Factor Model and Program Evaluation

Beyond DID

Jie Wei[a]

[a]School of Economics, Huazhong University of Science and Technology

School of Economics, HUST, Wuhan

Nov 23, 2017

Table of Contents

Reference

- Hsiao, C., Ching, H.S., Wan, S.K., 2012. A panel data approach for program evaluation: Measuring the benefit of political and economic integration of Hong Kong with mainland China. J. Appl. Econom. 27, 705-740.
- Du, Z., Zhang, L., 2015. Home-Purchase restriction, property tax and housing price in China: A counterfactual analysis. J. Econometrics 188, 558-568.
- Li, K., Bell, D., 2017. Estimation of average treatment effects with panel data: Asymptotic theory and implementation. J. Econometrics 197, 65-75.

Advantage of the Factor Model in Treatment Effect

- Flexible in modelling cross section dependence.
- No need to estimate estimate the number of factors, factors or factor loadings.
- Easy and fast to implement: just OLS!

The General Question in Program Evaluation

- Observation: $\{d_{it}, Y_{it}\}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$.
  $d_{it} = 1$ if the $i$the unit is under treatment at time t, and $d_{it} = 0$ otherwise. Y_{it}: outcome variable.
- Let Y_{it}^1 and Y_{it}^0 denote unit i's outcome in period t with and without treatment, respectively. We cannot simultaneously observe both. Thus, the observed data is in the form $Y_{it} = d_{it} Y_{it}^1 + (1 - d_{it}) Y_{it}^0$.
- Our point of interest: Treatment Effect as

$$\Delta_{it} \equiv Y_{it}^1 - Y_{it}^0$$

  Or, at least, the Average Treatment Effect (ATE).
- Difficulty: We need to address (model) unobserved heterogeneity appropriately; otherwise, estimate of ATE would be BIASED.

An Easy Tackle: DID under Fixed Effects

Heuristically, we assume $N = 2$ and $T = 2$. In particular, at time t (pre-treatment), no one receives treatment; at time s (post-treatment), unit 1 receives treatment while unit 2 does not.

Thus, we observe: $\{Y_{1t}^0, Y_{2t}^0, Y_{1s}^1, Y_{2s}^0\}$, and impose the following structure

$$\begin{aligned} Y_{1t}^0 &= \alpha_1 + \lambda_t + u_{1t} \\ Y_{2t}^0 &= \alpha_2 + \lambda_t + u_{2t} \\ Y_{1s}^1 &= \alpha_1 + \lambda_s + \Delta_{1s} + u_{1s} \\ Y_{2s}^0 &= \alpha_2 + \lambda_s + u_{2s} \end{aligned}$$

Naturally,

$$\widehat{\Delta}_{1s} = \left(Y_{1s}^1 - Y_{1t}^0\right) - \left(Y_{2s}^0 - Y_{2t}^0\right)$$

will be consistent if $N \to \infty$

Table of Contents

A Detour: a brief introduction of factor models

- The basic model

$$Y_{it} = X_{it}'\beta + b_i'f_t + u_{it} \tag{1}$$

  b_i $(K \times 1)$ is a vector of (unobserved) factor loadings, and f_t $(K \times 1)$ is a vector of (unobserved) factors, so that $b_i'f_t = b_{i1}f_{1t} + \ldots + b_{iK}f_{Kt}$. u_{it} is idiosyncratic error.

- Interpretation of f_t and b_i, see Bai (2009, Econometrica).
- Note when f_t is constant, model (1) turns to an individual fixed effect; when b_i is constant, model (1) turns to a time fixed effect; when

$$f_t = \begin{pmatrix} 1 \\ \lambda_t \end{pmatrix}, \quad b_i = \begin{pmatrix} \alpha_i \\ 1 \end{pmatrix},$$

  $b_i'f_t = \alpha_i + \lambda_t$, specializing the two-way fixed effect model.

Heuristics

Again, assume $N = 2$ and $T = 2$.

Pre-treatment at time t,

$$Y_{1t}^0 = \alpha_1 + b_1' f_t + u_{1t} \tag{2}$$

$$Y_{2t}^0 = \alpha_2 + b_2' f_t + u_{2t} \tag{3}$$

Post-treatment at time s,

$$Y_{1s}^1 = \alpha_1 + b_1' f_s + \Delta_{1s} + u_{1s} \tag{4}$$

$$Y_{2s}^0 = \alpha_2 + b_2' f_s + u_{2s} \tag{5}$$

Ideally, we wish to have Y_{1s}^0 so that $\Delta_{1s} = Y_{1s}^1 - Y_{1s}^0$, and then job is done.

Heuristics

Again, assume $N = 2$ and $T = 2$.

Pre-treatment at time t,

$$Y_{1t}^0 = \alpha_1 + b_1' f_t + u_{1t} \tag{2}$$

$$Y_{2t}^0 = \alpha_2 + b_2' f_t + u_{2t} \tag{3}$$

Post-treatment at time s,

$$Y_{1s}^1 = \alpha_1 + b_1' f_s + \Delta_{1s} + u_{1s} \tag{4}$$

$$Y_{2s}^0 = \alpha_2 + b_2' f_s + u_{2s} \tag{5}$$

Ideally, we wish to have Y_{1s}^0 so that $\Delta_{1s} = Y_{1s}^1 - Y_{1s}^0$, and then job is done.

The bad news is Y_{1s}^0 is NOT directly observed. The good news is we know that

$$Y_{1s}^0 = \alpha_1 + b_1' f_s + u_{1s} \tag{6}$$

so it is hopeful to estimate Y_{1s}^0 with information available.

The Main Idea

For simplicity, we further assume that $K = 1$. Then, by $(6) - (5) \times \frac{b_1}{b_2}$,

$$Y_{1s}^0 - \frac{b_1}{b_2} Y_{2s}^0 = \alpha_1 - \frac{b_1}{b_2}\alpha_2 + u_{1s} - \frac{b_1}{b_2} u_{2s}$$

Or equivalently,

$$Y_{1s}^0 = \left(\alpha_1 - \frac{b_1}{b_2}\alpha_2\right) + \frac{b_1}{b_2} Y_{2s}^0 + u_{1s} - \frac{b_1}{b_2} u_{2s} \tag{7}$$

Meanwhile, by $(2) - (3) \times \frac{b_1}{b_2}$, we find that at pretreatment

$$Y_{1t}^0 = \left(\alpha_1 - \frac{b_1}{b_2}\alpha_2\right) + \frac{b_1}{b_2} Y_{2t}^0 + u_{1t} - \frac{b_1}{b_2} u_{2t} \tag{8}$$

The Main Idea

For simplicity, we further assume that $K = 1$. Then, by $(6) - (5) \times \frac{b_1}{b_2}$,

$$Y_{1s}^0 - \frac{b_1}{b_2} Y_{2s}^0 = \alpha_1 - \frac{b_1}{b_2}\alpha_2 + u_{1s} - \frac{b_1}{b_2} u_{2s}$$

Or equivalently,

$$Y_{1s}^0 = \left(\alpha_1 - \frac{b_1}{b_2}\alpha_2\right) + \frac{b_1}{b_2} Y_{2s}^0 + u_{1s} - \frac{b_1}{b_2} u_{2s} \tag{7}$$

Meanwhile, by $(2) - (3) \times \frac{b_1}{b_2}$, we find that at pretreatment

$$Y_{1t}^0 = \left(\alpha_1 - \frac{b_1}{b_2}\alpha_2\right) + \frac{b_1}{b_2} Y_{2t}^0 + u_{1t} - \frac{b_1}{b_2} u_{2t} \tag{8}$$

If we could collect data from multiple periods, say T_1, of pre-treatment, then equation (8) holds for $t = 1, \ldots, T_1$. So it is likely we can get a good estimator for $\left(\alpha_1 - \frac{b_1}{b_2}\alpha_2\right)$ and $\frac{b_1}{b_2}$, as intercept and slope, and thus derive $\widehat{Y_{1s}^0}$ from (7), and thus come to $\widehat{\Delta}_{1s} = Y_{1s}^1 - \widehat{Y_{1s}^0}$.

A Formal Framework

- Basics. Total periods: T. Pretreatment: $t = 1, \ldots, T_1$; post-treatment: $t = T_1 + 1, \ldots, T$. Individual 1 receives treatment at $t = T_1 + 1$, while others $j = 2, \ldots, N$ do not receive any treatment.
- Before treatment

$$Y_{it}^0 = \alpha_i + b_i' f_t + u_{it}, \quad i = 1, \ldots, N; \; t = 1, \ldots, T_1$$

Stacking $i = 1, \ldots, N$,

$$Y_t = Y_t^0 = \alpha + B f_t + u_t, \quad t = 1, \ldots, T_1 \tag{9}$$

where $Y_t = (Y_{1t}, \ldots, Y_{Nt})'$, $\alpha = (\alpha_1, \ldots, \alpha_N)'$, $B = (b_1, \ldots, b_N)'$ is $N \times K$, and $u_t = (u_{1t}, \ldots, u_{Nt})'$.

Also, define $\widetilde{Y}_t = (Y_{2t}, \ldots, Y_{Nt})'$ such that $Y_t = \left(Y_{1t}, \widetilde{Y}_t'\right)'$. Analogue definitions apply for $\widetilde{\alpha}$ and $\widetilde{u}_t$.

A Simple Derivation

Let $a = (1, -\gamma')'$, where a is an $N \times 1$ vector lying in the null space of B, $N(B)$, i.e., $a'B = 0$. The vector of a in the traditional fixed effect model is quite intuitive: $a = \left(1, -\frac{1}{N-1}, -\frac{1}{N-1}, \dots\right)'$.

Then by the definition of Y_t and equation (9),

$$Y_{1t} - \gamma' \widetilde{Y}_t = a' Y_t = a'\alpha + a' u_t$$

Or equivalently,

$$\begin{aligned} Y_{1t} =& a'\alpha + \gamma' \widetilde{Y}_t + a' u_t && (10) \\ \equiv& \gamma_1 + \gamma' \widetilde{Y}_t + u_{1t}^* && (11) \end{aligned}$$

where $\gamma_1 = a'\alpha$ and $u_{1t}^* = a' u_t = u_{1t} - \gamma' \widetilde{u}_t$.

Main Assumptions

- **Assumption 1** (i) $||b_i|| \leq M < \infty$ for all $i = 1, \ldots, N$; (ii) u_t is a weakly dependent process with $E(u_t) = 0$ and $E(u_t u_t') = V$, where V is an $N \times N$ diagonal matrix; (iii) $E(u_t f_t') = 0$; (iv) $E(u_{jt}|d_{1t}) = 0$ for all $j \neq 1$.
- **Assumption 2** (i) Let $x_t = (1, \widetilde{Y}_t')'$. Then $\{x_t\}_{t=1}^{T}$ is a weakly dependent and weakly stationary process. $T_1^{-1} \sum_{t=1}^{T_1} x_t x_t' \xrightarrow{p} E(x_t x_t')$ as $T_1 \to \infty$, and $E(x_t x_t')$ is invertible. (ii) $Rank(\widetilde{B}) = K$.
- **Assumption 3** (made in HCW, not necessary) $E(u_{1t}^*) = c_1 + c' \widetilde{Y}_t$.

Main Assumptions

- **Assumption 1** (i) $||b_i|| \leq M < \infty$ for all $i = 1, \ldots, N$; (ii) u_t is a weakly dependent process with $E(u_t) = 0$ and $E(u_t u_t') = V$, where V is an $N \times N$ diagonal matrix; (iii) $E(u_t f_t') = 0$; (iv) $E(u_{jt} | d_{1t}) = 0$ for all $j \neq 1$.
- **Assumption 2** (i) Let $x_t = (1, \widetilde{Y}_t')'$. Then $\{x_t\}_{t=1}^{T}$ is a weakly dependent and weakly stationary process. $T_1^{-1} \sum_{t=1}^{T_1} x_t x_t' \xrightarrow{p} E(x_t x_t')$ as $T_1 \to \infty$, and $E(x_t x_t')$ is invertible. (ii) $Rank(\widetilde{B}) = K$.
- **Assumption 3** (made in HCW, not necessary) $E(u_{1t}^*) = c_1 + c' \widetilde{Y}_t$.

Note Assumption 3 implies a decomposition

$$u_{1t}^* = E(u_{1t}^* | \widetilde{Y}_t) + u_{1t}^* - E(u_{1t}^* | \widetilde{Y}_t) \tag{12}$$

$$= c_1 + c' \widetilde{Y}_t + \eta_{1t} \tag{13}$$

where $E(\eta_{1t} \mid \widetilde{Y}_t) = 0$.

The Main Result of ATE

- Now plugging (13) into (11),

$$\begin{aligned} Y_{1t} =& (\gamma_1 + c_1) + (\gamma + c)' \widetilde{Y}_t + \eta_{1t} \\ =& \delta_1 + \delta' \widetilde{Y}_t + \eta_{1t}, \quad t = 1, \ldots, T_1 \end{aligned}$$

  Without Assumption 3, we can come to a slightly different result

$$Y_{1t} = \delta_1 + \delta' \widetilde{Y}_t + v_{1t}, \quad t = 1, \ldots, T_1 \tag{14}$$

  where $Cov(\widetilde{Y}_t, v_{1t}) = 0$. By OLS (over a time series of length T_1), we can obtain $\widehat{\delta}_1$ and $\widehat{\delta}$, consistently.

- For post-treatment $s = T_1 + 1, \ldots, T$, extrapolate $\widehat{Y}_{1s}^0 = \widehat{\delta}_1 + \widehat{\delta}' \widetilde{Y}_s$. With the observation Y_{1s}: $Y_{1s} = \delta_1 + \delta' \widetilde{Y}_s + \Delta_{1s} + v_{1s}$, a sensible estimator for treatment effect is $\widehat{\Delta}_{1s} = Y_{1s} - \widehat{Y}_{1s}^0$, and for ATE,

$$\widehat{\Delta}_1 = \frac{1}{T_2} \sum_{s=T_1+1}^{T} \widehat{\Delta}_{1s}$$

Selecting significant control units

1. Hsiao, Ching, and Wan (2012): AIC or AICC
2. Du and Zhang(2015): Leave-n_v-out cross validation.
3. Li and Bell(2017): Least absolute shrinkage and selection operator (LASSO).

   Consider a linear regression model

   $$Y_{1t} = x_t'\beta + v_{1t}, \quad t = 1, \ldots, T_1,$$

   where $x_t' = (1, \widetilde{Y}_t')$ and $\beta = (\delta_1, \delta')'$ is an $N \times 1$ unknown vector. The LASSO selects β to minimize

   $$\sum_{t=1}^{T_1} [Y_{1t} - x_t'\beta]^2 + \lambda \sum_{j=1}^{N} |\beta_j|$$

   - Works even if $N > T_1$.
   - Computationally efficient: 10^{-1}s v.s. 10^2s
   - has smaller out-of-sample prediction errors.

Consistency

Let $\Delta_1 = E(\Delta_{1s})$ be the ATE for the first unit (under weak stationary). By Assumption 1-3 in Li and Bell (2017),

$$\widehat{\Delta}_1 - \Delta_1 = O_p(T_1^{-1/2} + T_2^{-1/2})$$

- The consistency still holds for trend-stationary processes.
- The result above is actually quite intuitive. The consistency rates for ATE reflects in-sample and out-of-sample estimation accuracy. Recall that $\widehat{Y}_{1s}^0 = \widehat{\delta}_1 + \widehat{\delta}' \widetilde{Y}_s$, $Y_{1s} = \delta_1 + \delta' \widetilde{Y}_s + \Delta_{1s} + v_{1s}$, $\widehat{\Delta}_{1s} = Y_{1s} - \widehat{Y}_{1s}^0$, and for ATE, $\widehat{\Delta}_1 = \frac{1}{T_2} \sum_{s=T_1+1}^{T} \widehat{\Delta}_{1s}$.

The Asymptotic distribution

- **Assumption 4** Let $\beta = (\delta_1, \delta')'$ and the LS estimator $\widehat{\beta} = (\widehat{\delta}_1, \widehat{\delta'})'$. Then $\sqrt{T_1}(\widehat{\beta_1} - \beta_1) \xrightarrow{d} N(0, V_\beta)$.
- **Assumption 5** $T_2^{-1/2} \sum_{s=T_1+1}^{T} (\Delta_{1s} - E(\Delta_{1s}) + v_{1s}) \xrightarrow{d} N(0, \Sigma_2)$, where Σ_2 is the asymptotic variance of $T_2^{-1/2} \sum_{s=T_1+1}^{T} (\Delta_{1s} - E(\Delta_{1s}) + v_{1s})$.
- **Assumption 7** Let $\eta = \lim_{T_1, T_2 \to \infty} \frac{T_2}{T_1}$. We assume that $0 \leq \eta < \infty$.

The Asymptotic distribution

- **Assumption 4** Let $\beta = (\delta_1, \delta')'$ and the LS estimator $\widehat{\beta} = (\widehat{\delta}_1, \widehat{\delta}')'$. Then $\sqrt{T_1}(\widehat{\beta_1} - \beta_1) \xrightarrow{d} N(0, V_\beta)$.
- **Assumption 5** $T_2^{-1/2}\sum_{s=T_1+1}^{T}(\Delta_{1s} - E(\Delta_{1s}) + v_{1s}) \xrightarrow{d} N(0, \Sigma_2)$, where Σ_2 is the asymptotic variance of $T_2^{-1/2}\sum_{s=T_1+1}^{T}(\Delta_{1s} - E(\Delta_{1s}) + v_{1s})$.
- **Assumption 7** Let $\eta = \lim_{T_1, T_2 \to \infty} \frac{T_2}{T_1}$. We assume that $0 \leq \eta < \infty$.

Under Assumption 1-7,

$$\sqrt{T_2}(\widehat{\Delta}_1 - \Delta_1) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma = \Sigma_1 + \Sigma_2$, with $\Sigma_1 = \eta E(x_t)' V_\beta E(x_t)$. When $T_2/T_1 \to 0$, $\Sigma \approx \Sigma_2$.

We can easily estimate Σ_1. For Σ_2, one can estimate by Newey-West. With $\widehat{\Sigma}$, one can test significance of ATE.

Table of Contents

The Data and Background of Application 1

- Outcome variable: Hong Kong Economic growth (quarterly real GDP growth rate in use).
- $N = 24$ countries/regions: Australia, Austria, Canada, China, Denmark, Finland, France, Germany, Indonesia, Italy, Japan, Korea, Malaysia, Mexico, Netherlands, New Zealand, Norway, Philippines, Singapore, Switzerland, Taiwan, Thailand, the UK, and the USA.
- Two treatments (interventions): (i) reversion of sovereignty on 1 July 1997 from the UK to China, (ii) the implementation of Closer Economics Partnership Arrangement (CEPA) starting in 2004:Q1 between mainland China and Hong Kong.
- The observed data range from 1993:Q1 to 2008:Q1.
- We report the resulting with the selection criterion AIC.

The Impact of reversion of sovereignty

Table XVIII. AIC: weights of control groups for the period 1993:Q1 – 1997:Q2

	Beta	SD	T
Constant	0.0316	0.0164	1.9283
Japan	−0.69	0.1056	−6.5341
Korea	−0.3767	0.0688	−5.4721
USA	0.8099	0.2873	2.8193
Philippines	−0.1624	0.0999	−1.6248
Taiwan	0.6189	0.311	1.9902

$R^2 = 0.9438$; AIC $= -180.986$.

Table XIX. AIC: treatment effect of political integration 1997:Q3−2003:Q4

	Actual	Control	Treatment
1997:Q3	0.061	0.0839	−0.0229
1997:Q4	0.014	0.0811	−0.0671
1998:Q1	−0.032	0.1344	−0.1664
1998:Q2	−0.061	0.1438	−0.2048
1998:Q3	−0.081	0.1334	−0.2144
1998:Q4	−0.065	0.1472	−0.2122
1999:Q1	−0.029	0.0952	−0.1242
1999:Q2	0.005	0.0704	−0.0654
1999:Q3	0.039	0.0464	−0.0074
1999:Q4	0.083	0.0473	0.0357
2000:Q1	0.107	0.031	0.076
2000:Q2	0.075	0.0344	0.0406
2000:Q3	0.076	0.0394	0.0366
2000:Q4	0.063	0.0208	0.0422
2001:Q1	0.027	0.0155	0.0115
2001:Q2	0.015	−0.0101	0.0251
2001:Q3	−0.001	−0.0071	0.0061
2001:Q4	−0.017	0.0251	−0.0421
2002:Q1	−0.01	0.0375	−0.0475
2002:Q2	0.005	0.0473	−0.0423
2002:Q3	0.028	0.0593	−0.0313
2002:Q4	0.048	0.027	0.021
2003:Q1	0.041	0.0463	−0.0053
2003:Q2	−0.009	0.0302	−0.0392
2003:Q3	0.038	0.0593	−0.0213
2003:Q4	0.047	0.077	−0.03
Mean	0.018	0.0583	−0.0403
SD	0.0478	0.0435	0.0815
T	0.3761	1.3393	−0.4953

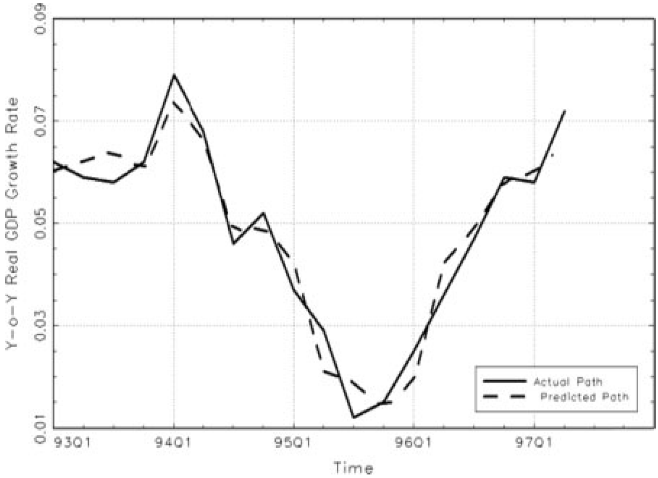


Figure 4. AIC: actual and predicted real GDP from 1993:Q1 to 1997:Q2

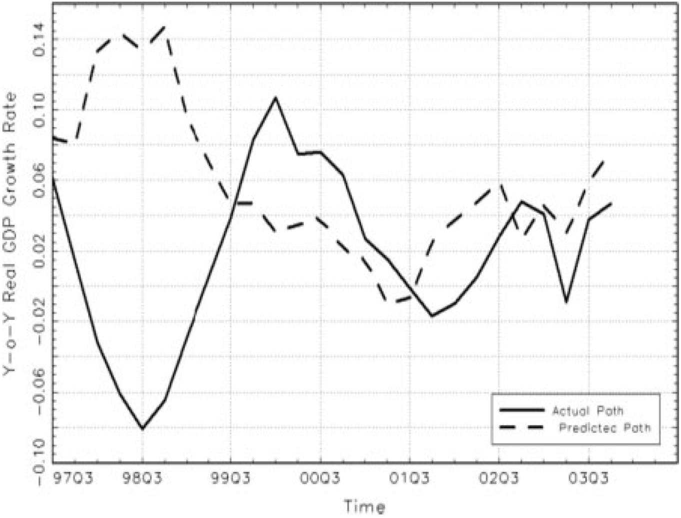


Figure 5. AIC: actual and counterfactual real GDP from 1997:Q3 to 2003:Q4

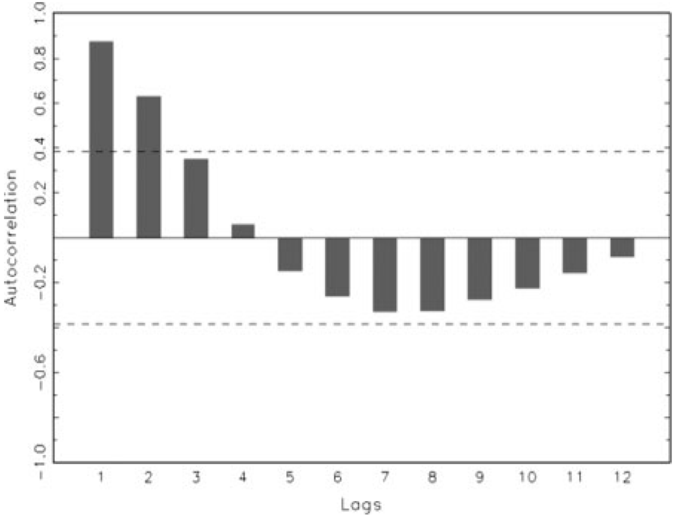


Figure 6. AIC: autocorrelation of treatment effect from 1997:Q3 to 2003:Q4

effects appear serially correlated (see Figure 6). The fitted AR(2) model takes the form

$$\hat{\Delta}_{1t} = \underset{(0.0078)}{-0.0066} + \underset{(0.1722)}{1.3821}\hat{\Delta}_{1,t-1} - \underset{(0.1722)}{0.5764}\hat{\Delta}_{1,t-2} + \hat{\eta}_t \tag{47}$$

The implied long-run effect is -0.033. However, the t-statistic is only -0.94, which is not statistically significant.

The lack of intervention effects is hardly surprising given the one country, two systems concept proposed by Deng Xiaoping. Moreover, the change of sovereignty was known 14 years in advance and the institutional arrangements were laid down in great detail in the Sino-British Joint Declaration of 1984. Presumably, all needed adjustments had already taken place before 1997.

The impact of CEPA

Table XXII. AIC: weights of control groups for the period 1993:Q1–2003:Q4

	Beta	SD	T
Constant	−0.003	0.0042	−0.7095
Austria	−1.2949	0.2181	−5.9361
Germany	0.3552	0.233	1.5243
Italy	−0.5768	0.1781	−3.2394
Korea	0.3016	0.0587	5.1342
Mexico	0.234	0.0609	3.8395
Norway	0.2881	0.0562	5.1304
Switzerland	0.2436	0.1729	1.4092
Singapore	0.2222	0.0553	4.0155
Philippines	0.1757	0.1089	1.6127

$R^2 = 0.9433$, AIC $= -385.7498$.

Table XXIII. AIC: treatment effect for economic integration 2004:Q1–2008:Q1

	Actual	Control	Treatment
2004:Q1	0.077	0.0559	0.0211
2004:Q2	0.12	0.0722	0.0478
2004:Q3	0.066	0.0446	0.0214
2004:Q4	0.079	0.0314	0.0476
2005:Q1	0.062	0.0121	0.0499
2005:Q2	0.071	0.0126	0.0584
2005:Q3	0.081	0.0314	0.0496
2005:Q4	0.069	0.0278	0.0412
2006:Q1	0.09	0.0436	0.0464
2006:Q2	0.062	0.0372	0.0248
2006:Q3	0.064	0.0292	0.0348
2006:Q4	0.066	0.0122	0.0538
2007:Q1	0.055	0.0051	0.0499
2007:Q2	0.062	0.0279	0.0341
2007:Q3	0.068	0.0255	0.0425
2007:Q4	0.069	0.0589	0.0101
2008:Q1	0.073	0.062	0.011
Mean	0.0726	0.0347	0.0379
SD	0.0149	0.0193	0.0151
T	4.8814	1.7929	2.5122

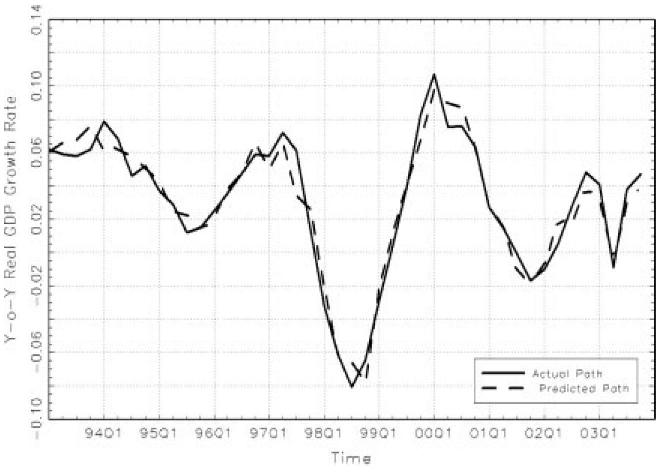


Figure 10. AIC: actual and predicted real GDP from 1993:Q1 to 2003:Q4

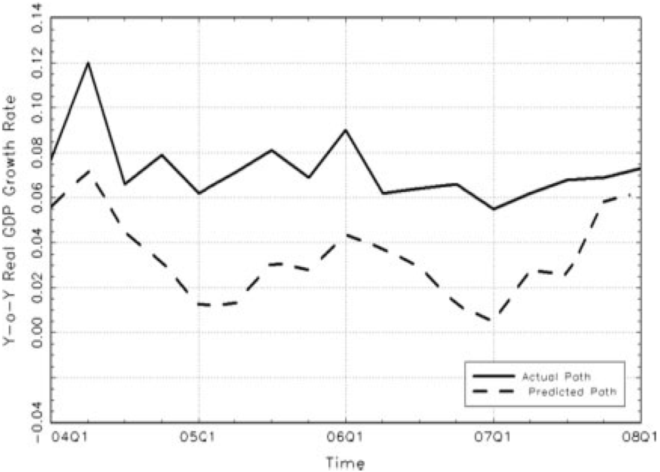


Figure 11. AIC: actual and counterfactual real GDP from 2004:Q1 to 2008:Q1

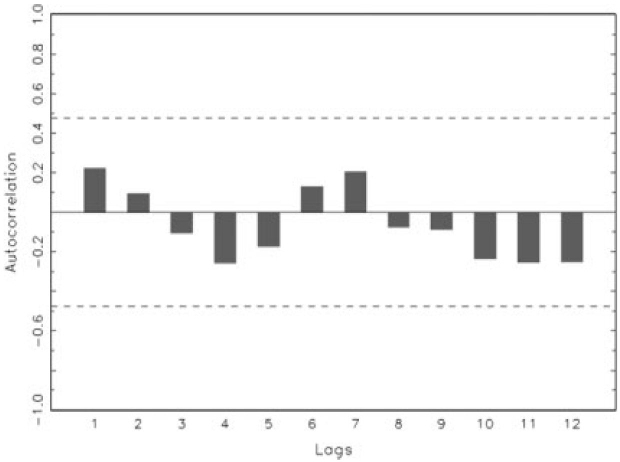


Figure 12. AIC: autocorrelation of treatment effect residuals from 2004:Q1 to 2008:Q1

Conclusion

The average actual growth rate from 2004:Q1 to 2008:Q1 is 7.26%. The average projected growth rate without CEPA is 3.47% using the group selected by AIC. The estimated average treatment effect is 3.79% with a standard error of 0.0151 based on the AIC group. The t-statistic is 2.5122.

Home-purchase Restrictions in Beijing

- The purchase restriction policy in Beijing entered into effect on April 30, 2010. Resident households were prohibited from buying more than two units of residential premises and non-resident households can buy at most one unit of residential premises with proof of local tax receipts or social security records of one year.
- The policy evaluation period is fixed from May 2010 to November 2011.
- We choose cities without purchase restrictions as the control group. Specifically, we include Tangshan, Qinhuangdao, Baotou, Jinzhou, Jilin, Yangzhou, Bengbu, Anqing, Quanzhou, Jiujiang, Ganzhou, Yantai, Jining, Luoyang, Pingdingshan, Yichang, Xiangfan, Yueyang, Changde, Guilin and Beihai into the control group.
- The data start from 2008:M1.

Table 2
Weights of control groups for Beijing, 2008:M1–2010:M4.

	Coefficient	St.Error	T-stat
Constant	3.2128	1.4228	2.2581
Jilin	0.2181	0.0729	2.9913
Yangzhou	0.1327	0.3473	0.3820
Bengbu	0.1182	0.1724	0.6859
Anqing	0.7689	0.0797	9.6447
Pingdingshan	−0.7321	0.1592	−4.5978
Xiangfan	0.4968	0.1962	2.5323
Guilin	0.8618	0.1884	4.5746
$R^2 = 0.99$			

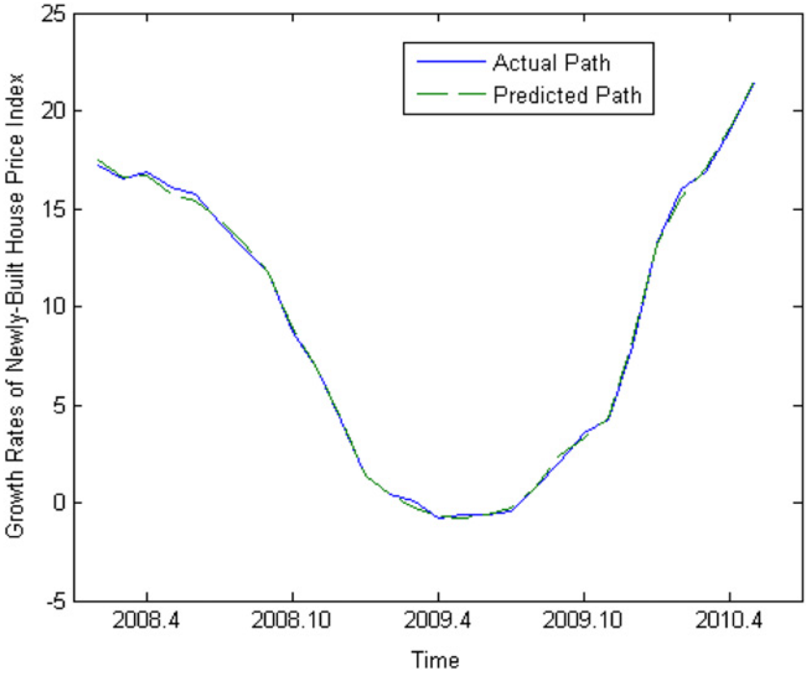

Actual Path
Predicted Path
Growth Rates of Newly-Built House Price Index
25
20
15
10
5
0
-5
2008.4
2008.10
2009.4
2009.10
2010.4
Time

Table 3
Treatment effects of Beijing home-purchase restriction, 2010:M5–2011:M11.

	Actual	Control	Treatment
2010.5	22.0	22.2575	−0.2575
2010.6	21.5	20.8147	0.6853
2010.7	20.1	21.1591	−1.0591
2010.8	18.6	20.1271	−1.5271
2010.9	18.2	19.7381	−1.5381
2010.10	17.5	20.8257	−3.3257
2010.11	14.3	19.6680	−5.3680
2010.12	9.9	20.9961	−11.0961
2011.1	6.8	19.5182	−12.7182
2011.2	6.8	18.7061	−11.9061
2011.3	4.9	18.2187	−13.3187
2011.4	2.8	15.9394	−13.1394
2011.5	2.1	14.2006	−12.1006
2011.6	2.2	14.0037	−11.8037
2011.7	1.9	12.7099	−10.8099
2011.8	1.9	12.0981	−10.1981
2011.9	1.8	10.6785	−8.8785
2011.10	1.7	11.0177	−9.3177
2011.11	1.3	9.7134	−8.4134
Average	9.2789	16.9679	−7.6890
T			−3.8437

Note: T-statistic in the last row is calculated using Theorem 3.2 of Li and Bell (2012), and it shows that the average treatment effect here is significantly less than 0 at 1%.

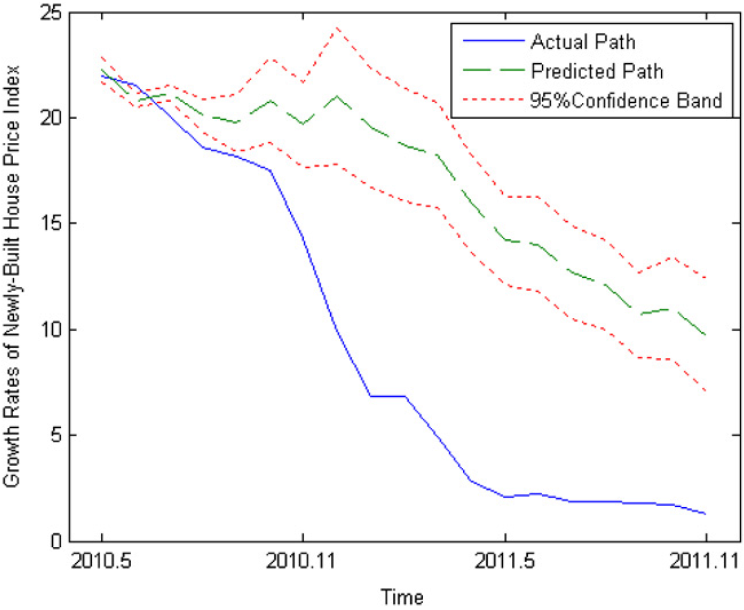


Fig. 4. Actual and predicted path of housing prices in Beijing, 2010:M5–2011:M11.

Trial property tax in Shanghai

- The trial property tax of Shanghai entered into effect on January 28, 2011.
- To separate the effect of property tax from that of purchase restrictions, we include 23 cities with purchase restrictions implemented about the same time as Shanghai into the control group, namely Shenzhen, Tianjin, Nanjing, Guangzhou, Xiamen, Ningbo, Fuzhou, Haikou, Wenzhou, Hangzhou, Dalian, Taiyuan, Zhengzhou, Wuhan, Hefei, Nanchang, Jinan, Qingdao, Kunming, Shijiazhuang, Xian, Wuxi and Jinhua.
- The policy evaluation period is fixed from February 2011 to February 2012.

Table 4
Weights of control groups for Shanghai, 2008:M1–2011:M1.

	Coefficient	St.Error	T-stat
Constant	0.1288	0.1754	0.7344
Hangzhou	0.4492	0.0395	11.3795
Hefei	0.2244	0.0664	3.3808
Fuzhou	0.0583	0.0554	1.0525
Jinan	0.0094	0.0806	0.1165
Guangzhou	0.2080	0.0134	15.5141
Haikou	−0.0225	0.0071	−3.1843
Wenzhou	−0.0679	0.0292	−2.3279
$R^2 = 0.99$			

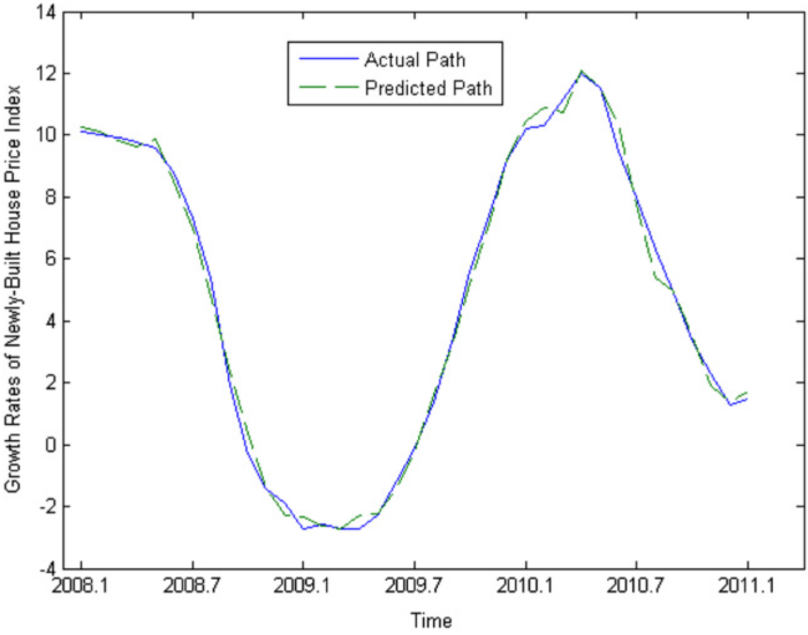
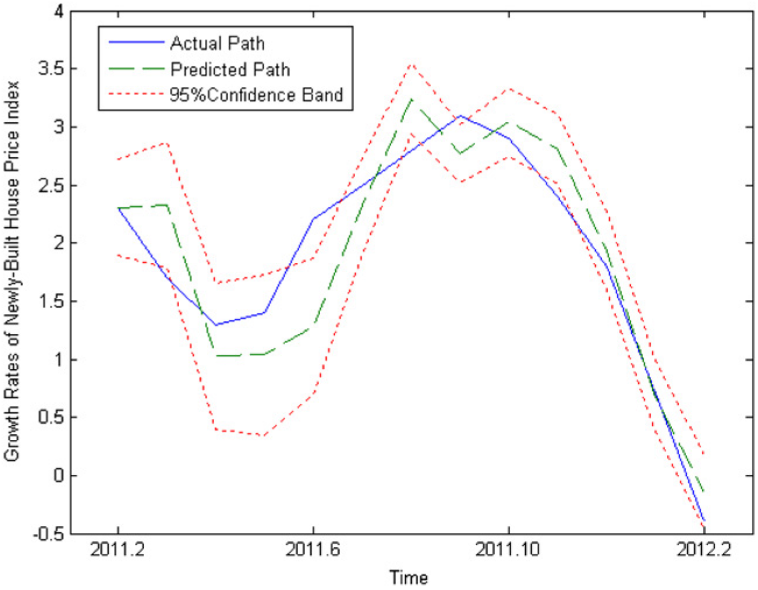


Fig. 5 Actual and predicted path of housing prices in Shanghai, 2008:M1–2011:M1

Table 5
Treatment effects of Shanghai property tax, 2011:M2–2012:M2.

	Actual	Control	Treatment
2011.2	2.3	2.3042	−0.0042
2011.3	1.7	2.3255	−0.6255
2011.4	1.3	1.0250	0.2750
2011.5	1.4	1.0395	0.3605
2011.6	2.2	1.2797	0.9203
2011.7	2.5	2.3214	0.1786
2011.8	2.8	3.2416	−0.4416
2011.9	3.1	2.7670	0.3330
2011.10	2.9	3.0404	−0.1404
2011.11	2.4	2.8092	−0.4092
2011.12	1.8	1.9329	−0.1329
2012.1	0.7	0.6746	0.0254
2012.2	−0.4	−0.1411	−0.2589
Average	1.9000	1.8938	0.0062
T			0.0324

Note: T-statistic in the last row is calculated using Theorem 3.2 of Li and Bell (2012), and it shows that the average treatment effect here is insignificant even at 10%.



Actual Path
Predicted Path
95%Confidence Band
Growth Rates of Newly-Built House Price Index
Time
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
2011.2
2011.6
2011.10
2012.2

Trial Property Tax in Chongqing

- The trial property tax of Chongqing entered into effect on January 28, 2011.
- We include 22 cities without purchase restrictions into the control group, namely Yangzhou, Bengbu, Anqing, Quanzhou, Jiujiang, Ganzhou, Jining, Luoyang, Pingdingshan, Yichang, Xiangfan, Yueyang, Changde, Huizhou, Zhanjiang, Shaoguan, Guilin, Beihai, Luzhou, Nanchong, Zunyi and Dali

Table 6
Weights of control groups for Chongqing, 2008:M1–2011:M1.

	Coefficient	St.Error	T-stat
Constant	2.9723	1.8832	1.5783
Bengbu	0.6609	0.2058	3.2110
Anqing	−0.8355	0.2677	−3.1215
Quanzhou	−0.0486	0.2789	−0.1742
Ganzhou	1.2605	0.3051	4.1315
Pingdingshan	−0.0579	0.4781	−0.1211
Yueyang	−0.4585	0.2018	−2.2719
Huizhou	0.0610	0.1321	0.4617
Shaoguan	−0.6897	0.3726	−1.8511
Guilin	0.5952	0.2843	2.0941
Nanchong	0.1779	0.3057	0.5819
$R^2 = 0.98$			

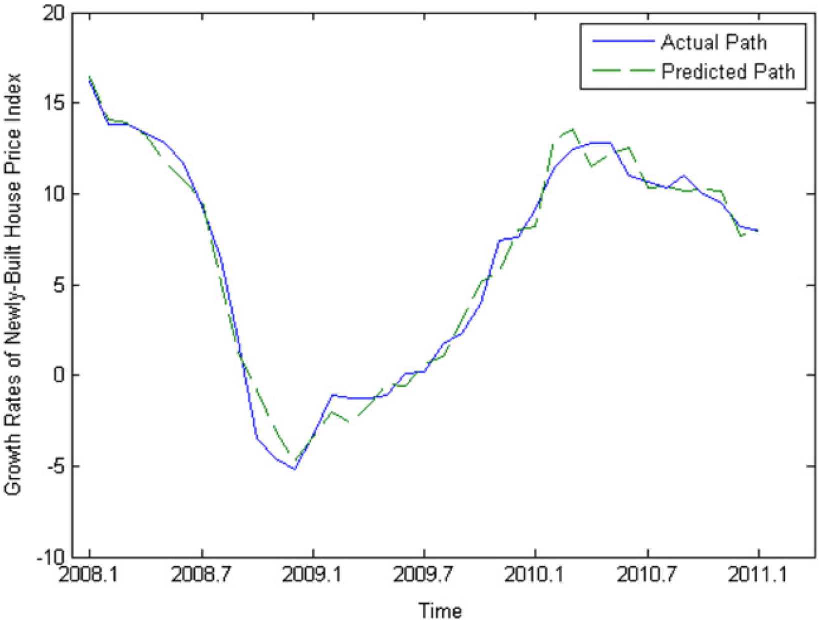


Fig. 7. Actual and predicted path of housing prices in Chongqing.

Table 7
Treatment effects of Chongqing property tax, 2011:M2–2012:M11.

	Actual	Control	Treatment
2011.2	6.2	5.2866	0.9134
2011.3	5.6	7.4988	−1.8988
2011.4	5.3	9.0804	−3.7804
2011.5	5.3	5.7708	−0.4708
2011.6	5.8	5.0611	0.7389
2011.7	5.6	5.6167	−0.0167
2011.8	4.2	5.8100	−1.6100
2011.9	2.2	5.9048	−3.7048
2011.10	1.2	4.2539	−3.0539
2011.11	0.1	3.8068	−3.7068
2011.12	−0.6	3.0193	−3.6193
2012.1	−0.7	3.2306	−3.9306
2012.2	−1.1	2.1803	−3.2803
2012.3	−1.5	1.4333	−2.9333
2012.4	−1.8	1.9552	−3.7552
2012.5	−1.9	2.2377	−4.1377
2012.6	−1.9	2.4340	−4.3340
2012.7	−1.5	2.5378	−4.0378
2012.8	−0.9	3.0689	−3.9689
2012.9	−0.5	1.9466	−2.4466
2012.10	−0.1	1.6278	−1.7278
2012.11	0.6	1.2181	−0.6181
Average	1.3455	3.8627	−2.5172
T			−1.5955

Note: T-statistic in the last row is calculated using Theorem 3.2 of Li and Bell (2012), and it shows that the average treatment effect here is significantly less than 0 at 10%.

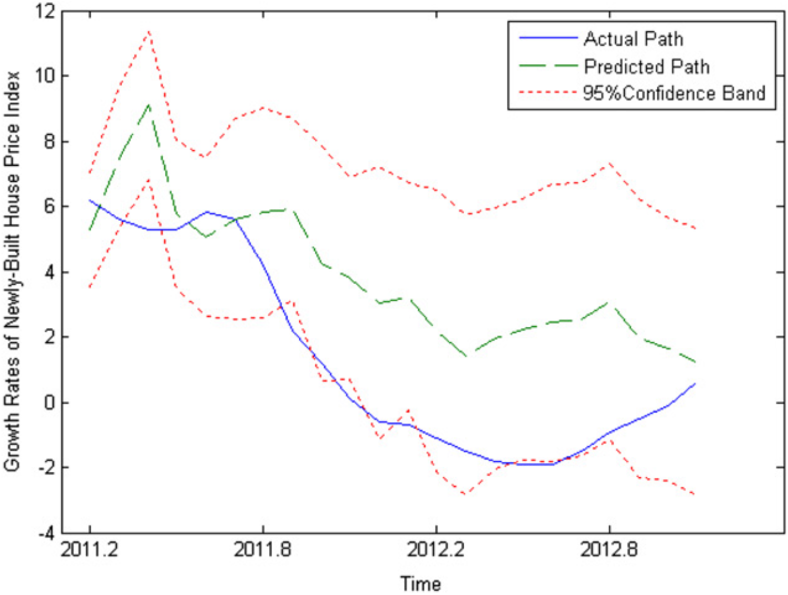


Fig. 8. Actual and predicted path of housing prices in Chongqing, 2011:M2–2012:M11.

Extensions

- Adding Covariates.
- Comparison with the Synthetic Control Methods (Abadie, Diamond and Hainmueller, 2010, JASA)

$$Y_{it} = \delta_t + \theta_t' Z_i + \lambda_t' \mu_i + \varepsilon_{it}$$

- ...

Thank you

Thank you
and Happy Thanksgiving!