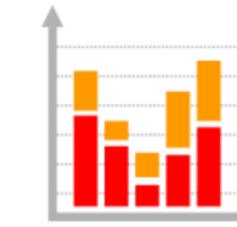
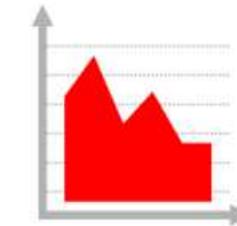
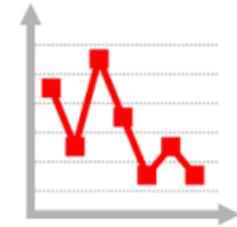


L'indice du coût de la construction

Tedjani TARAYOUN
Division Recueil et traitement de l'information
DMCSI



L'indice du coût de la construction

Plan de la présentation

1. Présentation générale de l'indice
2. Enquête PRLN
3. Le principe de calcul de l'indice
4. Discussion autour de la spécification
 - 4.1 Sélection des variables
 - 4.2 Variables d'aménités

1. Présentation générale de l'indice

Un indice de prix (et non de coût) de la construction de logements neufs à usage d'habitation

- Indice trimestriel, référence 100 au 4ème trimestre 1953
- Outil de connaissance de l'évolution des prix dans la construction
- ICC est une composante de l'ILC et de l'ILAT, les nouveaux indices de référence pour l'indexation des loyers des baux tertiaire
- Utilisé comme déflateur des comptes nationaux de la construction (avec l'IPEA)

Champ de l'indice

- Logements collectifs, individuels purs et individuels groupés.
- Le prix retenu est celui payé par le maître d'ouvrage
- L'indice porte uniquement sur les prestations de construction du bâtiment

2. Enquête PRLN

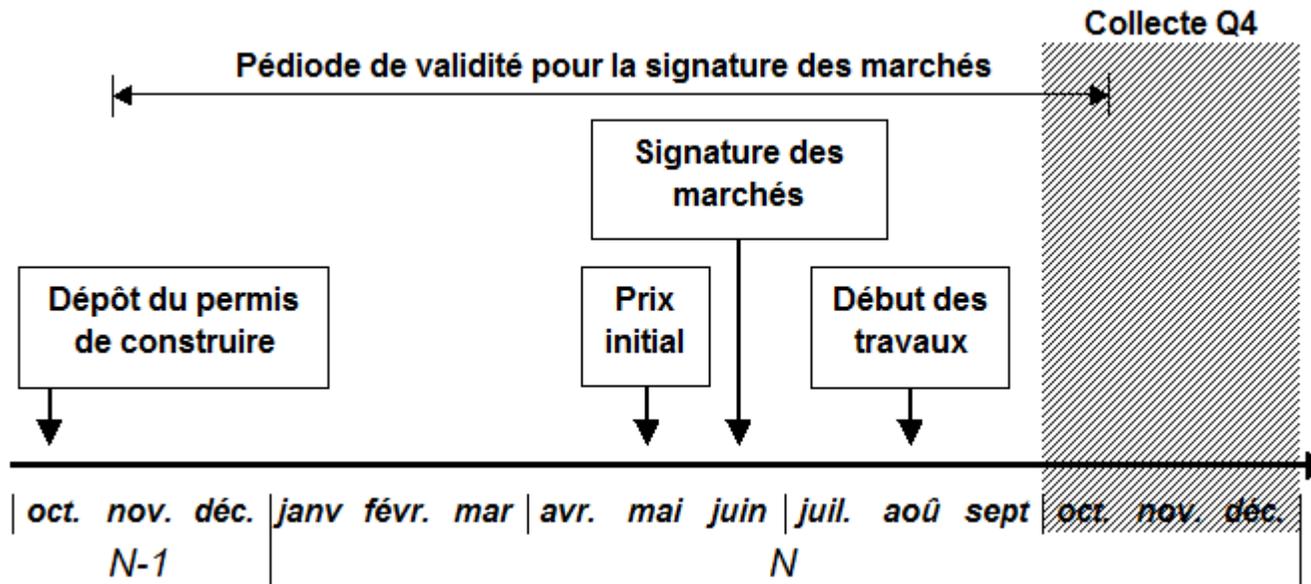
Chaque trimestre, l'enquête PRLN permet de recueillir les données de construction en 2 phases

- Une 1^{ère} phase : repérage des permis de construire correspondant aux critères de l'indice, 1500 permis
- Une 2^e phase : collecte des informations techniques et financières sur les permis de construire, 500 permis, dont 50 % individuels purs, 10 % individuels groupés et 40 % collectifs

On ne retrouve en 2^e phase que les opérations déclarées éligibles pour le calcul de l'ICC :

- Les permis relatifs à la construction d'un ou plusieurs logements neufs (hors rénovation, etc.)
- Les marchés signés dans un délai de 12 mois

Les prix sont ensuite tous ramenés à des prix centrés sur le deuxième mois du trimestre sous revue (en appliquant l'évolution mesurée au travers de l'Index Bâtiment BT01)



24/03/2017

3. Le principe de calcul de l'indice

3 modèles hédoniques sont estimés séparément pour les 3 grands types de construction : individuel pur, individuel groupé et collectif.

Les 3 modèles sont estimés chaque trimestre t en fenêtres glissantes à partir des données des 7 dernières enquêtes PRLN :

$$\log(\text{Prix}_i) = \alpha + X_i\beta + \sum_{j=0}^5 \gamma_{t-j} T_i^{t-j} + \epsilon_i \quad (\text{reference : } T_i^{t-6})$$

- **Prix_i** le prix de la construction du logement i (ou le prix total divisé par le nombre de logements pour les habitats collectifs et individuels groupés)
- **X_i** le vecteur des caractéristiques du logement i et
- **T_i^{t-j}** l'indicatrice valant 1 si le logement i a été observé au trimestre « $t-j$ »
- On fait une première régression qui permet d'identifier des observations « atypiques », celles dont la valeur absolue du Student du résidu > 2 ou la distance de Cook > 1
- On fait une seconde régression en affectant un poids nul aux observations « atypiques »

L'indice des prix hédoniques du trimestre t s'écrit : $\text{Indice}_{t/t-1} = \exp(\hat{\gamma}_t - \hat{\gamma}_{t-1})$

3. Le principe de calcul de l'indice

L'ICC hédonique pour l'ensemble de la construction de logements neufs résulte de l'**agrégation** des 3 sous-indices comme une moyenne harmonique pondérée par la valeur de chaque type de construction à la période courante (technique de Paasche)

- Valeur de l'individuel pur au T2 2016 = 3,1 milliards d'euros
- Valeur de l'individuel groupé au T2 2016 = 0,9 milliards d'euros
- Valeur du collectif au T2 2016 = 4 milliards d'euros

Une méthode de **chaînage** par recouvrement annuel est mise en œuvre

- En notant A l'année, t le trimestre et i le type de construction, la formule de l'ICC est :

$$ICC_{A,t} = ICC_{A-1} * \frac{\sum_{i=1}^3 Valeurs_{A,t}(i)}{\sum_{i=1}^3 Valeurs_{A,t}(i) * Indice_{A-1}^i / Indice_{A,t}^i}$$

- Et la moyenne annuelle du sous-indice i s'écrit :

$$Indice_A^i = \frac{\sum_{t \in A} Valeurs_{A,t}(i)}{\sum_{t \in A} Valeurs_{A,t}(i) * Indice_{A-1}^i / Indice_{A,t}^i}$$

4. Discussion autour de la spécification

Nous étudions la possibilité d'améliorer le dispositif de calcul

Par le passé, la liste des variables prises en compte dans le modèle a pu évoluer d'un trimestre à l'autre, d'une part, par l'intégration de variables issues d'évolutions réglementaires (conformité à la Réglementation Thermique 2012 notamment), et d'autre part, de façon à améliorer le R^2 ajusté

- Inconvénients : - les variables de qualité devraient avoir un impact structurel
 - risque d'instabilité de l'indice
- Proposition : construire un modèle stable dans le temps, qui permet de conserver les mêmes variables d'un trimestre à l'autre

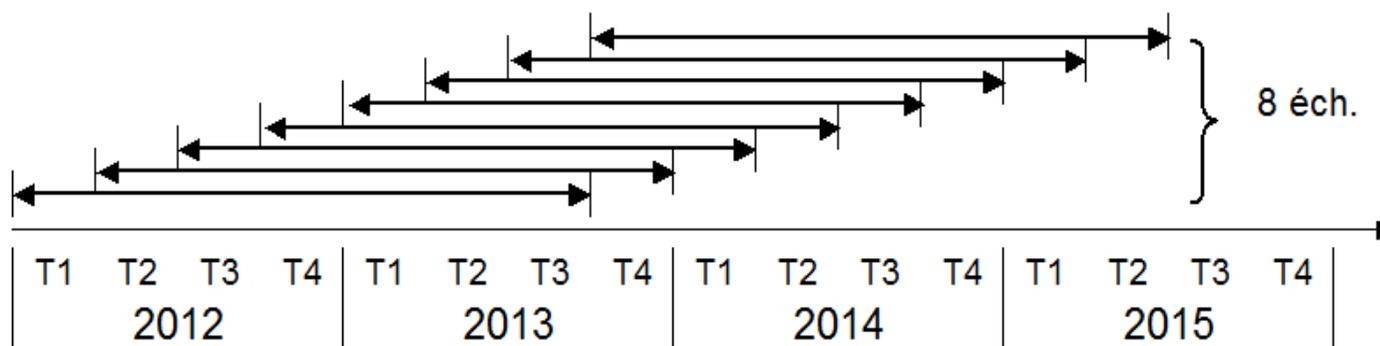
La méthode hédonique repose sur l'hypothèse que le prix total d'un logement est la somme des prix implicites que l'acheteur attribue aux différentes caractéristiques du logement, y compris les services et aménités à proximité

- Risques liés à l'omission de ces variables : si ces variables omises ne sont pas en moyenne les mêmes d'une période à la suivante, alors l'évolution de prix ne peut pas être considérée à champ constant
- Proposition : étudier la possibilité d'ajouter des variables d'aménités au modèle

4.1. Sélection des variables

Nous testons une procédure de sélection de variables de qualité qui permet d'obtenir un modèle d'estimation stable, qui ne nécessite pas d'être actualisé plusieurs fois par an

- On considère plusieurs échantillons en fenêtres glissantes, chacun ayant une taille de 7 trimestres
- On estime un modèle sur chaque échantillon via une procédure stepwise
- On compare les variables sélectionnées dans chaque modèle
- On élimine les variables les moins significatives et on relance les procédures stepwise
- On réitère l'étape précédente plusieurs fois de manière à obtenir une liste de variables significatives dans tous les modèles



4.1. Sélection des variables

Enfin, on vérifie l'absence de multicolinéarité entre les variables explicatives, en particulier avec les indicatrices temporelles

- La multicolinéarité augmente la variance des coefficients et peut rendre les valeurs et les signes des coefficients instables à l'ajout ou à la suppression de quelques observations

Pour détecter la multicolinéarité :

- On régresse chacune des variables explicatives par les autres
- Chaque R^2 issu des régressions mesure la part de variance expliquée par les autres variables, et le calcul de $1/(1-R^2)$ permet d'obtenir les « VIF » (Facteur d'inflation de la variance)
- Un $VIF > 2,5$ ($R^2 > 60\%$) peut être problématique

4.1. Sélection des variables

Résultats de la sélection de variables

Les variables de qualité sélectionnées impactent fortement le prix, de manière stable dans le temps, sans soucis de multicollinéarité

–Collectif : 11 variables de qualité

- Une augmentation de 1 % de la surface d'un appartement augmente son prix de 0,8 %
- Construire en île de France augmente le prix de la construction de 18 %

–Individuel : 14 variables de qualité

- Une augmentation de 1 % de la surface de la maison augmente son prix de 0,8 %
- La présence d'un sous-sol augmente le prix de 12 %

–Individuel groupé : 6 variables de qualité

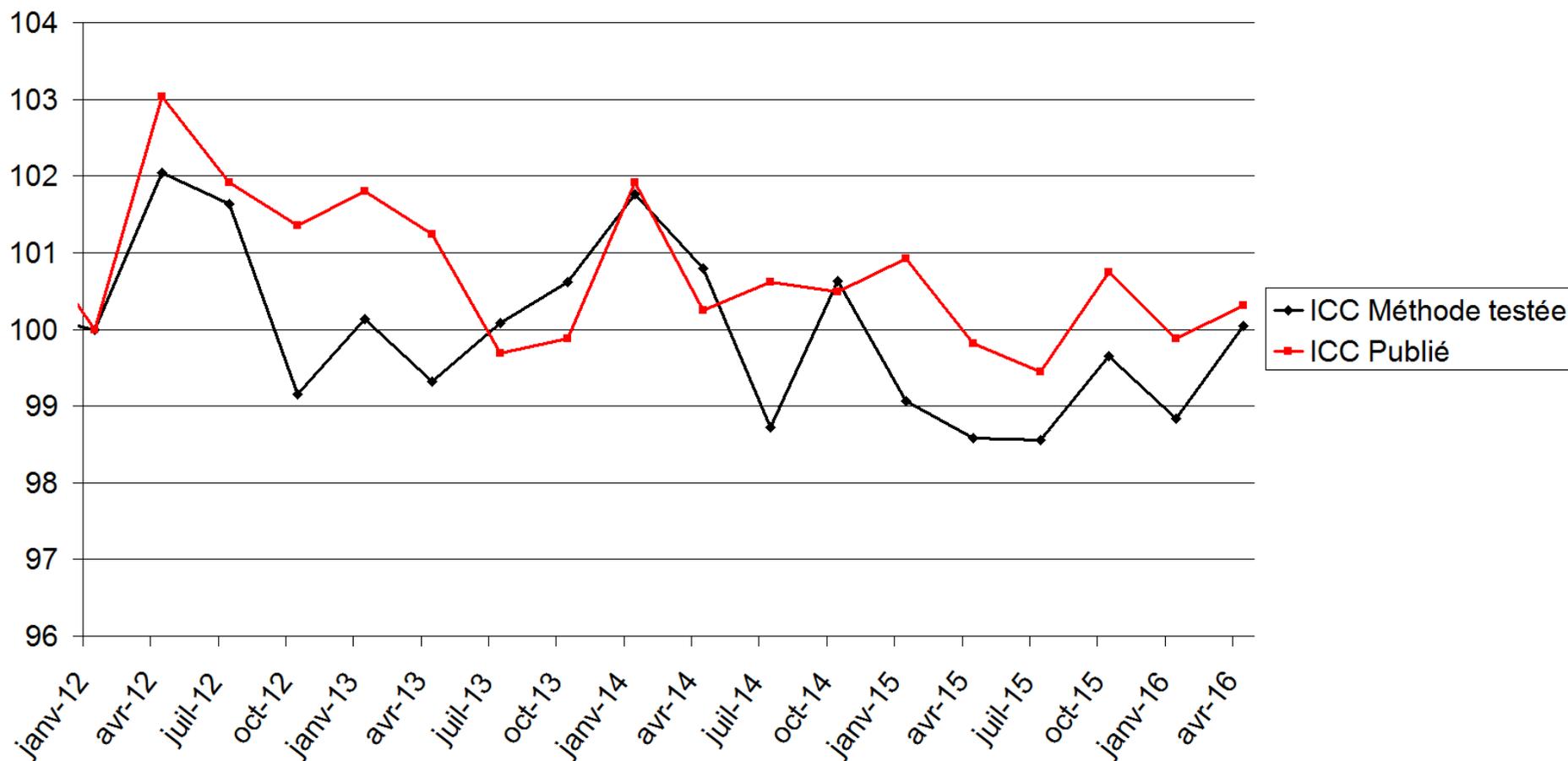
- Une augmentation de 1 % de la surface par logement fait croître le prix de 0,7 %
- L'absence de travaux de peinture réduit le prix de 12 %

Les indicatrices temporelles sont hors procédure de sélection des variables car les variations pures de prix sont calculées grâce aux coefficients qui y sont associées.

Quelle que soit la méthode de sélection des variables utilisée (méthode actuelle ou méthode testée), les coefficients associées aux indicatrices temporelles ne ressortent significatifs que très rarement.

4.1. Sélection des variables

Tracé de l'ICC publié et de l'ICC obtenu avec la nouvelle sélection de variables



4.2. Variables d'aménités

Nous disposons d'environ 60 variables décrivant les services et les aménités dans les communes :

- Source BPE 2012 : Nombre de site touristiques, nombre de crèches, etc.
- Source Corinne Land Cover 2006 : % Espaces verts, % de Terres agricoles, etc.
- Source GEOIDD 2011 : Risque inondation, séisme, etc.
- Source Observatoire territoires : Littoral, Montagne, Indice de pollution atmosphérique,
- Autre : Distance à la commune de plus de 50 000 habitants

Objectif : construire une variable synthétique qui traduirait un effet global des aménités sur le prix et l'utiliser dans la sélection des variables

4.2. Variables d'aménités

On réalise une régression stepwise pour chaque type de construction sur la base des observations passées, sur une période de référence de dix ans, de 2002 à 2011

$$\log(\text{Prix}_i) = \alpha + Z_i \zeta + X_i \beta + \sum_{t=2002}^{2010} \delta_t 1_{\{\text{année}=t\}} + e_i$$

- Avec Z_i le vecteur des aménités de la commune où le logement est construit, et X_i le vecteur des caractéristiques de la construction

Les valeurs des coefficients $\hat{\zeta}$ permettent d'établir l'effet (linéaire) des aménités sur le logarithme du prix

- On en déduit pour chaque commune, en fonction des aménités Z_c qui y sont présentes, la valeur d'un score $Z_c \hat{\zeta}$
- On suppose que la relation est valable pour l'ensemble des communes, mêmes celles n'ayant pas été concernées par l'estimation de $\hat{\zeta}$

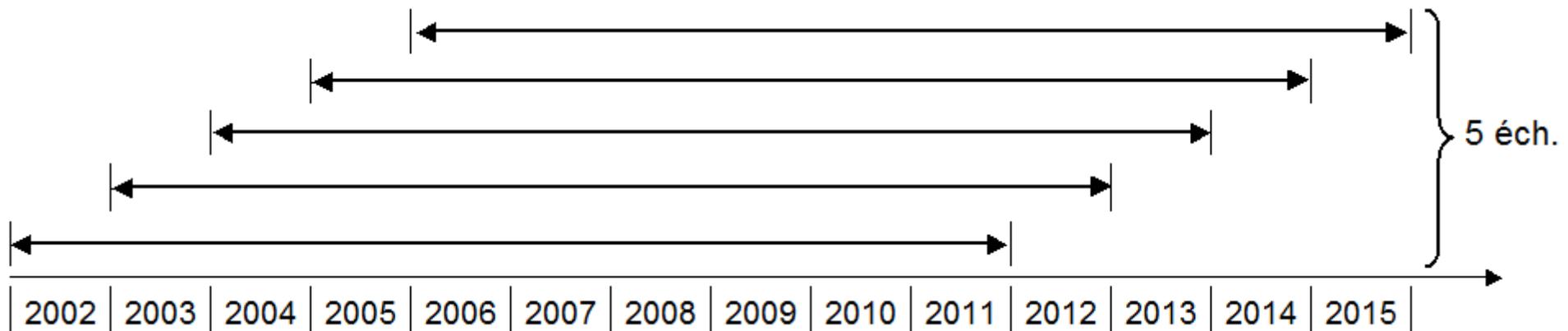
4.2. Variables d'aménités

Le score des communes est-il stable dans le temps ?

- Stabilité de la moyenne empirique ?
- Stabilité de la dispersion empirique ?
- Stabilité de la covariance empirique à écart de temps identique ?

Méthode :

- on décale la fenêtre de temps des échantillons
- on ré-estime les coefficients associés aux variables d'aménités
- on re-calcule le score de chaque commune
- on compare les résultats



4.2. Variables d'aménités

Stabilité du score

Statistiques simples

Variable	N	Moyenne	Écart-type
Score 02-11	34966	0,058	0,042
Score 03-12	34966	0,057	0,042
Score 04-13	34966	0,062	0,040
Score 05-14	34966	0,065	0,038
Score 06-15	34966	0,076	0,038

Matrice de covariance

	Score 02-11	Score 03-12	Score 04-13	Score 05-14	Score 06-15
Score 02-11	0,0018	0,0018	0,0017	0,0016	0,0016
Score 03-12		0,0018	0,0017	0,0016	0,0016
Score 04-13			0,0016	0,0015	0,0015
Score 05-14				0,0015	0,0015
Score 06-15					0,0015

Coefficients de corrélation de Pearson

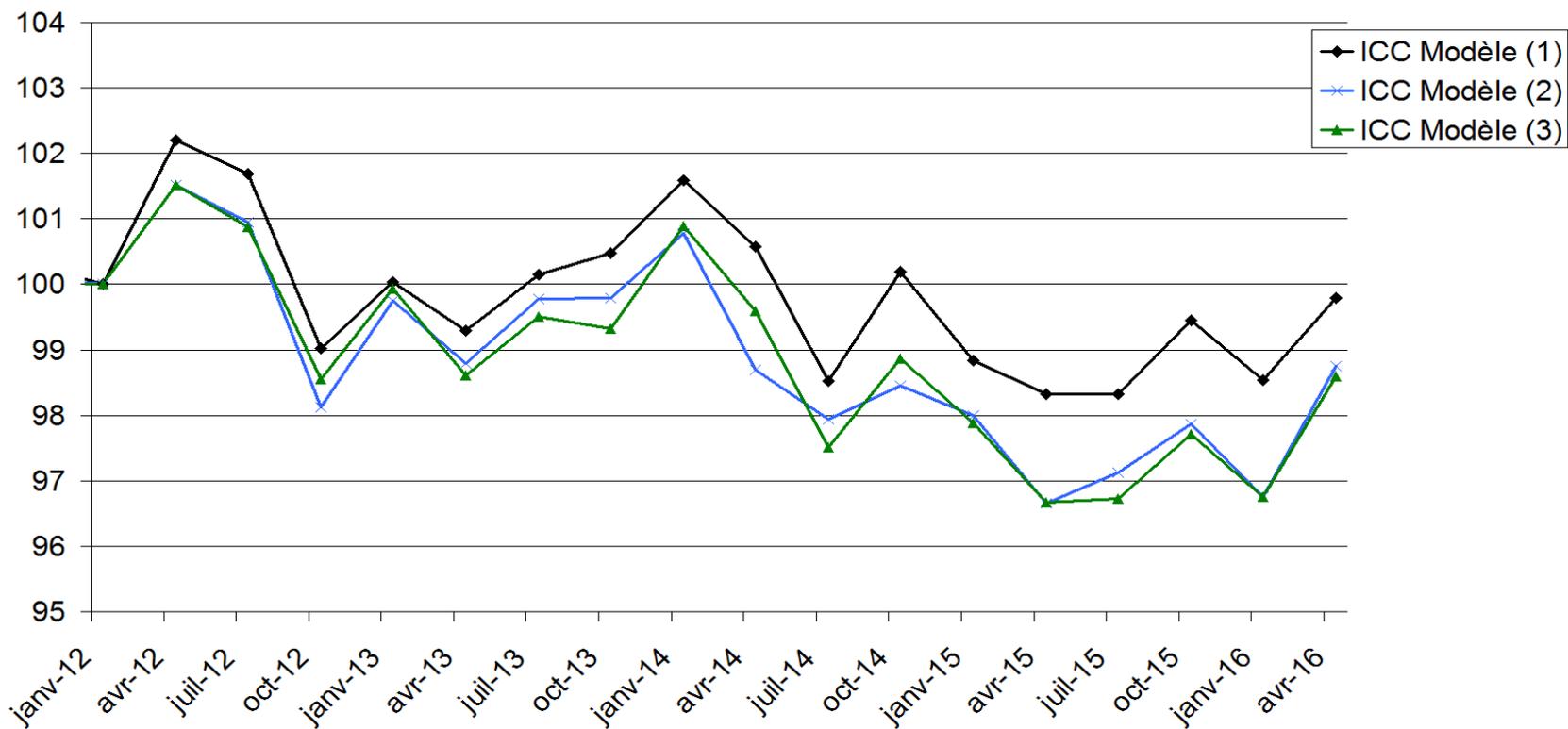
	Score 02-11	Score 03-12	Score 04-13	Score 05-14	Score 06-15
Score 02-11	1,00	1,00	0,99	0,99	0,99
Score 03-12		1,00	1,00	0,99	0,99
Score 04-13			1,00	1,00	1,00
Score 05-14				1,00	1,00
Score 06-15					1,00

4.2. Variables d'aménités

Le score calculé entre 2002 et 2011 a-t-il un impact sur l'indice hédonique ?

On construit 3 modèles :

- Modèle (1) où la variable score n'a pas été intégrée au processus de sélection des variables
- Modèle (2) où la variable score a été intégrée au processus de sélection des variables
- Modèle (3) = Modèle (2) sans la variable score



Bilan et conclusion

Nous proposons de ne pas intégrer la variable de score communal dans le calcul de l'indice :

- + : On peut considérer que l'effet des aménités communales est fixe dans le temps
- : Le gain associé est faible, l'ajout du score ne semble pas apporter d'informations supplémentaires
- : La maintenance de cette variable induit un coût supplémentaire

Nous proposons d'utiliser en production la méthode de sélection des variables que nous avons testée

Nous proposons un suivi de plusieurs indicateurs en production. Le producteur de l'indice doit en effet pouvoir relancer une sélection automatique des variables en cas d'instabilité des coefficients.

- Le tracé de l'indice hédonique (et de son intervalle de confiance à 95 %)
- Le R^2 ajusté qui mesure la qualité de l'ajustement
- La valorisation des coefficients estimés chaque trimestre
- La proportion de valeurs influencées et aberrantes supprimées

L'indice du coût de la construction

Merci de votre attention

Avez-vous des questions ?

Insee
www.insee.fr

Prénom Nom
N° téléphone
Mèl