JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **27**: 705–740 (2012)
Published online 5 January 2011 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1230

A PANEL DATA APPROACH FOR PROGRAM EVALUATION: MEASURING THE BENEFITS OF POLITICAL AND ECONOMIC INTEGRATION OF HONG KONG WITH MAINLAND CHINA

CHENG HSIAO,[a,b,c] H. STEVE CHING[b] AND SHUI KI WAN[d]*

[a] *University of Southern California, Los Angeles, CA, USA*
[b] *City University of Hong Kong, Kowloon, Hong Kong*
[c] *WISE, Xiamen University, Xiamen, China*
[d] *Hong Kong Baptist University, Kowloon, Hong Kong*

SUMMARY

We propose a simple-to-implement panel data method to evaluate the impacts of social policy. The basic idea is to exploit the dependence among cross-sectional units to construct the counterfactuals. The cross-sectional correlations are attributed to the presence of some (unobserved) common factors. However, instead of trying to estimate the unobserved factors, we propose to use observed data. We use a panel of 24 countries to evaluate the impact of political and economic integration of Hong Kong with mainland China. We find that the political integration hardly had any impact on the growth of the Hong Kong economy. However, the economic integration has raised Hong Kong's annual real GDP by about 4%. Copyright © 2011 John Wiley & Sons, Ltd.

1. INTRODUCTION

This paper proposes a panel data methodology to measure the impact of political and economic integration of Hong Kong with China. One of the difficulties of using nonexperimental data to measure the economic impact of a policy intervention is not being able to simultaneously observe the outcomes of an entity under the intervention and not under the intervention (e.g., Heckman and Hotz, 1989; Rosenbaum and Rubin, 1983). Panel data with observations for a number of individuals over time will often contain information on some individuals that are subject to policy intervention and some that are not. If the reactions of individuals towards policy changes are similar (e.g., Hsiao, 2003; Hsiao and Tahmiscioglu, 1997) or even if their responses are different, as long as they are driven by some common factors (e.g., Gregory and Head, 1999; Sargent and Sims, 1977), information on other individuals not subject to policy intervention can help to construct the counterfactuals of those who are subject to policy changes.

Hong Kong was a fishing village ceded to Britain after the Opium War in 1842. Many mainland Chinese migrated to Hong Kong after the establishment of the People's Republic of China in 1949. The population in 1950 was about 2.6 million. In the 1960s and 1970s Hong Kong experienced rapid economic growth and is now considered one of the four 'little dragons' in East Asia. In 1961, its per capita income was US $410–about 13.8% of that of the USA. By the eve of reverting sovereignty back to China on 1 July 1997, Hong Kong's population was 6.5 million, with per capita

* Correspondence to: Shui Ki Wan, Department of Economics, Hong Kong Baptist University, 5/F The Wing Lung Bank Building for Business Studies, 34 Renfrew Road, Kowloon Tong, Kowloon, Hong Kong, China.
E-mail: shuikiwan@gmail.com

income of US$21,441, which was 67.2% of that of the USA.[1] The Hang Seng stock market index was at 15,196. Because Hong Kong had been growing rapidly prior to the reversion of sovereignty to China, many questions have been raised about the impacts of the change of sovereignty on the growth of the Hong Kong economy (e.g., Sung and Wong, 2000). As a matter of fact, on the eve of Hong Kong's signing of the Closer Economic Partnership Arrangement (CEPA) with mainland China in June 2003, the growth rate for the second quarter of 2003 was -0.67%. The per capita income was US $22,673 in 2003. The Hang Seng Index fell to 8717 in April 2003.

The CEPA aimed to strengthen the linkage between mainland China and Hong Kong by liberalizing trade in services, enhancing cooperation in the area of finance, promoting trade and investment facilitation and mutual recognition of professional qualifications. The implementation of CEPA started on 1 January 2004 when 273 types of Hong Kong products could be exported to the mainland tariff free; another 713 types were added on 1 January 2005, 261 on 1 January 2006 and a further 37 on 1 January 2007. Chinese citizens residing in selected cities were also allowed to visit Hong Kong as individual tourists, from four cities in 2003 to 49 cities in 2007, including all 21 cities in Guangdong province.

In this paper we try to assess the impact of the political and economic integration of Hong Kong with mainland China on Hong Kong's economy by comparing what actually happened to Hong Kong's real GDP growth rates with what would have happened if there had been no change of sovereignty in July 1997 or no CEPA with mainland China in 2003. More specifically, we wish to analyze how these events have changed the growth rate of Hong Kong. However, to answer this question through conventional econometric modeling is not easy. We need to know how and why the Hong Kong economy has grown over time and how the China factor plays a role in Hong Kong's investment, labor migration and Hong Kong as an entrepot between China and the rest of the world, etc. Furthermore, most of the growth literature is highly abstract. Empirical analysis based on the theoretical literature would often require the imposition, as Sims (1980) claimed, of 'incredible' a priori identifying restrictions. Data demand will also be huge. Moreover, often when external conditions change, people's optimal decision rules also change. There simply may not be enough post-change observations to provide reliable inferences for the post-change outcomes. In addition, Hong Kong's economy has also been subject to many external shocks after the reversion of sovereignty. The Asian financial crisis broke out in October 1997. The Thai baht/US dollar exchange rate was 27 in June, 1997. It fell to 35.8 in September 1997 and further, to 44.4 baht to US$1 in December. The crisis in Thailand quickly spread to South Korea, Malaysia, Indonesia, Philippines, Singapore, Taiwan and other Pacific Rim countries with varying degrees of severity. Hong Kong was hit by international speculative attacks on four occasions in 1998. H5N1 Avian flu also broke out in December 1997, which caused five deaths and led to the slaughtering of more than a million chickens. By December 1997, the Hang Seng index had fallen to 10,722. In March 2003 severe acute respiratory syndrome (SARS) spread to Hong Kong from China. [2]

If we know the outcomes of a subject under intervention and not under intervention, the effect of a policy intervention is just the difference between the outcomes under intervention and in the absence of intervention. However, we rarely simultaneously observe the outcomes of an individual under intervention or in the absence of intervention. To properly evaluate the effect of a policy intervention on a subject or unit we need to construct the counterfactuals of the missing outcomes. Our approach to constructing the counterfactuals of the individual subject to intervention, say the ith unit, is to use other units that are not subject to intervention to predict what would have happened to the ith unit had it not been subject to policy intervention. The basic idea behind this approach is to rely on the correlations among cross-sectional units. We attribute the cross-sectional

[1] Hong Kong Census and Statistics Department website: http://www.censtad.gov.hk.
[2] For more information, see Jao (2001).

DOI: 10.1002/jae

dependence to the presence of common factors that drive all the relevant cross-sectional units. In Section 2 we set up the basic model. Section 3 proposes a panel approach to construct the counterfactuals without the need to identify the underlying model. Section 4 discusses a procedure to evaluate the time-varying treatment effects of a social program. Section 5 discusses strategies for selecting the most relevant cross-sectional units to construct counterfactuals. Section 6 discusses the data sources. Empirical results are presented in Section 7. Conclusions are in Section 8.

2. THE BASIC MODEL

The basic approach for constructing the counterfactuals is to rely on the correlations among cross-sectional units. We assume the correlations among cross-sectional units are due to some common factors that drive all cross-sectional units, although their impacts on each cross-sectional unit may be different. Let y_{it}^0 denote the outcome of the ith unit at time t without policy intervention. As in Forni and Reichlin (1998), Gregory and Head (1999), etc. we assume that y_{it}^0 is generated by a factor model of the form

$$y_{it}^0 = \underset{\sim}{b}_i' \underset{\sim}{f}_t + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T \tag{1}$$

where $\underset{\sim}{f}_t$ denotes the $K \times 1$ (unobserved) common factors that vary over time, $\underset{\sim}{b}_i'$ denotes the $1 \times K$ vector of constants that may vary across i, α_i denotes the fixed individual-specific effects, and ε_{it} denotes the ith unit random idiosyncratic component with $E(\varepsilon_{it}) = 0$.

Stacking $N \times 1$ y_{it}^0 into a vector yields

$$\underset{\sim}{y}_t^0 = B \underset{\sim}{f}_t + \underset{\sim}{\alpha} + \underset{\sim}{\varepsilon}_t \tag{2}$$

where $\underset{\sim}{y}_t^0 = (y_{1t}^0, \ldots, y_{Nt}^0)'$, $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_N)'$, $\underset{\sim}{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, and B is the $N \times K$ factor loading matrix $B = (\underset{\sim}{b}_1, \ldots, \underset{\sim}{b}_N)'$. We make the following assumptions.

Assumption 1. $||\underset{\sim}{b}_i|| = c < \infty$ for all i.

Assumption 2. $\underset{\sim}{\varepsilon}_t$ is $I(0)$ with $E(\underset{\sim}{\varepsilon}_t) = \underset{\sim}{0}$ and $E(\underset{\sim}{\varepsilon}_t \underset{\sim}{\varepsilon}_t') = V$, where V is a diagonal constant matrix.

Assumption 3. $E \underset{\sim}{\varepsilon}_t \underset{\sim}{f}_t' = \underset{\sim}{0}$.

Assumption 4. Rank $(B) = K$.

Remark 1. Model (1) assumes that the individual outcome is the sum of two components: a component of a function of some common time varying factors $\underset{\sim}{f}_t$ that drive all cross-sectional units and an idiosyncratic component consisting of a function of individual specific effects α_i and a random component ε_{it}. We assume the idiosyncratic components are uncorrelated across individuals.[3] The correlation across individuals are caused by the common factors, $\underset{\sim}{f}_t$. However, the impact of common factors $\underset{\sim}{f}_t$ on individuals can be heterogeneous by allowing $\underset{\sim}{b}_i \neq \underset{\sim}{b}_j$.

[3] As pointed out by a referee, the assumption about the idiosyncratic components being mutually uncorrelated rules out the possibility of local dependence. Although a distinction as to whether a factor is strong or weak could be useful (e.g., Chudik *et al.*, 2010), in practice it is difficult to know whether the extracted factors are strong or weak (or somewhere in between). Therefore, in this paper, we make the simplified assumption that the correlations among cross-sectional units are due to the presence of common factors, $\underset{\sim}{f}_t$ and the random idiosyncratic component for the ith unit, ε_{it}, merely represents the impacts of ith unit-specific factors. In other words, we assume the factors that create the 'local' dependence are also part of $\underset{\sim}{f}_t$.

DOI: 10.1002/jae

Remark 2. We made no assumption about the time series properties of $\underset{\sim}{f}_t$. It can be nonstationary or it can be stationary with $\lim \frac{1}{T}\sum_{t=1}^{T}||\underset{\sim}{f}_t||^2 \;= \text{constant}$.

Remark 3. Assumption 4 implies that the number of observable cross-sectional units, N, is greater than the number of common time-varying factors, $\underset{\sim}{f}_t$. The assumption is reasonable since it has been shown empirically by Sargent and Sims (1977), Giannone *et al.* (2005) (see also Watson's discussion of that paper in the same publication), Stock and Watson (2005) and Onatski (2009) that only a few common factors explain the bulk of the variance of macroeconomic data.

3. A PANEL APPROACH TO CONSTRUCTING COUNTERFACTUALS

Let y_{it}^1 denote the outcome of the ith unit at time t under treatment or intervention and y_{it}^0 denote the outcome of the ith unit in the absence of treatment or intervention at time t. Then the treatment effect for the ith unit at time t is

$$\Delta_{it} = y_{it}^1 - y_{it}^0 \tag{3}$$

However, often we do not simultaneously observe y_{it}^0 and y_{it}^1. The observed data, y_{it}, are in the form

$$y_{it} = d_{it}y_{it}^1 + (1-d_{it})y_{it}^0 \tag{4}$$

where

$$d_{it} = \begin{cases} 1, & \text{if the } i\text{th unit is under treatment at time } t \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Let $\underset{\sim}{y}_t = (y_{1t}, \ldots, y_{Nt})'$ be an $N \times 1$ vector of y_{it} at time t. Suppose there is no intervention before T_1, then the observed $\underset{\sim}{y}_t$ takes the form

$$\underset{\sim}{y}_t = \underset{\sim}{y}_t^0, \quad \text{for } t = 1, \ldots, T_1 \tag{6}$$

Suppose at time $T_1 + 1$, there is a policy change for the ith unit. Without loss of generality, let this be the first unit that receives the treatment at time $T_1 + 1$ and onwards:

$$y_{1t} = y_{1t}^1 \text{ for } t = T_1 + 1, \ldots, T \tag{7}$$

We assume other units are not affected by the policy intervention at the first unit, then

$$y_{it} = y_{it}^0 \text{ for } i = 2, \ldots, N, \text{ for } t = 1, \ldots, T \tag{8}$$

We make the following assumption.

Assumption 5. $E(\varepsilon_{js}|d_{it}) = 0$, for $j \neq i$.

Assumption 5 makes no claim about the relationship between d_{it} and ε_{it}. They can be correlated. If so, the observed data are subject to selection on unobservables (e.g., Heckman and Vytlacil, 2001). They can be independent; then the observed data satisfy the conditional independence assumption of Rosenbaum and Rubin (1983). All we need for our approach is that the jth unit's idiosyncratic components are independent of d_{it} for $j \neq i$.

Under Assumptions 1–5, we may predict y_{1t}^0 by $\hat{y}_{1t}^0 = \alpha_1 + \underset{\sim}{b}_1'\underset{\sim}{f}_t$ for $t = T_1 + 1, \ldots, T$, if we can identify α_1, $\underset{\sim}{b}_1$ and $\underset{\sim}{f}_t$. If both N and T are large, we may use the procedure of Bai and Ng

DOI: 10.1002/jae

(2002) to identify the number of common factors, K, and estimate $\underset{\sim}{f}_t$ by the maximum likelihood procedure. Often, neither N nor T is large. In this situation, we suggest using $\underset{\sim}{\tilde{y}}_t = (y_{2t}, \ldots, y_{Nt})'$ in lieu of $\underset{\sim}{f}_t$ to predict y_{1t}^0.

Let $\underset{\sim}{a}$ be a vector lying in the null space of B, $N(B)$. We normalize the first element of $\underset{\sim}{a}$ to be 1 and denote $\underset{\sim}{a}' = (1, -\underset{\sim}{\tilde{a}}')$. If $\underset{\sim}{a} \in N(B)$, then $\underset{\sim}{a}'B = 0$, and

$$y_{1t}^0 = \overline{\alpha} + \underset{\sim}{\tilde{a}}'\underset{\sim}{\tilde{y}}_t + \varepsilon_{1t} - \underset{\sim}{\tilde{a}}'\underset{\sim}{\tilde{\varepsilon}}_t \tag{9}$$

where $\overline{\alpha} = \underset{\sim}{a}'\underset{\sim}{\alpha}$, $\underset{\sim}{\tilde{y}}_t = (y_{2t}, \ldots, y_{Nt})'$, and $\underset{\sim}{\tilde{\varepsilon}}_t = (\varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. Equation (9) indicates that we can use $\underset{\sim}{\tilde{y}}_t$ in lieu of $\underset{\sim}{f}_t$ to predict y_{1t}^0.

Then, for any $\underset{\sim}{a} \in N(B)$,

$$y_{1t}^0 = E(y_{1t}^0|\underset{\sim}{\tilde{y}}_t) + \varepsilon_{1t}^* \tag{10}$$

where

$$\begin{aligned} E(y_{1t}^0|\underset{\sim}{\tilde{y}}_t) &= \overline{\alpha} + \underset{\sim}{\tilde{a}}'\underset{\sim}{\tilde{y}}_t + E(\varepsilon_{1t}|\underset{\sim}{\tilde{y}}_t) - E(\underset{\sim}{\tilde{a}}'\underset{\sim}{\tilde{\varepsilon}}_t|\underset{\sim}{\tilde{y}}_t) \\ &= \overline{\alpha} + \underset{\sim}{\tilde{a}}^{*\prime}\underset{\sim}{\tilde{y}}_t \end{aligned} \tag{11}$$

$$\underset{\sim}{\tilde{a}}^{*\prime} = \underset{\sim}{\tilde{a}}'(I_{N-1} - \text{cov}(\underset{\sim}{\tilde{\varepsilon}}_t, \underset{\sim}{\tilde{y}}_t)\text{var}(\underset{\sim}{\tilde{y}}_t)^{-1}) \tag{12}$$

$$\varepsilon_{1t}^* = \underset{\sim}{a}'\underset{\sim}{\varepsilon}_t + \underset{\sim}{\tilde{a}}'[\text{cov}(\underset{\sim}{\tilde{\varepsilon}}_t, \underset{\sim}{\tilde{y}}_t)\ \text{var}(\underset{\sim}{\tilde{y}}_t)^{-1}]\underset{\sim}{\tilde{y}}_t \tag{13}$$

The var$(\cdot)$ and cov$(\cdot)$ denote the long-run variance and covariance. The variance of y_{1t}^0 given $\underset{\sim}{\tilde{y}}_t$ for $\underset{\sim}{a} \in N(B)$ is equal to

$$\text{var}(y_{1t}^0|\underset{\sim}{\tilde{y}}_t) = \text{var}(\varepsilon_{1t}) + \underset{\sim}{\tilde{a}}'[\text{var}(\underset{\sim}{\tilde{\varepsilon}}_t) - \text{cov}(\underset{\sim}{\tilde{\varepsilon}}_t, \underset{\sim}{\tilde{y}}_t)\ \text{var}(\underset{\sim}{\tilde{y}}_t)^{-1}\ \text{cov}(\underset{\sim}{\tilde{y}}_t\ \underset{\sim}{\tilde{\varepsilon}}_t)]\underset{\sim}{\tilde{a}} \tag{14}$$

We note that $(\overline{\alpha}, \underset{\sim}{\tilde{a}}^*)$ depends on $\underset{\sim}{a}$ for any $\underset{\sim}{a} \in N(B)$. Since the minimum variance predictor depends on the choice of $\underset{\sim}{a}$ and covariance structure of $\underset{\sim}{\varepsilon}_t$, we propose to choose $\underset{\sim}{\tilde{a}}^*$ to minimize

$$\frac{1}{T_1}(\underset{\sim}{y}_1^0 - \underset{\sim}{e}\overline{\alpha} - Y\underset{\sim}{\tilde{a}}^*)'A(\underset{\sim}{y}_1^0 - \underset{\sim}{e}\overline{\alpha} - Y\underset{\sim}{\tilde{a}}^*) \tag{15}$$

where $\underset{\sim}{y}_1^0 = (y_{11}, \ldots, y_{1T_1})$, $\underset{\sim}{e}$ is a $T_1 \times 1$ vector of 1's, Y is a $T_1 \times (N-1)$ matrix of T_1 time series observations of $(\underset{\sim}{\tilde{y}}_t')$, and A is a $T_1 \times T_1$ positive definite matrix.

Assumption 6. For fixed K and N, there exists an $\underset{\sim}{a} \in N(B)$ such that in the neighborhood of $\underset{\sim}{a}$

$$E\left[\frac{1}{T_1}(\underset{\sim}{y}_1^0 - \underset{\sim}{e}\overline{\alpha} - Y\underset{\sim}{\tilde{a}}^*)'A(\underset{\sim}{y}_1^0 - \underset{\sim}{e}\overline{\alpha} - Y\underset{\sim}{\tilde{a}}^*)\right] \tag{16}$$

has a unique minimum.

Lemma 1. Under Assumptions 1–6, the solution of equation (15), $(\hat{\overline{\alpha}}, \hat{\underset{\sim}{\tilde{a}}}^*)$ converges to a $(\overline{\alpha}, \underset{\sim}{\tilde{a}}^*)$ that corresponds to an $\underset{\sim}{a} \in N(B)$ as $T_1 \to \infty$.

Proof. From equation (10) we have $y_{1t}^0 = \overline{\alpha} + \underset{\sim}{\tilde{a}}^{*\prime}\underset{\sim}{\tilde{y}}_t + \varepsilon_{1t}^*$ and $E(\varepsilon_{1t}^*|\underset{\sim}{\tilde{y}}_t) = 0$. Therefore, the minimum distance regression of y_{1t} on a constant and $\underset{\sim}{\tilde{y}}_t$ yields consistent estimators for $(\overline{\alpha}, \underset{\sim}{\tilde{a}}^{*\prime})$ when $T_1 \to \infty$ (e.g., Amemiya, 1985).

Remark 4. The null vectors in $N(B)$ are not unique. However, for given A or objective function (15), the solution is unique. When $A = I$, our objective is to obtain the minimum variance predictor of y_{1t}^0 given $\tilde{\underset{\sim}{y}}_t$. In other words, as pointed out by a referee, the conditional paths for the units under treatment are computed by exploiting only contemporaneous cross-sectional correlations. To allow the exploitation at leads and lags of the dynamic relationships among cross-sectional units, we may let A in equation (15) be a nondiagonal matrix, say the inverse of the covariance matrix of $\underset{\sim}{\varepsilon}_1^* = (\varepsilon_{11}^*, \ldots, \varepsilon_{1,\ T_1}^*)'$, and use the optimal forecasting formula of Goldberger (1962) to produce a counterfactual path that uses both current and lagged values of $\underset{\sim}{y}_t$. However, the so generated lead–lag relationships are restricted by the serial correlation patterns of $\underset{\sim}{\varepsilon}_1^*$ as compared to an unrestricted vector autoregressive model (VAR).

Remark 5. Although $\overline{\alpha} = \underset{\sim}{a}'\underset{\sim}{\alpha}$ depends on the choice of $\underset{\sim}{a}$, it is just an unknown finite constant under Assumption 1 in the regression model (15). Therefore it can be treated as an unknown in equation (15).

When $A = I$, Lemma 1 suggests that we can predict y_{1t}^0 by

$$\hat{y}_{1t}^0 = \hat{\overline{\alpha}} + \hat{\tilde{\underset{\sim}{a}}}{*}'\tilde{\underset{\sim}{y}}_t \tag{17}$$

Therefore, we may predict Δ_{1t} using

$$\hat{\Delta}_{1t} = y_{1t} - \hat{y}_{1t}^0 \text{ for } t = T_1 + 1, \ldots, T \tag{18}$$

Lemma 2. Under Assumptions 1–6,

$$E(\hat{\Delta}_{1t}|Y, \tilde{\underset{\sim}{y}}_t) = \Delta_{1t}, \quad t = T_1 + 1, \ldots, T \tag{19}$$

and

$$\operatorname{var}(\hat{\Delta}_{1t}) = \operatorname{var}(\varepsilon_{1t}^*) + (1, \tilde{\underset{\sim}{y}}_t')\operatorname{cov}\begin{pmatrix}\hat{\overline{\alpha}}\\ \hat{\tilde{\underset{\sim}{a}}}^*\end{pmatrix}\begin{pmatrix}1\\ \tilde{\underset{\sim}{y}}_t\end{pmatrix} \tag{20}$$

Proof. Under Assumptions 1–6,

$$E\begin{pmatrix}\hat{\overline{\alpha}}\\ \hat{\tilde{\underset{\sim}{a}}}^*|Y\end{pmatrix} = \begin{pmatrix}\overline{\alpha}\\ \tilde{\underset{\sim}{a}}^*\end{pmatrix}$$

Lemma 2 follows from equations (11) and (14).

Remark 6. The counterfactuals $y_{1t}^0, t = T_1 + 1, \ldots, T$ depend on the individual specific effects α_1, the common factors, $\underset{\sim}{f}_t$, the individual specific response to time-varying common factors $\underset{\sim}{f}_t, \underset{\sim}{b}_1$, and the idiosyncratic component ε_{1t}. However, the counterfactual predictor (17) does not require any such knowledge, nor the dimension of $\underset{\sim}{f}_t$. The information provided by $\underset{\sim}{f}_t$ is embedded in $\tilde{\underset{\sim}{y}}_t$. It follows that the predictor $\hat{\Delta}_{1t}$, (18), which uses $\tilde{\underset{\sim}{y}}_t$ in lieu of $\underset{\sim}{f}_t$ allows the evaluation of policy interventions without the need to identify $\underset{\sim}{f}_t$ or B, which may be difficult in a finite sample.

Remark 7. We do not make any assumption about ε_{is} and d_{it}. All we need is that the policy intervention on the ith unit has no bearing on ε_{jt} for $j \neq i$ (Assumption 5). Hence, if the process (2) satisfies Assumptions 1–5, our proposed approach allows us to bypass the selection issue, which has been a central concern in the program evaluation literature (e.g., Heckman and Hotz, 1989; Heckman and Vytlacil, 2001).

DOI: 10.1002/jae

Remark 8. When $(T - T_1)$ is large, given Assumption 5, one can reverse the procedure to predict y_{1t}^1 by $E(y_{1t}^1|\tilde{\underset{\sim}{y}}_t)$ for $t = 1, \ldots, T_1$, where $E(y_{1t}^1|\tilde{\underset{\sim}{y}}_t)$ may be approximated by

$$\hat{y}_{1t}^1 = \hat{\tilde{\alpha}}^* + \hat{\underset{\sim}{\delta}}' \tilde{\underset{\sim}{y}}_t, \quad t = 1, \ldots, T_1 \tag{21}$$

and construct the treatment effect had the policy intervention been in place before T_1, $\hat{\Delta}_{1t} = \hat{y}_{1t}^1 - y_{1t}$, $t = 1, \ldots, T_1$, where $\hat{\tilde{a}}^*$ and $\hat{\underset{\sim}{\delta}}$ are estimated using data from $T_1 + 1, \ldots, T$.

Remark 9. The synthetic control method for comparative case studies also uses information of other individuals to construct the counterfactuals of treated individuals (e.g., Abadie and Gardeazabal, 2003; Card and Krueger, 1994). However, the focus and the approach are different. The synthetic approach assumes that (e.g., Abadie *et al.*, 2010)

$$y_{it}^0 = \delta_t + \underset{\sim}{z}_i'\theta_t + \underset{\sim}{\mu}_i'\underset{\sim}{\lambda}_t + \varepsilon_{it}, \text{ for } i = 2, \ldots, N, t = 1, \ldots, T_1, T_1 + 1, \ldots, T \tag{22}$$

while for the first unit they assume y_{1t} follows equation (22) for $t = 1, \ldots, T_1$, and for $t = T_1 + 1, \ldots, T$, y_{1t} equals

$$y_{1t}^1 = \Delta_{1t} + \delta_t + \underset{\sim}{z}_1'\theta_t + \underset{\sim}{\mu}_1'\underset{\sim}{\lambda}_t + \varepsilon_{1t} \text{ for } t = T_1 + 1, \ldots, T \tag{23}$$

where δ_t is an unknown common factor with constant factor loading across units, $\underset{\sim}{z}_i$ is an $(r \times 1)$ vector of observed covariates (not affected by the intervention), $\underset{\sim}{\theta}_t$ is an $(r \times 1)$ vector of unknown parameters, $\underset{\sim}{\lambda}_t$ is a $(K \times 1)$ vector of unobserved common factors, $\underset{\sim}{\mu}_i$ is a $(K \times 1)$ vector of unknown factor loading and Δ_{1t} is the treatment effect for the first unit. If we let $\alpha_i = 0$, $b_i' = (1, \underset{\sim}{z}_i', \underset{\sim}{\mu}_i')$ and $\underset{\sim}{f}_t' = (\delta_t, \underset{\sim}{\theta}_t', \underset{\sim}{\lambda}_t')$ in equation (1), model (22) can be put in the form of equation (1). However, for equation (9) to hold, we need $\underset{\sim}{a}'B = 0$, which imposes the restriction $\sum_{i=1}^N a_i = 0$, $\sum_{i=1}^N a_i\underset{\sim}{z}_i' = \underset{\sim}{0}'$ and $\sum_{i=1}^N a_i\underset{\sim}{\mu}_i' = \underset{\sim}{0}$.

The synthetic control method constructs the predicted y_{1t}^0 through $\hat{y}_{1t}^0 = \sum_{i=2}^N a_i y_{it}$, where the weights $\tilde{\underset{\sim}{a}}' = (a_2, \ldots, a_N)$ are obtained by minimizing

$$(\underset{\sim}{x}_1 - X\tilde{\underset{\sim}{a}})'V(\underset{\sim}{x}_1 - X\tilde{\underset{\sim}{a}}) \tag{24}$$

subject to the constraints $a_i \geq 0$ for $i = 1, \ldots, N$ and $\sum_{j=2}^N a_i = 1$, where $\underset{\sim}{x}_1$ is an $(r + M) \times 1$ column vector consisting of the r observed covariates $\underset{\sim}{z}_1$ for the control (pre-treatment) period. To ensure unbiasedness of their estimator, under the assumption that the dimension of unknown common factors is K, they suggest including M other covariates in $\underset{\sim}{x}_1$ by constructing the time average of at least K such selectors through $\overline{y}_1^m = \sum_{s=0}^{T_1} k_{1s}^m y_{1s}$, $m = 1, \ldots, M$. X is an $(r + M) \times (N - 1)$ matrix, a collection of all other $\underset{\sim}{x}_j$, $j = 2, \ldots, N$ columns similarly constructed as $\underset{\sim}{x}_1$ with $\overline{y}_j^m = \sum_{s=0}^{T_1} k_{js}^m y_{js}$, $m = 1, \ldots, M$ (when $k_{js}^m = \frac{1}{T_1}$, then y_j^m is just a simple pre-intervention time average). Therefore, the cross-sectional units weight a_i will be sensitive to the prior choice of $\underset{\sim}{z}$, M, and k_{js}; hence the predicted $\hat{y}_{1t}^0$, or $\hat{\Delta}_{1t}$. Nor is the probability distribution of $\hat{y}_{1t}^0$ or $\hat{\Delta}_{1t}$ easily derivable. On the other hand, we suggest using the regression method to choose $\tilde{\underset{\sim}{a}}$ to mimic the behavior of treated individuals before the intervention as closely as possible, say, by minimizing equation (15). As long as N is fixed, our procedure yields a unique weight $\tilde{\underset{\sim}{a}}$ and unique $\hat{y}_{1t}^0$; hence unique $\hat{\Delta}_{1t}$ with known probability distribution. Neither do we need to impose the constraint $a_j \geq 0$, nor $\sum_{j=2}^N a_j = 1$. Our approach can also easily be adapted to accommodate the case that some exogenous variables $\underset{\sim}{z}_i$ also drive $\underset{\sim}{y}_t$ by treating equation (2) conditional on $\underset{\sim}{z}_i$.

DOI: 10.1002/jae

Remark 10. An alternative approach to exploit the correlation among the cross-sectional units is to construct a vector autoregressive model (VAR). A VAR can describe dynamic correlations generally. For instance, one can construct a VAR for $\underset{\sim}{y}_t$, as

$$\underset{\sim}{y}_t^0 = \underset{\sim}{c} + A_1 \underset{\sim}{y}_{t-1}^0 + \ldots + A_p \underset{\sim}{y}_{t-p}^0 + \underset{\sim}{u}_t \tag{25}$$

Pre-intervention data can then be used to estimate the parameters of the system (25), $\underset{\sim}{\theta}_{\text{pre-int}} = \text{vec}\ (\underset{\sim}{c}, A_1, \ldots, A_p)_{\text{pre-int}}$, $\hat{\underset{\sim}{\theta}}_{\text{pre-int}}$. The expected path of a cross-sectional unit in the absence of intervention, say y_{1t}^0, can then be constructed based on $\hat{\underset{\sim}{\theta}}_{\text{pre-int}}$. This approach has been adopted recently by Giannone *et al.* (2010) to evaluate the effect of policy intervention in the euro area. However, if there is a feedback relation from $y_{1,t-j}^0$ to $(y_{2t}^0, \ldots, y_{Nt}^0)$ (i.e., the elements of the first column of A_j are nonzero (e.g., Granger, 1969; Hsiao, 1982)), substituting y_{1t} in lieu of y_{1t}^0 in system (25) is likely to generate errors because the post-intervention $y_{1,t-j}$ is actually $y_{1,t-j}^1$, not $y_{1,t-j}^0$. Moreover, if $y_{1,t-j} = y_{1,t-j}^1$, then $(y_{2t}, \ldots, y_{Nt})$ cannot be $(y_{2t}^0, \ldots, y_{Nt}^0)$.

Remark 11. Model (2) with

$$\underset{\sim}{f}_t = H_1 \underset{\sim}{f}_{t-1} + \ldots + H_q \underset{\sim}{f}_{t-q} + v_t \tag{26}$$

can be transformed into a VAR model (25). If the number of factors, K, is less than N, then it is possible to transform equation (25) into a one-way causal model from $(y_{2t}^0, \ldots, y_{Nt}^0)$ to y_{1t}^0 (i.e., with all the elements of the first column of A_j equal to zero except for the first element for $j = 1, \ldots, p$; Granger, 1969). However, if $\underset{\sim}{a}'B = \underset{\sim}{0}'$, so are $\underset{\sim}{a}'A_j = \underset{\sim}{0}'$, $j = 1, \ldots, p$, then y_{1t}^0 will be just equal to equation (9).

Remark 12. If $\underset{\sim}{f}_t$ and $\underset{\sim}{b}_1$ are known, then $\text{var}(y_{1t}^0 | \underset{\sim}{f}_t) = \text{var}(\varepsilon_{1t})$ is smaller than (14). If N and T are large, one can use Bai's (2003), Bai and Ng's (2002) procedure to identify the number of unknown factors K and estimate $\underset{\sim}{b}_1$ and $\underset{\sim}{f}_t$. However, if T is small, then there could be sampling errors in identifying and estimating $\underset{\sim}{b}_1$ and $\underset{\sim}{f}_t$. It may be better to use $\tilde{\underset{\sim}{y}}_t$ and $\tilde{\underset{\sim}{a}}$ in lieu of $\underset{\sim}{b}_1$ and $\underset{\sim}{f}_t$ as demonstrated in our Monte Carlo studies in Section 5 (also see Pesaran *et al.*, 2007).

4. TESTS FOR SIGNIFICANCE OF POLICY INTERVENTION

The predictor for the effectiveness of social policy (18) allows the effects of such a policy to vary over time. From the estimated Δ_{1t}, we may use time series techniques to evaluate the evolution of policy effects over time.

Assumption 7. $\{\varepsilon_{it}\}$ is weakly dependent (mixing) for all i.

Suppose the treatment effects, Δ_{1t}, follow an autoregressive moving average model (ARMA) of the form

$$a(L)\Delta_{1t} = \mu + \theta(L)\eta_t \tag{27}$$

where L is the lag operator, η_t is an i.i.d. process with zero mean and constant variance and the roots of $\theta(L) = 0$ lie outside the unit circle. If the roots of $a(L) = 0$ all lie outside the unit circle, the treatment effect is stationary, and the long-term treatment effect is

$$\Delta_1 = a(L)^{-1}\mu = \mu^* \tag{28}$$

DOI: 10.1002/jae

If one of the roots of $a(L) = 0$ lies on the unit circle, the intervention effects are integrated of order 1, $I(1)$.

From the estimated $\hat{\Delta}_{1t}$, we can use the Box–Jenkins (1970) procedure to construct a time series model:

$$\tilde{a}(L)\hat{\Delta}_{1t} = \tilde{\mu} + \tilde{\theta}(L)v_t \tag{29}$$

where v_t is i.i.d. with mean zero and variance σ_v^2.

Lemma 3. Suppose the roots of $a(L) = 0$ lie outside the unit circle, under Assumptions 1–6, when both T_1 and $(T - T_1)$ go to infinity:

$$\text{plim } \tilde{a}(L)^{-1}\tilde{\mu} = \text{plim } \hat{\mu}^* = \mu^* = a(L)^{-1}\mu \tag{30}$$

and

$$\sqrt{T - T_1}(\hat{\mu}^* - \mu^*) \sim N(0, \sigma_{\mu^*}^2) \tag{31}$$

where

$$\sigma_{\mu^*}^2 = \frac{\partial \mu^*}{\partial \underset{\sim}{\gamma}'} \text{ cov}(\sqrt{T - T_1}\hat{\underset{\sim}{\gamma}})\frac{\partial \mu^*}{\partial \underset{\sim}{\gamma}} \tag{32}$$

and $\underset{\sim}{\gamma} = (\tilde{\mu}, \tilde{a}_1, \ldots, \tilde{a}_p)$, assuming $\tilde{a}(L)$ is of pth order.

Proof. If $\underset{\sim}{y}_t$ is stationary, the estimators of $(\hat{\bar{\alpha}}, \hat{\underset{\sim}{a}}^{*\prime})$ are $\sqrt{T_1}$-consistent. If $\tilde{y}_t \sim I(1)$ are not cointegrated, the estimator of $\hat{\bar{\alpha}}$ remains $\sqrt{T_1}$-consistent, but $\hat{\underset{\sim}{a}}^*$ is T_1-consistent (e.g., Phillips and Durlauf, 1986). Either way,

$$y_{1t}^0 - \hat{y}_{1t}^0 = \varepsilon_{1t}^* + O(T^{-\frac{1}{2}}) \tag{33}$$

Adding and subtracting yields

$$\hat{\Delta}_{1t} = y_{1t} - \hat{y}_{1t}^0 = \Delta_{1t} + \varepsilon_{1t}^* + o(1) \tag{34}$$

Substituting (34) into (27) yields

$$a(L)\hat{\Delta}_{1t} = \mu + \theta(L)\eta_t + a(L)\varepsilon_{1t}^* + o(1) \tag{35}$$

Since ε_{1t}^* is a mean zero $I(0)$ process, we obtain (28) by approximating $\theta(L)\eta_t + a(L)\varepsilon_{1t}^*$ by a qth-order moving average process, $\theta^*(L)v_t$. If the roots of $\theta^*(L)$ all lie outside the unit circle, $\hat{\Delta}_{1t}$ can also be approximated by an AR process:

$$\tilde{a}(L)\hat{\Delta}_{1t} = \tilde{\mu} + v_t \tag{36}$$

where $\tilde{a}(L) = \theta^*(L)^{-1}a(L)$ and $\tilde{\mu} = \theta^*(L)^{-1}\mu$.

Under fairly general conditions, the maximum likelihood estimator (MLE) of $a(L), \theta^*(L)$ and μ are consistent and asymptotically normally distributed. The asymptotic variance, $\sigma_{\mu^*}^2$, can then be derived by using the delta method (e.g., Rao, 1973, ch. 2).

If the treatment effects is a stationary process (i.e., the roots of $a(L) = 0$ all lie outside the unit circle), the long-term impact of the intervention can also be estimated by taking the simple average of the treatment effects.

DOI: 10.1002/jae

Lemma 4. Suppose all the roots of $a(L) = 0$ lie outside the unit circle, under Assumptions 1–6, when both T_1 and $(T - T_1)$ go to infinity,

$$\operatorname*{plim}_{(T-T_1)\to\infty} \frac{1}{T-T_1}\sum_{t=T_1+1}^{T}\hat{\Delta}_{1t} = \Delta_1 \tag{37}$$

The variance of equation (37) can be approximated by the heteroscedastic–autocorrelation consistent (HAC) estimator of Newey and West (1987).

Proof. Given equations (19) and (20), the law of large number holds.

5. CHOICE OF CROSS-SECTIONAL UNITS

5.1. Modeling Strategy

Often there are a large number of cross-sectional units that can be used to predict y_{1t}^0 (or that are generated according to equation (1) or (2)). Intuitively, it would appear to favor using as many available cross-sectional units as possible as long as $T > N$. This will be the case when the number of common factors, K, is fixed, T_1 goes to infinity, and $\frac{N}{T_1} \to 0$. However, if T_1 or $\frac{N}{T_1}$ is finite, there may be an advantage to using only a subset of available cross-sectional units to predict the counterfactuals, in particular, if the data-generating processes for cross-sectional units satisfy the condition of Lemma 5(ii) below.

Let there be m cross-sectional units that optimally predict y_{1t}^0 and $(N - m - 1)$ remaining cross-sectional units that could also be included to predict y_{1t}^0. Let Y_1 and Y_2 be the $T_1 \times m$ and $T_1 \times (N - m - 1)$ time series observations for these m cross-sectional unit and $(N - m - 1)$ cross-sectional units, respectively, then

$$Y_1 = FB_1' + \mathcal{E}_1 \tag{38}$$

and

$$Y_2 = FB_2' + \mathcal{E}_2 \tag{39}$$

where the tth row of F takes the form $(1, \underset{\sim}{f}_t')$ and the ith column of B_1' and B_2' takes the form $(\alpha_i, \underset{\sim}{b}_i')'$, and $\mathcal{E}_1$ and $\mathcal{E}_2$ denote the $T_1 \times m$ and $T_1 \times (N - m - 1)$ idiosyncratic components of Y_1 and Y_2, respectively.

Lemma 5. Under Assumptions 1–5:

(i) The optimal number of cross-sectional units for constructing y_{1t}^0 is $K \le m \le N - 1$.

(ii) If $B_2(B_1'\Omega_1^{-1}B_1' + I_K)^{-1}\begin{pmatrix}\alpha_1\\ b_1\end{pmatrix} = \underset{\sim}{0}$ then Y_2 yields no predictive power for $E(y_1|Y_1)$, where $\Omega_1 = E(\mathcal{E}_1'\mathcal{E}_1)$.

For proof, see Appendix A.

When N is fixed and $T_1 \to \infty$, the least squares estimator of $(y_1 - Y_1\underset{\sim}{a}_1 - Y_2\underset{\sim}{a}_2)'(y_1 - Y_1\underset{\sim}{a}_1 - Y_2\underset{\sim}{a}_2)$ yields $\hat{\underset{\sim}{a}}_2$ that will converge to 0 under the condition of Lemma 5(ii). In other words, one can use all $(N - 1)$ available cross-sectional units to predict y_{1t}^0. However, on many occasions, T_1 is finite. As more cross-sectional units are used, the variance of $\tilde{\underset{\sim}{a}}^*$ will also increase. To balance the within-sample fit with post-sample prediction error, we suggest the following model selection strategy (Hsiao and Wan, 2010):

Step 1. Use R^2 or likelihood values to select the best predictor for y_{1t}^0 using j cross-sectional units out of $(N-1)$ cross-sectional units, denoted by $M(j)^*$, for $j = 1, \ldots, N-1$.[4]

Step 2. From $M(1)^*, M(2)^*, \ldots, M(N-1)^*$, choose $M(m)^*$ in terms of a model selection criterion.

5.2. Monte Carlo Studies

Under the assumption that $\underset{\sim}{y}_t$ is generated by a factor model of the form (2), in this subsection we compare the predictive performance of our approach versus that of first determining the number of factors, K, and identifying α_1 and $\underset{\sim}{b}_1$, then using the estimated $\underset{\sim}{f}_t$, α_1, and $\underset{\sim}{b}_1$ to generate the counterfactuals when N and T are small.

First, we wish to see if there is a need to use all cross-sectional units using our approach. There are a number of model selection criteria one can use to select the best approximating model. In this section we conduct a small-scale Monte Carlo to examine the performance of the Akaike information criterion (AIC) (Akaike, 1973, 1974), and corrected Akaike information criterion (AICC) (Hurvich and Tsai, 1989), by comparing the post-intervention mean square prediction error:

$$\mathrm{PMSE}(p) = \frac{1}{T-T_1} \sum_{t=T_1+1}^{T} (y_{1t}^0 - \hat{y}_{1t}^0(p))^2 \tag{40}$$

where $\hat{y}_{1t}^0(p)$ is generated by using p cross-sectional unit data of y_{it} for $t = 1, \ldots, T_1$ to obtain the least square estimates of $\hat{\underset{\sim}{a}}^{*\prime}_p$, then $\hat{y}_{1t}^0(p) = \hat{\underset{\sim}{a}}^{*\prime}_p Y_t$, $t = T_1+1, \ldots, T$.

To see which model selection criterion works better, we generate model (1) with $N = 21$ countries, the sum of the dimension of both Y_1 and Y_2. We use $T_1 = 25$, 40, and 60 observations, the number of pre-intervention periods to approximate the path of y_1 before intervention. The OLS estimators are then used to predict y_{1t}^0 for the post-intervention period which has $T - T_1 = 10$ periods. Four different factor structures are used. The first one consists of two ($K = 2$) stationary factors:

$$\begin{aligned} f_{1t} &= 0.3 f_{1,t-1} + u_{1t} \\ f_{2t} &= 0.6 f_{2,t-1} + u_{2t} \end{aligned} \tag{41}$$

where the innovation for factor loadings u_t and the idiosyncratic errors ε_t are generated by $N(0,1)$ and $\sigma N(0, 1)$, respectively. The second one is another set of stationary factors:

$$\begin{aligned} f_{1t} &= 0.8 f_{1,t-1} + u_{1t} \\ f_{2t} &= -0.6 f_{2,t-1} + u_{2t} + 0.8 u_{2t-1} \\ f_{3t} &= u_{3t} + 0.9 u_{3t-1} + 0.4 u_{3t-2} \end{aligned} \tag{42}$$

[4] Given N, there are 2^N possible combinations. Step 1 is proposed to simplify the choice of the number of predictive models. An alternative is to use a boosting method (Buhlmann, 2006) or LASSO (Tibshirani, 1996) or to use some targeted predictors based on soft and hard-thresholding as suggested by a referee.

DOI: 10.1002/jae

The third has an i.i.d. factor. The last one has an almost non-stationary factor:

$$f_{1t} = 0.95 f_{1,t-1} + u_{1t} \tag{43}$$

In all the above cases, $\underset{\sim}{b}_i \sim N(1, 1)$.

Two model selection criteria are compared:

$$\text{AIC}(p) = T_1 \ln\left(\frac{\underset{\sim}{e}_0' \underset{\sim}{e}_0}{T_1}\right) + 2(p+2) \tag{44}$$

$$\text{AICC}(p) = \text{AIC}(p) + \frac{2(p+2)(p+3)}{T_1 - (p+1) - 2} \tag{45}$$

where p is the number of countries included; $\underset{\sim}{e}_0$ denotes the OLS residuals.

We repeat the experiment for each of the four data-generating processes 500 times. All four simulation results show that the pre-intervention MSE decreases when p increases, whereas post-intervention MSE decreases initially and then increases when p increases. Denote the optimal number of countries chosen as m. The results are summarized in Tables I–IV. For all the experiments, the average m is between K and $N-1$. This is consistent with the notion of the bias and variance tradeoff. The frequency distributions show that not once are all 20 cross-sectional units selected in terms of AICC and there is around a 1% chance that the 20 units models are chosen in terms of AIC. On average, between 4 and 6 cross-sectional units are chosen in terms of AICC and 6–12 cross-sectional units are chosen in terms of AIC. When T_1 is small, say 25 or 40, the PMSE in terms of models chosen by AICC and AIC are significantly smaller than the models using all 20 cross-sectional units. When T_1 becomes large, say 60, models using all 20 cross-sectional units have PMSE converge towards the PMSE of the optimally chosen models in terms of AICC or AIC, but the PMSE based on 20 cross-sectional units are still larger than those based only on m cross-sectional units. These Monte Carlo studies appear to support the theoretical finding that optimal m is between K and $(N-1)$ when T_1 is finite. When $\frac{N}{T_1} \to 0$, then using all available cross-sectional units will be fine because the estimated $\hat{\underset{\sim}{a}}_2$ will converge to zero.

We then compare the predictive performance of our approach based on AIC, and AICC with Bai and Ng (2002) PC and IC criteria. Tables V and Table VI provide the results of one factor model based on setting the maximum number of K equal to 8 and 20, respectively. As one can see, the predictive performance of Bai and Ng (2002) is very sensitive to the prior specified number of maximum K. The performance of the factor model also deteriorates when the number of factors increased from one to five (Tables VI and VII); when the average of $\underset{\sim}{b}_i$ changed from 0 to 0.3 or 1 (Tables VIII and IX); when the distribution of $\underset{\sim}{b}_i$ changed to uniform (−1,1) or N(2,2) (Tables X and XI); when the idiosyncratic components, ε_{it}, have heteroscedastic variances or serially correlated (Tables XII and XIII) and signal-to-noise ratio reduces (Table XIV). However, the performance of the factor model does improve when T increases (Table XV).

In short, when N and T are finite, the limited Monte Carlos show that generating counterfactuals based on a factor model using Bai and Ng (2002) model selection strategy appears to be sensitive to: (a) the signal-to-noise ratio; (b) the distribution of factor loading matrix, B; (c) the average values of B, $\frac{1}{N}\sum_{i=1}^{N} \underset{\sim}{b}_i'$; (d) the number of unknown factors, K; (e) the a priori assumed maximum number of unknown factors; (f) the serial correlations of the idiosyncratic components, ε_{it}; and (g) the heteroscedasticity of ε_{it}. On the other hand, our procedure of using $\tilde{y}_{it}$ in lieu of $\underset{\sim}{f}_t$ does not appear to be affected by any of these issues. On average, they yield much smaller prediction errors than the factor approach.

Table I(a). Optimal choice of m and the average PMSE, 2 stationary factors

	$\sigma^2 = 1$			$\sigma^2 = 0.5$			$\sigma^2 = 0.1$		
	AIC	AICC	20	AIC	AICC	20	AIC	AICC	20
$T_1 = 25$, $T = 35$									
Avg. #	11.726	4.432	—	11.418	4.468	—	11.39	4.664	—
Avg. R^2	0.9291	0.8412	0.9498	0.9436	0.881	0.9597	0.986	0.9699	0.9901
Avg. PMSE	6.1888	2.3592	7.9999	2.8749	1.2214	3.7846	0.6227	0.2434	0.8875
$T_1 = 40$, $T = 50$									
Avg. #	6.872	4.684	—	6.924	4.794	—	7.096	4.73	—
Avg. R^2	0.8175	0.7937	0.8531	0.8804	0.8644	0.904	0.9617	0.9563	0.9693
Avg. PMSE	1.8227	1.67	2.1695	0.9195	0.8492	1.0977	0.1903	0.1737	0.2197
$T_1 = 60$, $T = 70$									
Avg. #	6.236	5.098	—	6.266	5.056	—	6.206	5.108	—
Avg. R^2	0.7711	0.7607	0.8014	0.8549	0.8484	0.8738	0.9531	0.9516	0.9592
Avg. PMSE	1.4242	1.3901	1.5422	0.7205	0.7152	0.7781	0.1459	0.1425	0.1587

Table I(b). Frequency distribution of optimal number of m, 2 stationary factors

	$T_1 = 25$, $T = 35$						$T_1 = 40$, $T = 50$						$T_1 = 60$, $T = 70$					
	$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	1	12	1	14	0	2	5	10	0	4	1	6	3	4	1	5	0	0
2	4	59	3	55	5	47	13	34	8	36	8	30	14	31	10	23	7	16
3	9	103	4	89	15	104	23	81	22	82	20	75	28	61	38	64	35	60
4	17	107	16	112	11	109	44	122	42	110	49	125	59	95	50	105	69	110
5	14	81	15	96	18	90	72	97	77	113	79	120	88	120	88	112	89	126
6	17	73	21	65	27	65	82	86	104	73	65	74	96	91	98	95	84	94
7	28	33	37	42	33	43	77	41	66	41	71	42	86	45	81	56	88	60
8	25	19	26	19	29	25	62	20	56	29	57	17	56	29	64	24	55	20
9	34	7	47	3	29	7	46	5	47	10	54	10	34	18	35	10	43	12
10	37	6	40	3	46	5	28	4	30	0	45	1	15	2	19	5	19	2
11	36	0	44	2	29	2	20	0	25	1	27	0	10	3	8	1	6	0
12	45	0	43	0	41	1	17	0	13	1	10	0	9	1	6	0	5	0
13	46	0	35	0	37	0	6	0	4	0	12	0	1	0	2	0	0	0
14	39	0	37	0	44	0	2	0	3	0	2	0	1	0	0	0	0	0
15	45	0	44	0	38	0	1	0	3	0	0	0	0	0	0	0	0	0
16	35	0	28	0	39	0	1	0	0	0	0	0	0	0	0	0	0	0
17	38	0	23	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0
18	18	0	23	0	8	0	1	0	0	0	0	0	0	0	0	0	0	0
19	10	0	11	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0
20	2	0	2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0

A = AIC; B = AICC.

6. DATA

Because Hong Kong, by comparison, is a tiny city relative to other countries and regions, we believe whatever happened in Hong Kong will have no bearing on other countries. In other words, we expect Assumption 5 to hold. Therefore, we use the quarterly real growth rate of 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 to predict the quarterly real growth rate of Hong Kong in

DOI: 10.1002/jae

Table II(a). Optimal choice of m and the average PMSE, 3 stationary factors

	$\sigma^2 = 1$			$\sigma^2 = 0.5$			$\sigma^2 = 0.1$		
	AIC	AICC	20	AIC	AICC	20	AIC	AICC	20
$T_1 = 25$, $T = 35$									
Avg. #	11.19	4.438	—	11.696	4.58	—	11.704	4.738	—
Avg. R^2	0.9312	0.8601	0.9524	0.9587	0.9137	0.9705	0.9884	0.9737	0.9914
Avg. PMSE	6.0833	2.3475	8.3017	3.0528	1.243	4.1386	0.6458	0.2469	0.8117
$T_1 = 40$, $T = 50$									
Avg. #	6.692	4.614	—	7.01	4.714	—	7.05	4.834	—
Avg. R^2	0.8563	0.8368	0.8837	0.917	0.9048	0.933	0.9754	0.9716	0.9801
Avg. PMSE	1.8804	1.7127	2.2374	0.9663	0.874	1.1342	0.1857	0.1652	0.221
$T_1 = 60$, $T = 70$									
Avg. #	6.114	4.922	—	6.172	5.004	—	6.262	5.104	—
Avg. R^2	0.8255	0.8183	0.8485	0.8861	0.8809	0.9009	0.9679	0.9668	0.9721
Avg. PMSE	1.3887	1.3565	1.5384	0.7115	0.6971	0.7841	0.1397	0.1389	0.1552

Table II(b). Frequency distribution of optimal number of m, 3 stationary factors

	$T_1 = 25$, $T = 35$						$T_1 = 40$, $T = 50$						$T_1 = 60$, $T = 70$					
	$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	2	17	1	10	0	3	4	13	2	4	0	1	1	1	0	3	0	0
2	5	57	2	49	1	45	9	44	12	31	12	33	15	28	12	23	16	24
3	5	100	7	104	9	94	29	77	23	88	19	71	40	68	30	69	32	61
4	15	104	10	97	17	113	68	127	44	118	55	120	67	116	67	107	64	105
5	28	85	16	101	12	94	70	91	69	111	72	114	93	115	93	118	79	110
6	24	65	22	65	18	61	75	72	75	72	68	79	73	94	94	88	83	103
7	24	36	28	34	28	42	71	42	73	48	68	59	88	43	79	54	88	58
8	31	26	34	20	31	32	58	23	56	21	68	11	48	23	54	22	60	28
9	43	6	30	13	32	8	38	9	62	6	47	8	38	9	38	14	41	8
10	43	1	35	6	38	5	31	1	38	1	41	4	22	2	19	2	23	2
11	40	3	46	0	39	2	26	1	23	0	22	0	10	1	10	0	10	1
12	43	0	45	1	42	1	12	0	14	0	16	0	3	0	4	0	3	0
13	30	0	48	0	47	0	6	0	7	0	5	0	1	0	0	0	1	0
14	43	0	36	0	56	0	1	0	1	0	3	0	1	0	0	0	0	0
15	35	0	42	0	38	0	2	0	0	0	2	0	0	0	0	0	0	0
16	32	0	35	0	27	0	0	0	1	0	2	0	0	0	0	0	0	0
17	31	0	33	0	37	0	0	0	0	0	0	0	0	0	0	0	0	0
18	17	0	16	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0
19	7	0	11	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
20	2	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0

A = AIC; B = AICC.

the absence of intervention. All the nominal GDP and CPI are from OECD statistics, international financial statistics and CEIC database.

There are many ways to compute quarterly growth rates. One can either measure the change compared with the corresponding quarter in the previous year (year-on-year) or measure the change since the previous quarter (e.g., Neo, 2003). We note that the four quarters within one year have different numbers of working days and different countries have different seasonal effects on production and expenditure. For instance, Chinese New Year always falls in the first quarter and it is a big holiday for Hong Kong–virtually all business and government agencies are closed for

Table III(a). Optimal choice of m and the average PMSE, i.i.d. factor

	$\sigma^2 = 1$			$\sigma^2 = 0.5$			$\sigma^2 = 0.1$		
	AIC	AICC	20	AIC	AICC	20	AIC	AICC	20
$T_1 = 25$, $T = 35$									
Avg. #	11.614	4.92	—	11.738	5.03	—	11.904	4.972	—
Avg. R^2	0.9389	0.8714	0.9555	0.9657	0.9266	0.9759	0.99	0.9782	0.9929
Avg. PMSE	5.9299	2.601	8.0873	3.3748	1.3028	4.3716	0.6594	0.2617	0.8572
$T_1 = 40$, $T = 50$									
Avg. #	6.174	3.99	—	6.474	4.184	—	6.224	4.05	—
Avg. R^2	0.6729	0.6306	0.7402	0.7291	0.6883	0.7835	0.8651	0.8469	0.8932
Avg. PMSE	1.6635	1.4858	2.0566	0.8903	0.7674	1.0975	0.1696	0.1501	0.2107
$T_1 = 60$, $T = 70$									
Avg. #	6.82	5.61	—	7.016	5.744	—	7.19	5.868	—
Avg. R^2	0.8182	0.8101	0.8397	0.8919	0.8872	0.9048	0.9752	0.974	0.9781
Avg. PMSE	1.5597	1.5375	1.6305	0.8022	0.7854	0.8393	0.1590	0.156	0.1655

Table III(b). Frequency distribution of optimal number of m, i.i.d. factor

	$T_1 = 25$, $T = 35$						$T_1 = 40$, $T = 50$						$T_1 = 60$, $T = 70$					
	$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	0	4	1	3	0	0	8	26	8	27	9	21	1	3	0	0	0	0
2	5	33	1	20	2	21	33	77	23	67	25	72	5	10	4	6	0	3
3	3	73	6	77	5	76	50	113	35	104	36	95	18	34	12	28	12	27
4	11	121	10	110	8	134	63	104	65	108	66	128	38	75	35	80	23	62
5	13	104	18	110	19	98	72	84	68	80	71	95	67	112	67	123	55	114
6	18	69	23	81	22	72	73	58	73	56	82	51	94	126	89	109	111	134
7	35	49	32	49	29	55	54	23	71	33	65	25	103	78	106	88	99	90
8	24	26	22	34	19	32	44	9	42	14	47	10	77	44	74	35	87	45
9	42	13	37	12	29	5	28	3	37	7	41	3	46	14	54	23	45	16
10	37	5	52	2	48	5	36	1	30	2	27	0	30	4	33	5	36	8
11	41	2	35	2	40	2	18	2	24	0	15	0	13	0	16	2	18	1
12	50	0	45	0	47	0	11	0	12	1	11	0	5	0	3	1	10	0
13	59	1	34	0	46	0	6	0	6	1	3	0	2	0	6	0	3	0
14	46	0	41	0	44	0	1	0	4	0	2	0	1	0	0	0	0	0
15	37	0	42	0	36	0	1	0	0	0	0	0	0	0	1	0	1	0
16	27	0	33	0	35	0	0	0	2	0	0	0	0	0	0	0	0	0
17	18	0	22	0	32	0	2	0	0	0	0	0	0	0	0	0	0	0
18	18	0	25	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0
19	13	0	16	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
20	3	0	5	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0

A = AIC; B = AICC.

celebration, but not so for other countries. Since our data are non-seasonally adjusted and our interest is in finding the long-term trend, we compute the quarterly growth rate by measuring the change compared with the corresponding quarter in the previous year.

7. EMPIRICAL ANALYSIS

In this section we illustrate the use of our panel data approach for program evaluation by considering the impact on Hong Kong real GDP growth rate with the reversion of sovereignty on

DOI: 10.1002/jae

Table IV(a). Optimal choice of m and the average PMSE, nearly non-stationary factor

	$\sigma^2 = 1$			$\sigma^2 = 0.5$			$\sigma^2 = 0.1$		
	AIC	AICC	20	AIC	AICC	20	AIC	AICC	20
$T_1 = 25$, $T = 35$									
Avg. #	11.692	4.28	—	11.728	4.156	—	11.452	4.206	—
Avg. R^2	0.9123	0.8048	0.9374	0.9305	0.8391	0.9504	0.962	0.9222	0.9728
Avg. PMSE	6.2022	2.2714	8.4701	3.2219	1.1118	4.1398	0.7399	0.2213	0.935
$T_1 = 40$, $T = 50$									
Avg. #	6.358	4.16	—	6.342	4.114	—	6.488	4.214	—
Avg. R^2	0.7955	0.7696	0.8378	0.8528	0.8332	0.8827	0.9237	0.9147	0.9393
Avg. PMSE	1.7769	1.6214	2.1585	0.851	0.7574	1.0292	0.1786	0.156	0.2124
$T_1 = 60$, $T = 70$									
Avg. #	5.468	4.308	—	5.432	4.302	—	5.472	4.272	—
Avg. R^2	0.77	0.7602	0.8019	0.8239	0.8159	0.8476	0.9139	0.9101	0.9259
Avg. PMSE	1.3278	1.2805	1.4846	0.6737	0.6517	0.7699	0.1286	0.125	0.1457

Table IV(b). Frequency distribution of optimal number of m, nearly non-stationary factor

	$T_1 = 25$, $T = 35$						$T_1 = 40$, $T = 50$						$T_1 = 60$, $T = 70$					
	$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$		$\sigma^2 = 1$		$\sigma^2 = 0.5$		$\sigma^2 = 0.1$	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	0	25	2	32	3	25	4	22	2	19	6	15	11	17	6	11	8	20
2	3	73	7	72	7	74	28	75	30	72	22	73	30	55	40	65	24	56
3	12	99	12	101	5	99	47	99	39	98	41	94	53	101	58	102	75	94
4	9	106	11	103	16	102	57	105	64	130	52	124	87	113	70	116	75	115
5	22	78	12	73	24	86	65	84	80	77	76	95	90	100	86	91	87	106
6	25	41	24	62	31	56	74	63	71	59	72	42	81	58	96	56	79	65
7	25	39	19	24	19	26	66	32	58	26	67	32	56	30	56	33	68	21
8	33	23	36	19	33	18	51	13	51	13	51	15	39	20	43	19	32	15
9	27	8	39	11	31	10	33	5	42	5	38	6	31	6	24	5	26	5
10	22	6	32	2	44	2	35	2	24	1	37	2	12	0	15	1	14	3
11	48	1	37	1	33	1	24	0	14	0	14	1	6	0	4	1	7	0
12	40	1	32	0	34	1	11	0	11	0	8	1	2	0	2	0	3	0
13	44	0	47	0	36	0	3	0	6	0	12	0	1	0	0	0	1	0
14	40	0	40	0	38	0	1	0	6	0	4	0	1	0	0	0	1	0
15	46	0	43	0	40	0	1	0	2	0	0	0	0	0	0	0	0	0
16	39	0	34	0	33	0	0	0	0	0	0	0	0	0	0	0	0	0
17	35	0	28	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0
18	16	0	17	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0
19	11	0	21	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0
20	3	0	7	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0

A = AIC; B = AICC.

1 July 1997 from the UK to China and the implementation of CEPA starting in 2004:Q1 between mainland China and Hong Kong. (We present the evaluation using the factor approach in Appendix B.) We first wish to evaluate the impact of change of sovereignty on real GDP had Hong Kong stayed under British rule. Since there are only 18 observations between 1993:Q1 and 1997:Q2, we limit the countries under consideration for constructing counterfactuals to China, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, Taiwan, Thailand and the USA–countries that are either in the region or economically closely associated with Hong Kong. Using AICC, we select Japan, Korea, the USA and Taiwan to construct the hypothetical growth path of Hong Kong

DOI: 10.1002/jae

Table V. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 1$, $F \sim N(0, 1)$, $B \sim N(0, 1)$, $k_{\max} = 8$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.8246	5.5135	11.092
AICC	0.6386	2.0310	4.034
PC1	0.5245	1.5724	7.880
PC2	0.4992	1.4896	7.012
PC3	0.5278	1.5799	8
IC1	0.3346	1.1108	1.356
IC2	0.3262	1.0767	1.002
IC3	0.5278	1.5799	8

Table VI. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 1$, $F \sim N(0, 1)$, $B \sim N(0, 1)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.8396	5.6705	11.392
AICC	0.6497	2.0088	4.112
PC1	0.8882	7.9356	20
PC2	0.8882	7.9356	20
PC3	0.8882	7.9356	20
IC1	0.8882	7.9356	20
IC2	0.8882	7.9356	20
IC3	0.8882	7.9356	20

Table VII. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(0, 1)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.9349	7.2288	12.1520
AICC	0.8584	3.3378	5.366
PC1	0.9527	9.2582	20
PC2	0.9527	9.2582	20
PC3	0.9527	9.2582	20
IC1	0.9527	9.2582	20
IC2	0.9527	9.2582	20
IC3	0.9527	9.2582	20

Table VIII. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(0.3, 1)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.9386	8.5870	12.018
AICC	0.8655	3.0519	5.162
PC1	0.9549	11.6829	20
PC2	0.9549	11.6829	20
PC3	0.9549	11.6829	20
IC1	0.9549	11.6829	20
IC2	0.9549	11.6829	20
IC3	0.9549	11.6829	20

Table IX. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(1, 1)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.9590	8.4321	12.296
AICC	0.9074	3.7535	5.524
PC1	0.9691	11.1183	20
PC2	0.9691	11.1183	20
PC3	0.9691	11.1183	20
IC1	0.9691	11.1183	20
IC2	0.9691	11.1183	20
IC3	0.9691	11.1183	20

Table X. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim U(-1, 1)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.8820	7.8766	11.98
AICC	0.7355	2.8996	4.864
PC1	0.9147	9.9083	20
PC2	0.9147	9.9083	20
PC3	0.9147	9.9083	20
IC1	0.9147	9.9083	20
IC2	0.9147	9.9083	20
IC3	0.9147	9.9083	20

Table XI. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(2, 2)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.9821	9.4982	12.39
AICC	0.9612	5.0858	6.034
PC1	0.9865	11.8991	20
PC2	0.9865	11.8991	20
PC3	0.9865	11.8991	20
IC1	0.9865	11.8991	20
IC2	0.9865	11.8991	20
IC3	0.9865	11.8991	20

Table XII. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(0, 1)$, $k_{\max} = 20$, $\sigma \sim U(1, 4)$

	R^2	PMSE	Average number of regressors
AIC	0.8338	41.4836	11.57
AICC	0.6382	16.4713	4.466
PC1	0.8810	55.8442	20
PC2	0.8810	55.8442	20
PC3	0.8810	55.8442	20
IC1	0.8810	55.8442	20
IC2	0.8810	55.8442	20
IC3	0.8810	55.8442	20

DOI: 10.1002/jae

Table XIII. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(0, 1)$, $k_{\max} = 20$, $e_{i,t} = 0.5 \times e_{I,t-1} + v_{i,t}$, $\sigma_v = 1$

	R^2	PMSE	Average number of regressors
AIC	0.9445	10.5882	12.188
AICC	0.8779	4.3454	5.744
PC1	0.9589	13.8675	20
PC2	0.9589	13.8675	20
PC3	0.9589	13.8675	20
IC1	0.9589	13.8675	20
IC2	0.9589	13.8675	20
IC3	0.9589	13.8675	20

Table XIV. Prediction comparison of model with $N = 20$, $T_0 = 25$, $T = 35$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(0, 1)$, $k_{\max} = 20$, $\sigma = 5$

	R^2	PMSE	Average number of regressors
AIC	0.8510	32.0960	11.77
AICC	0.6615	13.5590	4.618
PC1	0.8919	42.9199	20
PC2	0.8919	42.9199	20
PC3	0.8919	42.9199	20
IC1	0.8919	42.9199	20
IC2	0.8919	42.9199	20
IC3	0.8919	42.9199	20

Table XV. Prediction comparison of model with $N = 20$, $T_0 = 60$, $T = 70$, $K = 5$, $F \sim N(0, 1)$, $B \sim N(1, 1)$, $k_{\max} = 20$, $\sigma = 1$

	R^2	PMSE	Average number of regressors
AIC	0.8865	1.8256	8.024
AICC	0.8799	1.8423	6.66
PC1	0.8986	1.8789	20
PC2	0.8986	1.8789	20
PC3	0.8986	1.8789	20
IC1	0.8986	1.8789	20
IC2	0.8986	1.8789	20
IC3	0.8986	1.8789	20

Table XVI. AICC: weights of control groups for the period 1993:Q1–1997:Q2

	Beta	SD	T
Constant	0.0263	0.017	1.5427
Japan	−0.676	0.1117	−6.0522
Korea	−0.4323	0.0634	−6.8211
USA	0.486	0.2195	2.2141
Taiwan	0.7926	0.3099	2.5576

$R^2 = 0.9314$; AICC $= -171.771$.

Table XVII. AICC: treatment effect of political integration 1997:Q3–2003:Q4

	Actual	Control	Treatment
1997:Q3	0.061	0.0798	−0.0188
1997:Q4	0.014	0.081	−0.067
1998:Q1	−0.032	0.1294	−0.1614
1998:Q2	−0.061	0.1433	−0.2043
1998:Q3	−0.081	0.1319	−0.2129
1998:Q4	−0.065	0.139	−0.204
1999:Q1	−0.029	0.0876	−0.1166
1999:Q2	0.005	0.067	−0.062
1999:Q3	0.039	0.04	−0.001
1999:Q4	0.083	0.0445	0.0385
2000:Q1	0.107	0.0434	0.0636
2000:Q2	0.075	0.0398	0.0352
2000:Q3	0.076	0.0524	0.0236
2000:Q4	0.063	0.0318	0.0312
2001:Q1	0.027	0.0118	0.0152
2001:Q2	0.015	−0.0177	0.0327
2001:Q3	−0.001	−0.0177	0.0167
2001:Q4	−0.017	0.0184	−0.0354
2002:Q1	−0.01	0.0314	−0.0414
2002:Q2	0.005	0.05	−0.045
2002:Q3	0.028	0.0577	−0.0297
2002:Q4	0.048	0.0346	0.0134
2003:Q1	0.041	0.0538	−0.0128
2003:Q2	−0.009	0.0251	−0.0341
2003:Q3	0.038	0.0628	−0.0248
2003:Q4	0.047	0.0761	−0.0291
Mean	0.018	0.0576	−0.0396
SD	0.0478	0.0429	0.0787
T	0.3761	1.3417	−0.5034

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 $= -180.986$.

had there been no change of sovereignty. The OLS weights based on 1993:Q1–1997:Q2 data are reported in Table XVI and the estimated treatment effects are reported in Table XVII. The actual and hypothetical growth paths for the period 1993:Q1–1997:Q2 and 1997:Q3–2003:Q4 are plotted in Figures 1 and 2, respectively. Because the treatment effects appear to be serially correlated (see Figure 3), we fit an AR(2) model for the estimated treatment effects:

$$\hat{\Delta}_{1t} = \underset{(0.0068)}{-0.0063} + \underset{(0.1559)}{1.459}\hat{\Delta}_{1,t-1} - \underset{(0.1558)}{0.6547}\hat{\Delta}_{1,t-2} + \hat{\eta}_t \tag{46}$$

where estimated standard errors are in parentheses. The implied long-run effects is -0.032. However, the t-statistic is only -1.04, which is not statistically significant.

DOI: 10.1002/jae

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

Using the AIC criterion, the selected countries are Japan, Korea, Philippines, Taiwan, and the USA. The OLS estimates of the weights are given in Table XVIII and treatment effects in Table XIX. The actual and hypothetical growth paths for 1993:Q1–1997:Q2 and 1997:Q3–2003:Q4 are plotted in Figures 4 and 5, respectively. Again, the estimated treatment 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 real GDP growth in Hong Kong appears to be approximated well by the chosen controls before treatment by either criterion. The estimated treatment effects are not statistically significant. Therefore, we may conclude that the political integration of Hong Kong with mainland China does not appear to have any significant impact on Hong Kong's economic growth. The lack of intervention effects is hardly surprising given the 'one country, two systems' concept proposed by Deng Xiaoping. It is generally recognized that, apart from change of national flags, Hong Kong's institutional arrangements were basically left untouched during this period. 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.

Given that we do not find any effect of the change of sovereignty, we can pool the data of 1993:Q1 with 2003:Q4 to examine the effect of the CEPA including Individual Travel Scheme

DOI: 10.1002/jae

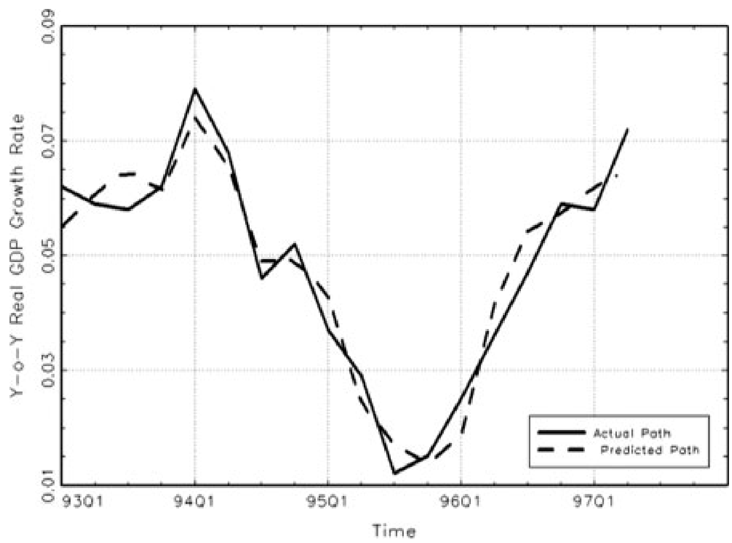

Figure 1. AICC: actual and predicted real GDP from 1993:Q1 to 1997:Q2

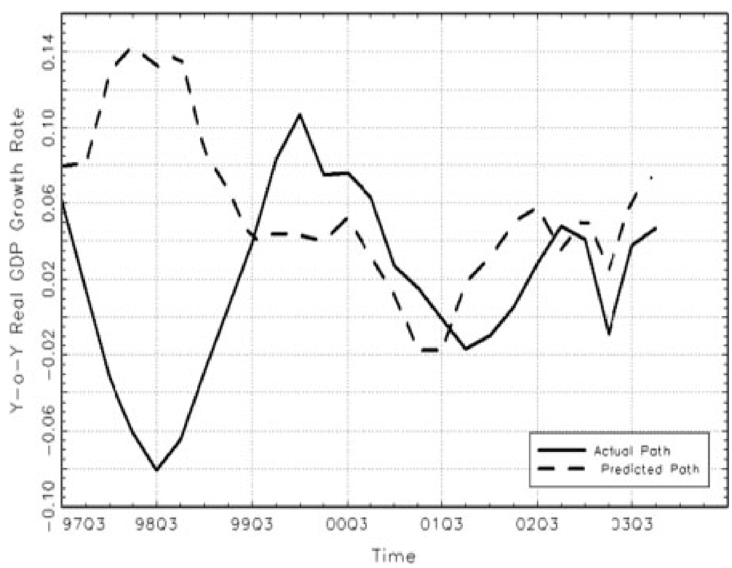

Figure 2. AICC : actual and counterfactual real GDP from 1997:Q3 to 2003:Q4

and Removal of Preferential Tariff which was signed on 29 June 2003, but implementation only started on 1 January 2004. Since we now have more degrees of freedom, we can use the model selection strategy discussed in Section 5 to generate the hypothetic growth path for Hong Kong had there been no CEPA with mainland China. Using the AICC criterion, the countries selected are Austria, Italy, Korea, Mexico, Norway and Singapore. OLS estimates of the weights are reported in Table XX. Actual and predicted growth path from 1993:Q1 to 2003:Q4 are plotted in Figure 7. The availability of more pre-intervention period data appears to allow more accurate estimates of the country weights and better tracing of the pre-intervention path. The estimated quarterly treatment effects are reported in Table XXI. The actual and predicted counterfactuals

DOI: 10.1002/jae

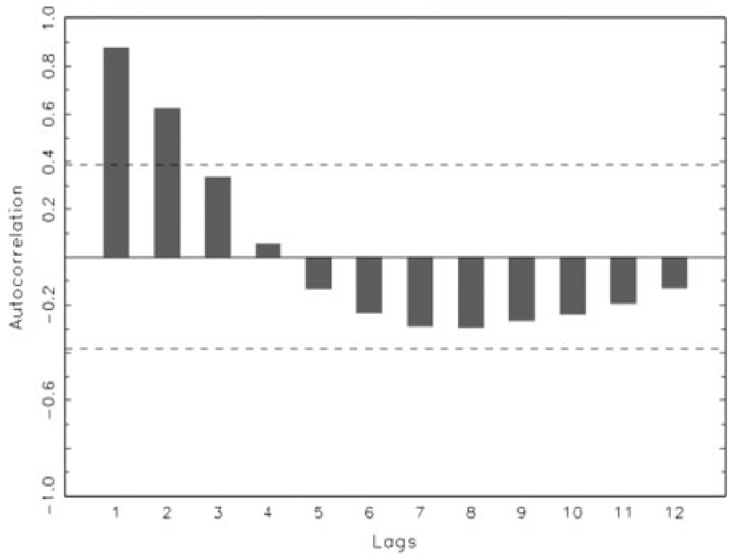

Figure 3. AICC: autocorrelation of treatment effect from 1997:Q3 to 2003:Q4

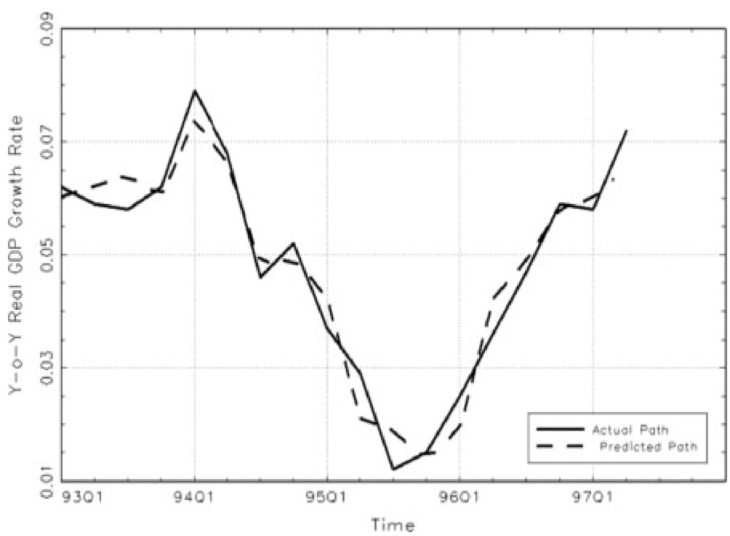

Figure 4. AIC: actual and predicted real GDP from 1993:Q1 to 1997:Q2

for the period 2004:Q1 to 2008:Q1 are presented in Figure 8. Figure 9 shows autocorrelation of the residuals of the treatment effect model. Using the AIC criterion, the selected group consists of Austria, Germany, Italy, Korea, Mexico, Norway, Philippines, Singapore and Switzerland. The OLS estimates of the weights are given in Table XXII and the estimated quarterly treatment effects are in Table XXIII. The pre- and post-intervention actual and predicted outcomes are plotted in Figures 10 and 11. It is notable that both groups of countries trace closely the actual Hong Kong path before the implementation of CEPA (with R^2 above 0.93). It is also quite remarkable that the post-sample predictions closely matched the actual turning points at a lower level for the treatment period even though no Hong Kong data were used. The CEPA effect at each quarter was all positive and the residuals of CEPA effect appeared to be serially uncorrelated (see Figures 8

DOI: 10.1002/jae

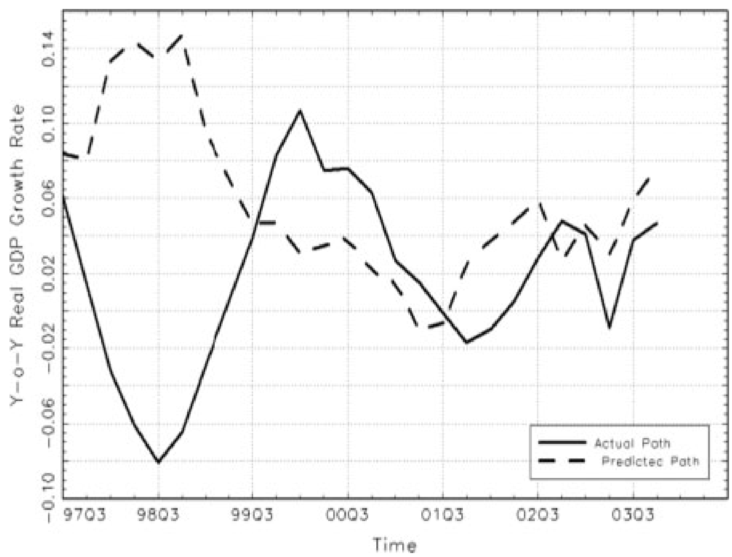

Figure 5. AIC: actual and counterfactual real GDP from 1997:Q3 to 2003:Q4

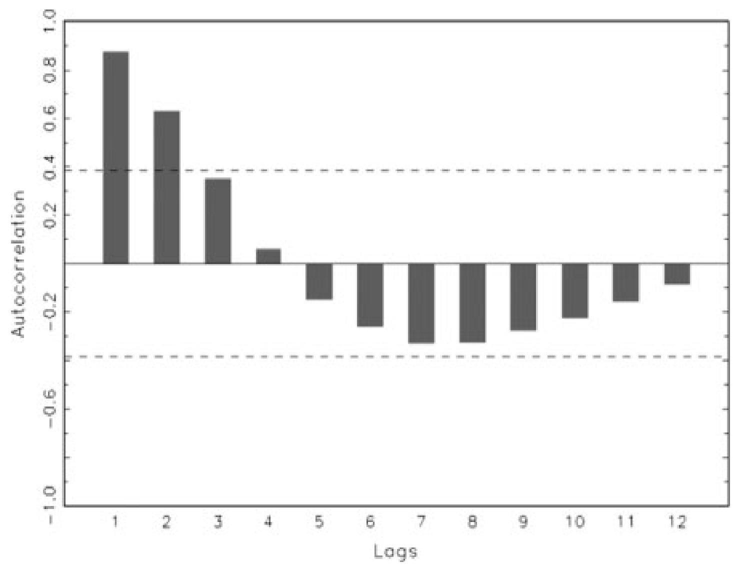

Figure 6. AIC: autocorrelation of treatment effect from 1997:Q3 to 2003:Q4

and 11; Figures 9 and 12). The average actual growth rate from 2004:Q1 to 2008:Q1 is 7.26%. The average projected growth rate without CEPA is 3.23% using the group of countries selected by AICC and 3.47% using the group selected by AIC. The estimated average treatment effect is 4.03% with a standard error of 0.016 based on the AICC group and 3.79% with a standard error of 0.0151 based on the AIC group. The *t*-statistic is 2.5134 for the former group and 2.5122 for the latter group. Either set of countries yields similar predictions and highly significant CEPA effects. In other words, through liberalization and increased openness with mainland China, the real GDP growth rate of Hong Kong is raised by more than 4% compared to the growth rate had there been no CEPA agreement with mainland China.

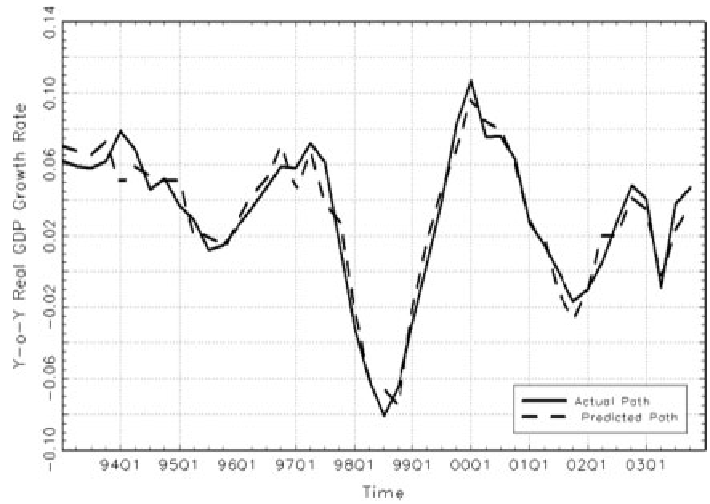

Figure 7. AICC: actual and predicted real GDP from 1993:Q1 to 2003:Q4

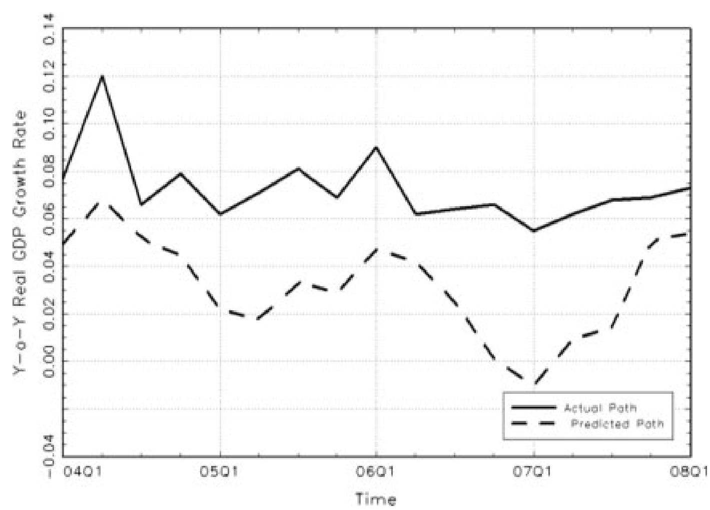

Figure 8. AICC: actual and counterfactual real GDP from 2004:Q1 to 2008:Q1

The Hong Kong government statistics appear to corroborate this finding. A recent paper (Hong Kong Legislative Council Panel on Commerce and Industry paper, 2006) shows that the tariff-free access of goods produced in Hong Kong has stimulated rising capital investment from HK \$103 million in 2005, to HK \$202 million in 2006 and HK \$239 million in 2007. Liberalization to trade in services has further stimulated capital investment. Capital investments in transport, logistics, distribution, advertising and construction stood at HK \$1.0 billion in 2004, but were at HK \$2.4 billion in 2007. The Individual Visit Scheme (IVS) has led to a substantial increase of tourism from China. From the implementation of the scheme to the end of 2006, mainland Chinese visitors have made 17.2 million trips to Hong Kong. IVS visitor spending in 2006 was HK \$9.3 billion–about 38% higher than 2004. Moreover, the implementation of CEPA also helped to rebuild confidence in the economy after a prolonged period of economic stagnation. For instance, the value of total receipts for the restaurant sector in 2008:Q1 was up by 15.8% compared with

DOI: 10.1002/jae

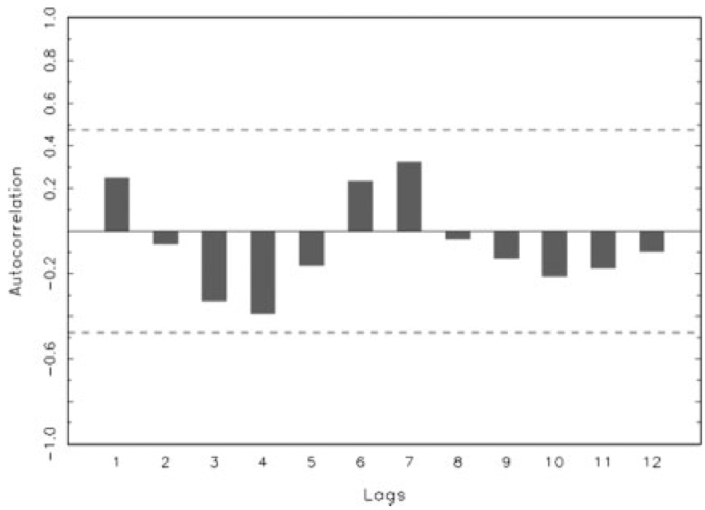

Figure 9. AICC: autocorrelation of treatment effect residuals from 2004:Q1 to 2008:Q1

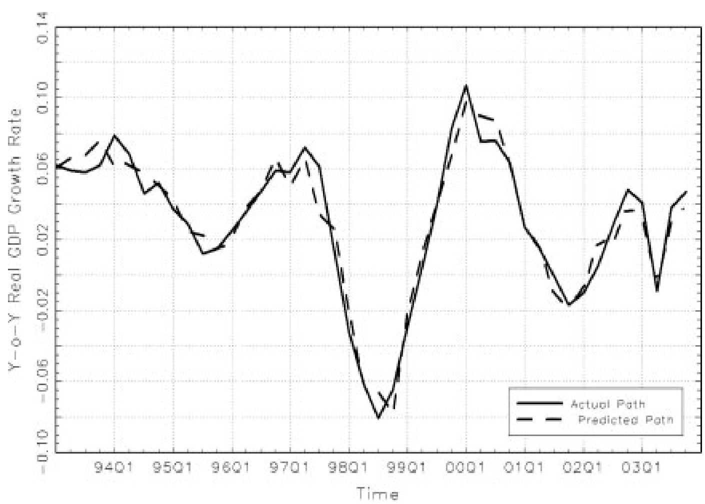

Figure 10. AIC: actual and predicted real GDP from 1993:Q1 to 2003:Q4

2007:Q1 and the value of total retail sales in March 2008 increased by 20% compared with a year earlier. If the fundamental relations between the aggregate and components stay the same before and after 2004:Q1, then one should expect the relative contribution of additional unit increase of each component to the aggregates to stay the same in those two periods, and the impact of CEPA would be just on the value of such component, not on the component's impact on the aggregate. However, a simple sectoral analysis of regressing log(real GDP) on log(re-export from China), log(import) and log(number of visitors) also shows a statistically significant change in the impact of log(number of Chinese visitors) from 0.0638 in the pre-CEPA period to 0.1663 for the post-CEPA period, with highly significant t-values. In other words, it appears that IVS is the most important component of CEPA; its impact is not just on increasing tourist revenue but also in serving to raise the confidence level of Hong Kong consumers and investors. As a result, the unemployment rate has dropped from 7.9% in 2003 to 4.8% in 2006 and 4.2% in September–November 2007.

DOI: 10.1002/jae

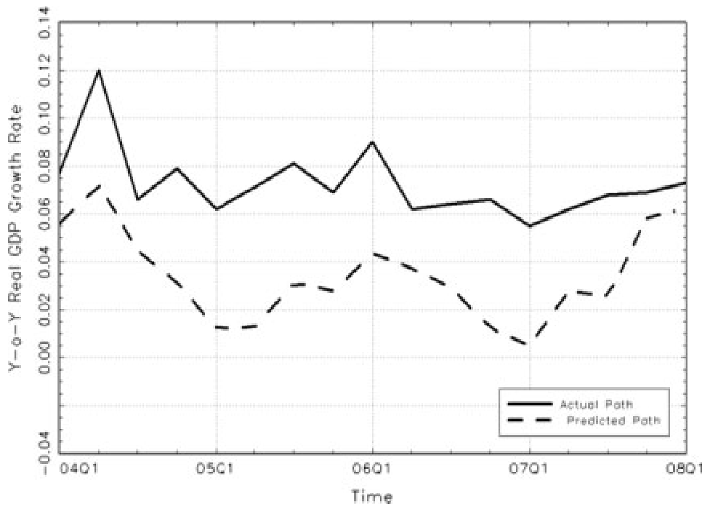

Figure 11. AIC: actual and counterfactual real GDP from 2004:Q1 to 2008:Q1

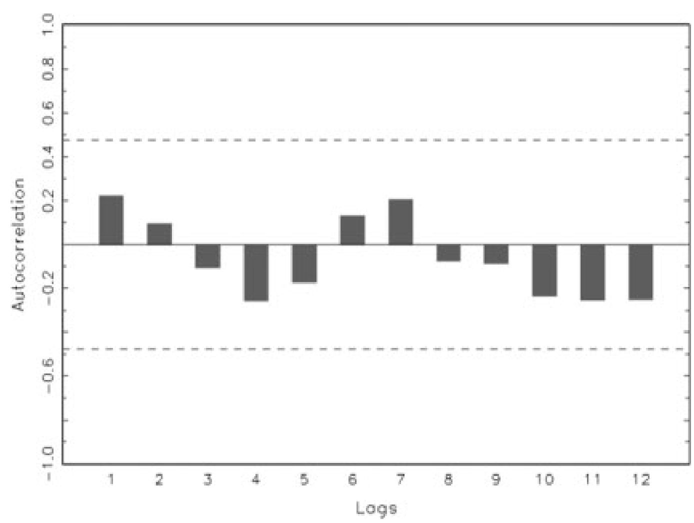

Figure 12. AIC: autocorrelation of treatment effect residuals from 2004:Q1 to 2008:Q1

The per capita income in 2006 reached US $27,604. The Hang Seng Index at the end of 2007 reached 27812.

8. CONCLUSION

In this paper we have proposed a panel data approach to assess the impact of a policy intervention. We demonstrate that the dependence among cross-sectional units can be utilized to construct the counterfactuals. We identify the source of cross-sectional correlations through a factor framework. However, if sample size is finite, there may be an advantage in just using observed $\tilde{y}_t$ because the impacts of unobserved factors, $\tilde{f}_t$, are already embedded in $\tilde{y}_t$. In this approach, there is no need to distill the fundamental factors and their factor loading matrix as in Bai (2003), Bai and Ng

DOI: 10.1002/jae

Table XX. AICC: weights of control groups for the period 1993:Q1–2003:Q4

	Beta	SD	T
Constant	−0.0019	0.0037	−0.524
Austria	−1.0116	0.1682	−6.0128
Italy	−0.3177	0.1591	−1.9971
Korea	0.3447	0.0469	7.3506
Mexico	0.3129	0.051	6.1335
Norway	0.3222	0.0538	5.9912
Singapore	0.1845	0.0546	3.3812

$R^2 = 0.931$, AICC $= -378.9427$.

Table XXI. AICC: treatment effect for economic integration 2004:Q1–2008:Q1

	Actual	Control	Treatment
2004:Q1	0.077	0.0493	0.0277
2004:Q2	0.12	0.0686	0.0514
2004:Q3	0.066	0.0515	0.0145
2004:Q4	0.079	0.0446	0.0344
2005:Q1	0.062	0.0217	0.0403
2005:Q2	0.071	0.0177	0.0533
2005:Q3	0.081	0.0333	0.0477
2005:Q4	0.069	0.029	0.04
2006:Q1	0.09	0.0471	0.0429
2006:Q2	0.062	0.0417	0.0203
2006:Q3	0.064	0.025	0.039
2006:Q4	0.066	0.0009	0.0651
2007:Q1	0.055	−0.0101	0.0651
2007:Q2	0.062	0.0092	0.0528
2007:Q3	0.068	0.0143	0.0537
2007:Q4	0.069	0.0508	0.0182
2008:Q1	0.073	0.0538	0.0192
Mean	0.0726	0.0323	0.0403
SD	0.0149	0.0213	0.016
T	4.8814	1.5132	2.5134

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 $= -385.7498$.

(2002), Bernanke and Boivin (2003), etc. The method is easy to implement and inference appears quite robust.

We illustrate our methodology by considering the political and economic intervention effects on Hong Kong's economy. We find that the change of sovereignty in 1997 hardly had any impact on

DOI: 10.1002/jae

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

Hong Kong's economy. On the other hand, the implementation of the CEPA agreement in 2004 has had a significant impact. Hong Kong's real GDP growth rate is 4% higher than what it would have been in the absence of CEPA.

ACKNOWLEDGEMENTS

We would like to thank T. S. Wang of the Chinese Academy of Social Sciences and Y. Fan and M. Guo of Xiamen University for assistance in obtaining data for China; also L. C. Lau of the Pearl River and Yangtze River Delta Collection, City University of Hong Kong, for assistance in obtaining the data for Hong Kong. We would also like to thank a co-editor, two referees and M. H. Pesaran for helpful comments on the early version.

REFERENCES

Abadie A, Gardeazabal J. 2003. The economic costs of conflict: a case study of the Basque Country. *American Economics Review* **93**: 113–132.

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**: 493–505.

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory*, Petrov BN, Csaki F (eds). Akadémiai Kiadó: Budapest, 267–281.

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC19**: 716–723.

Amemiya T. 1985. *Advanced Econometrics*. Harvard University Press: Cambridge, MA.

Anderson TW. 2003. *An Introduction to Multivariate Statistical Analysis*. Wiley: Hoboken, NJ.

Bai J. 2003. Inferential theory for factor models of large dimensions. *Econometrica* **71**: 135–171.

Bai J, Ng S. 2002. Determining the number of factors in approximate factor models. *Econometrica* **70**: 191–221.

Bernanke B, Boivin J. 2003. Monetary policy in a data rich environment. *Journal of Monetary Economics* **50**: 525–546.

Box GEP, Jenkins GM. 1970. *Time Series Analysis, Forecasting and Control*. Holden Day: San Francisco, CA.
Buhlmann P. 2006. Boosting for high-dimensional linear models. *Annals of Statistics* **54**: 559–583.
Card D, Krueger AB. 1994. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* **84**: 772–793.
Chudik A, Pesaran MH, Tosetti E. 2010. Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal* (forthcoming).
Forni M, Reichlin L. 1998. Let's get real: a factor-analytic approach to disaggregrated business cycle dynamics. *Review of Economic Studies* **65**: 453–473.
Giannone D, Reichlin L, Sala L. 2005. Monetary policy in real time. In *NBER Macroeconomics Annual 2004*, Gertler M, and Rogoff K (eds). MIT Press: Cambridge, MA; 161–200.
Giannone D, Lenza M, Reichlin L. 2010. Business cycles in the euro area. In *Europe, and the Euro*, Alesina A, Giavazzi F (eds). National Bureau of Economic Research: Cambridge, MA; 141–167.
Goldberger A. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* **57**: 369–375.
Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**: 424–438.
Gregory A, Head A. 1999. Fluctuations in productivity, investment, and the current account. *Journal of Monetary Economics* **44**: 423–452.
Heckman JJ, Hotz VJ. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *Journal of the American Statistical Association* **84**: 862–874.
Heckman JJ, Vytlacil EJ. 2001. Local instrumental variables. In *Nonlinear Statistical Modeling* Hsiao C, Morimune K, Powell JL (eds). Cambridge University Press: Cambridge, UK; 1–46.
Hong Kong Legislative Council Panel on Commerce and Industry. 2006. Mainland and Hong Kong Closer Economic Partnership Arrangement (CEPA): impact on the Hong Kong economy. LC Paper No. CD(1) 1849/06-07(04).
Hurvich and Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* **76**(2): 297–307.
Hsiao C. 1982. Autoregressive modelling and causal ordering of economic variables. *Journal of Economic Dynamics and Control* **4**: 243–259.
Hsiao C. 2003. *Analysis of Panel Data* (2nd edn). *Econometric Society Monograph* no. 34. Cambridge University Press: New York.
Hsiao C, Tahmiscioglu AK. 1997. A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* **92**: 455–465.
Hsiao C, Wan SK. 2010. Is there an optimal forecast combination? Mimeo. University of Southern California.
Jao YC. 2001. *The Asian Financial Crisis and the Ordeal of Hong Kong*. Quorum/Greenwood: Westport, CT.
Neo PC. 2003. Quarterly growth rates. *Statistics Singapore Newsletter* March: 7–10.
Newey WK, West KD. 1987. A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**: 703–708.
Onatski A. 2009. Testing hypotheses about the number of factors in large factor models. *Econometrica* **77**: 1447–1479.
Pesaran MH, Smith LV, Smith RP. 2007. What if the UK or Sweden had joined the euro in 1999? An empirical evaluation using a global VAR. *International Journal of Finance and Economics* **12**: 55–87.
Phillips PCB, Durlauf SN. 1986. Multiple time series with integrated processes. *Review of Economic Studies* **53**(4): 473–495.
Rao CR. 1973. *Linear Statistical Inference and its Applications* (2nd edn). Wiley: New York.
Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.
Sargent TJ, Sims CA. 1977. Business cycle modeling without pretending to have too much a-priori economic theory. In *New Methods in Business Cycle Research*, Sims C *et al*. (eds). Federal Reserve Bank of Minneapolis: Minneapolis, MN.
Sims CA. 1980. Macroeconomics and reality. *Econometrica* **48**: 1–48.
Stock JH, Watson MW. 2005. Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467, July 2005: 1–65.
Sung YW, Wong KY. 2000. Growth of Hong Kong before and after its reversion to China: the China factor. *Pacific Economic Review* **5**: 201–228.
Tibshirani R. 1996. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society* **B58**: 267–288.

APPENDIX A

In this appendix we prove Lemma 6. Using the notation of Section 5, decompose $Y = (Y_1, Y_2)$. Noting that $\underset{\sim}{a} = (BB' + \Omega)^{-1}B\tilde{\underset{\sim}{b}}_1'$ and $a_1 = (B_1B_1' + \Omega_1)^{-1}B_1\tilde{\underset{\sim}{b}}_1'$ minimize $E[\underset{\sim}{y}_1^0 - Y\underset{\sim}{a}]'[\underset{\sim}{y}_1^0 - Y\underset{\sim}{a}]$ and $E[\underset{\sim}{y}_1^0 - Y_1\underset{\sim}{a}_1]'[\underset{\sim}{y}_1^0 - Y_1a_1]$, respectively, where $\tilde{\underset{\sim}{b}}_1' = (1, \underset{\sim}{b}_1')$, and $\Omega = E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \begin{bmatrix} \Omega_1 & \underset{\sim}{0} \\ \underset{\sim}{0}' & \Omega_2 \end{bmatrix}$. Then

$$\text{MSE}(Y_1) = \sigma_1^2 + \tilde{\underset{\sim}{b}}_1'[I - B_1'(B_1B_1' + \Omega_1)^{-1}B_1]\tilde{\underset{\sim}{b}}_1 \tag{A.1}$$

$$\text{MSE}(Y) = \sigma_1^2 + \tilde{\underset{\sim}{b}}_1'[I - B'(BB' + \Omega)^{-1}B]\tilde{\underset{\sim}{b}}_1 \tag{A.2}$$

We first show that if $m < K$, then $\text{MSE}(Y) < \text{MSE}(Y_1)$, $\text{MSE}(Y) < \text{MSE}(Y_1)$ holds iff

$$\begin{aligned} & I - B'(BB' + \Omega)^{-1}B < I - B_1'(B_1B_1' + \Omega_1)^{-1}B_1 \\ \Leftrightarrow & \begin{pmatrix} (B_1B_1' + \Omega_1)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} - \begin{pmatrix} B_1B_1' + \Omega_1 & B_1B_2' \\ B_2B_1' & B_2B_2' + \Omega_2 \end{pmatrix}^{-1} < 0 \\ \Leftrightarrow & \begin{pmatrix} A & X \\ X' & C \end{pmatrix} < 0 \end{aligned} \tag{A.3}$$

where

$$\begin{aligned} A &= (B_1B_1' + \Omega_1)^{-1} - (B_1M_2B_1' + \Omega_1)^{-1} \\ X &= (B_1M_2B_1' + \Omega_1)^{-1}B_1B_2'(B_2B_2' + \Omega_2)^{-1} \\ X' &= (B_2B_2' + \Omega_2)^{-1}B_2B_1'(B_1M_2B_1' + \Omega_1)^{-1} \\ C &= -(B_2B_2' + \Omega_2)^{-1}B_2B_1'(B_1M_2B_1' + \Omega_1)^{-1}B_1B_2'(B_2B_2' + \Omega_2)^{-1} - (B_2B_2' + \Omega_2)^{-1} \\ M_2 &= I_K - B_2'(B_2B_2' + \Omega_2)^{-1}B_2 \end{aligned} \tag{48}$$

Remark 1. $M_2 = I_K - B_2'(B_2B_2' + \Omega_2)^{-1}B_2 = I_K - B_2'D^{-\frac{1}{2}}D^{-\frac{1}{2}}B_2 \equiv I_K - U'U$, where $U = D^{-\frac{1}{2}}B_2$ and $D = B_2B_2' + \Omega_2$.

By Theorem A.3.5 (Anderson, 2003, p. 639), the conditions for $I_K - U'U$ and $I_L - UU'$ to be positive definite are the same. Thus

$$I_L - UU' = I_L - D^{-\frac{1}{2}}B_2B_2'D^{-\frac{1}{2}} > 0 \Leftrightarrow D - B_2B_2' = \Omega_2 > 0$$

Since the last statement holds, we conclude that $M_2 > 0$.

$$P_2 \equiv I_K - M_2 = B_2'(B_2B_2' + \Omega_2)^{-1}B_2 > 0$$

Thus $(B_1B_1' + \Omega_1) - (B_1M_2B_1' + \Omega_1) = B_1(I_K - M_2)B_1' = B_1P_2B_1' > 0$. Then, $A = (B_1B_1' + \Omega_1)^{-1} - (B_1M_2B_1' + \Omega_1)^{-1} < 0$.

Remark 2. $-C = (B_2B_2' + \Omega_2)^{-1}B_2B_1'(B_1M_2B_1' + \Omega_1)^{-1}B_1B_2'(B_2B_2' + \Omega_2)^{-1} + (B_2B_2' + \Omega_2)^{-1} > 0$.

$$\Leftrightarrow B_2B_1'(B_1M_2B_1' + \Omega_1)^{-1}B_1B_2'(B_2B_2' + \Omega_2)^{-1} + I_L > 0$$

$$\Leftrightarrow B_2B_1'(B_1M_2B_1' + \Omega_1)^{-1}B_1B_2' + (B_2B_2' + \Omega_2) > 0$$
$$\Leftrightarrow B_2[B_1'(B_1M_2B_1' + \Omega_1)^{-1}B_1 + I_K]B_2' + \Omega_2 > 0$$

Therefore C is a negative definite matrix.

Remark 3. The Schur complement: $S = -C - (-X')(-A)^{-1}(-X) = -(C - X'A^{-1}X)$.

$$\begin{aligned} S &= (B_2B_2' + \Omega_2)^{-1}B_2B_1'G^{-1}B_1B_2'(B_2B_2' + \Omega_2)^{-1} + (B_2B_2' + \Omega_2)^{-1} \\ &\quad + (B_2B_2' + \Omega_2)^{-1}B_2B_1'G^{-1}[H^{-1} - G^{-1}]^{-1}G^{-1}B_1B_2'(B_2B_2' + \Omega_2)^{-1} \\ &\quad (B_2B_2' + \Omega_2)S(B_2B_2' + \Omega_2) \\ &= B_2B_1'G^{-1}B_1B_2' + (B_2B_2' + \Omega_2) + B_2B_1'G^{-1}(H^{-1} - G^{-1})^{-1}G^{-1}B_1B_2' \\ &= B_2B_1'G^{-1}B_1B_2' + (B_2B_2' + \Omega_2) + B_2B_1'[G(H^{-1} - G^{-1})G]^{-1}B_1B_2' \\ &= B_2B_1'G^{-1}B_1B_2' + (B_2B_2' + \Omega_2) + B_2B_1'[GH^{-1}G - G]^{-1}B_1B_2' \\ &= B_2B_1'[G^{-1} - (GH^{-1}G - G)^{-1}]B_1B_2' + (B_2B_2' + \Omega_2) \end{aligned}$$

where $G = B_1M_2B_1' + \Omega_1$ and $H = B_1B_1' + \Omega_1$. To see if RHS > 0, we need to check $G^{-1} - (GH^{-1}G - G)^{-1} > 0$.

Note that $GH^{-1}G - G = G(H^{-1}G - I_K) < 0$ because

$$H^{-1}G = (B_1I_KB_1' + \Omega_1)^{-1}(B_1M_2B_1' + \Omega_1) < I_K \text{ since } I_K - M_2 > 0.$$

Then, $G^{-1} - (GH^{-1}G - G)^{-1} > 0$

$$\Leftrightarrow G^{-1} > (GH^{-1}G - G)^{-1} = (H^{-1}G - I_K)^{-1}G^{-1}$$
$$\Leftrightarrow I_K > (H^{-1}G - I_K)^{-1}$$

which is always true. Therefore, $S > 0$. By the positivity of Schur complement, equation (A.3) holds. Therefore $m > K$.

We now show that given the optimal choice of m cross-sectional units, any additional cross-sectional units yield no predictive power.

Rewrite

$$\text{MSE}(Y) = E(\underset{\sim}{y}_1^0 - Y_1\underset{\sim}{a}_1 - Y_2\underset{\sim}{a}_2)'(\underset{\sim}{y}_1^0 - Y_1\underset{\sim}{a}_1 - Y_2\underset{\sim}{a}) \tag{A.4}$$

Minimizing (A.4) yields

$$\underset{\sim}{a}_1 = (B_1B_1' + \Omega_1)^{-1}(B_1\tilde{\underset{\sim}{b}}_1 - B_1B_2'\underset{\sim}{a}_2) \tag{A.5}$$

and

$$\underset{\sim}{a}_2 = (B_2B_2' + \Omega_2)^{-1}(B_2\tilde{\underset{\sim}{b}}_1 - B_2B_1'\underset{\sim}{a}_1) \tag{A.6}$$

Substituting (A.5) into (A.6) yields

$$\begin{aligned} &[(B_2B_2' + \Omega_2) - B_2B_1'(B_1B_1' + \Omega_1)^{-1}B_1]\underset{\sim}{a}_2 \\ &\quad = B_2(I - B_1'(B_1B_1' + \Omega_1)^{-1}B_1)\tilde{\underset{\sim}{b}}_1 \\ &\quad = B_2(B_1'\Omega_1^{-1}B_1 + I_K)^{-1}\tilde{\underset{\sim}{b}}_1 \end{aligned} \tag{A.7}$$

Therefore $\underset{\sim}{a}_2$ equals to zero iff the right-hand side of (A.7) is equal to zero.

$$\underset{\sim}{a}_1 = (B_1 B_1' + \Omega_1)^{-1}(B_1 \underset{\sim}{\alpha} - B_1 B_2' \tilde{\underset{\sim}{a}}_2) \tag{A.8}$$

Substituting (A.5) into (A.6) yields

$$\begin{aligned}
&[(B_2 B_2' + \Omega_2) - B_2 B_1'(B_1 B_1' + \Omega_1)^{-1} B_1 B_2']\underset{\sim}{a}_2 \\
&\quad = B_2[I_K - B_1'(B_1 B_1' + \Omega_1)^{-1} B_1] \\
&\quad = B_2(B_1' \Omega_1^{-1} B_1 + I_K)^{-1} \tilde{\underset{\sim}{b}}_1
\end{aligned} \tag{A.9}$$

APPENDIX B

In this appendix we present the predictions of Hong Kong's real economic growth rate had there been no change in sovereignty or no CEPA implementation with mainland China using the factor approach. IC1 and IC2 are used to estimate the number of underlying common factors with maximum number of K equal to 20. Both methods give $K = 20$ and therefore the same predicted path. Tables B.1 and B.2 present the estimated treatment effects of political and economic integration based on Bai and Ng (2002) the IC criterion of selecting K. Figures B.1–B.4 plot the within-sample and post-sample predictions under political and economic integration. As one can see from these figures, both the within-sample and post-sample predictions are a lot more volatile than using the observed data. The estimated treatment effects fitting an AR(1) model is 4.63% and is statistically significant with t-statistic equal to 8.9.

DOI: 10.1002/jae

B.1. Treatment effect: political integration

	Actual	Predicted	Treatment
1997:Q3	0.061	0.2526	−0.1916
1997:Q4	0.014	0.453	−0.439
1998:Q1	−0.032	−1.7748	1.7428
1998:Q2	−0.061	−0.8151	0.7541
1998:Q3	−0.081	−0.4678	0.3868
1998:Q4	−0.065	−1.1164	1.0514
1999:Q1	−0.029	1.4224	−1.4514
1999:Q2	0.005	0.1434	−0.1384
1999:Q3	0.039	1.2769	−1.2379
1999:Q4	0.083	1.5864	−1.5034
2000:Q1	0.107	0.224	−0.117
2000:Q2	0.075	1.0059	−0.9309
2000:Q3	0.076	−1.3426	1.4186
2000:Q4	0.063	−3.3762	3.4392
2001:Q1	0.027	−2.5358	2.5628
2001:Q2	0.015	−2.7242	2.7392
2001:Q3	−0.001	−1.2293	1.2283
2001:Q4	−0.017	−0.8478	0.8308
2002:Q1	−0.01	−0.7321	0.7221
2002:Q2	0.005	−0.4256	0.4306
2002:Q3	0.028	−1.5934	1.6214
2002:Q4	0.048	0.5178	−0.4698
2003:Q1	0.041	−0.5959	0.6369
2003:Q2	−0.009	0.357	−0.366
2003:Q3	0.038	0.5072	−0.4692
2003:Q4	0.047	0.0609	−0.0139
Mean	0.018	−0.4527	0.4706
SD	0.0478	1.281	1.2737

IC1 gives $\hat{K} = k_{\max} = 20$.

B.2. Treatment effect: economic Integration

	Actual	Predicted	Treatment
2004:Q1	0.077	0.0584	0.0186
2004:Q2	0.12	0.0755	0.0445
2004:Q3	0.066	0.0545	0.0115
2004:Q4	0.079	0.0432	0.0358
2005:Q1	0.062	0.0132	0.0488
2005:Q2	0.071	0.0072	0.0638
2005:Q3	0.081	0.0248	0.0562
2005:Q4	0.069	0.0272	0.0418
2006:Q1	0.09	0.0476	0.0424
2006:Q2	0.062	0.0401	0.0219
2006:Q3	0.064	0.0258	0.0382
2006:Q4	0.066	0.0174	0.0486
2007:Q1	0.055	0.013	0.042
2007:Q2	0.062	0.0032	0.0588
2007:Q3	0.068	0.0022	0.0658
2007:Q4	0.069	0.0205	0.0485
2008:Q1	0.073	0.017	0.056
Mean	0.0726	0.0289	0.0437
SD	0.0149	0.0211	0.0153

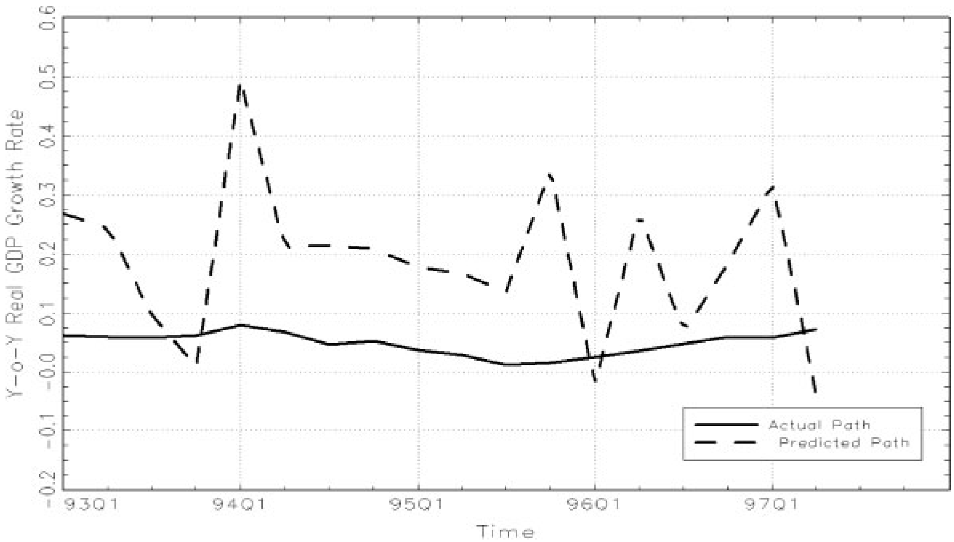

Figure B.1. Predicted counterfactual for political integration from 1993:Q1 to 1997:Q2

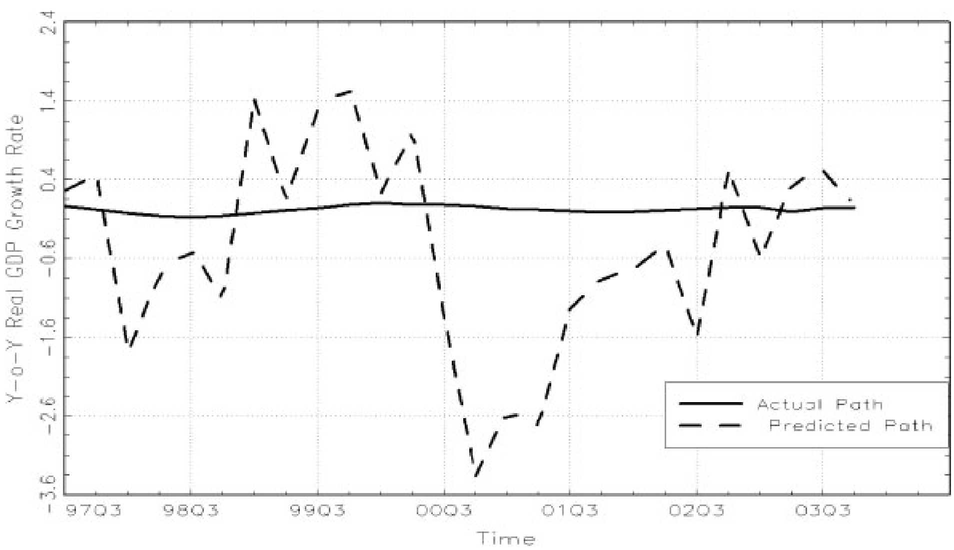

Figure B.2. Predicted counterfactual for political integration from 1997:Q3 to 2003:Q4

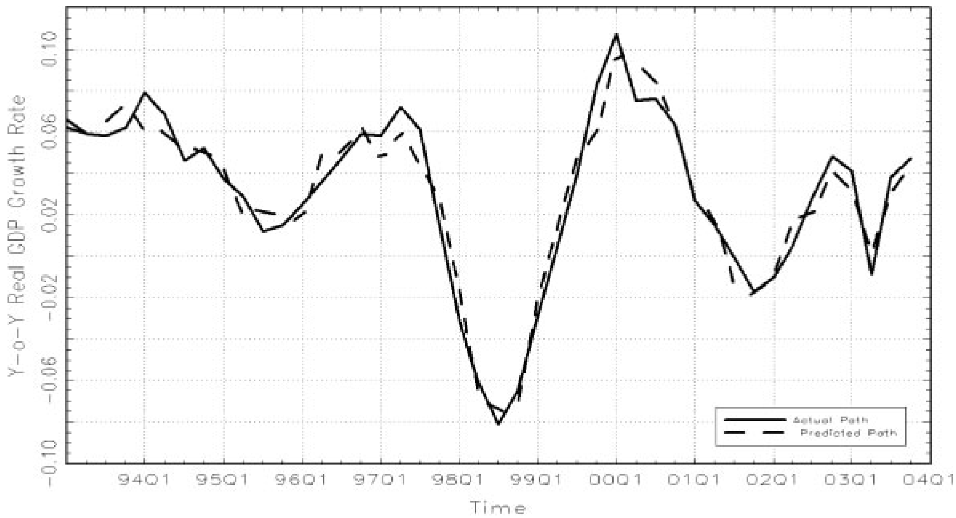

Figure B.3. Counterfactual for economic integration using approximate factor model

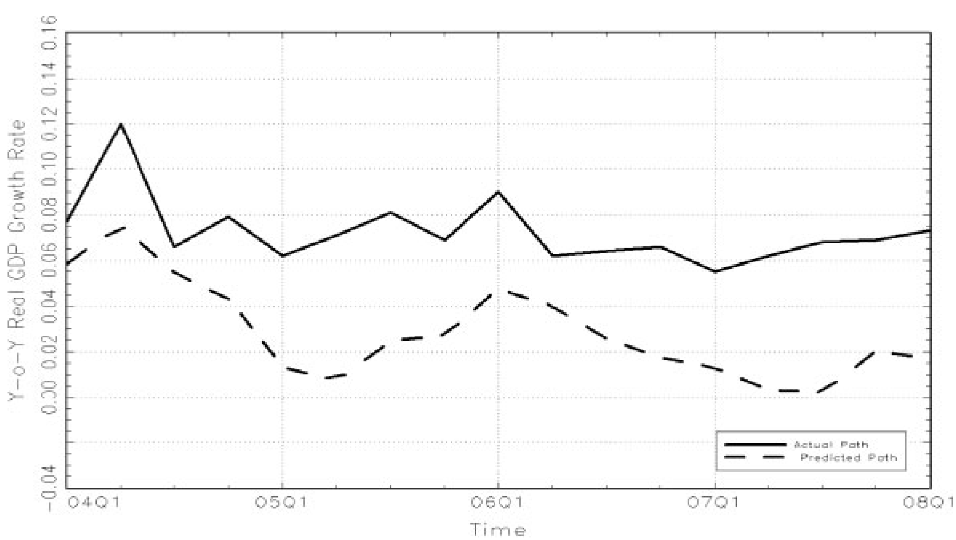

Figure B.4. Counterfactual for economic integration using approximate factor model