



## Mesurer l'impact économique d'une politique culturelle

*Étude de cas sur le programme Lille, Capitale Européenne de la Culture en 2004*

**Louise-Taos AIMENE**

**Édith DARIN**

**Marc GINESTET**

Sous la direction de Thibault Brodaty

ENSAE Paristech  
Projet de statistiques appliquées 2016

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Présentation de la méthode de contrôle synthétique</b>	<b>5</b>
2.1	Méthodologie . . . . .	5
2.2	Les avantages du modèle . . . . .	7
2.2.1	Une protection contre l'extrapolation . . . . .	7
2.2.2	Une garantie de non-dépendance au modèle . . . . .	7
2.2.3	Une explicitation de la contribution relative de chaque unité à la zone de contrôle . . . . .	8
2.2.4	Une systématisation du choix des unités de comparaison qui rend possible l'inférence . . . . .	8
2.3	Hypothèses fondamentales sous-jacentes au modèle . . . . .	9
2.3.1	Hypothèse de restriction dans la construction du groupe de donneurs . . . . .	9
2.3.2	Hypothèse d'exclusion dans la construction du groupe de donneurs . . . . .	9
<b>3</b>	<b>Implémentation : Estimation de l'impact économique de "Lille 2004"</b>	<b>9</b>
3.1	Les données . . . . .	9
3.2	Une approche en statistiques descriptives . . . . .	11
3.3	La construction du groupe de donneurs . . . . .	12
3.4	Les résultats : impact de <i>Lille 2004</i> sur le PIB . . . . .	13
3.5	Les résultats : impact de <i>Lille 2004</i> sur le chômage . . . . .	15
3.6	Limites d'une étude sur le tourisme à l'échelle départementale . . . . .	16
<b>4</b>	<b>La robustesse des résultats : inférence sur l'effet de "Lille 2004"</b>	<b>16</b>
4.1	Les tests de robustesse . . . . .	17
4.2	Les tests de falsification . . . . .	18
<b>5</b>	<b>Conclusion</b>	<b>20</b>
<b>6</b>	<b>Annexes</b>	<b>21</b>
6.1	Le contenu de <i>Lille 2004</i> . . . . .	21
6.2	Les résultats sur le modèle du PIB . . . . .	22
6.3	Le modèle et ses propriétés économétriques . . . . .	24
6.4	L'implémentation algorithmique . . . . .	25
6.5	Validation de l'hypothèse de restriction . . . . .	26
6.6	Validation de l'hypothèse d'exclusion . . . . .	27
6.7	Tableaux de résultat du placebo de sensibilité . . . . .	30
6.7.1	Modèle de base . . . . .	30
6.7.2	Sans le Poitou-Charentes . . . . .	30
6.7.3	Sans le Pays de la Loire . . . . .	31

# 1 Introduction

## - *Économie et culture : nécessité de l'évaluation, difficulté de l'estimation*

En 2010, 7,45 milliards d'euros ont été dépensés par les collectivités territoriales pour la culture (Delvainquière *et alii* (2014) [11]). Ce chiffre permet de mettre en évidence l'importance de la culture dans les politiques publiques, nécessitant donc de mettre en place son évaluation afin de garantir une utilisation optimale des dépenses publiques. Cependant les liens entre économie et culture ne se résument pas simplement à cette thématique du financement des actions culturelles.

En effet, comme le développe le courant d'économie de la culture représenté en France par E. Benhamou, la culture, et les activités qui l'environnent, peuvent être la source de retombées économiques pour le territoire sur lequel ont lieu les interventions, que ce soit en termes de croissance économique, de création d'entreprises ou de recul du chômage. Ces effets peuvent être directs, dans la mesure où l'organisation d'événements culturels implique le recrutement de travailleurs mais aussi la construction de bâtiments ou d'infrastructures – ayant un effet à court terme sur le contexte économique d'une région. Mais l'impact peut aussi être indirect sur d'autres secteurs d'activité non directement impliqués, principalement le secteur touristique, qui bénéficient ainsi de l'effet d'attractivité de l'événement. Or ceci peut se doubler d'un effet de long terme avec l'espérance que l'intervention culturelle change en profondeur l'image d'un territoire tant du point de vue touristique que du point de vue de l'implantation de nouvelles activités économiques.

A travers cette brève esquisse des enjeux économiques de la culture, on perçoit la nécessité de pouvoir précisément mesurer l'impact réel des actions culturelles afin d'évaluer l'efficacité de la politique publique et le retour sur investissement que cela engendre. Cependant évaluer une mesure d'intervention est une tâche méthodologiquement délicate car il s'agit de comparer la situation observée une fois la mesure mise en place avec la « situation contre-factuelle » inobservée (i.e ce qui serait advenu en l'absence de politique publique).

## - *La politique des « Capitales européennes de la culture » et son traitement dans la littérature économique*

Nous souhaitons nous positionner dans ce champ méthodologique en analysant les retombées économiques du programme culturel des « Capitales européennes de la culture », lancé en 1985 à l'initiative de la ministre grecque de la culture Melina Mercouri et destiné à mettre à l'honneur la diversité culturelle de l'Europe et à améliorer son patrimoine culturel commun grâce au développement de la coopération transfrontalière entre les acteurs et les institutions du secteur culturel.

Cet événement circonscrit dans le temps et dans l'espace, puisqu'ayant lieu à l'échelle d'une année dans une ville européenne, est un cadre d'analyse simple pour chercher à appréhender l'impact économique d'une politique culturelle. Les premières analyses menées raisonnaient de façon qualitative en dressant un panorama discursif de l'évolution de ce programme culturel (Mettler (2008) [22]) ou en se focalisant sur le ressenti des acteurs à travers des questionnaires (Palmer-Rae (2004), ou Richards et Wilson (2004) [24] ). D'autres études consistent à mesurer les dépenses des visiteurs, et à estimer comment cette dépense se diffuse dans l'économie via le mécanisme du multiplicateur ( Herrero *et alii* (2006) [14] <sup>1</sup>). Une des hypothèses principales de ces méthodes est de supposer qu'en l'absence d'événement aucun nouveau visiteur n'aurait été remplacé par d'autres visiteurs. Or des contraintes telles que la capacité d'accueil limitée, la probable congestion des transports ou des hausses de prix exceptionnelles

peuvent pousser certains visiteurs à renoncer à leur séjour <sup>2</sup>.

Pour contourner les limites de ces approches il est nécessaire d'appliquer des méthodes économétriques d'évaluations ex-post qui permettent de raisonner toutes choses égales par ailleurs et donc d'isoler des effets causaux pur et ainsi d'estimer la situation contrefactuelle. Ce cadre d'analyse renvoie à la notion d'estimation en différence de différence, méthode couramment employée depuis l'analyse du marché du travail par Card et Krueger(1994) et importée dans le domaine de la mesure d'impact d'événements et notamment sportif (cf. par exemple par Jasmand et Maennig (2008) [17] pour les Jeux Olympiques). Cette approche est celle de Gomes et Librero-Cano (2014) [13] qui estiment que le PIB des villes ayant accueilli l'initiative Capitale européenne de la culture augmente de 4,5% par rapport aux villes ne l'ayant pas accueillie, avec un effet perceptible durant 5 ans. Nous souhaitons vérifier ce résultat à l'échelle d'une seule ville, Lille, qui a expérimenté cette action culturelle en 2004.

#### *-Le cas d'étude : Lille Capitale Européenne de la Culture en 2004*

Les objectifs économiques de Lille 2004 étaient clairement définis: il s'agissait via la culture de changer durablement l'image des Flandresurement affectées par le phénomène de désindustrialisation pour attirer de nouveaux investisseurs et confirmer une évolution positive engagée par une reconversion économique et urbaine tournée vers l'Europe et l'innovation. Le budget total était de 76 millions d'euros, provenant de subventions publiques, de donations privées et de mécénat. 2500 manifestations ont été organisées (les deux tiers ayant eu lieu dans la métropole lilloise) attirant au total 9 000 000 participants.

Un bilan chiffré (Lille Horizon 2004) a été publié à la suite de l'événement ainsi que des études sur l'impact direct et de court terme (Maurence (2012) ou Nicolas (2007)). De façon plus systématique une thèse a été consacrée à l'évolution de nombreux indicateurs économiques afin d'évaluer un effet Lille 2004 à partir de statistiques descriptives (Werquin (2006) [26]). Cette méthodologie, bien qu'offrant un panorama exhaustif de la situation économique du territoire, ne permet pas de conclure sur un effet causal de l'événement. L'enjeu s'inscrit dans le courant des études comparatives : il s'agit, suivant le cadre conceptuel de Mills, de trouver des unités de comparaison qui pourraient reproduire le contrefactuel de Lille 2004, c'est-à-dire décrire comment aurait évolué l'économie du Nord-Pas-de-Calais en l'absence de cette politique culturelle. Pour combler cette faille, Brodaty (2011) a construit un modèle d'intervention à partir des séries d'offres d'emplois en incluant dans la régression des variables de contrôle, indicatrices de régions qui n'ont pas pu être affectées par l'événement, afin de raisonner en termes de zone d'impact/zone de contrôle. Il conclut que Lille 2004 a eu un impact globalement positif de 2004 à 2006 dans le Nord-Pas-de-Calais<sup>3</sup>. Nous souhaitons nous inscrire dans cette même démarche de réflexion tout en systématisant la notion de zone de contrôle puisque cette étape est cruciale dans la détermination d'un effet pur de l'intervention concernée.

#### *-La méthode de contrôle synthétique : extension du modèle de différence-de-différence*

La méthode de contrôle synthétique, développée par Abadie et Gardeazabal (2003) [2] puis formalisée par Abadie, Diamond et Hainmueller (2010) [1] s'inscrit dans un entre-deux entre les analyses en différence-de-différence et les analyses comparatives de cas. Comme les différences-de-différences, elle s'applique lorsque l'on dispose de données de panel à un niveau agrégé sur une intervention ayant affecté une entité alors qu'un grand nombre d'entités n'ont pas été affectées par cette intervention. Mais contrairement à cette analyse, la méthode d'Abadie et al. permet aux

variables inobservées affectant la variable d'intérêt de varier dans le temps, autrement dit elle surmonte le biais de variable omise tout en contournant la nécessité d'une hypothèse d'évolution commune (« common trend ») propre aux modèles de panel à effet fixe. En effet ces auteurs proposent de construire un contrefactuel à partir d'une moyenne pondérée des caractéristiques des unités de contrôle. Il s'agit donc d'une méthode fondée strictement sur l'observation des données qui systématisé le choix de l'unité de comparaison. Par ailleurs elle permet d'utiliser des tests d'inférence en dépit du petit nombre d'unités observées.

Après avoir été appliquée par Abadie et Gardeazabal sur le terrorisme dans le pays basque puis avec Hainmueller et Diamond sur la réunification allemande ou le passage d'une loi anti-tabac en Californie, cette approche a été largement reprise récemment que ce soit pour des études empiriques<sup>4</sup> ou dans des développements théoriques (Ando et Sävje (2013) [3] ou Kaul *et alii* (2016) [19]).

#### *- Résultats et organisation de l'article*

Nous appliquons donc cette méthode pour étudier l'effet de *Lille 2004*, au niveau régional du Nord-Pas-de-Calais à travers deux indicateurs économiques, le PIB et le chômage. Nous démontrons que cet événement a permis au PIB de cette région d'augmenter de 1,7% par rapport à ce qu'il aurait été sans cette intervention et a eu un impact négatif sur le chômage. Par ailleurs en utilisant des méthodes inférentielles nous prouvons la significativité de nos estimations. Notre mémoire est organisé comme suit : dans la première section (1.), nous abordons la méthodologie, ses hypothèses et ses implications; la deuxième section (2.) est consacrée à l'implémentation de la méthode donc aux données que nous utilisons, et aux résultats obtenus, tandis que la troisième (3.) expose les tests de robustesse et de falsification que nous avons menés sur les résultats concernant le PIB.

## 2 Présentation de la méthode de contrôle synthétique

### 2.1 Méthodologie

La méthode de contrôle synthétique (MCS) permet de créer un estimateur de l'effet d'un traitement sur une variable d'intérêt. Soit  $Y_{i,t}(0)$ , la variable d'intérêt pour l'unité  $i$  à la date  $t$ , en l'absence de traitement. De manière similaire on définit  $Y_{i,t}(1)$ , la variable d'intérêt pour l'unité  $i$  à la date  $t$ , en présence du traitement.

Soit  $\alpha_{i,t} = Y_{i,t}(1) - Y_{i,t}(0)$ , l'effet du traitement pour l'unité  $i$  à la date  $t$ , et soit  $D_{i,t}$  une variable indicatrice prenant la valeur 1 quand l'unité  $i$  est exposée au traitement à la date  $t$ , et la valeur 0 sinon. La variable d'intérêt pour l'unité  $i$  à la date  $t$  est:  $Y_{i,t}(0) + \alpha_{i,t}D_{i,t}$ .

L'échantillon considéré est un échantillon de panel, où toutes les régions sont observées aux même périodes temporelles,  $t = 1, \dots, T$ . Il comprend un nombre positif de périodes de pré-traitement,  $T_0$ , ainsi qu'un nombre positif de périodes post-intervention,  $T_1$ . On a ainsi  $T = T_0 + T_1$ . La région traitée est supposée unique et exposée à l'événement (le "traitement") de manière ininterrompue. On suppose que l'événement n'a aucun effet sur la période pré-traitement.

Le but de la méthode de contrôle synthétique est d'estimer  $\alpha_{i,t}$  pour l'unité traitée. Dans la mesure où  $Y_{i,t}(1)$  est observé, il suffit, pour estimer  $\alpha_{i,t}$ , d'estimer  $Y_{i,t}(0)$ , le contrefactuel. Abadie, Hainmueller et Diamond proposent pour ce faire d'utiliser une pondération d'autres unités.

Plus précisément, supposons que nous ayons un échantillon de  $J + 1$  unités parmi lesquelles  $j = 1$  est notre

unité d'intérêt et  $j = 2$  à  $j = J + 1$  les unités de comparaisons.  $J = 1$  devient donc la région traitée, i.e la zone exposée à l'événement "Lille, Capitale Européenne de la Culture", et les unités 2 à  $J + 1$  l'échantillon de contrôle - ou groupe de donneurs pour reprendre la terminologie médicale utilisée par Abadie *et alii*. Dans la mesure où seule la première région est exposée à l'intervention, et ce après la date  $T_0$  (avec  $1 \leq T_0 \leq T_1$ ), on a:

$$D_{i,t} = \begin{cases} 1 & \text{si } i = 1 \text{ et } t \leq T_0 \\ 0 & \text{sinon} \end{cases}$$

Un des éléments essentiels de leur méthode est de considérer que les caractéristiques de la région traitée peuvent être approximées avec davantage de précisions par une combinaison linéaire de régions non soumises au traitement que par une unique région non-traitée. Nous définissons donc la région de contrôle comme une moyenne pondérée des caractéristiques du groupe de donneurs. Ainsi on peut la représenter par un vecteur de poids de taille ( $J \times 1$ ),  $W$ , tel que  $W = (w_2, \dots, w_{J+1})$ , avec  $0 \leq w_j \leq 1$  pour  $j = 2, \dots, J$  et  $w_2 + \dots + w_{J+1} = 1$ . Le choix d'un vecteur  $W$  particulier est équivalent au choix d'une zone de contrôle synthétique particulière. Notons que la composition de la zone de contrôle synthétique est supposée stable au cours du temps.

Ainsi nous cherchons à trouver un vecteur  $W$  tel que les caractéristiques de la zone de contrôle synthétique se rapprochent le plus de celles de la région traitée avant le traitement. Soit  $X_1$  un vecteur de taille ( $k \times 1$ ) contenant les valeurs des caractéristiques pré-intervention de la région traitée. Soit  $X_0$  une matrice ( $k \times J$ ), contenant ces mêmes caractéristiques, cette fois pour toutes les régions du groupe de donneurs.<sup>5</sup>

La différence entre les caractéristiques pré-intervention de la région traitée et celles de la zone de contrôle synthétique est donnée par le vecteur  $X_1 - X_0 W$ . Pour construire la zone de contrôle, il s'agit donc de choisir le vecteur  $W^*$  qui minimise la distance entre  $X_1$  et  $X_0 W$ . Autrement dit, on recherche  $W^*$  qui minimise:

$$\|X_1 - X_0 W\|_v = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)} \quad (1)$$

où  $V$  est définie comme une matrice ( $k \times k$ ) symétrique et semi-définie positive qui permet d'allouer des poids différents à chaque variable selon leur capacité à bien reproduire la région traitée. Ces poids forment alors la diagonale de  $V$ , dont les termes non-diagonaux sont nuls.

Détaillons le choix de  $V$ : Abadie *et alii* (2010) proposent pour choisir  $V$  une méthode fondée uniquement sur les données qui consiste à choisir, parmi toutes les matrices diagonales définies positives, celle dont l'erreur quadratique moyenne de prédiction (MSPE) de la variable d'intérêt est minimisée au sein de l'ensemble des périodes de pré-intervention. Autrement dit, soit  $Z_{jt}$  la variable d'intérêt de la région  $j$  à la date  $t$ . De plus, soit  $Z_1$  un vecteur de taille ( $T_0 \times 1$ ), comprenant les valeurs pré-intervention de la variable d'intérêt pour la région traitée. On a ainsi:  $Z_1 = (Z_{1,1}, \dots, Z_{1,T_0})'$ . De manière similaire, on définit  $Z_0$  la matrice ( $T_0 \times J$ ), contenant les valeurs pré-intervention de la variable d'intérêt pour les régions de la zone de contrôle. Le vecteur  $V$  choisi est alors la solution du problème:

$$\arg \min_{V \in \nu} (Z_1 - Z_0 W^*(V))' (Z_1 - Z_0 W^*(V)) \quad (2)$$

avec  $\nu$  l'ensemble des matrices diagonales définies positives. Nous sommes donc face à une situation d'optimisation imbriquée où l'on cherche à minimiser à la fois sur  $W$  et sur  $V$ .

Il s'agit maintenant de définir l'estimateur. Soit  $Y_1$  un vecteur de taille  $(T_1 \times 1)$ , comprenant les valeurs post-intervention de la variable d'intérêt pour la région traitée. On a ainsi:  $Y_1 = (Y_{1,T_0+1}, \dots, Y_{1,T})'$ . De manière similaire, on définit  $Y_0$  une matrice  $(T_1 \times J)$ , dans laquelle la colonne  $j$  contient les valeurs post-intervention de la variable d'intérêt pour la région  $j+1$ . L'estimateur de contrôle synthétique de l'effet du traitement est alors donné par la comparaison de la variable d'intérêt post-intervention de la région exposée à l'événement d'une part, et de la zone de contrôle synthétique, non-exposée à l'événement d'autre part:  $Y_1 - Y_0 W^*$ . Autrement dit, pour une date post-intervention  $t$  (avec  $t \geq T_0$ ), il est donné à la date  $t$  par l'expression:

$$\hat{\alpha}_{1,t} = Y_{1,t} - \sum_{j=2}^{J+1} w_j^* Y_{j,t}$$

## 2.2 Les avantages du modèle

### 2.2.1 Une protection contre l'extrapolation

L'un des problèmes fréquents dans les méthodes de prédiction et de reconstitution d'un contrefactuel est celui de l'extrapolation. On entend par là le fait de déterminer la relation entre des variables  $X$  et  $Y$  au-delà des valeurs de  $X$  déjà observées dans les échantillons à notre disposition. Cela signifie dans notre cas chercher à construire notre région de contrôle à partir de valeurs non-observées des prédicteurs. De façon plus formelle, il nous faut introduire la notion d'enveloppe convexe des données, soit le plus petit convexe qui entoure l'ensemble des données. L'extrapolation désigne le fait de vouloir atteindre un point en-dehors de ce convexe, point qui, de fait, ne peut être issu d'une simple combinaison affine des observations. Or, afin d'obtenir ce point, il est nécessaire de postuler l'existence d'un modèle étendant l'espace des possibles. Nous renvoyons à King et Zen (2006) pour de plus amples développements mais on peut aisément comprendre qu'un biais d'extrapolation implique une totale dépendance au modèle postulé, ce qui affaiblit la robustesse du contrefactuel.

Par ailleurs ce biais d'extrapolation dans les modèles traditionnels de construction de contrefactuels (telles les régressions) n'est pas facilement repérable ainsi que le soulignent ces auteurs<sup>6</sup>. Face à cela, la méthode de construction d'une zone synthétique de contrôle apparaît véritablement révolutionnaire. En effet, le fait d'utiliser une moyenne pondérée d'unités de comparaison permet d'exclure les problèmes d'extrapolation puisqu'il s'agit de trouver des poids qui sont automatiquement dans l'intervalle  $[0,1]$ . Ainsi on ne sort pas du convexe formé par les observations des autres régions.

### 2.2.2 Une garantie de non-dépendance au modèle

Nous avons déjà évoqué le fait que grâce à une combinaison des données observées, la méthode d'Abadie *et alii* ne fait pas intervenir un modèle particulier dans la construction du contrefactuel, ce qui rend cette méthode plus robuste que le modèle d'intervention utilisé par Brodaty(2008).

Par ailleurs la MCS n'a jamais recours aux variables d'intérêt post-intervention. Elle permet donc aux chercheurs de concevoir leur modèle et de faire les choix optimaux de  $W^*$  et de  $V^*$  sans savoir comment leurs décisions affecteront les conclusions de leurs études. Or Rubin *et alii* (2001) affirment que la capacité de prendre des décisions sur les orientations de la recherche tout en restant aveugle à la manière dont ces décisions particulières affecteront les conclusions de l'étude est un élément fondamental pour promouvoir l'honnêteté intellectuelle dans la recherche.

### **2.2.3 Une explicitation de la contribution relative de chaque unité à la zone de contrôle**

Dans la mesure où la zone de contrôle est une moyenne pondérée des unités de contrôle disponibles, cela permet de rendre explicites: (1) la contribution relative de chaque unité de contrôle au contrefactuel de la région traitée, (2) les similitudes entre la région affectée par l'événement et la zone de contrôle synthétique, en termes de variable d'intérêt pré-intervention et d'autres prédicteurs de variables d'intérêt post-intervention.

La mise en évidence de la contribution relative de chaque unité permet ainsi aux chercheurs d'ajouter dans un deuxième temps une approche qualitative pour analyser les similitudes et les différences entre l'unité traitée et la zone de contrôle mise en valeur par une technique quantitative. Or de nos jours il apparaît nécessaire (entre autres Brady et Collier (2004)) de réunir méthodologie quantitative et méthodologie qualitative dans les Sciences Sociales. L'approche d'Abadie s'inscrit clairement dans cette voie, avec une méthode qui part des données mais cherche toujours à mener une réflexion économique à partir de chiffres trouvés.

### **2.2.4 Une systématisation du choix des unités de comparaison qui rend possible l'inférence**

L'utilisation d'inférence statistique dans les études de cas comparatives est souvent difficile: de fait de la petite taille des échantillons, absence de randomisation (sélection aléatoire des unités observées) et d'échantillonnage probabiliste (issus d'une procédure de sélection) lors du choix des unités de la zone de contrôle. Cependant, Abadie, Diamond et Heinmueller (2015) [1] démontrent que la principale barrière à l'inférence quantitative dans les études comparatives ne tient pas tant à ces limites, qu'à l'absence de mécanisme explicite de sélection des unités de comparaison (cf. King, Keohane et Verba (1994), Card et Krueger (1994) et Rosenbaum (2005)).

En systématisant le processus d'estimation du contrefactuel, la MCS permet aux chercheurs de recourir à une vaste gamme de tests de falsification, que nous appelons des "tests placebo". Ces tests deviennent ainsi la pierre angulaire d'un mode alternatif d'inférence à la fois qualitative et quantitative. Ce modèle d'inférence alternatif est fondé sur l'hypothèse que notre confiance dans le fait que la zone de contrôle reflète bien le contrefactuel serait fortement affaiblie si nous obtenions des effets estimés similaires, voire encore plus importants, dans des cas où le traitement n'avait pas eu lieu.

Lorsque la méthode est appliquée à des dates antérieures au traitement, ces tests de falsification sont nommés "placebos temporels". Ces tests sont faisables si les données disponibles couvrent une période temporelle sans choc sur la variable d'intérêt suffisamment large. Une autre manière de réaliser des tests placébo est de réassigner l'événement, non pas dans le temps, mais à des membres du groupe de donneurs. Nous faisons référence à ces tests sous le terme de "placébos spatiaux". Dans ce cas, l'hypothèse fondamentale est que notre confiance dans le fait que l'estimateur de contrôle synthétique reflète bien l'effet de l'événement disparaîtrait si des effets similaires apparaissaient lorsque l'événement était réassigné artificiellement à des unités du groupe de donneurs n'étant pas exposées à l'événement.

Comme en inférence statistique traditionnelle, une comparaison quantitative entre les effets placébo et la zone de contrôle synthétique estimée peut être menée grâce à l'utilisation de p-values. Dans notre contexte, une p-value peut être construite à partir d'un test estimant les effets placébo spatiaux pour chaque unité de l'échantillon et en calculant ensuite le ratio effets placébo/effet estimé pour la véritable unité traitée. Notons que ce test d'inférence revient, quand l'intervention se fait de manière aléatoire, à une inférence aléatoire classique. En absence de randomisation, la p-value peut s'interpréter comme la probabilité d'obtenir une estimation au moins aussi importante que celle obtenue pour la véritable unité traitée.

## **2.3 Hypothèses fondamentales sous-jacentes au modèle**

Il est important pour comprendre la portée de la méthode de bien mettre en évidence les hypothèses nécessaires à la convergence de l'estimateur. Nous avons donc choisi de les spécifier à part et de leur donner un nom, afin de les rendre plus manipulables.

### **2.3.1 Hypothèse de restriction dans la construction du groupe de donneurs**

Il s'agit à présent de discuter - après le biais d'extrapolation - du biais d'interpolation dans la construction d'un contrefactuel. En effet si la méthode d'Abadie permet de garantir que l'on se situe dans l'enveloppe convexe des observations, elle implique obligatoirement de l'interpolation puisqu'on cherche à construire un point qui n'existe pas auparavant.

Pour pouvoir limiter les biais résultant de cette construction, il est nécessaire de restreindre le groupe de donneurs aux unités dont les caractéristiques sont similaires à celles de l'unité traitée. Les caractéristiques de toutes les unités de la zone de contrôle doivent ainsi permettre de recouvrir celle de l'unité de traitement.

Par ailleurs, dans la mesure où l'on cherche à ce que l'échantillon de contrôle constitue la meilleure approximation possible du contrefactuel de la région d'intérêt, il est important de réduire l'échantillon de contrôle aux régions n'ayant pas subit de chocs structurels idiosyncratiques pendant la période de l'étude.

Enfin, une autre raison pour restreindre la taille du groupe de donneurs et uniquement considérer les unités similaires à l'unité traitée est d'éviter le sur-ajustement. Un sur-ajustement apparaît en effet lorsque les caractéristiques de l'unité affectée par l'événement s'apparent artificiellement par la combinaison de variations idiosyncratique avec les variables n'ayant pas subi l'intervention, ce qui signifierait que notre contrefactuel ne fasse que reproduire en réalité les spécificités propres à chaque régions de contrôle.

### **2.3.2 Hypothèse d'exclusion dans la construction du groupe de donneurs**

Il est primordial que les unités pouvant être affectées par l'événement d'intérêt ou par tout événement de nature similaire soient exclues du groupe de donneurs. Cette hypothèse est résumée par Rosenbaum (2007) sous le nom de "non-interférence entre les unités". Il y a interférence si le traitement que subit l'unité traitée n'affecte pas uniquement cette unité mais également les autres unités de la zone de contrôle, par les liens que l'unité traité peut entretenir avec elles (liens de proximité, liens commerciaux ...). En effet on cherche à isoler l'effet pur de l'événement d'intérêt. Or si une région appartenant à la zone de contrôle a pu être de quelque manière affectée par cet événement, son évolution observée après la date d'intervention sera également influencée par l'événement, ce qui fausse radicalement la reconstitution du contrefactuel de la région traitée.

## **3 Implémentation : Estimation de l'impact économique de "Lille 2004"**

### **3.1 Les données**

Les données que nous utilisons pour implémenter la méthode du contrôle synthétique sont des données de panel annuelles collectées par l'INSEE et portant sur la période 1990 – 2008. Le champ est toujours la France métropolitaine et l'échelle est régionale (du point de vue de l'ancien découpage administratif du territoire métropolitain en 22 régions). L'utilisation d'une échelle régionale dans l'évaluation de l'impact de Lille 2004 peut sembler de prime

abord incongrue et trop large, mais il apparaît dans les faits que la manifestation s'est étendue bien au-delà de zone urbaine lilloise. Des événements ont en effet eu lieu dans tout le Nord-Pas-de-Calais et même jusqu'en Belgique (cf. Figure 13). L'impact de la manifestation a donc été ressenti au niveau régional et peut être évalué à cette échelle.

Notre base comprend 35 variables permettant de décrire partiellement la structure économique d'une région française. Parmi ces 35 variables (voir Annexe Table 3), nous en avons sélectionné 19 pour mener notre étude (voir Table 1).

Table 1: Présentation des variables

Variables	N	Moyenne	Écart type	Min	Max
PIB	528	22,749.670	5,781.633	14,225	53,617
Chômage	528	8.827	1.860	4.500	14.600
Population	911	2,719,978	8,358,399	249,645	11,959,807
Nombre total de création d'entreprises	308	12,380.690	14,143.460	1,917	72,955
- <i>Entreprises industrielles</i>	308	674.058	769.858	106	5,188
- <i>Entreprises hors industrie</i>	528	6,828.871	11,750.020	0	70,107
Nombre total de défaillances d'entreprises	352	2,066.722	2,569.619	187	15,642
- <i>Industries</i>	352	211.915	301.876	7	2,422
- <i>hors industrie</i>	528	1,236.538	2,058.483	0	13,282
- <i>Fermeture de cafés</i>	308	227.513	202.278	26	1,252
Valeur ajoutée totale	443	57,539.880	74,642.210	3,021	501,013
Offres d'emploi	377	128.166	1,339.539	0.810	19,523.000
Aucun diplôme	44	461,356.400	330,094.000	51,583	1,849,977
CEP	44	358,538.600	237,258.800	29,727	1,282,169
BEPC	44	149,359.700	137,532.600	20,342	683,978
CAP-BEP	44	433,115.900	312,494.300	23,452	1,602,121
baccalauréat	44	215,482.200	211,324.500	21,840	1,088,024
diplôme supérieur	44	273,203.000	377,024	14,766	2,180,895

Nous avons ainsi choisi pour chaque région son PIB, ainsi que les valeurs ajoutées, en séparant le domaine industriel du reste, ce qui nous offre un indicateur des dynamiques de production de richesse dans les différents territoires. Dans cette même veine, le revenu disponible des ménages permet de décrire la richesse individuelle caractérisant notre région.

Nous avons souhaité aussi décrire le marché du travail régional à travers des variables comme le taux de chômage régional, les offres d'emploi ou les offres d'emploi durables (ces dernières désignant toutes les offres d'emploi de plus de 6 mois déposées dans les agences locales par les employeurs chaque année) ainsi que le tissu d'entreprises implantées sur le territoire (avec le nombre d'entreprises créées par secteur chaque année, et le nombre d'entreprises ayant

fermé chaque année par secteur). Nous disposons enfin de données sur certaines caractéristiques démographiques des différentes régions (population et niveau de diplôme), ce qui permet ainsi d'ajouter à notre analyse une composante de capital humain de la population.

Parmi ces variables, deux d'entre elles sont nos variables d'intérêt, i.e sur lesquelles nous menons notre étude: le PIB et le chômage.

Au total, nous disposons donc de données portant sur une période pré-intervention de 14 ans (1990 – 2004) et une période post-intervention de 4 ans, ce qui nous permet d'évaluer l'impact non seulement à court terme mais aussi à moyen-terme de l'événement Lille 2004.

En ce qui concerne la période pré-traitement, nous avons pris le parti de considérer que les effets d'annonce de l'événement Lille 2004 sur nos variables d'intérêt étaient négligeables. Dans l'implémentation de la MCS la date d'intervention est donc bien 2004 et non 1999 (date correspondant à la nomination de Lille comme capitale européenne de la culture pour 2004). La résolution du problème d'optimisation décrit en 2.1 dans le cadre de notre étude se fait sur la période 1990 – 2002. Le choix de la borne inférieure de cet intervalle temporel est évident: il correspond à la date à partir de laquelle nos données de panel sont disponibles. Le choix de 2002 comme borne supérieure l'est moins: il est fondé sur le fait que, comme à partir de 2003 l'État accorde des crédits d'investissement exceptionnels, notamment pour la réhabilitation du patrimoine et des opérations d'aménagement, on peut considérer que l'événement débute à cette date.

### 3.2 Une approche en statistiques descriptives

On peut tenter de mettre en valeur l'effet de Lille 2004 sur l'économie du Nord-Pas-de-Calais par une première approche fondée sur l'utilisation de statistiques descriptives.

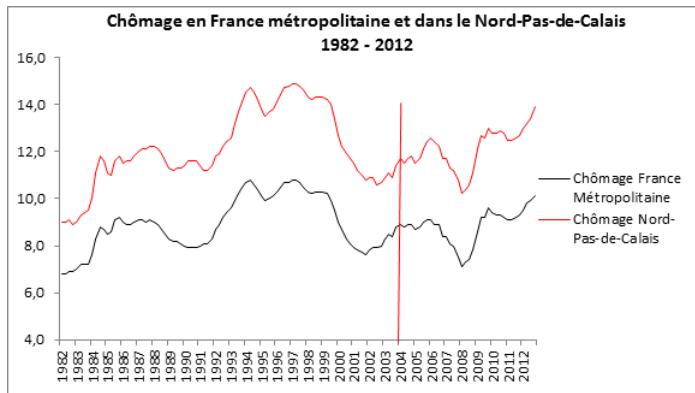


Figure 1: Comparaison de l'évolution du chômage entre 1990 et 2012

Le Nord-Pas-de-Calais est marqué sur toute la période par un taux de chômage largement supérieur à la moyenne de la France métropolitaine. Les variations du taux de chômage du Nord-Pas-de-Calais sont concomitantes à celles de la moyenne nationale mais avec environ deux points de pourcentage d'écart (cf. Figure 1). On pourrait supposer que le taux de chômage du Nord-Pas-de-Calais est structurellement plus élevé que le taux de chômage

métropolitain, ce qui confirmerait l'image du Nord-Pas-de-Calais comme une région ayant fortement souffert de la désindustrialisation. C'est à cette image que "Lille 2004" et les politiques culturelles menées dans cette région voudraient remédier. Mais on remarque entre 2004 et 2006 une forte tendance à la hausse du chômage dans le Nord-Pas-de-Calais, ce qui va à rebours de la tendance nationale qui est à la stabilisation. Pour résoudre ce paradoxe on pourrait émettre l'hypothèse que Lille 2004 a créé un flux migratoire vers le Nord-Pas-de-Calais, augmentant par là sa population active inoccupée et biaisant donc vers le haut son taux de chômage. Une autre hypothèse serait que Lille 2004 ait eu un impact positif sur le bien-être des habitants et ainsi poussé certains demandeurs d'emploi à rester plutôt que de tenter leur chance ailleurs, faisant par là augmenter le nombre de demandeurs d'emploi.

D'autres indicateurs tels que le PIB (cf. Figure 2), la création d'entreprises ou l'emploi (voir Annexe Figure 15)

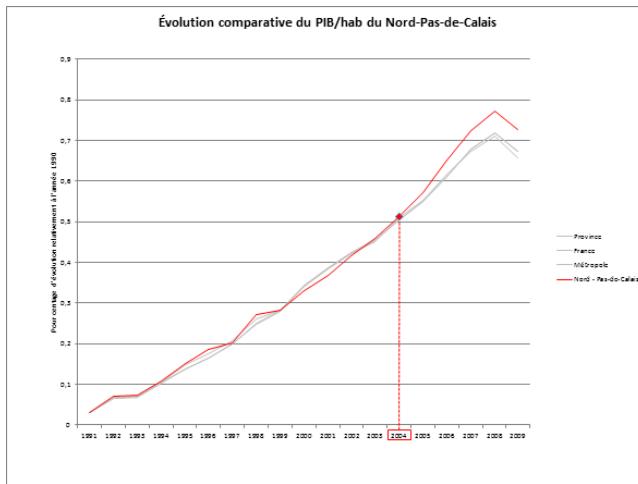


Figure 2: Comparaison de l'évolution du PIB par tête entre 1990 et 2012

semblent au contraire confirmer une amélioration de la santé économique du Nord et du Nord-Pas-de-Calais en 2004. En ce qui concerne l'évolution du PIB par tête, on constate que le Nord suit la même tendance haussière que la France métropolitaine mais de manière plus accentuée, avec un écart qui se creuse en 2004. Le même constat peut être fait sur l'évolution de la création d'entreprises (voir Annexe) où l'évolution à la hausse est plus forte dans le Nord-Pas-de-Calais par rapport à la moyenne nationale à partir de l'année 2004. De là à arguer de l'existence d'une corrélation voire d'une causalité entre Lille 2004 et l'amélioration de la situation retranscrite par ces indicateurs, il y a un pas que nous ne saurions encore franchir. On peut cependant supposer que Lille 2004 a joué un rôle – partiellement retranscrit par ces statistiques descriptives – dont l'évaluation précise est l'enjeu principal de ce mémoire.

### 3.3 La construction du groupe de donneurs

Le choix du groupe de donneurs (i.e des régions à partir desquelles on va construire le Nord-Pas-de-Calais synthétique) est fondé sur la pertinence de différentes hypothèses (cf. 2.3).

Premièrement, il faut s'assurer que le Nord-Pas-de-Calais ne prend pas des valeurs extrêmes par rapport aux

autres régions pendant la période de pré-intervention (cf. l'hypothèse de restriction) afin de lutter contre les biais d'interpolation. Les graphiques (présentés en Annexe Figure 16 et Figure 17) représentant la distribution de chaque variable pour chaque région nous permettent donc d'éliminer certaines variables pour lesquelles le Nord-Pas-de-Calais prend des valeurs extrêmes. Ainsi le revenu disponible des ménages dans le Nord-Pas-de-Calais (moyenne régionale des revenus disponibles annuels en euros) est largement inférieur à celui de toutes les autres régions. Cette variable ne peut donc pas être prise en compte dans la construction du Nord-Pas-de-Calais synthétique. Pour toutes les autres variables, le Nord-Pas-de-Calais est toujours encadré par deux régions et est donc à ce titre reproductible par une combinaison affine des régions de France métropolitaine.

Deuxièmement, l'hypothèse de non interférence entre les régions du groupe de donneur et la région traitée – qui est canonique dans les modèles utilisant un contrefactuel – doit être vérifiée (cf. hypothèse d'exclusion).

Pour déterminer la présence ou l'absence d'un impact économique notable de Lille 2004 sur d'autres régions que le Nord-Pas-de-Calais, nous avons analysé les échanges commerciaux interrégionaux. Les flux commerciaux entrants et sortants ainsi que les données sur le transport commercial de marchandises présentées en annexe (Figure 18 et Figure ??) nous ont conduit à éliminer les régions dont la contribution au commerce interrégional du Nord-Pas-de-Calais était supérieure à 5%. Ces régions sont l'Île-de-France, Champagne-Ardenne, la Picardie, le Centre et la Lorraine. On constate donc sans surprise que ce sont globalement les régions frontalières et les plus proches géographiquement du Nord-Pas-de-Calais qui sont retirées du groupe de donneurs.

### 3.4 Les résultats : impact de *Lille 2004* sur le PIB

Poids	Régions		Réel	Synthétique	Moyenne
0.000	Aquitaine	PIB	17125.692	17135.701	19103.790
0.000	Auvergne	Chômage	12.385	12.088	8.614
0.000	Basse-Normandie	VA totale	59270.308	33626.354	40420.077
0.000	Bourgogne	Population	3986750.462	2223140.511	2372136.103
0.000	Bretagne	Création total - Industrie	9752.300	13247.476	10284.653
0.000	Centre	Création Industrie	605.200	656.398	633.140
0.000	Franche-Comté	Défaillance Industrie	300.200	184.536	194.033
0.775	Languedoc-Roussillon	Défaillance : total - Industrie	2106.000	1941.395	1624.253
0.000	Limousin	Offre d'emplois	13.477	8.596	9.108
0.000	Midi-Pyrénées	Diplôme : aucun	755857.000	456474.945	409721.533
0.077	Pays de la Loire	Diplôme : BEPC	212074.500	157699.630	131475.833
0.148	Poitou-Charentes	Diplôme: CEP	563588.500	320522.437	329304.067
0.000	PACA	Diplôme: CAP, BEP	622105.000	350824.144	388108.200
0.000	Rhône-Alpes	Diplôme: baccalaureat	268481.500	200570.760	186505.500
		Diplôme: supérieur	298480.500	227540.186	214782.633

Figure 3: Régions de contrôle

Figure 4: Comparaison des constructions de contrefactuel

Nous présentons dans un premier temps les résultats liés à la variable d'intérêt du PIB. Nous avons donc sélectionné 16 régions pour le groupe de donneurs et 15 prédicteurs (cf. annexe 2 pour la l'ensemble des prédicteurs et leur pondération). La première étape qui consiste à trouver les coefficients de pondération associés à chaque variables (ceux de la diagonale de  $V$ ) nous indique que les prédicteurs les plus importants sont dans l'ordre : le PIB

(0.329), le chômage (0.259), le nombre d'entreprises créées (0.187 pour les entreprises industrielles et 0.189 pour les autres) et la proportion de personnes non diplômées (0.019). La figure 3 quant à elle montre le poids de chacune des régions de contrôle, autrement dit la zone de contrôle. On remarque que trois régions permettent de reproduire le Nord-Pas-de-Calais : principalement le Languedoc-Roussillon et le Poitou-Charentes et de façon secondaire le Pays de la Loire.

La figure 4 permet de comparer deux manières de construire le contrefactuel : notre méthode et une simple moyenne de l'échantillon. On remarque que sur les variables ayant le plus fort pouvoir de prédiction (chômage et PIB), la zone synthétique est plus proche du Nord-Pas-de-Calais réel. Il en est de même pour la variable qui mesure la part de personnes non-diplômées dans la population. Cependant la zone synthétique peine à approcher les créations d'entreprises du Nord-Pas-de-Calais. Ainsi cet appariement démontre qu'il existe bien une combinaison de régions françaises qui tend à reproduire les caractéristiques économiques du Nord-Pas-de-Calais avant l'événement culturel. Il est donc pertinent d'utiliser cette zone de contrôle pour appréhender l'impact de *Lille 2004*.

La figure 5 représente la trajectoire du PIB par tête du Nord-Pas-de-Calais et celle de son contrefactuel synthétique pour la période 1990-2008. Le Nord-Pas-de-Calais synthétique reproduit globalement le PIB par tête du Nord-Pas-de-Calais sur la période entière de pré-intervention.

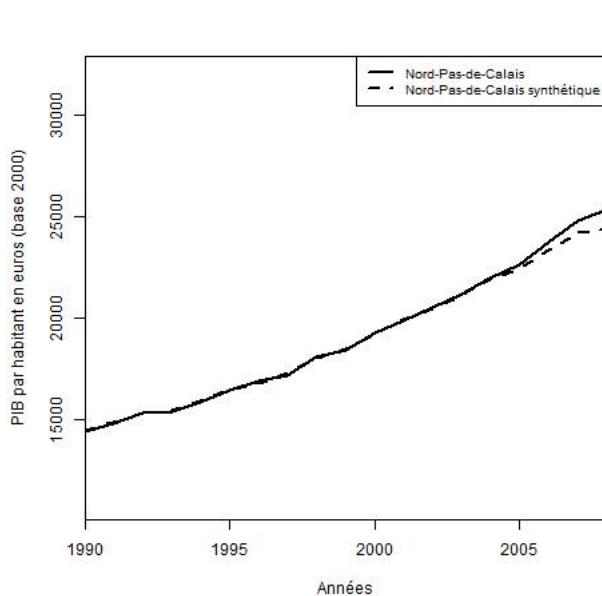


Figure 5: Évolution comparée du PIB par tête

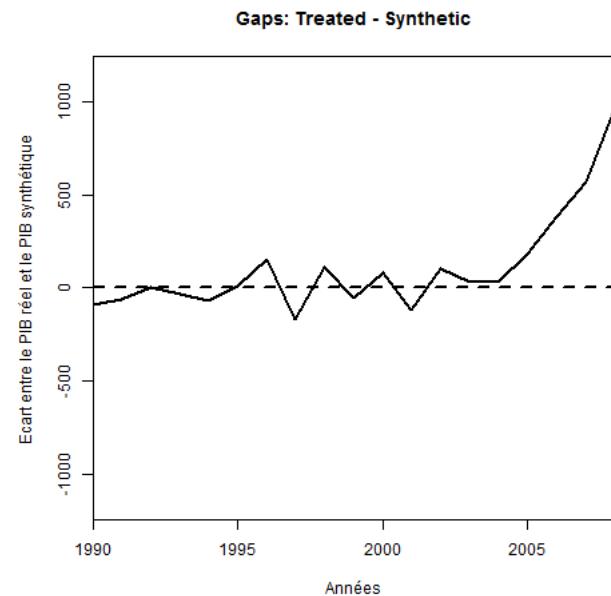


Figure 6: Écart entre PIB réel et PIB synthétique

Notre estimation de l'effet de l'événement "Lille, capitale Européenne de la Culture" est donnée par la différence entre le Nord-Pas-de-Calais réel et son contrefactuel, i.e l'écart entre le Nord-Pas-de-Calais actuel et le Nord-Pas-de-Calais tel qu'il aurait été sans l'événement culturel. Cette différence est perceptible sur la figure 6.

Nous estimons que l'événement culturel a eu un effet immédiat sur le PIB/tête du Nord-Pas-de-Calais, même si

cet effet reste assez faible l'année de l'événement. A partir de 2005, les deux trajectoires diffèrent substantiellement. Tandis que la croissance du PIB/tête s'accélère au Nord-Pas-de-Calais, le PIB/tête pour le Nord-Pas-de-Calais synthétique continue d'évoluer à une vitesse relativement similaire à celle de la période pré-événement. La différence entre les deux séries continue d'augmenter jusqu'à la fin de la période. Ainsi, nos résultats suggèrent un effet positif de l'événement "Lille, capitale Européenne de la Culture" sur le PIB de la région Nord-Pas-de-Calais.

Nous avons différencié nos résultats en résultats de court terme (2004) et résultats de moyen terme (2008). Passée cette date, nous estimons qu'interpréter l'écart entre le PIB réel et le PIB synthétique comme une conséquence de l'événement culturel s'avère incertain. Nous trouvons qu'en 2004, le PIB/tête du Nord-Pas-de-Calais est supérieur de 0,15% au PIB du Nord-Pas-de-Calais synthétique, et en 2008 de 3,9%. Sur la période post-intervention considérée, l'événement "Lille, Capitale Européenne de la Culture" aura ainsi permis au PIB d'augmenter en moyenne de 1,7% par rapport à ce qu'il aurait été sans l'événement.

### 3.5 Les résultats : impact de *Lille 2004* sur le chômage

Nous nous sommes également intéressés à l'impact de cet événement sur un indicateur décrivant le marché du travail: le chômage. En effet, s'il est vrai qu'on observe une augmentation du nombre d'emploi à court terme (cf. Figure 15) du fait du nombre de contrats qui ont été signés avec des intermittents du spectacle ou des intérimaires lors des manifestations culturelles, ce sont beaucoup d'emplois à durée déterminée. Il est donc tout à fait pertinent de s'interroger sur les retombées en termes de diminution du chômage. Néanmoins, compte tenu des statistiques descriptives (cf. Figure 1), nous nous attendons à observer plutôt une augmentation du chômage.

Pour ce faire nous reprenons la même base régionale que précédemment, en changeant de variable d'intérêt et donc de coefficients de pondération puisque que les poids  $v$  sont définies à partir de cette même variable. Cependant pour reproduire au mieux le chômage nous avons décidé de l'utiliser en variation vis-à-vis de 1990. Or le Nord-Pas-de-Calais ayant des valeurs en-dehors du convexe des autres données (voir Figure 16 en annexe) sur la période 1990-1993, nous faisons donc porter l'optimisation uniquement sur la période 1994 - 2002, soit huit années.

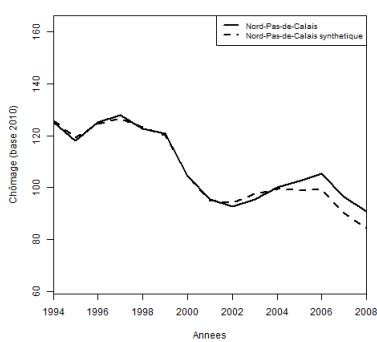


Figure 7: Évolution comparée du chômage

	Réel	Synthétique	Moyenne
PIB	17125.692	17688.132	19103.790
Chômage	111.574	112.813	114.695
VA totale	59270.308	40469.142	40420.077
Population	3986750.462	2573635.449	2372136.103
Création total - Industrie	9752.300	13330.275	10284.653
Création Industrie	605.200	701.585	633.140
Défaillance Industrie	300.200	197.242	194.033
Défaillance : total - Industrie	2106.000	1874.668	1624.253
Offre d'emplois	13.477	9.900	9.108
Diplôme : aucun	755857.000	560021.055	409721.533
Diplôme : BEPC	212074.500	206135.661	131475.833
Diplôme: CEP	563588.500	401087.574	329304.067
Diplôme: CAP, BEP	622105.000	453982.476	388108.200
Diplôme: baccalauréat	268481.500	267415.084	186505.500
Diplôme: supérieur	298480.500	307551.299	214782.633

Figure 8: Comparaison des contrefactuels du chômage

Cette fois-ci le prédicteur le plus important est la proportion de personnes ayant leur baccalauréat (0,405), suivie de la proportion de personnes ayant un diplôme du supérieur (0,184), et du nombre d'entreprises créées dans le secteur non industriel (0,074) (voir annexe pour l'ensemble des prédicteurs). La reproduction du Nord-Pas-de-Calais se fait à partir de quatre régions: Au Languedoc-Roussillon, Poitou-Charentes et Pays de Loire que l'on trouvait déjà pour le PIB s'ajoute la Bretagne. La figure 7 nous montre que le chômage augmente bien plus qu'il n'aurait dû à partir de 2004. Les raisons de cette augmentation qui va à l'encontre des créations d'emplois consécutives à l'évènement, sont celles déjà mentionnées dans la partie 3.2. Nous trouvons qu'en 2004, le chômage du Nord-Pas-de-Calais est supérieur de 0,5% au chômage du Nord-Pas-de-Calais synthétique, et en 2008 de 7,8%. Sur la période post-intervention considérée, l'évènement "Lille, Capitale Européenne de la Culture" aura ainsi contribuer à augmenter le chômage du Nord-Pas-de-Calais de 5,08% par rapport à ce qu'il aurait été sans l'évènement.

### **3.6 Limites d'une étude sur le tourisme à l'échelle départementale**

Afin d'avoir un aperçu des impacts micro-économiques d'un événement culturel de si grande ampleur, nous avions souhaité raisonner à l'échelle départementale sur les nuitées d'hôtel qui est indicateur lié aux retombées touristiques de la manifestation. Pour ce faire, nous avions réuni 11 variables: la population, le revenu net imposable par foyer fiscal, le taux de chômage, le nombre d'entreprises créées dans l'année par département, le nombre d'employés salariés de l'industrie dans le secteur de l'hébergement et restauration et dans le secteur du tertiaire marchand ainsi que le nombre d'offres d'emploi durables émises dans l'année.

Or nous avons rencontré plusieurs limites intrinsèques à nos données qui empêchaient d'appliquer avec pertinence la méthode de contrôle synthétique. Premièrement l'indicateur des nuitées n'était disponibles sans rupture qu'à partir de 1999. Or, comme nous devons utiliser la variable d'intérêt pour résoudre la minimisation sur  $V$ , cela reviendrait à n'avoir comme période d'optimisation que 1999-2002, soit quatre ans. La profondeur temporelle est donc très limitée. Deuxièmement, il se trouve que sur deux prédicteurs, le Nord est complètement en-dehors du convexe et que pour quatre autres, il n'y a que l'Île-de-France qui se situe au-dessus (et très largement au-dessus), ce qui pose un réel problème d'interpolation. Enfin la recherche d'un groupe de donneurs qui respecte l'hypothèse de non-interférence entre unités est rendue bien plus difficile car il faudrait une théorie sur les destinations touristiques en France. En effet il s'agit de déterminer quels sont les départements qui n'auraient pas été choisis par les touristes si l'événement n'avait pas eu lieu. Si l'on incluait dans notre groupe de donneurs des destinations de substitution au Nord, l'impact de la manifestation serait sur-estimé (car en l'absence de Lille 2004, les touristes de Lille 2004 seraient allés dans la zone de substitution).

## **4 La robustesse des résultats : inférence sur l'effet de "Lille 2004"**

Un des apports primordiaux de la méthode appliquée est, comme énoncé dans la partie 2.2, la possibilité de faire de la statistique inférentielle sur nos résultats. Nous distinguons deux types de tests, ceux portant sur la fiabilité des pondérations trouvées que nous appelons tests de robustesse, et ceux portant sur la confiance que nous pouvons avoir dans l'estimateur que nous appelons tests de falsification ou placebos et qui portent sur la fiabilité du modèle en entier. Ici ne seront représentés que les résultats concernant le modèle du PIB à l'échelle régionale.

## 4.1 Les tests de robustesse

### -Face à la profondeur temporelle

Nom de la région	Pondération
Alsace	0.000
Aquitaine	0.000
Auvergne	0.002
Basse-Normandie	0.008
Bourgogne	0.002
Bretagne	0.000
Centre	0.001
Franche-Comté	0.001
Languedoc-Roussillon	0.846
Limousin	0.001
Midi-Pyrénées	0.000
Pays de la Loire	0.131
Poitou-Charentes	0.008
PACA	0.000
Rhône-Alpes	0.000

Figure 9: Zone de contrôle

Nous voulons tester si notre modèle est sensible à la taille de notre période d'optimisation. En optimisant l'appariement pré-intervention des variables sur une période plus courte (1995-2002 au lieu de 1990-2002), nous aimerais retrouver des résultats similaires aux résultats issus de l'optimisation sur la période totale de pré-intervention. En effet, si les régions de la zone de contrôle précédemment trouvées venaient à changer, cela remettrait en question la robustesse de notre modèle. En optimisant l'appariement sur la période 1995-2002, nous retrouvons les mêmes régions de contrôle, avec des poids quasiment identiques.

### -Face au calcul des prédicteurs

Nous effectuons ensuite un test que nous nommons "sensibilité à la moyenne". Nous nous inspirons pour cela d'un commentaire de Kaul, Klössner, Pfeifer et Schieler (2016) [19] à la méthode d'Abadie. Kaul, Klössner, Pfeifer et Schieler montrent en effet que les variables d'intérêt ne doivent pas être intégrées telles quelles au reste des prédicteurs, sous peine de rendre ces derniers non-pertinents.

Ils préconisent d'appliquer deux fois la méthode de contrôle synthétique: une fois en intégrant aux prédicteurs, sous forme de prédicteur additionnel, la moyenne de la variable d'intérêt sur la période pré-intervention (ce que nous faisons dans la méthode que nous appliquons); une autre fois en intégrant la dernière valeur pré-intervention de la variable d'intérêt. Si le vecteur de poids  $W$  estimé est identique ou presque dans les deux cas, alors le choix d'intégrer la variable d'intérêt aux prédicteurs ne pose plus problème et la robustesse du modèle est assurée. Nous avons donc implémentée une nouvelle fois la méthode, en intégrant dans les prédicteurs uniquement la valeur du PIB en 2003, et non plus sa moyenne sur la période 1990-2003. Nous trouvons exactement les même régions de contrôle avec exactement les mêmes poids, ainsi que des MSPE identiques.

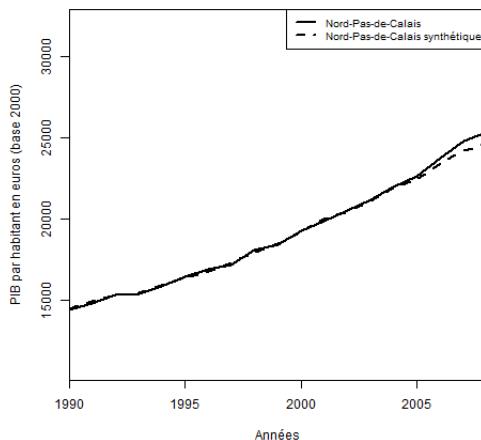


Figure 10: Test de sensibilité à la moyenne

### -Face au nombre de régions

Le test suivant est un test de robustesse qui consiste à retirer de l'analyse une région ayant un poids positif dans la reconstitution du Nord-Pas-de-Calais synthétique. Par ce retrait, on perd certes en précision dans la prédition du modèle, mais on est alors aussi capable de déterminer à quel point nos résultats sont influencés par la présence d'une région ou d'une autre.

Nous retirons de ce test que le modèle est très sensible à la présence du Languedoc-Roussillon dans le groupe de donneur. En effet, lorsque l'on supprime de l'analyse cette région, le modèle perd tout caractère prédictif et ne retient que le Poitou-Charentes pour créer la région de contrôle synthétique.

En revanche, lorsque que l'on retire la région Poitou-Charentes, notre région synthétique est cette fois-ci uniquement constituée d'une combinaison de Languedoc-Roussillon et du Pays de la Loire. Lorsque l'on retire le Pays de la Loire, la combinaison est formée d'une dominante Languedoc-Roussillon, de la région Poitou-Charentes mais aussi de la Bretagne avec un très faible pondération. Le modèle est donc finalement bien plus sensible à la région Languedoc-Roussillon à laquelle est en effet affecté le poids le plus important.

## 4.2 Les tests de falsification

Pour s'interroger sur la significativité de notre estimateur de l'effet *Lille 2004*, nous cherchons à réallouer l'événement à une autre date -placebo temporel - ou à une autre région -placebo spatial- et à voir si l'on détecte tout de même une divergence alors qu'a priori aucun événement n'a eu lieu. Si l'estimateur est élevé lorsque nous réaffectons l'événement cela indique que l'effet mis en lumière est plus lié à un manque de pouvoir de prédition que représentatif d'un impact de *Lille 2004*. Pour comparer l'estimateur de notre modèle face à celui du modèle réaffecté, nous utilisons la mesure du MSPE ("mean square prediction error") qui correspond à la différence pour une région entre la trajectoire réelle du PIB et son contrefactuel ( $MSPE = \frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{1,t} - \sum_{j=2}^{J+1} w_j^* Y_{j,t})^2$ ).

### -Placebo dans le temps

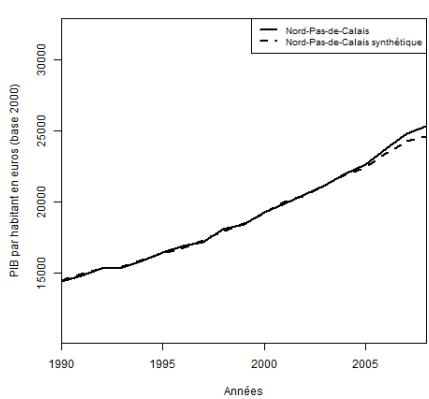


Figure 11: Placebo dans le temps

L'événement culturel a eu lieu en 2004, mais nous avons des données depuis 1990, ce qui nous permet de réaliser des placebos temporels, i.e de tester si la méthode montre l'existence d'effets importants lorsqu'elle est appliquée à des dates précédant l'événement culturel. Si nous trouvons que ces effets estimés sont identiques, voire plus grands que ceux estimés depuis l'événement culturel, notre confiance dans le fait que les effets estimés depuis l'événement peuvent être attribués à celui-ci diminue fortement. Dans un tel cas, les tests placebo suggèreraient que le contrôle synthétique ne fournit pas de bonnes prédictions pour la trajectoire du PIB du Nord Pas de Calais dans des périodes où l'événement culturel n'a pas eu lieu. Au contraire, en assignant artificiellement l'événement en 1997 nous trouvons qu'il n'y pas d'effet notable avant 2004. Le Nord-Pas-de-Calais synthétique reproduit parfaitement l'évolution du PIB réel sur la période

1990-1997, mais, plus important, les trajectoires du PIB réel et du PIB synthétique ne divergent pas considérablement sur la période 1997-2004. Autrement dit, contrairement au véritable événement, notre placebo-1997 n'a pas d'effet notable.

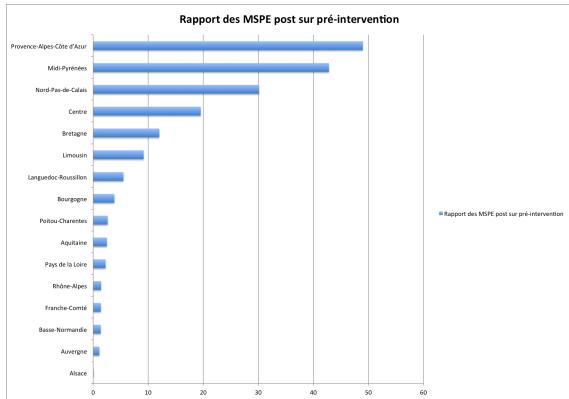


Figure 12: Placebo dans l'espace

Un MSPE post-intervention élevé signifie que l'écart entre la valeur prédite et la valeur réelle est important et semble donc indiquer un véritable effet de l'intervention sur la variable d'intérêt. La lecture du MSPE post-intervention doit cependant se faire à la lumière du MSPE pré-intervention, qui permet d'indiquer si le modèle reproduit bien la région concernée avant l'intervention. Nous comparons donc le rapport MSPE post-intervention/MSPE pré-intervention du Nord-Pas-de-Calais avec celui des autres régions. On remarque alors (cf Figure 12) que le Nord-Pas-de-Calais a un des ratios de MSPE les plus élevés, derrière les régions Midi-Pyrénées et Provence-Alpes-Côte d'Azur.

Il apparaît que les MSPE pré-intervention des régions Midi-Pyrénées et Provence-Alpes-Côte d'Azur sont relativement faibles, indiquant ainsi des modèles assez précis pour ces deux régions. Une interprétation possible serait alors qu'un événement susceptible d'influencer le PIB régional a eu lieu en 2004 dans ces deux régions. À des fins de vérification, nous avons aussi mené ce test de falsification spatial sur une période d'optimisation différente (1990 - 1997) et nous avons trouvé sensiblement les mêmes résultats (les régions Provence-Alpes-Côte d'Azur et Midi-Pyrénées étant sensiblement devant le Nord-Pas-de-Calais, voir figure en annexe). Dans les deux cas, on trouve une p-value importante de  $2/16 = 12,5\%$  ce qui pose ultimement question sur la significativité de notre modèle pour le Nord-Pas-de-Calais.

#### -Placebo dans l'espace

Pour mener un test de falsification spatial, on fait les calculs du modèle en considérant que la région traitée est non pas le Nord-Pas-de-Calais mais une des régions du groupe de donneurs. Ainsi on peut obtenir des estimateurs de la zone de contrôle synthétique pour des régions qui n'ont pas reçu le traitement qu'est Lille 2004.

On peut alors comparer l'estimation de l'impact de Lille 2004 fournie par le modèle avec la distribution des effets placebos obtenus pour les autres régions. Cette comparaison se fait via les MSPE. On prend soin de distinguer le MSPE global des MSPE calculés sur la période pré et post-intervention.

## 5 Conclusion

L'enjeu de notre mémoire était d'évaluer l'impact économique d'une politique culturelle de grande échelle spécifique. Pour ce faire, il nous a fallu adapter une méthode économétrique récente qui a eu de nombreuses répercussions dans le champ académique. Cette méthode a trouvé des applications dans des domaines divers (de l'évaluation des coûts d'un conflit à ceux d'une catastrophe naturelle en passant par l'effet du terrorisme sur le vote), La MCS appliquée au cas de Lille 2004, capitale européenne de la culture permet donc de conclure à un effet sur le PIB du Nord-Pas-de-Calais qui est à bien des égards significatif (+1,7% de PIB par an en moyenne sur la période 2004-2008). En ce qui concerne le chômage, l'interprétation des résultats est plus difficile et nous réduit à émettre des hypothèses. Mais elle est aussi soumises à des critiques et à des correctifs et on ne peut à cet égard s'empêcher de souligner les limites inhérentes à tout modèle basé sur l'évaluation d'un contrefactuel, qui est autant robuste que la plus faible de ses hypothèses. Ando et Savje dans un article de 2015 soulignent les limites de la méthode du contrôle synthétique lorsque le traitement n'est pas un événement ponctuel mais continu ou que les variables de traitement sont multiples. Or ces deux derniers points sont caractéristiques des politiques culturelles visant à changer durablement le visage d'un territoire, comme c'était le cas dans le Nord-Pas-de-Calais au milieu des années 2000. Lille 2004 a en effet été le point de départ d'une politique culturelle ambitieuse et durable dont l'évaluation des effets dans la durée reste à faire.

## Notes

<sup>1</sup>Ils se focalisent uniquement sur un événement, celui de Salamanque en 2002, et concluent qu'il a rapporté : 701,587,097€au niveau national par rapport à un investissement de 399,777,760 €

<sup>2</sup>Si tel est le cas, ces études d'impact dites in-itinere surestiment l'impact réel de l'événement étudié.

<sup>3</sup>Plus précisément selon lui La manifestation est responsable de 8,9% d'offres d'emplois supplémentaires en 2005 et 2006, de 6,7% de créations d'entreprises supplémentaires en 2005 et 2006 (principalement hors industrie et dans le Pas-de-Calais) et de huit défaillances d'entreprises en moins par trimestre (32,7%) dans l'industrie (principalement dans le Nord). L'impact sur les demandeurs d'emploi, les salaires et les heures travaillées a été neutre.

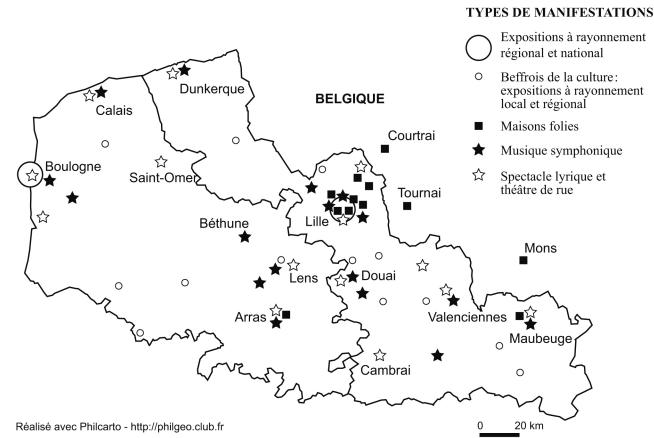
<sup>4</sup>On peut penser aux articles de Nannicini et Billmeier (2011) [23] ou Billmeier and Nannicini (2013) [7] (sur la croissance économique), Coffman et Noy (2012) [10] ou Cavallo et al. (2013) [9] (sur les catastrophes naturelles), Hinrichs [15](2012) (sur l'activité légale), Hosny [16](2012) (sur le libre-échange), Acemoglu et al. (2015) (les connections politiques), Jinjarkar et al. [18](2013), (entrée de capitaux), Kleven et al. (2013) (imposition des athlètes), Bauhoff [4](2014) (l'alimentation scolaire), Belot et Vandenberghe [5] (2014) (niveau d'instruction), Bohn et al. [8](2014), Gobillon et Magnac (2015) [12] (zones d'entreprise), Liu (2015) (les retombées de l'université), Bilgel et Galle (2015) [6](le don d'organes), Kreif et al. [20] (2015) (la mortalité à l'hôpital), et Stearns (2015) [25] (le congé maternité).

<sup>5</sup>Les caractéristiques pré-intervention de  $X_1$  et de  $X_0$  peuvent inclure la variable d'intérêt de l'étude. Attention tout de même à la manière d'intégrer la variable d'intérêt. Kaul, Klössner, Pfeifer et Schieler (2016) montrent qu'elles ne doivent pas être intégrées telles quelles au reste des prédicteurs, c'est-à-dire en tenant compte de toutes les dates, sous peine de rendre ces derniers non-pertinents. En effet le fait de considérer la variable d'intérêt à toutes les dates pré-intervention comme variable caractérisant les régions pré-traitement porte automatiquement la contribution des autres à zéro.

<sup>6</sup>"Identifying the hull with even a few explanatory variables can take an extraordinary amount of computational power. Doing it with more than about 10 variables appears nearly impossible. Moreover, the problem of locating whether a counterfactual point lies within or outside the hull is itself a difficult computational problem that also has no solution known in the statistical literature" p.190

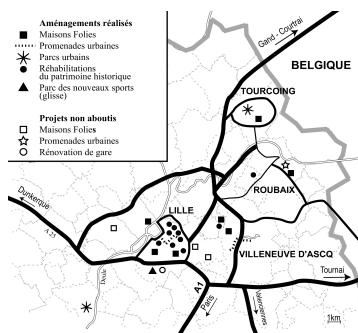
## 6 Annexes

### 6.1 Le contenu de *Lille 2004*



*Conception/réalisation: C. Liefooghe/J. Domont, TVES, Lille 1, 2009. D'après Lemaire C., 2004, Lille 2004 dans les villes moyennes du Nord-Pas-de-Calais, Master 1, Grégoris M.-T. (dir.), Lille 1*

Figure 13: Les réalisations dans le Nord-Pas-de-Calais de *Lille 2004*



*Conception/réalisation: C. Liefooghe / J. Domont, TVES, Lille 1, 2009*

Figure 14: Les réalisations à Lille de *Lille 2004*

A travers ces figures produites par Lefooghe (2010) [21], nous voyons que Lille 2004 est un événement qui, en plus de l'organisation de spectacles vivants et d'expositions - permettant ainsi de créer des emplois dans le tertiaire - s'est aussi accompagné d'investissements dans l'aménagement urbain. L'État et les collectivités locales ont fait un effort budgétaire pour améliorer les espaces publics (« promenades urbaines », parcs) et l'équipement culturel (opéra, musées), pour réhabiliter le patrimoine monumental (fortifications, églises, palais) ou faire de la régénération urbaine (Lille Sud, Maisons Folies).

	France	Nord Pas-de-Calais	Lille intramuros
<b>Activités culturelles</b>	+ 0,4 %	+ 4 %	+ 22 %
<b>Commerce de détail</b>	+ 1,2 %	+ 1,1 %	+ 3 %
<b>Hôtellerie</b>	+ 0,5 %	stable	+ 15 %
<b>Restauration - Débits boissons</b>	+ 1,4 %	+ 3 %	+ 7 %

Source: INSEE - Lille Horizon 2004, impact de Lille 2004 sur l'économie du tourisme

Figure 15: Impacts de *Lille 2004*: évolution de l'emploi entre octobre 2003 et septembre 2004

## 6.2 Les résultats sur le modèle du PIB

Prédicteurs	Poids
PIB	0.29
Chômage	0.259
VA totale	0
Population	0
Création total - Industrie	0.189
Création Industrie	0.187
Défaillance Industrie	0.002
Défaillance : total - Industrie	0.001
Offre d'emplois	0
Diplôme : aucun	0.019
Diplôme : BEPC	0.005
Diplôme: CEP	0
Diplôme: CAP, BEP	0.008
Diplôme: baccalauréat	0.002
Diplôme: supérieur	0

Table 2: Pondération des prédicteurs

Table 3: Présentation des variables

Variables	N	Moyenne	Écart type	Min	Max
PIB	528	22,749.670	5,781.633	14,225	53,617
Chômage	528	8.827	1.860	4.500	14.600
Total de création d'entreprises	308	12,380.690	14,143.460	1,917	72,955
-cafés	294	1,412.762	1,189.800	284	5,504
-entreprises industrielles	308	674.058	769.858	106	5,188
-entreprises hors industrie	528	6,828.871	11,750.020	0	70,107
Total de défaillances d'entreprises	352	2,066.722	2,569.619	187	15,642
-industries	352	211.915	301.876	7	2,422
-hors industrie	528	1,236.538	2,058.483	0	13,282
-cafés	308	227.513	202.278	26	1,252
Valeur ajoutée totale	443	57,539.880	74,642.210	3,021	501,013
-agriculture	443	95.017	15.000	27.360	137.620
-industrie biens agricoles	443	88.691	12.941	35.000	119.390
-industrie biens de consommation	443	100.825	19.320	33.430	153.220
-industrie automobile	443	104.371	39.642	0.000	536.840
-industrie biens d'équipement	443	82.325	16.286	17.610	149.380
-industrie biens intermédiaires	443	87.228	13.202	18.650	121.040
-énergie	443	85.326	15.725	18.170	139.680
-constructions	443	69.560	19.297	30.160	113.950
-commerce	443	81.302	15.729	26.740	106.240
-transport	443	76.666	17.817	27.650	108.420
-finance.immo	443	71.248	18.625	25.900	107.930
-service aux entreprises	443	70.914	19.824	22.900	106.710
-service aux particuliers	443	72.967	19.492	26.940	109.490
-social	443	76.480	17.576	34.240	110.250
Revenu disponible des ménages	440	16,300.710	2,803.427	10,493	24,201
Offres d'emploi	377	128.166	1,339.539	0.810	19,523.000
Population	911	3,916,178.000	8,358,399.000	249,645	62,134,866
Aucun diplôme	44	461,356.400	330,094.000	51,583	1,849,977
CEP	44	358,538.600	237,258.800	29,727	1,282,169
BEPC	44	149,359.700	137,532.600	20,342	683,978
CAP-BEP	44	433,115.900	312,494.300	23,452	1,602,121
baccalauréat	44	215,482.200	211,324.500	21,840	1,088,024
diplôme supérieur	44	273,203.000	377,024.000	14,766	2,180,895

### 6.3 Le modèle et ses propriétés économétriques

Pour justifier la pertinence de la méthode de contrôle synthétique, nous empruntons à Abadie *et alii* leur modèle mathématique qui illustre la convergence de l'estimateur. Nous reprenons la terminologie utilisée au cours de notre mémoire.

Supposons pour ce faire que  $Y_{i,t}(0)$  est donné par un modèle à facteurs communs:

$$Y_{i,t}(0) = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{i,t} \quad (3)$$

où  $\delta_t$  est un paramètre non-observable correspondant à la date  $t$ ,  $Z_i$  un vecteur ( $r \times 1$ ) contenant les covariables observées (non affectées par l'intervention),  $\theta_t$  le vecteur ( $1 \times r$ ) des paramètres inconnus correspondants,  $\lambda_t$  un vecteur ( $1 \times F$ ) contenant les variables inobservées communes associées à la date  $t$ <sup>7</sup>,  $\mu_i$  le vecteur ( $F \times 1$ ) contenant les variables inobservées associées à chaque région  $i$  et  $\epsilon_{i,t}$  les termes d'erreur représentant le choc temporaire spécifique à une unité et une date données, supposés iid et de moyenne nulle.

Soit  $W$  un vecteur ( $J \times 1$ ) de poids  $= (w_2, \dots, w_{J+1})$ , tel que  $w_j \leq 0$  pour  $j = 2, \dots, J+1$  et  $w_2 + \dots + w_{J+1} = 1$ . Chaque vecteur  $W$  particulier représente donc une potentielle zone de contrôle.

Supposons qu'il existe un vecteur  $W^*$  tel que :

$$Z_1 = \sum_{j=2}^{J+1} w_j^* Z_j \quad \text{et} \quad \mu_1 = \sum_{j=2}^{J+1} w_j^* \mu_j \quad (4)$$

On dit qu'une zone de contrôle synthétique (parfaite) alors existe puisque la somme pondérée par  $W^*$  des unités de contrôle reproduit parfaitement  $Y_{1,t}(0)$  jusqu'au choc transitoire. En d'autres termes, l'utilisation de  $W^*$  produit un estimateur non-biaisé de la variable d'intérêt du contrefactuel de l'unité traitée.

Cependant, deux questions restent en suspens : non seulement concernant l'existence de  $W^*$ , mais aussi  $\mu_i$  n'est pas observé, ce qui ne nous permet pas de savoir si (4) tient.

Les équations de (4) sont justes uniquement si le vecteur  $(Z_1, \mu_1)$  est à l'intérieur de l'enveloppe convexe des vecteurs correspondants des unités de contrôle. Si le vecteur est en dehors, alors  $W^*$  ne peut pas exister, et par conséquent ne pourra être construit d'unité de contrôle synthétique. Abadie, Diamond et Heinmueller affirment cependant que si (4) est vraie de manière approximative, cette approximation débordera sur celle de la zone de contrôle synthétique, qui ne reproduira donc qu'approximativement  $Y_{1,t}(0)$ . La manière dont (4) est respectée (i.e la manière dont les variables préintervention s'apparent) peut ainsi être interprétée comme une mesure du biais de l'estimateur.

En ce qui concerne la seconde question, Abadie, Diamond et Heinmueller montrent que s'il existe  $W$  tel que :

$$\forall i \in [1, \dots, T_0], \quad Y_{1,t} = \sum_{j=2}^{J+1} w_j^* Y_{j,t} \quad \text{et} \quad Z_1 = \sum_{j=2}^{J+1} w_j^* Z_j$$

alors ces poids approchent  $W^*$  lorsque  $T_0$  augmente. Abadie, Diamond et Heinmueller affirment ainsi enfin que si la période de préintervention  $T_0$  est raisonnablement large, l'estimateur est sans biais (voir Appendice 3 dans Abadie *et alii* (2010) [1]).

## 6.4 L'implémentation algorithmique

Abadie, Diamond et Hainmueller ont développé un Rpackage appelé «Synth» qui permet d'implémenter leur méthode de contrôle synthétique. Nous allons revenir sur leur approche afin d'expliquer les adaptations que nous avons faites.

Nous avons déjà indiqué dans notre partie 1, le problème d'optimisation imbriquée auquel nous faisons face: la procédure cherche à la fois le vecteur  $W$  qui minimise la distance entre les prédicteurs pré-intervention des régions deu groupe de donneurs et ceux de la région d'intérêt assignant ainsi différents poids aux régions de contrôle et la matrice  $V$  semi-définie, positive qui minimise la distance entre les valeurs de la variable d'intérêt pré-intervention de notre groupe de donneurs et celles de notre région d'intérêt, assignant ainsi différents poids aux prédicteurs.

Plutôt que de créer ex-nihilo la matrice  $V$  (choix fait lorsque les chercheurs ont déjà une idée préconçue des prédicteurs qu'ils veulent mettre en avant), nous avons choisi d'obtenir cette matrice uniquement à partir des données et donc à partir d'un problème d'optimisation cherchant à minimiser la moyenne de la racine carrée des erreurs de prédiction (MSPE pour «Mean square prediction error») sur la période de préintervention (cf. équation (2)).

Il existe plusieurs méthodes pour résoudre ce problème d'optimisation. On pourrait d'abord penser à une validation-croisée à partir d'une partition en deux de la période de pré-traitement (comme appliquée par Abadie *et alii* sur la réunification [citeAbadie2015comparative](#)). Cependant Kaul *et alii* (2016) [19] ont démontré que ceci n'est pas pertinent dans la mesure à la solution produite pour le vecteur  $W$  n'est pas unique. Nous avons donc plutôt utilisé l'optimisation offerte par le Rpackage «Kernlab». Ainsi nous optimisons le premier programme eqrefoptimW grâce à l'algorithme Broyden-Fletcher-Goldfarb-Shanno (BFGS) de quasi-Newton, qui suppose que la fonction peut être approchée localement par un développement limité quadratique autour de l'optimum. Afin de ne pas se retrouver dans un problème de minimum local non pris en compte dans cet algorithme, nous l'exécutons par deux fois en utilisant deux matrices  $V$  différentes, suivant la proposition du Rpackage «Synth». Il était possible de mettre une option qui combine cet algorithme avec des méthodes évolutionnaires issues du Rpackage «Rgenoud» cependant les résultats n'étaient pas très différents et l'implémentation était très longue.

## 6.5 Validation de l'hypothèse de restriction

Figure 16: Représentation de l'enveloppe convexe des données par variable

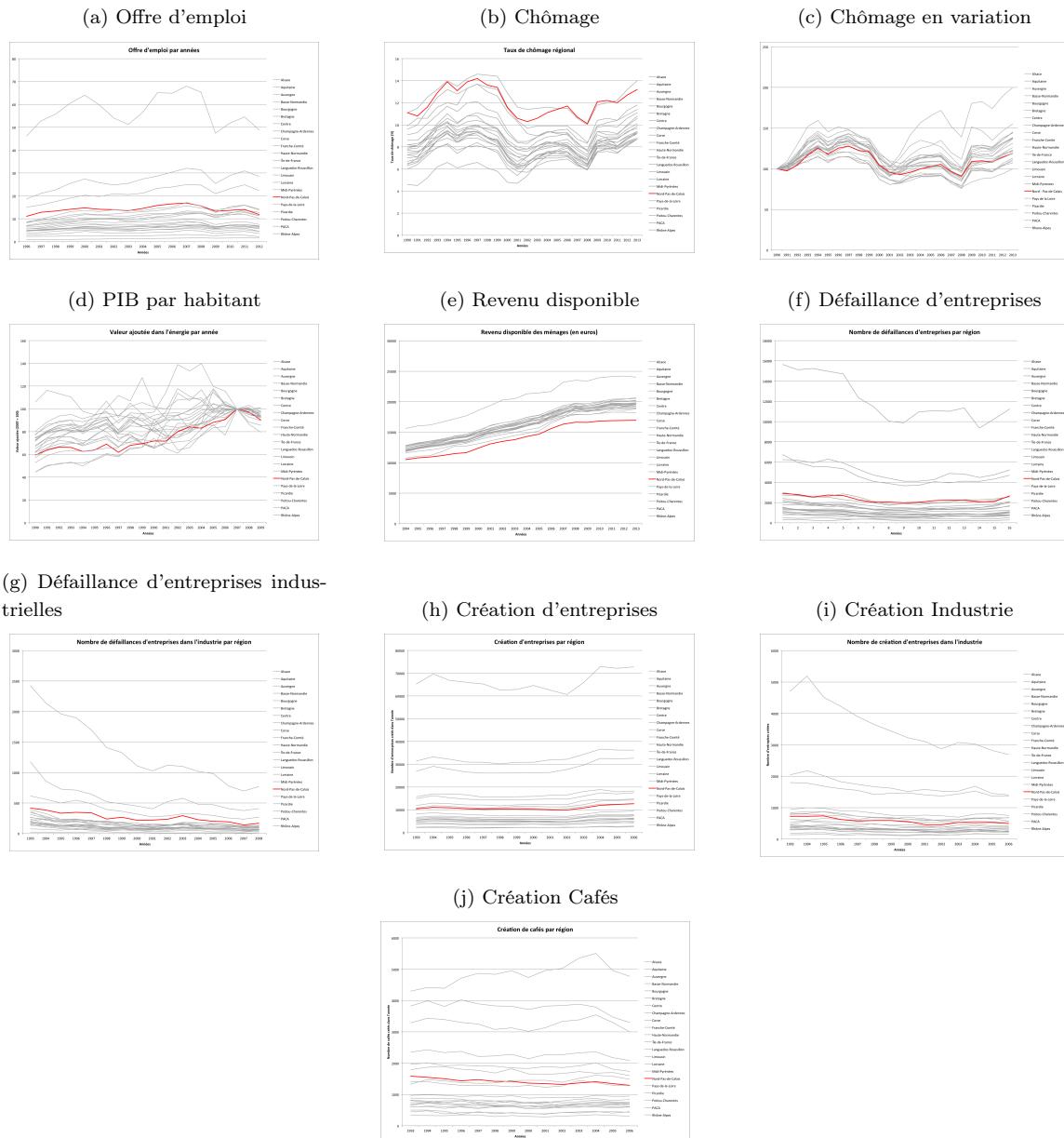
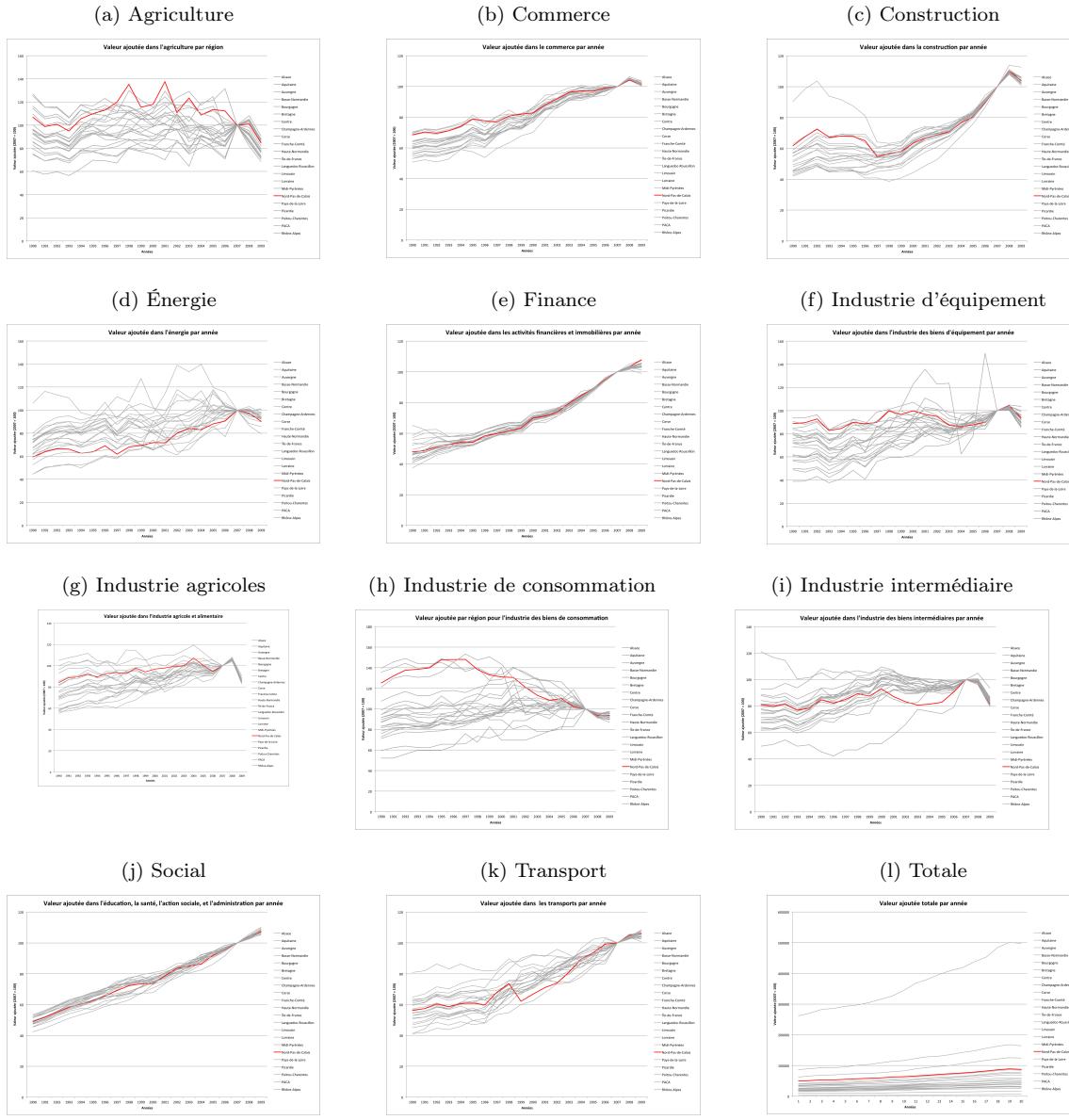


Figure 17: Représentation de l'enveloppe convexe des données pour les valeurs ajoutées



## 6.6 Validation de l'hypothèse d'exclusion

**TRANSPORT INTERRÉGIONAL PAR ROUTE ET VOIE NAVIGABLE AVANT POUR ZONE DE CHARGEMENT LE NORD-PAS-DE-CALAIS**

Régions	Île-de-France	Champagne-Ardenne	Picardie	Haute-Normandie	Centre	Haute-Normandie	Basse-Normandie	Bruxelles	Lorraine	Alsace	Franche-Comté	Pays-de-la-Terre
<i>Transport en milliers de tonnes par année:</i>												
2003	6101	2557	10592	2176	1044	437	683	5811	598	533	1659	-
2004	6255	2992	10873	2272	1202	698	713	5978	594	453	1698	-
2005	6056	2862	10678	2187	1190	526	722	5909	781	423	1656	-
2006	7215	2192	11795	2511	1147	688	720	6292	895	384	1352	-
2007	6005	2256	8607	2192	1265	639	764	525	528	246	1178	-
2008	5615	1780	8763	2213	1221	483	1273	448	284	1153	-	-
2009	5179	1892	7470	1956	1064	543	545	581	393	291	1027	-
<b>Total</b>	<b>42426</b>	<b>16531</b>	<b>68578</b>	<b>15807</b>	<b>8133</b>	<b>3861</b>	<b>4865</b>	<b>27769</b>	<b>4937</b>	<b>2614</b>	<b>9723</b>	-
<i>Contributions:</i>												
2003	16%	7%	27%	6%	3%	1%	2%	15%	2%	1%	4%	-
2004	16%	8%	26%	6%	3%	2%	2%	15%	3%	1%	4%	-
2005	16%	7%	28%	6%	3%	1%	2%	15%	2%	1%	4%	-
2006	18%	5%	29%	6%	3%	1%	2%	16%	2%	1%	3%	-
2007	21%	8%	25%	8%	4%	2%	3%	3%	2%	1%	4%	-
2008	20%	6%	31%	8%	4%	2%	3%	5%	2%	1%	4%	-
2009	21%	8%	38%	8%	4%	2%	2%	4%	2%	1%	4%	-
Contribution sur les échanges totaux	18%	7%	29%	7%	3%	2%	2%	11%	2%	1%	4%	-
Régions	Brétagne	Poitou-Charentes	Aquitaine	Madère-Pyrénées	Limousin	Rhône-Alpes	Auvergne	LangUEDOC-PIACA-Corse	Total en provenance du NPPC.			
<i>Transport en milliers de tonnes par année:</i>												
2003	938	436	584	356	147	1499	237	479	916	37888	-	-
2004	872	597	669	341	118	1427	250	545	922	39482	-	-
2005	854	341	576	328	128	1397	282	514	762	38177	-	-
2006	835	342	641	362	190	1604	311	362	772	40559	-	-
2007	653	386	441	309	149	1618	334	225	485	29275	-	-
2008	878	397	225	86	171	308	171	637	28074	-	-	-
2009	813	350	369	188	84	1203	178	122	376	25084	-	-
<b>Total</b>	<b>5844</b>	<b>2852</b>	<b>3677</b>	<b>2109</b>	<b>902</b>	<b>9979</b>	<b>1980</b>	<b>2613</b>	<b>4870</b>	<b>238495</b>	-	-
<i>Contributions:</i>												
2003	2%	1%	2%	1%	0%	4%	1%	1%	2%	100%	-	-
2004	2%	1%	1%	2%	1%	0%	4%	1%	1%	2%	100%	-
2005	2%	1%	2%	1%	0%	4%	1%	1%	1%	2%	100%	-
2006	2%	1%	2%	1%	0%	4%	1%	1%	1%	2%	100%	-
2007	2%	1%	2%	1%	1%	0%	4%	1%	1%	2%	100%	-
2008	3%	1%	1%	1%	0%	4%	1%	1%	1%	2%	100%	-
2009	3%	2%	1%	1%	0%	5%	1%	0%	1%	100%	-	-
Contribution sur les échanges totaux	2%	1%	2%	1%	0%	4%	1%	1%	2%	100%	-	-

Données issues de

Figure 18: Destinations des marchandises issues du Nord-Pas-de-Calais

**TRANSPORT INTERRÉGIONAL PAR ROUTE ET VOIE NAVIGABLE AVANT POUR ZONE DE DÉCHARGEMENT LE NORD-PAS-DE-CALAIS**

Régions	Ile-de-France	Champagne-Ardenne	Picardie	Haute-Normandie	Centre	Basse-Normandie	Bourgogne	Lorraine	Alsace	Franche-Comté	Pays de la Loire
<i>Transport en milliers de tonnes par année</i>											
2003	4161		2321	8127	1843	1043	413	598	1857	246	334
2004	3997		1942	7333	1973	1769	587	476	1864	668	265
2005	4254		2320	7467	1691	1733	545	645	1857	639	408
2006	4216		2224	7101	1853	1768	531	755	1827	885	302
2007	3914		1820	7552	1720	1281	635	712	1833	565	232
2008	4229		1767	6404	1980	1128	511	632	1142	425	170
2009	3408		1716	6158	1775	1161	423	457	1069	405	254
<b>Total</b>	<b>28171</b>		<b>14110</b>	<b>50742</b>	<b>12755</b>	<b>10423</b>	<b>3665</b>	<b>4315</b>	<b>10840</b>	<b>4253</b>	<b>1965</b>
<i>Contributions</i>											
2003	15%		8%	29%	7%	6%	1%	2%	6%	3%	1%
2004	15%		7%	27%	7%	7%	2%	2%	7%	2%	1%
2005	16%		9%	26%	6%	6%	2%	2%	6%	2%	4%
2006	15%		8%	26%	7%	6%	2%	3%	7%	5%	1%
2007	16%		8%	31%	7%	5%	3%	4%	2%	1%	4%
2008	19%		8%	29%	9%	5%	2%	3%	2%	1%	4%
2009	16%		8%	30%	9%	5%	2%	2%	5%	1%	5%
<b>Contribution sur les échanges totaux</b>	<b>16%</b>		<b>8%</b>	<b>29%</b>	<b>7%</b>	<b>6%</b>	<b>2%</b>	<b>2%</b>	<b>6%</b>	<b>1%</b>	<b>4%</b>
<i>Régions</i>											
	Bretagne	Poitou-Charentes	Aquitaine	Limousin	Pyrénées	Occitanie	Rhône-Alpes	Auvergne	Langue d'Auvergne	Corse	Total en direction du Nord-Pas-de-Calais
<i>Transport en milliers de tonnes par année</i>											
2003	630		404	457	227	127	1281	304	357	650	2763
2004	635		439	457	211	111	1311	475	638	630	2697
2005	597		294	526	180	133	1420	376	494	544	26977
2006	526		243	552	179	93	1390	473	338	594	27537
2007	578		293	367	216	71	1155	359	277	365	24130
2008	652		117	297	147	93	926	176	206	434	22311
2009	705		269	320	155	64	961	220	122	294	20847
<b>Total</b>	<b>4444</b>		<b>2069</b>	<b>3016</b>	<b>1315</b>	<b>692</b>	<b>8444</b>	<b>2383</b>	<b>2834</b>	<b>3520</b>	<b>17650</b>
<i>Contributions</i>											
2003	2%		1%	2%	1%	0%	5%	1%	2%	2%	100%
2004	3%		2%	2%	1%	0%	5%	2%	2%	2%	100%
2005	2%		1%	2%	1%	0%	5%	1%	2%	2%	100%
2006	2%		1%	2%	1%	0%	5%	2%	1%	2%	100%
2007	2%		1%	2%	1%	0%	5%	1%	2%	2%	100%
2008	3%		1%	2%	1%	0%	4%	1%	1%	2%	100%
2009	3%		1%	2%	1%	0%	5%	1%	1%	1%	100%
<b>Contribution sur les échanges totaux</b>	<b>3%</b>		<b>1%</b>	<b>2%</b>	<b>1%</b>	<b>0%</b>	<b>5%</b>	<b>1%</b>	<b>1%</b>	<b>2%</b>	<b>100%</b>

Données issues de

Figure 19: Provenance des marchandises ayant pour destination le Nord-Pas-de-Calais

## 6.7 Tableaux de résultat du placebo de sensibilité

### 6.7.1 Modèle de base

Poids	Régions
0.000	Aquitaine
0.000	Auvergne
0.000	Basse-Normandie
0.000	Bourgogne
0.000	Bretagne
0.000	Centre
0.000	Franche-Comté
0.775	Languedoc-Roussillon
0.000	Limousin
0.000	Midi-Pyrénées
0.077	Pays de la Loire
0.148	Poitou-Charentes
0.000	PACA
0.000	Rhône-Alpes

### 6.7.2 Sans le Poitou-Charentes

Poids	Régions
0.000	Aquitaine
0.000	Auvergne
0.000	Basse-Normandie
0.000	Bourgogne
0.000	Bretagne
0.000	Centre
0.000	Franche-Comté
0.852	Languedoc-Roussillon
0.000	Limousin
0.000	Midi-Pyrénées
0.147	Pays de la Loire
0.000	PACA
0.000	Rhône-Alpes

### 6.7.3 Sans le Pays de la Loire

#### Poids Régions

0.000	Aquitaine
0.000	Auvergne
0.000	Basse-Normandie
0.000	Bourgogne
0.121	Bretagne
0.004	Centre
0.000	Franche-Comté
0.757	Languedoc-Roussillon
0.000	Limousin
0.000	Midi-Pyrénées
0.117	Poitou-Charentes
0.000	PACA
0.000	Rhône-Alpes

## List of Figures

1	Comparaison de l'évolution du chômage entre 1990 et 2012 . . . . .	11
2	Comparaison de l'évolution du PIB par tête entre 1990 et 2012 . . . . .	12
3	Régions de contrôle . . . . .	13
4	Comparaison des constructions de contrefactuel . . . . .	13
5	Évolution comparée du PIB par tête . . . . .	14
6	Écart entre PIB réel et PIB synthétique . . . . .	14
7	Évolution comparée du chômage . . . . .	15
8	Comparaison des contrefactuels du chômage . . . . .	15
9	Zone de contrôle . . . . .	17
10	Test de sensibilité à la moyenne . . . . .	17
11	Placebo dans le temps . . . . .	18
12	Placebo dans l'espace . . . . .	19
13	Les réalisations dans le Nord-Pas-de-Calais de <i>Lille 2004</i> . . . . .	21
14	Les réalisations à Lille de <i>Lille 2004</i> . . . . .	21
15	Impacts de <i>Lille 2004</i> : évolution de l'emploi entre octobre 2003 et septembre 2004 . . . . .	22
16	Représentation de l'enveloppe convexe des données par variable . . . . .	26
17	Représentation de l'enveloppe convexe des données pour les valeurs ajoutées . . . . .	27
18	Destinations des marchandises issues du Nord-Pas-de-Calais . . . . .	28
19	Provenance des marchandises ayant pour destination le Nord-Pas-de-Calais . . . . .	29

## References

- [1] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 2010.
- [2] Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American economic review*, pages 113–132, 2003.
- [3] Michihito Ando and Fredrik Sävje. Inference with the synthetic control method. In *the 67th European Meeting of the Econometric Society*. Institutet för bostads-och urbanforskning, Uppsala universitet, 2013.
- [4] Sebastian Bauhoff. The effect of school district nutrition policies on dietary intake and overweight: a synthetic control approach. *Economics & Human Biology*, 12:45–55, 2014.
- [5] Michèle Belot and Vincent Vandenberghe. Evaluating the ‘threat’ effects of grade repetition: exploiting the 2001 reform by the french-speaking community of belgium. *Education Economics*, 22(1):73–89, 2014.
- [6] Firat Bilgel and Brian Galle. Financial incentives for kidney donation: A comparative case study using synthetic controls. *Journal of health economics*, 43:103–117, 2015.

- [7] Andreas Billmeier and Tommaso Nannicini. Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001, 2013.
- [8] Sarah Bohn, Magnus Lofstrom, and Steven Raphael. Did the 2007 legal arizona workers act reduce the state’s unauthorized immigrant population? *Review of Economics and Statistics*, 96(2):258–269, 2014.
- [9] Eduardo Cavallo, Sebastian Galiani, Ilan Noy, and Juan Pantano. Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, 95(5):1549–1561, 2013.
- [10] Makenna Coffman, Ilan Noy, and autres. Hurricane iniki: measuring the long-term economic impact of a natural disaster using synthetic control. *Environment and Development Economics*, 17(2):187–205, 2012.
- [11] Jean-Cédric Delvainquièrre and Bruno Dietsch. Les dépenses culturelles des collectivités locales en 2006: près de 7 milliards d'euros pour la culture. *Culture chiffres*, (3):1–32, 2009.
- [12] Laurent Gobillon and Thierry Magnac. Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, (0), 2015.
- [13] Pedro Gomes and Alejandro Librero-Cano. Evaluating 3 decades of the european capital of culture programme: a difference-in-differences approach. 2014.
- [14] Luis César Herrero, José Ángel Sanz, María Devesa, Ana Bedate, and María José Del Barrio. The economic impact of cultural events a case-study of salamanca 2002, european capital of culture. *European urban and regional studies*, 13(1):41–57, 2006.
- [15] Peter Hinrichs. The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics*, 94(3):712–722, 2012.
- [16] Amr Sadek Hosny. Algeria’s trade with gafta countries: A synthetic control approach. *Transition Studies Review*, 19(1):35–42, 2012.
- [17] Stephanie Jasmand and Wolfgang Maennig. Regional income and employment effects of the 1972 munich summer olympic games. *Regional Studies*, 42(7):991–1002, 2008.
- [18] Yothin Jinjarak, Ilan Noy, and Huanhuan Zheng. Capital controls in brazil–stemming a tide with a signal? *Journal of Banking & Finance*, 37(8):2938–2952, 2013.
- [19] Stefan Klößner, Ashok Kaul, Gregor Pfeifer, and Manuel Schieler. Comparative politics and the synthetic control method revisited: A note on abadie et al.(2015). Technical report, Working Paper, 2016.
- [20] Noémi Kreif, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton. Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics*, 2015.
- [21] Christine Liefooghe. Lille 2004, capitale européenne de la culture ou la quête d'un nouveau modèle de développement. *Méditerranée. Revue géographique des pays méditerranéens/Journal of Mediterranean geography*, (114):35–45, 2010.

- [22] Elisabeth Mettler and J Mittag. Nachhaltige effekte oder strohfeuer für ein jahr. die kulturhauptstadtjahre glasgow 1990, luxemburg 1995 und weimar 1999. *Mittag (2008)*, pages 125–143, 2008.
- [23] Tommaso Nannicini and Andreas Billmeier. Economies in transition: How important is trade openness for growth?\*. *Oxford Bulletin of Economics and Statistics*, 73(3):287–314, 2011.
- [24] Greg Richards and Julie Wilson. The impact of cultural events on city image: Rotterdam, cultural capital of europe 2001. *Urban studies*, 41(10):1931–1951, 2004.
- [25] Jenna Stearns. The effects of paid maternity leave: Evidence from temporary disability insurance. *Journal of health economics*, 43:85–102, 2015.
- [26] Thomas Werquin. *Impact de l'infrastructure culturelle sur le développement économique local-Elaboration d'une méthode d'évaluation ex-post et application à Lille2004 Capitale européenne de la Culture*. PhD thesis, Université des Sciences et Technologie de Lille-Lille I, 2006.