

LES DÉLAIS DE LA CONSTRUCTION : ASPECTS EXPLORATOIRES

Magda Tomasini

L'évaluation de la production de logements pour l'élaboration des comptes annuels et trimestriels se heurte à un problème de date de disponibilité des données. En effet, les mises en chantier et les achèvements de travaux sont pris en compte avec retard par le système informatique car le maître d'ouvrage ne transmet pas nécessairement l'information immédiatement aux services administratifs concernés. C'est pourquoi, la production réelle n'est connue qu'avec un délai de 2 ans, lors des enquêtes annuelles d'entreprises dans le BTP qui permettent de calculer la production du bâtiment dans les comptes annuels définitifs.

Pour l'évaluation des comptes trimestriels et des comptes annuels semi-définitifs, les comptables nationaux tentent d'estimer la production de logements pendant une année ou un trimestre donné notamment à partir des déclarations d'ouverture de chantier [1]¹. Ainsi, un logement mis en chantier en octobre 1994 qui n'est la plupart du temps pas achevé la même année contribue cependant à la production de logements de 1994. Cette contribution est appelée "nombre d'équivalents-logements" puisqu'elle représente en quelque sorte un équivalent physique des travaux réalisés. Ce passage des ouvertures à la production "réelle" estimée est résumé par des grilles appelées communément "grilles délais", grilles fortement dépendantes des délais de la construction.

Le suivi statistique de la construction d'un logement comporte 3 étapes : l'autorisation de construire (délivrée par les mairies ou les directions départementales de l'Équipement), l'ouverture du chantier et l'achèvement des travaux. Ainsi, deux types de délais entrent en ligne de compte [2] : le délai qui s'écoule entre l'autorisation et l'ouverture du chantier d'une part et entre l'ouverture du chantier et l'achèvement des travaux d'autre part. Ils n'ont pas nécessairement la même nature car leurs déterminants ne sont pas toujours du même ordre. Par exemple, les délais entre ouverture et achèvement des travaux sont fortement influencés par la taille des chantiers alors que ceux qui s'écoulent entre l'autorisation et l'ouverture du chantier peuvent dépendre de la conjoncture du secteur du BTP et plus particulièrement de la demande de logements neufs perçue par les promoteurs.

1. Les nombres entre crochet renvoient à la bibliographie figurant à la fin du document

Par ailleurs, la taille du fichier SICLONE qui compte plus de 3 millions d'évènements ne permet pas une approche quantitative immédiate dans l'étude de ces délais, qui consisterait à tester l'adéquation de la distribution des délais à telle ou telle loi, et ce d'autant plus qu'on ne sait pas a priori quelles lois tester. Ainsi, en l'absence d'idée préconçue et face à un fichier assez "lourd" à manipuler, l'analyse exploratoire paraît mieux appropriée. L'objet de cette étude est de présenter la démarche adoptée pour appréhender les délais de la construction dans ce contexte de fichier important et d'absence d'information préalable.

Seule l'étude des délais entre l'ouverture et l'achèvement des travaux (ou délais de la construction) est présentée ici. Etant donnée la taille importante du fichier, des méthodes de "validation croisée" ont été largement utilisées. Elles consistent à explorer des échantillons de taille réduite et à valider ensuite les hypothèses retenues soit sur des échantillons de taille importante, soit sur le fichier exhaustif [3].

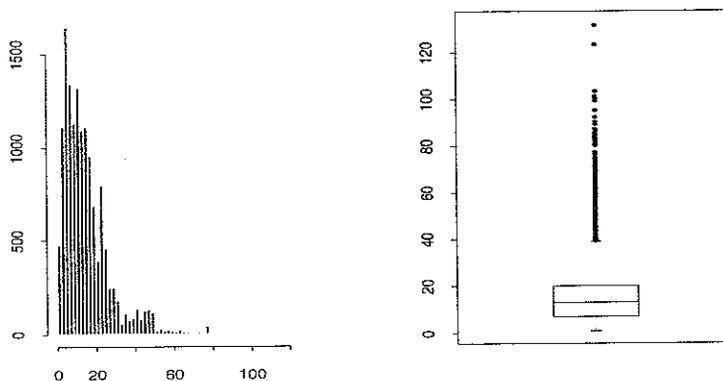
À travers quelques exemples, on montrera comment les techniques exploratoires, illustrées par des outils simples et robustes, permettent de vérifier ou d'échafauder des hypothèses sur la structure des données.

À quelle loi peut on ajuster les délais de la construction ?

La représentation graphique de la distribution des délais est asymétrique avec un mode à gauche (figure 1). Parmi les lois connues, il semble que ce soit la loi log-normale qui présente les caractéristiques les plus proches de cette distribution empirique. Cette hypothèse est confirmée par le tracé du *qq-plot*¹ du logarithme de la variable par rapport à une loi normale (figure 2). En effet, la courbe quantile – quantile s'ajuste correctement à une droite.

Figure 1

Histogramme et Box Plot² des durées de la construction (en mois) sur la période 1980-1991



1 Voir encadré n° 1

2 Voir encadré n° 2

Encadré 1 : Le QQ Plot

Le QQ Plot [3] (graphique quantile-quantile) est une méthode graphique qui permet de comparer deux distributions en représentant les valeurs des quantiles de la première série sur l'axe des abscisses et les valeurs des quantiles correspondants de la deuxième série sur l'axe des ordonnées. On peut ainsi comparer deux séries empiriques ou comparer une distribution empirique à une distribution théorique. Cette méthode permet donc de vérifier de visu la présomption d'adéquation d'une distribution à une loi connue.

Si les deux distributions sont identiques, les points s'alignent sur la première bissectrice. Si les points s'alignent sur une droite parallèle à la première bissectrice, une constante additive permet de passer d'une distribution à l'autre : la distribution empirique est compatible avec la distribution théorique mais elles ont une position différente (estimée par la médiane ou la moyenne). Si les points s'alignent sur une droite quelconque passant par l'origine, une constante multiplicative permet de passer d'une distribution à l'autre : les distributions sont compatibles mais diffèrent par leur étalement (estimé par l'écart – type ou l'intervalle interquantile).

Le critère fondamental pour juger l'adéquation distributionnelle de deux séries est l'existence d'une droite, les translations ou inclinaisons par rapport à la droite ou leur combinaison n'indiquent que les variations de paramètres de tendance centrale et de dispersion. Cette invariance permet de comparer un jeu de données à certaines familles de distributions en un seul graphique. On peut par exemple comparer une distribution empirique à toute loi normale quels que soient sa moyenne et son écart-type.

Les déviations par rapport à la première bissectrice (points écartés, courbure aux extrémités, courbure centrale, segments horizontaux) permettent de suggérer la nature de la différence entre deux distributions.

QQ Plot ou test d'adéquation ?

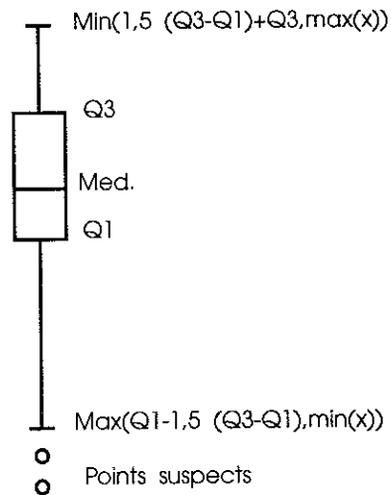
Un test d'adéquation de la distribution à la loi log-normale aurait certainement refusé l'hypothèse d'adéquation. En effet, la distribution empirique comporte un pic à 12 mois qui ne reflète pas la réalité mais plutôt un phénomène "administratif" : l'administration peut en effet engager des contrôles de chantier à 12 mois pour constater si le chantier est terminé ou non. De plus, dès que la taille d'un échantillon est trop importante, les tests statistiques sont trop sévères. Ainsi, le test de Shapiro – Wilk mis en œuvre avec la procédure UNIVARIATE de SAS sur un tirage aléatoire simple au 100^e a refusé l'hypothèse d'adéquation des délais à la loi log-normale.

Encadré 2 - Box Plot (boîte à pattes)

Le Box plot [3] est un graphique représentant la distribution d'une variable par ses quantiles.

Le corps de la boîte est formé par le premier et troisième quartile et coupé par le deuxième quartile (médiane), les moustaches figurent les valeurs en-deçà de une fois et demi l'intervalle inter-quartile. Les valeurs suspectes potentiellement "aberrantes" sont situées au-delà des moustaches.

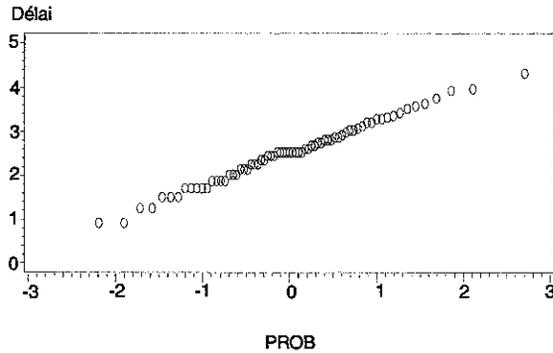
Ainsi la taille de la boîte symbolise l'étendue inter-quartile, la position de la médiane est un bon indicateur de la symétrie de la distribution, la taille des moustaches traduit la dispersion, les valeurs suspectes sont immédiatement repérées.



Le diagramme permet aussi de comparer facilement plusieurs séries de données, en termes de médiane, quartile et valeurs extrêmes. Toutefois ce "résumé numérique" rend compte de la dispersion et de la tendance centrale mais peu de la forme de la distribution, et doit être complété par d'autres méthodes (QQ Plot, histogramme, stem and leaf,...).

Figure 2

QQ Plot du logarithme de la variable délai / loi normale sur la période 1980-1991



La distribution est-elle stable dans le temps ?

La série de *box-plots* des logarithmes de la durée des travaux par semestre (figure 4) suggère une relative stabilité dans le temps. Ces *box-plots* sont symétriques pour la plupart des semestres. Néanmoins, cette stabilité de la loi ne signifie pas que les paramètres soient conservés d'une année sur l'autre. Ainsi, les lois annuelles, semestrielles et trimestrielles peuvent aussi être assimilées à des lois log-normales dont les paramètres moyenne et écart-type varient selon l'année, le semestre ou le trimestre. Par exemple, la figure 3 représentant les *QQ-Plots* du logarithme des délais par rapport à la loi normale en 1980 et 1990 montre un bon alignement pour chacune des deux années. En revanche, il apparaît que les 2 droites n'ont pas les mêmes équations : les paramètres des 2 lois sont différents.

Figure 3

QQ Plot du logarithme de la variable délai / loi log-normale pour les années 1980 et 1990

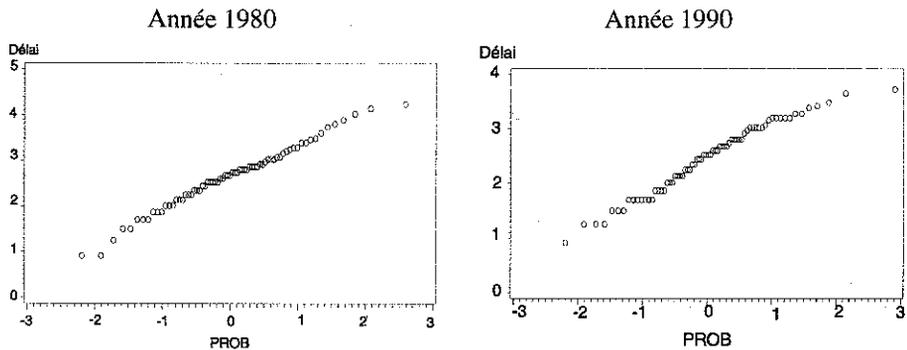
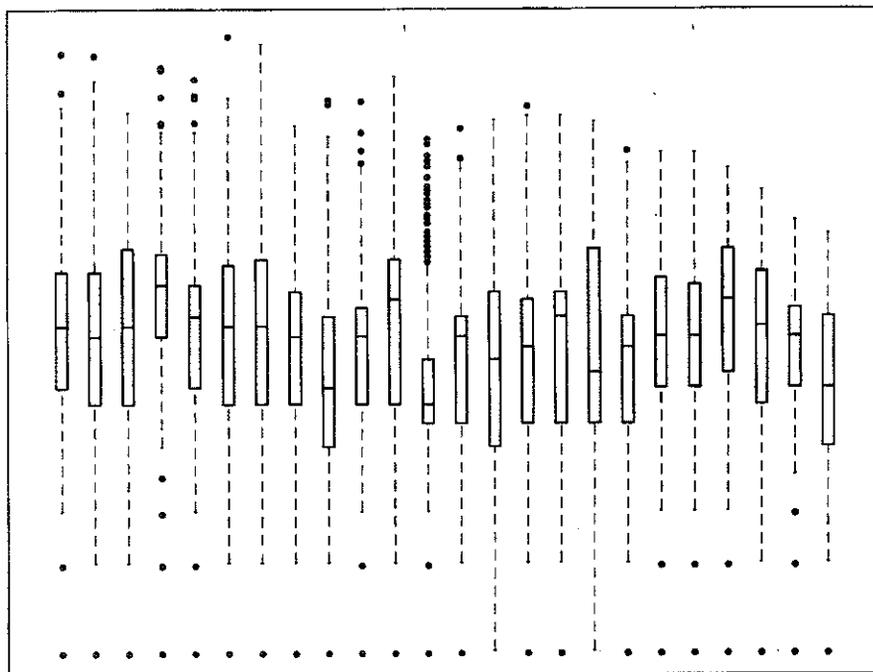


Figure 4

Box plot du logarithme des durées de la construction par semestre sur la période 1980-1991



Recherche des déterminants de la construction

Ici, il existe un certain nombre d'*a priori* sur les déterminants. En effet, il semble évident que la durée de construction dépende de la taille du permis : une maison individuelle se construit plus vite qu'un immeuble. D'autre part, le temps devrait aussi jouer un rôle dans la détermination des délais traduisant ainsi les gains de productivité enregistrés dans la construction neuve. Mais on pourrait aussi se poser la question de savoir si l'une de ces variables est plus influente que l'autre et s'il existe un effet-croisé entre le type de construction et le temps dans la détermination des délais, c'est-à-dire si chaque type de construction a bénéficié ou non du raccourcissement des délais de la construction de la même façon. L'analyse exploratoire offre un outil permettant de répondre à cette question : le *Median-polish*.

Encadré 3 - Median Polish ("polissage" d'un tableau croisé)

Le Median polish [4] est une méthode exploratoire robuste et simple qui permet d'analyser un tableau croisé en dissociant l'effet commun de deux variables qualitatives de l'effet individuel de chacune sur une variable de réponse X. On postule que les valeurs du tableau se décomposent suivant un modèle additif de la forme $X = \text{constante} + \text{effet ligne} + \text{effet colonne} + \text{résidu aléatoire}$. L'analyse du tableau des résidus permet de juger de la qualité de l'ajustement des données au modèle et de déceler d'éventuels effets croisés des variables lignes et colonnes.

En pratique, à chaque valeur du tableau on soustrait la médiane de la ligne, puis la médiane de la colonne. On réitère le processus sur le tableau résiduel, jusqu'à obtenir des médianes nulles. La même opération effectuée sur les marges permet alors d'estimer les différents effets.

L'examen des résidus s'effectue par exemple à l'aide de graphiques stem and leaf et box plot. La présence d'effets croisés peut être détectée sur le graphique des résidus observés et des résidus théoriques (calculés sous l'hypothèse d'absence d'effet croisé). La détection d'une structure linéaire dans ce nuage permet alors d'exhiber et de chiffrer ces effets.

Le median polish est une technique robuste, permettant des diagnostics puissants, mais qui ne permet pas d'évaluer la précision des estimations trouvées.

Median Polish ou Analyse de la Variance ?

Le median polish présente plusieurs avantages par rapport à l'analyse de la variance : cette technique est robuste (car elle est basée sur une analyse par la médiane) et elle permet des diagnostics par l'analyse des effets croisés. Cette analyse robuste est ici intéressante parce qu'elle est peu sensible à la présence de valeurs extrêmes, ce qui est fondamental puisque certaines cases du tableau sont peu significatives voire vides lorsque les variables sont très corrélées. L'analyse par median polish est également peu sensible à la non-normalité de la loi à l'intérieur de chaque sous-population (i.e chaque case du tableau).

Median Polish des délais moyens par année et par type de logement

Tableau B1

Délais moyens par année et par type de logement

Années	Individuel pur		Individuel groupé	Collectif		Ensemble
	non mixte	mixte	non mixte	non mixte	mixte	
1980	15,02	19,17	16,11	21,57	22,94	17,46
1981	13,93	25,57	16,36	19,18	21,54	16,49
1982	13,09	17,95	21,09	15,29	22,35	15,74
1983	12,58	19,14	17,71	18,88	20,30	15,41
1984	11,66	16,67	12,78	18,11	16,21	13,27
1985	10,64	21,42	15,16	17,69	18,71	14,13
1986	9,99	12,88	13,03	17,53	21,56	12,90
1987	9,30	19,29	13,85	17,09	18,57	12,93
1988	9,31	15,14	16,88	20,60	16,09	14,46
1989	9,53	11,33	11,25	19,18	18,74	14,43
1990	9,32	10,00	12,87	16,53	22,61	13,51
1991	9,03	11,69	7,94	14,45	14,22	11,86
Ensemble	11,42	17,66	15,45	17,92	19,63	16,37

Figure B1

Graphique STEM 1 LEAF des résidus

STEM	LEAVES	EFFECTIF	
LOW	-4,83 -4,39	2	la plus petite valeur représentée est : - 3.80
-3	1	1	la plus grande valeur représentée est : 3.80
-2	6	1	la médiane de la distribution est : 0
-1	*9965533	8	
0	*775433320000000000000000023556667	33	
1	2457	4	
2	126	3	
3	3668	4	
HIGH	4,97 5,34 5,59 6,91	4	
TOTAL	60		

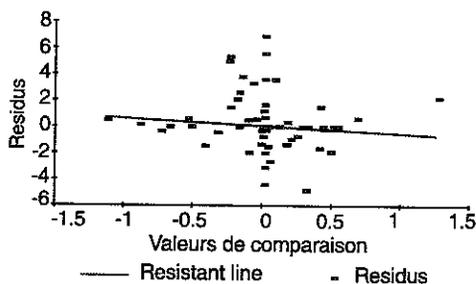
Tableau B2

Effets du type de logement (A) et de l'année (B), et résidus

Résidus	Individuel pur		Individuel groupé	Collectif		Effet année (A)
	non mixte	mixte	non mixte	non mixte	mixte	
1980	0,5	- 0,3	- 1,5	0,0	0,6	3,7
1981	0,2	6,9	- 0,5	- 1,6	0,0	2,9
1982	0,0	- 0,0	5,0	- 4,8	1,5	2,2
1983	0,0	1,7	2,1	- 0,7	- 0,0	1,7
1984	0,6	0,7	- 1,3	0,0	- 2,6	0,2
1985	- 0,3	5,6	1,2	- 0,3	0,0	0,1
1986	- 0,0	- 2,0	- 0,0	0,5	3,8	- 0,9
1987	- 1,5	3,6	0,0	- 0,7	0,0	- 0,1
1988	- 1,0	0,0	3,6	3,3	- 1,9	- 0,6
1989	0,0	- 3,1	- 1,3	2,6	1,4	- 1,3
1990	- 0,2	- 4,4	0,3	0,0	5,3	- 1,4
1991	2,2	0,0	- 1,9	0,6	- 0,4	- 4,1
Effet type (B)	- 4,9	0,0	- 1,8	2,1	2,9	15,8

Figure B2

Diagnostic-plot



ÉQUATION DE LA RESISTANT LINE

Résidu = 0,06 - 0,57 * valeur de comparaison

Les résidus sont faibles et bien centrés, mais sans nette structure, il n'y a donc pas d'effet croisé

Les résidus élevés correspondent à des points atypiques.

Les premiers commentaires que l'on peut faire sur une telle analyse sont les suivants :

- les délais sont plus courts dans l'individuel (pur et groupé) que dans le collectif et la construction simultanée de locaux (permis mixte) allonge la durée de construction du permis (effet B) ;

- les délais de la construction se sont raccourcis entre 1981 et 1991 (effet A), traduisant d'une part la réduction de la taille des permis et l'existence de gains de productivité réalisés dans la construction de logements neufs d'autre part ;
- l'influence du temps et du type de construction sur les délais de la construction sont de même ampleur ;
- le coefficient de la valeur de comparaison dans l'équation de la resistant line est faible : il n'y a pas d'effet croisé. Les différents types de construction ont bénéficié du raccourcissement des délais sensiblement de la même façon.

Conclusion

L'analyse exploratoire permet ainsi de définir les axes de futures investigations. Le fait que la loi des délais puisse être approximée par une loi log-normale autorise l'utilisation de nombreuses méthodes confirmatoires après transformation des données. Il sera par exemple possible d'effectuer des analyses de la variance sans s'écarter notablement des conditions de normalité que cette méthode impose.

Cette possible approximation des délais par une loi log-normale permet aussi l'utilisation des modèles de durée de vie dans l'estimation des délais de la construction. On pourra en effet réaliser des estimations paramétriques par maximisation de la vraisemblance. De plus, la mise en évidence du rôle important de certaines variables souligne l'existence de variables exogènes pour l'estimation des délais à partir des modèles de durée de vie.

Ces distributions établies, il restera la question difficile du passage aux grilles – délais qui permettent le calcul d'équivalent – logements pour estimer la production.

BIBLIOGRAPHIE

[1] Claudie LOUVOT, "La durée des chantiers de construction et son incidence sur l'estimation de la production des logements", *Note interne* n° 021/E331 du 22 septembre 1994.

[2] Hervé BACHERE, "Les délais de construction des logements neufs", *Notes vertes économiques* n° 312, juin 1993.

[3] John CHAMBERS, William S. CLEVELAND, Beat KLEINER and Paul A. TUKEY, "Graphical methods for Data Analysis", Wadsworth & Brooks/Cole Publishing Company, 1983.

[4] Michel GRUN-REHOMME et Dominique LADIRAY, "Une méthode de statistiques douces : le polissage d'un tableau croisé", Actes du Club SAS 1994.