

# ***Can Cross-Border E-Commerce Comprehensive Pilot Zones Enhance Urban Export Resilience? — Multi-Period DID Evidence from Prefecture-Level Cities in China***

**Ruochu Wang**

*School of Business, Zhengzhou University, Zhengzhou, China  
15533987044@163.com*

**Abstract.** Based on panel data of 282 prefecture-level cities in China from 2010 to 2022, this paper treats the phased establishment of cross-border e-commerce comprehensive pilot zones as a quasi-natural experiment, and uses the multi-period difference-in-differences method to examine the relationship between the construction of comprehensive pilot zones and urban export resilience. The study finds that the comprehensive export resilience index of pilot cities shows a significant increase, and the direction and magnitude of the TWFE and CSDID estimates are basically consistent. The decomposition of indicators shows that this change is mainly reflected in the improvement of trade support conditions such as economic base, residents' income, financial services, and technological level, while direct trade indicators such as export scale, growth rate, volatility, and resistance have not yet shown significant changes within the sample period. Further tests show that some early pilot cities already exhibited strong development foundations and first-mover advantages before approval, suggesting that policy selection and advance layout may affect the estimation results. Overall, the short-term effect of the comprehensive pilot zones is more reflected in the capacity building of export resilience, and its transformation into direct trade performance still requires a longer time. This paper provides city-level empirical evidence for distinguishing the 'capacity-building effect' from the 'trade-outcome effect' of digital trade policies.

**Keywords:** cross-border e-commerce comprehensive pilot zone, urban export resilience, multi-period difference-in-differences, parallel trends, composite indicator

## **1. Introduction**

Cross-border e-commerce is changing the way firms acquire overseas customers, organize cross-border logistics, and complete payment and settlement. Compared with traditional foreign trade, digital platforms can migrate the links of product display, supply-demand matching, credit evaluation, payment, and after-sales service online, thereby reducing the costs of information search and market entry [1, 2]. The World Trade Organization points out that digital technologies may not only compress trade costs, but also change trade structure, comparative advantage, and the organization of global value chains [3]. For the numerous small and medium-sized enterprises with

limited resources, digital channels may especially ease the constraints of information, certification, and fixed costs they face when entering international markets.

To form a regulatory and service system adapted to the development of cross-border e-commerce, the State Council approved the establishment of the China (Hangzhou) Cross-Border E-Commerce Comprehensive Pilot Zone in 2015, and then continuously expanded it in 2016, 2018, 2019, and 2020, and added 60 cities and regions in two batches in 2022. The comprehensive pilot zone is not a single subsidy policy, but centers on the online comprehensive service platform and the offline industrial park, coordinating the promotion of institutional innovation in information sharing, financial services, smart logistics, e-commerce credit, statistical monitoring, and risk prevention and control [4, 5]. Its policy implications include both the facilitation of customs clearance, tax rebates, and foreign exchange settlement, and cross-department data coordination, industrial agglomeration, and public service capacity building, and therefore may affect both urban trade activities and general development conditions.

At the same time, trade frictions, global demand fluctuations, supply chain disruptions, and public health events have made 'resilience' an important topic in the study of opening up. Export resilience is not equivalent to export scale; it involves at least the resistance, recovery, and adaptation capacities of cities under external shocks [6, 7]. Even a city with a large total export volume may lack resilience due to excessive market concentration, single product structure, or slow institutional adjustment; conversely, a city with limited export scale may also rely on market diversification, flexible supply chains, and stronger institutional response capacity to reduce shock losses. Therefore, the evaluation of comprehensive pilot zones should not only examine 'whether exports increase,' but also analyze whether they improve the city's capacity to respond to external changes.

Most existing studies find that comprehensive pilot zones can promote cross-border e-commerce exports [4, 5], wage income [8], industrial structure upgrading [9], regional coordinated development [10], or general economic growth [11], but few directly test urban export resilience; existing resilience studies mainly explain differences from export market diversification, economic structure, and regional adaptation capacity [12, 13]. At the methodological level, the phased implementation of comprehensive pilot zones may make conventional TWFE subject to heterogeneous treatment effects and inappropriate weights [14, 15]; although the entropy method composite index can aggregate multi-dimensional information, it may mix the general development level of GDP, income, finance, and technology with direct export performance. The above methodological and measurement issues set preconditions for the credibility of policy evaluation conclusions.

This paper focuses on answering two questions: first, whether the construction of comprehensive pilot zones is accompanied by the improvement of urban export resilience; second, whether this improvement is reflected in the increase of export scale and stability, or mainly in the enhancement of support capacities such as finance, technology, and economic base. To this end, this paper uses CSDID and TWFE to estimate policy-related effects, and further compares the composite index, component indicators, and direct export results. The study finds that the comprehensive export resilience index of pilot cities increases significantly, but this change mainly comes from the improvement of support conditions, and direct export performance has not yet improved synchronously. This paper thus summarizes the role of comprehensive pilot zones as 'capacity building first, trade performance transformation relatively lagging,' and discusses the impact of pilot selection and advance layout on result interpretation.

The rest of this paper is arranged as follows: Part 2 reviews the policy background and mechanism of action; Part 3 reviews the literature from three directions of digital trade, regional resilience, and policy evaluation methods; Part 4 describes the data, variables, and identification strategy; Part 5 reports the empirical results; Part 6 discusses the measurement and identification implications; Part 7 gives the conclusions and future research directions.

## **2. Policy background and theoretical mechanism**

### **2.1. Policy evolution and institutional content**

The cross-border e-commerce comprehensive pilot zone reflects China's policy path of promoting open institutional innovation through local pilots [4, 5]. The first batch of pilots in Hangzhou in 2015 focused on exploring regulatory, statistical, and service rules adapted to cross-border e-commerce; the second batch in 2016 expanded to 12 cities including Tianjin, Shanghai, and Chongqing; in 2018, 2019, and 2020, 22, 24, and 46 cities and regions were newly added respectively; in 2022, two more batches of 60 pilots were added. Policy coverage gradually extended from the eastern coast and central cities to the central and western regions, the northeast, and border areas, and different batches differ significantly in opening foundation, logistics conditions, industrial structure, and administrative capacity [11]. This phased expansion provides the spatiotemporal variation required by multi-period DID, and also means that treatment effects may vary by batch and implementation years.

From the perspective of policy instruments, comprehensive pilot zones usually take the 'six systems and two platforms' as the basic framework [4, 5]: the information sharing system aims to connect the data of customs, taxation, foreign exchange, market supervision, logistics, and financial institutions; the financial service system provides support around payment, settlement, financing, and insurance; the smart logistics system improves cross-border distribution and warehousing efficiency; the credit, statistical monitoring, and risk prevention and control systems are used to alleviate information asymmetry and regulatory risks in online transactions; the online comprehensive service platform and the offline industrial park undertake the functions of process integration and enterprise agglomeration. Therefore, the actual shock of the comprehensive pilot zone is not a homogeneous, instantly completed treatment, but a process in which institutional rules, digital infrastructure, and industrial ecology are gradually implemented. This paper takes the year after approval as the benchmark effective year, and takes the year of approval as the effective year for robustness checks.

### **2.2. Reducing transaction costs and expanding markets**

The primary mechanism through which digital technology affects international trade is the reduction of information and market entry costs. Freund and Weinhold [16] find that Internet development can promote export growth, and explain it as a decline in market-specific fixed costs. Goldfarb and Tucker [1] further summarize the cost changes brought by digitalization as declines in search, replication, transport, tracking, and verification costs. Cross-border e-commerce platforms improve the matching efficiency of buyers and sellers through search, recommendation, online communication, and credit evaluation, while comprehensive pilot zones extend the digital advantages at the platform end to the cross-border fulfillment link through the coordination of customs clearance, tax rebate, payment, and logistics institutions. Theoretically, this institutional and

technological complementarity can lower the threshold for firms to enter new markets, expand the scope of serviceable markets, and provide alternative channels when a single market is blocked.

However, digitalization does not completely eliminate distance and institutional boundaries. Blum and Goldfarb [17] show that even when the transport cost of digital products approaches zero, preference, language, and cultural distances still affect cross-border consumption; Lendle et al. [2] find that online platforms markedly weaken but do not eliminate the geographic distance effect. Thus, comprehensive pilot zones may improve market accessibility, but do not necessarily bring an immediate increase in export scale or growth, especially when firms face constraints of product quality, overseas compliance, brand awareness, and supply chain capacity, the effect of institutional facilitation may be more reflected in long-term adaptation capacity.

### 2.3. Enhancing adaptation capacity and potential offsetting effects

From a resilience perspective, policy may improve cities' adaptation capacity through three channels. First, cross-department information sharing and process standardization increase policy response speed, enabling firms to adjust customs declaration, logistics, and settlement arrangements faster when demand or rules change. Second, the agglomeration of platforms, logistics, payment, marketing, and professional service institutions can form knowledge spillovers, helping firms identify new markets and reorganize supply chains. Third, digital infrastructure, financial services, and talent training enhance the city's general development capacity, providing support for firms to shift to new products, new markets, or new business models. This is consistent with the adaptation and adaptive capacity emphasized by evolutionary economic geography: resilience is not only returning to the original path, but also forming new development paths [7].

Policy effects may also be affected by multiple factors. First, pilot qualification is not randomly assigned, and cities approved earlier usually have better opening foundations and growth trends. Second, the concentration of resources in pilot parks may distribute policy benefits unevenly within cities or between regions. Third, changes in external demand, exchange rate fluctuations, and trade frictions may also affect the release of policy effects. Finally, the construction content of comprehensive pilot zones is complex, and a single approval year is difficult to fully characterize the actual implementation intensity. Based on the above mechanisms, this paper proposes research hypothesis H1: the construction of cross-border e-commerce comprehensive pilot zones enhances urban export resilience by reducing cross-border transaction costs, improving public services, and strengthening market adaptation capacity.

## 3. Literature review

### 3.1. Digital technology, e-commerce and international trade

Early digital trade research mainly examined whether the Internet promotes international trade and whether it weakens the distance effect. Starting from Internet host growth, Freund and Weinhold [16] find a significant correlation between Internet development and export growth, supporting the explanation that digital technology reduces fixed entry costs in foreign markets. Taking digital product consumption as the object, Blum and Goldfarb [17] point out that even if traditional transport costs disappear, distance can still affect cross-border transactions through preference and information channels. Comparing eBay with traditional trade, Lendle et al. [2] find that the distance elasticity on online platforms is significantly smaller, and attribute it to the decline in search costs and credit evaluation technology. Goldfarb and Tucker [1] then explain from a more general digital

economy framework how data technology changes costs such as search, verification, and tracking. The above studies provide a micro-foundation for the comprehensive pilot zone to influence trade through digital platforms and institutional coordination, and also suggest that the judgment of 'going online' being equivalent to 'distance disappearing' is too simplistic.

In the Chinese context, Ma and Guo [4] use cross-border e-commerce logistics waybill data to study the institutional innovation of comprehensive pilot zones, and find that the policy can promote cross-border e-commerce exports and is related to the easing of financing constraints and the reduction of trade costs. Hu and Song [8] start from the wage income of employees of listed companies, showing that cross-border e-commerce reform may also have broader impacts through employment and enterprise operation channels. Wang and Rui [11] use the synthetic control and panel policy evaluation methods, emphasizing that the economic effects of different cities differ significantly. Zhang et al. [10] further discuss the regional coordination effect of comprehensive pilot zones from the synchronization of urban economic cycles; Peng and Zhang [9] find that the policy is related to industrial structure upgrading. These studies show that the impact of comprehensive pilot zones may go beyond the export scale itself, but also imply that when directly incorporating general development indicators such as GDP, income, and finance into the 'export resilience' composite index, the economic meaning of the explained variable needs to be defined with caution.

### 3.2. Regional resilience and export resilience

The regional resilience literature has evolved from 'restoration to the original state' to 'adaptation and transformation.' From an evolutionary perspective, Simmie and Martin [18] emphasize the adaptation capacity of regions in shocks; Hassink [19] warns that if the resilience concept lacks clear mechanisms and historical dimensions, it may obscure the long-term evolution process of regions. Martin [20] distinguishes resistance and recovery, and incorporates the path dependence after shocks into the analysis. Martin and Sunley [6] further point out that there are still large disputes over the definition, measurement, and determinants of regional resilience and its relationship with long-term growth. Boschma [7] emphasizes that the industrial, network, and institutional dimensions jointly determine whether a region can form a new growth path; the empirical study of Martin et al. [13] shows that both industrial structure and region-specific competitiveness affect resistance and recovery.

Export resilience is the concretization of regional resilience in the open economy field, usually requiring answers to three questions: what shock is faced, which export subject or spatial unit has resilience, and whether resilience is reflected in resistance, recovery, or structural adjustment. Based on Chinese customs data, Wang et al. [12] study the impact of trade barrier shocks, and find that export market-related diversification helps cities resist external shocks, and thereby link market structure with export resilience. Compared with this measurement centered on clear shocks and export performance, the composite index can reflect broader support capacity, but may mix in general economic growth. Therefore, the composite index and the direct shock response indicators should complement each other, rather than replace each other.

### 3.3. Phased policy evaluation and identification tests

Phased pilot policies make traditional DID face new identification problems. Goodman-Bacon [15] shows that the TWFE coefficient under staggered implementation can be written as a weighted average of different two-group two-period DID, and may produce hard-to-interpret weights when

treatment effects vary over time or groups. de Chaisemartin and D'Haultfoeuille [14] as well as Sun and Abraham [21] further point out that the lead and lag coefficients of dynamic TWFE may be contaminated by treatment effects of other periods. Based on group-time ATT, Callaway and Sant'Anna [22] allow researchers to use never-treated or not-yet-treated units as controls, and aggregate by group, calendar time, or event time, which is the main framework adopted in this paper. The imputation estimator proposed by Borusyak, Jaravel, and Spiess [23] also emphasizes that counterfactual prediction, estimation target, and treatment effect heterogeneity should be clearly distinguished.

The parallel trend test itself cannot be understood mechanically. Roth [24] points out that the pre-policy test may have insufficient power, and selecting a model based on pre-test results creates new inference biases; Rambachan and Roth [25] therefore propose a sensitivity analysis that allows limited deviations from parallel trends. This paper does not equate the insignificance of a certain average pre-policy coefficient with the joint zero of all pre-policy differences, but takes the joint pre-trend test of CSDID as the main diagnosis, and combines it with TWFE event studies, excluding early batches, and shortening event windows for comparison. The research gap is thus reflected in: not only asking whether the coefficient of the comprehensive pilot zone is significant, but also judging whether the significance is built on credible pre-treatment trends and outcome variables with clear economic meaning.

## 4. Research design

### 4.1. Data sources and sample matching

This paper uses a balanced panel of 282 prefecture-level cities from 2010 to 2022. The policy list and approval time are sorted out according to the seven batches of comprehensive pilot zone approvals by the State Council. Among the 165 pilot zones, 150 can be accurately matched at the prefecture-level city level; Yiwu, Suifenhe, Manzhouli, and Tongjiang are mapped to Jinhua, Mudanjiang, Hulunbuir, and Jiamusi respectively according to administrative affiliation. Since Jinhua is both the superior prefecture-level city of Yiwu and was approved in its own name in later batches, 153 unique treated cities are obtained after coding by the earliest treatment year. Another 11 pilot zones are excluded because their corresponding statistical units are not in the city panel; they are neither set as the treatment group nor wrongly classified into the control group.

The benchmark sample only retains the five batches of pilot cities and never-treated cities from 2015 to 2020. Calculated by the effective year of the following year after approval, the treatment group includes 102 cities, the never-treated group 129 cities, totaling 231 cities and 3,003 city-year observations. The two batches of 51 cities in 2022 have no complete post-policy years within the sample period, so they do not participate in the benchmark effect identification, and are only added as the not-yet-treated group in robustness checks. This treatment can reduce the contamination caused by treating cities that have been approved but have not yet formed complete annual effects as the control group for a long time.

### 4.2. Variable definition and measurement boundary

The core dependent variable is the urban export resilience index synthesized by the entropy method. The original indicator system includes 12 items: regional GDP, urbanization rate, per capita road area, foreign trade dependence, export trade, export deviation, financial development level, per capita disposable income of urban residents, fiscal investment intensity, science and technology

level, distance to the nearest port, and foreign capital dependence. Among them, the weights of export deviation, foreign trade dependence, GDP, and science and technology level are about 28.6%, 21.5%, 17.1%, and 11.6% respectively. The index covers multiple dimensions such as economic base, trade performance, infrastructure, finance, and innovation, and can reflect the comprehensive support conditions of cities in responding to external changes.

This indicator also has measurement boundaries that need to be faced squarely. First, GDP, income, finance, and technology are not direct manifestations of export resilience, but potential support capacities; second, the policy itself may affect these general development indicators, so that the composite index is significant but the direct export results are not; third, the distance to the port is a time-invariant variable and cannot be identified for its post-policy change in the city fixed-effects model. To avoid circular explanation, this paper does not use the 12 components as benchmark control variables, but uses them as outcome variables for diagnosis respectively; at the same time, it reports alternative results such as ln export value, export growth rate, three-year rolling volatility, export resistance, export deviation, and foreign trade dependence.

### 4.3. Model specification

As a comparison, this paper first estimates the TWFE model that includes city and year fixed effects:

$$\text{Resilience}_{it} = \beta \text{Did}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (1)$$

where  $\text{D\_id\_it}$  takes 1 after the city enters the policy effective period,  $\mu_i$  and  $\lambda_t$  are the city fixed effect and the year fixed effect respectively, and the standard errors are clustered at the city level. Given that the pilots are implemented in batches and the treatment effects may be heterogeneous, this paper does not take TWFE as the only evidence.

The main estimation adopts the group-time average treatment effect of Callaway and Sant'Anna [22]:

$$\text{ATT}(g, t) = E[Y_t(1) - Y_t(0) | G = g] \quad (2)$$

$G=g$  denotes the year when the city first enters the policy period. CSDID uses the not-yet-treated group and the never-treated group to construct controls respectively, adopts the dripw estimator, and aggregates the valid  $\text{ATT}(g, t)$ . It needs to be emphasized that the causal interpretation of point estimates relies on conditional parallel trends under no anticipatory response and corresponding control groups. If the pre-treatment  $\text{ATT}(g, t)$  jointly deviates significantly from zero, then even if the average ATT is significant, it cannot be automatically interpreted as the net policy effect.

### 4.4. Identification tests and robustness arrangements

The identification test takes the joint pre-policy trend test of CSDID as the core, rather than taking whether  $\text{Pre\_avg}$  is significant as a substitute. This paper also reports the caliber of only never-treated group, excluding Hangzhou and the first batch, shortening the event window, and reconstructing the index with pre-policy fixed entropy weights. On the TWFE side, it reports the joint test of pre-policy event-period coefficients, but given the heterogeneity contamination risk pointed out by Sun and Abraham [21], the TWFE event study is only auxiliary. Robustness checks also include the effective year of the approval year, including 2022 cities as the not-yet-treated group, exact matching only, 1%/99% winsorization, excluding municipalities directly under the central government, province x year fixed effects, and false policy timing.

## 5. Empirical results

### 5.1. Descriptive statistics

Table 1 shows that the mean of the urban export resilience index of the benchmark sample cities is 0.0959, the standard deviation is 0.0546, and the maximum is 0.3675, with relatively obvious differences between cities. Treatment group observations account for 44.16%, and post-policy observations account for 10.79%. The proportion of post-policy observations is relatively limited, mainly because the treatment periods of the 2019 and 2020 batches are short, which also means that long-term dynamic effects are more contributed by early batches.

Table 1. Descriptive statistics (benchmark sample)

| Variable                | N    | Mean   | Std. Dev. | Min    | Max    |
|-------------------------|------|--------|-----------|--------|--------|
| Urban export resilience | 3003 | 0.0959 | 0.0546    | 0.0230 | 0.3675 |
| Treatment group         | 3003 | 0.4416 | 0.4967    | 0      | 1      |
| Post-policy             | 3003 | 0.1079 | 0.3103    | 0      | 1      |
| DID                     | 3003 | 0.1079 | 0.3103    | 0      | 1      |

*Note: The treatment group and post-policy variables are both 0-1 dummy variables.*

### 5.2. Benchmark estimates

Table 2 reports the benchmark estimation results. The TWFE coefficient is 0.00734, about 7.7% of the sample mean of the resilience index; the average ATT of CSDID using the not-yet-treated group and the never-treated group is 0.00696 and 0.00720 respectively, and the estimation direction and magnitude are relatively close. The above results show that after the establishment of comprehensive pilot zones, the comprehensive export resilience index of pilot cities shows a significant increase, providing preliminary evidence for the positive relationship between the policy and the improvement of comprehensive resilience. Considering the phased implementation of pilots and city selectivity, the following further examines the interpretation boundary of this result through dynamic effects and parallel trend tests.

Table 2. Benchmark estimation results

| Estimator                 | Coefficient | Std. Error | p-value | N    |
|---------------------------|-------------|------------|---------|------|
| TWFE (city, year FE)      | 0.00734     | 0.00283    | 0.010   | 3003 |
| CSDID ATT (not-yet)       | 0.00696     | 0.00192    | <0.001  | 3003 |
| CSDID ATT (never-treated) | 0.00720     | 0.00201    | <0.001  | 3003 |

*Note: TWFE standard errors are clustered at the city level; CSDID uses the dripw estimator.*

### 5.3. Dynamic effects and parallel trend tests

Table 3 gives the joint pre-policy trend test. Both the benchmark not-yet caliber and the never-treated group caliber show that some policy batches already had significant differences before formal approval. After excluding Hangzhou and the first batch, the test statistic dropped significantly, but the pre-policy difference still did not completely disappear; similar conclusions are also obtained after shortening the event window and adopting the pre-policy fixed entropy-weight

index. This indicates that the first-mover foundation and advance layout of early pilot cities may affect the policy coefficient, so this paper mainly interprets the estimation results as the policy-related effect between the construction of comprehensive pilot zones and the change of urban export resilience.

Table 3. Joint pre-policy trend test

| Test specification                           | Statistic | df    | p-value | Judgment      |
|----------------------------------------------|-----------|-------|---------|---------------|
| CSDID: not-yet treated group                 | 357.75    | 38    | <0.0001 | Reject        |
| CSDID: never-treated group                   | 372.30    | 38    | <0.0001 | Reject        |
| CSDID: excluding Hangzhou/first batch        | 68.07     | 33    | 0.0003  | Reject        |
| CSDID: event window [-3,-1]                  | 84.55     | 15    | <0.0001 | Reject        |
| CSDID: pre-policy fixed entropy-weight index | 746.63    | 38    | <0.0001 | Reject        |
| TWFE: full pre-policy window                 | F=0.99    | 4,230 | 0.414   | Do not reject |
| TWFE: pre-policy terms of window [-3,2]      | F=0.69    | 2,230 | 0.502   | Do not reject |

*Note: The null hypothesis of CSDID is that the relevant pre-policy  $ATT(g,t)$  jointly equals zero. The dynamic coefficients of TWFE may be contaminated under staggered implementation and heterogeneous treatment effects, so they are only used as auxiliary evidence.*

Item-by-item results show that the significant pre-policy differences mainly come from the 2012, 2013, and 2015 of the 2016 group (Hangzhou), and the 2017 of the 2020 group and the 2012 of the 2021 group. Hangzhou is the only first-batch pilot; its economic openness, digital infrastructure, and policy readiness differ greatly from ordinary cities, and after exclusion the joint test statistic drops significantly, indicating that Hangzhou is indeed an important source; but the test still rejects, also showing that the problem cannot be fully attributed to one city. On the other hand, the joint test of TWFE pre-policy coefficients is not significant, but the two types of event studies use different comparison and weighting methods, so this result cannot offset the identification diagnosis of CSDID.

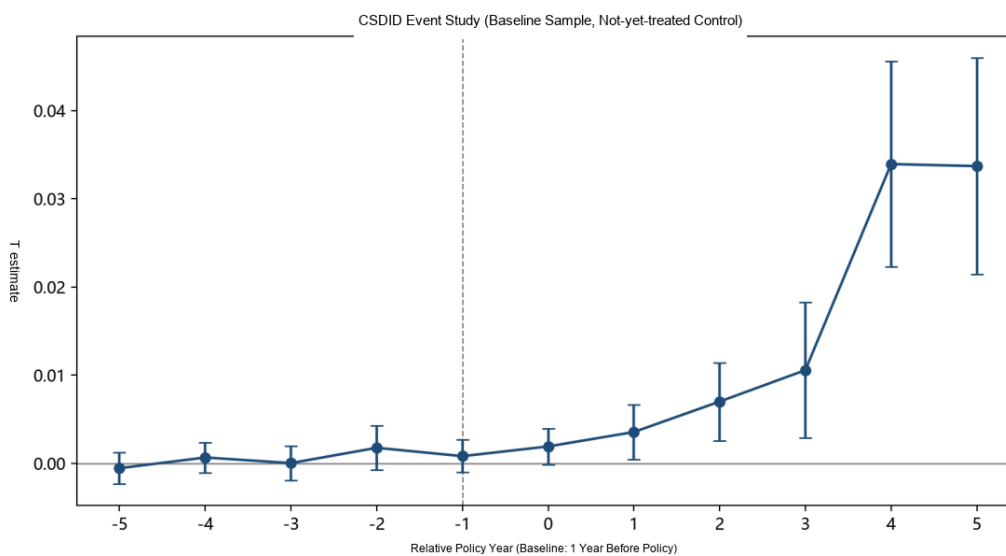


Figure 1. CSDID Event study results (base period is the year before policy)

As shown in Figure 1, the average pre-policy event-period coefficient is close to zero, but the event-period coefficients of different groups have positive and negative differences that cancel out after aggregation and averaging, which explains the phenomenon that the Pre\_avg statistic is not significant while the joint test is significant. The coefficients of the 4th and 5th years after the policy show a clear upward trend, and these long-term effects mainly come from the early treatment groups, among which the first batch treatment group contains only one city, Hangzhou.

#### 5.4. Supplementary tests and result stability

Table 4 summarizes the supplementary test results. After excluding Hangzhou city, the average coefficients of TWFE and CSDID are still significantly positive; after recalculating the entropy weights using the pre-policy implementation sample from 2010 to 2014, the estimated coefficient of the alternative index also remains positive. After adding province x year fixed effects, adjusting the policy effective time, including the newly added pilots in 2022, retaining only the exact matching sample, winsorizing, and excluding municipalities directly under the central government, the direction of the TWFE coefficient is generally consistent. The above results show that the positive change of the composite index is not driven by a single city, weight setting, or sample processing method, but its causal interpretation still needs to be judged in combination with the pre-policy trend.

Table 4. Key robustness results

| Test                                        | Coefficient | Std. Error | p-value | N    |
|---------------------------------------------|-------------|------------|---------|------|
| TWFE excluding Hangzhou                     | 0.00717     | 0.00287    | 0.013   | 2990 |
| CSDID excluding Hangzhou/first batch        | 0.00662     | 0.00193    | 0.001   | 2990 |
| TWFE pre-policy fixed entropy-weight index  | 0.01796     | 0.00356    | <0.001  | 3003 |
| CSDID pre-policy fixed entropy-weight index | 0.01311     | 0.00270    | <0.001  | 3003 |

*Note: The CSDID joint pre-policy test of the alternative index still significantly rejects parallel trends.*

#### 5.5. Decomposition of component indicators

Table 5 analyzes the 12 component indicators as dependent variables respectively. The results show that GDP, per capita disposable income of residents, financial development, and science and technology level all increase significantly, while export trade, foreign trade dependence, and some infrastructure indicators do not show synchronous improvement. The distance to the port, as a time-invariant variable, is absorbed by city fixed effects. Overall, the positive change of the composite index is mainly reflected in the enhancement of the city's economic base and trade support capacity.

Table 5. Decomposition of component indicators of the urban export resilience index

| Component indicator      | DID coefficient        | Std. Error             | p-value | N    |
|--------------------------|------------------------|------------------------|---------|------|
| Regional GDP (raw unit)  | 2.479x10 <sup>-7</sup> | 3.661x10 <sup>-6</sup> | <0.001  | 3003 |
| Urbanization rate        | -0.01825               | 0.00609                | 0.003   | 3003 |
| Per capita road area     | -2.188                 | 0.5791                 | <0.001  | 3003 |
| Foreign trade dependence | -0.04074               | 0.02163                | 0.061   | 3003 |
| Export trade             | -0.02928               | 0.01430                | 0.042   | 3003 |

Table 5. (continued)

|                                           |                  |          |        |      |
|-------------------------------------------|------------------|----------|--------|------|
| Export deviation                          | -1.635           | 1.386    | 0.239  | 3003 |
| Financial development level               | 0.1071           | 0.03859  | 0.006  | 3003 |
| Per capita disposable income of residents | 4818             | 508.9    | <0.001 | 3003 |
| Fiscal investment intensity               | -1.048           | 0.4308   | 0.016  | 3003 |
| Science and technology level              | 0.004746         | 0.001572 | 0.003  | 3003 |
| Distance to nearest port                  | Not identifiable | -        | -      | 3003 |
| Foreign capital dependence                | -0.001595        | 0.002567 | 0.535  | 2977 |

*Note: Each column is a separately estimated TWFE result, with standard errors clustered at the city level. The original indicators have different dimensions, and the coefficients are mainly used to judge the direction and driving source, and should not be directly compared in absolute size.*

The above decomposition shows that the short-term effect of comprehensive pilot zones may first be reflected in capacity building such as financial services, technological innovation, residents' income, and economic base, and there is a certain time lag in the transformation of these conditions into export scale, stability, and shock recovery capacity. Therefore, the evaluation of policy effects needs to distinguish the two levels of 'resilience capacity building' and 'direct trade performance,' and make judgments in combination with export market diversification, product structure, and recovery performance under clear external shocks.

## 5.6. Alternative outcome variables

Table 6 reports the direct trade results. The estimates of the policy on ln export value, export growth rate, three-year rolling volatility, export resistance, and export deviation all fail to reach the commonly used significance level, and foreign trade dependence shows a marginal negative change. Considering that the post-policy observation period of the 2019 and 2020 batches is short, and the later sample simultaneously experiences global supply chain interruptions and demand reorganization, the capacity-building effect of comprehensive pilot zones may not yet be fully transformed into direct export performance.

Table 6. Alternative outcome variables (TWFE)

| Dependent variable                   | Coefficient | Std. Error | p-value | N    |
|--------------------------------------|-------------|------------|---------|------|
| ln(export value+1)                   | -0.055998   | 0.063592   | 0.379   | 2986 |
| Export growth rate                   | 0.024233    | 0.025550   | 0.344   | 2744 |
| Export growth volatility (3-year SD) | 0.040520    | 0.029874   | 0.176   | 2261 |
| Export resistance                    | 0.063428    | 0.047165   | 0.180   | 2261 |
| Export deviation                     | -1.635449   | 1.386464   | 0.239   | 3003 |
| Foreign trade dependence             | -0.040745   | 0.021633   | 0.061   | 3003 |

*Note: All models absorb city and year fixed effects, with standard errors clustered at the city level. Foreign trade dependence is only used as an alternative outcome variable and does not enter the control variables of the composite index regression.*

## 6. Further discussion

### 6.1. Stage differences between capacity building and trade performance

The difference between the composite index and the direct export results is not contradictory. What comprehensive pilot zones may first improve is digital infrastructure, financial services, logistics systems, and the business environment; these conditions can enhance firms' ability to enter new markets and adjust supply chains, but their transformation into export growth and stability takes time. Especially the 2020 batch has only two complete post-policy years of 2021-2022, and this period simultaneously experiences global supply chain interruptions and demand reorganization. Therefore, the current results are more suitable to be interpreted as 'capacity building first, trade performance transformation relatively lagging.'

### 6.2. Pilot selection, advance layout and result interpretation

The approval of comprehensive pilot zones has a certain selectivity. Early pilots usually have a more mature foundation of foreign trade, logistics, e-commerce, and public services, and local governments may also carry out park construction, platform building, and policy preparation before formal approval. The pre-policy difference of Hangzhou is especially prominent, and the joint test statistic drops significantly after exclusion; some differences still exist in other batches, indicating that pilot selection and advance layout may jointly affect the estimation results. Subsequent research can further collect local implementation plans, platform launch, and key institutional landing time to more accurately characterize the actual policy start-up process.

In addition, policies such as the free trade pilot zone, comprehensive bonded zone, national e-commerce demonstration city, 'Broadband China,' innovative city, and logistics hub may overlap with comprehensive pilot zones and jointly affect GDP, technology, finance, and export performance. This paper conducts preliminary control by adding province  $\times$  year fixed effects; subsequent research still needs to further establish a city-level list of competing policies to improve the comparability between the treatment group and the control group.

### 6.3. Interpretation of results and applicability boundary

Combining the various results, this paper can confirm that pilot cities show a higher composite index before and after policy implementation, and this correlation remains stable under several model settings; but the pre-policy differences indicate that the coefficient may simultaneously contain the joint influence of pilot selection, advance layout, and policy implementation. Therefore, this paper interprets it as the policy-related effect between the construction of comprehensive pilot zones and the improvement of urban comprehensive export resilience, and focuses on discussing its economic composition and stage characteristics.

### 6.4. Research limitations and future prospects

This paper still has three limitations. First, the sample period ends in 2022, and the post-policy observation of late pilots is short; second, the approval year is difficult to fully reflect the real time of local platform launch and institutional implementation; third, the urban composite index cannot yet replace the direct resilience indicators at the firm, product, and destination country levels. In the future, longer time series, competing policy controls, and firm-product-destination data can be used

to further examine mechanisms such as firm entry, market diversification, supply chain recovery, and product structure upgrading.

## 7. Conclusions and policy implications

Based on data of 282 prefecture-level cities from 2010 to 2022, this paper examines the relationship between the construction of cross-border e-commerce comprehensive pilot zones and urban export resilience. Both TWFE and CSDID estimates show that the comprehensive export resilience index of pilot cities increases significantly, but the decomposition of indicators shows that this change mainly comes from the improvement of support conditions such as GDP, residents' income, financial development, and science and technology level, and the direct trade results such as export scale, growth stability, and shock resistance have not yet shown obvious changes. Thus, the short-term effect of comprehensive pilot zones may be more reflected in the capacity building of export resilience foundation, and its transformation into direct trade performance has a certain time lag.

From the perspective of research interpretation, some treatment batches already showed a strong development foundation before formal approval, indicating that pilot selection and advance layout may affect the policy coefficient. Based on this, this paper understands the results as the policy-related effect between the construction of comprehensive pilot zones and the improvement of urban export resilience, rather than the net causal effect after completely excluding other factors. In the future, the sample period can be extended, the actual implementation time refined, competing policies included, and more targeted export structure and shock response indicators used for further testing.

From the perspective of policy practice, the performance evaluation of comprehensive pilot zones should distinguish three levels: institution and capacity building, enterprise behavior change, and final trade results. While continuing to improve customs clearance, cross-border logistics, digital credit, and financial services, the addition of new export enterprises, destination country expansion, order stability, and supply chain recovery time should be tracked as a focus. For inland cities and later-batch pilots, differentiated policies should be implemented in combination with industrial base, port conditions, and target markets, to promote the more effective transformation of institutional investment into stable, diversified, and recoverable export performance.

The sample period of this paper is short, the policy intensity is only proxied by the approval year, and the simultaneous policies at the city level have not yet been fully controlled. The above limitations do not change the core finding of this paper that 'there is a stage difference between the capacity-building effect and the short-term trade performance,' but suggest that subsequent research needs to use longer time series and finer-grained data for further verification.

## References

- [1] Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1), 3-43. <https://doi.org/10.1257/jel.20171452>
- [2] Lendle, A., Olarreaga, M., Schropp, S., & Vézina, P. L. (2016). There goes gravity: eBay and the death of distance. *The Economic Journal*, 126(591), 406-441. <https://doi.org/10.1111/eoj.12286>
- [3] World Trade Organization. (2018). *World trade report 2018: The future of world trade—How digital technologies are transforming global commerce*. WTO.
- [4] Ma, S. Z., & Guo, J. W. (2022). How institutional innovation affects China's cross-border e-commerce exports? Empirical evidence from the establishment of comprehensive pilot zones. *Management World*, 38(8), 83-102.
- [5] Wang, R., Gu, Q. Y., & Zhong, B. P. (2020). What kind of trade facilitation does cross-border e-commerce need? Evidence from 35 cross-border e-commerce comprehensive pilot zones in China. *Zhejiang Academic Journal*, (4), 100-110.

- [6] Martin, R., & Sunley, P. (2015). On the notion of regional economic resilience: Conceptualization and explanation. *Journal of Economic Geography*, 15(1), 1-42. <https://doi.org/10.1093/jeg/lbu015>
- [7] Boschma, R. (2015). Towards an evolutionary perspective on regional resilience. *Regional Studies*, 49(5), 733-751. <https://doi.org/10.1080/00343404.2014.959481>
- [8] Hu, H. R., & Song, Y. Q. (2022). Cross-border e-commerce reform and wage income: A new openness perspective. *Journal of Finance and Economics*, 48(5), 49-63.
- [9] Peng, H. A., & Zhang, C. M. (2024). Cross-border e-commerce comprehensive pilot zones and industrial structure upgrading: Empirical analysis based on urban panel data. *Journal of Hainan University (Social Sciences Edition)*, 42(3), 106-115. <https://doi.org/10.15886/j.cnki.hnus.202303.0233>
- [10] Zhang, B. B., Chen, Y. J., Zhu, J., & Yan, Z. J. (2023). Cross-border e-commerce comprehensive pilot zones and regional coordinated development: Window radiation or siphon effect. *Journal of Finance and Economics*, 49(7), 34-47.
- [11] Wang, L. R., & Rui, L. L. (2022). The impact and heterogeneity analysis of cross-border e-commerce comprehensive pilot zones on regional economy: Based on the counterfactual perspective. *South China Journal of Economics*, (3), 53-73.
- [12] Wang, W. Y., Ren, Z. R., Li, W., & He, C. F. (2021). Trade barriers, market-related variety and urban export resilience. *Geographical Research*, 40(12), 3287-3301. <https://doi.org/10.11821/dlyj020210162>
- [13] Martin, R., Sunley, P., Gardiner, B., & Tyler, P. (2016). How regions react to recessions: Resilience and the role of economic structure. *Regional Studies*, 50(4), 561-585. <https://doi.org/10.1080/00343404.2015.1136410>
- [14] de Chaisemartin, C., & D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964-2996. <https://doi.org/10.1257/aer.20181169>
- [15] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254-277. <https://doi.org/10.1016/j.jeconom.2021.03.014>
- [16] Freund, C. L., & Weinhold, D. (2004). The effect of the Internet on international trade. *Journal of International Economics*, 62(1), 171-189. [https://doi.org/10.1016/S0022-1996\(03\)00059-X](https://doi.org/10.1016/S0022-1996(03)00059-X)
- [17] Blum, B. S., & Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, 70(2), 384-405. <https://doi.org/10.1016/j.jinteco.2005.10.002>
- [18] Simmie, J., & Martin, R. (2010). The economic resilience of regions: Towards an evolutionary approach. *Cambridge Journal of Regions, Economy and Society*, 3(1), 27-43. <https://doi.org/10.1093/cjres/rsp029>
- [19] Hassink, R. (2010). Regional resilience: A promising concept to explain differences in regional economic adaptability? *Cambridge Journal of Regions, Economy and Society*, 3(1), 45-58. <https://doi.org/10.1093/cjres/rsp033>
- [20] Martin, R. (2012). Regional economic resilience, hysteresis and recessionary shocks. *Journal of Economic Geography*, 12(1), 1-32. <https://doi.org/10.1093/jeg/lbr019>
- [21] Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175-199. <https://doi.org/10.1016/j.jeconom.2020.09.006>
- [22] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200-230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- [23] Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6), 3253-3285. <https://doi.org/10.1093/restud/rdae007>
- [24] Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3), 305-322. <https://doi.org/10.1257/aeri.20210236>
- [25] Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5), 2555-2591. <https://doi.org/10.1093/restud/rdad018>

## Appendix: caliber and purpose of main variables

The figures in this paper come from the batch processing results of Stata/SE 16.1; the policy matching uses city codes and the earliest treatment year, and the 2022 batch is not included in the benchmark causal estimation because of the lack of post-policy observations. Table A1 shows the analytical caliber of the main variables.

Table A1. Caliber and purpose of main variables

| Variable                           | Calculation or definition                                                                                                      | Purpose                            |
|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| Composite export resilience index  | Synthesized from export performance and urban support-condition indicators; higher value means stronger comprehensive capacity | Benchmark outcome variable         |
| $\ln(\text{export value}+1)$       | Natural logarithm of city annual export value plus 1                                                                           | Direct trade result                |
| Export growth rate                 | Year-on-year change rate of export value                                                                                       | Direct trade result                |
| Export growth volatility           | Three-year rolling standard deviation of export growth rate                                                                    | Stability test                     |
| Export resistance                  | City export change relative to national shock resistance performance                                                           | Shock response test                |
| Export deviation                   | Actual export deviation from its own trend path                                                                                | Trend deviation test               |
| Foreign trade dependence           | Ratio of total import and export to regional GDP                                                                               | Alternative outcome variable       |
| Treatment group and treatment year | Whether and when the city was approved (earliest)                                                                              | Policy identification variable     |
| Event time                         | Observation year minus the city earliest approval year; year before policy as base                                             | Dynamic effect and trend diagnosis |