

## Article

# Un estimateur par la régression généralisée de la variation des prix des logements fondé sur des évaluations foncières

par Jan de Haan et Rens Hendriks

Janvier 2014



## Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Téléphone** entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

## Programme des services de dépôt

Service de renseignements 1-800-635-7943  
Télécopieur 1-800-565-7757

## Comment accéder à ce produit

Le produit n° 12-001-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca) et de parcourir par « Ressource clé » > « Publications ».

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.gc.ca](http://www.statcan.gc.ca) sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2014

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada (<http://www.statcan.gc.ca/reference/licence-fra.html>).

This publication is also available in English.

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, ses entreprises, ses administrations et les autres établissements. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

## Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0<sup>s</sup> valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- P provisoire
- r révisé
- X confidentiel en vertu des dispositions de la *Loi sur la statistique*
- E à utiliser avec prudence
- F trop peu fiable pour être publié
- \* valeur significativement différente de l'estimation pour la catégorie de référence ( $p < 0,05$ )

# Un estimateur par la régression généralisée de la variation des prix des logements fondé sur des évaluations foncières

Jan de Haan et Rens Hendriks<sup>1</sup>

## Résumé

Statistics Netherlands s'appuie sur la méthode du ratio prix de vente-évaluation ou méthode SPAR (pour *Sale Price Appraisal Ratio*) pour produire son indice des prix des logements. Cette méthode combine les prix de vente aux évaluations foncières faites par l'administration publique. Le présent article décrit une approche de rechange dans laquelle les évaluations foncières servent d'information auxiliaire dans un cadre de régression généralisée (GREG). Une application aux données des Pays-Bas montre que, même si l'indice GREG est plus lisse que le ratio des moyennes d'échantillon, il donne une série très semblable à la série SPAR. Pour expliquer ce résultat, nous montrons que l'indice SPAR est un estimateur de notre indice GREG plus général et qu'en pratique, il est presque aussi efficace.

**Mots clés :** Estimation par la régression généralisée; indice des prix des logements; évaluations foncières; échantillonnage.

## 1 Introduction

Lorsqu'ils essaient de construire des indices des prix des logements de qualité constante, les organismes statistiques doivent résoudre plusieurs problèmes. Premièrement, l'appariement exact des biens immobiliers au fil du temps est problématique, car la qualité de ces biens aura probablement évolué; les logements se déprécient et peuvent également avoir subi des réparations, des rénovations ou des rajouts importants. Autrement dit, chaque bien immobilier peut être considéré à chaque période comme un bien unique. Deuxièmement, le roulement des logements est généralement faible comparativement au parc de logements et la composition des biens immobiliers vendus évolue au fil du temps, ce qui pose un problème de qualité. Troisièmement, les données sur les caractéristiques des logements font souvent défaut, ce qui a des répercussions sur le choix de la méthode de mesure.

Trois grandes catégories d'indices des prix des logements sont décrites dans la littérature, à savoir les indices médians ou moyens, les indices fondés sur la méthode des ventes répétées et les indices hédoniques. Un indice médian (moyen) suit l'évolution du prix du logement médian (moyen) négocié d'une période à la suivante. Cette méthode pose problème du fait que les caractéristiques, disons, du logement médian évolue au cours du temps. On contourne souvent ce problème en stratifiant les échantillons selon la région, le type de logement, *etc.*, procédure que l'on appelle aussi ajustement de la composition (*mix adjustment*). Évidemment, la stratification nécessite des données supplémentaires.

Les *méthodes des ventes répétées* abordent le problème de la composition qualitative (*quality mix*) en limitant le jeu de données aux logements qui ont été vendus au moins deux fois au cours de la période d'échantillonnage. On s'assure de cette façon de « comparer les mêmes choses », en supposant que la qualité des logements individuels ne change pas. Les méthodes des ventes répétées sont fondées sur des régressions dans lesquelles sont regroupées les données sur les ventes répétées pour différentes périodes.

---

1. Jan de Haan, OTB Research Institute for the Built Environment, Delft University of Technology and Division of Process Development, IT and Methodology, Statistics Netherlands, P.O. Box 24500, 2490 HA La Haye, Pays-Bas. Courriel : j.dehaan@cbs.nl; Rens Hendriks, Division of Economic and Business Statistics and National Accounts, Statistics Netherlands. Courriel : r.hendriks@cbs.nl.

Un inconvénient éventuel de cette approche est la révision; lorsque de nouvelles données sont ajoutées à l'échantillon, les indices calculés antérieurement changent. La méthode des ventes répétées a été proposée au départ par Bailey, Muth et Nourse (1963). Case et Shiller (1987, 1989) soutiennent que les variations des prix des logements comprennent des composantes dont la variance augmente avec l'intervalle entre les ventes et proposent une approche par les moindres carrés pondérés pour corriger ce genre d'hétéroscédasticité. Une autre méthode pondérée a été proposée par Calhoun (1996). Jansen, de Vries, Coolen, Lamain et Boelhouwer (2008), comparent, en se servant de données pour les Pays-Bas, la méthode des ventes répétées non pondérée à diverses méthodes pondérées et concluent que la méthode non pondérée donne des résultats satisfaisants.

Contrairement aux méthodes des ventes répétées, les *méthodes de régression hédoniques* permettent, en principe, de corriger les données pour tenir compte des changements de qualité des biens individuels (en plus des variations de composition qualitative). Ces méthodes requièrent des données sur les caractéristiques des logements, comme le nombre de chambres à coucher, et la taille et la localisation du terrain, pour estimer par la régression des indices de prix corrigés pour tenir compte de la qualité. Aujourd'hui, de nombreux pays calculent des indices hédoniques des prix des logements. Ainsi, un indice hédonique est produit par l'institut de statistique de la France (INSEE) en collaboration avec le Conseil supérieur du notariat (Gouriéroux et Laferrère 2009), ainsi que par Statistics Finland (Saarnio 2006). Au Royaume-Uni, trois indices hédoniques des prix des logements sont produits par différents instituts. En Australie, RPDData-Rismark calcule des indices hédoniques pour les capitales (Hardman 2011). On distingue deux grandes catégories d'indices hédoniques. La méthode des variables indicatrices temporelles (*time dummies*) consiste à modéliser le logarithme du prix en fonction des caractéristiques du bien et d'un jeu de variables indicatrices représentant les périodes. Comme les données pour toutes les périodes sont regroupées, cette méthode souffre aussi de révisions. Les méthodes d'imputation hédoniques, qui consistent à estimer les « prix manquants », n'ont pas cet inconvénient. Hill et Melser (2008) discutent de nombreuses méthodes d'imputation hédoniques dans le contexte du logement. Diewert, Heravi et Silver (2009) et de Haan (2010) donnent une comparaison des indices de prix fondés sur la méthode des variables indicatrices temporelles et sur la méthode d'imputation hédonique.

Une quatrième approche en vue d'estimer l'indice des prix des logements est celle de l'*utilisation de données d'évaluation foncière*. Une option consiste à augmenter le jeu de données sur les ventes répétées en utilisant des données d'évaluation foncière comme estimation des valeurs passées ou courantes des biens immobiliers qui n'ont pas été revendus durant la période d'échantillonnage. Certaines données sur lesquelles repose l'indice fondé sur les ventes répétées seraient alors des pseudo-données plutôt que des données réelles sur ces ventes. Pour en savoir davantage sur l'utilisation des valeurs d'évaluation dans un indice des prix fondés sur les ventes répétées et sur l'élimination du biais d'évaluation, voir, par exemple, Geltner (1996), Edelstein et Quan (2006), et Leventis (2006). Une autre option, qui tient également compte des effets des variations de composition qualitative, consiste à combiner les prix de vente de la période courante avec les évaluations pour une période antérieure afin de calculer des prix relatifs (ratio de prix) dans un cadre classique d'appariement de biens. L'un des avantages par rapport à l'approche des ventes répétées tient au fait que les indices ne seront pas révisés. Cette méthode dite du ratio prix de vente-évaluation (*Sale Price Appraisal Ratio* ou SPAR) est appliquée depuis longtemps en Nouvelle-Zélande et est également utilisée maintenant aux Pays-Bas et dans quelques autres pays européens. Bourassa, Hoesli et Sun (2006) décrivent l'indice SPAR de la Nouvelle-Zélande qui est produit par Quotable Value, une

entreprise publique d'évaluation foncière. D'autres études de la méthode SPAR comprennent Rossini et Kershaw (2006), van der Wal, ter Steege et Kroese (2006), de Vries, de Haan, van der Wal et Mariën (2009), de Haan, van der Wal et de Vries (2009), Shi, Young et Hargreaves (2009), et Grimes et Young (2010).

Dans le présent article, nous décrivons une autre méthode fondée sur les évaluations foncières pour mesurer la variation des prix des logements. Les évaluations foncières servent d'information auxiliaire dans un cadre d'estimation par la *régression généralisée* (GREG). La régression GREG est une technique assistée par modèle qui peut être utilisée pour augmenter l'efficacité comparativement aux estimateurs plus simples tels que les moyennes d'échantillon (Särndal, Swensson et Wretman 1992), à condition que l'information sur la population soit connue pour une ou plusieurs variables présentant une forte corrélation linéaire avec la variable étudiée. Dans notre cas, nous calculons la régression des prix de vente à chaque période sur les évaluations foncières. Les valeurs d'évaluation sont disponibles aux Pays-Bas pour tous les biens faisant partie du parc immobiliers durant une période de référence donnée, et nous nous attendons à ce qu'elles présentent une forte colinéarité avec les prix de vente. Bien que la méthode repose sur la régression, l'indice des prix résultant n'est pas un indice hédonique, car le modèle de régression est descriptif plutôt qu'explicatif.

La présentation de l'article est la suivante. Pour préparer le terrain, à la section 2, nous décrivons la méthode SPAR et les liens entre cette dernière et les moyennes d'échantillon des prix de vente ainsi que les évaluations foncières. En raison du changement de composition et du nombre relativement faible de transactions, la série SPAR des Pays-Bas est caractérisée par une forte volatilité, surtout pour les petits créneaux du marché. À la section 3, nous décrivons un estimateur GREG simple et deux options de rechange. La première option est une version stratifiée de l'indice original, tandis que la deuxième repose sur une autre spécification du modèle. À la section 4, nous présentons des constatations empiriques obtenues en utilisant les données des Pays-Bas. Les indices GREG s'avèrent fort semblables aux indices SPAR et sont tout aussi volatils. À la section 5, nous expliquons ce résultat en montrant que l'indice SPAR est en fait un estimateur de l'indice GREG et qu'il est presque aussi efficace. À la section 6, nous présentons nos conclusions et proposons un sujet de future étude dans ce domaine.

## 2 Estimateurs de Horvitz-Thompson et l'indice SPAR

Habituellement, l'objectif de l'échantillonnage est d'estimer le total ou la moyenne (arithmétique) d'une variable donnée pour une population finie. Dans le contexte du logement, nous pourrions vouloir estimer la valeur totale du parc de logements, disons, à la période 0. Soit  $U^0$  le parc de logements de taille  $N^0$  et  $p_n^0$  la valeur du logement  $n$  ( $n = 1, \dots, N^0$ ). La valeur cible que l'on veut estimer est

$$V^0 = \sum_{n \in U^0} p_n^0. \quad (2.1)$$

Supposons que nous ayons un échantillon  $S^0$  comprenant  $n^0$  logements vendus durant la période de référence. Si les logements sont sélectionnés par échantillonnage aléatoire simple dans le parc de

logements  $U^0$ , où chaque logement possède la même probabilité d'inclusion, alors l'estimateur de Horvitz-Thompson

$$\hat{V}^0 = (N^0/n^0) \sum_{n=1}^{n^0} p_n^0 \quad (2.2)$$

est un estimateur sans biais de (2.1); voir, par exemple, Cochran (1977).

Une cible naturelle – quoi qu'il ne s'agisse pas de la seule possibilité – pour un indice des prix des logements serait la variation de la valeur d'un parc de logements fixe. Le conditionnement sur le *parc à la période de référence* a deux implications : les ajouts au parc (principalement des logements neufs) doivent être exclus et les variations des biens immobiliers existants doivent être ajustées pour tenir compte des changements de qualité, c'est-à-dire l'effet de la dépréciation, des rénovations et des rajouts. Pour simplifier, nous supposons que ces changements de qualité sont négligeables. Dans ces conditions, l'indice des prix cible en passant de la période de référence 0 à la période de comparaison  $t (> 0)$  est défini comme

$$P^{0t} = \frac{\sum_{n \in U^0} p_n^t}{\sum_{n \in U^0} p_n^0}, \quad (2.3)$$

la notation étant évidente. Supposons que nous ayons aussi un échantillon  $S^t$ , constitué de  $n^t$  logements vendus à la période  $t$  et supposons qu'il s'agit d'un tirage aléatoire indépendant fait dans le parc de logements à la période de référence. Le ratio des estimateurs de Horvitz-Thompson (les moyennes d'échantillon) aux deux périodes

$$\hat{P}^{0t} = \frac{(N^0/n^t) \sum_{n \in S^t} p_n^t}{(N^0/n^0) \sum_{n \in S^0} p_n^0} = \frac{\sum_{n \in S^t} p_n^t/n^t}{\sum_{n \in S^0} p_n^0/n^0} \quad (2.4)$$

peut sembler être un estimateur naturel de notre indice cible (2.3). Toutefois, si les échantillons  $S^0$  et  $S^t$  sont tirés indépendamment, la variance de l'estimateur (2.4) risque d'être considérable. En outre, un ratio estimé tel que (2.4) présente un biais qui dépend de la variance du numérateur et de la covariance du numérateur et du dénominateur (Cochran 1977). Dans la perspective d'un indice, le problème important est que la composition des biens négociés à la période  $t$  n'est pas la même qu'à la période 0. Autrement dit, nous ne comparons pas les mêmes choses.

L'approche classique d'estimation des indices de prix repose sur les méthodes d'appariement de modèles où les prix  $p_n^0$  et  $p_n^t$  sont observés pour un panel fixe d'articles. L'utilisation de données de panel permet de s'assurer que l'on compare des articles qui sont les mêmes, ce qui réduit la variance de l'estimateur par le ratio, parce que  $p_n^0$  et  $p_n^t$  sont habituellement corrélés positivement. Cependant, à moins que les échantillons  $S^0$  et  $S^t$  soient extraordinairement grands, on n'obtient que quelques appariements de logements, si tant qu'il y en ait. Donc, alors que les prix  $p_n^t$  sont observés pour les logements appartenant à  $S^t$ , les prix à la période de référence  $p_n^0$  « manquent » pour la plupart de ces

logements. Les données qui pourraient par contre être disponibles sont les évaluations foncières de l'administration publique  $a_n^0$ . Nous pourrions utiliser ces évaluations comme valeurs à la période de référence et construire l'estimateur par « pseudo » appariement de modèles qui suit de la variation des prix des logements :

$$\tilde{P}^{0t} = \frac{\sum_{n \in S^t} p_n^t / n^t}{\sum_{n \in S^t} a_n^0 / n^t}. \quad (2.5)$$

Un problème que pose l'estimateur (2.5) est que l'indice à la période de référence ne sera pas égal à 1, parce que les évaluations  $a_n^0$  diffèrent des prix de vente  $p_n^0$ . Le rééchantillonnage de l'estimateur (2.5) en le divisant par sa valeur à la période de référence est une solution évidente, qui donne

$$\hat{P}_{\text{SPAR}}^{0t} = \frac{\sum_{n \in S^t} p_n^t / n^t}{\sum_{n \in S^t} a_n^0 / n^t} \left[ \frac{\sum_{n \in S^0} p_n^0 / n^0}{\sum_{n \in S^0} a_n^0 / n^0} \right]^{-1} = \frac{\sum_{n \in S^t} p_n^t / n^t}{\sum_{n \in S^0} p_n^0 / n^0} \left[ \frac{\sum_{n \in S^0} a_n^0 / n^0}{\sum_{n \in S^t} a_n^0 / n^t} \right]. \quad (2.6)$$

Notons que le facteur de rééchantillonnage est stochastique, car il s'agit d'un ratio de moyennes d'échantillon pour la période de référence, ce qui augmentera la variance de (2.6) comparativement à l'estimateur donné par (2.5), en fonction des corrélations entre les évaluations foncières et les prix de vente. Des renseignements détaillés figurent dans de Haan (2007). Cependant, nous ne pouvons pas contourner le rééchantillonnage, puisqu'un indice de prix dont la valeur initiale n'est pas égale à 1 n'aurait pas de sens.

L'expression (2.6) est appelée indice du ratio prix de vente-évaluation (*sale price appraisal ratio*) ou indice SPAR. La méthode SPAR est appliquée aux Pays-Bas depuis janvier 2008 pour mesurer le changement de prix des logements occupés par le propriétaire. Comme il est mentionné plus haut, nous supposons que l'indice SPAR a pour objectif de suivre l'évolution du prix du *parc de logements*, qui est une mesure de la variation du patrimoine. Par ailleurs, dans le contexte de l'Indice harmonisé des prix à la consommation, l'indice des prix des logements doit mesurer le changement de prix des *logements vendus* durant la période de référence (Makaronidis et Hayes 2006; Eurostat 2010). Sous ce dernier concept, aucun échantillonnage ne doit être effectué si toutes les transactions sont enregistrées et utilisées dans le calcul de l'indice, comme cela est le cas aux Pays-Bas.

Le deuxième membre de l'équation (2.6) exprime l'indice SPAR sous la forme du produit de deux facteurs, le ratio des moyennes d'échantillon et un facteur entre crochets. Comme l'indice SPAR est essentiellement fondé sur la méthode d'appariement de modèles (en utilisant des évaluations à la place des prix de vente à la période de référence), ce facteur rajuste le ratio des moyennes d'échantillon pour tenir compte des changements de composition qualitative des échantillons qui ont lieu entre la période 0 et la période  $t$ . Un problème éventuel est que l'indice SPAR *n'est pas un estimateur de type panel*. Par conséquent, une série chronologique SPAR, disons pour les périodes  $t = 0, \dots, T$ , pourrait souffrir d'une volatilité dans le court terme due à des changements de composition, surtout si le nombre de ventes est faible.

### 3 Estimation par la régression généralisée

#### 3.1 Une méthode GREG simple

A la présente section, nous décrivons une approche de rechange pour mesurer le changement de prix des logements qui s'appuie sur des données d'évaluation. Les évaluations foncières servent maintenant d'information auxiliaire dans un cadre de régression généralisée (GREG). Considérons le simple modèle de régression linéaire à deux variables suivant :

$$p_n^0 = \alpha^0 + \beta^0 a_n^0 + \varepsilon_n^0, \quad (3.1)$$

où  $\varepsilon_n^0$  est le terme d'erreur. Contrairement aux modèles de régression hédonique, qui postulent une relation causale entre le prix de vente  $p_n^0$  et un jeu de caractéristiques ayant trait à la structure et à la localisation des unités de logement, ce modèle ne dit rien sur la façon dont les prix des logements sont produits; l'équation (3.1) est simplement un modèle descriptif.

L'estimation du modèle (3.1) par la régression par les moindres carrés sur les données de l'échantillon  $S^0$  donne les prix prédits

$$\hat{p}_n^0 = \hat{\alpha}^0 + \hat{\beta}^0 a_n^0. \quad (3.2)$$

Les résidus de la régression pour  $n \in S^0$  sont  $e_n^0 = p_n^0 - \hat{p}_n^0$ . En supposant un échantillonnage aléatoire, comme auparavant, nous pouvons écrire l'estimateur de Horvitz-Thompson  $\sum_{n \in S^0} p_n^0 / n^0$  de la valeur moyenne  $\sum_{n \in U^0} p_n^0 / N^0$  sous la forme

$$\sum_{n \in S^0} p_n^0 / n^0 = \sum_{n \in S^0} \hat{p}_n^0 / n^0 + \sum_{n \in S^0} e_n^0 / n^0 = \hat{\alpha}^0 + \hat{\beta}^0 \sum_{n \in S^0} a_n^0 / n^0 + \sum_{n \in S^0} e_n^0 / n^0. \quad (3.3)$$

Le remplacement de la moyenne d'échantillon des évaluations,  $\sum_{n \in S^0} a_n^0 / n^0$ , par son équivalent pour la population,  $\sum_{n \in U^0} a_n^0 / N^0$  donne l'estimateur par la régression généralisée (GREG) :

$$\hat{p}_{\text{GREG}}^0 = \hat{\alpha}^0 + \hat{\beta}^0 \sum_{n \in U^0} a_n^0 / N^0 + \sum_{n \in S^0} e_n^0 / n^0 = \sum_{n \in U^0} \hat{p}_n^0 / N^0 + \sum_{n \in S^0} e_n^0 / n^0. \quad (3.4)$$

La théorie de l'échantillonnage assisté par modèle montre que les estimateurs GREG sont *asymptotiquement sans biais sous le plan de sondage* (Särndal et coll. 1992), quel que soit le choix des variables explicatives. À moins que l'échantillon soit petit, le biais peut être négligé. Il est évident que l'estimateur GREG (3.4) sera plus efficace – au sens où sa variance est plus faible – que l'estimateur de Horvitz-Thompson (3.3). Par conséquent, l'estimateur GREG donnera habituellement de meilleurs résultats que l'estimateur de Horvitz-Thompson en termes d'erreur quadratique moyenne (la somme de la variance et du carré du biais).

La même procédure peut être appliquée à la période de comparaison  $t$ . Après avoir estimé le modèle

$$p_n^t = \alpha^t + \beta^t a_n^0 + \varepsilon_n^t \quad (3.5)$$

par la régression par les moindres carrés sur les données de l'échantillon de la période courante  $S^t$ , nous obtenons les prix prédits



$$\hat{p}_n^t = \hat{\alpha}^t + \hat{\beta}^t a_n^0, \quad (3.6)$$

ce qui mène à l'estimateur GREG de la valeur moyenne du parc de logements à la période  $t$  :

$$\hat{p}_{\text{GREG}}^t = \hat{\alpha}^t + \hat{\beta}^t \sum_{n \in U^t} a_n^0 / N^t + \sum_{n \in S^t} e_n^t / n^t = \sum_{n \in U^t} \hat{p}_n^t / N^t + \sum_{n \in S^t} e_n^t / n^t, \quad (3.7)$$

où  $e_n^t = p_n^t - \hat{p}_n^t$  désigne les résidus de la régression à la période  $t$ . Pour un parc de logements fixe, nous avons  $U^t = U^0$ , d'où  $\sum_{n \in U^t} a_n^0 / N^t = \sum_{n \in U^0} a_n^0 / N^0$ , et il s'ensuit que

$$\hat{p}_{\text{GREG}}^t = \hat{\alpha}^t + \hat{\beta}^t \sum_{n \in U^0} a_n^0 / N^0 + \sum_{n \in S^t} e_n^t / n^t = \sum_{n \in U^0} \hat{p}_n^t / N^0 + \sum_{n \in S^t} e_n^t / n^t. \quad (3.8)$$

L'estimateur GREG du changement de prix des logements s'obtient simplement en prenant le ratio des équations (3.8) et (3.4):

$$\hat{p}_{\text{GREG}}^{0t} = \frac{\hat{p}_{\text{GREG}}^t}{\hat{p}_{\text{GREG}}^0} = \frac{\hat{\alpha}^t + \hat{\beta}^t \bar{a}^0 + \sum_{n \in S^t} e_n^t / n^t}{\hat{\alpha}^0 + \hat{\beta}^0 \bar{a}^0 + \sum_{n \in S^0} e_n^0 / n^0} = \frac{\sum_{n \in U^0} \hat{p}_n^t / N^0 + \sum_{n \in S^t} e_n^t / n^t}{\sum_{n \in U^0} \hat{p}_n^0 / N^0 + \sum_{n \in S^0} e_n^0 / n^0}, \quad (3.9)$$

où  $\bar{a}^0 = \sum_{n \in U^0} a_n^0 / N^0$ . Un certain biais de petit échantillon supplémentaire sera introduit en raison de la structure non linéaire (ratio). Lorsque l'on utilise la régression par les moindres carrés ordinaires (MCO) pour estimer les modèles (3.1) et (3.5), les moyennes d'échantillon non pondérées des résidus de la régression dans (3.9),  $\sum_{n \in S^0} e_n^0 / n^0$  et  $\sum_{n \in S^t} e_n^t / n^t$ , sont égales à 0 et l'indice GREG se réduit à

$$\hat{p}_{\text{GREG,MCO}}^{0t} = \frac{\sum_{n \in U^0} \hat{p}_n^t / N^0}{\sum_{n \in U^0} \hat{p}_n^0 / N^0} = \frac{\hat{\alpha}^t + \hat{\beta}^t \bar{a}^0}{\hat{\alpha}^0 + \hat{\beta}^0 \bar{a}^0} = \frac{\hat{\alpha}^t / \bar{a}^0 + \hat{\beta}^t}{\hat{\alpha}^0 / \bar{a}^0 + \hat{\beta}^0}. \quad (3.10)$$

Comme l'indique la première expression dans le membre de droite de l'équation (3.10), l'approche GREG (MCO) consiste essentiellement à imputer les prix pour la période de référence et pour la période courante en utilisant les équations (3.2) et (3.6). La différence par rapport à la méthode hédonique d'*imputation double* tient à deux aspects : nous utilisons un modèle descriptif, et non un modèle hédonique, pour estimer les prix prédits – de sorte que nous ne pouvons pas parler de prix prédits sans biais – et nous imputons les prix de tous les logements faisant partie du parc au lieu de ceux du sous-ensemble de logements échantillonnés.

## 3.2 Propriétés de l'indice GREG

L'indice GREG (MCO) possède plusieurs propriétés intéressantes. Premièrement, son calcul est très simple. Une fois que l'on a calculé la moyenne de population des évaluations foncières  $\bar{a}^0$  et les coefficients de régression pour la période de référence  $\hat{\alpha}^0$  et  $\hat{\beta}^0$ , il suffit d'exécuter chaque mois une

régression des prix de vente en fonction des évaluations, puis d'introduire les valeurs des coefficients  $\hat{\alpha}^t$  et  $\hat{\beta}^t$  dans (3.10). Notons que l'indice GREG peut s'écrire sous la forme d'un pseudo indice-chaîne :

$$\hat{P}_{\text{GREG,MCO}}^{0t} = \frac{\hat{\alpha}^t / \bar{a}^0 + \hat{\beta}^t}{\hat{\alpha}^0 / \bar{a}^0 + \hat{\beta}^0} = \prod_{\tau=1}^t \frac{\hat{\alpha}^\tau / \bar{a}^0 + \hat{\beta}^\tau}{\hat{\alpha}^{\tau-1} / \bar{a}^0 + \hat{\beta}^{\tau-1}}. \quad (3.11)$$

Cela peut être utile en pratique, surtout quand de nouvelles données d'évaluation deviennent disponibles. Les nouvelles évaluations sont souvent fournies à l'organisme statistique avec un délai important, pouvant dépasser un an. Les évaluations les plus récentes doivent être utilisées pour deux raisons. La qualité des évaluations peut s'améliorer avec le temps, ce qui semble avoir été le cas aux Pays-Bas (de Vries et coll. 2009). En outre, l'hypothèse d'un parc de logements fixe peut être relâchée afin que les logements nouvellement construits puissent être intégrés par enchaînement; l'indice GREG chaîné tient compte de la dynamique du parc de logements. Les mêmes avantages de l'enchaînement s'appliquent à la méthode SPAR. Supposons que de nouvelles évaluations foncières, se rapportant à la période  $T$  ( $0 < T \leq t$ ), soient disponibles à la période  $t + 1$ . La série chronologique peut alors être mise à jour par enchaînement, c'est-à-dire en multipliant  $\hat{P}_{\text{GREG,MCO}}^{0t}$  par la variation d'un mois à l'autre  $(\tilde{\alpha}^{t+1} / \bar{a}^T + \tilde{\beta}^{t+1}) / (\tilde{\alpha}^t / \bar{a}^T + \tilde{\beta}^t)$ , où les coefficients sont maintenant ceux d'une régression des prix de vente sur les évaluations foncières à la période  $T$ .

Deuxièmement, les *erreurs-types* de l'indice GREG peuvent être estimées assez facilement en utilisant la matrice de variance-covariance des coefficients de régression, qui est une sortie standard de la plupart des progiciels statistiques. Une expression de l'erreur-type approximative est dérivée en annexe. L'erreur-type de l'indice GREG dépend de la qualité de l'ajustement ( $R^2$ ) du modèle de régression. Il est fort probable que la valeur de  $R^2$  pour la régression à la période de référence soit plus élevée que pour les régressions à la période courante. Nous nous attendons en effet à observer une forte relation linéaire entre les évaluations foncières et les prix de vente à la période de référence des évaluations, mais une relation probablement plus faible aux périodes ultérieures en raison des différences de tendance des prix selon le type de logement ou la région. Il est un peu plus compliqué d'établir une expression pour les erreurs-types approximatives dans le cas de l'indice SPAR, parce que la variabilité d'échantillonnage des évaluations moyennes est une source additionnelle d'erreur d'échantillonnage; voir de Haan (2007).

Cette dernière remarque nous mène à la troisième propriété de l'indice GREG, c'est-à-dire sa dépendance à l'égard de la *qualité des données d'évaluation*. Pour au moins deux raisons, il peut arriver que les évaluations foncières ne représentent pas exactement les prix de transaction durant la période de référence, de sorte que l'ajustement du modèle n'est pas parfait ( $R^2 < 1$ ). Les organismes chargés des évaluations pourraient ne pas avoir accès (en temps réel) aux prix de vente réels et, par conséquent, être obligés d'exercer leur propre jugement en se basant sur d'autres renseignements. Toutefois, même s'ils connaissaient les prix de vente, ces organismes pourraient encore décider de faire des ajustements lorsqu'ils déterminent la valeur des biens immobiliers. On peut soutenir que le prix de vente ne mesure pas toujours correctement la valeur de marché inconnue – laquelle peut être considérée comme une variable latente – et a tendance à être plus volatile. À cet égard, Francke (2010) et d'autres ont utilisé le terme de bruit de transaction.

La manière dont les évaluations foncières ont été déterminées aura une incidence sur l'erreur-type de l'indice GREG. À condition que la qualité des données d'évaluation soit la même pour tous les logements

compris dans le parc, il n'existe aucun biais, puisque les évaluations servent seulement de variables auxiliaires dans les régressions exécutées sur les échantillons  $S^0$  et  $S^t$  de biens immobiliers vendus aux périodes 0 et  $t$  ( $t = 1, \dots, T$ ). Cependant, en général, nous nous attendons à ce que la qualité des évaluations soit meilleure pour les biens appartenant à l'échantillon de la période de référence où a eu lieu l'évaluation  $S^0$ , quoique cela varie fort probablement en fonction de la méthode d'évaluation. Aux Pays-Bas, les biens immobiliers sont évalués aux fins de l'impôt (impôt sur le revenu ainsi que les impôts municipaux). Les municipalités sont chargées des évaluations. Plusieurs d'entre elles évaluent les logements qui sont vendus durant la période de référence (janvier) au moyen du prix de vente. Les logements qui n'ont pas été vendus sont parfois évalués en les comparant à des logements négociés similaires. Il semble que certaines municipalités utilisent une forme de régression hédonique pour évaluer les logements, mais la méthodologie n'a malheureusement pas été rendue publique. Pour plus de renseignements sur le système d'évaluation foncière des Pays-Bas, voir de Vries et coll. (2009).

Jusqu'à présent, nous avons supposé que la qualité des logements individuels ne varie pas au fil du temps. Cette hypothèse est forte. Donc, la quatrième propriété – et l'inconvénient le plus important – de la méthode GREG est que l'indice des prix résultants est entaché d'un *biais de changement de qualité* puisque l'on n'effectue pas d'ajustement explicite de la qualité. La méthode SPAR ainsi que la méthode classique fondée sur les ventes répétées présentent le même inconvénient. En principe, les méthodes de régression hédonique permettent de traiter le problème du changement de qualité, quoi qu'il puisse s'avérer difficile d'utiliser des variables de contrôle pour toutes les caractéristiques influant sur le prix pertinentes, en particulier la microlocalisation. La méthode SPAR tient compte automatiquement de la microlocalisation, à condition naturellement que les évaluations foncières en tiennent suffisamment compte, puisqu'elle est basée sur la méthode d'appariement de modèles pour laquelle l'appariement est effectué au niveau de l'adresse.

### 3.3 Estimateur GREG de rechange

Statistics Netherlands calcule les indices des prix des logements non seulement pour l'ensemble du pays, mais aussi pour certains créneaux du marché du logement, selon le type de logement (logements familiaux et appartements) et la région (provinces et grandes villes), principalement pour répondre aux besoins des utilisateurs. L'échantillon peut aussi être stratifié afin d'atténuer l'effet du *biais de sélection dans l'échantillon*. Ce type de biais peut survenir si l'ensemble de logements vendus durant une période particulière n'est pas une sélection aléatoire provenant du parc de logements. L'indice national doit alors être calculé indirectement sous forme d'une moyenne pondérée des indices de strate plutôt que directement d'après toutes les observations.

Supposons que le parc total de logements  $U^0$  est subdivisé en  $K$  strates non chevauchantes  $U_k^0$  de taille  $N_k^0$  ( $\sum_{k=1}^K N_k^0 = N^0$ ). L'indice des prix cible (2.3) peut alors être réécrit sous la forme

$$P^{0t} = \frac{\sum_{n \in U^0} p_n^t}{\sum_{n \in U^0} p_n^0} = \frac{\sum_{k=1}^K \sum_{n \in U_k^0} p_n^t}{\sum_{k=1}^K \sum_{n \in U_k^0} p_n^0} = \sum_{k=1}^K S_k^0 P_k^{0t}, \quad (3.12)$$

où  $P_k^{0t} = \sum_{n \in U_k^0} p_n^t / \sum_{n \in U_k^0} p_n^0$  est l'indice des prix cible pour la strate  $U_k^0$  ( $k = 1, \dots, K$ ). Les parts de la valeur du parc de logements à la période de référence  $s_k^0 = \sum_{n \in U_k^0} p_n^0 / \sum_{n \in U^0} p_n^0$ , qui servent de pondérations pour les indices de strate, sont inconnues et doivent être estimées. En supposant que l'on connaît les variables qui définissent les strates pour tout  $n \in U^0$ , un choix naturel pour les pondérations serait les parts fondées sur l'évaluation foncière  $\hat{s}_k^0 = \sum_{n \in U_k^0} a_n^0 / \sum_{n \in U^0} a_n^0 = (N_k^0 / N^0) (\bar{a}_k^0 / \bar{a}^0)$ . Manifestement, les variables de logement qui définissent les strates doivent être incluses dans le jeu de données d'évaluation. Aux Pays-Bas, l'adresse et le type de logement sont inclus. Cela permet une subdivision de la population en strates obtenues par classification croisée de la localisation et du type de logement. Les évaluations foncières ne sont peut-être pas toujours des estimations exactes de la valeur de marché « réelle » des biens immobiliers individuels, mais au niveau de la strate, nous nous attendons à ce que l'exactitude des évaluations moyennes soit suffisante pour le calcul des pondérations.

Des techniques statistiques telles que l'estimation GREG sont habituellement appliquées pour estimer les totaux ou les moyennes pour de petits domaines pour lesquels le nombre d'observations est si faible que les erreurs-types lorsque l'on utilise les estimateurs classiques (de Horvitz-Thompson) – ici le ratio des moyennes d'échantillon – deviendraient inacceptablement grandes. Il convient de mentionner que, même avec la méthode GREG, le schéma de stratification ne doit pas être trop détaillé, car cela pourrait accroître excessivement la variance des indices de strate, et donc, de l'indice agrégé. Fait peut-être encore plus important, le biais de petit échantillon augmentera au point de devenir éventuellement non négligeable pour les très petits échantillons.

Les régressions par les MCO des prix de vente sur les évaluations foncières doivent maintenant être exécutées à chaque période pour chaque strate afin de calculer l'indice GREG agrégé. L'indice GREG (MCO) stratifié est donné par

$$\hat{P}_{\text{StrGREG}}^{0t} = \sum_{k=1}^K \hat{s}_k^0 \hat{P}_{k,\text{GREG,MCO}}^{0t} = \sum_{k=1}^K \hat{s}_k^0 \left( \frac{\hat{\alpha}_k^t / \bar{a}_k^0 + \hat{\beta}_k^t}{\hat{\alpha}_k^0 / \bar{a}_k^0 + \hat{\beta}_k^0} \right); \quad (3.13)$$

Les écarts entre les coefficients de pente  $\hat{\beta}_k^s$  ( $s = 0, t$ ) d'une strate à l'autre pourraient résulter de l'erreur d'échantillonnage ou refléter un phénomène réel. Celui-ci peut avoir une importance particulière pour les périodes  $t$  très éloignées de la période 0, car les différents créneaux du marché du logement ont tendance à présenter des tendances des prix variables. On pourrait effectuer un test afin de savoir si tout écart entre les coefficients de pente reflète un phénomène réel.

Un modèle de rechange, à estimer sur le jeu complet de données, comprendrait un terme d'ordonnée à l'origine unique, mais des coefficients  $\beta$  pouvant varier d'une strate à l'autre. Soit  $D_{n,k}$  une variable indicatrice binaire qui prend la valeur 1 si le bien immobilier  $n$  appartient à la strate  $k$  et 0 autrement. À la période  $s$  ( $s = 0, t$ ), le modèle

$$p_n^s = \alpha^s + \sum_{k=1}^K \beta_k^s D_{n,k} a_n^0 + \varepsilon_n^s \quad (3.14)$$

est estimé par la régression par les MCO sur les données de l'échantillon  $S^s$ , ce qui donne les prix prédits  $\tilde{p}_n^s = \tilde{\alpha}^s + \tilde{\beta}_k^s a_n^0$  pour  $n \in U_k^0$ . De nouveau, la somme des résidus est égale à zéro et le nouvel indice GREG (OMC) (non stratifié) devient

$$\tilde{P}_{\text{GREG, MCO}}^{0t} = \frac{\sum_{n \in U^0} \tilde{P}_n^t / N^0}{\sum_{n \in U^0} \tilde{P}_n^0 / N^0} = \frac{\sum_{k=1}^K \sum_{n \in U_k^0} \tilde{P}_n^t / N^0}{\sum_{k=1}^K \sum_{n \in U_k^0} \tilde{P}_n^0 / N^0} = \frac{\tilde{\alpha}^t + \sum_{k=1}^K \left( \frac{N_k^0}{N^0} \right) \tilde{\beta}_k^t \bar{a}_k^0}{\tilde{\alpha}^0 + \sum_{k=1}^K \left( \frac{N_k^0}{N^0} \right) \tilde{\beta}_k^0 \bar{a}_k^0}. \quad (3.15)$$

Le modèle (3.14) est plus souple que le modèle original donnée par les équations (3.1) et (3.5), et pourrait être utile si la proportionnalité entre les prix de vente et les évaluations foncières n'est pas respectée. L'estimateur (3.15) se réduit à l'indice GREG original (3.10) si les coefficients  $\tilde{\beta}_k^s$  sont tous égaux. En pratique, cela n'arrivera pas et (3.15) et (3.10) donneront des réponses différentes. Une raison fréquemment avancée pour justifier l'utilisation des estimateurs GREG est que, étant asymptotiquement sans biais, ils sont relativement *robustes au choix du modèle*. Donc, nous nous attendrions à ce que l'effet de la spécification du modèle de rechange (3.15) soit modéré. Par ailleurs, il est généralement reconnu dans la littérature que l'indépendance à l'égard du modèle peut être un problème dans des circonstances particulières, notamment lorsqu'on a affaire à des populations très variables et ayant tendance à présenter des valeurs aberrantes. Par exemple, Hedlin, Falvey, Chambers et Kopic (2001) soulignent qu'il est important de procéder à une recherche minutieuse des spécifications du modèle, tandis que Beaumont et Alavi (2004) se concentrent sur le traitement des valeurs aberrantes. Il serait donc utile d'examiner l'effet de la spécification de ce modèle de rechange.

## 4 Illustration empiriques

Pour l'étude empirique, nous avons utilisé deux jeux de données de sources différentes. Le premier contient les prix de vente pour presque toutes les transactions concernant des logements existants (à l'exclusion des logements neufs) aux Pays-Bas effectuées entre janvier 2003 et mars 2009, telles qu'elles ont été enregistrées par le bureau du cadastre des Pays-Bas. Le nombre total d'observations est de 1 126 242, soit environ 15 000 par mois. Les ventes ont été enregistrées au moment de la signature de l'acte de vente au cabinet du notaire, en moyenne six semaines après la signature du compromis de vente. Le deuxième jeu de données contient les évaluations foncières de l'administration publique ayant trait à janvier 2003 pour tous les logements occupés par le propriétaire compris dans le parc de logements. Comme les adresses sont disponibles dans les deux jeux de données, nous connaissons le prix de vente et la valeur d'évaluation pour chaque transaction. Comme le type de logement est également disponible, nous avons pu stratifier l'échantillon selon le type et la localisation du logement.

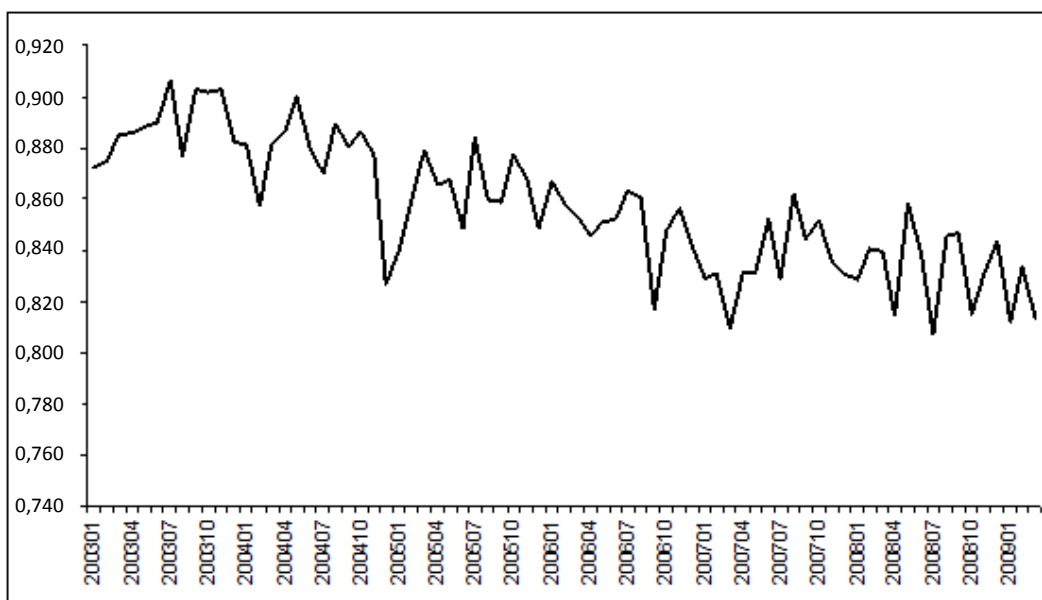
Pour commencer, nous avons exécuté des régressions par les MCO non stratifiées des prix de vente sur les évaluations foncières, en utilisant le modèle (3.8), pour chacun des 75 mois. Certains de ces résultats sont présentés au tableau 4.1; les résultats empiriques détaillés peuvent être obtenus sur demande auprès des auteurs. Évidemment, les coefficients  $\hat{\beta}^t$  ne sont pas nuls à de très faibles seuils de signification. Dans la plupart des cas, les ordonnées à l'origine  $\hat{\alpha}^t$  diffèrent de manière significative de zéro au seuil de

signification de 5 %. Environ 80 % à 90 % de la variation des prix de vente est « expliquée » par la variation des évaluations, comme l'indique les valeurs de  $R^2$ . Autrement dit, le coefficient de corrélation entre les prix de vente et les évaluations foncières à la période de référence varie de 0,89 à 0,95. La figure 4.1 montre que  $R^2$  diminue légèrement au cours du temps. Comme nous l'avons mentionné plus haut, cela pourrait tenir au fait que les prix varient différemment dans différents créneaux du marché. Nous avons été un peu surpris de constater que la valeur de  $R^2$  n'était pas la plus élevée en janvier 2003, qui est la période de référence.

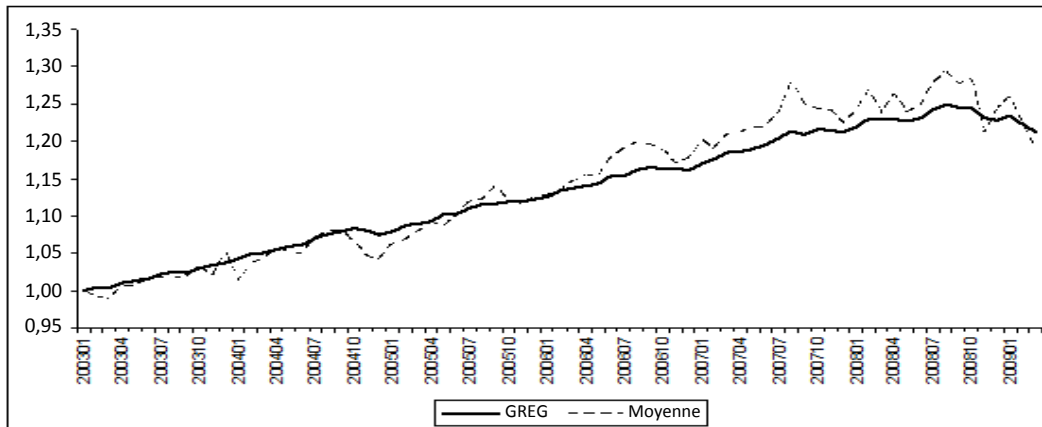
Sur la base des résultats de régression susmentionnés, nous avons calculé les indices de prix GREG en utilisant l'équation (3.10). De janvier 2003 jusqu'au milieu 2008, les prix des logements ont augmenté de quelque 25 % aux Pays-Bas, puis ont commencé à baisser, probablement en raison de la crise financière et économique. Fait important, l'indice GREG est beaucoup plus lisse que le simple ratio des moyennes d'échantillon, comme en témoigne la figure 4.2, ce qui est précisément ce pourquoi l'indice a été conçu.

**Tableau 4.1**  
**Résultats de régression**

Mois	Alpha	t	Bêta	t	R carré
Janvier 2003	1 900,49	2,26	0,98	275,19	0,87
Janvier 2004	5 039,16	5,96	1,01	269,26	0,88
Janvier 2005	-2 555,12	2,43	1,08	237,54	0,84
Janvier 2006	1 282,14	1,41	1,11	286,39	0,87
Janvier 2007	-7 567,99	6,36	1,19	243,72	0,83
Janvier 2008	11 007,39	8,48	1,26	231,93	0,83
Janvier 2009	16 677,31	9,83	1,30	184,24	0,81



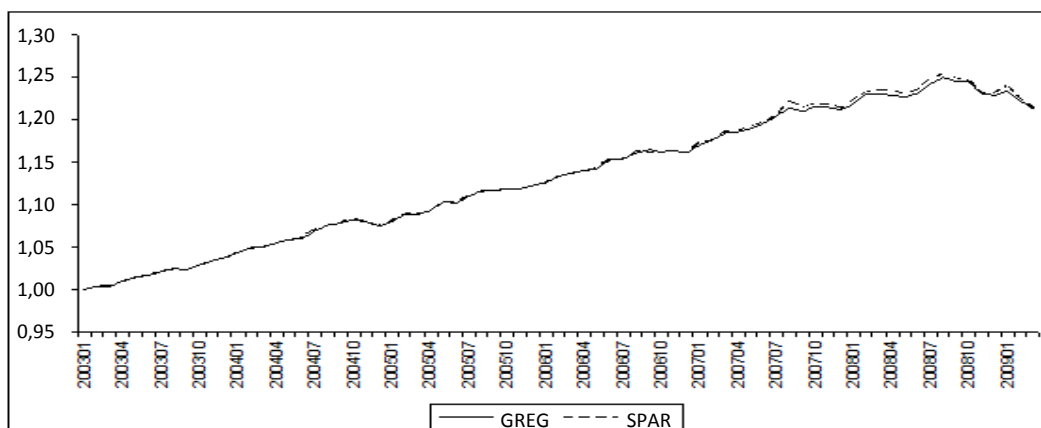
**Figure 4.1** Valeur de R carré



**Figure 4.2** Indice GREG et ratio des moyennes d'échantillon

À la figure 4.3, l'indice GREG est comparé à l'indice SPAR. En général, la tendance des deux indices est très semblable, quoiqu'il semble exister une petite différence à la fin de la période. La figure 4.4 montre que les variations d'un mois à l'autre des indices GREG et SPAR ne diffèrent pas beaucoup non plus, l'indice GREG étant juste un peu moins volatil. Donc, nous pouvons conclure qu'à l'échelle nationale, les deux méthodes produisent des résultats plus ou moins équivalents. Notons que dans les figures 4.3 et 4.4, l'indice SPAR n'est pas l'indice SPAR officiel publié par Statistics Netherlands. Nous avons calculé un indice à base fixe en utilisant les évaluations foncières pour janvier 2003 uniquement, alors que l'indice officiel est un indice-chaîne, basé sur les évaluations pour diverses périodes de référence; voir aussi la section 5.3.

Ensuite, nous avons stratifié les données en fonction des 13 provinces et de 5 types de logements, avant d'exécuter les régressions par les MCO pour chaque mois pour les 65 strates résultantes et de calculer les indices GREG ainsi que les ratios des moyennes d'échantillon. La figure 4.5 donne les résultats pour une strate, celle des appartements dans la province de Frise. Étant donné le nombre relativement faible d'observations, on observe quelques pics importants, par exemple en septembre 2009, quand le ratio des moyennes d'échantillon a augmenté de 50 %. De nouveau, l'indice GREG est plus lisse que le ratio des moyennes d'échantillon (mais néanmoins très volatil) et étonnamment similaire à l'indice SPAR. Le même tableau se dégage pour les autres strates, de sorte que nous ne présentons pas ces résultats.



**Figure 4.3** Indices GREG et SPAR

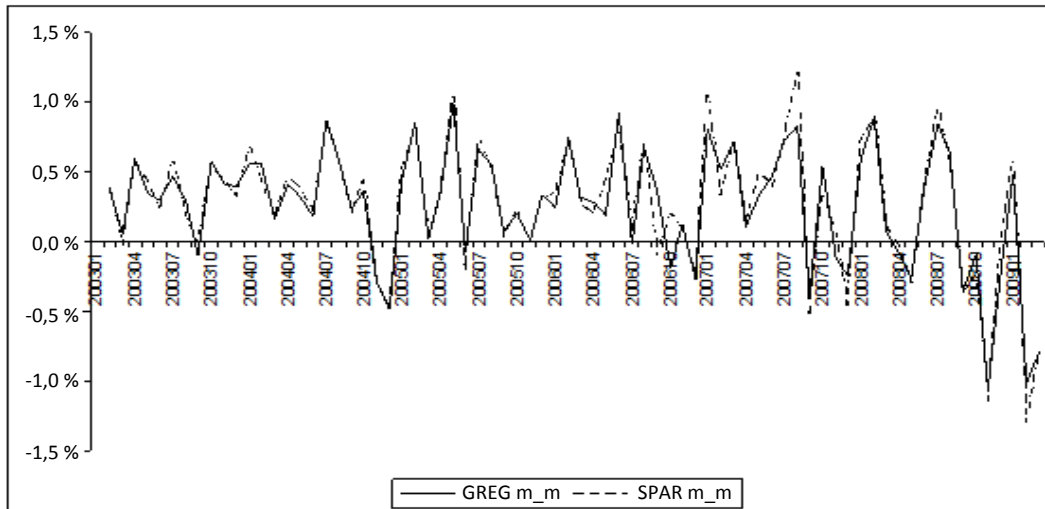


Figure 4.4 GREG et SPAR : variations en pourcentage d'un mois à l'autre

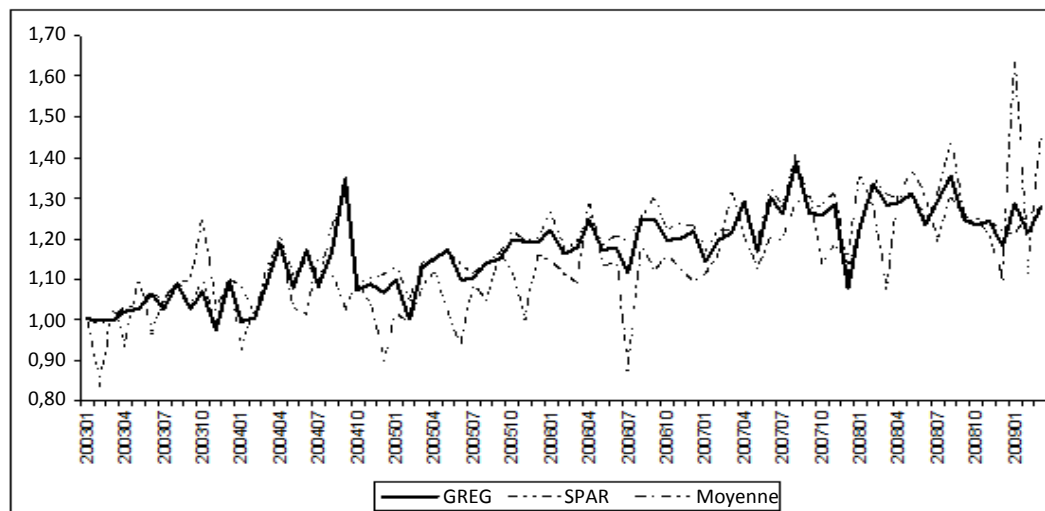


Figure 4.5 Indices GREG et SPAR et ratio des moyennes d'échantillon, appartements dans la province de Frise

Enfin, en utilisant les résultats par strate, nous avons calculé des indices GREG stratifié pour l'ensemble du pays en appliquant l'équation (3.13), où les parts de la valeur d'évaluation à la période de référence servent de pondérations pour la valeur du parc de logements. Comme le montre la figure 4.6, il n'y a pratiquement aucune différence entre les indices GREG stratifié et non stratifié, ce qui donne à penser que le biais de sélection dans l'échantillon n'est pas un problème important. La figure 4.6 montre aussi un deuxième indice de prix GREG de rechange, calculé selon l'équation (3.15), qui est fondé sur des



régressions par les MCO du modèle avec les variables indicatrices temporelles (3.14). De nouveau, les différences par rapport à l'indice GREG original paraissent faibles.

Il convient de mentionner que, même à l'intérieur des strates, certains logements étaient plus susceptibles d'être vendus que d'autres, en particulier durant la crise d'après 2008, de sorte qu'un certain biais de sélection dans l'échantillon persiste dans les indices GREG et SPAR. La direction et la grandeur de ce biais ne peuvent être prédites que si l'on dispose de données sur les caractéristiques des biens immobiliers pour estimer la probabilité que les logements soient vendus. En outre, comme nous l'avons mentionné plus haut, une stratification trop détaillée augmente à la fois la variance d'échantillonnage et le biais d'échantillon si le nombre de logements vendus est extrêmement faible, et peut accroître plutôt que réduire l'erreur quadratique moyenne des estimateurs.

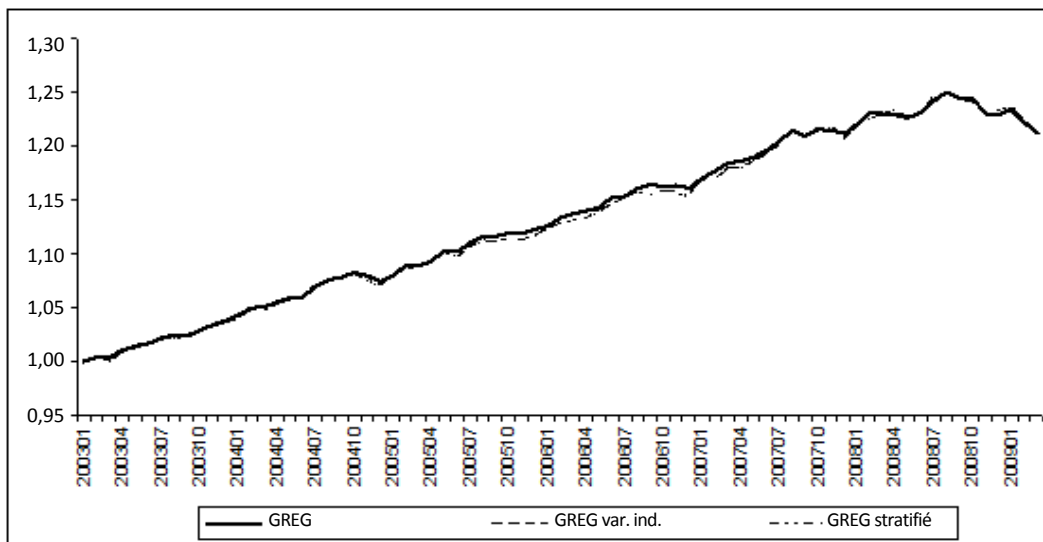


Figure 4.6 Indices GREG, GREG stratifié et GREG avec variables indicatrices

## 5 Discussion

### 5.1 Comparaisons de GREG et SPAR

La question la plus intéressante que suscite la section 4 est celle de savoir pourquoi les indices GREG et SPAR sont si semblables malgré leurs méthodes de construction très différentes. Il n'est pas étonnant que les tendances soient similaires : même si l'indice GREG ne s'appuie pas sur la méthode d'appariement de modèles, sa cible est la même que celle de l'indice SPAR. Si les tailles d'échantillon  $n^0$  et  $n^t$  s'approchaient de la taille de la population  $N^0$  – ce qui naturellement n'arrive jamais dans la réalité – les deux indices des prix s'approcheraient du changement de valeur du parc fixe de logements. Autrement dit, les deux méthodes sont asymptotiquement sans biais ou « convergentes ».

Ce qui peut paraître surprenant est que le degré de volatilité l'indice GREG au cours du temps est à peu près le même que celui de l'indice SPAR. Pour en comprendre la raison, rappelons que, sous les MCO, la somme des résidus de régression est nulle à chaque période. Cela implique que  $\sum_{n \in S^0} p_n^0/n^0 = \sum_{n \in S^0} \hat{p}_n^0/n^0$  et  $\sum_{n \in S^t} p_n^t/n^t = \sum_{n \in S^t} \hat{p}_n^t/n^t$ . Pour les modèles de régression élémentaires (3.1) et (3.5), l'indice SPAR peut donc s'écrire aussi sous la forme

$$\hat{P}_{\text{SPAR}}^{0t} = \frac{\sum_{n \in S^t} \hat{p}_n^t/n^t \left[ \frac{\sum_{n \in S^0} a_n^0/n^0}{\sum_{n \in S^t} a_n^0/n^t} \right]}{\sum_{n \in S^0} \hat{p}_n^0/n^0 \left[ \frac{\sum_{n \in S^0} a_n^0/n^0}{\sum_{n \in S^t} a_n^0/n^t} \right]} = \frac{(\hat{\alpha}^t + \hat{\beta}^t \bar{a}^{0(t)})/\bar{a}^{0(t)}}{(\hat{\alpha}^0 + \hat{\beta}^0 \bar{a}^{0(0)})/\bar{a}^{0(0)}} = \frac{\hat{\alpha}^t/\bar{a}^{0(t)} + \hat{\beta}^t}{\hat{\alpha}^0/\bar{a}^{0(0)} + \hat{\beta}^0}, \quad (5.1)$$

en utilisant (3.2) et (3.6) pour  $n \in S^0$  et  $n \in S^t$ , respectivement, où  $\bar{a}^{0(0)} = \sum_{n \in S^0} a_n^0/n^0$  et  $\bar{a}^{0(t)} = \sum_{n \in S^t} a_n^0/n^t$  pour être bref. Il existe une similarité frappante entre la dernière expression des deuxièmes membres de (5.1) et (3.10). La seule différence est que, dans l'indice SPAR (5.1), les coefficients  $\hat{\alpha}^0$  et  $\hat{\alpha}^t$  sont divisés par les moyennes d'échantillon des évaluations foncières,  $\bar{a}^{0(0)}$  et  $\bar{a}^{0(t)}$ , tandis que dans l'indice GREG (3.10), ils sont tous les deux divisés par la moyenne de population non stochastique, fixe,  $\bar{a}^0$ . Essentiellement, l'indice SPAR est un estimateur entièrement fondé sur échantillon de l'indice GREG.

Comparativement à la méthode SPAR, l'approche GREG élimine une source d'erreur d'échantillonnage, c'est-à-dire la variabilité d'échantillonnage des évaluations moyennes. Conformément à la théorie de la régression généralisée, nous nous attendrions intuitivement à ce que la méthode GREG réduise l'erreur d'échantillonnage de l'indice des prix et produise une série chronologique moins volatile (sous l'hypothèse raisonnable que  $\bar{a}^{0(t)}$  et  $\hat{\alpha}^t$  ne sont pas corrélées entre les périodes  $t = 0, \dots, T$ ). En d'autres mots, alors que la méthode GREG a été conçue comme une amélioration du ratio des moyennes d'échantillon, nous aurions pu nous attendre également à ce qu'elle joue le rôle de procédure de lissage de l'indice SPAR. Toutefois, comme nous l'avons montré à la section 4, en pratique, cela n'est guère le cas. Ce résultat peut s'expliquer comme il suit.

La réduction de la variance de l'indice GREG comparativement à l'indice SPAR dépend de la valeur des termes d'ordonnée à l'origine des régressions aux périodes 0 et  $t$ . Si les droites de régression passaient exactement par l'origine ( $\hat{\alpha}^t = \hat{\alpha}^0 = 0$ ), les indices GREG et SPAR seraient tous deux égaux au ratio des coefficients de pente  $\hat{\beta}^t/\hat{\beta}^0$  et aucune réduction de la variance n'aurait lieu. Dans le cas moins extrême où  $\hat{\alpha}^t$  et  $\hat{\alpha}^0$  sont proches de 0 et où les ratios  $\hat{\alpha}^t/\bar{a}^0$ ,  $\hat{\alpha}^t/\bar{a}^{0(t)}$ ,  $\hat{\alpha}^0/\bar{a}^0$  et  $\hat{\alpha}^0/\bar{a}^{0(0)}$  dans (3.10) et (5.2) sont très faibles comparativement à  $\hat{\beta}^t$  et  $\hat{\beta}^0$ , les indices GREG et SPAR ne différeront que légèrement, et la réduction de la variance sera marginale; voir aussi l'annexe.

Cette dernière situation est ce que l'on constate effectivement en pratique, comme le montrent les figures 5.1 et 5.2, où les valeurs de  $\hat{\alpha}^t/\bar{a}^0$  et  $\hat{\alpha}^t/\bar{a}^{0(t)}$  et celles de  $\hat{\beta}^t$  sont représentées en fonction du temps. Les ratios  $\hat{\alpha}^t/\bar{a}^0$  et  $\hat{\alpha}^t/\bar{a}^{0(t)}$  sont remarquablement similaires et petits comparativement aux  $\hat{\beta}^t$ . Bien que nous ne puissions pas ignorer ces ratios, c'est la variation de  $\hat{\beta}^t$  qui dicte principalement les indices GREG et SPAR. L'indice SPAR est non seulement un estimateur entièrement fondé sur échantillon de l'indice GREG, comme nous l'avons mentionné plus haut, mais il semble être presque aussi efficace.

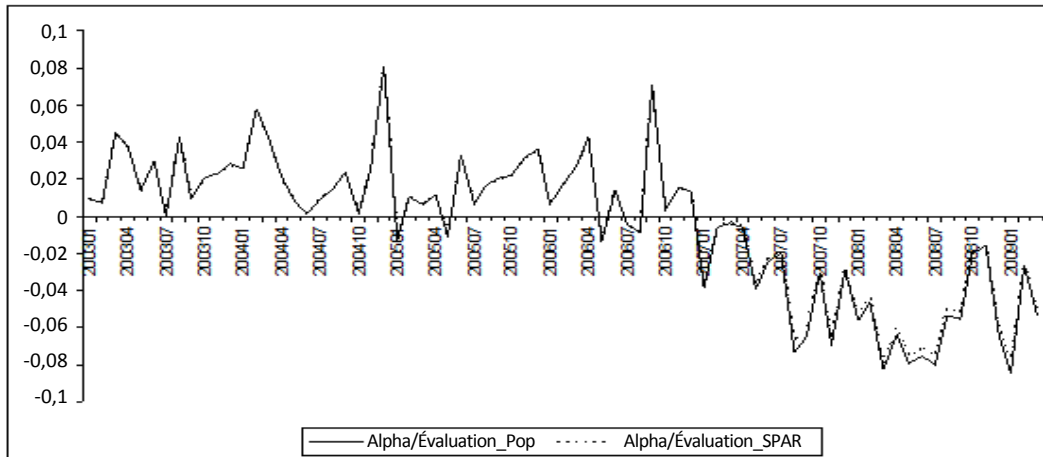


Figure 5.1 Ordonnées à l'origine divisées par les moyennes des évaluations

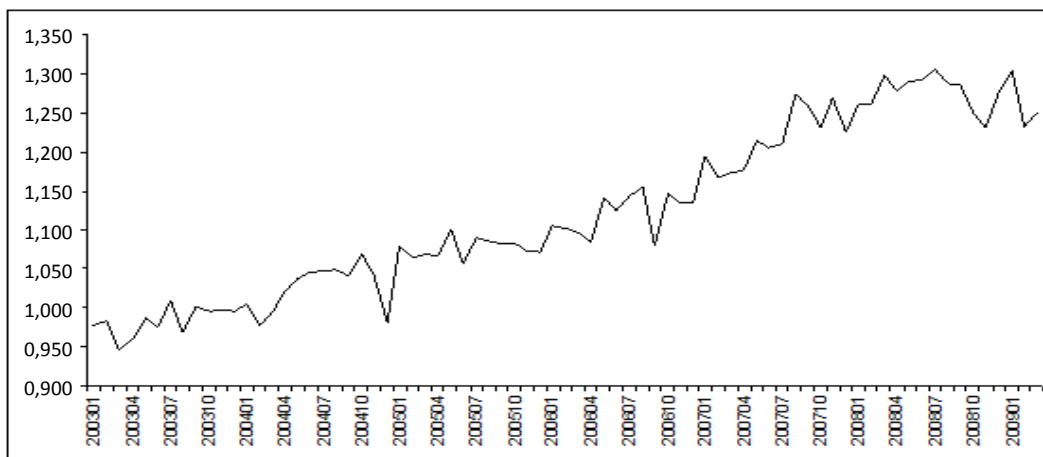


Figure 5.2 Coefficients de pente

## 5.2 Volatilité du coefficient de pente

Plusieurs facteurs peuvent avoir contribué à la volatilité des coefficients de pente  $\hat{\beta}'$  dans nos régressions des prix de vente sur les évaluations foncières, et donc sur les indices de GREG et SPAR. Nous allons discuter brièvement de trois de ces facteurs, à savoir le changement de composition de l'échantillon, l'hétéroscédasticité et les valeurs aberrantes.

Un échantillon de logements peut être considéré comme un échantillon de localisations, ou adresses, puisque les logements sont attachés au terrain sur lequel ils sont construits. Un changement de composition de l'échantillon n'est rien d'autre qu'un changement dans les localisations au niveau le plus bas. Un *changement de composition des localisations* influe sur la composition de l'échantillon en ce qui concerne les caractéristiques de qualité moyennes des biens, telles que le nombre de pièces, la superficie, etc. Dans notre cadre simple, où nous observons une seule caractéristique (non physique), à savoir la valeur d'évaluation, un changement de composition des localisations se résume à un changement de la distribution d'échantillon des évaluations. Cela, conjugué à toute variation des changements de prix selon

le créneau du marché, induit un changement dans la distribution d'échantillons des ratios  $p_n^t/a_n^0$ , qui à son tour entraîne un changement de  $\hat{\beta}^t$  dans le modèle de régression à deux variables (3.5).

Hormis la stratification, nous ne pouvons pas faire grand-chose quant à l'effet des changements de composition des localisations dans l'échantillon (mais la stratification par province et par type de logement n'a pas été très utile), de sorte qu'il est difficile de réduire la volatilité de  $\hat{\beta}^t$  et, par conséquent, des indices GREG et SPAR. Il est également impossible d'introduire une variable de contrôle pour la localisation au niveau de l'adresse dans les méthodes d'imputation hédoniques. Dans ces dernières, l'effet du changement de composition (des localisations) est atténué par l'ajout de variables de contrôle pour la région ainsi qu'une gamme de caractéristiques physiques. Cependant, cela ne signifie pas nécessairement que l'imputation hédonique produira une série d'indices plus stable que les méthodes GREG ou SPAR. La plupart des modèles hédoniques classiques sont moins bien ajustés aux données transversales que notre modèle, et les coefficients des caractéristiques présentent habituellement une forte variabilité au cours du temps. Donc, il n'est peut-être pas étonnant que Bourassa, Hoesli et Sun (2006) constatent que [traduction] « l'indice SPAR [...] suit fiablement les variations de prix des logements mais est moins volatil que les indices produits par des méthodes qui requièrent plus d'estimations de paramètres. »

Nous pouvons aussi examiner la variabilité du coefficient de pente d'un point de vue purement statistique. Il est bien connu que, dans un modèle à deux variables, l'estimateur par les MCO  $\hat{\beta}^t$  peut s'écrire sous la forme

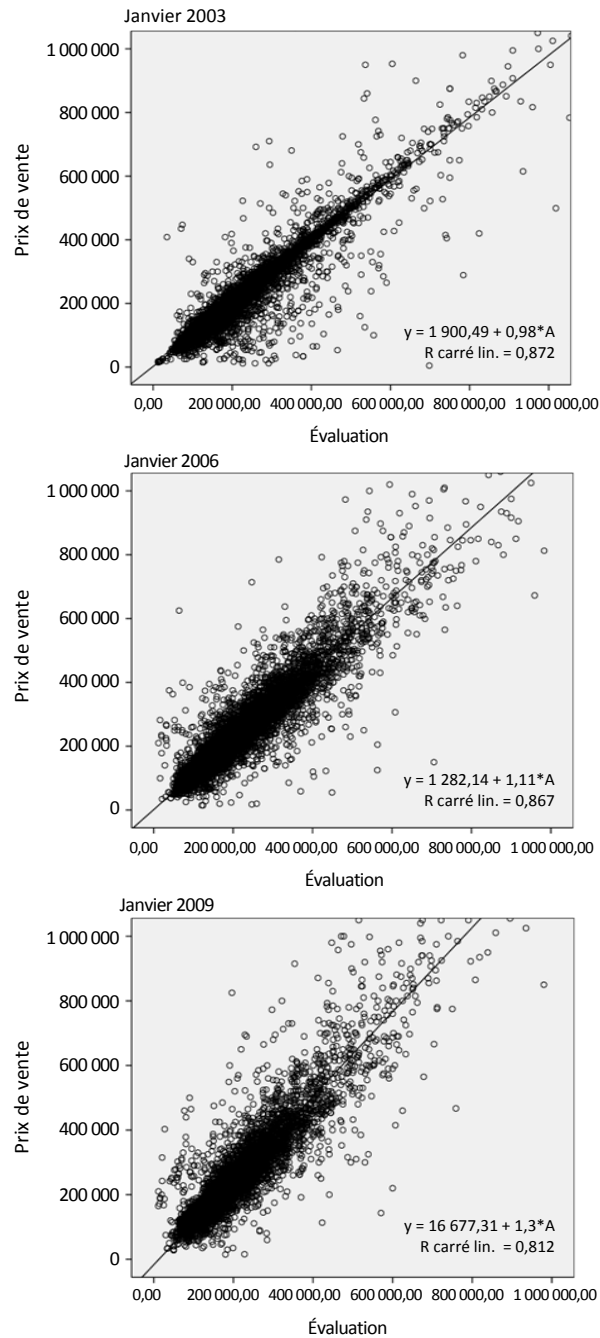
$$\hat{\beta}^t = r(p^t, a^0) \frac{s(p^t)}{s(a^0)}, \quad (5.2)$$

où  $r(p^t, a^0)$  désigne le coefficient de corrélation dans l'échantillon à la période  $t$  entre les prix de vente et les évaluations foncières, qui est égal à la racine carrée de  $R^2$ ;  $s(p^t)$  et  $s(a^0)$  sont les écarts-types d'échantillon correspondants. Une comparaison des figures 4.1 et 5.2 laisse entendre que des variations subites de  $R^2$  sont en grande partie responsables de la volatilité de  $\hat{\beta}^t$ . Ainsi, en décembre 2004, une diminution importante de  $R^2$  coïncide avec une diminution importante de  $\hat{\beta}^t$  (et avec une diminution des indices GREG et SPAR, comme le montre la figure 4.4).

La régression par les moindres carrés peut être pondérée ou non pondérée. En l'absence d'hétéroscédasticité, c'est-à-dire quand la variance des erreurs est constante, il faut utiliser les MCO. En présence d'hétéroscédasticité, la préférence va aux moindres carrés pondérés (MCP); si l'on utilise les poids appropriés, les MCP donnent des coefficients plus stables que les MCO. Dans ce cas, la somme des résidus dans l'échantillon pondéré diffère de zéro, de sorte que l'estimateur (3.9), doit être utilisé. Pour faciliter l'interprétation de l'indice GREG et la comparaison avec l'indice SPAR, à la section 3, nous avons supposé qu'il n'y avait pas de problème d'hétéroscédasticité et nous nous sommes limités aux MCO. Donc, l'estimateur GREG (MCO) donné par (3.10) demeure asymptotiquement sans biais sous le plan en présence d'hétéroscédasticité.

La forme la plus intéressante d'hétéroscédasticité (classique) – et, étant donné notre jeu de données, la seule forme que nous serions capables de réduire – se présenterait si la variance des erreurs de notre modèle de régression (3.5) dépendait de la valeur d'évaluation, celle-ci étant la seule variable explicative. Cependant, les résidus de nos régressions par les MCO n'indiquent par la présence d'une

hétéroscédasticité de ce type importante. Cela est illustré à la figure 5.3, qui représente les prix de vente en fonction des évaluations, pour trois mois y compris la période de référence (janvier 2003); les droites de régression sont également données. En guise de confirmation, nous avons également effectué le test de White (1980), qui n'a pas indiqué cette forme d'hétéroscédasticité.



**Figure 5.3 Nuages de points et droites de régression**

Notre jeu de données initiales de prix de vente et d'évaluations foncières comprenait certaines *valeurs aberrantes* évidentes. Pour estimer l'indice GREG, nous avons par conséquent utilisé un jeu de données nettoyées qui a été préparé pour calculer l'indice officiel des prix des logements aux Pays-Bas. Statistics Netherlands applique plusieurs procédures de nettoyage des données. Les logements qui ont été vendus plus d'une fois durant un mois donné sont exclus du jeu de données. Pour éliminer les erreurs de saisies et les valeurs aberrantes qui pourraient influencer excessivement les résultats, les biens dont le prix de vente ou l'évaluation foncière est inférieur à 10 000 € ou supérieur à 5 000 000 € et ceux dont le ratio prix de vente-évaluation est « irréaliste » sont également supprimés. La suppression des observations « irréalistes » est faite en examinant la distribution du logarithme des ratios prix de vente-évaluation; sont supprimées toutes les observations pour lesquelles l'écart du logarithme du ratio par rapport à la moyenne est de plus de 5 écarts-types. Pour plus de renseignements, voir Statistics Netherlands (2008).

Ces procédures sont assez arbitraires. Pour les estimateurs par la régression, tels que l'estimateur GREG, il est plus approprié de supprimer les observations dont l'effet de levier est important, c'est-à-dire d'éliminer de l'échantillon les unités dont l'exclusion a un effet important sur les coefficients de régression. Une mesure bien connue dans ce contexte est le DFBETA d'une unité de l'échantillon (Cook et Weisberg 1982). Puisque l'indice SPAR peut s'écrire sous forme d'un indice fondé sur la régression, cette mesure pourrait également être utilisée pour déceler et supprimer les valeurs aberrantes. Les nuages de points de la figure 5.3 montrent que le jeu de données nettoyé contient encore certaines valeurs aberrantes importantes. Il reste à déterminer si ces valeurs ont un effet de levier important et si leur élimination réduira la volatilité des  $\hat{\beta}^t$  dans les indices GREG et SPAR.

### 5.3 Certaines autres remarques

La méthode GREG part du principe que le parc de logements est fixe. Autrement dit, nous avons supposé qu'il ne se produit pas d'entrées (par exemple logements nouvellement construits) ni de sorties (logements mis aux rebuts) et que la qualité des logements demeure constante au cours du temps. Notre approche n'est pas symétrique en ce sens que nous nous conditionnons sur le parc de logements à la *période de référence*. Dans la perspective d'un indice, nous estimons un indice des prix de Laspeyres pour le parc de logements où les quantités sont toutes égales à 1 parce que chaque logement est traité comme un bien unique. Une approche tout aussi justifiable consisterait à mesurer la variation du parc de logements à la période courante, qui comprend les ajouts au parc durant chaque période, en utilisant un indice de Paasche. En calculant la moyenne géométrique des deux indices, on obtiendrait l'indice de Fisher. Ce dernier est une mesure privilégiée de la variation des prix en raison de sa forme symétrique. La construction d'un indice GREG de type Fisher est toutefois impossible, puisque la composante de Paasche requiert des valeurs d'évaluation en temps réel pour les logements neufs dans le parc, alors qu'elles ne sont manifestement pas disponibles.

L'hypothèse d'un parc de logements fixe (à la période de référence) peut être relâchée par enchaînement annuel, à condition que le parc de logements soit réévalué annuellement. Il s'agit de la situation actuelle aux Pays-Bas; dans le passé, les évaluations foncières étaient effectuées tous les trois ou quatre ans. Une mise à jour annuelle des évaluations pourrait également comprendre une correction pour les changements de qualité des biens, du moins dans une certaine mesure, parce que les évaluations mises

à jour tiennent vraisemblablement compte des réparations importantes, des rénovations et de la dépréciation.

Une remarque finale s'impose. À certaines fins, il est souhaitable de décomposer l'indice des prix des logements global en deux composantes : l'une qui mesure la variation de prix du bâtiment, et l'autre, la variation de prix du terrain. Ni notre méthode GREG ni les méthodes SPAR et des ventes répétées ne conviennent pour cela. Les méthodes d'imputation hédoniques pourraient convenir, malgré des problèmes pratiques tels que la multicolinéarité; voir Diewert, de Haan et Hendriks (2012) pour une première tentative. Si les données sur la taille du bâtiment, la taille du terrain et d'autres attributs déterminant le prix devenaient disponibles pour tous les biens inclus dans le parc de logements, nous serions capables d'estimer un « indice GREG avec imputation hédonique », comprenant la décomposition terrain-bâtiment. Les chances d'obtenir ce genre de données aux Pays-Bas sont malheureusement minces.

## 6 Conclusion

La simple méthode GREG décrite dans le présent article, qui est fondée sur la régression des prix de vente sur les évaluations foncières par les MCO, réduit considérablement la volatilité d'un indice des prix des logements comparativement aux ratios des moyennes d'échantillon. L'indice SPAR peut être considéré comme un estimateur de l'indice GREG (MCO) (lui-même un estimateur, évidemment) dans lequel la moyenne de population des évaluations à la période de référence est remplacée par les moyennes d'échantillon à la période de référence et à la période comparée. Nos résultats empiriques pour les Pays-Bas indiquent que l'indice SPAR est presque aussi efficace que l'indice GREG, même pour de petites sous-populations. Nous avons vérifié cela en tirant un échantillon aléatoire de 50 observations chaque mois du nombre total de ventes mensuelles (15 000 en moyenne). Les variations d'un mois à l'autre de l'indice SPAR sont à peine plus importantes que celles de l'indice GREG.

En raison du changement de composition de l'ensemble de logements vendus, la série chronologique GREG (et SPAR) présente une forte volatilité à court terme. Une augmentation durant un mois particulier est habituellement suivie d'une diminution le mois suivant. Autrement dit, les variations d'un mois à l'autre ne disent pas grand-chose au sujet de la variation réelle du prix du parc de logements qui, sauf dans des circonstances inhabituelles, devrait être harmonieuse. Une méthode améliorée de détection des valeurs aberrantes aiderait peut-être à réduire la volatilité de l'indice, mais l'effet serait vraisemblablement limité. L'application d'une procédure de lissage semble être une option. Cependant, ce lissage entraîne habituellement des révisions des indices de prix publiés antérieurement et l'absence de révision est l'un des points forts des approches GREG et SPAR. Une autre option consisterait à réduire la fréquence d'observation, en la choisissant par exemple trimestrielle, mais cela pourrait être indésirable également.

D'un point de vue purement statistique, dans notre modèles à deux variables, la variabilité de  $R^2$  semble être en grande partie à l'origine de la volatilité du coefficient de pente et, par conséquent, de celle de la série d'indices de prix. De futures études pourraient porter sur la relation entre les changements de composition ayant trait aux caractéristiques des biens immobiliers et les variations de  $R^2$ . Comme les données sur de nombreuses caractéristiques des logements ne sont pas disponibles, nous ne pouvons pas étudier cette question au moyen de nos données. Heureusement, Statistics Netherlands a accès à un jeu de données produit par la plus grande association d'agents immobiliers aux Pays-Bas qui pourrait être utile. Ce jeu de données couvre environ 70 % des ventes de logements qui ont eu lieu aux Pays-Bas de 1999 à

2008, comprend des données sur de nombreuses caractéristiques des biens et a été enrichi par les données d'évaluation foncière. Dans le passé, nous avons utilisé ce jeu de données pour comparer l'indice SPAR à divers types d'indices hédoniques.

## Remerciements

Les auteurs remercient les participants à l'Economic Measurement Group Workshop, qui s'est déroulé du 1<sup>er</sup> au 3 décembre 2010 à l'Université de New South Wales, à Sydney, en Australie, et les participants à un séminaire d'économie appliquée, tenu le 22 novembre 2011 à l'Université du Queensland, à Brisbane, en Australie, de leurs commentaires constructifs concernant des versions préliminaires de l'article. Les commentaires et suggestions faits par le rédacteur et deux examinateurs anonymes ont également contribué à améliorer l'article. Les auteurs remercient de son aide Erna van der Wal, qui leur a fourni les données. Les opinions exprimées dans l'article sont celles des auteurs et ne représentent pas forcément celles de Statistics Netherlands.

## Annexe

### Erreurs-types approximatives de l'indice GREG

L'indice GREG défini par l'équation (3.10) dans le corps du texte est un ratio de deux estimateurs,  $\hat{p}_{\text{GREG}}^t$  et  $\hat{p}_{\text{GREG}}^0$ ; pour simplifier, nous supprimons la notation « MCO ». En utilisant un développement du premier degré en séries de Taylor, la variance de l'indice peut être approximée par (voir, par exemple, Kendall et Stuart 1976)

$$\text{var}(\hat{P}_{\text{GREG}}^{0t}) \cong \left[ \frac{E(\hat{p}_{\text{GREG}}^t)}{E(\hat{p}_{\text{GREG}}^0)} \right]^2 \left[ \frac{\text{var}(\hat{p}_{\text{GREG}}^t)}{\{E(\hat{p}_{\text{GREG}}^t)\}^2} + \frac{\text{var}(\hat{p}_{\text{GREG}}^0)}{\{E(\hat{p}_{\text{GREG}}^0)\}^2} + \frac{\text{cov}(\hat{p}_{\text{GREG}}^t, \hat{p}_{\text{GREG}}^0)}{E(\hat{p}_{\text{GREG}}^t)E(\hat{p}_{\text{GREG}}^0)} \right], \quad (\text{A.1})$$

où  $E(\hat{p}_{\text{GREG}}^t)$  et  $E(\hat{p}_{\text{GREG}}^0)$  désignent les valeurs espérées.

Le terme de covariance dans (A.1) est égal à 0 puisque, par hypothèse, les échantillons aux périodes 0 et  $t$  sont tirés indépendamment. Le remplacement des valeurs espérées dans (A.1) par les estimateurs et le calcul subséquent de la racine carré donne l'expression qui suit pour l'erreur-type de  $\hat{P}_{\text{GREG}}^{0t}$  :

$$se(\hat{P}_{\text{GREG}}^{0t}) \cong \hat{P}_{\text{GREG}}^{0t} \left[ \frac{\text{var}(\hat{p}_{\text{GREG}}^t)}{(\hat{p}_{\text{GREG}}^t)^2} + \frac{\text{var}(\hat{p}_{\text{GREG}}^0)}{(\hat{p}_{\text{GREG}}^0)^2} \right]^{1/2}. \quad (\text{A.2})$$

L'équation (A.2) peut être estimée en pratique en utilisant  $\hat{p}_{\text{GREG}}^s = \hat{\alpha}^s + \hat{\beta}^s \bar{a}^0$  ( $s = 0, t$ ), d'où  $\text{var}(\hat{p}_{\text{GREG}}^s) = \text{var}(\hat{\alpha}^s) + (\bar{a}^0)^2 \text{var}(\hat{\beta}^s) + 2\bar{a}^0 \text{cov}(\hat{\alpha}^s, \hat{\beta}^s)$ . Les estimations des (co)variances sont obtenues facilement dans la plupart des progiciels statistiques à partir de la matrice de variance-covariance.



La division de (A.2) par  $\hat{P}_{\text{GREG}}^{0t}$  donne une expression pour l'erreur-type relative ou coefficient de variation,  $CV(\hat{P}_{\text{GREG}}^{0t}) = se(\hat{P}_{\text{GREG}}^{0t}) / \hat{P}_{\text{GREG}}^{0t}$ , de l'indice GREG :

$$CV(\hat{P}_{\text{GREG}}^{0t}) \cong \left[ \frac{\text{var}(\hat{p}_{\text{GREG}}^t)}{(\hat{p}_{\text{GREG}}^t)^2} + \frac{\text{var}(\hat{p}_{\text{GREG}}^0)}{(\hat{p}_{\text{GREG}}^0)^2} \right]^{1/2} = \left[ \{CV(\hat{p}_{\text{GREG}}^t)\}^2 + \{CV(\hat{p}_{\text{GREG}}^0)\}^2 \right]^{1/2}. \quad (\text{A.3})$$

Un élément plus important est l'erreur-type relative de la *variation en pourcentage* de l'indice, c'est-à-dire  $CV(\hat{P}_{\text{GREG}}^{0t} - 1) = se(\hat{P}_{\text{GREG}}^{0t} - 1) / (\hat{P}_{\text{GREG}}^{0t} - 1)$ . Celle-ci est généralement plus grande que  $CV(\hat{P}_{\text{GREG}}^{0t})$ , étant donné que  $se(\hat{P}_{\text{GREG}}^{0t} - 1) = se(\hat{P}_{\text{GREG}}^{0t})$  et  $\hat{P}_{\text{GREG}}^{0t} - 1 < \hat{P}_{\text{GREG}}^{0t}$ .

Si les deux droites de régression passent presque par l'origine, donc que  $\hat{\alpha}^s \cong 0 (s = 0, t)$ , nous avons  $\hat{P}_{\text{GREG}}^{0t} \cong \hat{\beta}^t / \hat{\beta}^0$  et (A.2) se simplifie pour donner

$$se(\hat{P}_{\text{GREG}}^{0t}) = se(\hat{P}_{\text{GREG}}^{0t} - 1) \cong \hat{P}_{\text{GREG}}^{0t} \left[ \frac{\text{var}(\hat{\beta}^t)}{(\hat{\beta}^t)^2} + \frac{\text{var}(\hat{\beta}^0)}{(\hat{\beta}^0)^2} \right]^{1/2}. \quad (\text{A.4})$$

Dans ce cas particulier, les indices GREG et SPAR coïncident presque, de sorte que l'expression (A.4) est également vérifiée pour l'indice SPAR (en utilisant  $\hat{P}_{\text{SPAR}}^{0t}$  au lieu de  $\hat{P}_{\text{GREG}}^{0t}$ ).

## Bibliographie

- Bailey, M.J., Muth, R.F. et Nourse, H.O. (1963). A regression method for real estate price construction. *Journal of the American Statistical Association*, 58, 933-942.
- Beaumont, J.-F., et Alavi, A. (2004). Estimation robuste par la régression généralisée. *Techniques d'enquête*, 30, 2, 217-231.
- Bourassa, S.C., Hoesli, M. et Sun, J. (2006). A simple alternative house price index method. *Journal of Housing Economics*, 15, 80-97.
- Calhoun, C.A. (1996). OFHEO House Price Indexes: HPI Technical Description. Office of Federal Housing Enterprise Oversight, Washington, DC.
- Case, K.E., et Shiller, R.J. (1987). Prices of single-family homes since 1970: New indexes for four cities. *New England Economic Review*, Septembre-Octobre, 45-56.
- Case, K.E., et Shiller, R.J. (1989). The efficiency of the market for single family homes. *The American Economic Review*, 79, 125-137.
- Cochran, W.G. (1977). *Sampling Techniques*, 3<sup>e</sup> Édition, New York : John Wiley & Sons, Inc.
- Cook, R.D., et Weisberg, S. (1982). *Residuals and Influence in Regression*, New York : Chapman and Hall.

- de Haan, J. (2007). *Formulae for the Variance of (Changes in) the SPAR Index*. Manuscrit non publié, Statistics Netherlands, Voorburg (Pays-Bas seulement, disponible auprès de l'auteur sur demande).
- de Haan, J. (2010). Hedonic price indexes: A comparison of imputation, time dummy and 'Re-Pricing' methods, *Journal of Economics and Statistics* (Jahrbucher fur Nationalokonomie und Statistik), 230, 772-791.
- de Haan, J., van der Wal, E. et de Vries, P. (2009). The measurement of house prices: A review of the sale price appraisal method. *Journal of Economic and Social Measurement*, 34, 51-86.
- de Vries, P., de Haan, J., van der Wal, E. et Mariën, G. (2009). A house price index based on the SPAR method. *Journal of Housing Economics*, 18, 214-223.
- Diewert, W.E., de Haan, J. et Hendriks, R. (2012). The decomposition of a house price index into land and structures components: A hedonic regression approach. *Econometric Reviews* (à venir).
- Diewert, W.E., Heravi, S. et Silver, M. (2009). Hedonic imputation versus time dummy hedonic indexes. Dans *Price Index Concepts and Measurement*, (Éds., W.E. Diewert, J. Greenlees et C. Hulten), NBER Studies in Income and Wealth, Chicago: Chicago University Press, 70, 161-196.
- Edelstein, R.H., et Quan, D.C. (2006). How does appraisal smoothing bias real estate returns measurement? *Journal of Real Estate Finance and Economics*, 32, 41-60.
- Eurostat (2010). *Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices*, Version 1.9. Disponible au [www.epp.eurostat.ec.europa.eu/portal/page/portal/hicp/documents/Tab/Tab/03\\_METH-OOH-TECHMANUAL\\_V1-9.pdf](http://www.epp.eurostat.ec.europa.eu/portal/page/portal/hicp/documents/Tab/Tab/03_METH-OOH-TECHMANUAL_V1-9.pdf).
- Francke, M.K. (2010). Repeat sales index for thin markets: A structural time series approach. *Journal of Real Estate Finance and Economics*, 41, 24-52.
- Geltner, D. (1996). The repeated-measures regression-based index: A better way to construct appraisal-based indexes of commercial property value. *Real Estate Finance*, 12, 29-35.
- Gouriéroux, C., et Laferrère, A. (2009). Managing hedonic house price indexes: The french experience. *Journal of Housing Economics*, 18, 206-213.
- Grimes, A., et Young, C. (2010). A Simple Repeat Sales House Price Index: Comparative Properties Under Alternative Data Generation Processes. Motu Working Paper 10-10, Motu Economic and Public Policy Research, New Zealand.
- Hardman, M. (2011). Calculating High Frequency Australian Residential Property Price Indices. Rismark Document technique, disponible au [www.rpdata.com/images/stories/content/PDFs/technical\\_method\\_paper.pdf](http://www.rpdata.com/images/stories/content/PDFs/technical_method_paper.pdf).
- Hedlin, D., Falvey, H., Chambers, R. et Kokic, P. (2001). Does the model matter for GREG estimation? A business survey example. *Journal of Official Statistics*, 17, 527-544.
- Hill, R.J., et Melser, D. (2008). Hedonic imputation and the price index problem: An application to housing. *Economic Inquiry*, 46, 593-609.

- Jansen, S.J.T., de Vries, P., Coolen, H.C.C.H., Lamain, C.J.M. et Boelhouwer, P. (2008). Developing a house price index for the Netherlands: A practical application of weighted repeat sales. *Journal of Real Estate Finance and Economics*, 37, 163-186.
- Kendall, M., et Stuart, A. (1976). *The Advanced Theory of Statistics – Volume 1: Distribution Theory*, 4<sup>e</sup> Édition, Londres : Charles Griffin & Company.
- Leventis, A. (2006). Removing Appraisal Bias from a Repeat Transactions House Price Index: A Basic Approach. Document présenté à l'atelier OECD-IMF on Real Estate Price Indexes, Paris, 6 au 7 novembre 2006.
- Makaronidis, A., et Hayes, K. (2006). Owner Occupied Housing for the HICP. Document présenté à l'atelier OECD-IMF on Real Estate Price Indexes, Paris, 6 au 7 novembre 2006.
- Rossini, P., et Kershaw, P. (2006). Developing a Weekly Residential Price Index Using the Sales Price Appraisal Ratio. Document présenté à la twelfth Annual Pacific Rim Real Estate Society Conference, Auckland, 22 au 25 janvier 2006.
- Saarnio, M. (2006). Housing Price Statistics at Statistics Finland. Document présenté à l'atelier OECD-IMF on Real Estate Price Indexes, Paris, 6 au 7 novembre 2006.
- Särndal, C.-E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*, New York : Springer-Verlag.
- Shi, S., Young, M. et Hargreaves, B. (2009). Issues in measuring a monthly house price index in New Zealand. *Journal of Housing Economics*, 18, 336-350.
- Statistics Netherlands (2008). Price Index Owner-occupied Existing Dwellings; Method Description. Statistics Netherlands, La Haye, disponible au [www.cbs.nl/NR/rdonlyres/A49D8542-26EC-40FD-9093-82A519247F4B/0/MethodebeschrijvingPrijsindexBestaandeKoopwoningene.pdf](http://www.cbs.nl/NR/rdonlyres/A49D8542-26EC-40FD-9093-82A519247F4B/0/MethodebeschrijvingPrijsindexBestaandeKoopwoningene.pdf).
- van der Wal, E., ter Steege, D. et Kroese B. (2006). Two Ways to Construct a House Price Index for the Netherlands: The Repeat Sale and Sale Price Appraisal Ratio. Document présenté à l'atelier OECD-IMF on Real Estate Price Indexes, Paris, 6 au 7 novembre 2006.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.