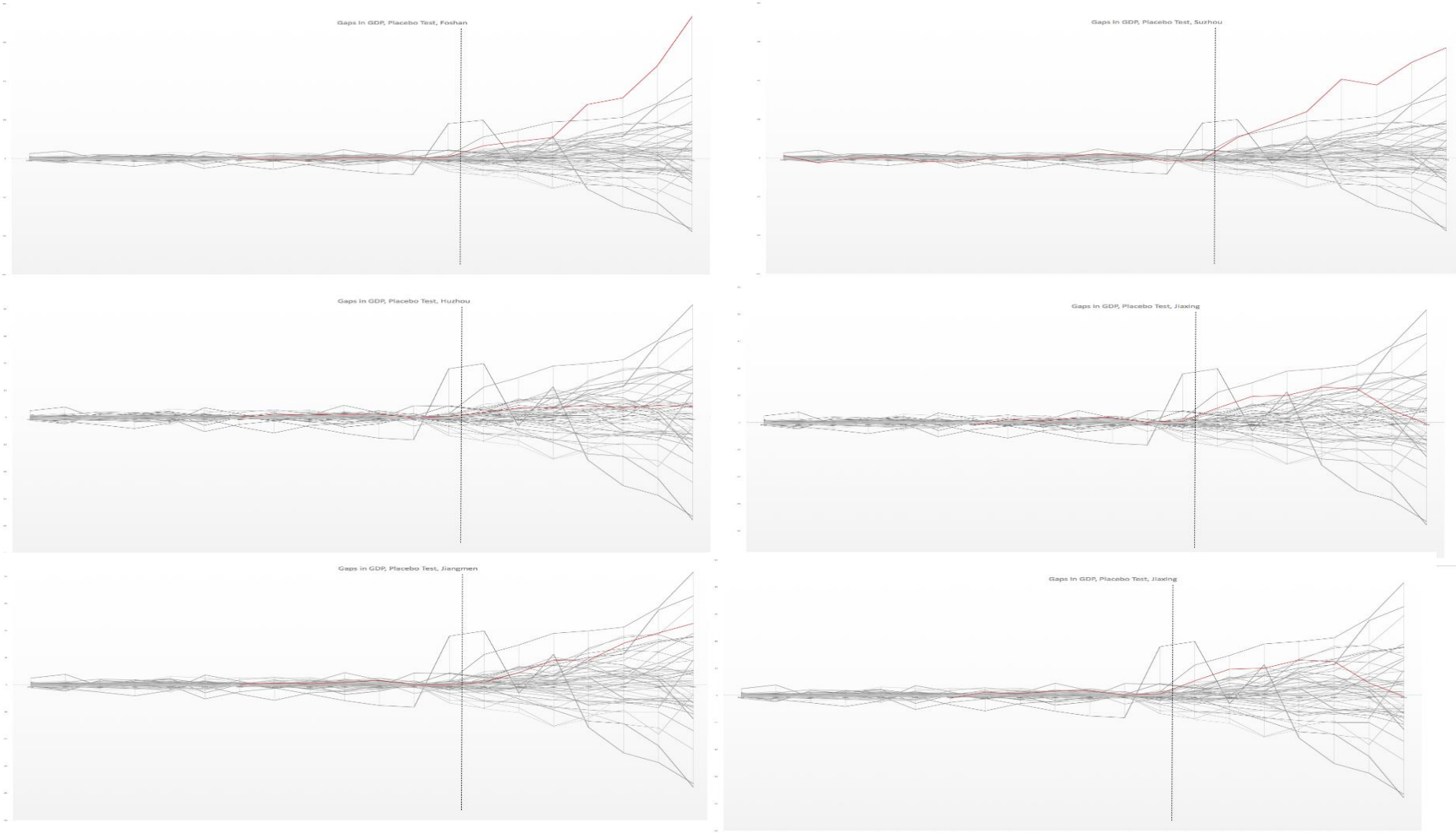


Figure 4 Placebo Tests for Open Costal Economic Areas



Bibliography

Abadie, A. & Gardeazabal, J. (2003), The economic costs of conflict: A case study of the Basque Country, *American Economic Review,* 93(1): 113-132.

Abadie, A. (2005), Semiparametric difference-in-differences estimators, *The Review of Economic Studies,* 72(1): 1-19.

Abadie, A., Diamond, A. & Hainmueller, J. (2010), Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program, *Journal of the American Statistical Association* 105(490):493-505.

Abadie, A., Diamond, A. & Hainmueller, J. (2011), Synth: An R package for synthetic control methods in comparative studies, *Journal of Statistical Software,* 42(13): 1-17.

Abadie, A., Diamond, A. & Hainmueller, J. (2015), Comparative Politics and the Synthetic Control Method, *American Journal of Political Science,* 59(2):495-510.

Bhatti, Y., A.L. Olsen and L.H. Pedersen, Administrative Professionals and the Diffusion of Innovations: The case of citizen service centres. Public Administration, 2011. 89(2): p. 577-594.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004), How much should we trust difference-in-difference estimators, *Quarterly Journal of Economics,* 199(1): 249-275.

Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*,*95*(3), 983-1001.

Chung, J.H., Central Control and Local Discretion in China: Leadership and Implementation during Post-Mao Decollectivization. 2000: Oxford University Press.

Dolowitz, D. and D. Marsh, Who learns what from whom: a review of the policy transfer literature. Political studies, 1996. 44(2): p. 343-357.

Fremeth, A. R., Holburn, G. L. & Richter, B. K., (2016), Bridging qualitative and quantitative methods in organizational research: Applications of synthetic control methodology in the US automobile industry, *Organizational Science*, 27(2): 462-482.

Guo, S., & Fraser, M. W. (2014). *Propensity Score Analysis: Statistical Methods and Applications: Statistical Methods and Applications* (Vol. 11): Sage Publications.

Hacker, J.S., Privatizing risk without privatizing the welfare state: The hidden politics of social policy retrenchment in the United States. American Political Science Review, 2004. 98(02): p. 243-260.

Hall, P.A., Policy paradigms, social learning and the state. Comparative politics, 1993. 25(3): p. 275-296.

Heilmann, S., Policy experimentation in China's economic rise. Studies in Comparative International Development, 2008. 43(1): p. 1-26.

Heilmann, S., From local experiments to national policy: The origins of China's distinctive policy process. The China Journal, 2008: p. 1-30.

Karl, P., The Problem of Demarcation, in Popper Selections. 1985, Princeton University Press. p. 118-130.

Kuhn, T.S., The structure of scientific revolutions. 2012: University of Chicago press.

Lakatos, I., Falsification and the Methodology of Scientific Research Programmes, in Criticsm and The Growth of Knowledge. 1970: Cambridge At the University Press. p. 91-195.

Lin, J.Y., Rural reforms and agricultural growth in China. The American economic review, 1992: p. 34-51.

Lindblom, C.E., The science of" muddling through". Public administration review, 1959: p. 79-88.
Mei, C. and Z. Liu, Experiment-based policy making or conscious policy design? The case of urban housing reform in China. Policy Sciences, 2014. 47(3): p. 321-337.
Mintrom, M. and P. Norman, Policy entrepreneurship and policy change. Policy Studies Journal, 2009. 37(4): p. 649-667.
Park, S. and F. Berry, Successful Diffusion of a Failed Policy: The case of pay-for-performance in the US federal government. Public Management Review, 2014. 16(6): p. 763-781.
Pierson, P., Increasing returns, path dependence, and the study of politics. American political science review, 2000. 94(02): p. 251-267.
Pierson, P., Politics in time: History, institutions, and social analysis. 2004: Princeton University Press.
Ramo, J.C., The Beijing Consensus. 2004: Foreign Policy Centre London.
Rawski, T.G., Reforming China's economy: what have we learned? China Journal, 1999(41): p. 139.
Teodoro, M.P., Bureaucratic job mobility and the diffusion of innovations. American Journal of Political Science, 2009. 53(1): p. 175-189.
Thelen, K., How institutions evolve: The political economy of skills in Germany, Britain, the United States, and Japan. 2004: Cambridge University Press.
Woo, W.T., The real reasons for China's growth. The China Journal, 1999. 41(1): p. 115-37.
Xu, C., The fundamental institutions of China's reforms and development. Journal of Economic Literature, 2011: p. 1076-1151.
Xu, Y. (2015), Generalized synthetic control method for casual inference with time-series cross-sectional data, *Working Paper* (No. 2015-1), MIT.
Zeng, J., Did Policy Experimentation in China Always Seek Efficiency? A case study of Wenzhou financial reform in 2012. Journal of Contemporary China, 2015. 24(92): p. 338-356.
Zhu, X., Mandate versus championship: Vertical government intervention and diffusion of innovation in public services in authoritarian China. Public Management Review, 2013(ahead-of-print): p. 1-23.
Zhu, X. and Y. Zhang, Political Mobility and Dynamic Diffusion of Innovation: The Spread of Municipal Pro-Business Administrative Reform in China. Journal of Public Administration Research and Theory, 2015: p. muv025.