

Prévoir l'investissement des entreprises

Un indicateur des révisions dans l'enquête Investissement

Nicolas Ferrari*

L'enquête trimestrielle sur les investissements dans l'industrie de l'Insee est une source d'information très importante sur les évolutions conjoncturelles de l'investissement productif. Toutefois, la nature annuelle des questions posées rend difficile son utilisation pour des prévisions selon un rythme trimestriel. Pour cela, cet article propose un indicateur trimestriel des révisions d'anticipations d'investissement des industriels. Cet indicateur mesure les adaptations des projets d'investissement au cours de l'année, en fonction des évolutions conjoncturelles. Bien corrélé aux évolutions trimestrielles de l'investissement des entreprises, il est disponible environ trois mois avant la publication des premiers résultats des comptes nationaux trimestriels.

Les distributions étudiées ne vérifiant pas certaines caractéristiques de la loi normale (les valeurs extrêmes y sont plus fréquentes, et on observe de fortes concentrations autour de la valeur zéro), il est nécessaire de mettre en œuvre une méthode d'estimation robuste aux révisions à valeurs extrêmes. En prenant également en compte la présence d'hétéroscédasticité, il a été choisi d'utiliser la méthode dite des « M-estimateurs Quasi Généralisés ».

* Nicolas Ferrari travaillait au Département de la conjoncture de l'Insee lors de la rédaction de cet article. Il remercie Olivier Biau, Michel Devilliers, Eric Dubois, Hélène Erkel-Rousse, Philippe Scherrer et Fabien Toutlemonde pour leurs nombreux conseils. Courriel : nicolas.ferrari@m4x.org

Les indicateurs conjoncturels relatifs à l'investissement productif sont rares et leur qualité prédictive est souvent faible (cf. Ferrari, 2005a, et aussi pour un panorama global de l'analyse conjoncturelle de l'investissement des entreprises, l'ouvrage de Fayolle, 1987). L'enquête sur les investissements dans l'industrie (nommé ici enquête *Investissement*), menée trimestriellement (depuis 2003) par l'Insee, est une des rares sources conjoncturelles relatives aux dépenses d'équipement des entreprises. Elle permet de prévoir l'investissement industriel annuel de manière relativement fiable et avancée dans le temps. L'investissement industriel est également bien corrélé avec celui des autres secteurs d'activité. L'enquête *Investissement* offre donc une information intéressante concernant l'investissement productif de l'ensemble des secteurs d'activité.

Toutefois, cette enquête se prête à première vue mal à une utilisation conjoncturelle. En effet, les entreprises sont interrogées sur leurs investissements de l'ensemble de l'année, si bien qu'on ne sait pas *a priori* dire à quelle période de l'année ont été ou seront effectués ces investissements. Une deuxième analyse conduit toutefois à réviser ce premier jugement négatif. Intuitivement, lorsqu'une entreprise révisé en cours d'année à la baisse sa prévision d'investissement, ce peut être parce qu'elle est confrontée à une dégradation imprévue de son environnement économique. Dans ces conditions, on s'attend à ce qu'elle ajuste ses plans d'investissement pour les trimestres qui suivent : une révision à la baisse de la croissance prévue des investissements devrait donc être associée à un ralentissement des investissements sur le reste de l'année, et inversement une révision à la hausse de la croissance prévue des investissements serait liée à une accélération sur le reste de l'année.

Cet article s'attache à extraire l'information contenue dans ces révisions, sans s'intéresser précisément aux canaux de transmission qui lient les évolutions de l'environnement de l'entreprise aux révisions d'investissement. Un examen plus attentif de ces éléments nécessiterait un modèle structurel exprimant des arbitrages inter-temporels des entreprises. Cette approche, plus « économique », au sens où elle identifierait les déterminants du comportement des entreprises pour ce qui concerne leurs investissements, n'entre pas dans le champ du travail présenté ici. En effet, dans cet article, l'accent est mis sur la méthode retenue pour extraire une informa-

tion conjoncturelle de questions qui s'y présentent mal à première vue.

Il est possible de construire un indicateur de révision de l'investissement simplement en calculant la différence entre le taux de croissance de l'investissement prévu à la dernière enquête et celui prévu à l'enquête précédente. Un tel indicateur se révèle toutefois décevant car peu corrélé aux évolutions trimestrielles de l'investissement. Une première explication à cette déception vient du fait que les révisions des prévisions de taux de croissance proviennent pour partie d'une modification de l'échantillon des entreprises répondantes. Il est ainsi nécessaire de ne retenir que les seules entreprises ayant répondu aux deux enquêtes. En outre, parmi les entreprises ayant répondu aux deux enquêtes, les révisions de taux de croissance découlent dans certains cas de modifications des évaluations par les entreprises du montant de leur investissement au cours de l'année antérieure : ces révisions n'ont aucune raison de modifier les investissements sur le reste de l'année en cours. Il apparaît alors plus efficace de raisonner sur les révisions apportées aux montants mêmes d'investissements prévus par les entreprises.

Mais cela ne suffit pas. En effet, la révision moyenne calculée sur l'ensemble des entreprises est très affectée par les valeurs extrêmes : une part non négligeable des entreprises interrogées connaît en effet des fluctuations importantes de ses rythmes d'investissement sans relation interprétable avec les modifications de son environnement économique. Ces fluctuations, importantes pour une entreprise donnée, ne sont pas pour l'essentiel liées à un mouvement économique global. Il est donc nécessaire de recourir à des techniques de statistique « robuste » aux révisions extrêmes (au sens précis où l'influence que sont susceptibles d'avoir les révisions « extrêmes » de projets d'investissement est limitée de manière à ce que le retrait d'une entreprise de l'échantillon ne modifie pas beaucoup le résultat global). L'utilisation de statistiques robustes permet de diminuer le poids de ces révisions extrêmes. Cet article propose ainsi un indicateur trimestriel des révisions d'anticipations d'investissement des industriels basé sur la méthode d'estimation robuste des M-estimateurs proposée par Huber (1964).

L'ampleur des révisions étant très corrélée à la taille, nous devons également tenir compte de cet effet « taille ».

L'indicateur des révisions ainsi construit se révélera alors être bien corrélé aux variations trimestrielles de la FBCF des ENF (1) en valeur. Il apparaît ainsi comme un indicateur utile pour prévoir les variations de cette variable.

Dans un premier temps, nous présentons l'enquête *Investissement* et nous proposons un guide de lecture des résultats publiés pour l'analyse conjoncturelle de l'investissement. Ces résultats apportent une information annuelle pertinente ; mais ils sont difficilement utilisables pour prévoir les évolutions de l'investissement à un rythme trimestriel. En revanche, les anticipations étant révisées tous les trimestres, ces révisions fournissent une information intéressante à ce rythme trimestriel. La construction de l'indicateur des révisions est alors exposée dans une seconde partie. Une attention particulière est portée au choix de la fonction de coût des M-estimateurs ainsi qu'à la correction de l'hétéroscédasticité liée à l'effet « taille ». La dernière partie propose deux exemples d'utilisation de la série construite à l'aide de modèles multivariés autorégressifs (VAR) : le premier est relatif à la prévision des évolutions de la FBCF des ENF. Les projets d'investissement des industriels étant liés étroitement à leurs anticipations de demande, le second porte sur les évolutions de l'activité industrielle mesurée par la production manufacturière. Dans les deux cas, l'apport d'information de l'indicateur est vérifié par des tests de causalité au sens de Granger.

Comment lire les résultats de l'enquête *Investissement* ?

L'enquête sur les investissements dans l'industrie est une enquête trimestrielle, réalisée au cours des mois de janvier, avril, juillet (2) et octobre. Ses résultats sont publiés une quinzaine de jours après la fin du mois considéré. L'enquête porte sur un échantillon d'environ 4 000 entreprises représentatives de l'ensemble de l'industrie en dehors de la production et de la distribution d'eau, d'électricité et de gaz. Deux types de questions permettent d'évaluer les perspectives d'évolution de l'investissement pour les industriels interrogés :

- d'une part, les entrepreneurs indiquent les montants annuels d'investissement réalisés ou prévus pour trois années civiles consécutives ;

- d'autre part, ils émettent une opinion sur les évolutions passées et prévues de leurs dépenses semestrielles d'investissement. Ces opinions sont formulées par un choix entre les modalités « en hausse », « stable » et « en baisse ». Elles sont agrégées et publiées sous formes de soldes d'opinion (différences des proportions pondérées de réponses « en hausse » et de réponses « en baisse »).

Ces deux types de questions sont présents à chaque enquête. De plus, d'autres questions sont ajoutées au questionnaire selon le trimestre de l'enquête afin d'affiner la perception de l'investissement. Le présent article se rapporte uniquement aux questions quantitatives annuelles. Pour disposer de plus de renseignements sur l'enquête *Investissement*, le lecteur pourra se reporter à la fiche méthodologique de l'enquête (3) et à l'ouvrage de la série *Insee Méthodes* relatif à cette enquête (Rosenwald, 1994).

Les révisions des entrepreneurs interrogés à l'enquête *Investissement* apportent une information conjoncturelle

Les questions relatives aux montants annuels d'investissement permettent de juger des prévisions de dépenses d'équipement des industriels. Pour une année donnée, l'enquête fournit, campagne après campagne, plusieurs estimations des évolutions – prévues puis réalisées – de l'investissement industriel (cf. graphique I). La méthode utilisée pour l'agrégation est une méthode particulière, dite « méthode des Grands Investisseurs », exposée et discutée dans Ravalet (1996). Il s'agit de la méthode dite « du ratio » (4), légèrement modifiée : certaines entreprises sont sous-pondérées lorsque leurs réponses sont atypiques ou trop influentes.

Les révisions entre les estimations successives du taux de croissance des investissements industriels pour une même année sont impor-

1. Formation Brute de Capital Fixe des Entreprises Non Financières.

2. Depuis juillet 2003 uniquement.

3. « Fiche méthodologique : Enquête sur les investissements dans l'industrie ». Disponible sur le site www.insee.fr, rubrique « Conjoncture », puis « Indicateurs de conjoncture », « Enquête sur les investissements dans l'industrie » et enfin « Ouvrir la note méthodologique ».

4. La méthode du ratio est la méthode simple qui consiste à estimer le taux de croissance d'une grandeur pour la population par le taux de croissance de cette même grandeur sur l'échantillon observé. Par exemple, le taux de croissance de l'investissement industriel serait estimé par celui de la somme des investissements des entreprises de l'échantillon.

tantes (cf. graphique I). Au cours des différentes occurrences de l'enquête, les résultats convergent vers la dernière estimation donnée par l'enquête d'avril de l'année suivante (5). Cette dernière estimation, dite « définitive », s'avère proche des évolutions réelles des investissements industriels, mesurés en particulier par l'Enquête Annuelle d'Entreprise (EAE) dans l'industrie (cf. graphique II). Cette convergence est un critère important pour juger de la qualité de l'enquête : les évolutions des estimations successives ne sont pas le fait d'un aléa statistique, mais bien de modifications au cours du temps des projets d'investissement des industriels.

Ainsi, pour l'année N , en avril de l'année $N + 1$, c'est-à-dire environ trois mois avant les résultats de l'EAE dans l'industrie et un an avant la publication des comptes nationaux annuels dits « semi-définitifs », l'enquête permet de juger des évolutions passées de la FBCF de l'industrie manufacturière. Toutefois, bien que cette information soit précoce par rapport aux autres sources, elle reste bien trop tardive dans la perspective de prévisions conjoncturelles. En revanche, les estimations précédentes de l'enquête Investissement (à partir d'octobre $N - 1$) sont publiées assez tôt pour les exercices de prévision de l'année N .

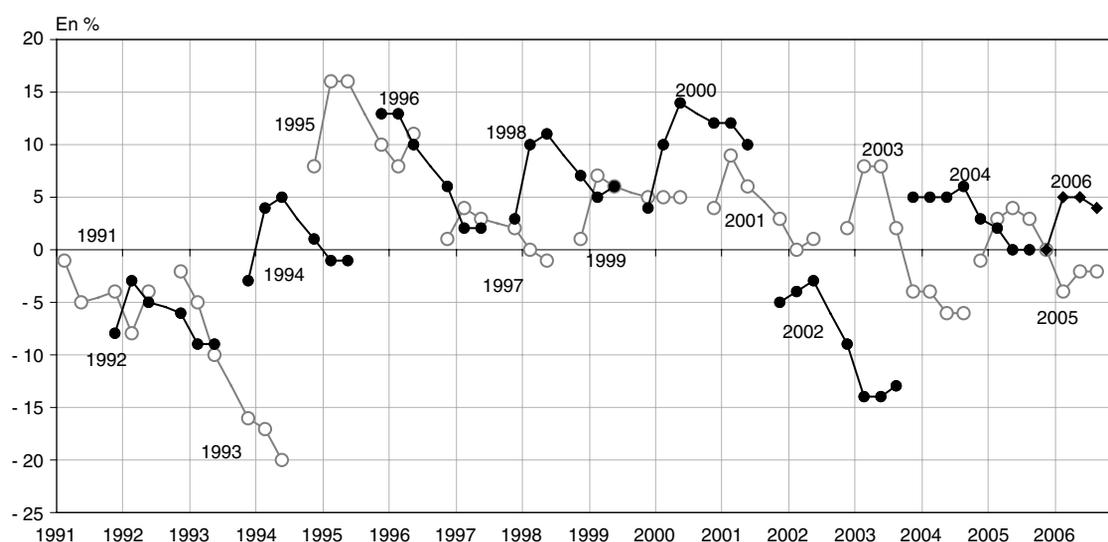
sions agrégées des industriels correspondent aux évolutions qui seront effectivement réalisées. Il apparaît toutefois qu'en moyenne les taux de croissance, prévus puis réalisés, sont ajustés selon un profil relativement stable (cf. graphique I). Il en résulte ainsi une révision systématique pour chaque estimation en fonction de la date de l'enquête par rapport à l'année considérée. Sur les années 1990-2004, les premières estimations des industriels se révèlent être surestimées par rapport à celle d'avril de l'année suivante. En moyenne sur cette période, cette surestimation est de 1,7 point pour la première estimation lors de l'enquête d'octobre de l'année précédente, de 6,1 points, 5,7 points et 2,3 points lors des enquêtes de janvier, avril et octobre de l'année considérée. Enfin, l'estimation en janvier de l'année suivante se révèle très proche de celle d'avril de cette même année, légèrement plus basse de - 0,1 point en moyenne (cf. tableau 1). Au regard de ces éléments, il apparaît essentiel de tenir compte de ces révisions systématiques pour analyser de manière rigoureuse les estimations successives du taux d'investissement.

Jusqu'à 1994, cette correction des révisions moyennes était intégrée directement dans les résultats publiés. Cependant, cette correction

La manière la plus naturelle d'utiliser cette information avancée est de supposer que les prévi-

5. Depuis 2003, l'enquête Investissement est trimestrielle avec une occurrence en juillet. La dernière estimation pour l'année N est désormais donnée par l'enquête de juillet $N + 1$ et non plus celle d'avril $N + 1$.

Graphique I
Estimations successives par les industriels du taux de croissance annuel de leurs investissements



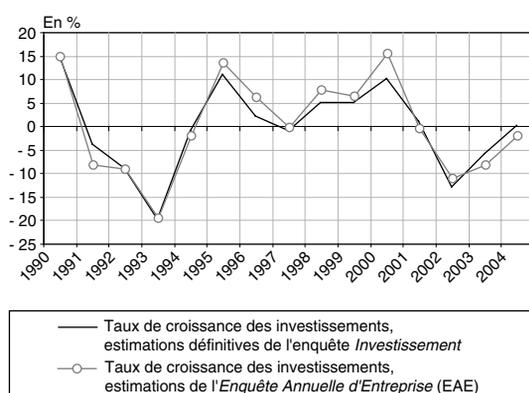
Lecture : chaque courbe représente les estimations successives pour l'année indiquée à côté. L'axe des abscisses représente les dates d'enquêtes. Ainsi, la première estimation relative à l'année 1994 avait lieu en octobre 1993. Les industriels prévoient alors une baisse de - 3 % de leurs investissements. En janvier 1994, toujours pour cette même année, ils prévoient une hausse de 4 %. Jusqu'en 2002, il n'y avait pas d'enquête en juillet.
Champ : industrie manufacturière.
Source : enquête Investissement, Insee.

s'est révélée problématique pour l'année 1993 : la situation économique s'est très fortement dégradée et la première anticipation en octobre 1992 n'a pas été révisée vers le haut mais vers le bas lors des enquêtes suivantes (enquêtes de janvier et d'avril 1993). Alors que la première estimation d'octobre 1992 donnait une information relativement neutre, les enquêtes successives ont convergé en indiquant au final une chute de 20 % de l'investissement industriel en 1993. Il est alors apparu qu'il était insuffisant de corriger simplement d'une révision moyenne : les révisions entre les enquêtes successives dépendent de la position dans le cycle conjoncturel. Par exemple, lors de l'enquête d'octobre de l'année 1999, la prévision de croissance pour l'année 2000 était de 4 %. Alors que l'activité était en haut de cycle, la dernière estimation, lors de l'enquête d'avril 2001, a été de 10 %.

prise en compte du biais moyen aurait conduit à ne prévoir que 5 % ou 6 % de croissance en 2000 lors de l'enquête d'octobre 1999.

En revanche, les révisions entre deux enquêtes successives fournissent une information particulièrement intéressante. Ainsi, pour 1993, la baisse entre les enquêtes d'octobre 1992 et de janvier 1993 du taux de croissance prévu des investissements indiquait un changement important des perspectives d'investissement des industriels (cf. graphique I). Il s'agit donc de considérer en même temps le niveau des anticipations des industriels (relatives à la croissance de leurs dépenses d'équipement) et les révisions de ces anticipations au cours des estimations successives. Par exemple, la révision à la baisse de seulement 2 points entre les enquêtes d'avril et d'octobre 2004 apparaît comme faible par rapport à la moyenne (3,5 points à la baisse). Il s'agit donc d'une information positive, concordante avec le dynamisme de la FBCF des entreprises au quatrième trimestre de 2004. Une telle utilisation des données permet une analyse qualitative très informative mais se prête mal à une utilisation quantitative comme l'autoriserait un indicateur trimestriel.

Graphique II
Comparaison des estimations définitives de l'enquête Investissement et des résultats de l'Enquête Annuelle d'Entreprise (EAE) dans l'industrie



Champ : industrie manufacturière.
 Sources : enquête Investissement, Insee, et Enquête Annuelle d'Entreprise dans l'industrie (Sessi et Insee).

Un indicateur trimestriel des révisions des montants annuels

Si la comparaison des révisions des taux de croissance annuels agrégés de l'investissement se révèle intéressante, ces révisions ne peuvent être correctement rapprochées, d'une manière ou d'une autre, des évolutions trimestrielles de l'investissement. D'une part, même si la méthode d'agrégation

Tableau 1
Les révisions moyennes des estimations des industriels sur la croissance annuelle de leurs investissements

Dates des enquêtes par rapport à l'année N considérée	Moyenne des taux de croissance annuelle de l'investissement estimé par l'enquête (en %)	Différence moyenne par rapport aux estimations de l'enquête d'avril N + 1 (en points)
Octobre N - 1	1,3	1,7
Janvier N	5,7	6,1
Avril N	5,3	5,7
Octobre N	1,9	2,3
Janvier N+1	- 0,5	- 0,1
Avril N+1	- 0,4	0,0

Lecture : de 1990 à 2004, lors des enquêtes de janvier (année N), les industriels ont estimé – en moyenne sur cette période – que leurs investissements augmenteraient de 5,7 % durant l'année courante. Finalement, lors des enquêtes d'avril de l'année suivante (année N + 1), ils estimaient en moyenne que leurs investissements avaient diminué de 0,4 %, soit une révision moyenne de 6,1 points entre l'enquête de janvier courante et l'enquête d'avril de l'année suivante.

Champ : industrie manufacturière, période : 1990-2004.
 Source : enquête Investissement, Insee, calculs de l'auteur.

tion utilisée (dite donc des *Grands Investisseurs*) permet d'estimer avec une certaine efficacité les prévisions de taux de croissance annuels, elle est peu adaptée à la mesure des révisions de taux de croissance. En effet, avec cette méthode, le poids accordé à une même entreprise peut changer entre deux enquêtes successives. D'autre part, certaines révisions des taux de croissance sont davantage dues à des modifications des montants relatifs aux années de référence qu'aux années d'intérêt. Par exemple, lors de l'enquête d'avril 2005, une entreprise révisé à la hausse son estimation de dépenses d'investissement pour 2004, ceci sans modifier sa prévision relative à 2005. Cette entreprise fait décroître le taux de croissance des investissements de 2005. Elle n'a pas pour autant modifié les perspectives de 2005.

À titre illustratif, le graphique III présente un indicateur trimestriel construit à partir des taux de croissance annuels de l'investissement publiés par l'Insee. La corrélation avec les évolutions trimestrielles de la FBCF des ENF (6) est faible : 31 %.

Les révisions des anticipations des industriels se révèlent plus informatives en considérant directement les révisions des montants d'investissement et non pas les révisions des évolutions annuelles. Cette approche avait déjà été retenue par Houdebine (1998). Ici, à chaque enquête et pour chaque entreprise, nous calculons la révision d'investissement rapportée au chiffre d'affaires. L'année considérée est l'année la plus « avancée » possible : lors de l'enquête d'octobre $N - 1$, les entreprises sont interrogées pour

la première fois sur leurs anticipations d'investissement pour l'année N . Aussi, dès l'enquête de janvier N , il est possible de calculer la différence entre le montant déclaré en janvier N et celui déclaré en octobre $N - 1$.

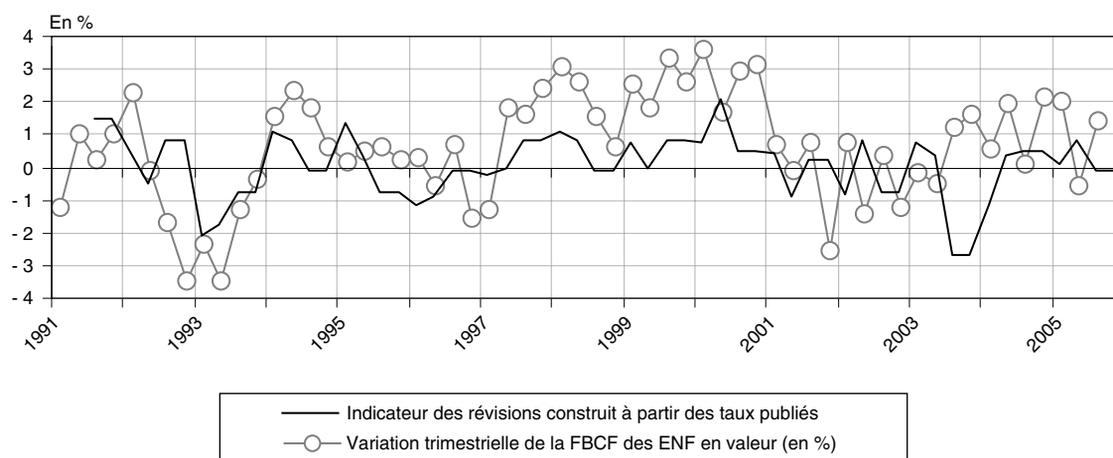
À chacune des enquêtes de l'année N (janvier, avril, juillet et octobre), nous pouvons ainsi calculer, entreprise par entreprise, l'évolution relative des réponses par rapport à l'enquête antérieure (cf. tableau 2).

L'enquête de juillet n'existe que depuis 2003. Jusqu'en 2002, l'indicateur ne peut donc être calculé en juillet. De même, jusqu'en 2002, l'indicateur en octobre ne peut être calculé que comme la révision entre l'enquête d'avril et l'enquête d'octobre. En attendant de disposer de suffisamment d'enquêtes en juillet et par souci de cohérence dans le temps des séries relatives à octobre, nous calculons de la même manière les révisions à partir de l'année 2003. Il n'y a donc pas d'indicateur des révisions pour les enquêtes de juillet et les réponses aux enquêtes d'octobre sont toutes comparées aux réponses données six mois plus tôt en avril. Ceci est une faiblesse – provisoire – de l'indicateur, qui n'est en fait disponible que trois trimestres sur quatre (cf. tableau 3).

Les révisions individuelles sont rapportées à la taille de l'entreprise, mesurée par son chiffre d'affaires. Pour toute enquête t entre l'enquête d'avril

6. Les comptes trimestriels utilisés dans l'ensemble de cet article sont les premiers résultats du troisième trimestre de 2005 en base 2000.

Graphique III
Comparaison de la série construite à partir des taux de croissance annuels publiés avec la série des taux de croissance trimestriels de l'investissement des entreprises



Lecture : les révisions des taux de croissance publiés sont centrées à zéro et réduites à une variance unitaire, ceci pour chaque saison d'enquête (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre).
 Champ : industrie manufacturière (indicateur des révisions) et entreprises non financières (FBCF).
 Sources : enquête Investissement, Insee (indicateur des révisions) et comptes nationaux trimestriels (FBCF). Calculs de l'auteur.

N et celle de janvier $N + 1$, il s'agit du chiffre d'affaires de l'année $N - 1$. Celui-ci est noté $CA_{i,A(t)}$ pour l'entreprise i , avec $A(t)$ représentant l'année $N - 1$ pour t compris entre avril N et janvier $N + 1$. En notant $I_{i,t}^a$ et $I_{i,t-1}^a$ les montants d'investissement pour l'année a déclarés par l'entreprise i lors des enquêtes t et $t - 1$, l'indicateur individuel des révisions est ainsi défini :

$$d_{i,t} = \frac{\text{Investissement}_{i,t}^a - \text{Investissement}_{i,t-1}^a}{\text{Chiffre d'affaires}_{i,A(t)}} = \frac{I_{i,t}^a - I_{i,t-1}^a}{CA_{i,A(t)}} \quad (1)$$

$d_{i,t}$: indicateur individuel (entreprise i) des révisions des projets d'investissement pour l'année a entre les deux enquêtes en t et $t - 1$.

Il est alors possible de construire un indicateur basé sur la moyenne (pondérée par strate et au sein de chaque strate par le chiffre d'affaires (7) des révisions des montants prévus d'investissement. Mais cet indicateur se révèle encore décevant : sa corrélation avec le taux de croissance trimestriel de l'investissement reste encore faible : 42 % (cf. graphique IV).

Le choix de la méthode d'agrégation : les M-estimateurs

Un examen plus attentif de la distribution des révisions permet de comprendre pourquoi la moyenne des révisions ne constitue pas un bon indicateur. En effet, la distribution des révisions

7. Sans pondérer par la taille des entreprises, la corrélation est beaucoup plus faible, de 28 %.

Tableau 2
Calendrier théorique de calcul de l'indicateur

Enquêtes	Montants demandés lors des enquêtes successives	Indicateurs individuels calculés sur la différence...
Janvier N	Année $N-2$... entre l'enquête d'octobre $N-1$
	Année $N-1$... et celle de janvier N
	Année N	... concernant l'année N
Avril N	Année $N-2$... entre l'enquête de janvier N
	Année $N-1$... et celle d'avril N
	Année N	... concernant l'année N
Juillet N	Année $N-2$... entre l'enquête d'avril N
	Année $N-1$... et celle de juillet N
	Année N	... concernant l'année N
Octobre N	Année $N-1$... entre l'enquête de juillet N
	Année N	... et celle d'octobre N
	Année $N+1$... concernant l'année N

Tableau 3
Calendrier provisoire de calcul de l'indicateur

Enquêtes	Montants demandés lors des enquêtes successives	Indicateurs individuels calculés sur la différence...
Janvier N	Année $N-2$... entre l'enquête d'octobre $N-1$
	Année $N-1$... et celle de janvier N
	Année N	... concernant l'année N
Avril N	Année $N-2$...entre l'enquête de janvier N
	Année $N-1$... et celle d'avril N
	Année N	... concernant l'année N
Octobre N	Année $N-1$... entre l'enquête d'avril N
	Année N	... et celle d'octobre N
	Année $N+1$... concernant l'année N

paraît soumise de manière disproportionnée aux valeurs extrêmes. D'une enquête à l'autre certaines entreprises peuvent fortement réviser leurs montants d'investissement. Par exemple, une petite entreprise prévoit d'acheter un bâtiment. Celui-ci peut être d'une valeur équivalente à plusieurs années de chiffre d'affaires. Si ce projet n'aboutit pas, la révision aura alors une amplitude de plusieurs fois le chiffre d'affaires. Ces fortes révisions créent des répartitions à « queues très épaisses » (cf. graphique V).

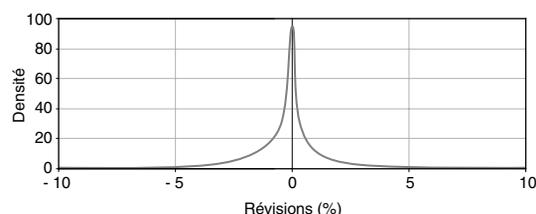
Une telle révision, sans lien avec l'environnement économique, n'est pas représentative d'un mouvement économique global : elle ne donne donc que peu d'information sur l'investissement de l'ensemble des entreprises. Il est par conséquent nécessaire de recourir à des méthodes d'estimation qui accordent un poids soit nul (ce

qui peut être excessif) soit (ce qui est sans doute préférable) moins important que dans la distribution initiale à ces variations extrêmes : le recours à des estimateurs « robustes » s'impose donc : un estimateur est dit « robuste » lorsque le retrait d'une observation, quelle qu'elle soit, modifie peu la valeur de l'estimation (8).

L'estimateur robuste le plus naturel est la médiane. Toutefois, cet estimateur ne peut pas non plus convenir : en plus d'être à queues épaisses, les distributions sont aussi très concentrées autour de zéro (cf. graphique V). En effet, de nombreuses entreprises ne révisent pas leurs anticipations d'investissement entre deux enquêtes (9) : la médiane est presque systématiquement nulle.

La méthode choisie est celle des M-estimateurs, due à Huber (1964). Cette méthode est une généralisation de la méthode des MCO (Moindres Carrés Ordinaires). Pour estimer le « centre » (10) de la distribution, au lieu de minimiser la somme des carrés des résidus comme dans le cas des MCO, nous minimisons la somme d'une autre fonction de coût appliquée aux résidus. Cette fonction est notée ρ .

Graphique V
Répartition empirique des révisions individuelles d'investissement



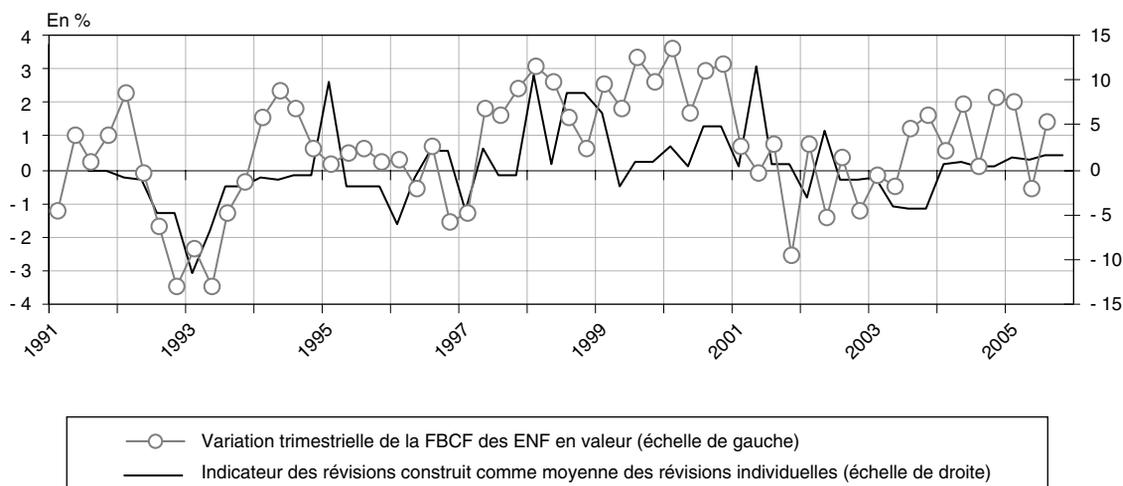
Lecture : estimation non paramétrique de la densité par la méthode du noyau (fenêtre gaussienne).
Champ : industrie manufacturière, toutes strates et toutes saisons confondues.
Source : enquête Investissement, Insee. Calculs de l'auteur.

8. Découlant de cette définition générale, plusieurs définitions mathématiques existent. Cf. Lecoutre et Tassi (1987) pour plus d'éléments sur ce sujet.

9. 36 % des révisions sont nulles ou d'une amplitude très faible, inférieure à un millième du chiffre d'affaires.

10. En statistique robuste, le terme consacré est « paramètre de position » (« location parameter »). Il désigne toute statistique telle que, lorsque la distribution est translatée, la statistique est translatée d'autant. La moyenne et la médiane sont des exemples classiques de paramètres de position.

Graphique IV
Comparaison de la série construite par moyenne des indicateurs individuels de révision avec la série des taux de croissance trimestrielle de l'investissement des entreprises



Lecture : moyennes pondérées par strate et par taille d'entreprise des indicateurs individuels de révision $d_{i,r}$. Les séries obtenues sont centrées à zéro et réduites à une variance unitaire, ceci pour chaque saison d'enquête (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre).
Champ : industrie manufacturière (indicateur des révisions) et entreprises non financières (FBCF).
Sources : enquête Investissement, Insee (indicateur des révisions) et comptes nationaux trimestriels (FBCF). Calculs de l'auteur.

Un grand nombre de M-estimateurs (définis par leur fonction de coût ρ) existent. Les plus classiques sont énumérés dans l'encadré 1.

L'estimateur utilisé ici fait partie de la famille de M-estimateurs proposée par Moberg, Ramberg et Randles en 1980.

Encadré 1

CHOISIR UNE FONCTION DE COÛT ρ

Nous présentons ici quelques exemples de M-estimateurs. Les premiers (MCO, médiane), exposés à titre d'exemples simples, ne possèdent pas des propriétés adaptées aux distributions étudiées ici. En revanche, les autres familles proposées sont susceptibles de convenir. Dans chaque cas, l'estimateur est défini par la fonction de coût ρ et/ou par sa dérivée, c'est-à-dire par la fonction de score ψ .

Les MCO (Moindres Carrés Ordinaires)

$$\rho(x) = \frac{1}{2}x^2 \text{ et } \psi(x) = x$$

L'estimateur des MCO est un cas particulier des M-estimateurs. Sous l'hypothèse de distributions gaussiennes, il est l'estimateur optimal au sens de l'efficacité, c'est-à-dire qu'il est sans biais et que sa variance est minimale. Toutefois, le caractère quadratique de la fonction de coût le rend beaucoup trop sensible aux valeurs extrêmes.

Les distributions empiriques envisagées ici ne sont pas gaussiennes et les queues très épaisses des distributions rendent les MCO beaucoup trop sensibles aux valeurs extrêmes.

La médiane

$$\rho(x) = |x| \text{ et } \psi(x) = \text{signe}(x)$$

avec :

$$\begin{cases} \text{signe}(x) = -1 & \text{si } x < 0 \\ \text{signe}(x) = 0 & \text{si } x = 0 \\ \text{signe}(x) = 1 & \text{si } x > 0 \end{cases}$$

L'utilisation de la médiane dans les régressions linéaires a été envisagée dès le début du XIX^e siècle par Pierre Simon de Laplace. Cet estimateur a l'avantage d'être très robuste aux valeurs extrêmes. Cependant, il ne capte aucune information en dehors du point médian. Ici, la très grande concentration en zéro des distributions empiriques conduit à des médianes la plupart du temps nulles. Il n'est donc pas pertinent ici.

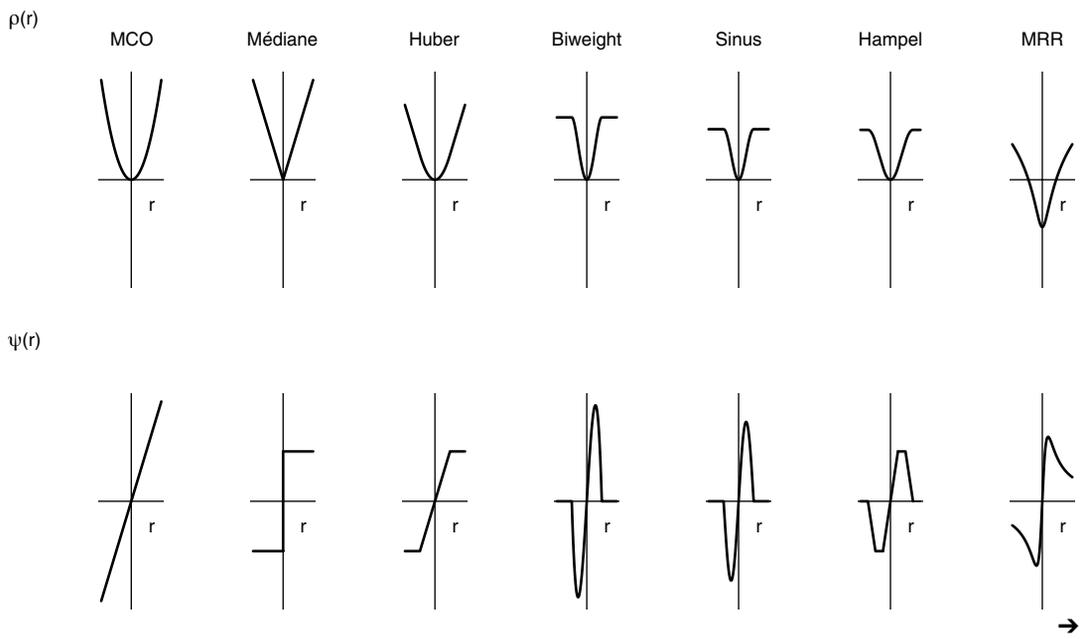
Huber

$$\psi(x) = x \text{ si } |x| < c \text{ et } \psi(x) = c \cdot \text{signe}(x) \text{ si } |x| > c$$

Le M-estimateur proposé par Huber (1964) est équivalent aux MCO autour du centre de la distribution et équivalent à la médiane sur les queues de distribution. Il a l'avantage par rapport aux autres M-estimateurs présentés *infra* de garantir l'existence et l'unicité de la solution du problème de minimisation associé. Toutefois, dans le cas des distributions envisagées ici (queues très épaisses et très fortes concentrations en zéro), il se révèle être trop peu robuste. Nous lui préférons alors des M-estimateurs avec des fonctions de score dites « redescendantes », c'est-à-dire des fonctions de score qui tendent vers zéro à l'infini.

Schéma

Exemples de fonctions objectif ρ et fonctions de score ψ associées



L'échantillon de l'enquête *Investissement* est stratifié par secteur et par taille d'entreprise. L'estimateur « robuste » est d'abord calculé par strate, puis nous agrégeons ces estimations par strate, en tenant compte de coefficients de redressement. La méthode d'agrégation utilisée au sein de chaque strate nécessite de disposer d'un nombre suffisant d'observations. Le niveau de stratification utilisé doit être relativement agrégé : l'échantillon est ainsi divisé par secteur NES 16 (Nomenclature Économique de Synthèse en 16 postes) et par taille d'effectif salarié (3 tranches de taille : moins de 100 salariés, de 100 à 499 salariés et 500 salariés et plus). L'industrie manufacturière (11) comporte quatre secteurs NES 16. Toutefois, le secteur de l'automobile est trop petit et trop concentré pour être divisé en trois tranches de taille avec la méthode d'agrégation utilisée. Il est regroupé en une seule strate ; si bien que le calcul est finalement réalisé sur dix strates.

La construction précise de l'indicateur au sein de chaque strate est expliquée dans l'annexe 1 et le choix de la fonction de coût ρ est détaillé dans l'annexe 2.

La construction ainsi décrite n'est toutefois pas la plus efficace (l'efficacité est le fait que l'intervalle de confiance de l'estimation soit de largeur minimale, sous couvert que cette estimation soit sans biais). En effet, il appa-

raît une forte dépendance de l'amplitude des résidus à la taille des entreprises : la variance diminue lorsque la taille des entreprises augmente. Deux raisons l'expliquent. D'une part, les grandes entreprises sont susceptibles de disposer de meilleurs systèmes d'information interne et d'avoir des systèmes de contrôle de gestion plus performants que les petites entreprises. D'autre part, la multiplicité des activités d'une grande entreprise et la diversité de ses projets d'investissement rendent possibles des compensations partielles : la réduction de certains investissements peut être compensée par l'apparition d'autres investissements.

Cet effet taille crée de l'hétéroscédasticité (12), qui diminue l'efficacité de l'estimation par les M-estimateurs : à taille donnée de l'échantillon, l'intervalle de confiance de l'estimation n'est pas de largeur minimale. Il est alors plus efficace de transformer le modèle afin de réaliser les estimations sur des distributions vérifiant

11. Jusqu'en 2004 les secteurs de l'énergie et des industries agroalimentaires étaient relativement mal couverts par l'enquête. De ce fait, ils sont exclus du champ de l'indicateur, ainsi restreint à la seule industrie manufacturière.

12. L'hétéroscédasticité est le fait que les résidus ne soient pas « homoscedastiques », c'est-à-dire qu'ils n'aient pas même variance. L'hétéroscédasticité ne crée pas de biais, mais elle limite la rapidité de convergence des estimateurs. Ainsi, pour le même nombre d'observations, l'intervalle de confiance de l'estimation sera plus grand en présence d'hétéroscédasticité que dans le cas homoscedastique.

Encadré 1 (suite)

Plusieurs fonctions « redescendantes »

Biweight de Tukey :

$$\psi(x) = \frac{x}{c} \left(1 - \frac{x^2}{c^2} \right)^2 \text{ si } |x| < c, 0 \text{ sinon}$$

Sinus d'Andrew :

$$\psi(x) = \frac{1}{\pi} \sin\left(\frac{\pi \cdot x}{c}\right) \text{ si } |x| < c, 0 \text{ sinon}$$

Hampel :

$$\left\{ \begin{array}{ll} \psi(x) = x & \text{si } |x| < a \\ \psi(x) = a \cdot \text{signe}(x) & \text{si } a < |x| < b \\ \psi(x) = a \frac{\text{signe}(x)c - x}{c - b} & \text{si } b < |x| < c \\ \psi(x) = 0 & \text{si } c < |x| \end{array} \right.$$

Toutes les fonctions de score données par les formules ci-dessus sont à valeurs « redescendantes ». De plus, elles s'annulent à partir d'un certain seuil ; si bien que les points trop éloignés du centre de la distribution sont complètement rejetés de l'estimation.

Les M-estimateurs de MRR (famille retenue ici)

$$\rho(x) = \frac{c}{2} \log(x^2 + c) \text{ et } \psi(x) = \frac{c \cdot x}{x^2 + c}$$

Pour des distributions telles qu'observées ici, Moberg, Ramberg et Randles (1980) proposent de choisir une fonction de score dans la famille définie ci-dessus. Nous donnons à cette famille de M-estimateurs les noms de ces auteurs (MRR en abrégé). Ces fonctions ont l'avantage d'être à valeurs redescendantes sans pour autant rejeter totalement des points de l'estimation : contrairement aux trois familles de fonctions précédentes (Biweight, Sinus et Hampel), les fonctions de score ne s'annulent pas à partir d'un certain seuil. Elles sont toujours proches de la proportionnalité autour de 0, si bien que les M-estimateurs correspondants s'apparentent toujours aux MCO autour du centre de la distribution. Ces estimateurs se révèlent être proches de l'efficacité – au sens où l'intervalle de confiance est peu large et l'estimation est sans biais – pour des distributions très concentrées et à queues très épaisses. Ce sont ces estimateurs qui seront retenus par la suite.

l'hypothèse d'homoscédasticité, cette correction étant réalisée par une procédure en deux temps (cf. annexe 1).

En s'inspirant de la méthode des MCQG (Moindres Carrés Quasi Généralisés), on réalise une première estimation par la méthode des M-estimateurs. Elle permet d'extraire les résidus. L'amplitude de ces derniers varie avec le chiffre d'affaires de chaque observation. La dépendance de la variance des résidus à la taille de l'entreprise est alors estimée. Dans un deuxième temps, le modèle est transformé en le divisant par la racine carrée de la variance ainsi estimée. Dans ce modèle transformé, les résidus sont alors d'égale variance. L'estimation est alors conduite une nouvelle fois, non plus sur le modèle initial mais sur ce modèle transformé, qui vérifie l'hypothèse d'homoscédasticité. La méthode est détaillée dans la deuxième partie de l'annexe 1.

Série trimestrielle des indicateurs de révision d'investissement

Pour chaque strate H , nous disposons en réalité de trois séries, une pour chaque saison S (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre). Pour être comparables, ces séries sont centrées en zéro et réduites à une variance unitaire.

Dans chaque strate H , les trois séries à pas annuel sont alors réunies en une unique série trimestrielle : les révisions de l'enquête de janvier N correspondent au premier trimestre de

l'année N , celles d'avril au deuxième trimestre, celles d'octobre aux troisième et quatrième trimestres.

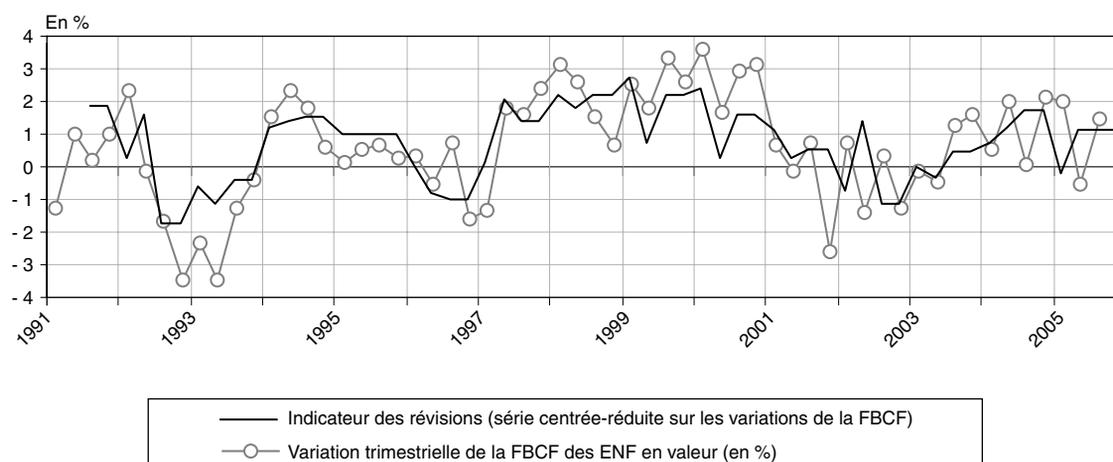
Les séries trimestrielles des révisions par strate sont ensuite agrégées en une unique série pour l'ensemble de l'industrie manufacturière. Pour cela, nous appliquons des coefficients de redressement constants, calculés comme les montants d'investissement par strate issus de l'EAE dans l'industrie de 2002.

Par construction, la série ainsi construite ne peut avoir de saisonnalité constante. Toutefois, il est possible qu'une saisonnalité évolutive subsiste. Les tests de saisonnalité de la procédure X11-ARIMA infirment cette hypothèse.

Cette série trimestrielle des révisions d'investissement (nommée par la suite « indicateur des révisions » et notée REV) apparaît fortement corrélée aux variations trimestrielles de la FBCF des ENF en valeur (cf. graphique VI). Sur la période qui s'étend du troisième trimestre de 1991 au troisième trimestre de 2005, la corrélation s'établit à 68 %. Pour un trimestre donné, l'indicateur est disponible au milieu de ce même trimestre. Il permet de prévoir l'investissement de ce trimestre pour la *Note de conjoncture* de l'Insee qui paraît à la fin du trimestre (13).

13. Ceci n'est pas encore possible pour les troisièmes trimestres et les Points de conjoncture d'octobre qui leur correspondent. Toutefois, à terme, l'enquête de juillet permettra de construire un indicateur disponible aussi fin août et permettant de prévoir le troisième trimestre, ceci en particulier lors du Point de conjoncture d'octobre.

Graphique VI
Comparaison de la série construite des indicateurs de révision avec la série des taux de croissance trimestrielle de l'investissement des entreprises



Champ : industrie manufacturière (indicateur des révisions) et entreprises non financières (FBCF).
 Sources : enquête Investissement, Insee (indicateur des révisions) et comptes nationaux trimestriels (FBCF). Calculs de l'auteur.

L'indicateur des révisions dans l'enquête *Investissement* apporte une information précieuse au conjoncturiste

L'indicateur des révisions étant donc bien corrélé avec les évolutions en valeur des investissements des entreprises, il nous faut maintenant vérifier que cette série est susceptible d'apporter un surcroît d'information pour les prévisions d'investissement. Plus que les évolutions en valeur, ce sont celles en volume qui intéressent le conjoncturiste. Nous allons donc maintenant porter notre intention sur les variations trimestrielles à prix constants de la FBCF des ENF (14).

Un dossier (15) dans la *Note de conjoncture* de l'Insee de mars 2005 proposait un panorama des indicateurs disponibles pour prévoir cette variable. Parmi ces indicateurs (16), le plus utile pour cette prévision est le solde d'opinion relatif aux intentions de commande en biens d'équipement dans l'enquête bimestrielle dans le commerce de gros de l'Insee (notée par la suite enquête *Commerce de gros*) (17). Nous voulons comparer l'apport d'information de ces deux indicateurs (indicateur des révisions et intentions de commande) pour la prévision des évolutions de l'investissement.

La trimestrialisation du solde d'opinion des intentions de commande en biens d'équipement est exposée dans l'encadré 2. La série trimestrialisée du solde d'opinion est moins pertinente

que ses propres évolutions pour prévoir les évolutions de l'investissement. Cette série des évolutions du solde trimestrialisé est notée dIC .

Nous construisons un modèle d'étalement multivarié autorégressif (modèle *VAR*) qui lie les taux de croissance trimestriel de la FBCF des ENF (série notée \dot{FBCF}_t) à l'indicateur des révisions (noté REV_t) et à la série dIC (18). Le vecteur du modèle est donc :

$$X_t = \begin{pmatrix} \dot{FBCF}_t \\ REV_t \\ dIC_t \end{pmatrix} \quad (2)$$

Dans ce vecteur :

- \dot{FBCF}_t est le taux de croissance trimestriel de la FBCF des ENF en volume au trimestre t ;

14. Rappelons que les comptes trimestriels utilisés dans cet article sont les premiers résultats du troisième trimestre de 2005 en base 2000.

15. Ferrari (2005a).

16. L'indicateur des investissements globaux de l'enquête de la Banque de France auprès des établissements de crédit n'est plus disponible.

17. Pour plus d'information, se reporter à la fiche méthodologique de cette enquête sur le site www.insee.fr (rubrique « Conjoncture », puis « Indicateurs de conjoncture », « Autres indicateurs » et enfin « méthodologie » dans la ligne « Enquête bimestrielle dans le commerce de gros »).

18. L'indicateur des révisions étant a priori davantage lié aux évolutions en valeur de l'investissement qu'à ces mêmes évolutions en volume, il serait naturel d'ajouter également une variable relative aux évolutions de prix de la FBCF. Toutefois, nous avons vérifié qu'une telle variable (croissance trimestrielle ou glissement annuel des prix de la FBCF des ENF) n'améliore pas sensiblement la prévision.

Encadré 2

LES ÉVOLUTIONS D'INTENTIONS DE COMMANDE EN BIENS D'ÉQUIPEMENT DE L'ENQUÊTE COMMERCE DE GROS (dIC)

L'enquête *Commerce de gros* est bimestrielle. Elle recueille en particulier les intentions de commande des grossistes en biens d'équipement. Plus précisément, ceux-ci indiquent si les commandes qu'ils ont l'intention de passer au cours des deux prochains mois pour ce type de biens sont « supérieures à la normale », « normales » ou « inférieures à la normale ». Avec ces réponses, l'Insee calcule un solde d'opinion, qui se révèle être bien corrélé au taux de croissance trimestriel de l'investissement des entreprises.

Nous trimestrialisons cette série bimestrielle : nous affectons la valeur de l'indicateur au mois de l'enquête et au mois suivant. L'indicateur retenu (IC) est la moyenne sur les trois mois du trimestre correspondant. Ainsi, l'indicateur du quatrième trimestre est pour 1/3 la valeur à l'enquête de septembre et pour 2/3

celle à l'enquête de novembre, publiée à la fin de ce même mois. À titre de comparaison, l'indicateur des révisions est lui disponible mi-novembre. L'indicateur pour le troisième trimestre est pour 2/3 la valeur à l'enquête de juin et pour 1/3 celle à l'enquête de septembre, publiée à la fin de ce même mois. L'indicateur des révisions est lui disponible au cours du mois d'août. Dans tous les cas, l'indicateur des révisions est donc disponible plus tôt que l'indicateur des intentions de commande ainsi trimestrialisé.

L'indicateur des intentions de commande portant sur le niveau des commandes par rapport à la normale, son évolution prévoit mieux les évolutions de l'investissement que son propre niveau. Ceci a été vérifié empiriquement. Nous notons alors dIC la différence première de l'indicateur ($dIC_t = IC_t - IC_{t-1}$).

- REV_t est l'indicateur des révisions affecté au trimestre t . Par exemple, pour le quatrième trimestre de 2005, il s'agit de l'indicateur calculé avec l'enquête *Investissement* d'octobre 2005 ;

- le calendrier exact de dIC_t est donné dans l'encadré 2.

Remarquons que ce modèle est estimé avec les comptes définitifs : les comptes trimestriels de l'année N sont calés sur les comptes annuels (19) en avril $N + 2$. Avant cette date, les comptes trimestriels peuvent connaître des révisions. Une fois ce calage réalisé, les comptes trimestriels sont quasiment figés et ils sont dits « définitifs ». Alors que le modèle *VAR* est estimé avec les comptes définitifs, il est utilisé en prévision avec des comptes provisoires. Cette faiblesse est générale aux étalonnages et il est impossible d'éviter cette difficulté. En ce qui concerne l'investissement, les révisions des comptes trimestriels sont parfois fortes et il s'agit donc ici d'une limite importante.

L'estimation du processus et les tests associés sont donnés en annexe 3. Un seul retard est retenu au vu des résultats des tests usuels.

L'indicateur REV est disponible un peu plus de trois mois avant la publication des comptes du trimestre correspondant. L'indicateur dIC est lui disponible environ deux mois avant. De manière à pouvoir utiliser les valeurs des deux indicateurs REV_t et dIC_t au moment de la prévision du trimestre courant t , nous écrivons le modèle sous sa forme dite bloc-réursive. Cette forme peut être obtenue en transformant le système des trois équations du modèle *VAR* à l'aide d'une décomposition de Cholesky de la matrice de variance-covariance. Cependant, une manière plus simple et parfaitement équivalente d'obtenir cette forme est d'effectuer directement la régression de la série $FBCF$ sur son premier retard, sur les séries REV et dIC ainsi que sur le premier retard de ces deux séries (cf. l'équation 3.6 de l'annexe 3). La forme bloc-réursive permet alors d'utiliser toute l'information disponible pour prévoir la $FBCF$ du trimestre concomitant aux indicateurs.

Par exemple, à la fin du mois de novembre 2005, les comptes trimestriels sont disponibles jusqu'au troisième trimestre de 2005. Les indicateurs REV et dIC sont eux disponibles jusqu'au quatrième trimestre de 2005 (enquête *Investissement* d'octobre 2005 et enquête

Commerce de gros de novembre 2005). Pour ce même trimestre, l'indicateur REV vaut 1,5 et dIC vaut 4,0. Sur la base de l'information disponible en décembre 2005, la forme bloc-réursive prévoit alors une évolution de la $FBCF$ des ENF de 1,2 % au quatrième trimestre de 2005. La forme canonique permet de prolonger la prévision : le taux de croissance de la $FBCF$ des ENF serait de 1,3 % au premier trimestre de 2006 puis de 1,2 % au deuxième trimestre.

Sur la période allant du premier trimestre de 1992 au troisième trimestre de 2005, l'écart-type des résidus à l'horizon d'un trimestre (forme bloc-réursive) vaut 1,06. En prolongeant à l'horizon de deux trimestres la prévision, l'écart-type s'établit à 1,17. Enfin, à l'horizon de trois trimestres, il est de 1,46. Ces valeurs sont à comparer à l'écart-type du taux de croissance de la $FBCF$ sur cette même période : 1,68. En terme de corrélation, les prévisions à l'horizon d'un trimestre sont corrélées à 75 % avec les réalisations, celles à l'horizon de deux trimestres à 69 % et celles à l'horizon de trois trimestres à 40 %. Au regard de ces statistiques, il apparaît que ce modèle est un bon outil pour prévoir les évolutions d'investissement sur deux trimestres. Toutefois, comme tous les modèles d'étalonnage, il devient insuffisant pour prévoir les évolutions sur trois trimestres.

Le graphique VII représente les prévisions des taux de croissance à l'horizon d'un trimestre (forme bloc-réursive) ainsi que les réalisations et les résidus. Le graphique VIII présente ces mêmes prévisions en niveau.

Dans le passé, les prévisions du modèle *VAR* captent correctement les points de retournement dans le cycle conjoncturel de l'investissement. Par exemple, la reprise en 1997 de la $FBCF$ apparaît dans les comptes à partir du deuxième trimestre. Le modèle aurait parfaitement prévu cette reprise à partir des résultats de l'enquête *Investissement* d'avril 1997 et l'enquête *Commerce de gros* de mai 1997. De la même manière, la nette décélération des investissements au premier trimestre de 2001, marquant la sortie de la « bulle Internet », correspond bien à la prévision du modèle. Toutefois, le modèle sous-estime systématiquement la $FBCF$ sur la période allant du troisième trimestre de 2002 au quatrième trimestre de 2003.

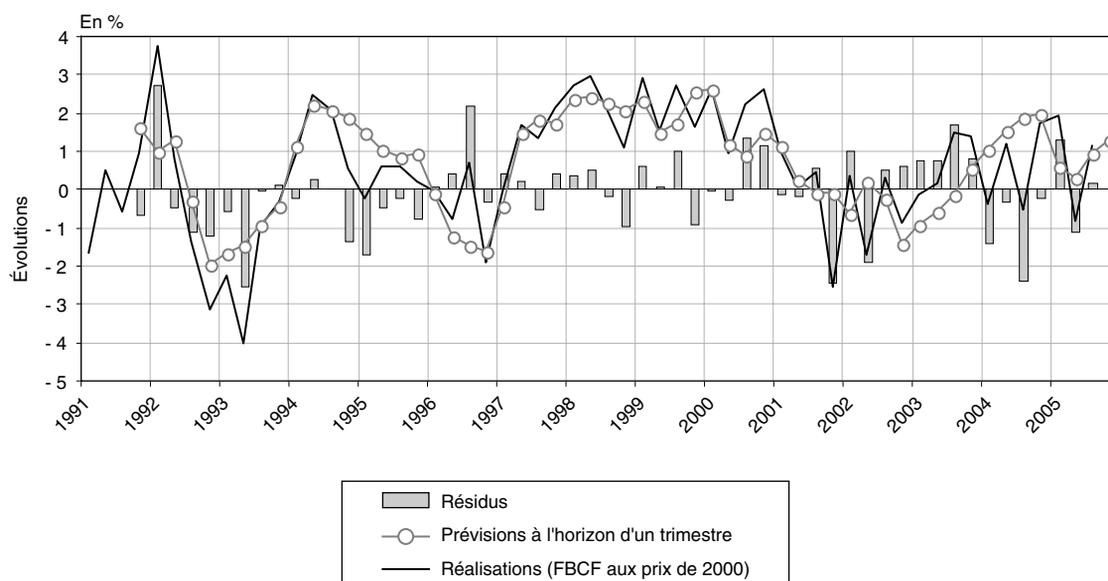
19. Il s'agit alors des comptes annuels dits « semi-définitifs », publiés un peu plus d'un an après la fin de l'année considérée.

L'encadré 3 présente des tests de causalité au sens de Granger pour les deux séries *REV* et *dIC*. L'indicateur des révisions *REV* apporte très significativement de l'information pour la prévision des évolutions courantes et futures de la FBCF des ENF. La série *dIC* apporte également de l'information pour la prévision des évolutions futures, mais un peu moins significativement que dans le cas de la série *REV*. Au contraire, la série

dIC n'apporte pas d'information supplémentaire pour la prévision des évolutions courantes.

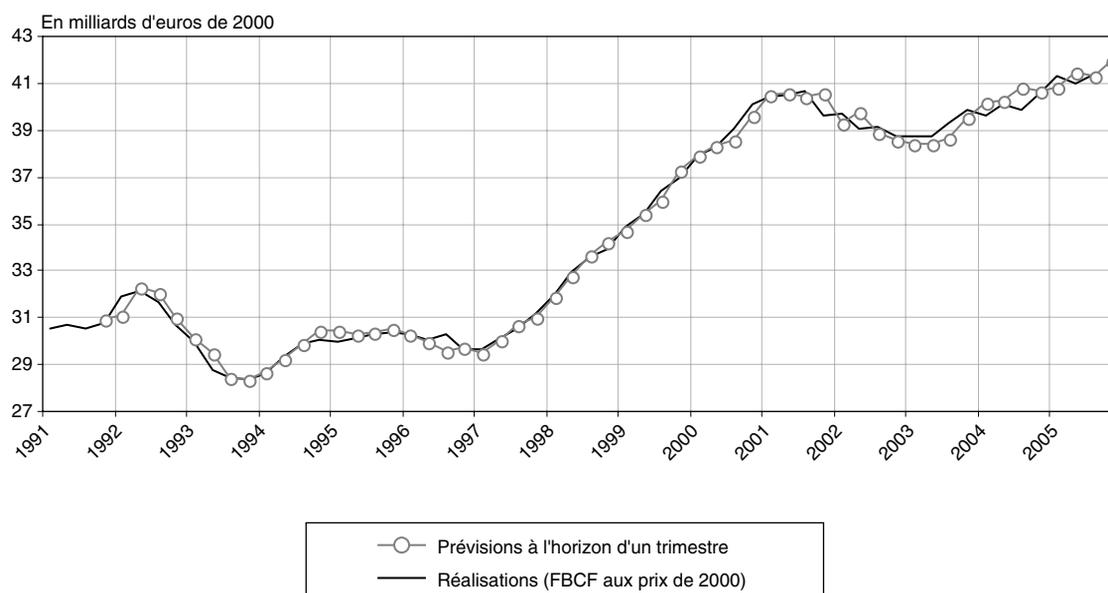
De plus, les indicateurs des révisions sont disponibles en moyenne un mois avant les intentions de commande. La série des indicateurs des révisions est donc une variable précieuse pour la prévision des évolutions de l'investissement des entreprises.

Graphique VII
Prévision des variations trimestrielles de la FBCF des ENF en volume à l'horizon d'un trimestre



Champ : entreprises non financières.
Sources : enquête Investissement, Insee et comptes nationaux trimestriels. Calculs de l'auteur.

Graphique VIII
Prévision de la FBCF des ENF en volume à l'horizon d'un trimestre



Champ : entreprises non financières.
Sources : enquête Investissement, Insee et comptes nationaux trimestriels. Calculs de l'auteur.

Encadré 3

TESTS DE CAUSALITÉ AU SENS DE GRANGER POUR LA PRÉVISION DE LA FBCF DES ENF

Nous voulons vérifier si la série des révisions (*REV*) d'une part, et la série des évolutions des intentions de commande en biens d'équipement (*dIC*) d'autre part, apportent bien un surcroît d'information pour la prévision de l'évolution courante et celle de l'évolution future de la FBCF des ENF. Nous réalisons pour cela des tests de causalité au sens de Granger. Ces tests ne sont valables que sous l'hypothèse de normalité des séries. Ici, la normalité de chaque série n'est jamais rejetée par les tests de Doornick et Hansen (1994) et par ceux de Jarque et Bera (1987).

Les premiers tests sont des tests de causalité non-instantanée. Par exemple, pour tester l'apport d'information de la série *REV*, le test vérifie que le passé de la série *REV* apporte bien de l'information par rapport aux évolutions passées de la FBCF et aux valeurs passées de l'indicateur *dIC*. Nous testons donc dans l'équation du modèle *VAR* relative à la variable *FBCF* (cf. annexe 3, équation 3.3) la nullité du coefficient relatif à REV_{t-1} par un test de Fisher à un seul degré de liberté. La statistique de test vaut 8,69. Elle est à comparer à une loi de Fisher $F(1,44)$. La *P-value* est de 5.10^{-3} . La nullité du coefficient du retard de la série explicative *REV* est donc rejetée. Le passé de la série des révisions apporte donc significativement de l'information pour la prévision de la FBCF.

De la même manière, le test de causalité non-instantanée sur la variable *dIC* indique que cette série apporte également de l'information pour la prévision de l'investissement. Toutefois, cet apport d'information est un peu moins significatif que celui de la série des révisions : la *P-value* est plus importante (0,02).

Les seconds tests vérifient la causalité instantanée : par exemple, pour la série *REV*, le test vérifie que les valeurs courantes de cette série apportent bien de l'information par rapport aux évolutions passées de l'investissement ($FBCF_{t-1}$), aux retards de la série de révision (REV_{t-1}) et aux valeurs courantes et passées de la série *dIC* (dIC_t et dIC_{t-1}). Nous testons dans l'équation de la forme bloc-réursive (cf. annexe 3, équation 3.6) la nullité du coefficient relatif au terme REV_t . Le test est alors un test de Fisher à un seul degré de liberté. La statistique de test vaut 10,70. Elle est à comparer à la loi de Fisher $F(1,42)$. La *P-value* est de 2.10^{-3} . La série *REV* des révisions apporte donc de l'information supplémentaire pour la prévision des variations concomitantes de l'investissement.

Au contraire, le test de causalité instantanée pour les évolutions des intentions de commande suggère que les valeurs courantes de cette série *dIC* n'apportent pas d'information supplémentaire pour la prévision de l'investissement (*P-value* de 0,38). Nous complétons ce test avec un autre test qui n'entre pas dans le cadre strict des tests de causalité au sens de Granger : il teste l'apport d'information supplémentaire de la série *dIC* et de son retard pour la prévision des évolutions courantes de l'investissement. La *P-value* de ce test est de 0,21. La nullité commune des coefficients de dIC_t et dIC_{t-1} dans l'équation de la forme bloc-réursive (cf. annexe 3, équation 3.6) n'est donc également pas rejetée.

La série *dIC* n'apporte donc pas significativement d'information supplémentaire par rapport à la série *REV* et au retard de la FBCF pour la prévision concomitante de la FBCF.

Tests de causalité au sens de Granger des variables *REV* et *dIC* pour la prévision des évolutions de la FBCF des ENF

Tests de causalité non-instantanée (équation 3.3)

Variabes testées	Statistique de Fisher	Degrés de liberté	<i>P-value</i>
REV_{t-1}	8,69	1 et 44	5.10^{-3}
dIC_{t-1}	5,64	1 et 44	0,02

Tests de causalité instantanée (équation 3.6)

Variabes testées	Statistique de Fisher	Degrés de liberté	<i>P-value</i>
REV_t	10,70	1 et 42	2.10^{-3}
dIC_t	0,78	1 et 42	0,38
REV_t et REV_{t-1}	10,47	2 et 42	2.10^{-4}
dIC_t et dIC_{t-1}	1,60	2 et 42	0,21

Lecture : la nullité du coefficient relatif au terme REV_{t-1} dans l'équation (3.3) est testée par un test de Fisher à 1 et 44 degrés de liberté. La statistique de test est 8,69. La *P-value* associée est 5.10^{-3} .

Période : du 1^{er} trimestre de 1992 au 4^e trimestre de 2003.

Sources : enquête Investissement, Insee, enquête Commerce de gros et comptes trimestriels. Calculs de l'auteur.

Les révisions d'investissement sont liées à des modifications de demande et prévoient l'activité industrielle

Les changements de plan d'investissement reflètent des modifications d'anticipations. Celles-ci peuvent être diverses : modifications des perspectives de demande, évolution des prix relatifs de la valeur ajoutée et des biens d'équipement, changements dans les conditions de financement. La littérature économique a mis en évidence que la demande adressée aux entreprises a un rôle prépondérant parmi ces différents facteurs. Ceci est dû au mécanisme appelé « effet accélérateur » : les entrepreneurs accumulent du capital pour ajuster leurs capacités de production au niveau de la demande anticipée. S'ils s'attendent à une hausse de la demande, leur niveau de capital idéal s'accroît (20).

Comme cela avait déjà été montré par Houdebine (1998), une révision à la hausse dans l'enquête *Investissement* signale ainsi une amélioration des perspectives de demande. Elle laisse alors présager une augmentation de l'activité industrielle dans la mesure où il n'y a pas de contraintes trop fortes sur les capacités de production. L'indicateur présenté dans cet article pourrait donc aider à la prévision de la production manufacturière, qui répond à court terme aux évolutions de la demande.

Nous voulons vérifier l'apport d'information de cet indicateur par rapport aux indicateurs d'enquête utilisés habituellement pour la prévision de la production manufacturière. Trois indicateurs ont été envisagés. Tous trois sont issus de l'enquête mensuelle de conjoncture dans l'industrie de l'Insee (21) (notée ici enquête *Industrie*) : le solde relatif à la tendance passée de la production (noté *TPPA*), celui relatif aux perspectives personnelles de production (noté *TPPRE*) et l'indicateur synthétique du climat des affaires (22). Nous ajoutons à ces trois indicateurs la série des révisions *REV*. Une méthode itérative de sélection des variables significatives nous a fait retenir cet indicateur et le solde *TPPRE* comme les deux séries les plus informatives pour la prévision. L'annexe 4 présente la construction d'un modèle *VAR* qui lie les taux de croissance trimestriels de la production manufacturière *PROD* à la série des révisions *REV* et à cette série *TPPRE*. Ce modèle *VAR* est inspiré de Reynaud et Scherrer (1997), qui utilisaient les soldes *TPPA* et *TPPRE* pour prévoir les évolutions de la production manufactu-

rière. Dans notre modélisation, un seul retard a été retenu par les tests usuels.

Les tests de causalité au sens de Granger (cf. encadré 4) montrent que la série des révisions aide à la prévision des évolutions courantes et surtout des évolutions futures de la production manufacturière. En ce sens, l'indicateur des révisions apporte une information avancée sur l'activité industrielle.

* *
*

Alors que les indicateurs d'enquête pour la prévision de la FBCF des entreprises sont rares, il apparaît que l'enquête *Investissement* apporte une information conjoncturelle précieuse pour juger de l'évolution future de l'investissement des entreprises. Les résultats publiés sous forme de taux de croissance annuelle sont très utiles pour une approche qualitative mais présentent certaines limites pour mettre en œuvre des outils quantitatifs de prévision à un rythme trimestriel. Cette limite est dépassée par l'indicateur des révisions : cet indicateur apporte une information pertinente pour la prévision des variations trimestrielles de l'investissement.

Certaines pistes restent néanmoins à étudier pour améliorer la qualité de l'information issue des révisions. Cet article ne s'est intéressé qu'au centre des révisions. Mais l'amplitude des révisions pourrait être également informative : la présence de plus fortes révisions pourrait être le signe avant-coureur d'un retournement conjoncturel. L'étude de l'amplitude des révisions a conduit jusqu'ici à un échec mais des méthodes économétriques plus fines permettraient éventuellement d'extraire une telle information. D'autre part, l'étude de l'asymétrie des distributions pourrait être une méthode concurrente pour juger du dynamisme de l'investissement.

Plus en aval, des travaux supplémentaires devraient être mis en œuvre pour utiliser au

20. Cf. par exemple Artus et Muet (1984), Epaulard (2001), Mairesse, Mulkay et Hall (2001) et Herbert (2001).

21. Cf. la fiche méthodologique de cette enquête disponible sur le site www.insee.fr, rubrique « Conjoncture », puis « Indicateurs de conjoncture », « Enquête mensuelle de conjoncture dans l'industrie » et enfin « Ouvrir la note méthodologique ».

22. Pour trimestrialiser ces trois séries mensuelles, nous affectons au trimestre courant le résultat de l'enquête du premier mois du trimestre. Ces indicateurs sont donc disponibles environ deux semaines avant l'indicateur des révisions. Pour plus d'information sur l'indicateur synthétique, le lecteur pourra se reporter à l'article de Doz et Lengart (1999).

mieux l'indicateur dans des étalonnages d'investissement et d'activité industrielle. Les modèles d'étalonnage sont présentés dans cet article à titre illustratif mais ils n'ont pas la prétention d'être les meilleurs possibles.

Enfin, la qualité des résultats macroéconomiques encourage l'utilisation de l'enquête *Investissement* pour des études microéconométriques des comportements d'investissements des industriels. Cette voie a déjà été explorée

par Naboulet et Raspiller (2007). En particulier, des travaux pourraient relier les révisions individuelles d'investissement à des facteurs tels que les anticipations de demande des industriels et leurs opinions sur les conditions de financement mesurées par l'enquête *Industrie* et l'enquête *Trésorerie* (23). □

23. La très grande majorité des entreprises de l'enquête Investissement est également interrogée par l'enquête Industrie et l'enquête Trésorerie de l'Insee.

Encadré 4

**TESTS DE CAUSALITÉ AU SENS DE GRANGER
POUR LA PRÉVISION DE LA PRODUCTION MANUFACTURIÈRE**

Nous voulons vérifier que l'indicateur trimestriel *REV* des révisions d'anticipations d'investissement des industriels apporte bien un surcroît d'information pour la prévision de l'évolution courante et celle de l'évolution future de la production manufacturière par rapport à l'information apportée par la série des perspectives personnelles de production (*TPPRE*). Nous réalisons pour cela des tests de causalité au sens de Granger de la même manière que dans l'encadré 3. Ces tests ne sont valables que sous l'hypothèse de normalité des séries. La normalité de chaque série n'est jamais rejetée par les tests de Doornick et Hansen (1994) ni par ceux de Jarque et Bera (1987).

Le test de causalité non-instantanée ne rejette pas l'apport d'information de la série *REV* pour la prévision des évolutions futures de la production manufacturière (*P-value* de 0,03). De même, le test de causalité instantanée montre que la série apporte très significativement de l'information pour la prévision concomitante (*P-value* de 3.10^{-3}).

Tests de causalité au sens de Granger des variables *REV* et *TPPRE* pour la prévision des évolutions de la production manufacturière

Tests de causalité non-instantanée (équation 4.2)

Variables testées	Statistique de Fisher	Degrés de liberté	<i>P-value</i>
<i>REV</i> _{<i>t</i>-1}	4,86	1 et 44	0,03
<i>TPPRE</i> _{<i>t</i>-1}	0,04	1 et 44	0,84

Tests de causalité instantanée (équation 4.5)

Variables testées	Statistique de Fisher	Degrés de liberté	<i>P-value</i>
<i>REV</i> _{<i>t</i>}	9,99	1 et 42	3.10^{-3}
<i>TPPRE</i> _{<i>t</i>}	45,55	1 et 42	3.10^{-8}
<i>REV</i> _{<i>t</i>} et <i>REV</i> _{<i>t</i>-1}	7,68	2 et 42	1.10^{-3}
<i>TPPRE</i> _{<i>t</i>} et <i>TPPRE</i> _{<i>t</i>-1}	23,47	2 et 42	1.10^{-7}

Lecture : la nullité du coefficient relatif au terme *REV*_{*t*-1} dans l'équation (4.2) est testée par un test de Fisher à 1 et 44 degrés de liberté. La statistique de test est 4,86. La *P-value* associée est 0,03.

Période : du 1^{er} trimestre de 1992 au 4^e trimestre de 2003.

Sources : enquête Investissement, Insee, enquête Industrie et comptes trimestriels. Calculs de l'auteur.

BIBLIOGRAPHIE

- Artus P. et Muet P.-A. (1984)**, « Un panorama des développements récents de l'économétrie de l'investissement », *Revue économique*, vol. 35, n° 5, pp. 791-830.
- Doornik J.A. et Hansen H. (1994)**, « An Omnibus Test for Univariate and Multivariate Normality », *Discussion Paper*, n° 5, Nuffield College, Oxford University.
- Doz C. et Lenglart F. (1999)**, « Analyse factorielle dynamique : test du nombre de facteurs, estimation et application à l'enquête de conjoncture dans l'industrie », *Annales d'Économie et de Statistique*, n° 54, pp. 91-127.
- Epaulard A. (2001)**, « À la recherche des déterminants de l'investissement des entreprises », *Économie et statistique*, n° 341-342, pp. 3-14.
- Fayolle J. (1987)**, *Pratique contemporaine de l'analyse conjoncturelle*, Économica.
- Ferrari N. (2005a)**, « Prévoir l'investissement des entreprises ? Un indicateur des révisions d'anticipations dans l'enquête Investissement dans l'industrie », *Note de conjoncture*, Insee, mars 2005. Téléchargeable sur le site www.insee.fr.
- Ferrari N. (2005b)**, « Prévoir l'investissement des entreprises, un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie », *document de travail*, Insee, n° G2005/09. Téléchargeable sur le site www.insee.fr.
- Gouriéroux C. et Monfort A. (1996)**, *Statistique et modèles économétriques*, vol. 1, Économica (2^e édition).
- Herbet J.-B. (2001)**, « Peut-on expliquer l'investissement à partir de ses déterminants traditionnels au cours de la décennie 90 ? », *Économie et statistique*, n° 341-342, pp. 85-106.
- Hogg R.V. (1974)**, « Adaptive Robust Procedures : a Partial Review and Some Suggestions for Future Applications and Theory », *Journal of the American Statistical Association*, vol. 69, n° 348, pp. 909-923.
- Hogg R.V., Bril G.K., Han S.M. et Yul L. (1988)**, « An Argument for Adaptive Robust Estimation », *Probability and Statistics Essays in Honor of Franklin A. Graybill*.
- Houdebine M. (1998)**, « Prévisions d'investissement et utilisation des capacités », *Note de conjoncture*, Insee, juin 1998. Téléchargeable sur le site www.insee.fr.
- Huber P.J. (1964)**, « Robust Estimation of a Location Parameter », *Annals of Mathematical Statistics*, vol. 35, n° 2, pp. 73-101.
- Jarque C.M. et Bera A.K. (1987)**, « A Test for Normality of Observations and Regression Residuals », *International Statistical Review*, vol. 55, n° 2, pp. 163-172.
- Lecoutre J.-P. et Tassi P. (1987)**, *Statistique non paramétrique et robustesse*, Économica.
- Mairesse J., Mulkay B. et Hall B.H. (2001)**, « Investissement des entreprises et contraintes financières en France et aux États-Unis », *Économie et Statistique*, n° 341-342, pp. 67-84.
- Moberg T.F., Ramberg J.S. et Randles R.H. (1980)**, « An Adaptive Multiple Regression Procedure Based on M-estimators », *Technometrics*, vol. 22, n° 2, pp. 213-224.
- Naboulet A. et Raspiller S. (2007)**, « Déterminants de la décision d'investir et destination économique des équipements », *Économie et Statistique*, n° 395-396, ce numéro.
- Ravalet P. (1996)**, « L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration », *document de travail*, Insee, série Méthodologie statistique, n° 9605. Téléchargeable sur le site www.insee.fr.
- Reynaud M. et Scherrer S. (1997)**, « L'utilisation de l'enquête mensuelle auprès des chefs d'entreprise de l'industrie dans le diagnostic conjoncturel », *Note de conjoncture*, Insee, décembre 1997. Téléchargeable sur le site www.insee.fr.
- Rosenwald F. (1994)**, *L'enquête sur l'investissement industriel*, Insee Méthodes, n° 45.
-

CONSTRUCTION DE L'INDICATEUR

Régression robuste : méthode des M-estimateurs

Après stratification de l'échantillon, pour chaque strate H et pour chaque saison d'enquête S (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre), nous réalisons une régression robuste des indicateurs individuels de révision sur des indicatrices temporelles. Pour toutes les enquêtes t de la saison S , nous calculons ainsi les $m_{H,t}^e$, estimateurs des $m_{H,t}^0$, centres des distributions des révisions $d_{i,t}$ dans la strate H . Nous notons $1_{\{\tau\}}$ la fonction indicatrice de x , qui vaut 1 en x et 0 ailleurs. Pour chaque strate H et chaque saison S , le modèle s'écrit selon l'équation 1.1 :

$$d_{i,t} = \sum_{\tau \in S} m_{H,\tau}^0 1_{\{\tau\}}(t) + \varepsilon_{i,t} \quad (1.1)$$

pour chaque entreprise i de la strate H et chaque date t de la saison S .

Les perturbations $\varepsilon_{i,t}$ suivent des lois aléatoires que nous supposons indépendantes (l'absence d'effets fixes par entreprise n'est en fait pas assurée ; néanmoins, par mesure de simplicité et afin de retenir l'ensemble des réponses des entreprises, nous n'avons pas intégré de tels effets dans la modélisation).

Les M-estimateurs ne sont pas linéaires. De ce fait, ils ne sont pas invariants par homothétie. Il est donc nécessaire de diviser par un paramètre d'échelle $\sigma_{H,S}$ de la dispersion des $d_{i,t}$ de la strate H et de la saison S considérées :

$$\frac{d_{i,t}}{\sigma_{H,S}} = \sum_{\tau \in S} m_{H,\tau}^0 \frac{1_{\{\tau\}}(t)}{\sigma_{H,S}} + \tilde{\varepsilon}_{i,t} \quad (1.2)$$

pour chaque entreprise i de la strate H et chaque date t de la saison S .

Plusieurs statistiques sont possibles pour estimer les paramètres d'échelle $\sigma_{H,S}$. Celle retenue est l'écart absolu médian qui a l'avantage d'être robuste aux valeurs extrêmes. Cette statistique, notée MAD (*Median Absolute Deviation*), est donnée par la formule 1.3. Nous notons $d_{H,S}$ l'ensemble des $d_{i,t}$ pour les entreprises i de la strate H et les dates t de la saison S . Le coefficient 1,48 permet à la statistique MAD d'être égale à l'écart-type dans le cas d'une loi normale.

$$\begin{aligned} \sigma_{H,S} &= MAD(d_{H,S}) \\ &= 1,48 \cdot \text{Médiane} \left| d_{H,S} - \text{Médiane}(d_{H,S}) \right| \end{aligned} \quad (1.3)$$

Le problème de minimisation associé à l'estimation pour la strate H et la saison S s'écrit alors :

$$(m_{H,t}^e)_{t \in S} = \text{Arg} \min_{(m_t)_{t \in S}} \sum_{\tau \in S} \tilde{\rho}_{H,S}[\tau, d_{i,\tau}, (m_t)_{t \in S}] \quad (1.4)$$

avec

$$\tilde{\rho}_{H,S}[\tau, d_{i,\tau}, (m_t)_{t \in S}] = \rho \left(\frac{d_{i,\tau} - m_{t=\tau}}{MAD(d_{H,S})} \right) \quad (1.5)$$

Sous certaines conditions théoriques qui sont supposées vérifiées ici, les $m_{H,t}^e$ convergent vers des limites finies $m_{H,t}^0$ lorsque le nombre d'observations tend vers l'infini. Le lecteur pourra se reporter à l'ouvrage de Gouriéroux et Monfort (1996) pour une présentation précise de ces conditions (cf. également Ferrari, 2005b).

La méthode adaptative de choix de la fonction de coût ρ est détaillée dans l'annexe 2. Cette méthode nous conduit à choisir le M-estimateur dit « de MRR » (cf. encadré 1) avec une constante de réglage $c = 0,10$:

$$\rho(x) = \frac{0,10}{2} \log[x^2 + 0,10] \quad (1.6)$$

Correction de l'hétéroscédasticité : la méthode des M-estimateurs Quasi-Généralisés

L'estimateur ainsi décrit peut toutefois être rendu plus efficace. En effet, il apparaît une forte hétéroscédasticité dans les résidus. Leur variance diminue lorsque la taille des entreprises augmente. Autrement dit, rapportée aux chiffres d'affaires, l'amplitude des révisions décroît lorsque la taille de l'entreprise augmente.

La présence d'hétéroscédasticité diminue la vitesse de convergence des M-estimateurs $m_{H,t}^e$ vers leur valeur limite $m_{H,t}^0$. Il est donc plus efficace de transformer le modèle afin de réaliser les estimations sur des distributions vérifiant l'hypothèse d'homoscédasticité. Afin de corriger l'hétéroscédasticité, une procédure en deux temps est utilisée, à l'exemple de la méthode des MCQG (Moindres Carrés Quasi-Généralisés) : une première estimation par la méthode des M-estimateurs est réalisée. Elle permet d'extraire les résidus. L'amplitude de ces derniers varie avec le chiffre d'affaires de chaque entreprise. La dépendance de la variance des résidus $\Sigma_{H,S}^2(\cdot)$ à la taille de l'entreprise est alors estimée pour chaque strate H et chaque saison S . Dans un deuxième temps, le modèle est transformé en divisant par la racine carrée de la variance $\Sigma_{H,S}^e(CA_{i,A(t)})$ ainsi estimée. L'équation 1.2 est alors remplacée par l'équation 1.7, où les $\tilde{\varepsilon}_{i,t}$ sont alors de même variance. Ils peuvent alors être considérés de même loi au sein d'une strate et pour une saison donnée.

Pour chaque entreprise i de la strate H et chaque date t de la saison S :

$$\begin{aligned} \frac{d_{i,t}}{\sigma_{H,S} \cdot \Sigma_{H,S}^e(CA_{i,A(t)})} \\ = \sum_{\tau \in S} m_{H,\tau}^0 \frac{1_{\{\tau\}}(t)}{\sigma_{H,S} \cdot \Sigma_{H,S}^e(CA_{i,A(t)})} + \tilde{\varepsilon}_{i,t} \end{aligned} \quad (1.7)$$

L'équation 1.5 devient alors :

$$\tilde{\rho}_{H,S}[\tau, d_{i,\tau}, (m_t)_{t \in S}] = \rho \left(\frac{d_{i,\tau} - m_{t=\tau}}{MAD(d_{H,S}) \cdot \Sigma_{H,S}^e(CA_{i,A(t)})} \right) \quad (1.8)$$

L'estimation de la dépendance de la dispersion des résidus à la taille de l'entreprise est réalisée par strate et par saison d'enquête (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre). Pour cela, nous ré-ordonnons les résidus selon le chiffre d'affaires des entreprises. La dispersion des résidus est alors estimée sur une fenêtre glissante de 100 observations à l'aide de la médiane de la valeur absolue des écarts (fonction MAD, *Median Absolute Deviation*).

Le logarithme de la dispersion des résidus ainsi estimée est ensuite régressé sur le logarithme du chiffre d'affaires. Pour chaque entreprise i de la strate H et chaque date t de la saison S , avec les $\mu_{i,t}$ indépendants et de moyenne nulle, le modèle retenu s'écrit alors :

$$\log \Sigma_{H,S}^e(CA_{i,A(t)}) = \alpha_{H,S} + \beta_{H,S} \log CA_{i,A(t)} + \mu_{i,t} \quad (1.9)$$

Une dispersion théorique est ainsi estimée en fonction de la taille de l'entreprise (mesurée par le chiffre d'affaires), ceci pour chaque strate H et pour chacune des saisons S (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre) :

$$\log \Sigma_{H,S}^e(CA_{i,A(t)}) = \alpha_{H,S}^e + \beta_{H,S}^e \log CA_{i,A(t)} \quad (1.10)$$

La régression est également réalisée par les M-estimateurs. La fonction de coût retenue est alors celle de Huber (cf. encadré 1).

Après correction de l'hétéroscédasticité et pour la seconde estimation, la procédure de choix de la fonction de coût ρ (décrite dans l'annexe 2) conduit à choisir toujours un M-estimateur de MMR, mais avec une constante plus importante : $c = 0,12$ (pour plus de précisions, cf. annexe 2).

Série trimestrielle des indicateurs de révision d'investissement

Pour chaque strate H , nous disposons de trois séries $(m_{H,t}^e)_{t \in S}$, une pour chaque saison S (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre). Pour être comparables, ces séries sont centrées en zéro et réduites à une variance unitaire.

$$\tilde{m}_{H,t} = \left(m_{H,t}^e - \overline{(m_{H,t}^e)_{t \in S}} \right) / \sqrt{V \left((m_{H,t}^e)_{t \in S} \right)} \quad (1.11)$$

Dans chaque strate H , les trois séries à pas annuel $(\tilde{m}_{H,t})_{t \in S}$ sont alors réunies en une unique série trimestrielle $(REV_{H,t})$: les révisions de l'enquête de janvier N correspondent au premier trimestre de l'année N , celles d'avril au deuxième trimestre, celles d'octobre aux troisième et quatrième trimestres.

$$t = 1^{\text{er}} \text{ trimestre de l'année } N : REV_{H,t} = \tilde{m}_{H,\text{janvier } N}$$

$$t = 2^{\text{e}} \text{ trimestre de l'année } N : REV_{H,t} = \tilde{m}_{H,\text{avril } N} \quad (1.12)$$

$$t = 3^{\text{e}} \text{ trimestre de l'année } N : REV_{H,t} = \tilde{m}_{H,\text{octobre } N}$$

$$t = 4^{\text{e}} \text{ trimestre de l'année } N : REV_{H,t} = \tilde{m}_{H,\text{octobre } N}$$

Les séries trimestrielles des révisions par strate sont ensuite agrégées en une unique série pour l'ensemble de l'industrie manufacturière. Pour cela, nous appliquons des coefficients de redressement constants, calculés comme les montants d'investissement par strate issus de l'EAE dans l'industrie de 2002 (notés $INV_{H,2002}$).

$$REV_t = \frac{\sum_H INV_{H,2002} \cdot REV_{H,t}}{\sum_H INV_{H,2002}} \quad (1.13)$$

Cette série trimestrielle des révisions d'investissement est nommée par la suite « indicateur des révisions ». Elle est notée REV .

LE CHOIX DE LA FONCTION DE COÛT ρ

Le choix de la fonction de coût ρ est le point difficile de la méthode des M-estimateurs. Cette fonction doit être adaptée à la forme générale de la distribution des résidus $\tilde{\varepsilon}_{i,t}$ (cf. annexe 1, équation 1.2). De plus, une fonction de coût dérivable rend plus facile la résolution numérique du problème de minimisation. Dans ce cas, la fonction de coût peut alors être définie par sa dérivée, appelée fonction de score et notée ψ .

Pour des distributions gaussiennes, on choisirait naturellement une fonction quadratique et le centre de la distribution serait la moyenne arithmétique. De manière plus générale, il s'agit de choisir une fonction de coût telle que le M-estimateur associé se comporte « bien » dans la famille des distributions considérées, l'estimateur devant être proche de l'efficacité et robuste. L'efficacité signifie que l'estimateur est sans biais et de variance minimale. La robustesse assure que retirer une observation modifie peu l'estimation (pour une définition précise de la robustesse, cf. par exemple l'ouvrage de Lecoutre et Tassi, 1987).

Le choix des M-estimateurs (c'est-à-dire de la fonction de coût ρ) influe non seulement sur la manière dont les centres des distributions sont mesurés mais aussi sur la définition même de ce que nous considérons être le centre d'une distribution. Autrement dit, nous choisissons la fonction de coût ρ afin que les M-estimateurs $m_{H,t}^e$ estiment bien les $m_{H,t}^0$, mais la fonction de coût définit simultanément les $m_{H,t}^0$ comme les valeurs assurant le minimum du problème limite associé au problème de minimisation 1.4 (cf. annexe 1). Ce problème limite est donné par la formule 2.1 ci-dessous où les $d_{i,t}$ sont indépendants et suivent la loi aléatoire des D_τ^H pour la strate H et la date τ considérées.

$$(m_{H,t}^0)_{t \in S} = \text{Arg min}_{(m_t)_{t \in S}} \sum_{\tau \in S} E \left[\tilde{\rho}_{H,S}[\tau, D_\tau^H, (m_t)_{t \in S}] \right] \quad (2.1)$$

Il existe un grand nombre de M-estimateurs (pour les plus classiques, cf. encadré 1).

Il n'y a pas de choix optimal a priori de la fonction de score. Le choix se fait au regard de la famille des distributions étudiées. Une procédure de choix de la fonction de coût est dite adaptative lorsque la fonction ρ est ainsi adaptée aux distributions empiriques. Cette approche est classée parmi le domaine des statistiques dites « non paramétriques » car il n'est pas nécessaire de faire d'hypothèses sur la loi des perturbations. Bien que depuis l'article de Huber (1964) les M-estimateurs fassent partie intégrante des statistiques usuelles, la fonction de score appropriée est choisie ici selon une méthode adaptative originale qui s'inspire de Ravalet (1996). Cette méthode se décompose en trois étapes :

1. Tout d'abord des statistiques robustes sont calculées sur les distributions empiriques afin de mesurer l'épaisseur des queues des distributions et leur concentration autour de zéro.
2. Dans un second temps, une loi théorique correspondant au mieux, au regard de ces statistiques, aux distri-

butions empiriques est construite.

3. Enfin nous simulons des échantillons à l'aide de la loi théorique ainsi choisie. Les centres des échantillons sont estimés à l'aide de différents M-estimateurs. Nous retenons le M-estimateur pour lequel le moment d'ordre deux des erreurs d'estimation est le plus faible en moyenne.

Nous détaillons ici la procédure adaptative retenue pour choisir la fonction de coût ρ , ceci dans le cas de la première estimation par la méthode des M-estimateurs Quasi-Généralisés. Le choix de la fonction de score pour la seconde estimation – après correction de l'hétéroscédasticité – suit la même procédure et les résultats intermédiaires concernant cette seconde estimation seront donnés plus loin. Enfin, la sensibilité des résultats à la constante de réglage est décrite dans la dernière partie de cette annexe.

Procédure de choix de la fonction de coût ρ

1. Les statistiques utilisées afin de mesurer l'épaisseur des queues et la concentration en zéro doivent être invariantes par translation et par homothétie de la distribution. Les indicateurs utilisés sont inspirés de ceux retenus par Ravalet (1996), mais ils sont modifiés afin d'être robustes à des valeurs extrêmes de très grandes amplitudes.

L'indicateur d'épaisseur de queue τ est inspiré de Hogg (1974). Nous notons $q(p)$ le quantile d'ordre p . Si nécessaire, nous réalisons une interpolation entre les deux valeurs les plus proches. Plus les queues d'une distribution sont épaisses, plus l'indicateur τ est grand. Dans le cas d'une distribution gaussienne, τ vaut 1,3.

$$\tau = \frac{q(95\%) - q(5\%)}{q(90\%) - q(10\%)} \quad (2.2)$$

L'indicateur de concentration autour de la médiane, inspiré de Hogg *et al.* (1988), est noté P_k . Plus P_k est grand, plus la distribution est concentrée autour de la médiane. P_k vaut 1,9 dans le cas d'une loi normale.

$$P_k = \frac{q(90\%) - q(10\%)}{q(75\%) - q(25\%)} \quad (2.3)$$

Les statistiques calculées varient peu selon les strates et les saisons. Nous retenons comme valeurs $\tau = 1,9$ et $P_k = 4,0$.

2. Nous recherchons alors une loi théorique présentant une distribution proche des distributions empiriques considérées. Cette proximité est mesurée à l'aide des statistiques déjà définies : elles doivent être égales à celles des distributions empiriques. Il est évident que ceci ne constitue en rien une proximité absolue comme celle qui pourrait être mesurée à l'aide d'une distance dans l'espace des distributions (par exemple la distance de Kullback). Il sera donc nécessaire de confirmer cette proximité par des représentations des distributions à l'aide de graphiques de comparaison quantiles-quantiles. Nous cherchons donc une loi symétrique présentant des statistiques τ et P_k égales à 1,9 et 4,0. Pour cela,

différentes lois sont envisagées (pour les statistiques correspondantes, cf. tableau A).

La loi de Slash est définie par le rapport entre une loi normale de moyenne nulle et de variance unitaire et une loi uniforme sur $[0, 1]$. La loi double-exponentielle est définie par la différence de deux distributions exponentielles. Enfin, la « loi normale puissance p » désigne ici la loi de la distribution donnée par $\text{signe}(N)|N|^p$ lorsque N suit une loi normale centrée et de variance unitaire.

Au regard de ces résultats, nous cherchons une distribution construite à l'aide la formule 2.4 :

$$X = \alpha D + (1 - \alpha) \text{signe}(N) \frac{|N|^p}{K} \quad (2.4)$$

où D suit une loi double-exponentielle et N une loi normale centrée et de variance unitaire. Avec $\alpha = 0,61$, $K = 0,60$ et $p = 2,70$ (obtenus par un algorithme de Newton), la loi théorique ainsi construite a les mêmes statistiques τ et P_k (1,9 et 4,0) que les lois empiriques des révisions obtenues par strate.

Des graphiques inter-quantiles permettent de vérifier la proximité entre les distributions empiriques des résidus et cette distribution théorique. À titre d'exemple, le graphique A compare la loi théorique à la distribution des résidus pour la strate des grandes entreprises du secteur des biens intermédiaires et pour les révisions lors des enquêtes d'octobre. La courbe étant proche d'une droite, les deux lois peuvent bien être considérées comme similaires (à une transformation affine près).

3. La forme particulière des distributions (queues très épaisses et très forte concentration en 0) oblige à choisir des fonctions de score dites « redescendantes ». Parmi celles envisagées dans l'encadré 1, nous choisissons les M-estimateurs de MRR car ils ont le double avantage d'avoir une seule constante de réglage (notée c) et de ne pas exclure à partir d'un certain seuil les valeurs extrêmes.

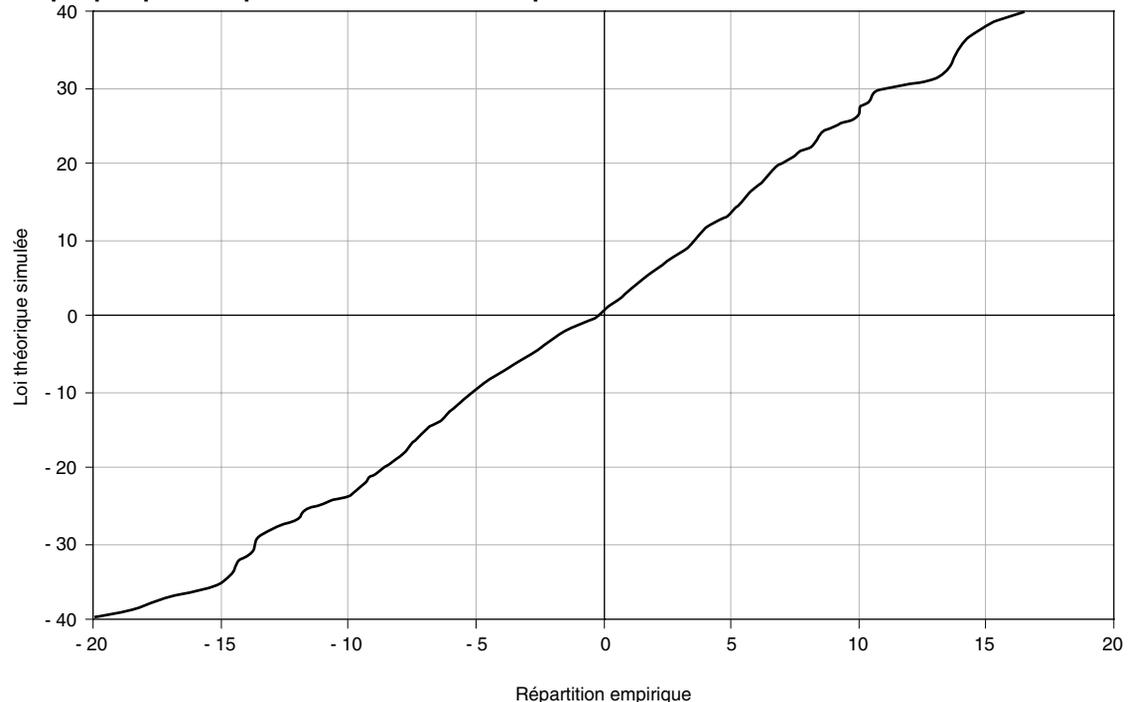
Afin de choisir le paramètre de réglage, nous simulons avec la loi théorique – définie ci-dessus – 1 000 échantillons d'une taille de 200 observations. Pour chaque échantillon, nous estimons le centre de la distribution avec le M-estimateur de MRR et avec la constante de réglage c testée. La racine carrée du moment empirique d'ordre 2 des 1 000 centres estimés est choisie comme critère de manque d'efficacité de l'estimation. Quelques

Tableau A
Indicateur d'épaisseur de queue τ et indicateur de concentration autour de la médiane P_k de quelques lois symétriques

Lois théoriques considérées	τ	P_k
Loi normale	1,3	1,9
Loi de Slash	2,0	2,7
Loi double-exponentielle	1,4	2,3
« Loi normale puissance 2 »	1,6	3,6
« Loi normale puissance 3 »	2,1	6,8
« Loi normale puissance 6 »	4,5	46,6

Source : calculs de l'auteur.

Graphique A
Graphique quantile-quantile entre la loi théorique et les résidus observés



Lecture : cette courbe fait correspondre les fractiles de la répartition empirique à ceux de la loi théorique simulée.

Champ : biens intermédiaires, entreprises de 500 salariés et plus, révisions en octobre. Période : de l'enquête d'octobre 1991 à celle d'octobre 2005.

Source : enquête Investissement, Insee. Calculs de l'auteur.

exemples ($c = 0,05, 0,1, 1$ et 4) sont présentés dans le tableau B. Pour comparaison, ils sont accompagnés de simulations similaires avec des M-estimateurs qui ne sont pas à fonction de score redescendantes : moyenne arithmétique (MCO), médiane et M-estimateur avec la fonction de Huber comme fonction de score.

Le paramètre c retenu au final est celui qui minimise ce critère de manque d'efficacité. Nous retenons ainsi $c = 0,10$ et la racine carrée du moment empirique d'ordre deux de la distribution des erreurs est alors de $0,054$.

Cette méthode présente certaines limites. En particulier, notre définition de la proximité ne repose pas sur une distance (il s'agit en fait d'une semi-distance). Le choix de la fonction théorique est donc relativement arbitraire puisqu'elle dépend des statistiques d'épaisseur de queues et de concentration retenues. Toutefois, lorsque ces statistiques sont construites sur des fenêtres de quantiles différentes (cf. les formules 2.2 et 2.3), cela ne modifie pas significativement la loi théorique choisie.

Choix de la fonction de coût ρ pour la seconde estimation par les MCQG

Le choix de la fonction de coût pour la seconde estimation par les M-estimateurs Quasi-Généralisés suit la même procédure que celle décrite ci-dessus.

Après correction de l'hétéroscédasticité, les paramètres d'épaisseur de queues et de concentration sont désormais proches de $\tau = 1,9$ et $P_k = 3,9$. La loi théorique correspondante est donnée par $\alpha = 0,59$, $K = 0,60$ et $p = 2,61$. Ceci conduit à choisir une constante plus importante pour la seconde estimation : $c = 0,12$.

La correction de l'hétéroscédasticité réduit la concentration en zéro des révisions des grandes entreprises. Ce sont elles qui connaissent en général les plus faibles révisions. Ceci explique la plus faible concentration des distributions après correction de l'hétéroscédasticité. Cette plus faible concentration conduit au choix d'une constante de réglage plus importante.

Sensibilité des résultats à la constante de réglages c

Afin de tester l'éventuelle fragilité des résultats à la constante de réglage c , nous faisons varier cette constante pour les deux étapes de l'estimation et nous comparons ensuite les résultats obtenus. Ceci est réévalué en deux exercices séparés.

- Dans un premier temps, nous faisons varier la constante pour la première étape sans modifier la constante pour la seconde étape ($c = 0,12$). Lorsque la constante de réglage pour la première étape prend les valeurs $0,02, 0,05, 0,10, 0,30$ et $1,00$, les séries calculées sont très proches : elles sont corrélées à plus de 98% . La première estimation ne sert qu'à disposer d'une première mesure des résidus pour la correction de l'hétéroscédasticité : son impact est donc faible sur le résultat final.
- Dans un second temps, la constante de réglage pour la première étape est fixée à sa valeur centrale ($c = 0,10$) et

nous faisons varier la constante de réglage uniquement pour la deuxième étape de l'estimation. Cette constante prend les valeurs : $0,02, 0,05, 0,12, 0,30$ et $1,00$. Il apparaît alors que la constante dans la seconde étape a un impact important sur la série calculée. La série construite avec une constante de $0,02$ et celle construite avec une constante de $1,00$ ne sont corrélées que de 80% . La série retenue ($c = 0,12$) est corrélée de respectivement $85\%, 94\%, 98\%$ et 94% avec celles construites avec $c = 0,02, c = 0,05, c = 0,30$ et $c = 1,00$. La lecture de ces différentes séries (cf. graphique B) met en évidence des mouvements parfois très différents pour les deux séries extrêmes ($c = 0,02$ et $c = 1,00$) par rapport aux autres séries.

Nous pouvons également étudier la qualité prédictive de ces différentes séries en calculant les corrélations entre les séries construites et la série des taux de croissance trimestriels de la FBCF des ENF en valeur (cf. tableau C). Ces corrélations sont calculées sur la période allant du

Tableau B
Résultats des simulations

M-estimateur	Racine carré du moment empirique d'ordre 2 des erreurs
MCO (i.e. la moyenne arithmétique)	0,095
Médiane	0,063
Huber avec $c = 0,1$	0,059
Huber avec $c = 1$	0,086
Huber avec $c = 3$	0,134
MRR avec $c = 0,05$	0,056
MRR avec $c = 0,1$	0,054
MRR avec $c = 1$	0,064
MRR avec $c = 4$	0,080

Source : calculs de l'auteur.

Tableau C
Corrélations des indicateurs de révisions avec les taux de croissance trimestriels de la FBCF des ENF en valeur

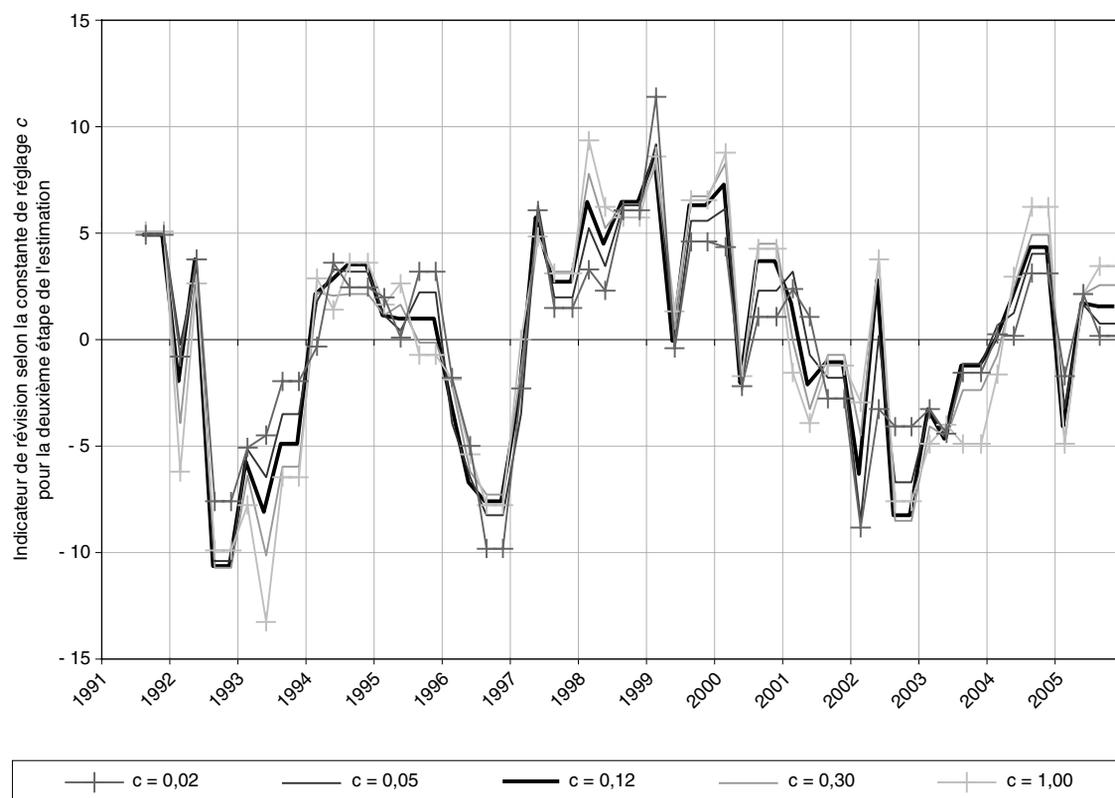
Constante de réglage	Corrélation (en %)
0,02	58,5
0,05	66,4
0,12	68,3
0,30	68,6
1,00	67,4

Lecture : cette corrélation est donnée selon la constante de réglage pour la seconde étape de l'estimation.
Champ : industrie manufacturière (indicateurs de révisions) et entreprises non financières (FBCF). Période : du 3^e trimestre de 1991 au 3^e trimestre de 2005.
Source : enquête Investissement, Insee, (indicateurs de révisions), et comptes trimestriels (FBCF). Calculs de l'auteur.

troisième trimestre de 1991 au troisième trimestre de 2005. La série la mieux corrélée des cinq est celle avec une constante de réglage de 0,30 (corrélation de 68,6 %), mais la série avec une constante de réglage de 0,12 a une corrélation proche (68,3 %). Pour une très faible constante de réglage ($c = 0,02$), il apparaît une perte significative de qualité prédictive (corrélation de 58,5 %).

Il semble donc que la constante de réglage ait bien un impact sur la série construite, mais il n'y a pas de perte importante de qualité prédictive de la série des révisions pour une fenêtre assez large de valeurs de la constante de réglage.

Graphique B
Sensibilité à la constante de réglage c pour la seconde étape de l'estimation



Champ : industrie manufacturière.
 Source : enquête Investissement, Insee. Calculs de l'auteur.

ESTIMATION DU MODÈLE VAR POUR LA PRÉVISION DE L'INVESTISSEMENT DES ENTREPRISES

On vérifie par des tests de racine unité la stationnarité des trois séries \dot{FBCF} , REV et dIC : les tests *ADF* (Augmented Dickey-Fuller) rejettent très significativement leur non-stationnarité. Nous estimons alors sous la forme d'un processus *VAR* le vecteur X_t de ces trois séries. En notant L l'opérateur retard, X_t est modélisé par l'équation 3.2, où Φ est un polynôme, c_0 un vecteur constant et v_t est un bruit blanc gaussien. L'absence d'autocorrélation des perturbations v_t et leur normalité seront testées *a posteriori*.

Équation 3.1

$$X_t = \begin{pmatrix} \dot{FBCF}_t \\ REV_t \\ dIC_t \end{pmatrix}$$

Équation 3.2

$$X_t = c_0 + \Phi(L)X_{t-1} + v_t$$

Le processus est estimé sur la période allant du premier trimestre de 1992 au dernier trimestre de 2003. La méthode itérative des tests du rapport de maximum de vraisemblance comme la minimisation du critère Akaike (*AIC*) proposent de ne retenir qu'un seul retard. Les trois équations de la dynamique sont alors estimées par les MCO, les taux de croissance trimestriels de la *FBCF* étant exprimés en pourcentage. Les statistiques de Student sont données entre parenthèses.

Équation 3.3

$$\dot{FBCF}_t = 0,49 + 0,19 \cdot \dot{FBCF}_{t-1} + 0,14 \cdot REV_{t-1} + 0,09 \cdot dIC_{t-1}$$

(2,4) (1,2) (3,0) (2,4)

$$R^2 = 0,51 \quad R^2_{ajusté} = 0,48 \quad RMSE = 1,23 \quad DW = 1,79$$

Équation 3.4

$$REV_t = -1,06 + 1,25 \cdot \dot{FBCF}_{t-1} + 0,23 \cdot REV_{t-1} + 0,23 \cdot dIC_{t-1}$$

(-1,7) (2,6) (1,5) (2,0)

$$R^2 = 0,49 \quad R^2_{ajusté} = 0,46 \quad RMSE = 3,81 \quad DW = 1,92$$

Équation 3.5

$$dIC_t = 0,21 - 0,09 \cdot \dot{FBCF}_{t-1} + 0,07 \cdot REV_{t-1} + 0,47 \cdot dIC_{t-1}$$

(0,3) (-0,2) (0,4) (3,4)

$$R^2 = 0,23 \quad R^2_{ajusté} = 0,18 \quad RMSE = 4,59 \quad DW = 1,98$$

Nous vérifions *a posteriori* que les résidus suivent bien un bruit blanc : on réalise un test du portmanteau multivarié. L'hypothèse nulle est l'absence d'autocorrélation des résidus. La statistique de test vaut 119,6. Elle est à comparer à un χ^2 de 135 degrés de liberté, soit une *p-value* de 0,82. Il n'est donc pas possible de rejeter l'absence d'autocorrélation des résidus.

Nous vérifions également *a posteriori* que les résidus sont bien distribués selon une loi gaussienne multivariée : appliquée aux résidus, la statistique de Skewness vaut 0,86 et celle de Kurtosis 2,80. La statistique jointe pour le test de Doornick et Hansen (1994) vaut 3,66, soit une *P-value* de 0,30 pour l'hypothèse de normalité des résidus.

La forme bloc récursive est donnée par les équations 3.4 et 3.5 et par l'équation 3.6 ci-dessous, utilisée pour les prévisions concomitantes.

Équation 3.6

$$\dot{FBCF}_t = 0,64 + 0,01 \cdot \dot{FBCF}_{t-1} + 0,14 \cdot REV_t + 0,11 \cdot REV_{t-1} + 0,03 \cdot dIC_t + 0,04 \cdot dIC_{t-1}$$

(3,4) (0,1) (3,3) (2,4) (0,9) (1,0)

$$R^2 = 0,62 \quad R^2_{ajusté} = 0,58 \quad RMSE = 1,11 \quad DW = 1,72$$

ESTIMATION DU MODÈLE VAR POUR LA PRÉVISION DE LA PRODUCTION MANUFACTURIÈRE

On vérifie par des tests de racine unité la stationnarité des trois séries \dot{PROD} , REV et $TPPRE$: les tests *ADF* (Augmented Dickey-Fuller) rejettent très significativement leur non-stationnarité. Nous estimons alors sous la forme d'un processus VAR le vecteur Y_t de ces trois séries.

Équation 4.1

$$Y_t = \begin{pmatrix} \dot{PROD}_t \\ REV_t \\ TPPRE_t \end{pmatrix}$$

Ce processus est estimé sur la période allant du premier trimestre de 1992 au dernier trimestre de 2003. La méthode itérative des tests du rapport de maximum de vraisemblance comme la minimisation du critère Akaike (*AIC*) proposent de ne retenir qu'un seul retard. Les trois équations de la dynamique sont alors estimées par les MCO. Les taux de croissance trimestriels de la production manufacturière \dot{PROD} sont exprimés en pourcentage. Les statistiques de Student sont données entre parenthèses.

Équation 4.2

$$\dot{PROD}_t = 0,63 + 0,20 \cdot \dot{PROD}_{t-1} + 0,11 \cdot REV_{t-1} - 0,01 \cdot TPPRE_{t-1}$$

(2,8) (1,1) (2,2) (-0,2)

$$R^2 = 0,26 \quad R^2_{ajusté} = 0,21 \quad RMSE = 1,27 \quad DW = 1,89$$

Équation 4.3

$$REV_t = -1,54 + 1,20 \cdot \dot{PROD}_{t-1} + 0,27 \cdot REV_{t-1} + 0,07 \cdot TPPRE_{t-1}$$

(-2,2) (2,1) (1,8) (1,1)

$$R^2 = 0,46 \quad R^2_{ajusté} = 0,42 \quad RMSE = 3,94 \quad DW = 2,06$$

Équation 4.4

$$TPPRE_t = 0,76 + 1,49 \cdot \dot{PROD}_{t-1} + 0,25 \cdot REV_{t-1} + 0,68 \cdot TPPRE_{t-1}$$

(0,7) (1,6) (1,0) (6,5)

$$R^2 = 0,76 \quad R^2_{ajusté} = 0,75 \quad RMSE = 6,47 \quad DW = 1,74$$

Nous vérifions *a posteriori* que les résidus suivent bien un bruit blanc : on réalise un test du portmanteau multivarié. L'hypothèse nulle est l'absence d'autocorrélation des résidus. La statistique de test vaut 148,8. Elle est à comparer à un χ^2 de 135 degrés de liberté, soit une *P-value* de 0,20. Il n'est donc pas possible de rejeter l'absence d'autocorrélation des résidus.

Nous vérifions également *a posteriori* que les résidus sont bien distribués selon une loi gaussienne multivariée : appliquée aux résidus, la statistique de Skewness vaut 1,74 et celle de Kurtosis 0,65. La statistique jointe pour le test de Doornik et Hansen (1994) vaut 2,39, soit une *P-value* de 0,50 pour l'hypothèse de normalité des résidus.

La forme bloc récursive est donnée par les équations 4.3 et 4.4 et par l'équation 4.5 ci-dessous, utilisée pour les prévisions concomitantes :

Équation 4.5

$$\dot{PROD}_t = 0,68 - 0,11 \cdot \dot{PROD}_{t-1} + 0,10 \cdot REV_t + 0,05 \cdot REV_{t-1} + 0,13 \cdot TPPRE_t - 0,10 \cdot TPPRE_{t-1}$$

(4,6) (-0,9) (3,2) (1,5) (6,7) (-5,5)

$$R^2 = 0,73 \quad R^2_{ajusté} = 0,70 \quad RMSE = 0,78 \quad DW = 1,93$$