Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls

Laurent Gobillon
INED and Paris School of Economics

Thierry Magnac
Toulouse School of Economics

First version: October 2012
This version November 5, 2014

Abstract

In this paper, we investigate the use of interactive effect or linear factor models in regional policy evaluation. We contrast treatment effect estimates obtained by Bai (2009)'s least squares method with the popular difference in differences estimates as well as with estimates obtained using synthetic control approaches as developed by Abadie and coauthors. We show that difference in differences are generically biased and we derive the support conditions that are required for the application of synthetic controls. We construct an extensive set of Monte Carlo experiments to compare the performance of these estimation methods in small samples. As an empirical illustration, we also apply them to the evaluation of the impact on local unemployment of an enterprise zone policy implemented in France in the 1990s.

Keywords: Policy evaluation, Linear factor models, Synthetic controls, Economic geography, Enterprise zones

JEL Classification: C21, C23, H53, J64, R11

1 Introduction[1]

It is becoming more and more common to evaluate the impact of regional policies using the tools of program evaluation derived from micro settings (see Blundell and Costa-Dias, 2009, or Imbens and Wooldridge, 2011 for surveys). In particular, enterprise and empowerment zone programs have received a renewed interest over recent years (see for instance, Busso, Gregory and Kline, 2013, Ham, Swenson, Imrohoroglu and Song, 2012, Gobillon, Magnac and Selod, 2012). Those programs consist in a variety of locally targeted subsidies aiming primarily at boosting local employment or the employment of residents. Their evaluations use panel data and methods akin to difference in differences that offer the simplest form of control of local unobserved characteristics that can be correlated with the treatment indicator. Nonetheless, specific issues arise when studying regional policies and the tools required to evaluate their impact or to perform a cost-benefit analysis are different from the ones used in more usual micro settings.

The issue of spatial dependence between local units is important in the evaluation of regional policies. Outcomes are likely to be spatially correlated in addition to the more usual issue of serial correlation in panel data. There is thus a need for a better control of spatial dependence and more generally of cross-section dependence when evaluating regional policies. This is why more elaborate procedures than difference in differences are worth exploring and the use of factors or interactive effects proved to be attractive and fruitful in micro studies (Carneiro, Hansen and Heckman, 2003). Interactive effect models facilitate the control for cross-section dependence not only because of spatial correlations but also because areas can be close in economic dimensions which depart from purely geographic characteristics. This is the case for instance when two local units are affected by the same sector-specific shocks because of sectoral specialisation even if these units are not neighbors.

Second, a key issue in policy evaluation is that treatment and outcomes might be correlated because of the presence of unobservables. It should also be acknowledged when using regional data that those unobservables differencing local units might be multidimensional because the

[1]We are grateful to two referees and to the coeditor for their suggestions and participants at seminars in Duke University, INED-Paris, Toulouse School of Economics, CREST, ISER at Essex, the 2012 NARSC conference in Ottawa, ESEM 2013 and 8th IZA Conference on Labor Market Policy Evaluation in London for useful comments as well as to Alberto Abadie and Sylvain Chabé-Ferret for fruitful discussions. We also thank DARES for financial support. The usual disclaimer applies.

underlying cycles of economic activities of local units are likely to be multiple. Interactive effect models are aimed precisely at allowing the set of unobserved heterogeneity terms or factor loadings that are controled for to have a moderately large dimension.

Moreover, the estimation of linear factor models in panels is relatively easy and asymptotic properties of estimates are now well known (Pesaran, 2006, Bai, 2009). Yet, there are only a few earlier contributions in the literature that conduct regional policy evaluations using factor models (Kim and Oka, 2014) or using a kindred conditional pseudo-likelihood approach (Hsiao, Ching and Wan, 2012).

The contributions of this paper are threefold. We first provide results concerning the theoretical set-up. We clarify restrictions in linear factor models under which the average treatment on the treated parameter is identified. We analytically derive the generic bias of the difference-in-differences estimator when the true data generating process has interactive effects and the set of factor loadings is richer than the standard single-dimensional additive local effect. Moreover, we derive from extant literature conditions on the number of treatment and control groups as well as on the number of periods under which factor model estimation delivers consistent estimates of the average treatment on the treated parameter.

Contrasting the estimation of linear factor models with the alternative method of synthetic controls is our second contribution. This alternative method was proposed by Abadie and Gardeazabal (2003) and its properties have been developed and vindicated in a model with factors (Abadie, Diamond and Hainmueller, 2010). Under the maintained assumption that the true model is a linear factor model, we show that synthetic controls are equivalent to interactive effect methods whenever matching variables (i.e. factor loadings and exogenous covariates) of all treated areas belong to the support of matching variables of control areas, which is assumed to be convex, a case that we call the *interpolation* case. This is not true any longer in the *extrapolation* case, that is, when matching variables of one treated area at least, do not belong to the support of matching variables in the control group.

Our third contribution is that we evaluate the relevance and analyze the properties of interactive effect, synthetic control and difference-in-differences methods by Monte Carlo experiments. We use various strategies for interactive effect estimation. First, a direct method estimates the counterfactual for treated units by linear factor methods in a restricted sample where post-

treatment observations for treated units are excluded. The second method estimates a linear factor model which includes a treatment dummy and uses the whole sample. Propensity score matching underlies the third method in which the score is conditioned by factor loading estimates obtained using the first method. Imposing common support constraints on factor loadings when estimating the counterfactual for treated units by linear factor methods provides the fourth method. We contrast these Monte Carlo estimation results with the ones we obtain by using synthetic controls and difference in differences.

We finally provide the results of an empirical application of these methods to the evaluation of the impact of a French enterprise zone program on unemployment exits at the municipality level in the Paris region. This extends our results in Gobillon et al. (2012) in which we were using conditional difference-in-differences methods. We show that the estimated impact is robust to the presence of factors and therefore to cross-section dependence. We also look at other empirical issues of interest such as the issue of missing data about destination when exiting unemployment and the more substantial issue of the impact of the policy on entries into unemployment.

In the next Section, we briefly review the meager empirical literature in which factor models are used to evaluate regional policies. We construct in Section 3 the theoretical set-up and write restrictions leading to the identification of the average treatment on the treated in linear factor models. Next, we derive the bias of difference in differences and describe the linear factor model estimation procedures. We derive the conditions that contrast their properties with those of synthetic control methods. Monte Carlo experiments reported in Section 4 are used to evaluate the small sample properties of the whole range of our estimation procedures. The empirical application and estimation results are presented in Section 5 and the last section concludes.

2 Review of the literature

To our knowledge, there are only two earlier empirical contributions by Hsiao, Ching and Wan (2012) and Kim and Oka (2014) applying factor models to the evaluation of regional policies. Interestingly, both papers motivate the use of factor models by contrasting them to the difference-in-differences approach. Hsiao et al. (2012) use an interactive effect model to study the effect on Hong Kong's domestic product of two policies of convergence with mainland China that were implemented at the turn of this century. Their observations consist in various macroeconomic

variables measured every quarter over ten years for Hong Kong and countries either in the region or economically associated with Hong-Kong. The authors argue that interactive models can be rewritten as models in which interactive effects can be replaced by summaries of outcomes for other countries at the same dates using a conditioning argument. Indeed, common factors can be predicted using this information but this entails a loss of information since information at the current period only is used to construct these predictions.

Interestingly, Ahn, Lee and Schmidt (2013) analyze an interactive effect model and their method**,** that consists in differencing out factor loadings**,** provides potential efficiency improvements over the procedure of Hsiao, Ching and Wan (2012). The authors indeed show that the parameters of interest are solutions of moment restrictions that do not depend on individual factor loadings. Assuming out any remaining spatial correlation, they show that their GMM estimates are consistent for fixed T.

Kim and Oka (2014) estimate an interactive effect model following Bai (2009) and provide a policy evaluation of the impact of changes in unilateral divorce state laws on divorce rates in the US. They find that interactive effect estimates are smaller than difference-in-differences estimates. Furthermore, they estimate their model varying the number of factors and find that the model selection procedures proposed by Bai and Ng (2002) are not informative.

Overall, in a large N and T environment, the most prominent estimation methods were proposed by Pesaran (2006) who uses regressions augmented with cross section averages of covariates and outcomes, and by Bai (2009) who uses principal component methods. Westerlund and Urbain (2011) review quite extensively differences between these methods.

3 Theoretical Set-Up

Consider a sample composed of $i = 1, ..., N$ local units observed at dates $t = 1, ..., T$. A simple binary treatment, $D_i \in \{0, 1\}$, is implemented at date $T_D < T$ so that for $t \geqslant T_D > 1$, the units $i = 1, ..., N_1$ are treated ($D_i = 1$). Units $i = N_1 + 1, ..., N$ are never treated ($D_i = 0$). For each unit, we observe outcomes, y_{it}, which might depend on the treatment and our parameter of interest is the average effect of the treatment on the treated. In Rubin's notation, we denote by $y_{it}(d)$ the outcome at date t for an individual i whose treatment status is d (where $d = 1$ in case of treatment, and $d = 0$ in the absence of treatment). This hypothetical status should

be distinguished from random variable D_i describing the actual assignment to treatment in this experiment.

The average effect of the treatment on the treated can be written when $t \geq T_D$:

$$E\left(y_{it}\left(1\right) - y_{it}\left(0\right)|D_i = 1\right) = E\left(y_{it}\left(1\right)|D_i = 1\right) - E\left(y_{it}\left(0\right)|D_i = 1\right) \tag{1}$$

A natural estimator of the first right-hand side term is its empirical counterpart since the outcome in case of treatment is observed for the treated at periods $t \geqslant T_D$. In contrast, the second right-hand side term is a counterfactual term since the outcome in the absence of treatment is not observed for the treated at periods $t \geqslant T_D$. The principle of evaluation methods relies on using additional restrictions to construct a consistent empirical counterpart to the second right-hand side term (e.g. Heckman and Vytlacil, 2007). For instance, it is well known that difference-in-differences methods are justified by an equal trend assumption:

$$E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1) = E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0) \text{ for } t \geq T_D. \tag{2}$$

under which the counterfactual can be written as:

$$E\left(y_{it}\left(0\right)|D_i = 1\right) = E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0) + E(y_{i,T_D-1}(0) \mid D_i = 1) \text{ for } t \geq T_D,$$

in which all terms on the right-hand side are directly estimable from the data.

The object of this section is to generalize the usual set-up in which difference in differences provide a consistent estimate of the effect of the treatment on the treated (TT) to a set-up allowing for higher-dimensional unobserved heterogeneity terms. Local units treated by regional policies could indeed be affected by various common shocks describing business cycles related for instance to different economic sectors. Associated factor loadings would describe the heterogeneity in the exposure of local units to these common shocks. A single dimensional additive local effect as in the set up underlying difference-in-differences estimation is unlikely to describe this rich economic environment. Furthermore, we know that difference in differences can dramatically fail when heterogeneity is richer than what is modelled (Heckman, Ichimura and Todd, 1997).

In this paper, we restrict our attention to linear models because the number of units is rather small although extensions to non-linear settings could follow the line of Abadie and Imbens (2011) at the price of losing the simplicity of linear factor models. The route taken by Conley and Taber (2011) to deal with small sample issues might also be worth extending to our setting. More

specifically, linearity makes one wary of issues of interpolation and extrapolation that we shall highlight in the general framework of linear factor models as well as in the approach of synthetic controls proposed in the seminal paper by Abadie and Gardeazabal (2003).

We present in the first subsection the specification of a linear factor data generating process which is maintained thorughout the paper and we discuss identifying assumptions. We show that the conventional difference-in-differences estimate is generically biased. Next, for a linear factor model that includes a treatment indicator, we derive a rank condition for the identification of the average treatment on the treated. We also propose a direct method whereby we construct the counterfactual term in equation (1) using the samples of control and treated units albeit the latter before treatment only (see Heckman and Robb, 1985 or Athey and Imbens, 2006). Finally, we describe the approach of synthetic controls and analyze its properties when the true model has interactive effects.

3.1 Interactive linear effects and restrictions on conditional means

In the conventional case of difference in differences (DID) (see for instance Blundell and Costa-Dias, 2009), the outcome in the absence of treatment is specified as a linear function:

$$y_{it}(0) = x_{it}\beta + \tilde{\lambda}_i + \widetilde{\delta}_t + \varepsilon_{it} \tag{3}$$

in which x_{it} is a $1 \times K$ vector of individual covariates, and $\tilde{\lambda}_i$ and $\widetilde{\delta}_t$ are individual and time effects. A limit to this specification is that individuals are all affected in the same way by the time effects. To allow for interactions and make the specification richer, we specify the outcome in the absence of treatment as a function of the interaction between factors varying over time and heterogenous individual terms called factor loadings as:

$$y_{it}(0) = x_{it}\beta + f_t'\lambda_i + \varepsilon_{it} \tag{4}$$

in which β are the effects of covariates, λ_i is a $L \times 1$ vector of individual effects or *factor loadings*, and f_t is a $L \times 1$ vector of time effects or *factors*. Note that this specification embeds the usual additive model which is obtained when $\lambda_i = \left(\tilde{\lambda}_i, 1\right)'$ and $f_t = \left(1, \widetilde{\delta}_t\right)'$ as, in that case, $f_t'\lambda_i = \tilde{\lambda}_i + \widetilde{\delta}_t$.

The true process generating the data is supposed to be given by equation (4) and is completed

by the description of the outcome in case of treatment:

$$y_{it}(1) = y_{it}(0) + \alpha_{it} \tag{5}$$

which, in contrast to the linear specification above, is not restrictive.

There are a few usual assumptions that complete the description of the true data generating process (DGP) maintained throughout the paper. First, we shall assume that we know the number of factors in the true DGP described by equation (4). It might be useful to implement tests regarding the number of factors (Bai and Ng, 2002, Moon and Weidner, 2013b) but these tests are fragile (Onatski, Moreira and Hallin, 2013). Moreover, we adopt the assumption that factors are sufficiently strong so that the consistency condition for the number of factors and consequently for factors and factor loadings is satisfied (for alternative views see Onatski, 2012 or Pesaran and Tosetti, 2011). This condition reflects the fact that factor loadings can be separated from the idiosyncratic random terms at the limit.[2]

Moreover, we do not specify the dynamics of factors in the spirit of Doz, Giannone and Reichlin (2011). Their specification imposes more restrictions on the estimation and inference is more difficult to develop. This is why we stick to the limited information framework which does not impose conditions on the dynamics of factors although it could be done in the way explained by Hsiao, Ching and Wan (2012). Furthermore, the only available explanatory variables are not varying over time in our empirical application. This corresponds to the *low rank regressor* assumption as defined by Moon and Weidner (2013a) and under which identifying assumptions are of a particular form. At this stage however we prefer to stick to the more general format.

A final comment is worth making. In treatment evaluation, lagged endogenous variables are at times included as matching covariates in order to control for possible ex-ante differences. In spirit, this is very close to a model with interactive effets because it is well known that a simple linear dynamic panel data model like:

$$y_{it} = \alpha y_{it-1} + \eta_i + u_{it}$$

can be rewritten as a static model:

$$y_{it} = \alpha^t y_{i0} + \left(1 - \alpha^t\right) \frac{\eta_i}{1-\alpha} + \nu_{it}$$

[2]It does not mean that the treatment parameter is not identified under alternative assumptions.

in which ν_{it} is an AR(1) process. Factors are α^t and $1-\alpha^t$, and factor loadings are y_{i0} and $\frac{\eta_i}{1-\alpha}$. This argument could be generalized to more sophisticated dynamic linear models.

3.1.1 Restrictions on conditional means

To complete the description of the true data generating process, we now present and comment the main restrictions on random terms. To keep notation simple and conform with the usual panel data set up, we generally consider that factors f_t are fixed while factor loadings λ_i are supposed to be correlated random effects.

We first assume that idiosyncratic terms ε_{it} are "orthogonal" to factor loadings and that explanatory variables are strictly exogenous:[3]

$$\varepsilon_{it} \perp (\lambda_i, x_i)$$

in which $x_i' = (x_{i1}', ...x_{iT}')'$ is a $[T, K]$ matrix. This would be without loss of generality when orthogonality is defined as the absence of correlation as in Bai (2009). Because of the next assumption we will adopt, we prefer to interpret orthogonality as mean independence and the formal translation of the informal statement above is therefore that:

Assumption A1: $$E(\varepsilon_{it} \mid \lambda_i, x_i) = 0.$$

Second, we extend the usual assumption made in difference-in-differences estimation by assuming that the conditioning set now includes unobserved factor loadings:

$$y_{it}(0) \perp D_i \mid (x_i, \lambda_i) \Leftrightarrow \varepsilon_{it} \perp D_i \mid (x_i, \lambda_i)$$

and we write this condition as a mean independence restriction:

Assumption A2: $$E(\varepsilon_{it} \mid D_i, \lambda_i, x_i) = E(\varepsilon_{it} \mid \lambda_i, x_i).$$

Note that we do not suppose that (λ_i, x_i) and D_i are uncorrelated and selection into treatment can freely depend on observed and unobserved heterogeneity terms.

Finally, define the average treatment effect over the periods after treatment as:

$$\alpha_i = \frac{1}{T - T_D + 1} \sum_{t=T_D}^{T} \alpha_{it}$$

[3]The extension to the case with weakly exogeneous regressors would follow Moon and Weidner (2013a) for instance.

so that our main parameter of interest is the average treatment on the treated over the periods after treatment defined as:[4]

Definition ATT:

$$\alpha = E(\alpha_i \mid D_i = 1) = \frac{1}{T - T_D + 1} \sum_{t=T_D}^{T} E(\alpha_{it} \mid D_i = 1).$$

Assumptions A1 and A2 are the main restrictions in our set-up and Definition ATT defines our parameter of interest.

3.2 The generic bias of difference-in-differences estimates

If the true data generating process comprises interactive effects, we now show that the difference-in-differences estimator is generically biased although we exhibit two interesting specific cases in which the bias is equal to zero. For simplicity, we omit covariates or, since covariates are assumed to be strictly exogenous, implicitly condition on them in this subsection. We also assume for simplicity that the probability measure of factor loadings in the treated population, $dG(\lambda_i \mid D_i = 1)$, and in the control population, $dG(\lambda_i \mid D_i = 0)$, are dominated by the Lebesgue measure so that both distributions are absolutely continuous.

We shall show that the condition which is implied by Assumption A2:[5]

$$E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1, \lambda_i) = E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0, \lambda_i) \text{ for } t \geqslant T_D \tag{6}$$

does not imply equation (2) under which the difference-in-differences estimator is consistent.

Indeed:

$$\begin{aligned} E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1) &= E\left[E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1, \lambda_i) \mid D_i = 1\right], \\ &= \int E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1, \lambda_i) dG(\lambda_i \mid D_i = 1). \end{aligned}$$

Replacing the integrand using equation (6) yields:

$$E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1) = \int E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0, \lambda_i) dG(\lambda_i \mid D_i = 1). \tag{7}$$

[4] In the case $T \to \infty$, those definitions should be interpreted as limits. Note also that it is generally easy to design estimates for time-specific treatment parameters such as $E(\alpha_{it} \mid D_i = 1)$ by restricting the post-treatment observations to period t only.

[5] This condition is slightly weaker than A2 because it considers differences between periods.

Two special cases are worth noting. Firstly, the integrand in the previous expression does not depend on λ_i in the restricted case in which there is a single factor $f_t = 1$ and a single individual effect associated with this factor. In this case, equation (7) can be written as:

$$\begin{aligned} E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1) &= E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0) \int dG(\lambda_i \mid D_i = 1) \\ &= E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0), \end{aligned}$$

which yields equation (2) describing equality of trends.

Alternatively, (perfectly) controled experiments also enables identification through difference in differences in spite of using the alternative argument that $dG(\lambda_i \mid D_i = 1) = dG(\lambda_i \mid D_i = 0)$. The same equation (2) holds and the treatment parameter is consistently estimable by difference in differences.

This implication is not true in general and we can distinguish two cases. If the conditional distribution of λ_i in the treated population is dominated by the corresponding measure in the control population i.e.:

$$\forall \lambda_i \text{ such that } dG(\lambda_i \mid D_i = 0) = 0 \text{ we have } dG(\lambda_i \mid D_i = 1) = 0, \tag{8}$$

the support of treated units is included in the support of non treated units. We shall describe from now on cases in which support condition (8) holds as an instance of *interpolation* and if such a condition is not satisfied, as an instance of *extrapolation*.

In the *interpolation* case, let:

$$r(\lambda_i) = \frac{dG(\lambda_i \mid D_i = 1)}{dG(\lambda_i \mid D_i = 0)} < \infty$$

which is well defined because of the support condition (8) and because distributions are absolutely continuous. Write equation (7) as:

$$E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1) = \int E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0, \lambda_i) r(\lambda_i) dG(\lambda_i \mid D_i = 0) \tag{9}$$

which in turn implies that:

$$\begin{aligned} E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 1) &= E(y_{it}(0) - y_{i,T_D-1}(0) \mid D_i = 0) \\ &\quad + Cov\left(y_{it}(0) - y_{i,T_D-1}(0), r(\lambda_i) \mid D_i = 0\right) \end{aligned}$$

The second term in the right hand side can be interpreted as the differential trend in outcomes which is due to the time varying effects of factors interacted with unobserved factor loadings. If

there is indeed a factor loading associated to a time-varying factor, the second term is not equal to zero except under special circumstances as seen above. In the interpolation case, the second term describes the bias in DID estimates.

In the alternative case of *extrapolation*, the bias term is derived in a similar way although its interpretation is less clear since it mixes issues of non inclusive supports with the time varying effect of factors.

3.3 Interactive Effect Estimation in the Whole Sample

We now explore interactive effect methods and exhibit conditions under which these methods allow the identification of the average treatment on the treated parameter. The observed outcome verifies:

$$y_{it} = y_{it}(0)(1 - I_t D_i) + y_{it}(1) I_t D_i,$$

in which D_i is the treatment indicator, and $I_t = 1\{t \geq T_D\}$ is a time indicator of treatment. Using equations (4) and (5) yields:

$$y_{it} = \alpha_{it} I_t D_i + x_{it}\beta + f_t'\lambda_i + \varepsilon_{it} \tag{10}$$

We maintain Assumptions A1 and A2 that allow the correlation between D_i and λ_i to be unrestricted so that selection into treatment can depend on factor loadings. Similarly, the correlation between I_t and f_t is unrestricted so that the implementation of the treatment can be correlated with economic cycles which are described here by factors.

We shall rewrite equation (10) as:

$$y_{it} = \alpha I_t D_i + x_{it}\beta + f_t'\lambda_i + \varepsilon_{it} + (\alpha_{it} - \alpha) I_t D_i \tag{11}$$

in which α is the average treatment on the treated parameter as in Definition ATT. If the number of periods after treatment is greater than 1 however, this model would not deliver unbiased estimates because of omitted variables. Indeed, we could rewrite model (10) as:

$$y_{it} = \alpha_t I_t D_i + x_{it}\beta + f_t'\lambda_i + \varepsilon_{it} + (\alpha_{it} - \alpha_t) I_t D_i, \tag{12}$$

allowing for a time varying treatment effect:

$$\alpha_t = E(\alpha_{it} \mid D_i = 1).$$

The omitted variables in equation (11) would be the $T - T_D$ period indicators interacted with the treatment indicator (except one). For the sake of simplicity, we develop our analysis in this section in the simple case in which we have:

Assumption A3: $\forall t \geq T_D, \qquad \alpha_t = \alpha$

so that equation (11) is correctly specified.[6]

We now exhibit further conditions under which α can be estimated using interactive effect procedures as proposed by Bai (2009). We start with the case $\beta = 0$ which requires a weak rank condition and then extend it to the general case with covariates which requires an additional assumption that is stronger albeit easy to interpret.

3.3.1 Average Treatment Effect on the Treated in the Absence of Covariates

We shall prove that the parameter of interest α is identified under the two conditions that I_t is not equal to a linear combination of factors f_t and that the probability of treatment is positive.

We keep considering that T is fixed as well as factors f_t and treatment I_t and we analyze identification as if factors f_t were known. This argument extends to the case in which T tends to infinity by taking limits.

Stack individual observations in individual vectors of dimension $[T, 1]$:

$$y_i = \alpha D_i I_{[1:T]} + F'\lambda_i + \varepsilon_i + \Delta_i I_{[1:T]} D_i \tag{13}$$

in which $y_i = (y_{i1}, ., y_{iT})'$, $I_{[1:T]} = (I_1, ., I_T)', F = (f_1, ., f_T)$, $\varepsilon_i = (\varepsilon_{i1}, ., \varepsilon_{iT})'$ and Δ_i is a diagonal matrix of dimension $[T, T]$ whose diagonal terms are $(\alpha_{i1} - \alpha, ..., \alpha_{iT} - \alpha)$. Set $M_F = I - F'(FF')^{-1}F$ and multiply the previous equation to obtain:

$$M_F y_i = \alpha D_i M_F I_{[1:T]} + M_F \varepsilon_i + M_F \Delta_i I_{[1:T]} D_i. \tag{14}$$

A necessary condition for identifying α using equation (14) stacked over the different individual units is therefore:

$$I'_{[1:T]} M_F I_{[1:T]} > 0 \text{ and } E(D_i) > 0. \tag{15}$$

This means that $I_{[1:T]}$ is not equal to a linear combination of factors and that the probability of being treated is positive. This is related to the rank condition underlying the identification of

[6] The identification of equation (12) can be established using very similar developments. The proof is available upon request.

parameters in Proposition 3 in Bai (2009, p.1259). Furthermore, this condition is also necessary in equation (13) because the correlation between λ_i and D_i is unrestricted.

This condition is also sufficient. This is because $E(D_i)I'_{[1:T]}M_F I_{[1:T]}$ is invertible using condition **(15)** and because we can then show that:

$$\alpha = (E(D_i I'_{[1:T]} M_F I_{[1:T]}))^{-1} E(D_i I'_{[1:T]} M_F y_i) = (E(D_i) I'_{[1:T]} M_F I_{[1:T]})^{-1} E(D_i I'_{[1:T]} M_F y_i). \quad (16)$$

Indeed, the covariance between the two right-hand side terms of equation (14), the regressor $D_i M_F I_{[1:T]}$ and the error term $M_F \varepsilon_i + M_F(\alpha_i - \alpha) I_{[1:T]} D_i$, is equal to zero. There are two terms in this correlation that we analyze in turn.

The first term is equal to 0 by construction (Assumption A2) because:

$$E(I'_{[1:T]} M_F D_i M_F \varepsilon_i) = E(I'_{[1:T]} M_F D_i M_F E(\varepsilon_i \mid D_i)) = 0 \quad (17)$$

since D_i is a scalar random variable and variables in the time dimension are supposed to be fixed.

The second term of the correlation above is more interesting and involves:

$$E(I'_{[1:T]} M_F D_i M_F \Delta_i I_{[1:T]} D_i) = E(I'_{[1:T]} M_F D_i M_F E(\Delta_i \mid D_i) I_{[1:T]} D_i), \quad (18)$$

which is equal to zero by construction of Δ_i since $E(\Delta_i \mid D_i = 1)$ is a diagonal matrix whose diagonal terms are:

$$E(\alpha_{it} - \alpha \mid D_i = 1) = \alpha_t - \alpha = 0,$$

by Assumption A3. The correlation in equation (18) is then equal to zero.

Finally, multiplying (14) by $I'_{[1:T]} M_F D_i$ and taking the expectation gives (16). This ends the proof that the average treatment on the treated parameter α is identified under rank condition (15).

3.3.2 The Case with Covariates

In the general case with covariates, we can write equation (11) as:

$$y_i = \alpha D_i I_{[1:T]} + x_i \beta + F' \lambda_i + \varepsilon_i + \Delta_i I_{[1:T]} D_i$$

Multiplying this equation by M_F, we obtain:

$$M_F y_i = \alpha D_i M_F I_{[1:T]} + M_F x_i \beta + M_F \varepsilon_i + M_F \Delta_i I_{[1:T]} D_i. \quad (19)$$

Denote the linear prediction of D_i as a function of x_i as:

$$D_i = vec(x_i)'\gamma + D_{ix},$$

and rewrite equation (19) as:

$$M_F y_i = \alpha D_{ix} M_F I_{[1:T]} + M_F \tilde{\varepsilon}_i + M_F \Delta_i I_{[1:T]} D_i, \tag{20}$$

in which $\tilde{\varepsilon}_i = \varepsilon_i + x_i\beta + \alpha.vec(x_i)'\gamma I_{[1:T]}$. Because x_i and $vec(x_i)$ are uncorrelated with D_{ix}, the same non correlation condition as in equation (17) is valid since we have from Assumptions A1 and A2 that $E\left(\varepsilon_i \,|D_i, x_i\right) = 0$. Thus, the second condition derived from equation (18) that remains to be checked refers to the equality to zero of:

$$E(\Delta_i I_{[1:T]} D_i D_{ix}) = E(\Delta_i I_{[1:T]} D_i D_i) - E((\Delta_i I_{[1:T]} D_i vec(x_i)'\gamma) = -E(\Delta_i I_{[1:T]} D_i vec(x_i)'\gamma),$$

because of the argument employed after equation (18) that uses Definition ATT. This term is equal to zero under the sufficient condition given by:

$$\forall t \geq T_D, \qquad E(\alpha_{it} \mid D_i = 1, x_i) = E(\alpha_{it} \mid D_i = 1),$$

since it implies that:

$$E(\Delta_i \mid D_i = 1, x_i) = E(\Delta_i \mid D_i = 1) = 0,$$

by Assumption A3 and Definition ATT as above. This condition is stronger than necessary as it would be sufficient to condition on the scalar variable $vec(x_i)\gamma$.[7] Note also that the linear interactive model could be generalized by conditioning on covariates in an unrestricted way or interacting covariates with the treatment indicator and this would substantially weaken this condition as in the static evaluation case (Heckman and Vytlacil, 2007).

Consistency and other asymptotic properties of this method can be derived from Bai (2003) when $N \longrightarrow \infty$ and $T \to \infty$. Note also that condition (15) also implies that N_1 tends to ∞ when $N \longrightarrow \infty$. Estimation could also proceed with the estimation method proposed by Ahn et al. (2013) and thence dispense with the assumption that $T \to \infty$. Note that when T is small, Bai's estimator is inconsistent unless errors are white noise (Ahn, Lee and Schmidt, 2001).

[7] In this case, developments following Wooldridge (2005) might be appropriate but we do not follow up this route in this paper.

3.3.3 Remarks[8]

First, when we let the number of periods grow, it is interesting to consider again the difference-in-differences estimator that might be consistent when $T \to \infty$ even if the sufficient conditions of Section 3.2 are not fulfilled. In the absence of covariates, the difference-in-differences estimator is the OLS estimator of the demeaned equation:

$$y_{it} - y_{.t} - y_{i.} + y_{..} = \alpha(D_i - D_.)(I_t - I_.) + (f_t - f_.)'(\lambda_i - \lambda_.) + \tilde{\varepsilon}_{it}$$

in which the notation ".", which replaces an index, points at the average of the variable running over this index, say for instance $y_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\tilde{\varepsilon}_{it}$ is the demeaned version of the errors. When $N \to \infty$, the bias in the OLS estimator of this equation converges to a term which is proportional to:

$$\begin{aligned} & plim_{N\to\infty} \frac{1}{NT} \sum_{i,t} (D_i - D_.)(I_t - I_.)(f_t - f_.)'(\lambda_i - \lambda_.) \\ = \quad & \frac{1}{T} \sum_{t} (I_t - I_.)(f_t - f_.)' \; plim_{N\to\infty} \frac{1}{N} \sum_{i,t} (D_i - D_.)(\lambda_i - \lambda_.). \end{aligned} \tag{21}$$

As assumed above, we generically have $plim_{N\to\infty} \frac{1}{N} \sum_{i,t} (D_i - D_.)(\lambda_i - \lambda_.) \neq 0$ because the correlation between D_i and λ_i is different from zero. Even in this case, the DID estimate can nonetheless be consistent when $T \to \infty$ if:

$$plim_{T\to\infty} \frac{1}{T} \sum_{t} (I_t - I_.)(f_t - f_.)' = 0.$$

This condition states that, in the long run, treatment and factors are uncorrelated and this is not an assumption that one would like to make in all policy evaluations.

Second, it is interesting to develop the reverse of the underspecified case developed in Section 3.2. Overspecification arises when a factor model is estimated while the true data generating process is that of a standard panel with additive individual and time effects. We speculate that results of Moon and Weidner (2013b) might be used to show that not only there is no bias but also that there is no loss of precision **when** using a greater number of factors than necessary, at least asymptotically.

[8] We address here additional points made by referees who we thank for their suggestions.

3.4 Direct Estimation of the Counterfactual

Assumptions A1 and A2 imply that a direct estimation strategy for the effects of treatment on the treated can also be adopted. Estimate first the interactive effect model (4) using the sample composed of non treated observations over the whole period *and* of treated observations before the date of the treatment $t < T_D$. Orthogonality assumption A2 makes sure that excluding observations (i,t) with $i \in \{1,.,N_1\}$ and $t \geq T_D$ does not generate selection. Second, orthogonality assumption A1 renders conditions stated by Bai (2009) valid and the derived asymptotic properties of linear factor estimates hold.

Various asymptotics can be considered:

- If N and T tend to ∞, then β, f_t and λ_i for the non treated are consistently estimated (Bai, 2009).
- If additionally the number of periods before treatment T_D tends to ∞, then λ_i for the treated units are consistently estimated.

As for the counterfactual term to be estimated in equation (1), we have for $t \geqslant T_D$:

$$E\left(y_{it}\left(0\right)|D_i=1\right)=E\left(x_{it}\beta+\lambda_i'f_t\,|D_i=1\right) \tag{22}$$

To estimate this quantity, we replace parameters λ_i, $i=1,...,N_1$, β and f_t when $t \geqslant T_D$ by their consistently estimated values in the right-hand side expression (computed as detailed in the online Appendix), and take the empirical counterpart of the expectation. Namely, the treatment on the treated at a given period is derived by using equation (1) and can be written as:

$$E\left(y_{it}\left(1\right)-y_{it}\left(0\right)|D_i=1\right)=E(\alpha_{it}\mid D_i=1)=E\left(y_{it}\left(1\right)|D_i=1\right)-E\left(x_{it}\beta+\lambda_i'f_t\,|D_i=1\right) \tag{23}$$

and its estimate is obtained by replacing unknown quantities by their empirical counterparts. The average treatment on the treated effect is then obtained by exploiting Definition ATT and averaging equation (23) over the periods after treatment.[9]

An additional word of caution about identification is in order since the rank condition (15) developed in the previous section is also necessary although it is not as simple to derive. This is summarized in the next proposition:

[9] The variance of the estimator can be computed using formulas in Bai (2003) and Bai (2009).

Proposition 1 *Suppose that rank condition (15) does not apply and that the treatment vector $I_{1:T}$ is a linear function of factors:*

$$I_{1:T} = F'\delta$$

in which δ is a $[L, 1]$ vector and F is the matrix of factors as defined above. Then for any value of the treatment effect α, there exists an observationally equivalent factor model in which the value of the treatment effect is equal to zero.

Proof. Let α be any value and write equation (13) as

$$y_i = \alpha I_{1:T} D_i + F'\lambda_i + \tilde{\varepsilon}_i$$

in which $\tilde{\varepsilon}_i$ includes any idiosyncratic variation of the treatment effect across individuals and periods. By replacing $I_{1:T} = F'\delta$, we get:

$$\begin{aligned} y_i &= \alpha F'\delta D_i + F'\lambda_i + \tilde{\varepsilon}_i, \\ &= F'(\alpha\delta D_i + \lambda_i) + \tilde{\varepsilon}_i, \end{aligned}$$

which provides the alternative factor representation in which the value of the treatment effect is equal to zero. ■

This shows the necessity of condition (15) for the estimation method derived in this section as well as for any other estimation method analyzed below.

3.5 A single-dimensional factor model

It is well known since Rubin and Rosenbaum (1983) that conditions A1 and A2 imply the condition:

$$E(\varepsilon_{it} \mid D_i = 1, p(x_i, \lambda_i)) = 0$$

in which the distinction between observed variables x_i and unobserved variables λ_i does not matter. Let $\mu_i = p(x_i, \lambda_i)$ denote the propensity score.

The condition above suggests the following strategy:

1. Estimate factors and factor loadings using the sample of controls and the subsample of treated observations before treatment as detailed in Subsection 3.4.

2. Regress D_i on x_i and $\hat{\lambda}_i$ and construct the predictor of the score $\hat{\mu}_i$.

3. Match on the propensity score à la Heckman, Ichimura and Todd (1998) or, under some conditions, use a single factor model associated to $\hat{\mu}_i$.

3.6 Synthetic controls

The technique of synthetic controls proposed by Abadie and Gardeazabal (2003) and further explored by Abadie, Diamond and Hainmueller (2010, ADH thereafter) proceeds differently. It focuses on the case in which the treatment group is composed of a single unit and uses a specific matching procedure of this treated unit to the control units whereby a so-called synthetic control is constructed. We shall proceed in the same way although as we have potentially more treated units, we shall repeat the procedure for each of them and then aggregate the result over various synthetic controls to yield the average treatment on the treated.[10]

3.6.1 Presentation

We follow the presentation by ADH (2010). An estimator of $y_{it}(0)$ for a single treated unit $i \in \{1, ., N_1\}$ after treatment $t \geq T_D$ is the outcome of a synthetic control "similar" to the treated unit that is constructed as a weighted average of non-treated units. We impose similarity of characteristics x_{it} between treated units and synthetic controls, by weighting characteristics x_{jt} of control units, $j \in \{N_1 + 1, ., N\}$ in such a way that

$$\sum_{j=N_1+1}^{N} \omega_j^{(i)} x_{jt} = x_{it} \text{ for } t = 1, ., T \tag{24}$$

where $\omega_j^{(i)}$ is the weight of unit j in the synthetic control (such that $\omega_j^{(i)} \geqslant 0$ and $\sum_{j=N_1+1}^{N} \omega_j^{(i)} = 1$).

Similarity between pretreatment outcomes is also imposed in ADH (2010):

$$\sum_{j=N_1+1}^{N} \omega_j^{(i)} y_j^{(k)} = y_i^{(k)} \tag{25}$$

[10] An alternative would be to aggregate the treated units into a single unit first. By analogy with what is done in non-parametric matching, this procedure seems more restrictive because using a single synthetic control leads to less precise estimates than when constructing various synthetic controls. Nonetheless, support conditions for the validity of the synthetic control method that we find might justify such an approach because support requirements are weaker in the "aggregate" case.

where $y_j^{(k)} = \sum_{t=1}^{T_D-1} k_t y_{jt}$ is a weighted average of pretreatment outcomes in which $k = (k_1, ., k_{T_D-1})$ are weights differing across periods ($y_i^{(k)}$ for the treated unit is defined similarly). A set of such pretreatment outcome summaries can be generated using various vectors of weights, k. Nevertheless, the most general setting is when we consider all pretreatment outcomes, y_{jt}, for $t = 1, ..., T_D - 1$. Indeed, taking linear combinations of pretreatment outcomes or considering the original ones is equivalent in this general formulation and we dispense with the construction of $y_j^{(k)}$ and $y_i^{(k)}$.

The average treatment on the treated for unit i is estimated as:

$$\hat{\alpha}_i = \frac{1}{T - T_D + 1} \sum_{t \geq T_D} \left[y_{it} - \sum_{j=N_1+1}^{N} \omega_j^{(i)} y_{jt} \right]. \tag{26}$$

In practice, one needs to determine the weights that allow the construction of the synthetic control. Weights should ensure that the synthetic control is as close as possible to the treated unit i and thus that conditions (24) and (25) are verified. Denote $z_j = (y_{j1}, ., y_{j,T_D-1}, x_{j1}, ., x_{jT})'$ (resp. z_i) the list of variables over which the synthetic control is constructed (i.e. pretreatment outcomes and exogenous variables). Weights are computed using the following minimization program:

$$\underset{\omega_j^{(i)} \left| \omega_j^{(i)} \geqslant 0, \sum_{j=N_1+1}^{N} \omega_j^{(i)} = 1 \right.}{Min} \left(\sum_{j=N_1+1}^{N} \omega_j^{(i)} z_j - z_i \right)' M \left(\sum_{j=N_1+1}^{N} \omega_j^{(i)} z_j - z_i \right) \tag{27}$$

in which M is a weighting matrix.[11] Note that the resulting weight $\omega^{(i)}$ is a function of the data $(z_i, z_{N_1+1}, ., z_N)$.

3.6.2 Synthetic controls and interactive effects

We now describe this procedure in an interactive effect model setting as first suggested by ADH (2010). Nonetheless, we show that the absence of bias implies constraints on the supports of factor loadings and exogenous variables, and is related to the developments in Section 3.2 above.

To proceed, we need to introduce additional notation. Our linear factor model can be written

[11] M can be chosen in various ways (see Abadie et al, 2010, for some guidance). In our case we set M to the identity matrix. There could also exist multiple solutions to this program if the treated observation belongs to the convex hull of the controls. Abadie, Diamond and Hainmueller (2014) suggest to use a refinement by selecting the convex combination of the specific points that are the closest to the treated observation (see their footnote 12).

at each time period as:

$$\begin{aligned} Y_t(0) &= \beta' X_t' + f_t' \Lambda_U + \varepsilon_t \quad \text{for the untreated,} \\ y_{it}(0) &= \beta' x_{it}' + f_t' \lambda_i + \varepsilon_{it} \quad \text{for each treated individual} \end{aligned} \tag{28}$$

where $\Lambda_U = (\lambda_{N_1+1}, ..., \lambda_N)$ is $(L, N - N_1)$ and f_t is a L column vector. Similarly, $Y_t(0)$ and ε_t are $(N - N_1)$ row vectors and X_t is a $(N - N_1, K)$ matrix.

Weights $\omega^{(i)} = \left(\omega_{N_1+1}^{(i)}, ..., \omega_N^{(i)}\right)$ are obtained by equation (27) and we have:

$$\begin{cases} y_{it}(0) = Y_t(0)\,\omega^{(i)} + \eta_{it} \text{ for } t < T_D, \\ x_{it}' = X_t' \omega^{(i)} + \eta_{itX} \text{ for } t = 1, ..., T \end{cases} \tag{29}$$

Note that the construction of the synthetic control by equation (29) is allowed to be imperfectly achieved and the discrepancy is captured by the terms η_{it} and η_{itX}. We thus acknowledge that characteristics of the treated unit, $z_i = (y_{i1}, ., y_{i,T_D-1}, x_{i1}, ., x_{iT})'$, might not belong to the convex hull, C_U, of the characteristics of control units. First, there are small sample issues when the number of pre-treatment periods, $T_D - 1$, and of covariates, KT, is larger than the number of untreated units, $N - N_1$. In other words, the convex hull C_U lies in a space whose dimension is lower than the number of vector components, $T_D - 1 + KT$. Second and more importantly, even if $T_D - 1 + KT < N - N_1$, vector z_i might not belong to this convex hull because supports of characteristics for treated and control units differ. Terms η_{it} and η_{itX} capture this discrepancy.

We now analyze what consequences this construction has on the estimation of the treatment effect. The estimated treatment effect given by equation (26) is a function of

$$\begin{aligned} y_{it} - \sum_{j=N_1+1}^{N} \omega_j^{(i)} y_{jt} &= y_{it}(1) - Y_t(0)\,\omega^{(i)} = \alpha_{it} + y_{it}(0) - Y_t(0)\,\omega^{(i)} \\ &= \alpha_{it} + \eta_{it}, \end{aligned}$$

in which we have extended definition (29) to all $t \geq T_D$. The absence of bias for the LHS estimate with respect to $E(\alpha_{it})$ can thus be written as $E(\eta_{it}) = 0$. To write this condition as a function of primitives, we need to replace dependent variables by their values in the model described by (28). This gives:

$$\begin{aligned} \eta_{it} &= y_{it}(0) - Y_t(0)\,\omega^{(i)} = \beta' x_{it}' + f_t' \lambda_i + \varepsilon_{it} - (\beta' X_t' + f_t' \Lambda_U + \varepsilon_t)\omega^{(i)}, \\ &= \beta'(x_{it}' - X_t' \omega^{(i)}) + f_t'(\lambda_i - \Lambda_U \omega^{(i)}) + \varepsilon_{it} - \varepsilon_t \omega^{(i)}. \end{aligned}$$

Considering that β and f_t are fixed and taking expectations yields:

$$\begin{aligned} E(\eta_{it}) &= \beta' E(x'_{it} - X'_t\omega^{(i)}) + f'_t E(\lambda_i - \Lambda_U\omega^{(i)}) + E(\varepsilon_{it} - \varepsilon_t\omega^{(i)}), \\ &\simeq \beta' E(x'_{it} - X'_t\omega^{(i)}) + f'_t E(\lambda_i - \Lambda_U\omega^{(i)}), \end{aligned}$$

in which we have used the result derived by ADH (2010) that $E(\varepsilon_{it} - \varepsilon_t\omega^{(i)})$ tends to 0 when the number of pretreatment periods T_D tends to ∞.[12] This expression should be true for any value of β and f_t and the absence of bias thus implies that:

$$E(x'_{it} - X'_t\omega^{(i)}) = 0 \text{ and } E(\lambda_i - \Lambda_U\omega^{(i)}) = 0. \tag{30}$$

The following sufficient condition is established in the Appendix:

Lemma 2 *If the support of exogenous variables and factor loadings of the treated units is a subset of the support of exogenous variables and factor loadings of the non treated units and this latter set is convex and bounded then condition (30) is satisfied at the limit when* $N - N_1 \to \infty$.

We call this case the *interpolation* case and this relates to the familiar support condition in the treatment effect literature and to the domination relationship between probability measures in the treated and control groups seen in equation (8) above.

If the support of controls does not contain the support of treated observations, the synthetic control method is based on *extrapolation* since it consists in projecting λ_i and x_{it} onto a convex set to which they do not belong and this generates a bias. For instance, to compute the distance between λ_i and the convex hull of the characteristics of the controls denoted $conv\,(\Lambda_U)$, we could use the support function (see Rockafellar, 1970) and show that:

$$d\,(\lambda_i, conv\,(\Lambda_U)) = \inf_{q \in \mathbb{R}^L} \left[\max_{j \in \{N_1+1,\ldots,N\}} (q'\lambda_j) - q'\lambda_i \right]$$

in which λ_j is the j-th column of Λ_U. Statistical methods to deal with inference in this setting could be derived from recent work by Chernozhukov, Lee and Rosen (2013) but this is out of the scope of this paper.

More generally, synthetic control is a method based on convexity arguments and thus needs assumptions based on convexity. The case of discrete regressors is a difficult intermediate case between interpolation and extrapolation that inherits the "bad" properties of extrapolation. In consequence, we conjecture that the synthetic cohort estimate is generically biased.

[12] The main difficulty there is to take into account that ω is a random function of z_i and z_j.

4 Monte Carlo experiments

4.1 The set-up

The data generating process is supposed to be given by a linear factor model:

$$y_{it} = \alpha_i I_t D_i + f_t' \lambda_i + \varepsilon_{it}$$

in which the treatment effect, α_i, is homogeneous or heterogenous across local units but not time and the number of factors L is variable. We always include additive individual and time effects, i.e. $\lambda_i = (\lambda_{i1}, 1, \lambda_{i2}, ...)'$ and $f_t = (1, f_{t1}, f_{t2}...)'$ as most economic applications would require. We did not include any other explanatory variables than the treatment variable itself.

The data generating process is constructed around a baseline experiment and several alternative experiments departing from the baseline in different dimensions such as the distribution of disturbances, the assumption that they are identically and independently distributed**,** the number of local units and periods, the correlation of treatment assignment and factor loadings, the structure of factors, the support of factor loadings and the heterogeneity of the treatment effect, α_i. Experiments are described in detail below or in the online Appendix. The Monte Carlo aspect of each experiment is given by drawing new values of $\{\varepsilon_{it}\}_{i=1,.,N,t=1,.,T}$ only and the number of replications is set to 1000.

In the baseline, individual and period shocks ε_{it} are independent and identically distributed and drawn in a zero-mean and unit-variance normal distribution.

The numbers of treated units, N_1 (resp. total, N) and the numbers of periods before treatment, T_D, (resp. total, T) as well as the number of factors L are fixed at relatively small values in line with our empirical application developed in the next section and more generally with data used in the evaluation of regional policies. In the baseline experiment, we fix $(N_1, N) = (13, 143)$, $(T_D, T) = (8, 20)$ and $L = 3$ (including one additive factor). We also experiment with L varying in the set $\{2, 4, 5, 6\}$.

The values of factors f_t and factor loadings λ_i are drawn once and for all in each experiment. Factors f_t, for $t = 1, ., T$, are drawn in a uniform distribution on $[0, 1]$ (except the first factor which is constrained to be equal to 1). Alternatively, we also experiment by fixing the second factor in f_t to the value $a.\sin(180.t/T)$ with $a > 0$ large enough.

The support of factor loadings, λ_i, is the same for treated units as for untreated units in our baseline experiment. They are drawn in a uniform distribution on $[0, 1]$ (except the second factor loading which is constrained to be equal to 1). In an alternative experiment, we construct overlapping supports for treated and untreated units. This is achieved by shifting the support of factor loadings of treated units by .5 or equivalently by adding .5 to draws. In another experiment, supports of treated and untreated units are made disjoint by shifting the support of treated units by 1. Because the original support is $[0, 1]$, this means that the intersection of the supports of treated and non-treated units is now reduced to a point. Note that adding .5 (resp. 1) to draws of treated units spawns a positive correlation between factor loadings and the treatment dummy D_i equal to .446 (resp. .706).

In the baseline experiment, the treatment effect is fixed to a constant, $\alpha_i = .3$ which is a value close to ten times the one obtained in our empirical application.

4.2 Estimation methods

We evaluate six estimation methods:

1. A direct approach using pretreatment period observations for control and treated units and post-treatment periods for the non treated only to estimate factors f_t and λ_i in the equation:

$$y_{it}(0) = f_t'\lambda_i + \varepsilon_{it} \tag{31}$$

   as in Section 3.4. The estimation procedure follows Bai's method and is based on an EM algorithm which is detailed in the Online Appendix A.1. A parameter estimate of α is then recovered from equation (23) replacing the right-hand side quantities by their empirical counterparts. This estimator is labelled "Interactive effects, counterfactual".

2. An approach whereby we estimate parameter α applying Bai's method to the linear model in which a treatment dummy is the only regressor:

$$Y_{it} = \alpha I_t D_i + f_t'\lambda_i + \varepsilon_{it}$$

   as in Section 3.3. The resulting estimator is labelled "Interactive effects, treatment dummy".

3. A matching approach (Subsection 3.5) by which equation (31) is first estimated as in the first estimation method. This yields estimates of λ_i from which a propensity score discriminating

treated and untreated units is computed. We use a logit specification for the score and construct the counterfactual outcome in the treated group in the absence of treatment at periods $t \geqslant T_D$ using the kernel method proposed by Heckman, Ichimura and Todd (1998). If we denote the score predicted by the logit model by $\hat{\mu}_i$, the counterfactual of the outcome for a given treated local unit i at a given post-treatment period is constructed as:

$$\widehat{E}\left(y_{it}\left(0\right)|D_i=1\right)=\sum\nolimits_{j=N_1+1}^{N}K_h\left(\hat{\mu}_i-\hat{\mu}_j\right)y_{jt}\Big/\sum\nolimits_{j=N_1+1}^{N}K_h\left(\hat{\mu}_i-\hat{\mu}_j\right)\ \text{ for } t\geqslant T_D$$

where $K_h\left(\cdot\right)$ is a normal kernel whose bandwidth is chosen using a rule of thumb (Silverman, 1986). An estimator of the average treatment on the treated is the average of $y_{it}-\widehat{E}\left(y_{it}\left(0\right)|D_i=1\right)$ over the population of treated local units for dates $t \geqslant T_D$. The resulting estimator is labelled "Interactive effects, matching".

4. An approach similar to "Interactive effects, counterfactual" in which we impose the constraint $\lambda_i = \Lambda_U \omega^{(i)}$ for any unit i when estimating (31). Λ_U is the $L \times (N - N_1)$ matrix comprising untreated factor loadings and $\omega^{(i)}$ are weights obtained in the synthetic control method. The estimation method is detailed in the Online Appendix A.2 and the estimator of α is recovered from (23) replacing right-hand side quantities by their empirical counterpart. This estimator is labelled "Interactive effects, constrained".

5. The synthetic control approach (Subsection 3.6) whereby the average treatment on the treated is obtained by averaging equation (26) over the population of treated units. The resulting estimator is labelled "Synthetic controls".

6. A standard difference-in-differences approach whereby we compute the FGLS estimator taking into account the covariance matrix of residuals (written in first difference). Recent research presented in Brewer, Crossley and Joyce (2013) suggests that this is the appropriate procedure if assumptions underlying difference in differences are satisfied. The resulting estimator is labelled "Diff-in-diffs".

In our simulations, the number of iterations for Bai's method involved in methods (1) to (4) is fixed to 20, and the number of iterations for the EM algorithm involved in method (1) and (4) is fixed to 1. When an estimation method using Bai's approach is implemented, we use the true

number of factors.[13]

4.3 Results

Our parameter of interest is α and we report the empirical mean, median and standard error of each estimator for every Monte-Carlo experiment. Results in the baseline case are presented in column 1 of Table 1, and unsurprisingly, show that the estimated treatment parameter exhibits little bias for all methods controlling for interactive factors: "Interactive effects, counterfactual", "Interactive effects, treatment dummy", "Interactive effects, matching", "Interactive effects, constrained" and "Synthetic controls". Similarly, the method of "Diff-in-diffs" is unbiased in spite of not accounting for interactive factors since factor loadings are orthogonal to the treatment indicator in the baseline experiment.

Interestingly, among methods allowing for interactive factors, those with constraints are the ones achieving the lowest standard errors ("Interactive effects, constrained" and "Synthetic controls") since using constraints that bind in the true model increases (*identification)* power. Note also that the standard error is larger when using the method "Interactive effects, counterfactual" than when using the method "Interactive effects, treatment dummy" as the structure of the true model after treatment in the treated group is not exploited. "Diff-in-diffs" standard errors lie between those values.

In Columns 2 and 3 of Table 1, we report results when shifting by .5 or 1 the support of individual factors for the treated. These shifts have two consequences. First, the validity conditions are now violated for interactive effect estimation which uses support constraints ("Interactive effects, constrained") and for synthetic controls. Second, they make factor loadings correlated with the treatment dummy. Results show that all methods are severely biased except "Interactive effects, counterfactual", "Interactive effects, treatment dummy" and more surprisingly "Diff-in-diffs". The two first methods are designed to properly control for interactive effects and factor loadings whatever the assumption about supports or about correlations between factor loadings and treatment. The bias for "Diff-in-diffs" is close to zero because the correlation between the factors and time indicators of treatment is close to 0 (see equation (21)). We investigate further

[13] Monte-Carlo simulations are implemented in *R*. Weights $\omega^{(i)}$ in methods (4) and (5) are computed using the *R* procedure *lsei* and the minimization algorithm *solve.QP*.

below the bias in a case in which they are correlated.

The method "Interactive effects, matching" does not work well because non-treated units close to treated units in the space of factor loadings are hard to find since the support for the treated has been shifted. We thus abstain from reporting the related results. As expected, the bias obtained for "Interactive effects, constrained" and "Synthetic controls" is large. These methods indeed impose that individual effects of treated units can be expressed as a linear combination of individual effects of non-treated units. These constraints are violated with a positive probability when the treated unit support is shifted by .5, and always violated when the support is shifted by 1.

[*Insert Table* 1]

To investigate further the cause of the surprising small bias of "Diff-in-diffs" in the previous Table, we modified the structure of factors in the experiment. The first factor in f_t is now fixed to $5.\sin(180.t/T)$ and this implies that factors and time indicators of treatment, I_t, are now correlated.[14] Table 2 shows that the "Diff-in-diffs" method can generate much larger biases in this alternative setting while biases of other methods remain the same. It is even the case that small sample biases of "Interactive effects, counterfactual", "Interactive effects, treatment dummy" become smaller in this alternative experiment.

[*Insert Table* 2]

We then make the number of factors vary between two and six (including individual and time additive effects) to assess to what extent the accuracy of estimates decreases with the number of factors. Results reported in Table 3 show that for the first three methods "Interactive effects, counterfactual", "Interactive effects, treatment dummy" and "Interactive effects, matching", the bias does not vary much and remains below 10%. Interestingly, whereas the standard error markedly increases with the number of factors for the method "Interactive effects, counterfactual", it increases much more slowly for the method "Interactive effects, treatment dummy". This occurs because factor loadings of the treated are estimated using pre-treatment periods only in the former case whereas in the latter case all periods contribute to the estimation of factor loadings. When using methods with constraints "Interactive effects, constrained" and "Synthetic controls",

[14] This correlation disappears when $T \to \infty$ as noted by a referee.

the bias can be larger than 10% but standard errors remain small. As in the baseline case, the bias of "Diff-in-diffs" is rather small although we know from the previous analysis that changing the structure of factors could make the bias larger.

[*Insert Table* 3]

There are two interesting conclusions in this analysis which bear upon our empirical application. First, the method of "Interactive effect, counterfactual" seems to be dominated in terms of bias and precision by the method "Interactive effect, treatment dummy" in all experiments and we shall thus retain only the second method. Second, the three methods "Interactive effects, matching", "Interactive effects, constrained" and "Synthetic controls" seem to behave similarly. Therefore, we shall retain only one method, synthetic controls, for our application.

4.4 Other experiments

In the Online Appendix B, we detail additional Monte-Carlo simulation results when the distribution of errors is uniform, when there are fewer pre-treatment and post-treatment periods, and when the number of local units is larger. Results conform with intuition.

We also report there, results when disturbances are not identically and independently distributed. Heteroskedasticity is introduced by drawing variances from a distribution with two points of support, with probability 1/2 for each point. We change the ratio of the two variance values across experiments. Alternatively, serial dependence is modelled as autoregressive of order 1 and we change serial correlation across experiments. This allows us to show that the number of periods that we considered $T = 20$ in line with our empirical application below is sufficiently large for asymptotic results developed in Bai (2009) to be valid. We find very little evidence of bias and the asymptotic variance of estimates obtained in the iid setting is a rather good approximation to the experimental variance. In other words, small sample biases shown by Ahn et al. (2001) can be neglected when $T = 20$.

5 Empirical Application

Our application is motivated by the evaluation reported in Gobillon et al. (2012) of an enterprise zone program implemented in France on January 1, 1997.

A survey of enterprise zone programs in the US and the UK is presented in this article as well as many particulars that we do not have the space to develop here. The fiscal incentives given by the program to enterprise zones were uniform across the country and consisted in a series of tax reliefs on property holding, corporate income, and above all on wages. The key measure was that firms needed to hire at least 20% of their labor force locally (after the third worker hired) in order to be exempted from employers' contributions to the national health insurance and pension system. This is a significant tax exemption that represents around 30% of whole labor costs (gross wage). It was expected that this measure would affect labor demand for residents of these zones and decrease unemployment. This is why we analyze the impact of such a program on unemployment entries and exits over this period.

We restrict our analysis to the Paris region in which 9 enterprise zones ("*Zones Franches Urbaines*") were created in 1997. Municipalities or groups of municipalities had to apply to the program and projects were selected taking into account their ranking given by a synthetic indicator. This indicator whose values have never been publicly released**,** aggregates five criteria: the population of the zone, its unemployment rate, the proportion of youngsters (less than 25 years old), the proportion of workers with no skill, and finally the income level in the municipality in which the enterprise zone would be located. An additional criterium is that the proposed zone should have at least 10,000 inhabitants. Nevertheless, the views of local and central government representatives who intervened in the geographic delimitation of the zones also played a role in the selection process. It thus suggests that although the selection of treated areas should be conditioned on the criteria of the synthetic indicator, it is likely that there is sufficient variability in the selection process due to political tampering. As a consequence, assumptions underlying matching estimates are not a priori invalid if observed heterogeneity is controlled for. Indeed, the supports of the propensity score in treated and non treated municipalities largely overlap though there are some outliers as shown in the Online Appendix C.3.

In Gobillon et al. (2012), we provided evidence that controlling for the effect of individual characteristics of the unemployed when studying unemployment exits only moderately affect the treatment evaluation. This is why we use raw data at the level of each municipality in the present empirical analysis. Furthermore, the destination after an unemployment exit, either to a job or to non employment, is quite uncertain in the data since unemployment spell is often terminated

because the unemployed worker is absent at a control. Many exits to a job might be hidden in the category "Absence at a control". The empirical contribution of our paper is that we investigate not only exits to a job as in Gobillon et al. (2012) but also unknown exits, as well as entries into unemployment. More generally, we assess the robustness of the results when using estimation methods which deal with the presence of a larger set of unobserved heterogeneity terms than difference in differences.

5.1 Data

We use the historical file of job applicants to the National Agency for Employment (*"Agence Nationale pour l'Emploi"* or ANPE hereafter) for the Paris region. This dataset covers the large majority of unemployment spells in the region given that registration with the national employment agency is a prerequisite for unemployed workers to claim unemployment benefits in France. We use a flow sample of unemployment spells that started between July 1989 and June 2003 and study exits from unemployment between January 1993 and June 2003. This period includes the implementation date of the enterprise zone program (January 1, 1997) and allows us to study the effect of enterprise zones not only in the short run but also in the medium run. These unemployment spells may end when the unemployed find a job, drop out of the labor force, leave unemployment for an unknown reason or when the spell is right censored.

Regarding the geographic scale of analysis, given that enterprise zones are clusters of a significant size within or across municipalities, it would be desirable to try to detect the effect of the policy at the level of an enterprise zone and comparable neighboring areas. Nevertheless, our data do not let us work at such a fine scale of disaggregation and we retain municipalities as our spatial units of analysis. Municipalities have on average twice the population of the enterprise zone they contain. As a consequence, any effect at the municipality level measures the effect of local job creation net of within-municipality transfers.

The Paris metropolitan region on which we focus is inhabited by 10.9 million people and subdivided into 1,300 municipalities. We only use municipalities which have between 8,000 and 100,000 inhabitants as every municipality comprising an enterprise zone has a population within this range. Using propensity score estimation, we select as controls municipalities whose score is close to the support of the score for treated municipalities and this restricts further our working

sample to 148 municipalities (135 controls and 13 treatments). On average, about 300 unemployed workers find a job each half-year in each of those municipalities. In view of these figures, we chose half years as our time intervals since using shorter periods would generate too much sampling variability.

Descriptive statistics relative to exits to a job, exits to non-employment, and exits for unknown reasons can be found in the Online Appendix C.2.

5.2 Results

In Table 4, we report estimation results of the enterprise zone treatment effect obtained with the most promising methods that were evaluated in the Monte Carlo experiments.[15] As explained at the end of the previous section, we use the interactive effect model with a treatment dummy and the synthetic control approach, and contrast them with the most popular method of difference in differences. Standard errors of the "Interactive effect, treatment dummy" estimates are computed using independently and identically distributed disturbances, an assumption we justify below.

We also derive a confidence interval for the synthetic control estimate which, as far as we know, has not been derived in the literature. We construct this confidence interval by inverting a test statistic whose distribution is obtained by using permutations between local units under the (admittedly strong) assumption of exchangeable disturbances across local units. The procedure is as follows. Subtract the synthetic control estimate $\hat{\alpha}$ from post treatment outcomes of treated units. Next, draw 1000 times without replacement 13 units in the whole population (treated and controls) and consider them as the new treated units while the other 135 are the new controls. Construct synthetic controls in each sample and estimate the average treatment effect. Derive the estimated quantiles $\hat{q}_{0.025}$ and $\hat{q}_{0.975}$ from the empirical distribution of estimates. Consider now any null hypothesis $H_0 : \alpha = \alpha_0$ and reject it at level 5% when $\alpha_0 - \hat{\alpha}$ does not belong to the interval bounded by those quantiles. Inverting this test yields the confidence interval, $[\hat{\alpha} + \hat{q}_{0.025}, \hat{\alpha} + \hat{q}_{0.975}]$, that is reported in Table 4. Note that we use a non pivotal statistic in the absence of any result about asymptotic standard errors of the synthetic control estimates. As a consequence, the confidence interval has no refined asymptotic properties.

[15] The only slight modification is that for the FGLS first difference estimate, the covariance matrix is kept general enough to allow for serial correlation of unknown form.

We analyze three outcomes at the level of municipalities constructed for each 6-month period between July 1993 and June 2003: exit from unemployment to a job, exit from unemployment for unknown reasons and entry into unemployment. The outcome describing unemployment exits (to a job or for unknown reasons) is defined as the logarithm of the ratio between the number of unemployed workers exiting during the period and the number of unemployed at risk at the beginning of the period. Entries are defined in the same way. Table 4 reports results using our three estimation methods for each outcome.

Starting with exits to a job, we find a small positive and significant treatment effect using the interactive effect method in line with the "Diff-in-diffs" estimate and with the findings in Gobillon et al. (2012) in which we used difference in differences but with a more limited number of periods.[16] The size of the interactive effect estimate is slightly larger than the difference-in-differences estimate and tends to increase with the number of factors that are included in the estimation. In contrast, the "Synthetic control" estimate is negative although the estimated confidence interval is so large that this estimate is not significantly different from zero at a level of 5%.

[*Insert Table* 4]

In the Monte Carlo experiment, differences between interactive effect estimates and other estimates were interpreted as an issue of disjoint supports. We plot in Figure 1, the additive local effect (i.e. the factor loading associated to the constant factor) and the multiplicative factor loading for each control unit (circle) and each treated unit (triangle) in the case in which the model includes two factors only. This graph does not exhibit any evidence against the hypothesis that the support of factor loadings for the treated units is included in the corresponding support for the controls. We tried to construct a test using permutation techniques (Good, 2005) and we failed to reject the null hypothesis of inclusion of the supports. In the absence of formal analyses of this test in the literature, we do not know however if this result is due to the low power of such a test.

[*Insert Figure* 1]

Another cause of the discrepancy between synthetic controls and interactive effects could be the

[16] This was based on an analysis distinguishing short-run and long-run effects of the program.

presence of serial correlation. When a single local effect is considered as in the difference-in-differences method, serial correlation is still substantial and the estimate of the autocorrelation of order 1 is around .35. In contrast, estimates of the serial correlation in the interactive effect model are close to zero. Factor models "exhaust" serial time dependence and this is also true for spatial dependence.[17] By contrast, we do not know much about the behavior of synthetic controls when serial correlation and spatial correlation are substantial. Interestingly, the within estimate without any correction for serial correlation is also on the negative side and close to the synthetic control estimate.

Results for other outcomes confirm the diagnostic that synthetic control estimates seem to have a behavior different from interactive effect estimates and difference-in-differences estimates. While interactive effect estimates of the treatment effect are undistinguishable from zero when we analyze exits from unemployment for unknown reasons, difference in differences yield a positive but insignificant estimate and synthetic controls a positive and significant estimate. As we have reasons to believe that the treatment effect should be larger for the outcome recording exits to a job than for the outcome recording exits for unknown reasons, synthetic control estimates seem slightly incoherent. Nonetheless, it is also true that synthetic control and interactive effect estimates for the effect of treatment on entries are very similar while difference-in-differences estimates seem surprisingly positive and nearly significant.

As a robustness check, we report in the Online Appendix C.4 the treatment effect estimates when the propensity score is introduced as a regressor. Results are very similar with those presented in the text.

6 Conclusion

In this paper, we compared different methods of estimation of the effect of a regional policy using time-varying regional data. Spatial and serial dependence are captured by a linear factor structure that permits conditioning on an extended set of unobserved local effects when applying

[17] This result is obtained using a Moran test when the distance matrix is constructed using the reciprocal of the geographical distance. Other contiguity schemes (for instance, when using discrete distance matrices constructed using 5 and 10km thresholds) capture positive spatial correlations although they diminish with the number of factors.

methods of policy evaluation. We show how difference-in-differences estimates are biased and how interactive effect methods following Bai (2009) can be applied. We compare different versions of these interactive effect methods with a synthetic control approach and with a more traditional difference-in-differences approach in Monte Carlo experiments. We finally apply the different methods to the evaluation of an entreprise zone program introduced in France in the late 1990s. In both Monte Carlo experiments and the empirical application, interactive effect estimates behave well with respect to competitors.

There are quite a few interesting extensions worth exploring in empirical analyses.

First, there is a tension between two empirical strategies in regional policy evaluations (Blundell, Costa-Dias, Meghir and van Reenen, 2004). On the one hand, choosing areas in the neighborhood of treated areas as controls might lead to biased estimates since neighbors might be affected by spillovers or contamination effects of the policy. On the other hand, non neighbors might be located too far away from the treated areas to be good matches and therefore good controls. This paper tackles this issue in a somewhat automatic way by letting factor loadings pick out spatial correlation in the data. A richer robustness analysis would allow the modification of the populations of controls and treatments by playing on the distance between municipalities and locally treated areas as was done in Gobillon et al. (2012).

Second, it is easy to extend the interactive effect procedures we have analyzed to the case in which the treatment date varies with time. This is particularly easy in the linear factor model and this set-up is used by Kim and Oka (2014). In addition, the variability of treatment dates facilitates the identification of the treatment effect since the rank condition (15) used in Section 3.3 for identification purposes is no longer needed although endogeneity issues might become more severe. The synthetic control approach can also be adapted when the treatment date varies across treated units by using a variable number of pre-treatment outcomes to construct the synthetic control.

A word of caution is also in order in case of extrapolation. When supports of exogenous variables and factor loadings of the treated units are not included in the corresponding supports of the control units, we have seen that unconstrained interactive effect estimation methods perform better than matching methods such as a constrained Bai method or synthetic controls. This conclusion is nonetheless due to our Monte Carlo setting in which the true data generating process

has linear factors. If it was non linear, this asymmetry between methods would disappear and no method would be likely to dominate each other. Extrapolation is indeed a case in which any technique needs some untestable assumptions to achieve identification. Bounds on outcome variations might however lead to partial identification of treatment effects.

REFERENCES

Abadie, A. and J., Gardeazabal, 2003, "The Economic Costs of Conflict: a case study of the Basque country", *American Economic Review*, 93, 113-132.

Abadie, A., A., Diamond and J., Hainmueller, 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.

Abadie, A., A., Diamond and J., Hainmueller, 2014, "Comparative Politics and the Synthetic Control Method", *American Journal of Political Science*, forthcoming.

Abadie, A. and G. Imbens, 2011, "Bias-Corrected Matching Estimators for Average Treatment Effects", Journal of Business & Economic Statistics, 29(1), 1-11.

Ahn, S., Y., Lee and P., Schmidt, 2001, "GMM estimation of linear panel data models with time-varying individual effects", *Journal of Econometrics,* 101, 219–255

Ahn, S., Y., Lee and P., Schmidt, 2013,"Panel Data Models with Multiple Time-Varying Individual Effects", *Journal of Econometrics*, 174, 1–14.

Athey, S. and G.Imbens, 2006, "Identification and Inference in Nonlinear Difference-in-Differences Models", *Econometrica*, 74(2), 431-497.

Bai, J., 2003, "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71(1), 135-171

Bai, J., 2009, "Panel Data Models With Interactive Fixed Effects", *Econometrica*, 77(4), 1229-1279.

Bai, J., and S. Ng, 2002, "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), pp. 191-221.

Blundell, R. and M. Costa-Dias, 2009, "Alternative Approaches to Evaluation in Empirical Microeconomics", *Journal of Human Resources*, 44, 565-640.

Blundell, R., M. Costa-Dias, C.Meghir and J. Van Reenen, 2004, "Evaluating the Employment Impact of a Mandatory Job Search Assistance Program", *Journal of European Economic Association*, 2(4), 596-606.

Brewer, M., T.F., Crossley and R., Joyce, 2013, "Inference with Differences in Differences Revisited", IZA Discussion Paper No. 7742.

Busso, M., Gregory J. and P. Kline, 2013, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy", *American Economic Review*, 103(2), 897-947.

Carneiro, P., K. T. Hansen, and J. J. Heckman, 2003, "2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44(2), 361-422.

Chernozhukov, V., S. Lee and A.M. Rosen, 2013, "Intersection Bounds: Estimation and Inference", *Econometrica,* 81(2), 667–737.

Conley, T.G, and C.R. Taber, 2011, "Inference with "Difference in Differences" with a small number of policy changes", *Review of Economics and Statistics*, 93(1), 113-125.

Doz, C., D., Giannone and L. Reichlin, 2012, "A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models", *Review of Economics and Statistics,* 94(4), 1014-1024.

Dumbgen L. and G. Walther, 1996, "Rates of Convergence for Random Approximations of Convex Sets", *Advanced Applied Probability*, 28, 384-393.

Gobillon, L., T., Magnac and H. Selod, 2012, "Do unemployed workers benefit from enterprise zones? The French experience", *Journal of Public Economics*, 96(9-10), 881-892.

Good, P.I., 2005, *Permutation, Parametric and Bootstrap Tests of Hypotheses*, Springer: New York.

Ham, J., C.W. Swenson, A. Imrohoroglu and H.Song, 2012, "Government Programs Can Improve Local Labor Markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Communities", *Journal of Public Economics*, 95(7-8), 779-797.

Heckman, J.J., H.,Ichimura and P.E.Todd, 1997, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economic Studies,* 64, 605-654.

Heckman, J.J., H.,Ichimura and P.E.Todd, 1998, "Matching as an econometric evaluation estimator", *Review of Economic Studies,* 65(223), 261–294.

Heckman J.J. and R.Robb, 1985, "Alternative Methods for Evaluating the Impact of Interventions " in *Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman and B.Singer, New York: Cambridge University Press, 156-245.

Heckman, J.J. and E.J. Vytlacil, 2007, "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation", In: James J. Heckman and Edward E. Leamer, Editor(s), *Handbook of Econometrics*, Volume 6, Part B, 4779-4874.

Hsiao, C., H.S.Ching and S.K.Wan, 2012, "A Panel Data Approach for Program Evaluation: Measuiring the Benefits of Political and Economic Integration of Hong Kong with Mainland China", *Journal of Applied Econometrics*, 27(5), 705-740.

Imbens, G., and J.M., Wooldridge, 2011, "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1), 5-86.

Kim, D., and T. Oka, 2014, "Divorce Law Reforms and Divorce Rates in the U.S.: An Interactive Fixed-Effects Approach", *Journal of Applied Econometrics*, 29(2), 231-245.

Moon, H.R. and M., Weidner, 2013a, "Dynamic Linear Panel Regression Models with Interactive Fixed Effects", CEMMAP WP 63/13.

Moon, H.R. and M., Weidner, 2013b, "Linear Regression for Panel with Unknown Number

of Factors as Interactive Effects", CEMMAP WP 49/13.

Onatski, A., 2012, "Asymptotics of the principal components estimator of large factor models with weakly influential factors", *Journal of Econometrics,* 168, pp. 244-258.

Onatski, A., Moreira M. and M. Hallin, 2013, "Asymptotic Power of Sphericity Tests for High-dimensional Data", *The Annals of Statistics*, 41(3), 1204-1231.

Pesaran, M., 2006, "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica*, 74(4), 967–1012.

Pesaran, M. and E. Tosetti, 2011, "Large panels with common factors and spatial corrrelation", *Journal of Econometrics*, 161, 182-202.

Rockafellar, R.T., 1970, *Convex Analysis,* Princeton University Press: Princeton, 472p.

Rosenbaum P. and D. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-55.

Silverman B., 1986, *Density Estimation for Statistics and Data Analysis*, Chapmal & Hall, 175p.

Westerlund, J. and J.P. Urbain, 2011, "Cross-Sectional Averages or Principal Components?", Maastricht University, Working Paper RM/11/053.

Wooldridge, J.M., 2005, "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models", *Review of Economics and Statistics*, 87(2), 385-390.

Appendix: Proof of Lemma 2

Let Y and X be some real random vectors whose supports denoted S_Y and S_X are included in $\mathbb{R}^K$. Assume that S_X is convex and bounded.

Denote D the distance between Y and its projection on the convex hull generated by n independent copies of X. Namely, let this convex hull be defined as:

$$\hat{S}_{X,n} = \{Z; Z = \sum_{j=1}^{n} \omega_j X_j, \omega_j \geq 0, \sum_{j=1}^{n} \omega_j = 1\},$$

so that:

$$D = \left\| Y - Proj_{\hat{S}_{X,n}}(Y) \right\|.$$

We shall use the result that if $n \to \infty$, $\hat{S}_{X,n} \to S_X$ in probability in the Hausdorf sense that is:

$$d_H(\hat{S}_{X,n}, S_X) = o_P(1),$$

in which d_H is the Hausdorf distance. The proof of this result is to be found in Dumbgen and Walther (1996).

Assume that $S_Y \subset S_X$. Consider any realization y of Y and a realization $\hat{S}_{x,n}$ of $\hat{S}_{X,n}$. If $y \in \hat{S}_{x,n}$ then the realization of D is zero. If $y \notin \hat{S}_{x,n}$ then the realization of D is bounded since S_X is bounded. As by the result above $d_H(\hat{S}_{X,n}, S_X) = o_P(1)$ and $y \in S_X$ then:

$$\begin{aligned} E(D) &= E(D \mid Y \in \hat{S}_{X,n}) \Pr(Y \in \hat{S}_{X,n}) + E(D \mid Y \notin \hat{S}_{X,n}) \Pr(Y \notin \hat{S}_{X,n}) \\ &= E(D \mid Y \notin \hat{S}_{X,n}) \Pr(Y \notin \hat{S}_{X,n}) \to 0 \text{ when } n \to \infty. \end{aligned}$$

Table 1: Monte-Carlo results, variation of support

Support difference	0	.5	1
Interactive effects,	**0.009**	**-0.045**	**-0.115**
counterfactual	*0.004*	*-0.046*	*-0.122*
	[0.174]	[0.204]	[0.248]
Interactive effects,	**0.009**	**-0.043**	**-0.093**
treatment dummy	*0.005*	*-0.046*	*-0.100*
	[0.155]	[0.172]	[0.284]
Interactive effects,	**0.007**	n.a.	n.a.
matching	*0.006*	n.a.	n.a.
	[0.154]	n.a.	n.a.
Interactive effects,	**-0.008**	**0.413**	**0.732**
constrained	*-0.005*	*0.418*	*0.720*
	[0.107]	[0.128]	[0.238]
Synthetic controls	**-0.017**	**0.661**	**1.510**
	-0.018	*0.660*	*1.510*
	[0.104]	[0.121]	[0.185]
Diff-in-diffs	**0.016**	**-0.052**	**-0.130**
	0.020	*-0.044*	*-0.134*
	[0.136]	[0.135]	[0.134]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, time and individual effects of the non treated drawn in a uniform distribution $[0, 1]$, individual effects of the treated drawn in a uniform distribution $[0 + s, 1 + s]$ with $s \in \{0, .5, 1\}$ reported at the top of column, errors drawn in a normal distribution with mean 0 and variance 1.

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The empirical standard error is reported in brackets.

Results for "Interactive effects, matching" are not reported when $s \in \{.5, 1\}$ as, in some simulations, some treated and non treated observations might be completely separated. As a consequence, the logit model used to construct the propensity score is not identified.

Table 2: Monte-Carlo results, variation of support, one sinusoidal factor

Support difference	0	.5	1
Interactive effects,	**0.004**	**0.007**	**0.030**
counterfactual	*0.010*	*0.014*	*0.026*
	[0.158]	[0.166]	[0.233]
Interactive effects,	**0.002**	**-0.009**	**-0.002**
treatment dummy	*0.006*	*-0.015*	*-0.007*
	[0.143]	[0.154]	[0.209]
Interactive effects,	**0.002**	n.a.	n.a.
matching	*0.006*	n.a.	n.a.
	[0.136]	n.a.	n.a.
Interactive effects,	**0.005**	**0.426**	**0.798**
constrained	*0.009*	*0.425*	*0.805*
	[0.104]	[0.119]	[0.213]
Synthetic controls	**0.010**	**0.633**	**1.420**
	0.013	*0.637*	*1.420*
	[0.102]	[0.120]	[0.206]
Diff-in-diffs	**-0.087**	**0.209**	**0.518**
	-0.087	*0.204*	*0.519*
	[0.134]	[0.134]	[0.137]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, one interactive time effect is the deterministic sinusoid $5. \sin(180.t/T)$, other time effets and individual effects of the non treated drawn in a uniform distribution $[0, 1]$, individual effects of the treated drawn in a uniform distribution $[0 + s, 1 + s]$ with $s \in \{0, .5, 1\}$ reported at the top of column, errors drawn in a normal distribution with mean 0 and variance 1.

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The empirical standard error is reported in brackets.

Results for "Interactive effects, matching" are not reported when $s \in \{.5, 1\}$ as, in some simulations, some treated and non treated observations might be completely separated. As a consequence, the logit model used to construct the propensity score is not identified.

Table 3: Monte-Carlo results, variation of the number of factors

Number of factors	1	2	3	4	5
Interactive effects,	**0.020**	**0.020**	**0.022**	**0.016**	**0.010**
counterfactual	*0.019*	*0.024*	*0.020*	*0.019*	*-0.011*
	[0.160]	[0.173]	[0.226]	[0.301]	[0.610]
Interactive effects,	**0.021**	**0.019**	**0.013**	**0.015**	**0.013**
treatment dummy	*0.020*	*0.022*	*0.015*	*0.019*	*0.010*
	[0.147]	[0.147]	[0.167]	[0.182]	[0.192]
Interactive effects,	**0.018**	**0.015**	**0.011**	**0.021**	**0.015**
matching	*0.018*	*0.017*	*0.010*	*0.016*	*0.025*
	[0.149]	[0.157]	[0.174]	[0.206]	[0.234]
Interactive effects,	**0.009**	**-0.005**	**-0.027**	**-0.011**	**-0.028**
constrained	*0.009*	*-0.007*	*-0.029*	*-0.014*	*-0.031*
	[0.111]	[0.107]	[0.109]	[0.112]	[0.118]
Synthetic controls	**0.003**	**-0.016**	**-0.045**	**-0.022**	**-0.040**
	0.004	*-0.017*	*-0.047*	*-0.023*	*-0.04*
	[0.110]	[0.105]	[0.105]	[0.110]	[0.116]
Diff-in-diffs	**0.023**	**0.020**	**0.018**	**0.028**	**0.024**
	0.022	*0.023*	*0.019*	*0.024*	*0.021*
	[0.137]	[0.132]	[0.136]	[0.136]	[0.136]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L \in \{2, 3, 4, 5, 6\}$ with L reported at the top of column, treatment parameter: $\alpha = .3$, time and individual effects drawn in a uniform distribution $[0, 1]$, errors drawn in a normal distribution with mean 0 and variance 1.

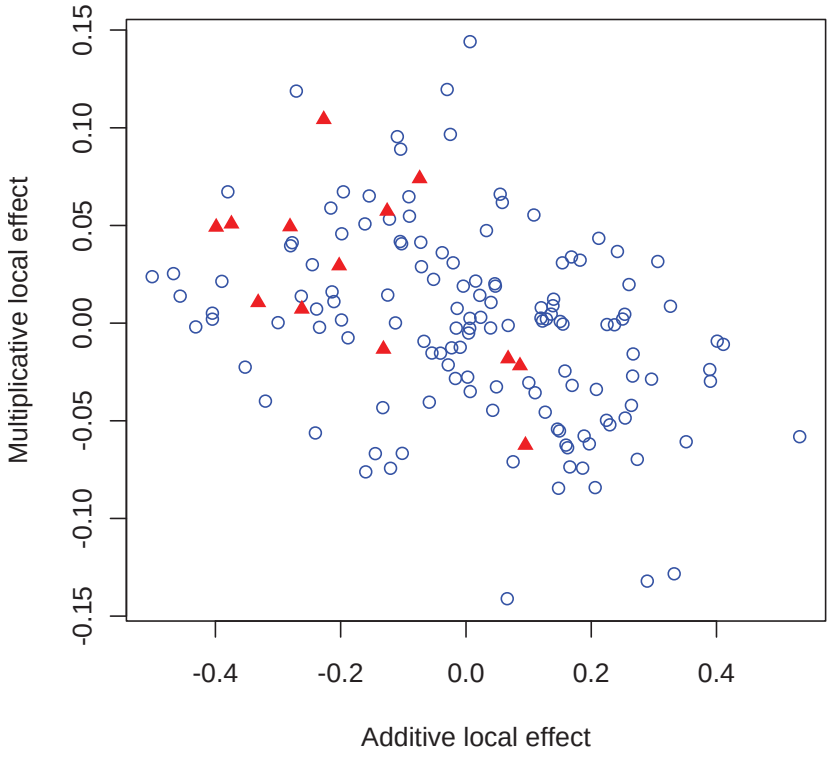
Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The empirical standard error is reported in brackets.

Table 4: Estimated enterprise zone program effects on unemployment exits and entry

Number of factors	2	3	4	5	6
Exit rate to a job					
Interactive effects,	0.032	0.036	0.039	0.043	0.046
treatment dummy	[-0.001 ; 0.065]	[-0.001 ; 0.073]	[0.006 ; 0.072]	[0.010 ; 0.076]	[0.015 ; 0.077]
Synthetic controls			-0.026		
			[-0.081 ; 0.013]		
Diff-in-diffs			0.028		
			[-0.003 ; 0.059]		
Exit rate for unknown reasons					
Interactive effects,	0.025	0.003	0.002	0.004	0.005
treatment dummy	[-0.012 ; 0.062]	[-0.032 ; 0.038]	[-0.029 ; 0.033]	[-0.027 ; 0.035]	[-0.024 ; 0.034]
Synthetic controls			0.046		
			[0.000 ; 0.091]		
Diff-in-diffs			0.019		
			[-0.012 ; 0.050]		
Entry rate					
Interactive effects,	0.007	0.006	0.004	0.008	0.007
treatment dummy	[-0.022 ; 0.036]	[-0.021 ; 0.033]	[-0.021 ; 0.029]	[-0.023 ; 0.039]	[-0.022 ; 0.036]
Synthetic controls			0.007		
			[-0.019 ; 0.034]		
Diff-in-diffs			0.020		
			[-0.004 ; 0.044]		

Notes: Outcomes are computed in logarithms at the municipality level. The number of observations are $(N1, N) = (13, 148)$ and the number of periods are $(T_D, T) = (8, 20)$. The estimated coefficient is the first reported figure. Its 95% confidence interval is given below in brackets. For the estimation method *Interactive effects, treatment dummy*, the confidence interval is computed considering that errors are independently and identically distributed. For the estimation method *Diff-in-diffs*, the feasible general least square estimator is computed assuming a constant within-municipality unrestricted covariance matrix. For *Synthetic controls*, the confidence interval is computed as explained in the text under the assumption of exchangeable errors.

Figure 1: Additive and multiplicative local effects, exit to a job



Note: Local effets are estimated using the method *Interactive model, treatment dummy* for the specification including the treatment dummy, an additive local effect and one multiplicative local effect only. Blue circle: control municipalities, red triangle: treated municipalities.

Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls

Laurent Gobillon
INED-Paris School of Economics

Thierry Magnac
Toulouse School of Economics

Online Appendix

A Proofs and estimation method

A.1 Our implementation of Bai (2009) estimation method

Notations in this subsection slightly depart from the notations used in the text since for explaining Bai's method it is not necessary to distinguish treated and untreated observations or include a treatment indicator.

The interactive model can be rewritten in vector form at the individual level as:

$$y_i = x_i\beta + F'\lambda_i + \varepsilon_i, \tag{1}$$

where:

$$\begin{aligned} y_i &= (y_{i1}, ..., y_{iT})' \text{ of dimension } T \times 1, \\ x_i &= (x'_{i1}, ..., x'_{iT})' \text{ of dimension } T \times K, \\ \varepsilon_i &= (\varepsilon_{i1}, ..., \varepsilon_{iT})' \text{ of dimension } T \times 1, \\ F &= (f_1, ..., f_T) \text{ of dimension } L \times T. \end{aligned}$$

Model (1) can also be rewritten in matrix form:

$$Y = X \circledcirc \beta + F'\Lambda + \varepsilon, \tag{2}$$

where:

$$\begin{aligned} &Y = (y_1, ..., y_N) \text{ is a } T \times N \text{ matrix}, \\ &\Lambda = (\lambda_1, ..., \lambda_N) \text{ is a } L \times N \text{ matrix}, \\ &\varepsilon = (\varepsilon_1, ..., \varepsilon_N), \\ &X \text{ is a three-dimensional } T \times K \times N \text{ matrix}, \end{aligned}$$

and in which the sign $\circledcirc$ defines in an adhoc way the operation of X with β. Matrix $X \circledcirc \beta = [x_1\beta, ., x_N\beta]$ is of dimension $T \times N$.

Restrictions are needed to identify Λ and F since for any invertible $L \times L$ matrix Q, we can always write:

$$F'\Lambda = F'Q^{-1}Q\Lambda \tag{3}$$

and prove that (Λ, F) is observationally equivalent to $(Q\Lambda, Q^{-1\prime}F)$. Following the literature (e.g. Bai, 2009), we set:

$$\begin{array}{lcll} FF'/T & = & I_L & L(L+1)/2 \text{ restrictions}, \\ \Lambda\Lambda' & \text{is} & \textit{diagonal} & L(L-1)/2 \text{ restrictions}. \end{array} \tag{4}$$

Parameters are estimated minimizing the least-square objective:

$$SSR\left(\beta,\Lambda,F\right)=\sum_{i=1}^{N}\left(y_i-x_i\beta-F'\lambda_i\right)'\left(y_i-x_i\beta-F'\lambda_i\right). \tag{5}$$

The minimization program can be solved using an iteration procedure. A minimizer in β can be computed given a value of F and Λ and minimizers in F and Λ can be computed given a value of β. It can also be shown that choosing initial values of F and Λ, and iterating leads to one solution which is the unique global minimizer. In line with Bai (2009), define the minimizer in β given parameters F and Λ as:

$$\beta\left(\Lambda,F\right)=\left(\sum_{i=1}^{N}x_i'x_i\right)^{-1}\left(\sum_{i=1}^{N}x_i'\left(y_i-F'\lambda_i\right)\right). \tag{6}$$

Note that the inverse of the matrix $\sum_{i=1}^{N}x_i'x_i$ can be computed once and for all, as it does not depend on the value of the parameters.

Conversely, given a value of β, values for F and Λ can be computed as follows. Let:

$$Z=Y-X\circledcirc\beta \text{ of dimension } T\times N. \tag{7}$$

The least square objective function (5) can be rewritten as:

$$Trace\left[\left(Z-F'\Lambda\right)'\left(Z-F'\Lambda\right)\right], \tag{8}$$

and the least-squares solution of Λ verifies:

$$\Lambda=\left(FF'\right)^{-1}FZ=FZ/T, \tag{9}$$

in which we have used the normalization $FF'/T=I_L$**.** Substituting (9) into (8) gives:

$$\begin{aligned}Trace\left[\left(Z-F'\Lambda\right)'\left(Z-F'\Lambda\right)\right] &= Trace\left[Z'(I-\frac{F'F}{T})(I-\frac{F'F}{T})Z\right]=Trace\left[Z'(I-\frac{F'F}{T})Z\right]\\ &= Trace\left(Z'Z\right)-Trace\left(Z'\frac{F'F}{T}Z\right)\\ &= Trace\left(Z'Z\right)-\frac{1}{T}Trace\left(FZZ'F'\right),\end{aligned} \tag{10}$$

in which we have used the invariance of the operator $Trace$ with respect to permutations of its arguments.

On the right-hand side, only the second term depends on F. Hence, an estimator of F is given by the maximization of $Trace\left(FZZ'F'\right)$. The estimator for F is the first L eigenvectors (multiplied

by $\sqrt{T}$ because of the restriction $FF'/T = I_L$) associated with the L first largest values of the matrix:

$$ZZ' = \sum_{i=1}^{N} (y_i - x_i\beta)(y_i - x_i\beta)'.$$

Hence, given a value of β, F is obtained as a set of eigenvectors for a given matrix. Reciprocally, for a given value of F, β can be obtained from (6) after replacing λ_i using equation (9). The minimization program becomes:

$$\overline{SSR}(\beta, F) = \sum_{i=1}^{N} (y_i - x_i\beta)' (I - \frac{F'F}{T}) (y_i - x_i\beta),$$

and the least-squares estimator of β is given by:

$$\beta(F) = \left(\sum_{i=1}^{N} x_i'(I - \frac{F'F}{T})x_i \right)^{-1} \left(\sum_{i=1}^{N} x_i'(I - \frac{F'F}{T})y_i \right).$$

Choosing some initial values for β or F, an iteration algorithm allows us to recover the OLS estimates of β and F. We finally estimate λ_i using equation (9).

A.2 Imposing constraints in estimation

We use an EM algorithm to obtain estimators of all parameters when using all observations of untreated units and pre-treatment observations of treated units only. Constraints on parameters can be imposed at further steps. The approach proceeds in the following way:

1. Step 0: Initialisation.

   We first apply Bai's estimation method developed above in the sample of the non treated group and we get preliminary and consistent estimates $\widehat{f_t}$, $\widehat{\Lambda}_U$, $\widehat{\beta}$. For the treated units before treatment, we perform the following OLS estimation:

   $$y_{it} - x_{it}\widehat{\beta} = \widehat{f_t'}\lambda_i + u_{it},$$

   and recover estimates of individual effects $\widehat{\lambda}_i$ for the treated group. We stack $\widehat{\lambda}_i$ into $\widehat{\Lambda}_T$ a consistent estimate of the corresponding matrix $\Lambda_T = (\lambda_1, ..., \Lambda_{N_1})$.

2. Step 1: Expectation Maximization Algorithm.

   - (E) We construct an expected value for the outcome in the treated group after treatment as if they were non treated (counterfactual):

     $$y_{it}(0) = x_{it}\widehat{\beta} + \widehat{f_t'}\widehat{\lambda}_i,$$

- (M) We reestimate Bai's model on the whole sample where the outcomes for the treatment group after treatment has been replaced by the (E) step. We thus recover consistent estimates $\widehat{f}_t^{(1)}$, $\widehat{\Lambda}_U^{(1)}$, $\widehat{\beta}^{(1)}$, $\widehat{\Lambda}_T^{(1)}$.

3. Step 2: Imposing constraints.

   Suppose that the set of weights derived for the synthetic control is $\omega^{(i)}$.

   We write for the control and treated groups:

$$\begin{aligned} y_{it} &= x_{it}\beta + \widehat{f}_t^{\prime(1)}\lambda_i + \varepsilon_{it}^{(1)},\ t = 1,.,T \text{ if } i \text{ not treated,} \\ y_{it} &= x_{it}\beta + \widehat{f}_t^{\prime(1)}\Lambda_U\omega^{(i)} + \varepsilon_{it}^{(1)},\ t = 1,.,T_D - 1 \text{ if } i \text{ treated.} \end{aligned}$$

   Estimating these equations by OLS, we recover estimates of β and Λ_U that we denote $\widehat{\beta}^{(2)}$ and $\widehat{\Lambda}_U^{(2)}$. To improve efficiency of the estimated factors f_t, an additional Newton Raphson step can be:

   (a) Step 3: Estimate f_t in the equations for the control and treated groups that are:

$$\begin{aligned} y_{it} - x_{it}\widehat{\beta}^{(2)} &= f_t'\widehat{\Lambda}_U^{(2)} + \varepsilon_{it}^{(2)}, t = 1,.,T \text{ if } i \text{ not treated,} \\ y_{it} - x_{it}\widehat{\beta}^{(2)} &= f_t'\widehat{\Lambda}_U^{(2)}\omega^{(i)} + \varepsilon_{it}^{(2)}, t = 1,.,T_D - 1 \text{ if } i \text{ treated.} \end{aligned}$$

B Monte-Carlo simulations

In this appendix, we present additional results of Monte-Carlo simulations when the DGP differs from the benchmark which is the following. We consider a model with both additive and interactive effects. The equation includes a treatment indicator whose coefficient is equal to .3. We have $N = 143$, $N_0 = 13$, $T = 20$, $T_D = 8$. All additive and interactive factors are drawn in a uniform distribution on [0,1]. When considering different supports for treated and non-treated units, individual factors of treated units, whether additive or multiplicative, are incremented by .5 (overlapping support) or 1 (disjoint support). The error is drawn in a zero mean and unit variance normal distribution.

Table B.1 presents the results when the error is drawn in a $\left[-\sqrt{3}, \sqrt{3}\right]$ uniform distribution (the variance is equal to 1) instead of a normal one. Results are qualitatively unchanged. Results when using a smaller number of periods T and T_D are reported in Table B.2. Biases remain similar except for the method "Interactive effects, counterfactual" for which the bias is large (due

to the fact that the number of pre-treatment periods is small). As expected, standard errors are larger. In Table B.3, we report the results when using a larger number of observations N and N_0, and obtain lower standard errors but slightly larger biases for the methods "Interactive effects, constrained" and "Synthetic controls".

[*Insert Figures B.*1, *B.*2 and *B.*3]

We also experimented with heterogeneous treatments. The treatment effect is assumed to be correlated with the second individual factor, the first factor loading, λ_{i1} being the standard linear fixed effect. The heterogenous treatment parameter is thus written as $\alpha_i = \alpha + \left(\lambda_{i2} - \frac{1}{N_1}\sum_{i=1}^{N_1}\lambda_{i2}\right)$. This specific form is retained to make sure that the average treatment on the treated is equal to α by construction as we have: $\frac{1}{N_1}\sum_{i=1}^{N_1}\alpha_i = \alpha$. This yields results close to those with homogenous treatments.

In a different vein, we also experimented with departures from iid errors. First, to generate results in Table B.4, we consider that idiosyncratic errors follow an AR(1) process $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + \varsigma_{i,t}$ in which shocks $\varsigma_{i,t}$ are iid and drawn in a normal distribution with zero mean and unit variance. We fix $\varepsilon_{i,1} = 0$. We experiment with different values of ρ from .1 to .9. The value of ρ is reported in the table at the top of the column. We find that biases are similar across simulations but standard errors increase as ρ increases.

[*Insert Figure B.*4]

Results in Table B.5 are obtained when considering two mass points for the variance of errors. Variance can be large, $V_h = 1 + a$, or small, $V_l = 1 - a$ with equal probability. Observations are still independently distributed but not identically. The ratio between the two variance values is $r = V_h/V_l$ and we have $a = (1-r)/(1+r)$. We experiment with values of r running from 1.2 to 3 reported in Table B.5 at the top of the column. We find that this kind of heteroskedasticity has nearly no effect on biases. It slightly affects standard errors but not much.

[*Insert Figure B.*5]

We also compare the asymptotic variance given in Bai (2009), and averaged across simulations, with the empirical variance computed from the simulations. Table B.6 gives the results when the number of factors varies. When there is only one additive individual effect and one interactive effect, the two variances take similar values. The discrepancy between the two variances increases as the number of interactive effects increases. Yet, as the average asymptotic variance remains

stable, it is the sample variance only which increases.

[*Insert Figure B.6*]

These last two experiments make it clear that the number of periods, $T = 20$, that we consider is sufficiently large to avoid the small sample biases that Ahn et al. (2001) describe.

C Data Appendix

C.1 The sample

We use the same data as in Gobillon et al. (2012). After eliminating the very few observations for which some socio-economic characteristics are missing, we are able to reconstruct 8,831,456 unemployment spells in the period of interest that runs from July 1993 to June 2003. We aggregate unemployment spell data by municipalities and half-years. The municipality rate of exit to a job for a given half-year is the ratio between the number of exits to a job during the half-year and the number of unemployed workers at the beginning of the half-year. The municipality rates of entry into unemployment and exit for unknown reasons are defined in the same way.

We restrict the sample by eliminating municipalities that are too small to be eligible for an enterprise zone and Paris districts whose population is much larger than that of any treated municipality. This restriction leaves us with a sample of 271 municipalities (258 controls and 13 treatments) having between 8,000 and 100,000 inhabitants. There are no noticeable differences between this restricted sample and the full sample except for population size. Roughly speaking, an average of 90,000 unemployed workers find a job each half-year and this corresponds to about 300 exits per half-year in each municipality.

C.2 Descriptive statistics

Figure C.1 describes the raw data and reports the evolution of exit rates to a job in the sample of treated municipalities and in three control groups: a sample composed by non-treated municipalities in a restricted population range and two subsamples of non-treated municipalities located respectively at a distance within 5 kilometers, and within a band of 5 to 10 kilometers around a treated municipality. For readability, we draw a vertical line at first half of 1997 (half-year 8) when the policy started to be implemented. The curves for the control groups are broadly decreasing and exhibit parallel trends throughout the period. The curve for the treatment group

slightly differs from those for the control groups between half-years 1 and 12 (second half of 1993 to first half of 1999). In particular, the exit rate to a job remains flat in the treatment group between half-years 7 and 8 (second half-year of 1996 and first half-year of 1997) when the policy is implemented whereas it is decreasing in the control groups. The estimation of the treatment parameter that we report in this paper is a way of formalizing and testing that these diverging trends are statistically significant.

[*Insert Figure C.*1]

None of these differences appears in the evolution of exit rates to non-employment and the evolution of exit rates for unknown reasons.

C.3 Propensity score

Next, we estimate a Probit model of enterprise zone designation as a function of municipality control variables among which are measures of physical job accessibility, the municipal composition of the population in terms of nationality or education, the unemployment rate, the proportion of young adults, and household income (proxying for the fiscal potential). We also include in the specification the smallest distance to another municipality comprising an enterprise zone. This is to account for the willingness of authorities to scatter enterprise zones more or less evenly throughout the region.

The results of our benchmark weighted Probit in which the weight is the (square root of the) number of unemployed workers in the municipality appear in the first column of Table C.1.

[*Insert Table C.*1]

In line with the selection criteria, the larger the average household income in the municipality or the smaller the proportion of persons without a high school diploma, the less likely the municipality comprises an enterprise zone although the effect is hardly significant for the latter variable. The higher the proportion of individuals below 25 years of age or the larger the size of the population, the larger the probability that the municipality contains an enterprise zone. We also find that the larger the density of jobs attainable in less than 60 minutes by private vehicle, the less likely it is that the municipality will be endowed with an enterprise zone. This is consistent with the targeting of places with relatively lower job accessibility. The effect of the distance to the nearest enterprise zone is not significant.

Using the results in the first column, we predicted the propensity score for each municipality. Interestingly, it reveals that the supports of the predicted propensity scores in the treated and control groups differ quite markedly as shown in Table C.2.

[*Insert Table C.2*]

The smallest predicted probability in the treatment group is equal to 0.08%, a low score which is consistent with political tampering in designation. In order to satisfy the common support condition (Smith and Todd, 2005), we further restrict the control group to municipalities whose predicted propensity scores are larger than the value 0.04% (see Table C.3). This restriction shrinks the control group by a factor of 2 and it now includes 135 municipalities (instead of 258), which is about ten times the number of treated municipalities (13).

In Table C.3, we report averages of explanatory variables in the treatment and control groups to assess whether those groups are balanced.

[*Insert Table C.3*]

Since the treatment group is small, it seems difficult to report averages in stratified samples defined by the propensity score levels (Smith and Todd, 2005). We rather report them globally even if results are less easy to interpret. The covariates of interest seem to be balanced in the two subsamples except for two variables: the proportion of college graduates and household income. This explains the differences in the propensity score averages between the control and treatment group. Nevertheless, the coefficient of designated municipalities in linear regressions of those covariates on the propensity score and the designation indicator is not significant even at the 10% level which indicates that samples are approximately balanced.

C.4 Robustness checks

Table C.4 reports the estimated effect of the enterprise zone program on entries and exits when controlling for the estimated propensity score. In the interactive effect and difference-in-differences approaches, this is done by including among the regressors the propensity score interacted with a trend t/T to mimic the presence of the propensity score in levels in the first difference equation as in Gobillon et al. (2012). We also include the propensity score among variables used in the construction of synthetic controls. Results obtained using the method "Interactive effect, treatment dummy" are very close to those obtained in the baseline case except when studying entries. In that case, the treatment effect estimate is larger when the specification includes four or less

factors (including an additive one). Treatment effects for the other outcomes – exits to a job and exits for unknown reason – when using synthetic controls are now close to zero. Results obtained with difference in differences are similar to those obtained previously. In summary, once again, synthetic controls estimates are more sensitive to the specification than factor model estimates and difference-in-differences estimates. Overall, results are in line with those reported in the article when there is no control for the propensity score.

[*Insert Table C.*4]

REFERENCES

Bai, J., 2009, "Panel Data Models With Interactive Fixed Effects", *Econometrica*, 77(4), 1229-1279.

Gobillon, L., T., Magnac and H. Selod, 2012, "Do unemployed workers benefit from enterprise zones? The French experience", *Journal of Public Economics*, 96(9-10):881-892.

Smith, J.A.. and P. Todd, 2005, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators", *Journal of Econometrics*, 125, 305-353.

Table B.1: Monte-Carlo results, variation of support, uniform errors

Support difference	0	.5	1
Interactive effects,	**0.024**	**-0.040**	**-0.092**
counterfactual	*0.021*	*-0.043*	*-0.101*
	[0.179]	[0.184]	[0.237]
Interactive effects,	**0.022**	**-0.040**	**-0.082**
treatment dummy	*0.024*	*-0.037*	*-0.090*
	[0.162]	[0.168]	[0.273]
Interactive effects,	**0.022**	n.a.	n.a.
matching	*0.026*	n.a.	n.a.
	[0.161]	n.a.	n.a.
Interactive effects,	**0.000**	**0.406**	**0.721**
constrained	*0.002*	*0.409*	*0.714*
	[0.111]	[0.128]	[0.237]
Synthetic controls	**-0.011**	**0.638**	**1.490**
	-0.014	*0.638*	*1.490*
	[0.106]	[0.120]	[0.179]
Diff-in-diffs	**0.026**	**-0.051**	**-0.125**
	0.028	*-0.048*	*-0.130*
	[0.143]	[0.136]	[0.135]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, time and individual effects of the non treated drawn in a uniform distribution $[0, 1]$, individual effects of the treated drawn in a uniform distribution $[0 + s, 1 + s]$ with $s \in \{0, .5, 1\}$ reported at the top of column, errors drawn in a uniform distribution $\left[-\sqrt{3}, \sqrt{3}\right]$.

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The empirical standard error is reported in brackets.

Results for "Interactive effects, matching" are not reported when $s \in \{.5, 1\}$ as, in some simulations, some treated and non treated observations might be completely separated. As a consequence, the logit model used to construct the propensity score is not identified.

Table B.2: Monte-Carlo results, variation of the number of periods

Number of periods	$T = 20, T_D = 8$	$T = 10, T_D = 4$
Interactive effects,	**0.022**	**-0.128**
counterfactual	*0.027*	*0.009*
	[0.175]	[19.80]
Interactive effects,	**0.025**	**0.022**
treatment dummy	*0.026*	*0.016*
	[0.153]	[0.256]
Interactive effects,	**0.016**	**0.021**
matching	*0.016*	*0.022*
	[0.158]	[0.264]
Interactive effects,	**-0.001**	**-0.009**
constrained	*0.001*	*-0.014*
	[0.110]	[0.142]
Synthetic controls	**-0.013**	**-0.020**
	-0.011	*-0.029*
	[0.108]	[0.132]
Diff-in-diffs	**0.028**	**0.023**
	0.025	*0.030*
	[0.137]	[0.202]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) \in \{(4, 10), (8, 20)\}$ with (T_D, T) reported at the top of column, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, time and individual effects drawn in a uniform distribution $[0, 1]$, errors drawn in a normal distribution with mean 0 and variance 1.

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The standard error is reported in brackets.

Table B.3: Monte-Carlo results, variation of the number of units

Number of individuals	$N = 143, N_0 = 13$	$N = 275, N_0 = 25$
Interactive effects,	**0.007**	**0.003**
counterfactual	*0.009*	*0.003*
	[0.174]	[0.126]
Interactive effects,	**0.010**	**0.002**
treatment dummy	*0.012*	*0.005*
	[0.155]	[0.107]
Interactive effects,	**0.003**	**0.006**
matching	*0.003*	*0.006*
	[0.157]	[0.117]
Interactive effects,	**-0.007**	**0.047**
constrained	*-0.011*	*0.045*
	[0.111]	[0.078]
Synthetic controls	**-0.016**	**0.069**
	-0.017	*0.069*
	[0.109]	[0.075]
Diff-in-diffs	**0.016**	**0.000**
	0.017	*0.000*
	[0.137]	[0.098]

Data generating process: number of observations: $(N_1, N) \in \{(13, 143), (25, 275)\}$ with (N_1, N) reported at the top of column, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, time and individual effects drawn in a uniform distribution $[0, 1]$, errors drawn in a normal distribution with mean 0 and variance 1.

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The standard error is reported in brackets.

Table B.4: Monte Carlo results: AR(1) errors

AR(1) autocorrelation	.1	.3	.5	.7	.9
Interactive effects,	**0.015**	**0.029**	**-0.026**	**-0.016**	**-0.039**
linear counterfactual	*0.020*	*0.022*	*-0.021*	*-0.044*	*-0.001*
	[0.235]	[0.436]	[0.660]	[0.838]	[1.530]
Interactive effects,	**0.013**	**0.023**	**0.020**	**0.027**	**0.050**
linear	*0.016*	*0.023*	*0.025*	*0.022*	*0.036*
	[0.188]	[0.301]	[0.395]	[0.437]	[0.477]
Interactive effects,	**0.012**	**0.011**	**-0.015**	**-0.02**	**-0.007**
matching	*0.015*	*0.017*	*-0.003*	*-0.009*	*0.006*
	[0.193]	[0.285]	[0.369]	[0.545]	[1.020]
Interactive effects,	**-0.003**	**-0.008**	**-0.021**	**-0.018**	**0.014**
constrained	*0.000*	*-0.007*	*-0.016*	*-0.022*	*0.006*
	[0.120]	[0.140]	[0.178]	[0.259]	[0.494]
Synthetic controls	**-0.012**	**-0.015**	**-0.023**	**-0.019**	**0.017**
	-0.01	*-0.01*	*-0.024*	*-0.02*	*0.005*
	[0.117]	[0.136]	[0.177]	[0.260]	[0.495]
Within	**0.024**	**0.027**	**0.022**	**0.014**	**0.035**
	0.029	*0.025*	*0.023*	*0.026*	*0.035*
	[0.147]	[0.183]	[0.236]	[0.312]	[0.490]
First difference	**0.024**	**0.027**	**0.022**	**0.014**	**0.035**
	0.029	*0.025*	*0.023*	*0.026*	*0.035*
	[0.147]	[0.183]	[0.236]	[0.312]	[0.490]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, time and individual effects drawn in a uniform distribution $[0, 1]$, errors drawn in an AR(1) process whose autocorrelation is reported at the top of column and innovations are white noise (normally distributed with mean 0 and variance 1).

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The standard error is reported in brackets.

Table B.5: Monte Carlo results: heteroskedastic errors

Variance ratio	1.2	1.5	2	2.5	3
Interactive effects,	**0.013**	**0.014**	**0.021**	**0.012**	**0.015**
linear counterfactual	*0.014*	*0.014*	*0.021*	*0.011*	*0.016*
	[0.170]	[0.158]	[0.140]	[0.122]	[0.123]
Interactive effects,	**0.016**	**0.015**	**0.016**	**0.011**	**0.016**
linear	*0.018*	*0.012*	*0.013*	*0.012*	*0.014*
	[0.148]	[0.137]	[0.126]	[0.109]	[0.106]
Interactive effects,	**0.010**	**0.013**	**0.015**	**0.009**	**0.010**
matching	*0.015*	*0.008*	*0.015*	*0.011*	*0.015*
	[0.152]	[0.144]	[0.128]	[0.116]	[0.114]
Interactive effects,	**-0.002**	**-0.001**	**0.000**	**-0.002**	**0.002**
constrained	*-0.003*	*-0.006*	*-0.002*	*-0.003*	*0.004*
	[0.107]	[0.100]	[0.093]	[0.081]	[0.078]
Synthetic controls	**-0.011**	**-0.009**	**-0.01**	**-0.009**	**-0.007**
	-0.014	*-0.01*	*-0.011*	*-0.008*	*-0.003*
	[0.106]	[0.098]	[0.092]	[0.081]	[0.077]
Within	**0.021**	**0.020**	**0.019**	**0.016**	**0.021**
	0.018	*0.018*	*0.015*	*0.016*	*0.020*
	[0.132]	[0.121]	[0.111]	[0.097]	[0.093]
First difference	**0.021**	**0.020**	**0.019**	**0.016**	**0.021**
	0.018	*0.018*	*0.015*	*0.016*	*0.020*
	[0.132]	[0.121]	[0.111]	[0.097]	[0.093]

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L = 3$, treatment parameter: $\alpha = .3$, time and individual effects drawn in a uniform distribution $[0, 1]$, errors drawn in a normal distribution with mean 0 and a two-point-of-support variance: $V_h = 1 + a$ and $V_l = 1 - a$. Observations are randomly allocated to large and small variance with equal probability 0.5. The variance ratio $r = V_h/V_l$ is reported at the top of the column.

Notes: Estimation methods are detailed in Section 4.1. $S = 1000$ simulations are used. The average (resp. median) estimated bias is reported in bold (resp. italic). The standard error is reported in brackets.

Table B.6: Monte-Carlo results: empirical variance and theoretically derived asymptotic variance

Number of factors	2	3	4	5	6
Mean bias	**0.017**	**0.011**	**0.010**	**0.019**	**0.022**
Empirical Standard error	*0.142*	*0.152*	*0.165*	*0.181*	*0.191*
Square root Asymp. Var	0.131	0.129	0.128	0.129	0.128

Data generating process: number of observations: $(N_1, N) = (13, 143)$, number of periods: $(T_D, T) = (8, 20)$, number of factors (including an additive one): $L \in \{2, 3, 4, 5, 6\}$ with L reported at the top of column, treatment parameter: $\alpha = .3$, time and individual effects drawn in a uniform distribution $[0, 1]$, errors drawn in a normal distribution with mean 0 and variance 1.

Notes: Mean bias: average bias computed across simulations; Empirical Standard error: empirical standard error computed across simulations; Square root Asymp. Var: square root of the average of the asymptotic variance computed across simulations.

Table C.1: Propensity score: effect of municipality characteristics on the probability of designation to receive an enterprise zone

	Weighted	Unweighted
Job density, 60 minutes by private vehicle	-3.999*	-4.171*
	(2.109)	(2.298)
Proportion of no diploma	37.779*	24.029
	(22.249)	(22.865)
Proportion of technical diplomas	20.998	0.974
	(28.215)	(28.900)
Proportion of college diplomas	38.978	17.299
	(29.889)	(31.336)
Distance to the nearest EZ	-0.027	-0.035
	(0.024)	(0.024)
Proportion of individuals below 25 in 1990	17.125***	11.834**
	(5.156)	(5.256)
Population in 1990	0.021**	0.019*
	(0.009)	(0.011)
Average net household income in 96	-4.975***	-2.033
	(1.563)	(1.593)
Constant	-32.115	-16.526
	(21.818)	(22.537)
Nb. observations	271	271
Pseudo-R^2	.542	.477

Note: ***: significant at 1% level; **: significant at 5% level; *: significant at 10% level. Probit estimation. The dependent variable is a dummy equal to one if the municipality is designated to receive an EZ (and zero otherwise). The sample is restricted to municipalities whose population is between 8,000 and 100,000 in 1990. The first column reports results when weighting by the square root of the number of unemployed workers at risk at the beginning of half-year 8 (1st half year of 1997), and the second column when using no weight.

Table C.2: Frequency of non-treated municipalities by propensity score brackets which bound are values of treated municipalities

Score bracket	Number of non-treated municipalities
[0, 0.0008)	125
[0.0008, 0.0119)	60
[0.0119, 0.1161)	52
[0.1161, 0.1772)	7
[0.1772, 0.3111)	6
[0.3111, 0.4404)	4
[0.4404, 0.4765)	0
[0.4765, 0.6091)	2
[0.6091, 0.7723)	2
[0.7723, 0.7933)	0
[0.7933, 0.8537)	0
[0.8537, 0.9032)	0
[0.9032, 0.9949)	0
[0.9949 , 1]	0
Total	258

Note: The observation unit is a municipality having between 8,000 and 100,000 inhabitants. The propensity score is computed as the predicted probability as derived from Probit estimates reported in Table C.1, column (1). Each bracket bound (excluding 0 and 1) corresponds to the propensity score of one treated municipality, with treated municipalities being sorted in ascending order of their estimated propensity score.

Table C.3: Average of municipality characteristics in treatment and control groups

	Treatment group	Control group, propensity score > 0.0004
Job density, 60 minutes by public transport	0.838	0.850
	(0.119)	(0.119)
Proportion of no diploma	0.536	0.465
	(0.041)	(0.074)
Proportion of technical diplomas	0.222	0.219
	(0.009)	(0.031)
Proportion of college diplomas	0.122	0.179
	(0.025)	(0.075)
Distance to the nearest EZ	9.074	11.016
	(12.193)	(8.051)
Proportion of individuals below 25 in 1990	0.416	0.372
	(0.038)	(0.043)
Population in 1990	45.201	43.578
	(18.226)	(26.357)
Average net household income in 96	0.375	0.509
	(0.087)	(0.125)
Number of observations	13	135

Note: Standard errors are reported in parenthesis under the means. The observation unit is a municipality between 8,000 and 100,000 inhabitants. Only municipalities whose estimated propensity score is greater than 0.0004 are considered in the control group. The propensity score is computed using the estimated coefficients of Table C.1, column (1).

Table C.4: Estimated enterprise zone program effects on unemployment exits and entry when controlling for the propensity score

Number of factors	2	3	4	5	6
Exit rate to a job					
Interactive effects,	0.033	0.034	0.037	0.039	0.046
treatment dummy	[-0.004 ; 0.070]	[-0.005 ; 0.073]	[0.002 ; 0.072]	[0.004 ; 0.074]	[0.013 ; 0.079]
Synthetic controls			-0.006		
			[-0.074 ; 0.036]		
Diff-in-diffs			0.041		
			[0.008 ; 0.074]		
Exit rate for unknown reasons					
Interactive effects,	0.022	0.008	0.006	0.015	0.014
treatment dummy	[-0.017 ; 0.061]	[-0.027 ; 0.043]	[-0.027 ; 0.039]	[-0.018 ; 0.048]	[-0.017 ; 0.045]
Synthetic controls			0.004		
			[-0.033 ; 0.081]		
Diff-in-diffs			0.021		
			[-0.012 ; 0.054]		
Entry rate					
Interactive effects,	0.021	0.028	0.029	0.008	0.008
treatment dummy	[-0.010 ; 0.052]	[-0.001 ; 0.057]	[0 ; 0.058]	[-0.023 ; 0.039]	[-0.021 ; 0.037]
Synthetic controls			0.003		
			[-0.016 ; 0.048]		
Diff-in-diffs			0.027		
			[0.002 ; 0.052]		

Notes: Outcomes are computed in logarithms at the municipality level. The number of observations are $(N1, N) = (13, 148)$ and the number of periods are $(T_D, T) = (8, 20)$. The estimated coefficient is the first reported figure. Its 95% confidence interval is given below in brackets. For the estimation method *Interactive effects, treatment dummy*, the confidence interval is computed considering that errors are independently and identically distributed. For the estimation method *Diff-in-diffs*, the feasible general least square estimator is computed assuming a constant within-municipality unrestricted covariance matrix. For *Synthetic controls*, the confidence interval is computed as explained in the text under the assumption of independently and identically distributed errors.

Figure C.1: Exit rate to a job, by group of municipalities

Note: The exit rates to a job are reported for half-years between 1 (2nd half year of 1993) and 20 (1st half year of 2003). Half-year 8 (1st half year of 1997) is the first half-year during which some municipalities are treated. Non-EZ: municipalities which do not include an EZ. 8,000-100,000: population between 8,000 and 100,000 in 1990. 0-5km: between 0 and 5km of a municipality including an EZ. Enterprise zones: municipalities which include an EZ.