COMMISSARIAT
GÉNÉRAL AU
DÉVELOPPEMENT
DURABLE

SERVICE
DE L'OBSERVATION
ET DES STATISTIQUES

Mars
2016

Document de travail n° 21

Logement-Construction

Méthodologie des estimations en date réelle des logements autorisés et commencés

Service de l'observation et des statistiques

www.statistiques.developpement-durable.gouv.fr

Résumé

Le suivi conjoncturel de la construction de logement a été réalisé pendant des années à partir des informations prises en compte dans la base Sit@del2, intégrant le délai de collecte des informations. Les séries « en date de prise en compte », qui recensent l'ensemble des informations (autorisations, mises en chantier) remontées chaque mois, présentent l'avantage d'être disponibles très rapidement tandis qu'il faut plusieurs mois pour collecter l'ensemble des permis autorisés ou mis en chantier un mois donné. Lorsque la collecte est régulière, ces séries permettent de suivre les évolutions de la construction de logement, les retournements conjoncturels s'observant toutefois avec un léger retard lié à la vitesse moyenne de remontée de l'information.

La production d'estimations en date réelle s'est avérée nécessaire suite à la dégradation progressive de la collecte des déclarations d'ouverture de chantier. Ainsi, début 2015, près de 80 000 logements autorisés au cours de l'année 2010 ne sont ni annulés ni commencés selon les informations contenues dans la base Sit@del2, soit 17 % des autorisations *a priori* valables deux ans. De plus, plusieurs chocs dans la collecte des autorisations de construire se sont produits ces dernières années, affectant le suivi conjoncturel dans certaines régions. Les nouveaux indicateurs visent à décrire l'activité réelle en matière d'autorisations de construire et de mises en chantier de logements sur le territoire, et non plus le recensement des formulaires collectés. L'objectif de ce document est de décrire les méthodes utilisées pour estimer en date réelle les nombres de logements autorisés et commencés.

Auteurs : François Limousin, Frédéric Minodier, Benoît Pentinat (*)

(*) Les auteurs tiennent à remercier Guillaume Houriez pour son implication dans le projet, ainsi que Karine Boutier, Emmanel Caidedo et Pauline Delance pour la qualité des échanges sur le sujet. Ils remercient également Sylvain Moreau et Guillaume Mordant (SOeS) et Olivier Sautory et Dominique Ladiray (Insee) pour leurs conseils sur ce projet.

Table des matières

Méthodologie des estimations de logements autorisés et commencés en date réelle

I) La source sitadel2

Sitadel2 (Système d'information et de traitement automatisé des données élémentaires sur les logements et les locaux) est à la fois un outil de collecte et une base de diffusion permettant au service de l'observation et des statistiques[1] (SOeS) de suivre la construction de logements et de locaux à partir des autorisations d'urbanisme.

Le suivi de la construction neuve à partir des permis de construire est ancien puisque le premier dispositif du système statistique public (Sirocco) remonte à 1972. Les dispositifs de collecte ont ensuite évolué en fonction de la déconcentration et de la décentralisation du service public d'une part, et des progrès informatiques d'autre part. Siclone (1986), Sitadel (1998) et Sitadel2 (2009) ont ainsi succédé au premier dispositif Sirocco.

Collecte des informations

Le dispositif de collecte des permis de construire s'appuie sur différents acteurs, depuis le pétitionnaire qui dépose son formulaire de permis de construire à la mairie du lieu des travaux jusqu'à la base nationale Sitadel2, gérée par l'administration centrale du ministère de l'Environnement, de l'Énergie et de la Mer (Meem).

La collecte pour le suivi statistique est directement réalisée auprès des centres instructeurs. L'article R1614-20 du code général des collectivités territoriales constitue le cadre réglementaire dans lequel est effectuée cette remontée d'information : « *les communes et établissements publics de coopération intercommunale qui instruisent eux-mêmes les actes d'urbanisme transmettent chaque mois aux services statistiques du ministère de l'équipement, pour l'établissement de statistiques, les informations statistiques prévues par les arrêtés pris pour l'application de l'article R 434-2 du code de l'urbanisme* ».

Jusqu'en juillet 2015, les communes compétentes en matière d'urbanisme de moins de 10 000 habitants et qui n'appartenaient pas à un EPCI de plus de 20 000 habitants pouvaient confier aux services de l'État (direction départementale des territoires) l'instruction des permis. Les autres communes compétentes devaient prendre en charge cette activité ou confier l'instruction à une autre collectivité (EPCI par exemple). Depuis le premier juillet 2015, le seuil d'autonomie concernant les EPCI a été abaissé à 10 000 habitants[2], générant une forte augmentation du nombre de centres instructeurs (de 1600 à 2 800 environ).

Le projet de construction est suivi tout au long de son cycle de vie. La première étape est l'autorisation délivrée par l'autorité compétente à l'issue du travail effectué par le centre instructeur. La remontée de cette information est relativement rapide puisque qu'environ 65 % de l'information est remontée au cours du premier mois et 95 % au bout de 6 mois.

[1] Service statistique ministériel du Meem et du ministère du Logement et de Habitat Durable (MLHD).
[2] L'article 134 de la loi n°2014-366 du 24 mars 2014 pour l'accès au logement et un urbanisme rénové « Alur » réserve la mise à disposition des moyens de l'État pour l'application du droit des sols aux seules communes compétentes appartenant à des EPCI qui comptent moins de 10 000 habitants.

La deuxième étape du projet de construction est la mise en chantier : le pétitionnaire doit remplir une déclaration d'ouverture de chantier, laquelle suivra ensuite le même processus en termes de suivi statistique. Le délai de remontée de l'information est plus long que celui des autorisations : il faut environ un trimestre pour obtenir la moitié des ouvertures de chantier.

En début de mois, les centres instructeurs transmettent dans l'outil de collecte Sitadel2 l'ensemble des évènements (dépôts, décision, ouverture de chantier, achèvement des travaux) reçus et traités le mois précédent. Vers le 20 de chaque mois, les permis collectés sont déversés dans la base de diffusion.

Deux dates sont affectées à chaque évènement déversé dans l'infocentre : sa date réelle d'événement (DR) et sa date de prise en compte (DPC) dans la base de diffusion.

Diffusion

Le SOeS publie chaque mois le nombre de logements autorisés et commencés à partir des données Sitadel2. Les chiffres sont ventilés par type de bâtiment (individuel, collectif...) et par zone géographique.

II) Les limites de l'ancienne méthode et les objectifs du projet DR+

Données précédemment publiées

Deux types de séries mensuelles d'autorisation de construire et de mises en chantier (logements et locaux non résidentiels) étaient publiés jusqu'en janvier 2015 à partir de la base Sitadel2 :
- des séries en dates réelles (DR) ;
- des séries en date de prise en compte (DPC).

Les autorisations

Les séries en date réelle reflètent in fine la réalité de la construction dans le temps et doivent être privilégiées aux séries en date de prise en compte pour la réalisation d'études structurelles. Le délai de mise à disposition de ces séries est relativement long car il dépend du délai de remontée des autorisations ou mises en chantier, des permis modificatifs et des annulations. Il faut environ 6 mois pour que le nombre d'autorisations d'un mois donné commence à se stabiliser et 18 mois pour les mises en chantier.

Les séries en date de prise en compte comptabilisent les flux réceptionnés chaque mois. Elles sont disponibles rapidement mais présentent l'inconvénient d'être sensibles aux aléas de collecte (réception et contenu des fichiers des centres instructeurs, relances mensuelles effectuées auprès des pétitionnaires pour obtenir des informations sur les mises en chantier).

Ces séries étaient les seules exploitables à des fins conjoncturelles. Les évolutions observées à partir des séries en date de prise en compte permettent en effet d'estimer celles des séries en date réelle sous réserve que la collecte soit régulière. Toutefois, les retournements conjoncturels s'observent avec un retard qui dépend de la vitesse moyenne de remontée de l'information. En outre, plusieurs chocs de collecte se sont produits ces dernières années, affectant fortement le suivi conjoncturel dans certaines régions et au niveau national.

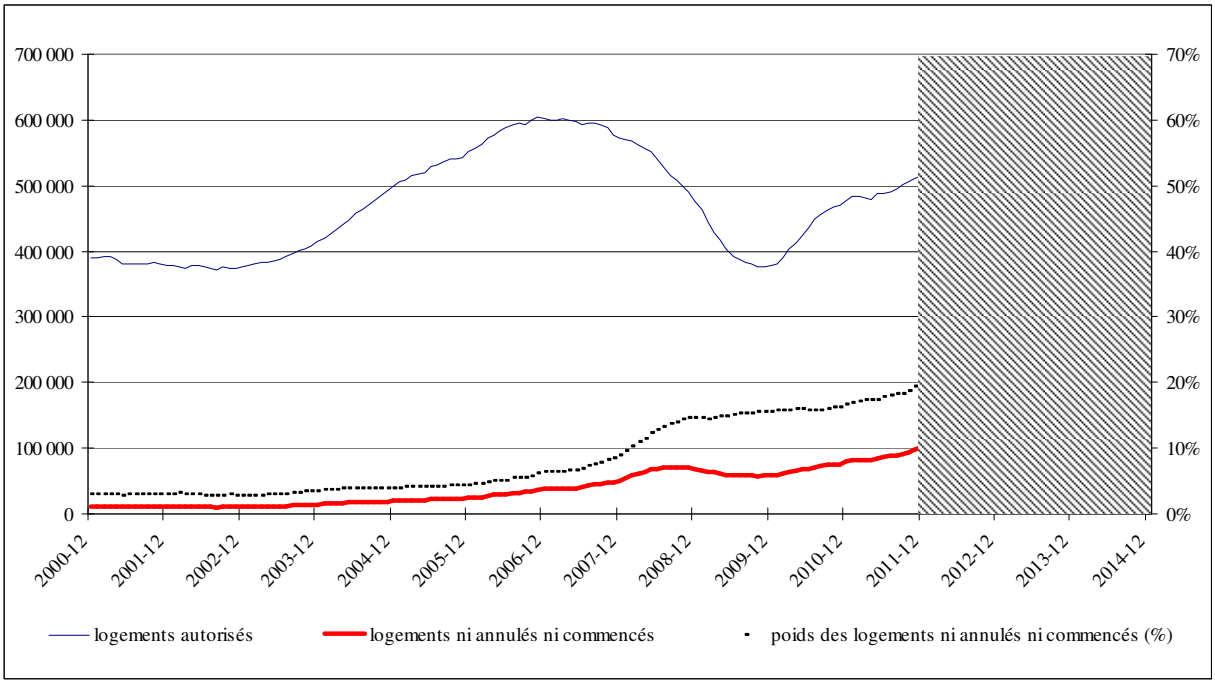
Les mises en chantier

Les séries des mises en chantier publiées reflètent les ouvertures de chantier réceptionnées dans le système de collecte. Mais il s'avère que toute l'information ne parvient plus jusqu'à la base Sitadel2. Certains projets autorisés restent dans un état hybride : on ne sait pas si la construction a effectivement commencé ou si le pétitionnaire n'a pas pu mettre en œuvre son projet de construction[3].

En effet la collecte des mises en chantier s'est dégradée au fil du temps depuis la réforme du droit des sols d'octobre 2007. La décentralisation de l'instruction a eu pour effet de multiplier les points de collecte de l'information et de fragiliser in fine le dispositif de collecte de l'information par les centres instructeurs. Le SOeS mène pourtant chaque mois une opération de relance par voie postale auprès de pétitionnaires autorisés à construire des logements afin de suivre leur projet de construction. Cette collecte complémentaire permet de récupérer une partie de l'information manquante mais pas la totalité. Ainsi, près de 80 000 logements autorisés au cours de l'année 2010 ne sont ni annulés ni commencés selon les informations contenues dans la base Sitadel à fin mars 2015 (*cf. figure 1*).

Figure 1 : poids des logements autorisés ni commencés, ni annulés

Par date d'autorisation



Source : SOeS, Sitadel2, données à fin mars 2015

[3] Les travaux doivent impérativement être commencés dans un délai de deux ans suivant l'obtention de l'autorisation d'urbanisme. Le titulaire peut demander le prolongement de son autorisation pour une durée d'un an si les travaux ne peuvent pas être commencés dans le délai de deux ans ou s'il prévoit d'interrompre le chantier pendant plus d'une année.

Objectifs du projet DR+

Un projet d'estimation des autorisations et des mises en chantier « en date réelle » a été engagé fin 2012 au sein du SOeS. Il s'agit de produire des estimations mensuelles robustes aux aléas de collecte, selon le même calendrier que les séries en date de prise en compte (diffusion du mois *m* en fin de mois *m+1*) et se faisant d'assurer une meilleure articulation entre les séries d'autorisations et de mises en chantier. Les estimations des mises en chantier produites doivent refléter les constructions effectivement commencées sur le territoire et non plus la simple addition des formulaires collectés. Ces estimations en date réelle ont vocation à se substituer aux séries en date de prise en compte dans l'analyse conjoncturelle.

Concernant les autorisations de construire, l'information est considérée comme exhaustive à terme. L'approche retenue consiste à traiter un problème de « vitesse de convergence ». À noter que l'estimateur en date réelle reflète le nombre de logements autorisés en date réelle y compris annulations, alors que la série en date réelle publiée auparavant est nette d'annulations. Les annulations sont toujours publiées permettant ainsi de reconstituer une série des autorisations nettes des annulations.

Concernant les mises en chantier, les estimations produites traduisent les constructions effectivement commencées sur le territoire. Le champ de la série, et donc les niveaux produits, diffèrent de ceux publiés précédemment en palliant le défaut d'exhaustivité de la collecte.

L'approche retenue s'inspire de la méthode mise en œuvre dans le cadre des comptes du logement. Le processus doit être toutefois décliné mensuellement et s'inscrire dans des délais de production très contraints. Des enquêtes sont menées sur des permis autorisés ni annulés ni commencés depuis plus de quatre ans (ils sont *a priori* caduques) afin d'estimer un volume de logements mis en chantier non recensés dans le cadre de la collecte Sitadel. Sur la période plus récente, c'est-à-dire pour les autorisations ayant moins de quatre ans d'ancienneté, les modélisations du taux d'annulation des permis autorisés et des délais de mise en chantier permettent d'estimer le nombre de mises en chantier. Les estimations produites complètent l'offre de diffusion sur la construction neuve qui était limitée jusqu'à présent au recensement des formulaires par la voie administrative.

III) Estimation du nombre de logements autorisés

L'analyse de la série en date réelle du nombre de logements autorisés montre qu'elle est globalement stabilisée à partir de 24 mois de collecte. La collecte des permis autorisés en *t* devient négligeable après *t* + 24 mois (moins de 2 % du nombre total de logements et moins de 0,1 point en évolution d'un mois sur l'autre).

Définition de l'indicateur à estimer

Soit A(*t*) le nombre de logements autorisés au mois *t* en date réelle, et A(*t, m*) le nombre de logements autorisés en *t* et collectés le mois *m*.

On peut décomposer A(*t*) de la manière suivante :

$$A(t) = A(t, t) + A(t, t+1) + \ldots + A(t, t+\infty).$$

En considérant que la série est stable au bout de 24 mois, l'approximation suivante est faite :

$$\forall t \qquad A(t) = A(t, t) + A(t, t+1) + \ldots + A(t, t+23) \qquad (1)$$

Pour faciliter les écritures, on considère dans la suite de ce document que la relation (1) est vraie.

L'estimateur s'écrit donc :

$$\forall t \qquad \hat{A}(t) = \hat{A}(t, t) + \hat{A}(t, t+1) + \ldots + \hat{A}(t, t+23)$$

Soit m le dernier mois de collecte, tous les $\hat{A}(m-i)$, pour i compris entre 0 et 23, doivent être estimés. La méthode développée va consister à estimer les évolutions estimées sur les 24 derniers mois et à les chaîner à partir du dernier point de la série stabilisée : $A(m-24)$.

On recherche donc un estimateur de la forme :

$\hat{A}^m(t)=A(m-24).\hat{E}^m(m-24,t)$ pour t compris entre $m-23$ et m, avec $\hat{E}^m(m-24,t)$ l'estimateur en m de l'évolution des autorisations entre $m-24$ et t.

Choix de l'estimateur sur les 24 derniers mois

En raison des délais de remontée de l'information, le calcul des évolutions directement observées à partir de la série en date réelle est biaisé, et ce notamment pour les mois les plus récents. En effet, plus le mois concerné est récent, plus le volume d'informations « restant à collecter » est important. Par exemple, comparer les nombres de logements autorisés en janvier 2015 et janvier 2014 en janvier 2015, revient à comparer un chiffre construit avec environ 70 % de l'information à un chiffre quasi-définitif.

Il est donc nécessaire de trouver un autre estimateur des évolutions du nombre de logements autorisés sur la fin de période.

Utilisation de la série en date de prise en compte

Une estimation « naturelle » des évolutions des autorisations en date réelle étant donnée le système de diffusion actuel, serait d'utiliser les évolutions observées à partir des données en date de prise en compte. Toutefois la série en date de prise en compte est sensible aux aléas de collecte qui peuvent impacter les évolutions.

Il est possible d'atténuer ces rattrapages de collecte en éliminant systématiquement les permis reçus avec un retard important. Ainsi, la série en date de prise en compte tronquée à d mois, écarte les permis dont le délai de réception de l'information (écart entre la date de prise en compte dans Sitadel2 et la date réelle d'autorisation) est supérieur à d mois. Cette série est notée DPC_d.

$$DPC_d(t) = A(t,t) + A(t-1,t) + \ldots\ldots + A(t-d+1,t)$$

En utilisant cette série DPC_d, l'estimation du nombre de logements autorisés s'écrit alors :

pour t tel que $m\text{-}23 \leq t \leq m$, $\quad \hat{A}^m(t) = A(m\text{-}24) \times \left(\frac{DPC_d(t)}{DPC_d(m-24)} \right)$

Cependant, les évolutions mensuelles de la série DPC_d restent souvent très éloignées de celles de la série en date réelle comme l'illustre la figure 2 (avec d=12).

Par ailleurs, la série en date de prise en compte est décalée temporellement par rapport à la série en date réelle en raison du délai de remontée de l'information, ainsi les retournements de tendance seraient décrits avec retard. Même si ce décalage s'atténue lorsque d diminue, il subsiste : pour d=12, le décalage s'élève à 0,7 mois.

Amélioration de la méthode d'estimation

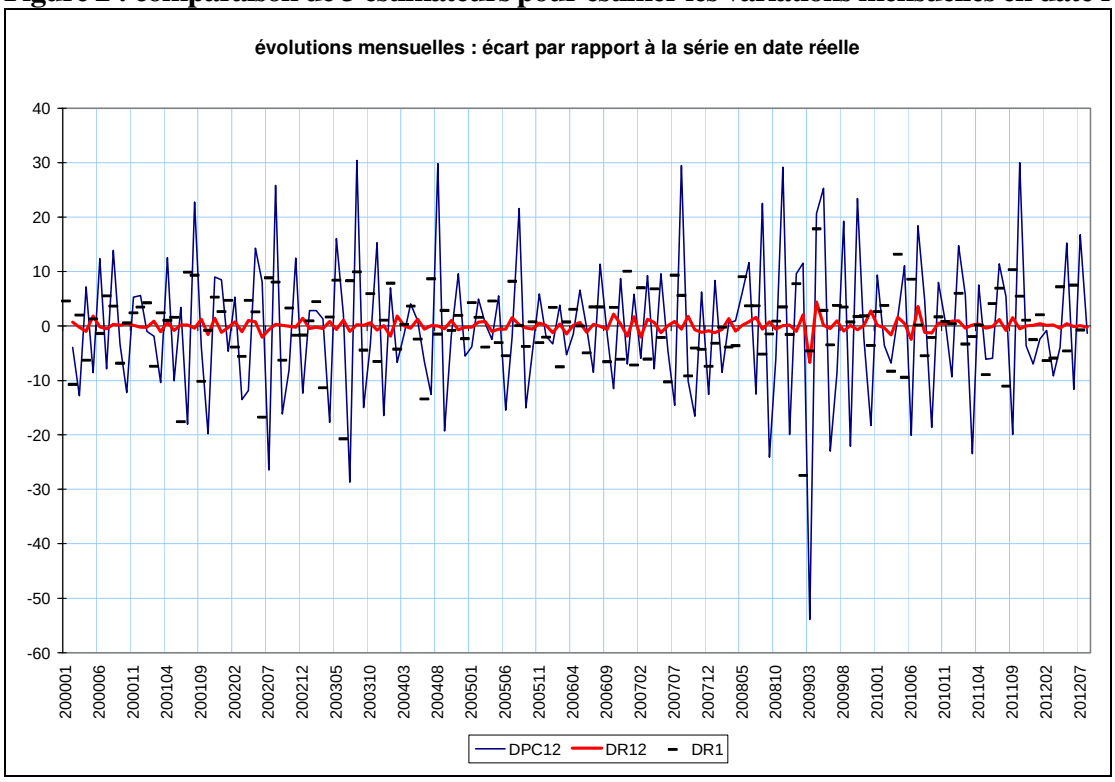
On définit la série du nombre de logements autorisés en date réelle tronquée à d, comme le nombre de logements autorisés en t et collectés avant $t+d$.

Cette série $DR_d(t)$ peut s'écrire de la manière suivante :

$DR_d(t) = A(t, t) + A(t, t+1) + \ldots + A(t, t+d-1)$.

Les séries en date réelle tronquée fournissent de meilleures estimations des évolutions en date réelle que les séries en date de prise en compte tronquée, mais dans des proportions très variables selon d. La figure 2 compare les écarts entre les évolutions de la série en date réelle par rapport à trois autres séries. La série DR_{12} est celle dont les écarts avec la série en date réelle sont les plus faibles.

Figure 2 : comparaison de 3 estimateurs pour estimer les variations mensuelles en date réelle

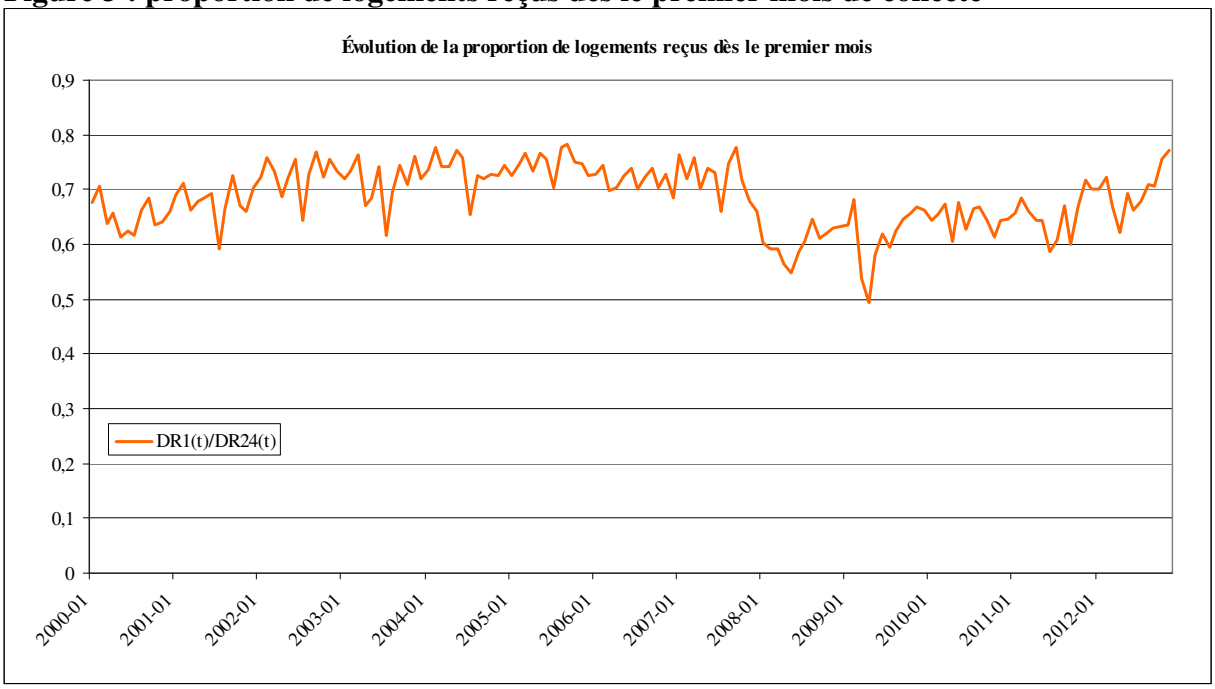


Les évolutions obtenues à partir de la série $DR_1(t)$ sont fréquemment supérieures de 5 points à celles observées in fine en date réelle, bien que $A(m, m)$ représente environ 70 % de $A(m)$. Cette proportion se révèle être relativement instable au cours du temps comme l'illustre la figure 3.

Parmi les différentes séries testées, celle dont les évolutions sont les plus proches de la série en date réelle est la série en date réelle tronquée à 12 mois. Cette série est retenue pour estimer les évolutions de la série en date réelle.

Cependant, les derniers points de cette série ne sont pas connus au moment de la diffusion et doivent être estimés.

Figure 3 : proportion de logements reçus dès le premier mois de collecte



Construction de la série en date réelle tronquée à 12 mois

La série DR_{12} s'écrit de la manière suivante :

$$DR_{12}(t) = A(t,t) + A(t,t+1) + \ldots + A(t,t+11) \qquad (2)$$

On peut relier la série $DR_{12}(t)$ avec la série des logements autorisés en date de prise en compte tronquée à 12 mois en considérant la part des logements autorisés en t dans la remontée d'information en m : $p(t,m)$ pour $m \leq t$:

$$p(t,m) = \frac{A(t,m)}{DPC_{12}(t)}$$

Ainsi chaque terme de la série $DR_{12}(t)$ se décompose de la façon suivante :

Pour i variant de 0 à 11 : $\quad A(t, t+i) = DPC_{12}(t+i) \times P(t, t+i)$

Avec $DPC_{12}(t+i)$: le nombre de logements collectés en $t+i$ dont le délai de réception de l'information est inférieur à douze mois ;

$P(t, t+i)$: la part des logements autorisés en t dans la remontée d'information en $t+i$.

Estimation de la série $DR_{12}(t)$

Soit m le dernier mois de collecte. Le terme $A(t, t+i)$ est directement observé si $t + i \leq m$. En revanche, si $t + i > m$, $A(t, t+i)$ doit être estimé. $A(t, t+i)$ est alors estimé de la manière suivante :

$$\hat{A}^m(t, t+i) = \hat{DPC}_{12}{}^m(t+i).\hat{p}^m(i)$$

$\hat{DPC}_{12}{}^m(t+i)$ est estimé via un modèle ARIMA (*cf. annexe 1*) qui prolonge la série en date de prise en compte tronquée à 12 mois.

$\hat{p}^m(i)$ est la moyenne, sur les 6 derniers mois de collecte, des parts de permis collectés i mois après l'autorisation : $\hat{p}^m(i) = \frac{1}{6}\sum_{k=0}^{5} p(m-k-i, m-k)$

La série en date réelle tronquée à 12 mois est ainsi estimée de la manière suivante :

Pour $t \leq m$ -12 : $$\hat{DR}_{12}{}^m(t) = DR_{12}(t) = \sum_{i=0}^{i=11} A(t, t+i) \qquad (4)$$

Pour m-12 < $t \leq m$ $$\hat{DR}_{12}{}^m(t) = \sum_{i=0}^{i=m-t} A(t, t+i) + \sum_{i=m+1-t}^{11} \hat{DPC}^m(t+i).\hat{p}^m(i) \qquad (5)$$

Chaînage des évolutions ou recalage de la série en date réelle tronquée

Les évolutions mensuelles estimées à partir de la série en date réelle tronquée sont chaînées au dernier point de la série stabilisée *m-24*.

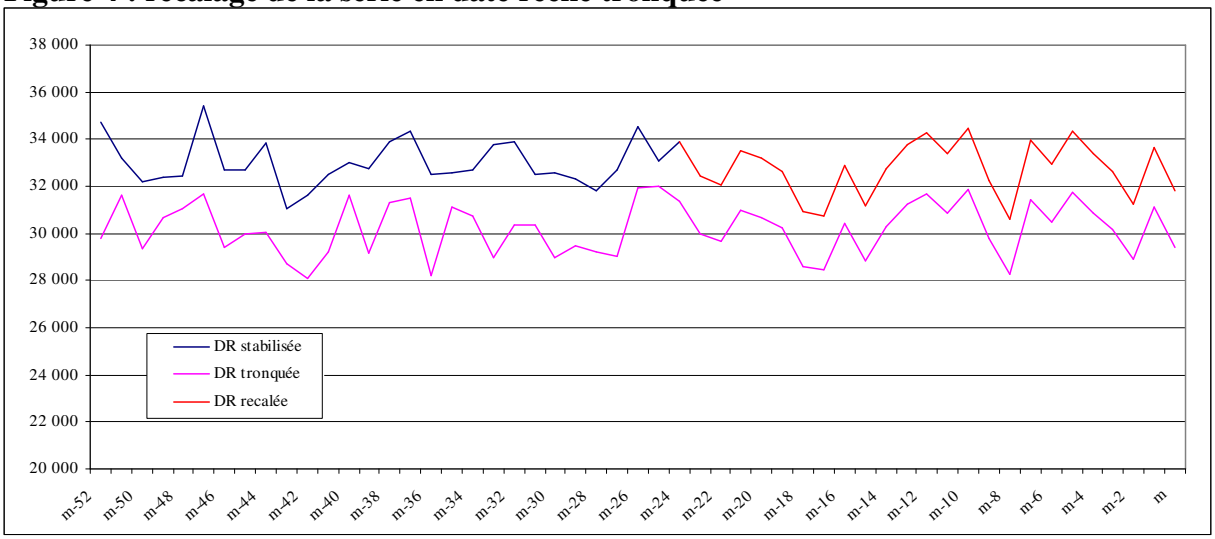
L'estimation du nombre de logements autorisés $\hat{A}(t)$ peut s'écrire de la manière suivante :

Pour t tel que $t \leq m$-24 $$\hat{A}^m(t) = A(t) = \sum_{i=0}^{24} A(t, t+i)$$

Pour t tel que m-23 $\leq t \leq m$, $$\hat{A}^m(t) = A(m-24).\left(\frac{\hat{DR}_{12}{}^m(t)}{DR_{12}(m-24)}\right) \qquad (6)$$

Ce chaînage revient à recaler la série estimée en date réelle tronquée à 12 mois au mois *m-24* comme le montre la figure 4.

Figure 4 : recalage de la série en date réelle tronquée



Mise en cohérence avec le système existant de remontée administrative

La publication d'estimations au niveau national et régional ne répond pas au besoin d'analyse sur des territoires plus petits, au niveau d'une commune par exemple. Or, pour certains territoires, la collecte administrative est satisfaisante : le système actuel répond à ces besoins. Il est donc prévu de maintenir la diffusion des séries actuelles à des niveaux géographiques fins. Ces informations seront publiées dans une partie « base administrative Sitadel » par opposition à sa partie « statistique » dans laquelle seront diffusées les nouvelles estimations.

Il sera donc possible pour un mois donné, à partir de la base administrative, de reconstituer les séries des autorisations en date réelle au niveau régional ou national et de les comparer aux estimations diffusées par ailleurs. La cohérence entre les deux systèmes d'informations (statistique et administratif) doit donc être recherchée. Afin d'assurer cette cohérence, une contrainte est imposée aux estimations en date réelle : elles ne peuvent être inférieures au niveau des autorisations en date réelle déjà collectées.

En d'autres termes, l'estimateur final retenu est le suivant :

$$\hat{A}_f{}^m(t) = \sup\left(\sum_{i=0}^{i=m-t} A(t,t+i), \hat{A}^m(t)\right) \qquad (7)$$

Où $\sum_{i=0}^{i=m-t} A(t,t+i)$ représente le nombre de logements autorisés en t d'après l'information collectée entre t et m.

Cette contrainte traduit également aussi la logique d'estimation : il s'agit de compléter la série en date réelle.

Révisions des estimations entre *m* et *m+1*

Au fil des mois, de nouvelles informations sont collectées, ce qui va générer des révisions successives des estimations. En effet, au mois m, le terme A($m,m+1$) a été estimé tandis qu'il

est directement observé en *m+1*. De plus les termes $\hat{DPC}_{12}^{\,m}(m+i)$ et $\hat{p}^{m}(i)$ sont ré-estimés. Enfin, le point de recalage est décalé de *m-24* à *m-23* ce qui est susceptible d'entrainer également des révisions.

Ainsi chaque mois, les données estimées sont remplacées, soit par des données collectées, soit par de nouvelles estimations prenant en compte les dernières informations collectées. Les révisions les plus fortes entre *m* et *m+1* sont concentrées sur le mois *m* compte tenu du poids du terme A(*m,m+1*) par apport aux autres termes A(*m,m+2*), … A(*m,m+11*). En effet, A(*m,m+1*) représente de 15 à 20 % de Â(t) contre un peu plus de 5 % pour la somme des autres termes.

Déclinaisons géographique et par type de logements

Les estimations en date réelle sont produites à différents niveaux :
- au niveau France entière pour estimer le nombre total de logements autorisés ;
- au niveau France entière par type de logement : individuels purs (IP), individuels groupés (IG), collectifs (COL) et résidences (RES) ;
- au niveau département pour trois types de logements : IP, IG, COL+RES.

La méthode est mise en œuvre au niveau national pour l'ensemble des logements d'une part, par type de bâtiment d'autre part ainsi que par type de bâtiment au niveau départemental. À noter que pour certains départements, les problèmes de collecte et/ou le faible nombre de permis rendent les estimations plus fragiles en termes de révisions potentielles.

Une cohérence entre les différentes estimations publiées est nécessaire : en effet, la somme des estimations départementales doit être égale à l'estimation nationale et la somme des estimations par type de bâtiment doit être égale au total des logements autorisés. Pour cela, deux calages successifs sont effectués (*cf. figure 5*).

Premier calage au niveau national sur le type de bâtiment

L'estimation en date réelle peut être décomposée en une partie collectée et un complément d'information « restant à collecter » estimé (*cf. formule 7*).

$$\hat{A}(t) = \hat{A}_f^{\,m}(t) = \sum_{i=0}^{i=m-t} A(t,t+i) + \hat{R}^m(t)$$

Les données collectées sont parfaitement calées entre elles : la somme du nombre de logements autorisées observés sur les 4 types de logements est égale au nombre de logements autorisés observés au niveau national.

Ainsi, pour assurer la cohérence des estimations entre le total des logements autorisés et la somme des estimations par type de logement au niveau national, il suffit de caler la partie estimée par type de logement sur la partie estimée du total national. Pour cela, la somme des logements estimés pour chaque type est calée proportionnellement sur le nombre de logements estimés sur la France entière afin d'obtenir l'égalité suivante :

$$\hat{R}^m_{total\_France}(t) = \hat{R}^m_{IP\_France}(t) + \hat{R}^m_{IG\_France}(t) + \hat{R}^m_{COL\_France}(t) + \hat{R}^m_{RES\_France}(t)$$

Pour cela, chaque estimateur initial de la partie « restant à collecter » par type de bâtiment est multiplié par le coefficient $\frac{\hat{R}^m_{total\_France}(t)}{\hat{R}^m_{IP\_France}(t) + \hat{R}^m_{IG\_France}(t) + \hat{R}^m_{COL\_France}(t) + \hat{R}^m_{RES\_France}(t)}$

Comparaison avant / après calage des estimations par type de bâtiment

Mois estimé : septembre 2015	**Partie collectée**	**Partie « restant à collecter » estimée avant calage**	**Partie« restant à collecter » estimée après calage**	**Correction de l'estimation mensuelle initiale** (%)
IP	7 432	2 788	2 814	0,3
IG	3 045	1 448	1 462	0,3
COL	13 800	3 760	3 796	0,2
RES	2 406	704	711	0,2

Estimations à fin septembre 2015 – unité : logements

Deuxième calage entre le niveau national et les départements pour chaque type de bâtiment

Le deuxième calage mis en œuvre a pour objectif de caler pour chaque type de bâtiment, la somme des estimations départementales sur l'estimation nationale.

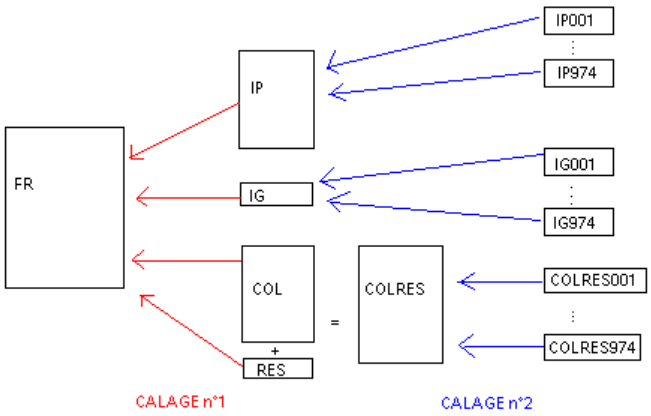
En notant IP001 par exemple l'estimation du nombre de logements individuels purs dans le département de l'Ain (code département 001), le calage mis en œuvre sur la partie estimée afin d'obtenir l'égalité suivante :

$$\hat{R}^m_{IP\_France}(t) = \hat{R}^m_{IP\_001}(t) + \hat{R}^m_{IP\_002}(t) + \ldots + \hat{R}^m_{IP\_974}(t)$$

Ce double calage assure aussi la cohérence pour l'ensemble des logements autorisés entre les départements et le niveau national.

Enfin, les données sont arrondies à l'unité. L'estimation nationale France est recalculée à l'issue du calage comme la somme des données départementales calées arrondies.

Figure 5 : calage des estimations



IV) Estimation des mises en chantier

L'estimation du nombre de logements commencés par date d'ouverture de chantier est réalisée en deux temps :

- la première étape consiste à estimer le nombre de logements réellement commencés par date d'autorisation des permis ;

- la seconde étape permet d'estimer le nombre de logements réellement commencés par date d'ouverture de chantier. Lors de cette phase, des délais d'ouverture de chantier sont affectés aux logements commencés. La plupart d'entre eux sont connus via les déclarations d'ouverture de chantier. En l'absence d'informations, des délais sont imputés.

L'ensemble des permis autorisés est segmenté en trois groupes :
- les permis « commencés » : ceux pour lesquels une déclaration d'ouverture de chantier a été réceptionnée ;
- les permis « annulés » : retrait du permis à l'initiative du pétitionnaire ou annulation administrative ;
- les permis autorisés ni annulés ni commencés.

À une date m donnée, on peut décomposer le nombre de logements autorisés en t ($t \leq m$) de la manière suivante :

$$A(t,m) = N(t,m) + C(t,m) + I(t,m)$$

Avec $N(t,m)$: logements autorisés en t puis annulés (entre t et m)
$C(t,m)$: logements autorisés en t dont les travaux ont commencé (entre t et m)
$I(t,m)$: logements autorisés en t ni annulés ni commencés au mois m

L'approche méthodologique retenue est différenciée selon la période traitée. En effet, compte tenu de la période de validité d'un permis de construire et des délais accordés pour démarrer le chantier[4], la collecte des informations « post-autorisation » (annulation ou retrait du permis, déclaration d'ouverture de chantier) des permis autorisés à une date donnée peut s'étaler sur environ trois ans.

La période d'estimation est ainsi scindée en deux parties :

- la période ancienne (ou révolue)

Pour ces permis, la collecte des informations « post autorisation » peut être considérée comme terminée : toute information sur l'annulation ou la mise en chantier aurait dû remonter dans la base Sitadel et il ne devrait plus y avoir de permis autorisé ni annulé ni commencé (soit $I(t,m)=0$). Les méthodes d'estimation mises en œuvre visent à traiter une problématique de non réponse.

[4] Les travaux doivent impérativement être commencés dans un délai de deux ans suivant l'obtention de l'autorisation d'urbanisme. Le titulaire peut demander le prolongement de son autorisation pour une durée d'un an si les travaux ne peuvent pas être commencés dans le délai de deux ans ou s'il prévoit d'interrompre le chantier pendant plus d'une année.

En pratique, cette période s'étale jusqu'à fin de l'année *N-4*. Elle est actualisée chaque année (allongement d'une année supplémentaire).

- la période « récente »

La collecte des informations n'est pas terminée au cours de cette période. L'exploitation de l'information en cours de collecte est délicate car le délai de remontée de l'information des différents évènements n'est pas homogène. Pour un ensemble de permis autorisés un même mois, les déclarations d'ouverture de chantier avec des délais courts parviennent plus rapidement que celles avec des délais importants.

Sur cette période qui s'étale de l'année N-3 à l'année en cours, les estimations seront réalisées avec des méthodes de modélisation. Des contrôles seront toutefois mis en œuvre par rapport aux données observées.

IV.1 – Détermination du nombre de logements commencés par date d'autorisation

Estimation sur période ancienne

L'objectif est d'estimer dans un premier temps la part des permis annulés (respectivement commencés) parmi les permis autorisés ni annulés ni commencés.

Des informations fiscales sur l'achèvement des travaux (formulaires H1 et H2 pour la fiscalité sur les propriétés bâties) sont mobilisées pour réduire le nombre de permis autorisés ni annulés ni commencés. Puis une enquête est réalisée annuellement auprès d'un échantillon de permis ni annulés ni commencés autorisés en *N-4*, afin d'obtenir un taux d'annulation pour ces permis.

L'exploitation de l'enquête, combinée aux informations fiscales, permet in fine de calculer deux paramètres sur le champ des permis autorisés, ni commencés ni annulés :

- le taux d'annulation des logements individuels purs (le MOA est généralement un particulier) ;
- le taux d'annulation des autres logements (le MOA n'est généralement pas un particulier).

Ces paramètres sont calculés pour la métropole d'une part et les DOM d'autre part. Pour chacun de ces territoires, les mêmes paramètres sont appliqués à l'ensemble des départements. Pour les départements d'outre-mer, le faible nombre de répondants conduit à établir des paramètres communs pour l'ensemble des départements d'outre mer et pour l'ensemble de la période 2004-2009. Faute d'enquête qualité sur les périodes les plus anciennes, c'est le paramètre 2004 qui est appliqué aux années antérieures à 2004. Le poids relativement modéré des logements autorisés ni annulés ni commencés jusqu'en 2006 rend cette approximation recevable : ils représentent moins de 5 % des logements autorisés. Le suivi des déclarations d'ouverture de chantier était de bonne qualité jusqu'à la vague de décentralisation de l'instruction de 2006 et la réforme du droit des sols intervenue fin 2007.

Des précisions (plan de sondage, estimateur utilisé, …) sur ces enquêtes sur les vieux permis sont données en annexe 2.

Les taux d'annulation obtenus via ces enquêtes sont des taux annuels. Ceux-ci sont mensualisés par une méthode de type Denton (*voir annexe 3*) qui consiste à générer une série mensuelle en minimisant les variations mensuelles sous contrainte de calage annuel. Les enquêtes et les informations fiscales permettent ainsi d'estimer un nombre de logements annulés pour la période ancienne, qui correspond à la somme des logements annulés collectés et des logements en attente estimés annulés.

On obtient alors : $\hat{N}^m(t) = N(t,m) + \hat{\alpha}^m(t).I(t,m)$ avec $\hat{\alpha}^m(t)$ le taux d'annulation estimé au mois m des permis autorisés en t et ni annulés, ni commencés.

Le nombre de logements autorisés en t et mis en chantier se déduit directement en considérant que $\hat{A}^m(t) = \hat{N}^m(t) + \hat{C}^m(t)$:

$\hat{C}^m(t) = \hat{A}^m(t) - \hat{N}^m(t)$ qu'on peut également écrire sous la forme :

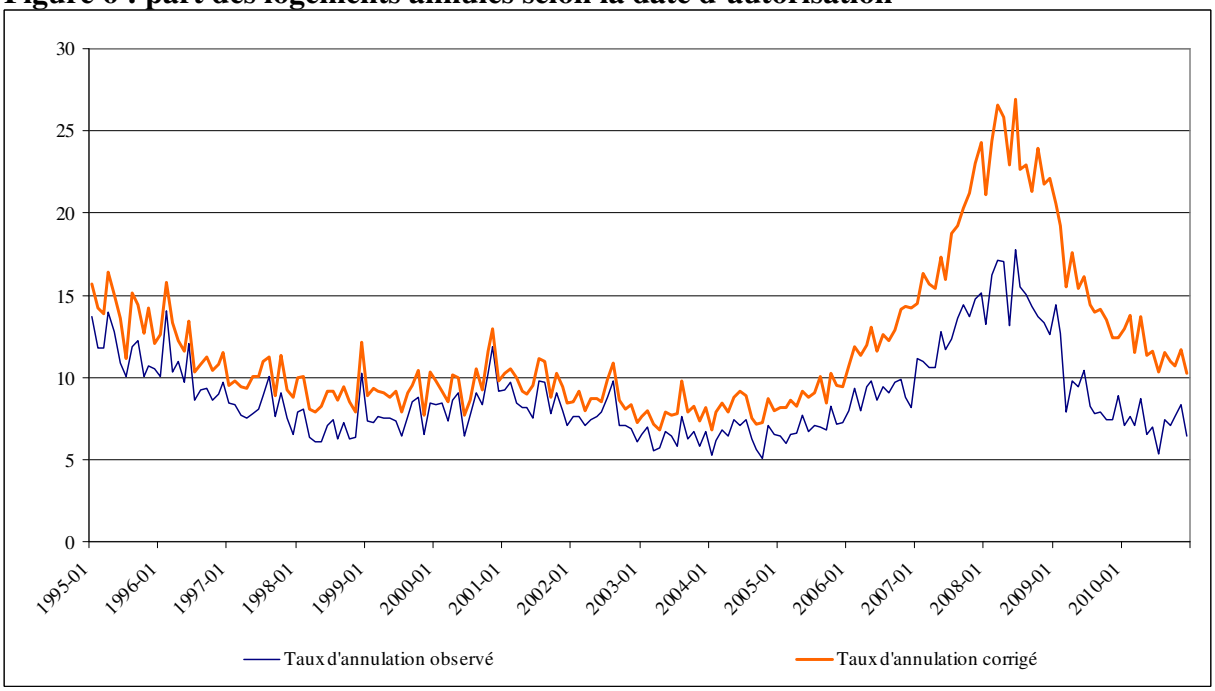
$\hat{C}^m(t) = \left(A(t,m) - N(t,m)\right) - \hat{\alpha}^m(t).I(t,m)$.

Soit :

$\hat{C}^m(t) = C(t,m) + I(t,m).(1 - \hat{\alpha}^m(t))$

Le taux d'annulation est égal à : $\hat{a}^m(t) = \frac{\hat{N}^m(t)}{\hat{A}^m(t)} = \frac{N(t,m)}{\hat{A}^m(t)} + \hat{\alpha}^m(t).\frac{I(t,m)}{\hat{A}^m(t)}$

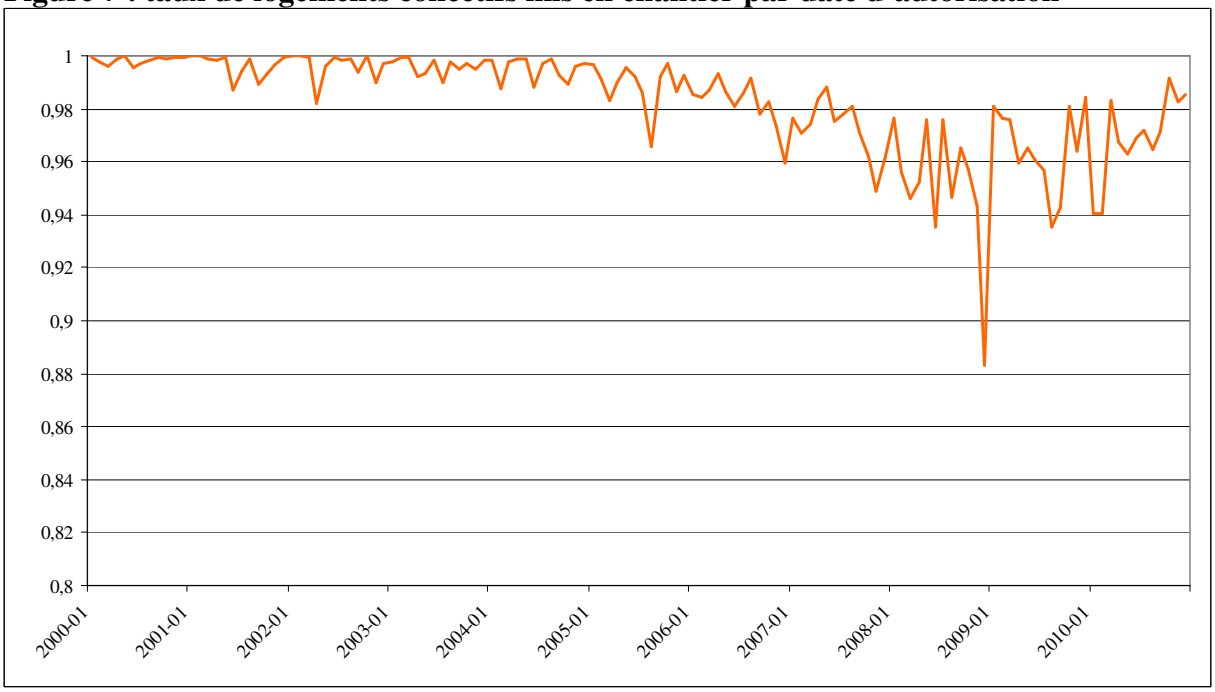
Figure 6 : part des logements annulés selon la date d'autorisation



Tranches de travaux abandonnées

Chaque projet a été comptabilisé jusqu'à présent avec la totalité du nombre de logements autorisés. Une correction supplémentaire est apportée afin de tenir compte de tranches de travaux qui ne débuteront pas. Un taux de logements mis en chantier, calculé comme le rapport entre le nombre de logements déclarés commencés et le nombre de logements autorisés de ces permis observés à partir des données disponibles dans Sitadel, est appliqué à chaque génération de logement commencés ($\hat{C}^m(t)$). Ce taux de logements mis en chantier est d'environ 95 % sur l'année 2010.

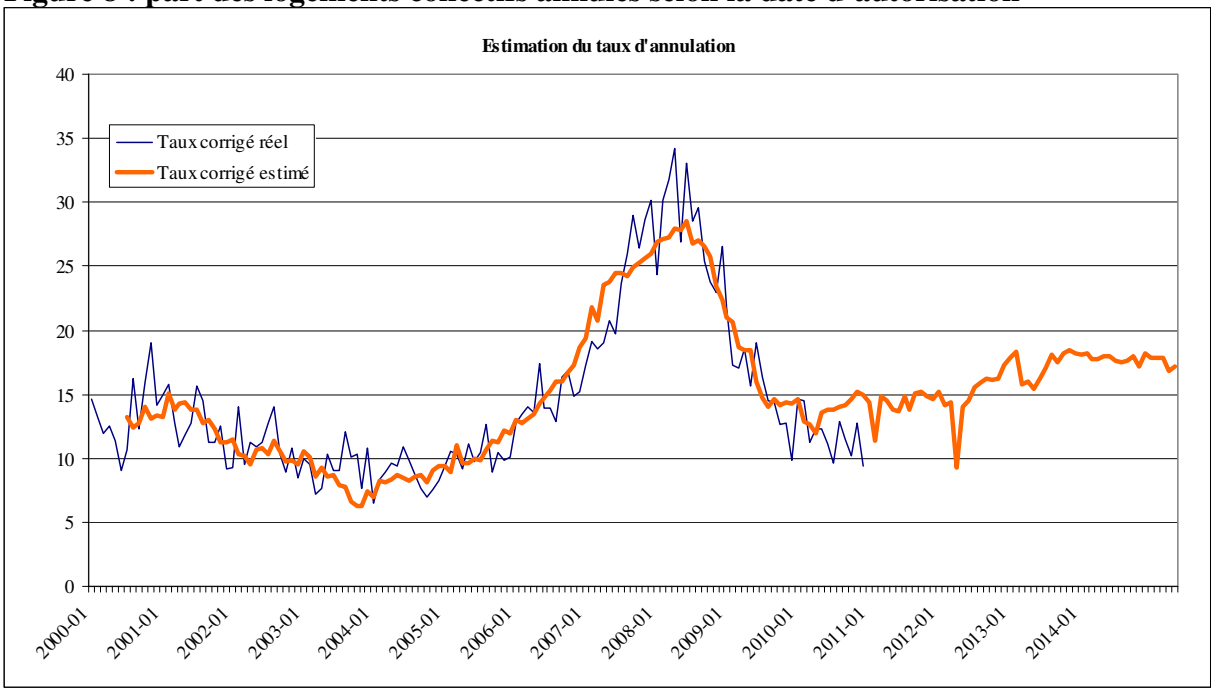
Figure 7 : taux de logements collectifs mis en chantier par date d'autorisation



Estimation sur période récente

Un modèle de régression est ensuite utilisé pour prolonger le taux d'annulation corrigé sur période récente. Quatre modèles sont établis au niveau national par type de logement : individuel pur (IP), individuel groupé (IG), collectif (COL) et résidence (RES). Ces modélisations sont construites à l'aide de variables conjoncturelles. Parmi les variables testées, ont notamment été retenus l'indice de confiance des ménages, l'indicateur de retournement conjoncturel et l'encours de logements proposés à la vente issu de l'enquête sur la commercialisation des logements neufs. Pour plus d'information sur les modèles utilisés, voir l'annexe 4.

Figure 8 : part des logements collectifs annulés selon la date d'autorisation



La figure 8 illustre les résultats de la modélisation dans le cas des logements collectifs. Le modèle est estimé à partir des taux d'annulation estimés sur la période ancienne ($\hat{a}^m(t)$).allant jusqu'à fin *N-4* (ici fin 2010). Le taux d'annulation est ensuite prédit à partir des variables conjoncturelles explicatives.

Pour s'assurer de ne pas annuler moins de logements que ce que la collecte indique, ce taux d'annulation estimé est ensuite comparé au taux d'annulation observé et corrigé le cas échéant. Dans ce cas, on applique le taux d'annulation observé.

IV.2 – Détermination du nombre de logements commencés par date de mise en chantier

Après avoir déterminé la part des permis qui se concrétiseront par des logements, il s'agit d'estimer la date à laquelle les travaux commenceront. Des délais d'ouverture de chantier sont estimés pour chaque cohorte de logements commencés ($\hat{C}^m(t)$). La plupart d'entre eux sont connus via les déclarations d'ouverture de chantier remontées dans Sitadel (il s'agit des permis entrant dans l'agrégat $C(t,m)$). En l'absence de déclaration d'ouverture de chantier, des délais sont imputés.

Imputation des délais de mise en chantier sur période ancienne

Des délais de mise en chantier sont imputés aux logements commencés estimés pour lesquels la date d'ouverture de chantier n'est pas connue (aire représentée en jaune sur la figure 9).

Sur période ancienne, la distribution des délais de mise en chantier pour la cohorte des permis autorisés à une date t est stable et l'imputation est réalisée en appliquant la distribution des délais de mise en chantier observée à partir des déclarations d'ouverture de chantier réceptionnées pour chaque mois d'autorisation *t* de la période ancienne (*cf. figure 10*). Les délais de la distribution sont compris entre 0 et 36 mois, ce qui correspond aux délais légaux

prévus pour le lancement des chantiers. La fréquence des permis mis en chantier plus de trois ans après la date d'autorisation est marginale[5] (moins de 0,5 % des logements ont une date de chantier supérieure à 36 mois sur la période 2000-2010).

Figure 9 : imputation des délais d'ouverture de chantier selon la date d'autorisation

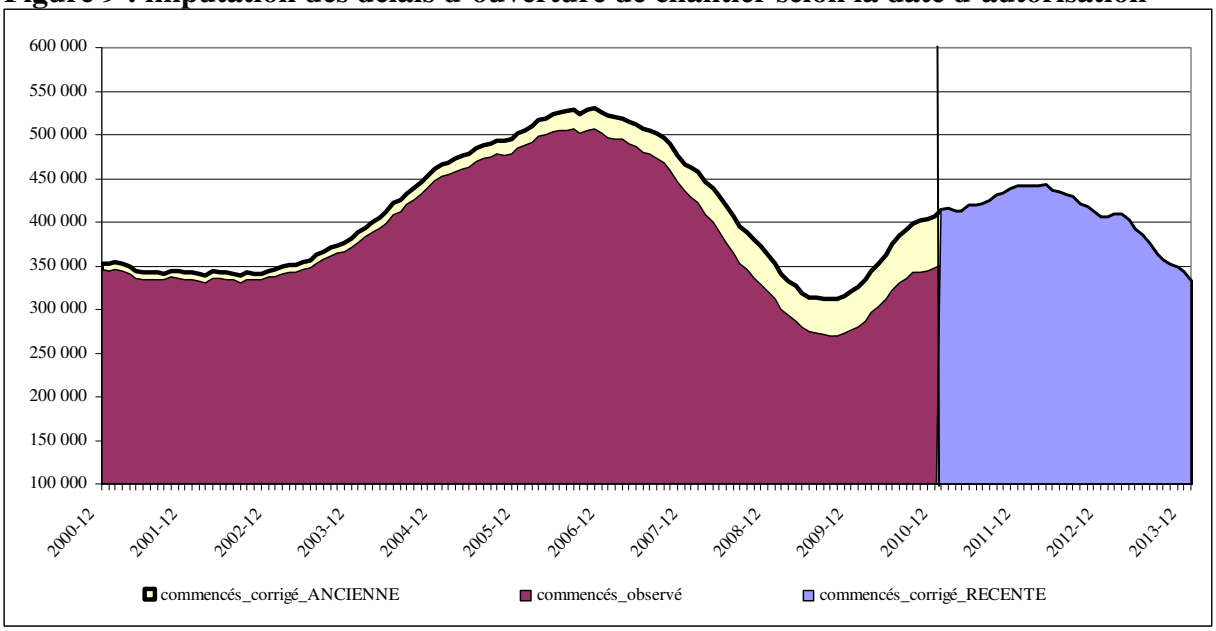
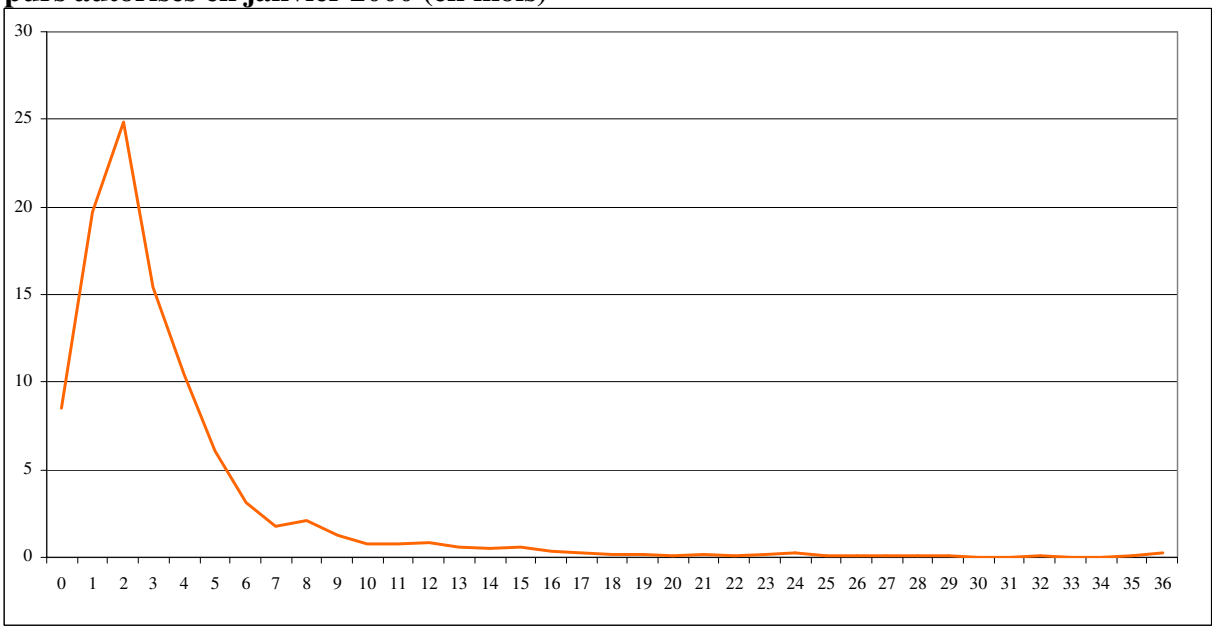


Figure 10 : distribution des délais d'ouverture de chantier des logements individuels purs autorisés en janvier 2000 (en mois)



Le nombre de logements réellement commencés ayant été autorisés en *t* se décompose de la manière suivante :

$$\hat{C}^m(t) = \sum_{k=0}^{k=36} \hat{d}_k . \hat{C}^m(t) = \sum_{k=0}^{k=36} \hat{C}_k^{\ m}(t)$$

[5] La distribution du délai de mise en chantier est recalée sur la plage 0-36 mois.

avec $\hat{d}_k$: l'estimation à partir des données collectées de la part des logements autorisés en *t* dont les travaux ont commencés avec un délai de *k* mois ($\sum_{k=0}^{k=36} \hat{d}_k = 1$)

On note $\hat{C}_k{}^m(t)$ l'estimation du nombre de logements autorisés en *t* commencés avec un délai *k*.

Imputation des délais de mise en chantier sur période récente

Sur la période récente, l'information réceptionnée concernant les ouvertures de chantier est incomplète : pour les permis autorisés depuis peu de temps, les déclarations pour lesquelles les délais d'ouverture de chantier sont importants ne sont pas encore faites et elles sont donc sous-représentées (elles vont arriver plus tardivement) par rapport à celles dont les chantiers ont déjà commencé et qui ont des délais de mise en chantier plus courts.

Le traitement mis en place consiste à imputer des délais d'ouverture de chantier à l'ensemble de la population des logements mis en chantier (aire représentée en bleu à droite sur la figure 9). L'imputation des délais de mise en chantier est réalisée en appliquant une distribution de délais (grille-délai) estimée aux logements commencés en fonction de la date d'autorisation du permis. Le délai moyen d'ouverture de chantier est estimé sur la période récente pour chaque mois d'autorisation *t*. La grille-délai est ensuite construite à partir d'une grille-délai de référence, laquelle est adaptée de façon à obtenir le délai moyen estimé.

Le délai moyen d'ouverture de chantier est estimé sur la période récente à partir de modèles économétriques établis sur période ancienne. Plus précisément, trois modèles sont établis au niveau national, les logements en résidence étant regroupés avec les logements collectifs. Parmi les variables retenues figurent le taux de l'OAT 10 ans ou le stock de logements neufs invendus. Les figures 11 illustrent les résultats des modèles pour les différents types de logements. Pour plus d'information sur les modèles utilisés, voir l'annexe 5.

Figure 11a : estimation du délai moyen pour les logements individuels purs

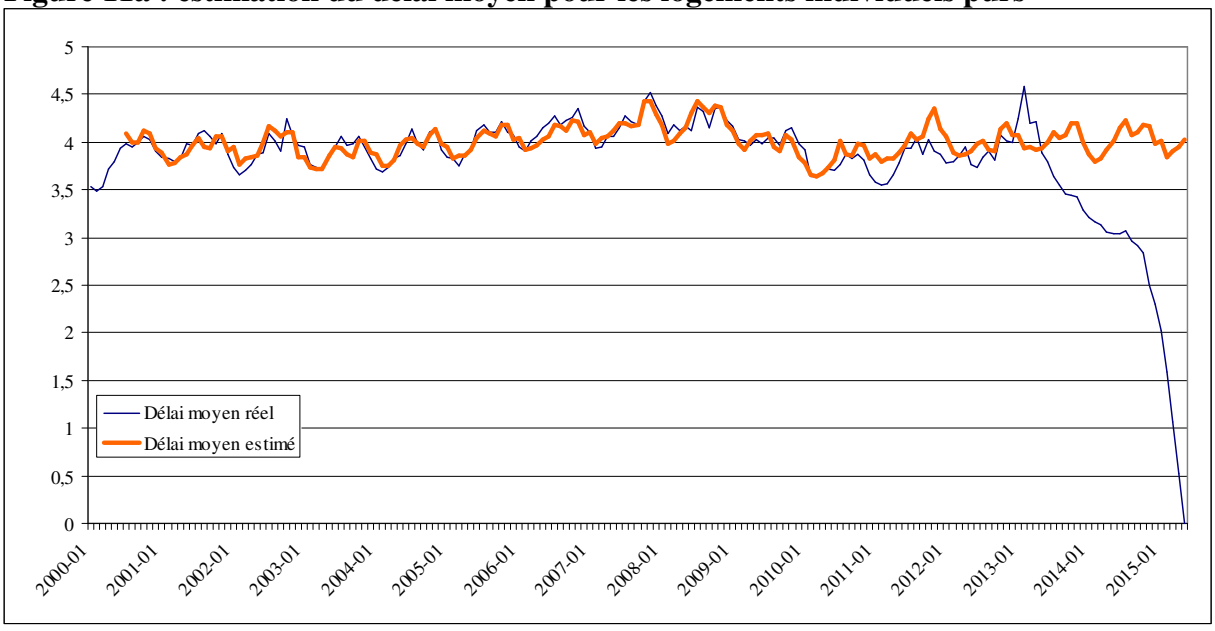


Figure 11b : estimation du délai moyen pour les logements individuels groupés

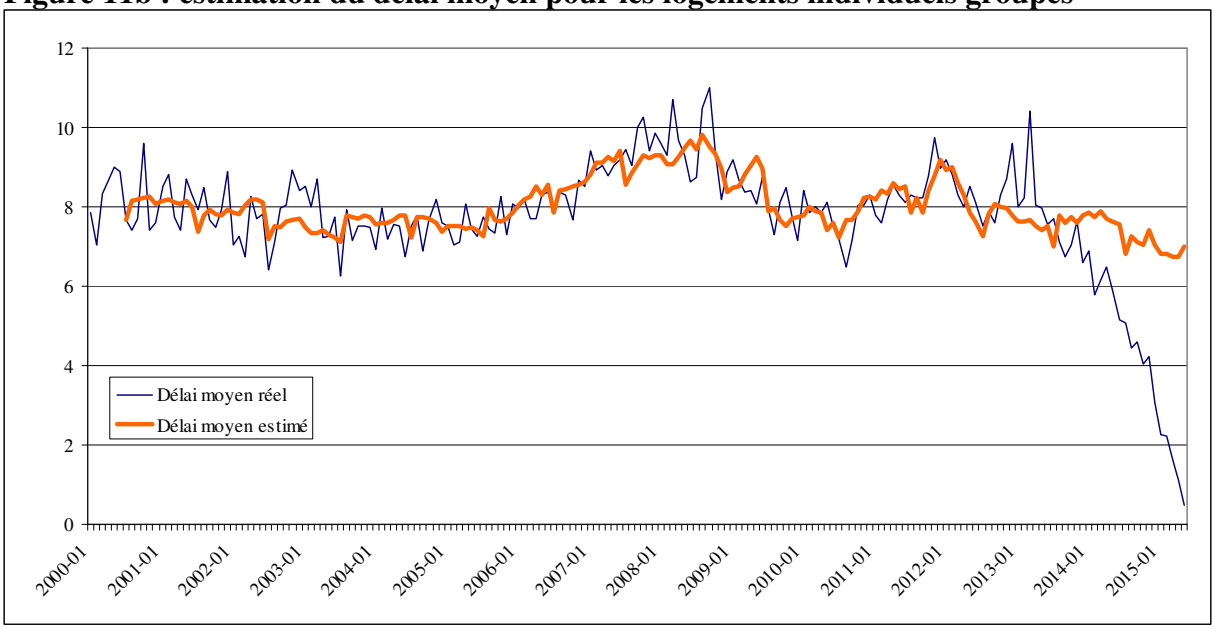
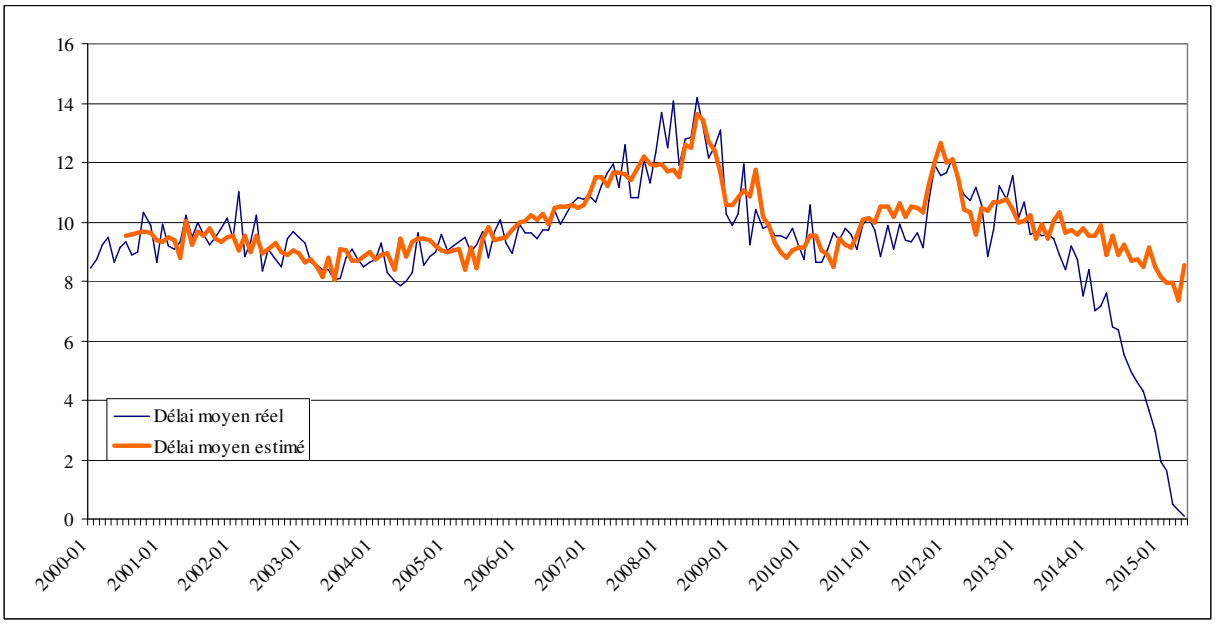


Figure 11c : estimation du délai moyen pour les logements collectifs, y compris résidences



Note de lecture : le délai moyen observé tombe mécaniquement à 0 en toute fin de période. En effet, parmi les logements autorisés en mai 2015, seuls des logements mis en chantier avec un délai de 0 mois ont pu être observés.

La distribution de délais ($\hat{d}_k$) à appliquer aux logements autorisés au mois *t* et commencés ($\hat{C}^m(t)$) est calculée en calant la distribution moyenne observée sur longue période au niveau national pour le mois t correspondant (janvier, … décembre), sur le délai moyen estimé.

La distribution observée sur longue période ($d^{ref}{}_k$) est ainsi déformée de manière homothétique à partir d'un point d'inflexion afin de se caler sur le délai moyen imposé. Le mois correspondant au point d'inflexion, noté $m_{inflexion}$ est fixé à mi distance entre le délai moyen sur longue période et le délai moyen estimé.

$\forall k < m_{inflexion}$, $\hat{d}_k = K_1 . d^{ref}{}_k$

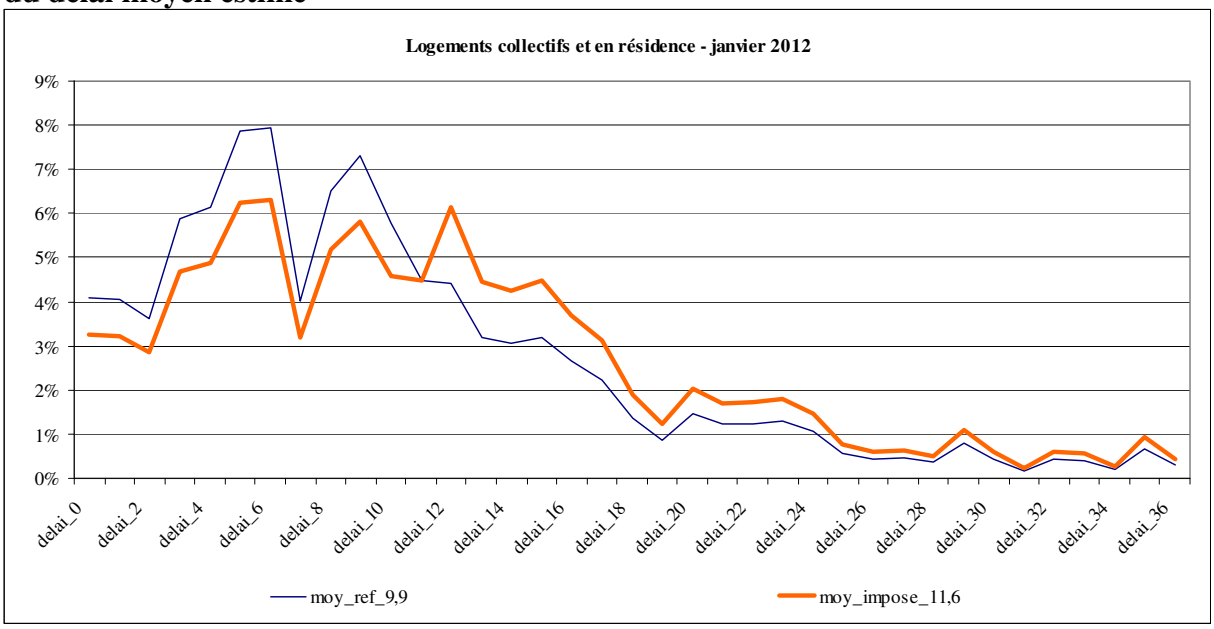
$Si\ k = m_{inflexion}$ $\hat{d}_k = d^{ref}{}_k$

$\forall k > m_{inflexion}$, $\hat{d}_k = K_2 . d^{ref}{}_k$

K_1 et K_2 sont déterminés de telle sorte que $\sum_{k=0}^{k=36} \hat{d}_k = 1$ et $\sum_{k=0}^{k=36} k.\hat{d}_k = délai\_moyen\_estimé$

Voir annexe 6 pour plus d'informations sur la fonction de déformation.

Figure 12 : déformation de la fonction de répartition moyenne pour tenir compte du délai moyen estimé



Note de lecture : dans cet exemple, la distribution de délais d'ouverture de chantier observée sur longue période (moyenne de 9,9 mois) est déformée pour être calée sur un délai moyen de 11,6 mois. Le point d'inflexion est donc positionné à 11 mois

Contrôle entre données collectées et estimées

Un contrôle entre les données observées et imputées est réalisé après la phase d'imputation (*cf. figure 13*) :

- pour un mois d'autorisation *t* donné, si le nombre de logements commencés imputés avec le délai *k* (noté $\hat{C}_k{}^m(t)$) est inférieur au nombre de logements commencés observé avec ce délai (noté $\overset{obs}{C}_k{}^m(t)$), alors le nombre de logements commencés imputé avec le délai k est corrigé pour être égal au nombre de logements observés ;

- la correction à la hausse apportée sur certains délais est répartie proportionnellement dans le sens contraire aux autres délais.

On note $\hat{\hat{C}}_k{}^m(t)$ l'estimation corrigée du nombre de logements autorisés en *t* commencés avec un délai *k*.

Soit la fonction $\mathbb{1}_k(t)$: $\forall k$ de 0 à 36 $\quad \mathbb{1}_k(t) = 1$ si $\hat{C}_k^m(t) < \overset{obs}{C}_k^m(t))$

$\mathbb{1}_k(t) = 0$ sinon

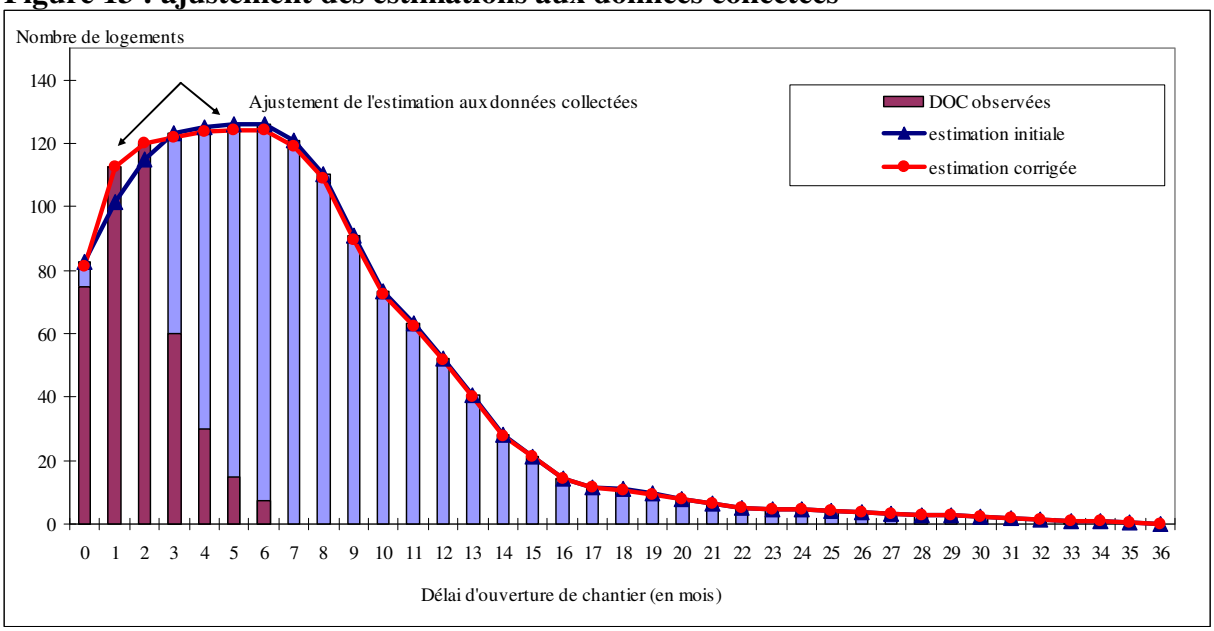
Soit $\hat{S}_1(t) = \sum_{k=0}^{k=36} (\overset{obs}{C}_k^m(t) - \hat{C}_k^m(t)) . \mathbb{1}_k(t)$ et $\hat{S}_2(t) = \sum_{k=0}^{k=36} (\hat{C}_k^m(t) - \overset{obs}{C}_k^m(t)) . (1-\mathbb{1}_k(t))$

si $\hat{C}_k^m(t) < \overset{obs}{C}_k^m(t))$, $\quad \hat{\hat{C}}_k^m(t) = \overset{obs}{C}_k^m(t)$

Sinon $\quad \hat{\hat{C}}_k^m(t) = \hat{C}_k^m(t) - \hat{S}_1(t) . (\hat{C}_k^m(t) - \overset{obs}{C}_k^m(t)) / \hat{S}_2(t)$

Cette phase de contrôle permet de garantir qu'en cas de collecte parfaite, l'estimation est in fine égale à la donnée collectée.

Figure 13 : ajustement des estimations aux données collectées



Calcul du nombre de logements mis en chantier par date de mise en chantier

Le nombre de logements commencés à la date d'ouverture de chantier *t* est la somme des chantiers autorisés à la date *t-k* avec un délai *k (avec* $0 \leq k \leq 36$).

Soit $\hat{K}(t)$ l'estimation du nombre de logements commencés en *t*,

$$\hat{K}^m(t) = \sum_{k=0}^{k=36} \hat{C}_k^m(t-k)$$

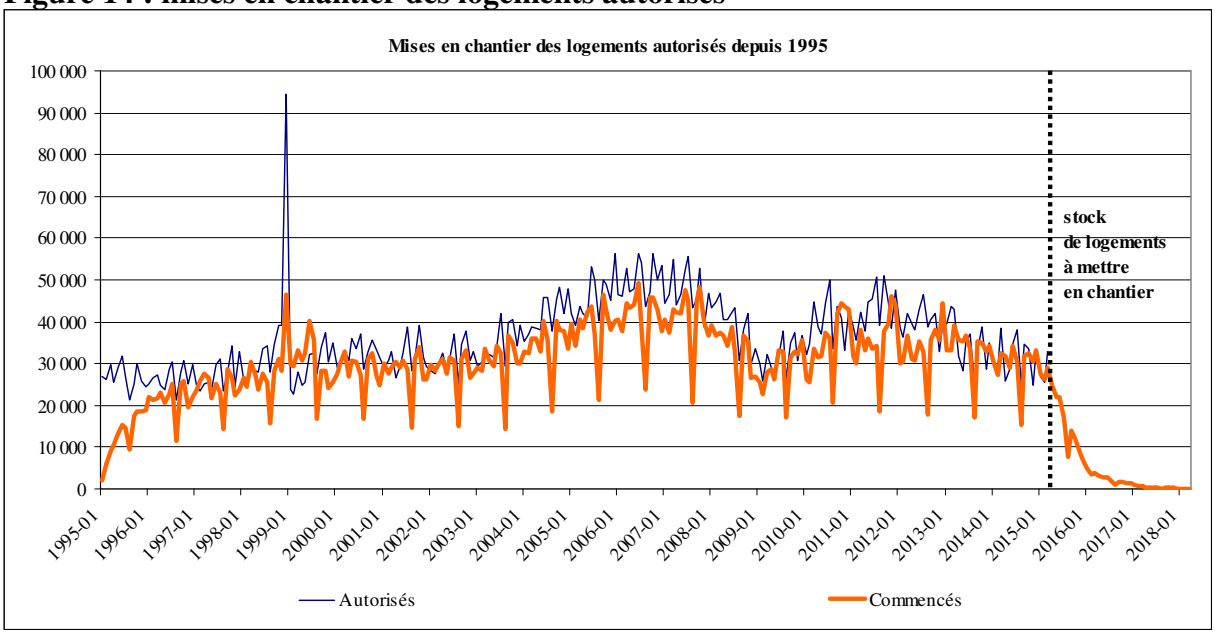
Les estimations sont réalisées à partir des permis autorisés en janvier 1995. Les estimations des logements commencés par date d'ouverture de chantier sont diffusées à partir de janvier 2000 compte tenu des délais de mises en chantier.

La méthode mise en œuvre assure un équilibre comptable dans le temps entre autorisations de construire, annulations et mises en chantier. Ce bouclage entre les autorisations et les mises en chantier permet de construire un nouvel indicateur : **le stock de logements restant à mettre**

en chantier (*cf. figure 14*). Il reflète le stock de chantiers autorisés qui seront mis en chantier dans les prochains mois, compte tenu des taux d'annulation estimés et des délais d'ouverture de chantier estimés.

$$\hat{Stock}(m) = \sum_{h=0}^{h=35} \sum_{k=m+1-h}^{k=36} \hat{C}_k^{\,m}(m-h)$$

Figure 14 : mises en chantier des logements autorisés



Estimations départementales

Le même processus est mis en œuvre pour produire des estimations départementales. Même si globalement la conjoncture est considérée comme identique dans l'ensemble des départements, les paramètres appliqués sont modulés selon des critères géographiques pour refléter des écarts structurels.

Pour chaque département et pour chaque type de bâtiment (individuel pur, individuel groupé, collectif & résidence), les taux d'annulation mensuels estimés sur la période ancienne sont prolongés par le profil de la série nationale sur la période récente.

Concernant les délais d'ouvertures de chantier, le nombre d'observations mensuelles par département ou par région peut être relativement faible pour certains types de logement. De ce fait, les distributions de délais de mise en chantier et le délai moyen mensuel observés au niveau local sont parfois peu robustes. Les paramètres de délais utilisés pour imputer des délais de mise en chantier sont ainsi calculés au niveau de la région (mêmes paramètres appliqués à l'ensemble des départements de la région).

Les grilles délais utilisées pour imputer des délais de mise en chantier au niveau régional seront construites selon le modèle suivant :

- des grilles de délais mensuelles « de référence » sont établies pour la métropole d'une part et pour l'ensemble des DOM d'autre part ;

- ces grilles délais sont déformées pour se caler sur un délai moyen régional calculé : il s'agit du délai moyen France (observé sur période ancienne, prédit sur période récente) corrigé de l'écart structurel observé pour la région.

Cet écart structurel est calculé par type de logement (individuel pur, individuel groupé, collectif & résidence) comme la différence entre le délai moyen de mise en chantier observé dans la région sur la période ancienne et celui observé au niveau national.

Des contrôles sont réalisés au niveau départemental pour chaque mois d'autorisation t de la période récente afin de tenir compte de l'information collectée, de manière analogue à ceux mis en œuvre pour les estimations nationales.

IV.3 – Calage des estimations départementales sur les estimations nationales

Les estimations départementales sont calées sur les estimations nationales.

Chaque mois, la somme des estimations des logements commencés par département doit être égale à l'estimation nationale par date d'ouverture de chantier.

Le calage mis en œuvre doit également veiller à ce que pour chaque département, le nombre de logements commencés ventilés par date d'ouverture de chantier sur l'ensemble de la période soit égal au nombre de logements autorisés nets des annulations (soit la somme des logements commencés par date d'autorisation sur la période estimée).

De plus, il faut veiller à ce que le nombre de logements commencés estimés par date d'ouverture de chantier soit supérieur ou égal au nombre de logements commencés observé chaque département.

Marges de calage des estimations départementales du nombre de logements commencés (estimations à fin mai 2015)

	1995-01	…	2018-05	Total période
dpt 001				$\sum_{t=199501}^{t=201505} \hat{C}_{dpt\_001}^{m}(t)$
…				…
dpt 100				$\sum_{t=199501}^{t=201505} \hat{C}_{dpt\_100}^{m}(t)$
Σ dpt	$\hat{K}(199501)$		$\hat{K}(201805)$	$\sum_{t=199501}^{t=201505} \hat{C}^{m}(t)$

⇨ Le calage est effectué avec la macro CALMAR sur la différence entre l'estimation du nombre de logements commencés en *t* et le nombre de logements commencés en *t* issus de la collecte par la voie administrative.

Les données sont arrondies à l'unité. L'estimation nationale France est recalculée à l'issue du calage comme la somme des données départementales calées arrondies.

V) Estimation des surfaces de logements autorisés et commencés

Une estimation des surfaces de plancher des logements est également produite. La même méthode est employée pour les autorisés et les commencés. Soit :

- $\hat{X}^{m}(t)$ la série du nombre de logements autorisés (ou commencés) estimés le mois m en date réelle
- $X(t,m)$ la série du nombre de logements autorisés (ou commencés) observés en date réelle le mois *m*
- $S(t,m)$ la série de surface de plancher des logements autorisés (ou commencés) observés en date réelle

L'objectif est de produire une série de surface de plancher des logements autorisés (ou commencés) estimés en date réelle notée $\hat{S}^{m}(t)$.

La surface de plancher d'une partie des logements autorisés estimés en date réelle est connue. Il convient d'estimer la surface des logements dont l'autorisation n'a pas encore été reçue. Une surface moyenne par logement est calculée à partir des données observées, elle est notée $\bar{s}^{m}(t)$ et calculée de la manière suivante : $\bar{s}^{m}(t) = \dfrac{S(t,m)}{X(t,m)}$

Cette série est ensuite lissée (tendance obtenue par procédure x11) afin de construire un paramètre d'imputation robuste aux aléas de collecte. La série lissée est notée $sl^{m}(t)$. Ce paramètre est ensuite appliqué aux logements dont l'autorisation n'a pas encore été reçue :

$$\hat{S}^{m}(t) = S(t,m) + \left(\hat{X}^{m}(t) - X(t,m)\right) sl^{m}(t)$$

Les coefficients de surface moyenne sont calculés au niveau France pour les types de logements suivants : individuels purs, individuels groupés, collectifs et résidences.

Les séries d'estimation des surfaces des logements autorisés (ou commencés) sont produites au niveau départemental, régional et national en utilisant les coefficients de surface moyenne France.

VI) Diffusion des séries d'estimations en date réelle

Ces nouveaux indicateurs sont désormais utilisés par le SOeS pour suivre la construction neuve de logement.

Chaque mois, le SOeS publie un « Chiffres & statistiques » contenant les derniers résultats. Cette diffusion s'accompagne de la mise en ligne des séries mensuelles depuis 2000. Ces séries sont intégralement révisées chaque mois pour tenir compte de l'évolution de la collecte. Les séries de logements autorisés et commencés sont disponibles aux niveaux départemental, régional et national. Les séries de taux d'annulation et de délais moyens sont disponibles par types de logement. Le stock de logements à mettre en chantier est diffusé uniquement au niveau national.

Les chiffres sont arrondis à la centaine la plus proche pour deux raisons. Tout d'abord, s'agissant d'une estimation statistique, ce choix indique la précision des indicateurs ; par ailleurs, cela permet de marquer la différence avec les résultats issus de la collecte administrative.

Ces données ne comprennent pas le département de Mayotte pour lequel la collecte est encore trop récente et trop instable pour permettre des estimations selon les méthodes décrites ici.

La diffusion de ces nouveaux indicateurs a conduit à réviser de manière significative le nombre de logements commencés sur la période 2010-2014. Pour préciser les raisons ayant conduit à ces changements, expliquer les différences de concepts, et décrire les principaux résultats, un numéro spécial a été publié en février 2015 dans la collection « Chiffres & statistiques ».

Ce document est disponible sur le site du SOeS :
www.statistiques.developpement-durable.gouv.fr
Rubrique Logement – construction/Construction/Logements/Premiers Résultats/Chiffres & statistiques
Numéro spécial : de nouveaux indicateurs pour suivre la construction de logements

Annexes

Annexe 1 : modélisation de la série des logements autorisés tronquée

Le prolongement de la série de logements autorisés en date de prise en compte tronquée à 12 mois, est estimé via un modèle ARIMA.

L'autocorrélation de la série non différenciée décroît lentement vers 0, donc la série n'est pas stationnaire (*voir figure 15*). Un test de Dickey-Fuller confirme cette non-stationnarité.

Figure 15 : autocorrélogramme de la série DPC tronquée à 12 mois.

Autocorrelations				
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	45678579	1.00000	|********************|	0
1	35148609	0.76948	|*************** . |	0.078811
2	33537111	0.73420	|*************** . |	0.116475
3	33462877	0.73257	|*************** . |	0.142347
4	30208297	0.66132	|************* . |	0.164101
5	29955925	0.65580	|************* . |	0.179895
6	28643038	0.62706	|************* . |	0.194177
7	26761715	0.58587	|************ . |	0.206371
8	25927989	0.56762	|*********** . |	0.216456
9	27822973	0.60910	|************ . |	0.225511
10	24983177	0.54693	|*********** . |	0.235508
11	23827087	0.52162	|********** . |	0.243270
12	26108696	0.57157	|*********** . |	0.250120
13	20445106	0.44759	|*********. . |	0.258106
14	18999840	0.41595	|******** . . |	0.262882
15	20156173	0.44126	|********* . . |	0.266939
16	14671388	0.32119	|****** . . |	0.271432
17	11857334	0.25958	|***** . . |	0.273782
18	10611672	0.23231	|***** . . |	0.275307
19	8344224	0.18267	|**** . . |	0.276521
20	7453785	0.16318	|*** . . |	0.277270
21	6394841	0.14000	|*** . . |	0.277866
22	2650419	0.05802	|* . . |	0.278304
23	2162956	0.04735	|* . . |	0.278379
24	4216723	0.09231	|** . . |	0.278429

L'autocorrélogramme de la série différenciée (*figure 16*) montre que les corrélations multiples de 12 décroissent lentement vers 0, preuve de la saisonnalité de la série. Lors de la procédure X12 permettant le prolongement de la série, une différenciation est effectivement pratiquée dans le modèle retenu : il s'agit d'un modèle de type Airline, (0,1,1) (0,1,1). Ce modèle est sélectionné car il minimise le critère BIC. Les paramètres sont présentés figure 17.

Figure 16 : autocorrélogramme de la série DPC tronquée à 12 mois différenciée

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	20469499	1.00000	|**********|	0
1	-8814412	-.43061	*********| . |	0.079057
2	-1095186	-.05350	. *| . |	0.092563
3	2998671	0.14649	. |***. |	0.092756
4	-2825965	-.13806	.***| . |	0.094191
5	727366	0.03553	. |* . |	0.095447
6	833716	0.04073	. |* . |	0.095530
7	-927270	-.04530	. *| . |	0.095638
8	-2433710	-.11889	. **| . |	0.095772
9	4414205	0.21565	. |**** |	0.096690
10	-1344475	-.06568	. *| . |	0.099651
11	-3858452	-.18850	****| . |	0.099921
12	7794433	0.38078	. |******** |	0.102119
13	-4366533	-.21332	****| . |	0.110638
14	-2522737	-.12324	. **| . |	0.113180
15	6560995	0.32053	. |****** |	0.114015
16	-2581739	-.12613	. ***| . |	0.119514
17	-1021923	-.04992	. *| . |	0.120343
18	755763	0.03692	. |* . |	0.120473
19	-1058289	-.05170	. *| . |	0.120543
20	-45631.837	-.00223	. | . |	0.120682
21	2666392	0.13026	. |*** . |	0.120682
22	-3069155	-.14994	. ***| . |	0.121558
23	-2942033	-.14373	. ***| . |	0.122708
24	5894901	0.28798	. |****** |	0.123756

Figure 17 : paramètres du modèle Airline retenu

Exact ARMA Maximum Likelihood Estimation					
For Variable logdpct					
Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal MA	1	0.59546	0.06573	9.06	<.0001
Seasonal MA	12	0.74891	0.07263	10.31	<.0001

Annexe 2 : l'enquête qualité sur les vieux permis autorisés ni commencés ni annulés

Le champ de l'enquête est constitué des permis autorisés l'année *N-4* qui ne sont ni commencés ni annulés (permis en attente) au moment de la réalisation de l'enquête.

Plan de sondage de l'enquête

L'échantillon des permis est tiré en deux phases (sondage à deux degrés) :

> 1er tirage : tirage des communes à probabilité inégale, directement proportionnelle au nombre de permis en attente à enquêter par commune
>
> 2e tirage : tirage aléatoire simple sans remise des permis à enquêter
>
> Nombre de permis à enquêter par commune : 20 permis maximum (seuil fixé pour optimiser les réponses des communes, des listes trop importantes à remplir pouvant s'avérer dissuasives).
>
> Parmi les permis tirés, on distingue les permis individuels purs des autres permis (groupés, collectifs et en résidence). S'il y a plus de 10 permis à enquêter par type de bâtiment, on sélectionne 10 permis de chaque type.

Soit N le nombre de permis en attente (*N* = 15 420 permis dans l'enquête 2010),
Soit nbcomtot le nombre de communes avec des permis en attente (*Enquête 2010 : nbcomtot* = 6 825),
Soit nbcomint le nombre de communes que l'on souhaite interroger (*Enquête 2010 : nbcomint = 335),*
Soit ni le nombre de permis en attente de la commune *i*,

La probabilité *Pi* de tirer la commune *i* est : $Pi = ni * nbcomint / N$

Calcul de la probabilité de sélection d'un permis *j*, en attente pour la commune *i* :

La commune *i* est sélectionnée, elle a *ni* permis en attente, dont *ni_ind* permis individuels purs et *ni_autres* permis autres (*ni = ni_ind + ni_autres*)

Nombre de permis interrogés pour la commune : Min(20,*ni)*
Nombre de permis individuels purs interrogés : Min(*ni_ind* ; Max (10, 20 - *ni_autres*))
Nombre de permis autres interrogés : Min(*ni_autres* ; Max (10, 20 - *ni_ind*))

⇨ $Pj\_ind \mid i = \text{Min}(ni\_ind ; \text{Max}(10, 20 - ni\_autres)) / ni\_ind$
⇨ $Pj\_autre \mid i = \text{Min}(ni\_autres ; \text{Max}(10, 20 - ni\_ind)) / ni\_autre$

La probabilité de sélection du permis *j* de la commune *i* est :

$Pi,j\_ind = Pi * Pj\_ind \mid i = (ni / N) * nbcomint * \text{Min}(ni\_ind ; \text{Max}(10, 20 - ni\_autres)) / ni\_ind$

$Pi,j\_autre = Pi * Pj\_autre \mid i = (ni / N) * nbcomint * \text{Min}(ni\_autres ; \text{Max}(10, 20 - ni\_ind)) / ni\_autre$

Estimation du nombre de logements en attente annulés

L'estimateur d'Horvitz-thompson du nombre de logements en attente annulés (pour chaque type) est :

$\hat{A}nnulés = \sum_{k=1 \text{ à } n} (A_k * Logements_k / P_k)$

où $Ak = 1$ si le permis k a été annulé, 0 sinon
$Logements_k$ est le nombre de logements autorisés du permis k
P_k la probabilité d'inclusion du permis k

Annexe 3 : mensualisation des taux d'annulation annuels

La solution mise en œuvre consiste à ajuster une série mensuelle sur jalons annuels. La technique choisie (méthode de Denton) consiste à déterminer la série de façon que son écart aux observations soit lisse.

Soit m le nombre de jalons annuels disponibles et soit n le nombre total d'observations (n=12*m).

Soit z'=[z1 ... zn] la série mensuelle originale, soit y'=[y1 ... ym] la série de jalons annuels, il faut trouver une série x'=[x1 ... xn] telle que :

$\sum_{3(T-1)+1}^{3T} x_t = y_t$ pour T [1 ... m] et qui minimise une fonction que l'on notera P(x,z).

P(x,z) correspond en fait à une forme quadratique qui peut se mettre sous la forme P(x,z)=(x-z)'A(x-z) avec A matrice carrée symétrique de taille n qui reste à définir.

La résolution de cette minimisation sous contraintes peut se résoudre à l'aide d'un Lagrangien de la forme : L= (x-z)'A(x-z) -2λ'(y-B'x) où λ'=[λ1 ... λm] et B une matrice de taille (n,m) telle que :

$$B' = \begin{bmatrix} 111111111111 & 000000000000 & & 000000000000 \\ 000000000000 & 111111111111 & & 000000000000 \\ & & ... & \\ 000000000000 & 000000000000 & & 111111111111 \end{bmatrix}$$

La solution est de la forme x = z + Cr avec r = y-B'z
C= $A^{-1}B(B'A^{-1}B)^{-1}$

Choix de A : $P(x,z) = \sum_{t=1}^{t=n} [\Delta(x_t - z_t)^2]$ avec $\Delta(x_t) = x_t - x_{t-1}$

Le vecteur différence du premier ordre peut s'écrire D(x-z) avec D, matrice carrée de taille n telle que :

$$D = \begin{bmatrix} 1 & 0 & 0 & . & & . \\ -1 & 1 & 0 & . & & . \\ 0 & -1 & 1 & . & & . \\ . & . & . & . & . & . \\ & & & . & -1 & 1 \end{bmatrix}$$ et A=D'D

Choix de Z : série constante égale au 1er jalon.

Référence : « Adjustment of Monthly or Quarterly series to annual totals : An Approach Based on Quadratic minimization », article de frank Denton, journal of the American Statistical Association, March 1971

Annexe 4 : modélisation du taux d'annulation corrigé sur période ancienne

On modélise le taux d'annulation mensuel corrigé au niveau national sur période ancienne par type de bâtiment (logements individuels purs, logements individuels groupés, logements collectifs, logements en résidence).

Plusieurs variables conjoncturelles ont été testées pour modéliser le taux d'annulation. Il s'agit de variables disponibles sur les sites de l'Insee, de la banque de France ou du SOeS telle que l'indice de confiance des ménages ou le niveau de l'en cours de logements proposés à la vente. L'option stepwise a été utilisée pour établir la liste des variables à prendre en compte dans le modèle. Parmi les variables explicatives utilisées, certaines sont « retardées » (principale raison : disponibilité de la variable mobilisée fin *m+1* pour estimer le mois *m*). Les variables trimestrielles mobilisées sont mensualisées au préalable par le processus décrit précédemment.

Logements individuels purs

Variables explicatives du taux d'annulation mensuel corrigé (mois t) :

Variable	Source	Période	Valeur coefficient	Erreur type
En cours de logements proposés à la vente (trim)	ECLN	t-6	0.00006776	0.00000358
Indice de confiance des ménages	Insee	t-1	0.04835	0.00845
Climat des affaires	Insee	t-1	-0.02783	0.00473
Stock de logements invendus (trim)	ECLN	t-3	-0.06306	0.01082
Création d'entreprise dans les activités immobilières	Insee	t-1	-0.00101	0.00037653
Constante			4.43620	0.99578

R^2 ajusté : 0,90

Logements individuels groupés

Variables explicatives du taux d'annulation mensuel corrigé (mois t) :

Variable	Source	Période	Valeur coefficient	Erreur type
En cours de logements proposés à la vente (trim)	ECLN	t-6	0.00032774	0.00001448
Indice de confiance des ménages	Insee	t-1	-0.24152	0.03438
Indice de retournement	Insee	t-1	-1.65209	0.35173
Climat des affaires	Insee	t-1	0.47374	0.03019
Constante			-27.01374	4.14036

R^2 ajusté = 0,89

Logements collectifs

Variables explicatives du taux d'annulation mensuel corrigé (mois t) :

Variable	Source	Période	Valeur coefficient	Erreur type
En cours de logements proposés à la vente (trim)	ECLN	t-6	0.00023099	0.00001457
IPI construction	Insee	t-2	0.30317	0.07218
Climat des affaires	Insee	t-1	0.28513	0.02695
Indice de retournement	Insee	t-1	-1.35752	0.25659
Constante			-60.55271	5.74370

R^2 ajusté = 0,88

Logements en résidence

Variables explicatives du taux d'annulation mensuel corrigé (mois t) :

Variable	Source	Période	Valeur coefficient	Erreur type
En cours de logements proposés à la vente (trim)	ECLN	t-6	0.00046137	0.00003998
Création d'entreprise dans la construction	Insee	t-1	-0.00177	0.00057275

Modèle sans constante.

Annexe 5 : modélisation du délai moyen de mise en chantier

Les délais moyens de mise en chantier sont modélisés au niveau national sur la période ancienne à partir de variables conjoncturelles ainsi que des variables indicatrices décrivant chaque mois de l'année (sauf décembre pour éviter la colinéarité) afin de tenir compte de la saisonnalité. Là encore, une sélection de variable par méthode stepwise est appliquée. Les délais moyens sont ensuite prédits sur la période récente à partir de ces modèles.

Modélisation du délai moyen d'ouverture de chantier - Logements individuels purs

Variables explicatives du Délai_moyen$_t$:

Variable	Source	Période	Valeur coefficient	Erreur type
Indicateur de retournement conjoncturel	Insee	t-1	-0.03530	0.01360
Carnet de commande Industrie du bâtiment	Insee	t-1	0.14140	0.02807
Perspectives mises en chantier	Insee	t-3	0.00486	0.00225
Délai d'écoulement des logements individuels	ECLN	t-3	0.07445	0.01389
Part d'entreprises désirant mettre à l'étude de nouveaux programmes	Insee	t-3	0.00886	0.00282
Création d'entreprise dans la construction	Insee	t-1	-0.00004404	0.00000785
Variable indicatrice si t est un mois de janvier, …, novembre			Mois 02 : -0.15088 Mois 03 : -0.14719 Mois 04 : -0.13456 Mois 05 : -0.08611 Mois 07 : 0.09402 Mois 10 : 0.14211 Mois 11 : 0.15122	Mois 02 : 0.03153 Mois 03 : 0.03137 Mois 04 : 0.03028 Mois 05 : 0.03037 Mois 07 : 0.03012 Mois 10 : 0.03012 Mois 11 : 0.03018
Constante			2.56864	0.21915

R^2 ajusté = 0,77

Modélisation du délai moyen d'ouverture de chantier - Logements individuels groupés

Variables explicatives du Délai_moyen$_t$:

Variable	Source	Période	Valeur coefficient	Erreur type
Stock de logements invendus	ECLN	t-3	0.02206	0.00542
Taux OAT 10 ans	BDF	t	0.70081	0.12287
Création d'entreprise dans les activités immobilières	Insee	t-1	0.00161	0.00028137
Délai d'écoulement des logements individuels	ECLN	t-3	0.12869	0.05078
Variable indicatrice si t est un mois de janvier, …, novembre			Mois 07 : -0.54140	Mois 07 : 0.18906
Constante			3.70263	0.86486

R^2 ajusté = 0,55

Modélisation du délai moyen d'ouverture de chantier - Logements collectifs et résidences

Variables explicatives du Délai_moyen$_t$:

Variable	Source	Période	Valeur coefficient	Erreur type
Carnet de commande Industrie du bâtiment	Insee	t-1	0.49172	0.17951
Stock de logements invendus	ECLN	t-3	0.02077	0.00467
Indicateur de retournement conjoncturel	Insee	t-1	-0.38160	0.08776
Taux OAT 10 ans	BDF	t	0.95023	0.13531
Création d'entreprise dans les activités immobilières	Insee	t-1	0.00295	0.00051330
Délai d'écoulement des logements collectifs	ECLN	t-3	0.15698	0.05904
Variable indicatrice si t est un mois de janvier, …, novembre			Mois 04 : -0.46445 Mois 06 : -0.48081	Mois 04 : 0.19892 Mois 06 : 0.19896

Modèle sans constante.

Annexe 6 : fonction de déformation de la distribution de délais de référence

La fonction F vise à déformer la distribution de délais de référence pour se caler sur un délai moyen donné.

F (distribution_référence, délai_moyen_imposé) = distribution_déformée

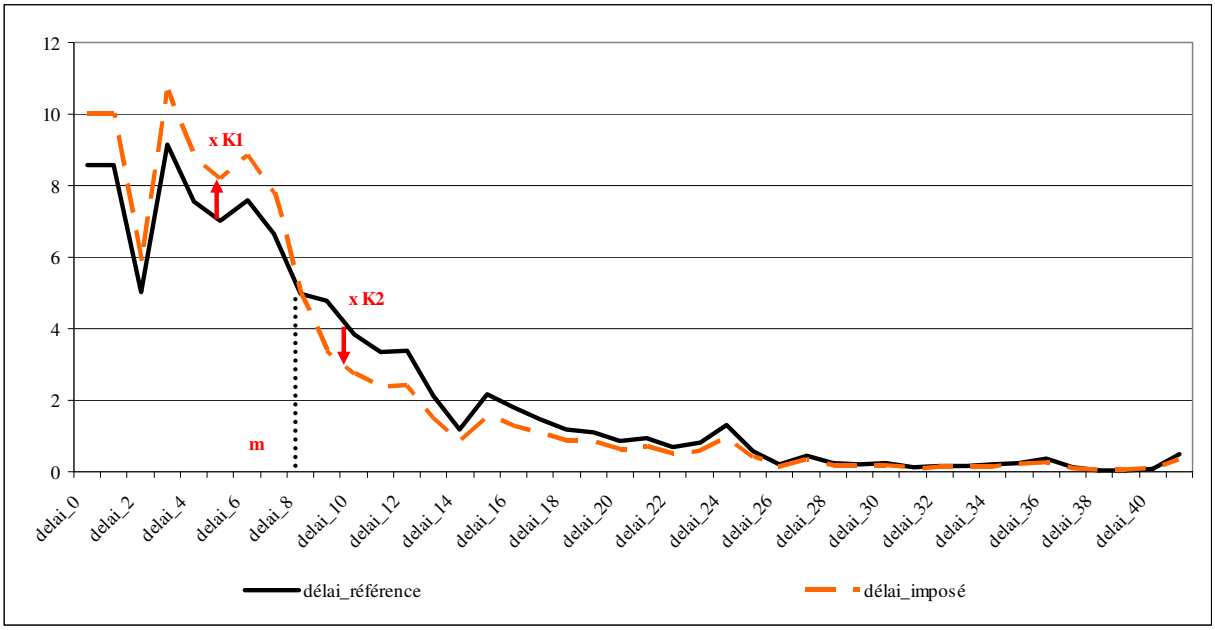
Les paramètres :

1) La distribution de délai d'ouverture de chantier « de référence » :
- délais i de 0 à 36 mois (variable discrète)
- fréquence associée au délai i : f_i ⇨ $\sum_i f_i = 1$
- délai moyen de la distribution de référence : *moy_ref* $moy\_ref = \sum_i \; i \, x \, f_i$

2) Délai moyen de la distribution « déformée » : il sera noté *moy_imposé*

Méthode :
La distribution initiale va être déformée de manière homothétique à partir d'un point d'inflexion noté m (à déterminer) afin de se caler sur le délai moyen imposé.

La fréquence associée à la distribution déformée sera notée h_i
$\forall$ i < m, $h_i = (1 + k_1) f_i = K_1 f_i$
Si i = m, $h_m = f_m$
$\forall$ i > m, $h_i = (1 + k_2) f_i = K_2 f_i$



Notations : Soit $a = \sum_{i<m} f_i$ $b = \sum_{i<m} \; i \, x \, f_i$
$c = \sum_{i>m} f_i$ $d = \sum_{i>m} \; i \, x \, f_i$

Propriétés de la distribution de référence :
$a + f_m + c = 1$ (1)
$b + m f_m + d = moy\_ref$ *(2)*

Propriétés de la distribution déformée :

$\sum_i h_i = 1 \Leftrightarrow \quad (1+k_1)\, a + f_m + (1+k_2)\, c = 1 \qquad (3)$

$\sum_i i \times h_i = moy\_imposé \Leftrightarrow \quad (1+k_1)\, b + m f_m + (1+k_2)\, d = moy\_imposé \qquad (4)$

$$\Rightarrow k_2 = \frac{\text{moy\_imposé - moy\_ref}}{\text{d - (bc/a)}}$$

$\Rightarrow k_1 = - k_2\, c / a$

On choisit par défaut *m* tel que *m* = Partie_entière[(*moy_imposé + moy_ref) / 2] + 1*

Si k1 < -1 alors on décale le point d'inflexion vers la droite (m = m+1). On recommence tant que k1<-1.
Si k2 < -1 alors on décale le point d'inflexion vers la gauche (m = m-1). On recommence tant que k2<-1.