Direction des Études et Synthèses Économiques

G 2005 / 09

# Prévoir l'investissement des entreprises

## Un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie

Nicolas FERRARI

Document de travail

INSEE

Institut National de la Statistique et des Études Économiques

**INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES**

*Série des documents de travail*
*de la Direction des Études et Synthèses Économiques*

**G 2005 / 09**

# Prévoir l'investissement des entreprises

## Un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie

Nicolas **FERRARI***

JUILLET 2005

L'auteur remercie Olivier BIAU, Karine BERGER, Matthieu CORNEC, Michel DEVILLIERS, Hélène ERKEL-ROUSSE, Philippe SCHERRER et Fabien TOUTLEMONDE pour leurs conseils et leurs relectures attentives.

---

* Département de la Conjoncture - Division « Enquêtes de Conjoncture » Timbre G120 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF Cedex

Département des Études Économiques d'Ensemble - Timbre G201 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX - France - Tél. : 33 (1) 41 17 60 68 - Fax : 33 (1) 41 17 60 45 - CEDEX - E-mail : d3e-dg@insee.fr - Site Web INSEE : http://www.insee.fr

# Prévoir l'investissement des entreprises

## Un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie


### Résumé

L'enquête trimestrielle sur les investissements dans l'industrie de l'INSEE est une source d'information majeure concernant les évolutions conjoncturelles de l'investissement productif. Toutefois, la nature annuelle des questions posées rend délicate son utilisation pour des prévisions selon un rythme trimestriel. Cet article propose un indicateur trimestriel des révisions d'anticipations d'investissement des industriels. Cet indicateur mesure les adaptations au cours de l'année des investissements en fonction des évolutions conjoncturelles. Il est très bien corrélé aux évolutions trimestrielles de l'investissement des entreprises mesurées par la comptabilité nationale. De plus, il est disponible environ trois mois avant la publication des premiers résultats des comptes trimestriels.

Les distributions étudiées ne vérifiant pas l'hypothèse classique de normalité (queues épaisses et fortes concentrations en zéro), il est nécessaire de mettre en oeuvre une méthode d'estimation robuste aux révisions extrêmes. En prenant également en compte la présence d'hétéroscédasticité, il a été choisi d'utiliser la méthode dite des "M-estimateurs Quasi-Généralisés".



**Mots-clés** : Investissement productif, prévisions conjoncturelles, enquêtes de conjoncture, valeurs extrêmes, procédure adaptative, régression par les M-estimateurs, méthode des M- estimateurs Quasi-Generalisés.


---

# Firm'investment forecast:

## An indicator of changes in expectations in industrial investment survey


### Abstract

The quarterly industrial investment survey constitutes one of the main sources of information for the short-term economic analysis of industrial firms' investment. However, its main questions are annual. Therefore, the use of this survey's results for the forecasting of investment on a quarterly basis requires some specific statistical treatment. This paper presents a quarterly indicator based on the changes in industrial entrepreneurs' expectations as regards annual investment. This indicator derives from the estimation of the successive adaptations of entrepreneurs' investment plans as times goes by, depending on the evolutions of short-term macroeconomic activity ; it proves to be strongly correlated with the fluctuations of the entrepreneurs' investment growth rate (as is measured in the French Quarterly Accounts). Moreover, the indicator is available about three months ahead with respect to the first results' release of the quarterly national accounts.

The probability distributions of changes in expectations are not gaussian (due to heavy tails and strong concentrations near zero). Consequently, robust estimation methods for extreme observations were performed. Due to the presence of heteroskedasticity, we choosed to apply the "Quasi-Generalized M-estimator» method.



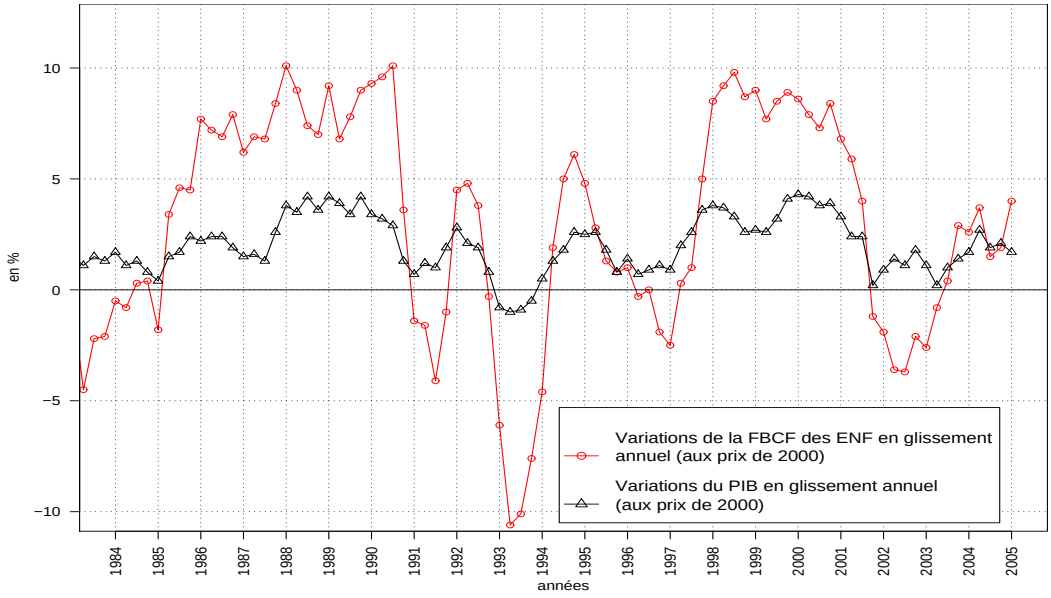
**Keywords**: Firms' investment, short-term forecasting, business tendency surveys, extreme values, adaptive M-regression, Quasi-Generalized M-estimator

**Classification JEL** : C14, C16, C42, C53, E22

# 1 Introduction

L'investissement des entreprises constitue une variable très importante de l'analyse conjoncturelle. S'il ne représente qu'une part assez modeste du PIB, entre 10% et 12% selon les années[2], il sur-réagit aux variations de l'activité (cf. figure 1). De ce fait, il contribue particulièrement aux variations du PIB. Sur période longue (de 1980 à 2003), la FBCF des ENF[3] a contribué pour 32% aux variations du PIB en glissement annuel[4]. En plus d'influer à court terme sur la demande, l'investissement permet le développement des moyens productifs des entreprises. Les efforts d'investissement engagent l'avenir avec des conséquences à moyen terme sur l'offre des entreprises.

FIG. 1 – Evolutions comparées du PIB et de l'investissement des entreprises




Source : Insee, comptes nationaux trimestriels (base 2000).

[2]Depuis le début des années 1990.

[3]Formation Brute de Capital Fixe des Entreprises Non Financières.

[4]Les comptes trimestriels utilisés dans cet article sont les premiers résultats du premier trimestre de 2005 en base 2000.

Les indicateurs conjoncturels concernant l'investissement sont rares[5]. Du fait de la forte hétérogénéité des comportements individuels, sa prévision s'avère délicate. Ainsi, l'enquête sur les investissements dans l'industrie (nommé ici *enquête investissement*), menée trimestriellement par l'Insee, est une des rares sources conjoncturelles relatives aux dépenses d'équipement des entreprises. Elle permet de prévoir l'investissement industriel de manière 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, la nature annuelle des questions rend relativement délicate son utilisation pour des prévisions trimestrielles. Les révisions entre deux enquêtes successives apportent une information pertinente, mais les taux de croissance agrégés ne se prêtent pas aisément à la construction d'un indicateur trimestriel du fait de leur caractère annuel.

En revanche, en étudiant les révisions individuelles des montants d'investissement anticipés, il apparaît que l'enquête apporte, trimestre après trimestre, une information très intéressante. Plus précisément, il est possible de construire un indicateur trimestriel des révisions d'anticipations individuelles des montants d'investissement. Cette série s'avère bien corrélée avec l'évolution trimestrielle de la FBCF des ENF disponible ultérieurement.

Cet indicateur capte les changements infra-annuels dans les projets d'investissement des industriels : avant le début d'année, les entrepreneurs prévoient le niveau et le rythme de leurs investissements en fonction de leurs projets de développement interne. Les évolutions conjoncturelles de l'année à venir sont alors encore très incertaines. Au cours de l'année, le rythme des investissements est affiné et adapté aux aléas conjoncturels. Les évolutions des anticipations sont alors reliées au dynamisme des investissements durant le restant de l'année courante. L'indicateur des révisions se révèle ainsi être bien corrélé aux variations trimestrielles de la FBCF en valeur des ENF. Il apparaît ainsi comme un très bon indicateur pour prévoir les variations de

[5] A ce sujet, le lecteur pourra se reporter au dossier dans la Note de conjoncture de l'Insee de mars 2005 [2].

cette variable.

Les révisions individuelles d'anticipation d'investissement ne suivent pas une loi aléatoire gaussienne. Les queues de distributions, particulièrement épaisses, nécessitent des méthodes d'estimations robustes aux valeurs aberrantes. Les révisions sont également très concentrées en zéro. Pour répondre à ces difficultés statistiques, les estimateurs de centre de distribution sont choisis dans la classe des M-estimateurs.

Les révisions présentent aussi une forte hétéroscédasticité : la variance des révisions rapportées aux chiffres d'affaires diminue significativement avec la taille des entreprises. Pour améliorer l'efficacité des estimateurs, une méthode en deux temps est mise en œuvre : elle s'apparente à la méthode des MCQG (Moindres Carrés Quasi-Généralisés) et elle est dite méthode des "M-estimateurs Quasi-Généralisés".

L'indicateur a déjà été présenté dans la note de conjoncture de l'Insee de mars 2005 avec le dossier "Prévoir l'investissement des entreprises ? Un indicateur des révisions d'anticipations dans l'enquête Investissement dans l'industrie" [2]. Cet article a pour objectif de détailler davantage la construction de l'indicateur. En particulier, il insiste sur la procédure adaptative de choix du M-estimateur et sur la méthode de correction de l'hétéroscédasticité.

Dans un premier temps (partie 2), nous présentons succinctement 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 en revanche difficilement utilisables pour prévoir les évolutions de l'investissement à un rythme trimestriel. En revanche, les anticipations étant révisées tous les trimestres, ces révisions apparaissent assez naturellement être une information pertinente à un rythme trimestriel. La construction de l'indicateur des révisions est ensuite exposée (partie 3). Une attention particulière est portée au choix de la fonction de score des M-estimateurs ainsi qu'à la correction de l'hétéroscédasticité. La partie 4 commente la série construite et en propose un exemple d'utilisation pour la prévision de l'investissement des entreprises à l'aide d'un modèle VAR. Enfin, la partie 5 conclut.

# 2 Présentation de l'*enquête investissement* et l'utilisation de ses résultats

## 2.1 Présentation 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[6] 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 (différences pondérées entre le nombre de réponses "en hausse" et celui des réponses "en baisse").

Ces deux types de questions sont présents à chaque enquête. De plus, d'autres questions sont ajoutées selon le trimestre de l'enquête afin d'affiner la perception de l'investissement. Un premier type de questions (enquêtes d'avril et d'octobre) permet de distinguer les investissements par justification économique[7]. D'autres questions concernent les facteurs économiques qui influencent les dépenses d'équipement (enquêtes d'octobre)[8]. D'autres encore sont relatives aux capacités de productions et aux déclassements d'équipe-

[6] Depuis juillet 2003 uniquement.

[7] On distingue 5 catégories d'investissements : les investissements destinés au renouvellement d'équipements usagés, à l'entretien et à la maintenance, ceux destinés à la modernisation et à la rationalisation, ceux qui concernent l'extension des capacités de production sur les produits existants, ceux qui concernent l'introduction de nouveaux produits et enfin les autres destinations.

[8] Les entrepreneurs émettent une opinion à propos du caractère stimulant ou limitant de huit facteurs : la demande intérieure, la demande étrangère, les perspectives de profit, l'autofinancement, l'endettement, les taux d'intérêt, les conditions globales de financement, les facteurs techniques et les autres facteurs.

ment (enquêtes d'avril). Naboulet-Raspiller [9] montrent que ces questions annexes permettent d'améliorer sensiblement la compréhension des comportements d'investissement des entreprises industrielles.

Le présent article se rapporte aux seules réponses aux questions quantitatives annuelles. Pour avoir davantage de renseignements sur l'*enquête investissement*, on pourra se reporter à sa fiche méthodologique [1] disponible sur le site Internet de l'Insee et à l'ouvrage de la série Insee-Méthode relatif à cette enquête (Rosenwald [11]).

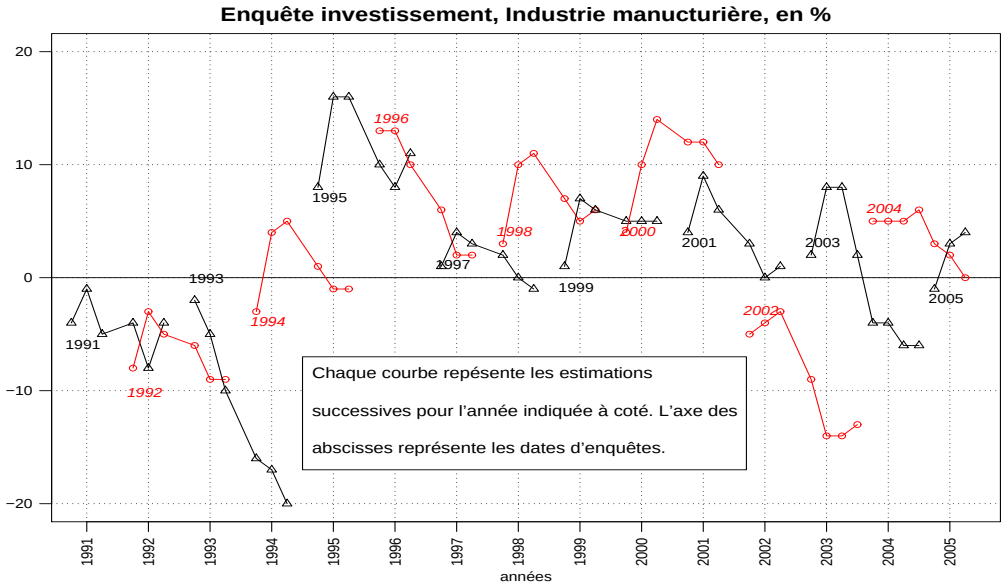
## 2.2 Les révisions des entrepreneurs interrogés à l'*enquête investissement* apportent une information conjoncturelle intéressante

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. figure 2). La méthode utilisée pour l'agrégation est une méthode particulière, dite "méthode des Grands Investisseurs". Cette méthode a été exposée et discutée dans Ravalet [10]. Il s'agit d'une méthode apparentée à la méthode dite "du ratio"¨ Cette méthode est légèrement modifiée : certaines entreprises sont sous-pondérées lorsque leurs réponses sont atypiques ou trop influentes.

Graphiquement, les taux estimés reflètent bien le dynamisme des dépenses d'équipement. Par exemple, les années 1998 à 2000 apparaissent comme des années où les efforts d'investissement des entreprises étaient particulièrement forts. Cette période a été suivie dès 2001 d'une phase de ralentissement économique et de rationalisation des projets de développement des entreprises industrielles.

La lecture de la figure 2 fait apparaître des révisions importantes des estimations successives du taux de croissance des investissements industriels pour une même année. Au cours des estimations successives, les résultats de l'enquête convergent vers les évolutions réelles des investissements industriels mesurés par la comptabilité nationale annuelle. Cette convergence est un critère important pour juger de la qualité de l'enquête : les évolutions des

FIG. 2 – Estimations successives des industriels de l'évolution de leurs investissements




Source : Insee, *enquête investissement.*

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 $N+1$ et environ un an avant la publication des comptes nationaux annuels dits "semi-définitifs", l'enquête permet de juger avec précision des évolutions passées de la FBCF du secteur industriel. Toutefois, bien que cette information soit précoce par rapport à la comptabilité nationale, elle reste bien trop tardive dans la perspective d'une utilisation pour la prévision conjoncturelle. En revanche, les estimations précédentes (à partir d'octobre $N-1$) sont publiées assez tôt pour être utilisées dans l'exercice de prévision conjoncturelle de l'année $N$.

La manière la plus naturelle d'utiliser cette information avancée est de supposer que les prévisions agrégées des industriels correspondent aux évolutions qui seront effectivement réalisées. Il apparaît qu'en moyenne les taux de crois-

TAB. 1 – Les révisions moyennes des estimations des industriels sur la croissance annuelle de leurs investissement, période 1990-2003, industrie manufacturière

| Dates des enquêtes par rapport à l'année $N$ considérée | Taux d'investissement moyen estimé par l'enquête | Différence moyenne par rapport aux estimations de l'enquête d'avril $N+1$ |
|---|---|---|
| Octobre $N-1$ | 1,1% | 1,5 point |
| Janvier $N$ | 5,8% | 6,2 points |
| Avril $N$ | 5,4% | 5,8 points |
| Octobre $N$ | 1,9% | 2,3 points |
| Janvier $N+1$ | -0,6% | -0,2 point |
| Avril $N+1$ | -0,4% | 0,0 point |

Source : *Enquête investissement.* Calculs de l'auteur.
*Lecture : Par exemple, de 1990 à 2003, lors des enquêtes de janvier (année N), les industriels ont estimés - en moyenne sur cette période - que leurs investissements augmenteraient de 5,8% 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 6,2 points entre l'enquête de janvier courante et l'enquête d'avril de l'année suivante.*

sance, prévus puis réalisés, sont révisés selon un profil relativement stable (cf. figure 2). 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 - 2003, les premières estimations des industriels se révèlent être sur-estimées par rapport à la dernière estimation de l'enquête en avril de l'année suivante. En moyenne sur cette période, cette sur-estimation est de 1,5 point pour la première estimation lors de l'enquêtes d'octobre de l'année précédente, de 6,2 points, 5,8 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 -0,2 point en moyenne (cf. tableau 1). Il est donc 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 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%. La simple 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. figure 2). 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.

# 3 Un indicateur trimestriel des révisions des montants annuels

## 3.1 Les révisions individuelles d'anticipation

La comparaison des révisions des taux de croissance annuels agrégés de l'investissement se révèle intéressante, mais 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

utilisée (dite des "Grands Investisseurs") permet d'estimer efficacement les prévisions de taux de croissance annuels, elle est peu adaptée à la mesure spécifique 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 références qu'aux années d'intérêt. Par exemple, lors de l'enquête d'avril 2005, une entreprise révise 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.

Les révisions des anticipations des industriels se révèlent plus informatives, ceci en considérant directement les révisions entre deux enquêtes successives des montants d'investissement et non pas les révisions des estimations d'évolution annuelle.

A chaque enquête, on calcule la révision moyenne d'investissement rapportée au chiffre d'affaires de l'entreprise. 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 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 $N-1$.

A chacune des enquêtes de l'année $N$ (janvier, avril, juillet et octobre), on peut 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, on calcule de la même manière les révisions pour les années 2003 et 2004. 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

TAB. 2 – Calendrier théorique de calcul de l'indicateur

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

tôt lors des enquêtes d'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 $N$ et celle de janvier $N+1$, il s'agit du chiffre d'affaires de l'année $N-1$ et celui-ci est noté $CA_{i,t}$ pour l'entreprise $i$[9]. En notant $I^a_{i,t}$ et $I^a_{i,t-1}$ 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 $d_{i,t}$ est donnée par la formule (1).

$$d_{i,t} = \frac{I^a_{i,t} - I^a_{i,t-1}}{CA_{i,t}} \tag{1}$$

## 3.2 Stratification

On agrège ces indicateurs des révisions individuelles de manière à calculer pour chaque date d'enquête un indicateur de position $m_t$ de la distribution

[9]Autrement dit, $CA_{i,t}$ est constant entre les enquêtes d'avril $N$ et de janvier $N+1$.

Tab. 3 – Calendrier provisoire de calcul de l'indicateur

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

des révisions individuelles $d_{i,t}$. L'échantillon est un échantillon stratifié par secteur et par taille d'entreprise. Il est donc naturel de calculer tout d'abord ces paramètres par strate puis de les agréger à l'aide de coefficients de redressement. Les coefficients de redressement choisis sont les montants annuels d'investissement calculés à partir de l'Enquête Annuelle d'Entreprise (EAE) dans l'industrie de 2002. 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). Jusqu'en octobre 2003[10], les secteurs de l'énergie et des industries agroalimentaires étaient relativement mal couverts par l'enquête. De ce fait, ils sont exclus de l'estimation. L'indicateur ne porte donc que sur le secteur de l'industrie manufacturière, soit quatre secteurs NES 16 : industrie des biens de consommation, industrie automobile, industrie des biens d'équipement et industrie des biens intermédiaires. Enfin, le secteur de l'automobile, trop petit et trop concentré pour être divisé en trois tranches de taille avec la méthode d'agrégation utilisée,

[10]L'*enquête investissement* est devenue obligatoire en 2004. Ceci a permis de relever très significativement les taux de réponses, en particulier dans les secteurs des industries agroalimentaires et de l'énergie.

est regroupé en une seule strate, si bien que le calcul est finalement réalisé sur dix strates.

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

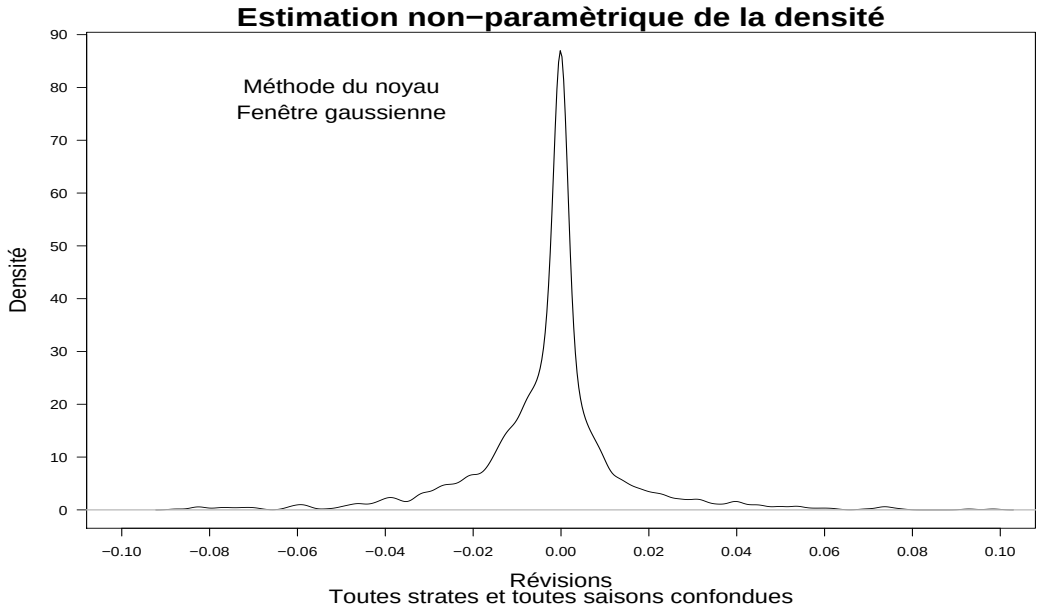
Au sein de chaque strate, le calcul de l'indicateur agrégé des révisions pose des difficultés techniques. Les distributions des indicateurs des révisions individuelles sont très étendues. En effet, certaines entreprises peuvent, d'une enquête à l'autre, fortement réviser leurs montants d'investissement. 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. Autre exemple, lors d'une enquête d'octobre, une très grande entreprise déclare sa prévision d'investissement pour l'année suivante. Cette anticipation découle des décisions stratégiques de l'entreprise. La prévision n'est pas modifiée dans les réponses successives à l'enquête jusqu'à ce que les comptes de l'entreprise soient publiés, c'est-à-dire quelques mois après la fin de l'année considérée. Une révision très importante a alors lieu, décalage entre ce qui était prévu avant le début de l'année et ce qui est constaté dans les comptes un an et demi après.

Au niveau statistique, ces fortes révisions se concrétisent par des répartitions à queues très épaisses (cf. figure 3). Il est alors nécessaire de mener une agrégation à l'aide d'estimateurs robustes et non pas à l'aide d'une moyenne arithmétique classique.

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. figure 3). En effet, de nombreuses entreprises ne révisent pas leurs anticipations d'investissement entre deux enquêtes : la médiane est presque systématiquement nulle.

La méthode choisie est celle des M-estimateurs, due à Huber [6]. Cette méthode est une généralisation de la méthode dite des MCO (Moindres Carrés Ordinaires). Pour estimer la position du centre de la distribution, au lieu de minimiser la somme des carrés des résidus comme dans le cas des MCO, on

FIG. 3 – Répartition empirique des révisions individuelles d'investissement




Source : Insee, *enquête investissement*. Calculs de l'auteur.

minimise la somme d'une autre fonction objectif appliquée aux résidus. Cette fonction est notée $\rho$.

Plus précisément, pour chaque saison d'enquête $S$ (enquêtes de janvier, enquêtes d'avril et enquêtes d'octobre) et pour chaque strate $H$, on effectue une régression robuste des indicateurs des révisions sur des indicatrices temporelles. Pour toutes les enquêtes $t$ de la saison $S$, on calcule ainsi les $m_{H,t}$, indicateurs agrégés des révisions de la strate $H$. On note $\mathbf{1}_{\{x\}}$ 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 :

$$d_{i,t} = \sum_{\tau} m_\tau \mathbf{1}_{\{\tau\}}(t) + \epsilon_{i,t} \tag{2}$$

Les perturbations $\epsilon_{i,t}$ suivent des lois aléatoires que nous supposons indépendantes. Dans un premier temps, nous supposons également que les perturba-

tions $\epsilon_{i,t}$ sont identiquement distribuées (par strate et par saison).

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

$$\frac{d_{i,t}}{\sigma_{H,S}} = \sum_{\tau} m_\tau \frac{\mathbf{1}_{\{\tau\}}(t)}{\sigma_{H,S}} + \tilde{\epsilon}_{i,t} \tag{3}$$

Plusieurs statistiques sont possibles pour les paramètres d'échelle $\sigma_{H,S}$. Celle retenue est la médiane de la valeur absolue de la médiane. Cette statistique, notée MAD (Median Absolute Deviation), est donnée par la formule (4). Cette mesure a l'avantage d'être robuste aux valeurs extrêmes. Le coefficient 1,48 permet à la statistique MAD d'être égale à l'écart-type dans le cas d'une loi normale.

$$\sigma\left(d_{H,S}\right) = MAD\left(d_{H,S}\right) = 1{,}48 \text{ Médiane}\left[d_{H,S} - \text{Médiane}\left(d_{H,S}\right)\right] \tag{4}$$

On note $n_{H,S}$ le nombre d'observations dans la strate $S$ pour les enquêtes de la saison $S$. Le problème de minimisation associé à l'estimation pour la strate $H$ et la saison $S$ s'écrit alors :

$$\left(m_{H,t}\right)_{t\in S} = \text{Arg} \min_{(m_t)_{t\in S}} \frac{1}{n_{H,S}} \sum_{\substack{i\in H \\ t\in S}} \tilde{\rho}_{H,S}\left[t, d_{i,t}, (m_t)_{t\in S}\right] \tag{5}$$

$$\text{avec} \qquad \tilde{\rho}_{H,S}\left[t, d_{i,t}, (m_t)_{t\in S}\right] = \rho\left(\frac{d_{i,t} - m_t}{MAD\left(d_{H,S}\right)}\right) \tag{6}$$

Sous certaines conditions théoriques qui sont supposées vérifiées ici, les $\left(m_{H,t}\right)_{t\in S}$ convergent vers une limite finie $\left(m^0_{H,t}\right)_{t\in S}$ lorsque le nombre d'observations $n_{H,S}$ tend vers l'infini. Le cadre théorique est défini précisément dans l'annexe A.

## 3.4 Le choix de la fonction objectif

Le choix de la fonction objectif $\rho$ est le point délicat de la méthode des M-estimateurs. Cette fonction doit être adaptée à la forme générale de la distribution des résidus. De plus, une fonction objectif dérivable rend plus facile la résolution numérique du problème de minimisation. Lorsqu'elle est

dérivable, la fonction objectif peut alors être définie par sa dérivée, appelée fonction de score et notée $\psi$.

Pour des distributions gaussiennes, on choisirait naturellement une fonction quadratique et l'estimateur de position de la distribution serait la moyenne arithmétique. De manière plus générale, il s'agit de choisir une fonction objectif telle que le M-estimateur associé se comporte bien dans la famille des distributions considérées. Plus précisément, l'estimateur doit être robuste et proche de l'efficacité. 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, on pourra se reporter par exemple à l'ouvrage de Lecoutre et Tassi [7].

## Encadré 2 : Choisir une fonction objectif $\rho$

La fonction objectif $\rho$ doit être choisie afin que les estimations soient à la fois robustes et proches de l'efficacité sur les distributions étudiées. La littérature statistique propose un grand nombre de familles de fonctions. Nous en présentons ici quelques exemples. Les premières (MCO, médiane) sont exposées à titre pédagogique mais 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 objectif $\rho$ et/ou par sa dérivée, c'est-à-dire par la fonction de score $\psi$.

**Les MCO (Moindres Carrés Ordinaires) :**

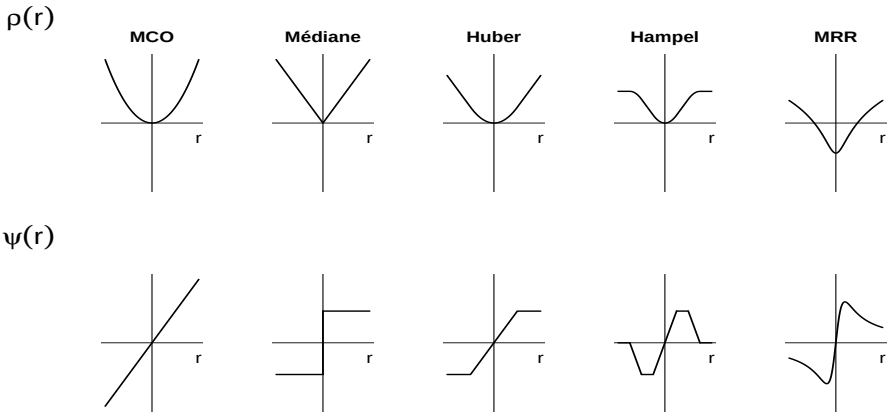
$$\rho(x) = \frac{1}{2}x^2 \quad \text{et} \quad \psi(x) = x$$

L'estimateur des MCO est un cas particulier des M-estimateurs. L'estimateur des MCO, extrêmement usité, a de nombreuses propriétés. En particulier, sous l'hypothèse de normalité des distributions, il est l'estimateur optimal au sens de l'efficacité. (Un estimateur est dit efficace lorsqu'il est sans biais et que sa variance est minimale.) En

revanche, le caractère quadratique de la fonction objectif le rend très sensible aux valeurs extrêmes.

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

FIG. A - Exemples de fonctions objectif $\rho$ et de fonctions de score $\psi$ associées




**La médiane :**

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

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

La médiane est un estimateur également très classique. Il 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, ce qui n'est absolument pas informatif.

**Huber :**

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

Le M-estimateur proposé par Huber [6] est équivalent aux MCO sur le centre de la distribution et équivalent à la médiane sur les queues de distribution. De plus, 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 forte concentration en zéro), il se révèle être trop peu robuste. On lui préfère 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.

**Plusieurs fonctions "redescendantes" :**

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

Sinus d'Andrew : $\psi(x) = \frac{1}{\pi}\sin\left(\frac{\pi x}{c}\right)$ si $|x| < c$, 0 sinon

Hampel :
$$\begin{cases} \psi(x) = x & \text{si} \quad |x| < a \\ \psi(x) = a \ \text{signe}(x) & \text{si} \quad a < |x| < b \\ \psi(x) = a\frac{\text{signe}(x)\ c - x}{c - b} & \text{si} \quad b < |x| < c \\ \psi(x) = 0 & \text{si} \quad c < |x| \end{cases}$$

Dans les trois familles de fonctions de score données par les formules ci-dessus, toutes sont à valeurs "redescendantes". Plus précisément elles s'annulent à partir d'un certain seuil. Ceci signifie que les points trop éloignés du centre de la distribution sont complètement rejetés de l'estimation. Toutes sont proches de la proportionnalité en 0. Les estimateurs s'apparentent donc aux MCO pour le centre de la distribution. L'utilisation des fonctions Biweight est due à Tukey. Celle des fonctions Sinus à Andrew. Enfin, la fonction de score affine par morceau est due à Hampel.

**Les fonctions MRR (famille de fonctions de score retenue) :**

$$\rho(x) = \frac{c}{2} \log \left[x^2 + c\right] \quad \text{et} \quad \psi(x) = \frac{cx}{x^2 + c}$$

Pour des distributions telles qu'observées ici, Moberg, Ramberg et Randles [8] proposent de choisir une fonction de score dans la famille définie ci-dessus. Nous donnons à cette famille de fonctions 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 pour des distributions très concentrées et à queues très épaisses. Ce sont ces estimateurs qui seront retenus par la suite.

La théorie du maximum de vraisemblance indique que, si $f$ est la densité de la vraie loi des perturbations, la fonction objectif optimale - au sens de l'efficacité définie *supra* - est donnée par $\rho = -\log f'/f$. Toutefois, par souci de robustesse des résultats et en acceptant en contre-partie une perte d'efficacité, il est possible de choisir une fonction objectif qui croisse moins vite avec l'amplitude des résidus.

Il faut noter que le choix des M-estimateurs (c'est-à-dire de la fonction objectif $\rho$) influe non seulement sur la manière dont les centres des distributions sont mesurés mais aussi sur la définition même de ces centres. Autrement dit, nous choisissons la fonction objectif $\rho$ afin que les M-estimateurs $\hat{m}_{H,t}$ estiment bien les $m^0_{H,t}$ mais la fonction objectif définit simultanément $m^0_{H,t}$ par l'équation (13).

Un grand nombre de M-estimateurs existent dans la littérature statistique. Les plus classiques sont énumérés dans l'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. Ici, on choisit la fonction de score appropriée selon une méthode adaptative qui s'inspire de Ravalet [10]. La méthode est réalisée en plusieurs 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, on construit une loi théorique qui correspond au mieux, au regard de ces statistiques, aux distributions empiriques.
3. Puis on simule des échantillons à l'aide de la loi théorique choisie. Les paramètres de localisation des échantillons sont estimés à l'aide des différentes familles de fonctions de score envisagées. On retient la famille pour laquelle les erreurs d'estimations sont les plus faibles.
4. Dans la famille retenue, une fonction de score particulière est choisie selon le même critère (minimisation d'une fonction de coût des erreurs d'estimations sur des simulations).

Ces différentes étapes sont détaillées dans l'annexe B. Nous retenons une fonction dite "MRR" (cf. encadré 1) avec une constante de réglage $c = 0{,}10$ :

$$\rho(x) = \frac{0{,}10}{2} \log\left[x^2 + 0{,}10\right] \tag{7}$$

L'algorithme de calcul du problème de minimisation suit la méthode dite de "re-pondération itérative des Moindres Carrés Ordinaires". Cet algorithme est décrit dans l'annexe C.

## 3.5 Les M-estimateurs Quasi-Généralisés

L'estimateur ainsi décrit peut toutefois être rendu plus proche de l'efficacité. 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. Deux raisons expliquent cela. D'une part, les grandes entreprises sont susceptibles d'être plus rationnelles dans leurs projets d'investissement 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.

L'égalité des lois de répartition des couples $(t, d_{i,t})$ est une condition nécessaire pour assurer la convergence des M-estimateurs (cf. annexe A). La présence d'hétéroscédasticité invalide cette hypothèse. Il est donc nécessaire de transformer le modèle afin de réaliser les estimations sur des distributions vérifiant l'égalité des lois de distributions.

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. La variance de ces derniers varie en amplitude avec le chiffre d'affaires de chaque observation. La dépendance de la variance des résidus $\Sigma^2_{H,S}(\cdot)$ 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 $\hat{\Sigma}^2_{H,S}(\cdot)$ ainsi estimée. L'équation (3) est alors remplacée par l'équation (8), où les $\tilde{\epsilon}_{i,t}$ sont alors de même variance. Il est alors légitime d'accepter l'hypothèse d'égalité des lois des couples $(t, d_{i,t})$.

$$\frac{d_{i,t}}{\sigma_{H,S} \cdot \hat{\Sigma}_{H,S}\left(CA_{i,t}\right)} = \sum_{\tau} m_\tau \frac{\mathbf{1}_{\{\tau\}}(t)}{\sigma_{H,S} \cdot \hat{\Sigma}_{H,S}\left(CA_{i,t}\right)} + \tilde{\epsilon}_{i,t} \tag{8}$$

L'équation (6) devient alors :

$$\tilde{\rho}_{H,S}\left[t, d_{i,t}, (m_t)_{t\in S}\right] = \rho\left(\frac{d_{i,t} - m_t}{MAD\left(d_{H,S}\right) \cdot \hat{\Sigma}_{H,S}\left(CA_{i,t}\right)}\right) \tag{9}$$

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, on ré-ordonne les résidus selon les chiffres 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, identiquement distribués et de moyenne nulle, le modèle retenu s'écrit alors :

$$\log \Sigma_{H,S}\left(CA_{i,t}\right) = \alpha_{H,S} + \beta_{H,S} \log CA_{i,t} + \mu_{i,t} \tag{10}$$

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) :

$$\log \hat{\Sigma}_{H,S}\left(CA_{i,t}\right) = \hat{\alpha}_{H,S} + \hat{\beta}_{H,S} \log CA_{i,t} \tag{11}$$

La régression est également réalisée par les M-estimateurs. La fonction de score retenue est alors la fonction de Huber.

Après correction de l'hétéroscédasticité et pour la seconde estimation, la procédure de choix de la fonction de score (décrite par la partie 3.4 et détaillée dans l'annexe B) conduit à choisir toujours une fonction MMR, mais avec une constante plus importante : $c = 0{,}12$ (cf. annexe B.2 pour plus de précisions).

## 3.6 Construction des séries trimestrielles par strates et redressement

Pour chaque strate $H$, on dispose ainsi de trois séries $(\hat{m}_{H,t})_{t \in S}$, 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.

Dans chaque strate $H$, les trois séries annuelles sont ensuite réunies en une 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 strates sont ensuite agrégées en une unique série pour l'ensemble de l'industrie manufacturière. Pour cela, on

applique des coefficients de redressement constants, calculés comme les montants d'investissement par strate issus de l'Enquête Annuelle d'Entreprise (EAE) dans l'industrie de 2002.

# 4 L'indicateur des révisions dans l'*enquête investissement* améliore sensiblement la qualité de la prévision de l'investissement

## 4.1 Description de la série des révisions

La série ainsi construite apparaît assez bien corrélée aux variations trimestrielles de la FBCF des Entreprises Non Financières en valeur[11] (cf. figure 4). Sur la période qui s'étend du troisième trimestre de 1991 au dernier trimestre de 2003, la corrélation s'établit à 69%. L'indicateur d'un trimestre donné 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. 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.

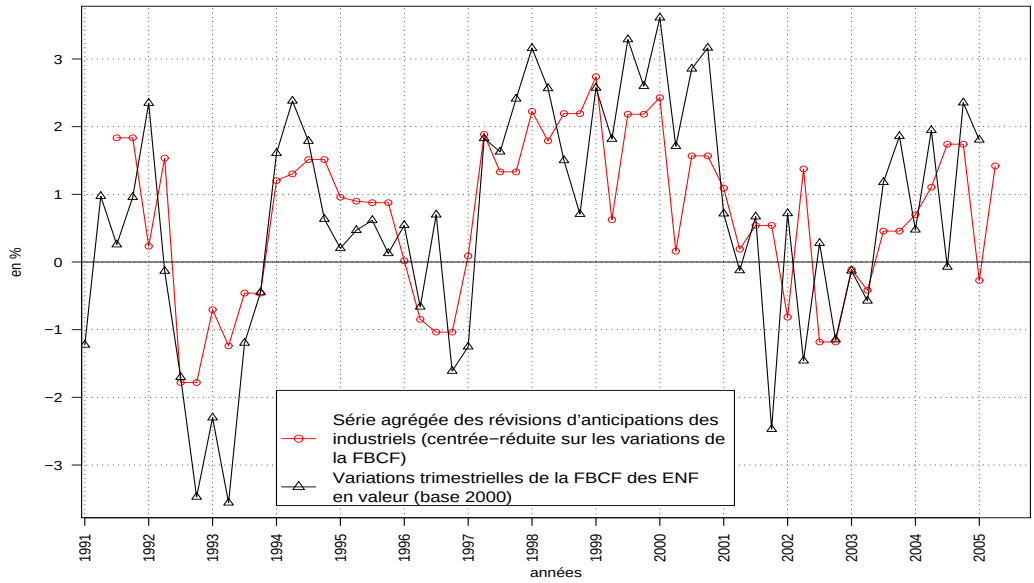
La série apparaît comme avancée par rapport aux comptes. Par exemple, la reprise de 1997 apparaît dans la série dès l'enquête de janvier 1997 alors qu'elle n'est effective qu'au deuxième trimestre dans les comptes trimestriels.

## 4.2 Un exemple de modèle d'étalonnage

Afin d'illustrer l'utilisation possible de l'indicateur des révisions pour la prévision de l'investissement à un rythme trimestriel, nous présentons un modèle d'étalonnage multivarié (modèle VAR). Celui-ci utilise pour unique variable explicative l'indicateur des révisions présenté dans cet article. En effet, les

[11] Les comptes trimestriels utilisés dans cet article sont les premiers résultats du premier trimestre de 2005 en base 2000.

FIG. 4 – Comparaison de la série construite et des évolutions de l'investissement des entreprises




Sources : Insee, *enquête investissement* et comptes nationaux trimestriels (base 2000). Calculs de l'auteur.

autres indicateurs prévoyant l'investissement[12] n'apportent pas d'information supplémentaire significative par rapport à celle contenue dans l'indicateur et ses retards.

Le processus vectoriel $X_t$ modélisé est formé des évolutions trimestrielles de la FBCF des ENF (Entreprises Non Financières) en volume[13] (notées FBCF) et des indicateurs des révisions (notés REV). $\text{FBCF}_t$ est l'évolution de la FBCF du trimestre $t$ et $\text{REV}_t$ est l'indicateur des révisions affecté au trimestre $t$. Par exemple, pour le deuxième trimestre de 2005, il s'agit de l'indicateur calculé avec l'*enquête investissement* d'avril 2005.

[12] Cf. le dossier dans la note de conjoncture de l'Insee [2].

[13] Les comptes trimestriels utilisés dans l'article sont les premiers résultats du premier trimestre de 2005 en base 2000.

$$X_t = \begin{pmatrix} \text{FBCF}_t \\ \text{REV}_t \end{pmatrix}$$

On vérifie par des tests de racine unité la stationnarité des deux séries FBCF et REV : les test ADF (Augmented Dickey-Fuller) rejettent très significativement la non-stationnarité des séries FBCF er REV. Il est alors naturel d'écrire $X_t$ sous la forme d'un processus VAR. En notant $L$ l'opérateur retard, on modélise $X_t$ par l'équation (12), où $\Phi$ est un polynôme, $c_0$ un vecteur constant et $\nu_t$ est un bruit blanc gaussien. L'absence d'autocorrélation des perturbations $\nu_t$ et leur normalité seront testées *a posteriori*.

$$X_t = c_0 + \Phi(L)X_{t-1} + \nu_t \tag{12}$$

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 deux équations de la dynamique sont alors estimées par les MCO[14] :

$$\left\{ \begin{array}{lllllll} \text{FBCF}_t & = & 0{,}46 & + & 0{,}26 \quad \text{FBCF}_{t-1} & + & 0{,}15 \quad \text{REV}_{t-1} \\ & & (2{,}2) & & (1{,}6) & & (2{,}9) \\ \text{avec} & R^2 = 0{,}45, & \overline{R}^2 = 0{,}42, & & RMSE = 1{,}30 & \text{et} & \text{Durbin-Watson} = 1{,}97 \end{array} \right.$$

$$\left\{ \begin{array}{lllllll} \text{REV}_t & = & \text{-}1{,}10 & + & 1{,}41 \quad \text{FBCF}_{t-1} & + & 0{,}24 \quad \text{REV}_{t-1} \\ & & (\text{-}1{,}7) & & (2{,}9) & & (1{,}5) \\ \text{avec} & R^2 = 0{,}44, & \overline{R}^2 = 0{,}42, & & RMSE = 3{,}89 & \text{et} & \text{Durbin-Watson} = 1{,}96 \end{array} \right.$$

Nous vérifions *a posteriori* que les résidus suivent bien un bruit blanc[15] et qu'ils sont biens distribués selon une loi gaussienne bivariée[16].

[14] Les variations de FBCF sont exprimées en pourcentage.

[15] On réalise un test du portmanteau multivarié. L'hypothèse nulle est l'absence d'autocorrélation des résidus. La statistique de test vaut 50,5. Elle est à comparer à un $\chi^2$ de 60 degrés de liberté, soit une p-value de 0,80. Il n'est donc pas possible de rejeter l'absence d'autocorrélation des résidus.

[16] Appliqués aux résidus, la statistique de Skewness vaut 2,64 et celle de Kurtosis 0,40. La statistique jointe pour le test de Doornick et Hansen (1994) vaut 3,04, soit une p-value de 0,21 pour l'hypothèse de normalité des résidus.

L'indicateur REV est disponible un peu plus de trois mois avant la publication des comptes du trimestre correspondant. De manière à pouvoir utiliser la dernière valeur de l'indicateur ($\text{REV}_t$) au moment de la prévision du trimestre courant $t$, il est utile d'écrire le modèle sous sa forme dite "bloc récursive". Cette forme peut-être obtenue en transformant le système des deux équations ci-dessus 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 la série REV ainsi que sur les deux séries retardées.

$$\begin{cases} \begin{array}{lcccccccccc} \text{FBCF}_t & = & 0{,}65 & + & 0{,}01 & \text{FBCF}_{t-1} & + & 0{,}17 & \text{REV}_t & + & 0{,}11 & \text{REV}_{t-1} \\ & & (3{,}4) & & (0{,}1) & & & (4{,}0) & & & (2{,}4) & \end{array} \\ \text{avec} \quad R^2 = 0{,}59, \quad \overline{R}^2 = 0{,}57, \quad RMSE = 1{,}12 \quad \text{et} \quad \text{Durbin-Watson} = 1{,}73 \end{cases}$$

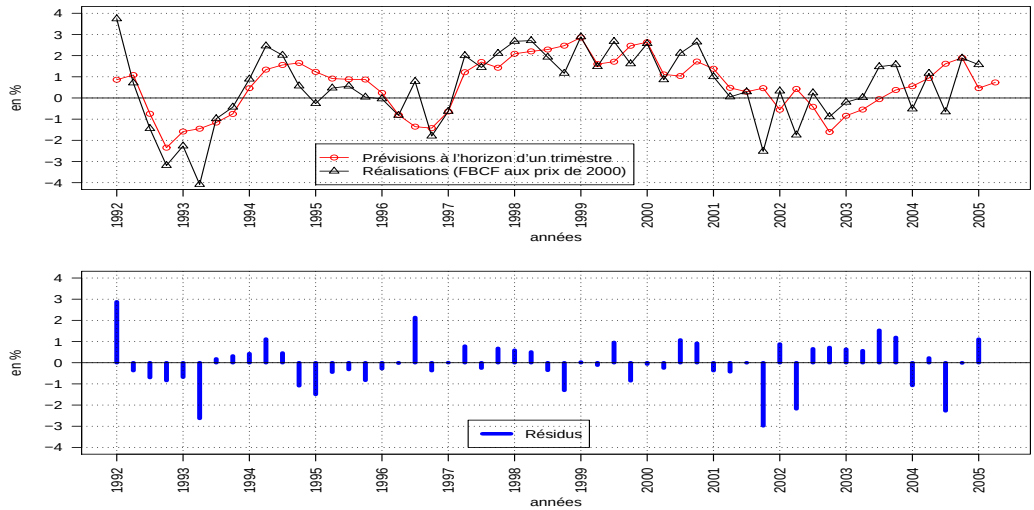
La forme "bloc récursive" permet alors d'utiliser toute l'information disponible pour prévoir la FBCF du trimestre concomitante à l'indicateur. Pour les trimestres suivants, la forme canonique est utilisée pour prolonger le processus $(X_t)_t$.

Par exemple, au mois de mai 2005, les comptes trimestriels sont disponibles jusqu'au premier trimestre de 2005. L'indicateur REV est lui disponible jusqu'au deuxième trimestre de 2005 (enquête d'avril 2005). Pour ce même trimestre, l'indicateur vaut 3,0. Sur la base de l'information disponible en mai 2005, la forme "bloc récursive" prévoit alors une évolution de la FBCF des ENF de 0,7% au deuxième trimestre de 2005. La forme canonique permet de prolonger la prévision : le taux de croissance de la FBCF des ENF serait de 1,1% au troisième trimestre de 2005 puis de 0,8% au quatrième trimestre de 2005.

Sur la période de début-1992 à début-2005, l'écart-type des résidus à l'horizon d'un trimestre (forme "bloc-récursive" ) vaut 1,10. Prolongé par la forme canonique à l'horizon de deux trimestres, l'écart-type s'établit à 1,26. Enfin, à l'horizon de trois trimestres, il est de 1,43. Ces valeurs sont à comparer à l'écart-type des variations de FBCF sur cette même période : 1,66. 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 65% et celles à l'horizon de trois trimestres à 44%. Au regard de ces statistique, il apparaît que l'indicateur est un bon outil pour prévoir les évolutions d'investissement sur

deux trimestres. Toutefois, comme tous les indicateurs d'enquête de conjoncture, il devient insuffisant pour prévoir les évolutions sur trois trimestres.

FIG. 5 – Prévision des variations trimestrielles de la FBCF des ENF en volume à l'horizon d'un trimestre




Sources : Insee, *enquête investissement* et comptes nationaux trimestriels (base 2000). Calculs de l'auteur.

La figure 5 représente les prévisions à horizon 1 (forme "bloc-récursive" ) ainsi que les réalisations et les résidus.

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 d'avril 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.

# 5 Conclusion

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 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 d'anticipations : cet indicateur apporte une information pertinente pour la prévision des variations trimestrielles de l'investissement.

# A Le cadre théorique des M-estimateurs

Le problème limite désigne le problème de minimisation lorsque le nombre d'observations tend vers l'infini. Le problème limite associé au programme d'optimisation (5) lorsque $n_{H,S}$ tend vers l'infini est équivalent au problème de minimisation donné par la formule (13). Dans cette formule, $(T, D)$ est un couple de variables aléatoires désignant respectivement la date d'une observation et la valeur de la révision associée à cette observation. $(T, D)$ suit la vraie loi de la distribution des $(t, d_{i,t})$ pour la strate considérée et la saison considérée. $E_0$ représente l'espérance sous la vraie loi de $D$ à $T$ fixé et $E_T$ l'espérance sur $T$, date de l'observation $(T, D)$.

$$\text{Arg} \min_{(m_t)_{t\in S}} \; E_T E_0 \; \tilde{\rho}\left[T, D, (m_t)_{t\in S}\right] \tag{13}$$

Certaines conditions sont nécessaires pour assurer la convergence des estimateurs vers la solution du problème limite. Ces conditions sont données dans Gouriéroux-Monfort [3] :

1. les couples $(t, d_{i,t})$ sont indépendants et de même loi,
2. l'espace d'optimisation est ouvert ($\mathbb{R}^T$ convient),
3. la fonction $\tilde{\rho}$ est continue par rapport à $(m_t)_{t\in S}$ et est intégrable par rapport à la vraie loi de $(T, D)$ pour tout $(m_t)_{t\in S}$. Ceci est équivalant à la continuité de $\rho$ et à son intégrabilité par rapport à la vraie loi des résidus.
4. $\frac{1}{n_{H,S}} \sum_{\substack{i\in H \\ t\in S}} \tilde{\rho}\left[t, d_{i,t}, (m_t)_{t\in S}\right]$ converge presque sûrement uniformément vers $E_T E_0 \; \tilde{\rho}\left[T, D, (m_t)_{t\in S}\right]$ lorsque le nombre d'observations $n_{H,S}$ tend vers l'infini.
5. le problème limite admet une solution unique $(m_t^0)_t$.

Sous ces conditions, il existe asymptotiquement une solution aux conditions du premier ordre du problème de minimisation à distance finie. Cette solution $(\hat{m}_t)_t$, appelée M-estimateur, converge presque sûrement vers $(m_t^0)_t$ lorsque le nombre d'observations $n_{H,S}$ tend vers l'infini.

Sous certaines conditions supplémentaires[17] , $\sqrt{n}\left((\hat{m}_t)_t - (m_t^0)_t\right)$ suit asymptotiquement la loi normale centrée et de matrice de variance $J^{-1}IJ^{-1}$ avec :

[17]Il suffit que $\tilde{\rho}$ soit deux fois continûment dérivable par rapport à $(m_t)_t$, que la matrice $J$ (définie ci-après) existe et qu'elle soit inversible.

$$I = E_T E_0 \left( \frac{\partial \tilde{\rho}\left[T, D, (m_t^0)_{t\in S}\right]}{\partial (m_t)_t} \frac{\partial \tilde{\rho}\left[T, D, (m_t^0)_{t\in S}\right]}{\partial (m_t)_t'} \right)$$

$$J = E_T E_0 \left( -\frac{\partial^2 \tilde{\rho}\left[T, D, (m_t^0)_{t\in S}\right]}{\partial (m_t)_t \ \partial (m_t)_t'} \right)$$

Le choix de la fonction objectif $\rho$ est contraint par la vérification des conditions énoncées *supra*. Concrètement, cette fonction est choisie parmi les fonctions continues, symétriques, minimales en 0 et croissantes sur $]0, +\infty[$. On vérifie *ex post* que l'intégrabilité est vérifiée, que le problème à distance finie a bien une solution sur les distributions empiriques observées et que le problème limite a une solution unique dans les familles de distributions théoriques envisagées.

En outre, les distributions doivent vérifier la condition 1. L'indépendance ne pose pas de difficulté. Cependant les couples $(t, d_{i,t})$ se révèlent ne pas être identiquement distribués, la variance de $d_{i,t}$ évoluant à l'inverse de la taille de l'entreprise. La partie 3.5 expose comment corriger le modèle afin de rentrer dans le cadre théorique énoncé *supra*.

# B Choix du M-estimateur

Un M-estimateur est défini par sa fonction objectif $\rho$ ou par la dérivée de cette fonction, appelée fonction de score et notée $\psi$. Il n'y a pas de choix optimal *a priori* et le M-estimateur retenu est choisi afin que l'estimation soit robuste et proche de l'efficacité sur les distributions empiriques. Une procédure de choix de la fonction de score est dite adaptative lorsque la fonction $\psi$ 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 l'hypothèse que la loi des perturbations appartienne à une famille paramétrique de lois.

## B.1 Procédure de choix de la fonction de score

Nous détaillons ici la procédure adaptative retenue pour choisir la fonction de score. Comme nous l'avons déjà précisé dans le corps du texte, cette procédure peut-être divisée en quatre 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, on construit une loi théorique qui correspond au mieux, au regard de ces statistiques, aux distributions empiriques.
3. Puis on simule des échantillons à l'aide de la loi théorique choisie. Les paramètres de localisation des échantillons sont estimés à l'aide des différentes familles de fonctions de score envisagées. On retient la famille pour laquelle les erreurs d'estimations sont les plus faibles.
4. Dans la famille retenue, une fonction de score particulière est choisie selon le même critère (minimisation d'une fonction des erreurs d'estimations sur des simulations).

Ces différentes étapes sont détaillées ci-après afin d'exposer le choix de la fonction de score pour la première estimation par la méthode des "M-estimateurs Quasi-Généralisés" (cf. partie 3.5). Le choix de la fonction de score pour la seconde estimation suit la même procédure et les résultats intermédiaires sont donnés dans la partie B.2 de l'annexe.

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 ceux retenus par Ravalet [10].

   La statistique d'épaisseur de queues a été proposée par Hogg [4]. On note $U(p)$ (respectivement $L(p)$) la moyenne des $np$ plus grandes (respectivement plus petites) statistiques d'ordre d'une distribution de taille $n$. On utilise une interpolation linéaire lorsque $np$ n'est pas entier. On calcule alors $\tau$, l'indicateur d'épaisseur de queue par la formule (14). Plus les queues d'une distribution sont épaisses, plus $\tau$ est grand. Dans le cas d'une distribution gaussienne, $\tau$ vaut 2,59.

   $$\tau = \frac{U(5\%) - L(5\%)}{U(50\%) - L(50\%)} \tag{14}$$

   On note $X(a, b)$ la moyenne des statistiques d'ordre entre la $na^{\text{ème}}$ et la $nb^{\text{ème}}$ (en interpolant si nécessaire). L'indicateur de concentration

TAB. 4 – Statistiques $\tau$ et $P_k$ sur quelques lois symétriques

| **Lois théoriques considérées** | $\tau$ | $P_k$ |
|---|---|---|
| Loi normale | 2,6 | 2,7 |
| Loi de Slash | 8,5 | 4,2 |
| Loi double-exponentielle | 3,3 | 3,4 |
| "Loi normale puissance 2" | 4,4 | 5,6 |
| "Loi normale puissance 3" | 6,1 | 10,0 |
| "Loi normale puissance 6" | 9,1 | 44,8 |

Calculs de l'auteur.

autour de la médiane est donné par Hogg et alii [5] et est noté $P_k$ (formule ci-dessous). Plus $P_k$ est grand, plus la distribution est concentrée autour de la médiane. $P_k$ vaut 2,74 dans le cas d'une loi normale.

$$P_k = \frac{X(85\%, 95\%) - X(5\%, 15\%)}{X(50\%, 85\%) - X(15\%, 50\%)} \tag{15}$$

Les statistiques calculées varient peu selon les strates et les saisons. Nous retenons comme valeurs $\tau = 5{,}4$ et $P_k = 6{,}0$.

2. On recherche alors une loi théorique qui présente 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 bien 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 est donc nécessaire de confirmer cette proximité par des représentations graphiques des distributions à l'aide de graphiques de comparaison quantiles-quantiles. Un tel exemple est donné par la figure 6. On cherche une loi symétrique qui présente des statistiques $\tau$ et $P_k$ égales à 5,4 et 6,0. Pour cela, on envisage différentes lois dont les statistiques sont données par la table 4.

   On rappelle que la loi de Slash est définie par le rapport entre une loi normale de moyenne nulle et de variance 1 et une loi uniforme sur $[0, 1]$. Par ailleurs, 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 signe$(N)|N|^p$ lorsque $N$ suit une loi normale centrée et de variance unitaire.

Au regard de ces résultats, on cherche une distribution construite à l'aide la formule (16) :

$$X = \alpha D + (1 - \alpha) \text{ signe}(N) \left| \frac{N}{K} \right|^p \tag{16}$$

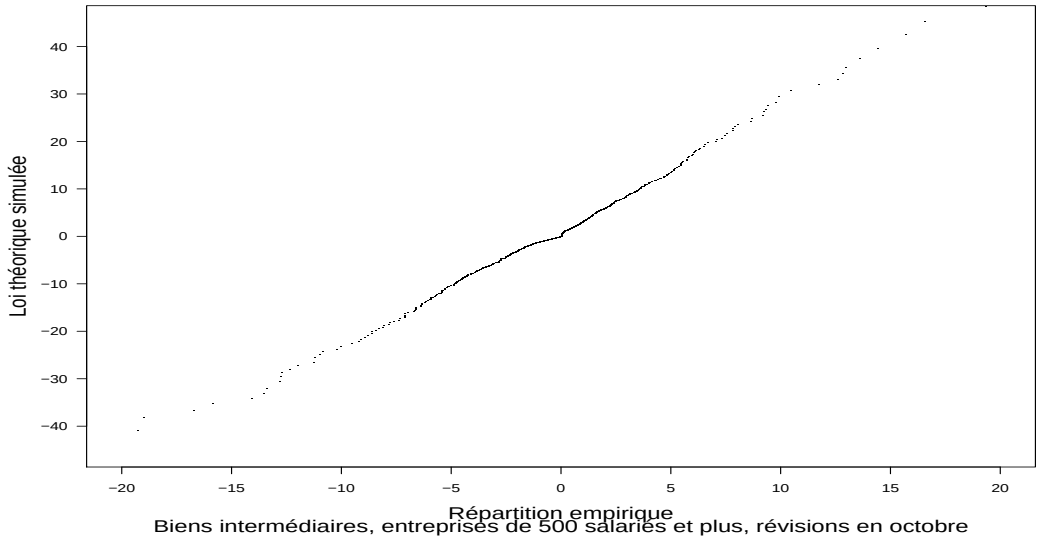
avec $D$ suivant une loi double-exponentielle et $N$ une loi normale centrée et de variance 1. Avec $\alpha = 0{,}60$, $K = 0{,}61$ et $p = 2{,}77$ (obtenus par un algorithme de Newton), la loi théorique ainsi construite a les mêmes statistiques $\tau$ et $P_k$ que les lois empiriques des révisions obtenues par strate.

Des graphiques inter-quantiles permettent de vérifier la proximité entre les distributions empiriques des résidus et cette distribution théorique. A titre d'exemple, la figure 6 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. Les fonctions de score envisagées (cf. encadré 1 page 17) sont testées en faisant varier le paramètre de réglage $c$ (ou les paramètres $a$, $b$ et $c$ dans le cas des fonctions de Hampel). On simule avec la loi théorique choisie 1 000 échantillons d'une taille de 200 observations. Dans chaque cas, on estime le paramètre de position de la distribution avec le M-estimateur correspondant à la fonction de score testée. La racine carrée du moment empirique d'ordre 2 des 1 000 paramètres estimés est choisie comme critère de manque d'efficacité de l'estimation. Les résultats sont présentés dans le tableau 5.

   Au vu de ces estimations, les M-estimateurs avec les fonctions MRR et avec une petite constante de réglage $c$ apparaissent être les meilleurs. Seuls les M-estimateurs avec les fonctions de Hampel semblent avoir une efficacité comparable. Le fait de n'avoir qu'une seule constante de réglage (ce qui a la vertu de la simplicité) et le fait de ne pas rejeter

FIG. 6 – Graphique quantile-quantile entre la loi théorique et les résidus observés




Source : Insee, *Enquête investissement*. Calculs de l'auteur.

totalement de l'estimation les points extrêmes incitent également à préférer les fonctions MRR aux fonctions de Hampel.

4. Il reste à choisir précisément le paramètre $c$ dans la fonction MRR. Pour cela, le critère est le même que pour le choix de la famille de fonction. Il conduit à choisir $c = 0{,}10$. Dans ce cas-là, la racine du moment empirique d'ordre deux de la distribution des erreurs est 0,30.

Cette méthode présente certaines limites. En particulier, notre définition de la proximité ne repose pas sur une distance[18]. 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 (14) et (15)), cela ne modifie pas significativement la loi théorique choisie.

[18]Il s'agit en fait d'une semi-distance.

TAB. 5 – Résultats des simulations

| M-estimateur | Racine carré du moment empirique d'ordre 2 |
|---|---|
| MCO (i.e. la moyenne arithmétique) | 0,053 |
| Médiane | 0,037 |
| Huber avec $c = 1$ | 0,046 |
| Huber avec $c = 3$ | 0,081 |
| Biweight avec $c = 1$ | 0,038 |
| Biweight avec $c = 5$ | 0,044 |
| Sinus de Andrew avec $c = 1$ | 0,036 |
| Sinus de Andrew avec $c = 5$ | 0,047 |
| Hampel avec $a = 0,5$, $b = 1$, $c = 2$ | 0,033 |
| Hampel avec $a = 2$, $b = 4$, $c = 8$ | 0,054 |
| MRR avec $c = 1$ | 0,032 |
| MRR avec $c = 4$ | 0,043 |
| MRR avec $c = 0,1$ | 0,030 |

Calculs de l'auteur.

Au total, la fonction de score MRR avec $c = 0,10$ apparaît un choix pertinent sur l'ensemble des strates et des saisons.

## B.2 Choix de la fonction de score pour la seconde estimation par les MCQG

Le choix de la fonction de score pour la seconde estimation par les "M-estimateurs Quasi-Généralisés" suit la même procédure que celle décrite ci-dessus (cf. B.1).

Après correction de l'hétéroscédasticité, les paramètres d'épaisseur de queues et de concentration sont désormais proches de $\tau = 5,3$ et $P_k = 6,1$. La loi théorique correspondante est donnée par $\alpha = 0,60$, $K = 0,60$ et $p = 2,70$. Ceci conduit à choisir une constante plus importante pour la seconde estimation : $c = 0,12$.

# C Algorithme de calcul

Les estimations ne sont pas calculées par la résolution directe des problèmes de minimisation associés mais par une méthode itérative de re-pondération de la méthode des MCO (Moindres Carrés Ordinaires). Cette procédure s'avère en effet converger plus rapidement. Nous en exposons ici le principe.

On se place dans le cas général de la régression d'une variable $Y$ sur $K$ variables explicatives $X^{(k)}$, $k$ compris entre 1 et $K$. Ces $K$ variables peut éventuellement comprendre une constante unitaire.

$$Y \approx X\beta$$

La variable $Y$ et les vecteurs $X = (X^{(k)})_{k\in[1..K]}$ sont observés $N$ fois et prennent les valeurs $y_i$ et $x_i = (x_i^{(k)})_{k\in[1..K]}$ avec $i$ variant de 1 à $N$.

On cherche à estimer $\min_\beta \sum_i \rho(y_i - x_i\beta)$. Si la solution existe, alors elle vérifie nécessairement les conditions du premier ordre du problème de minimisation. Ces conditions sont données par $0 = \sum_{i,t} x_i^{(k)} \psi(y_i - x_i\beta)$ pour tout $k$ de 1 à $K$.

En posant $w(r) = \psi(r)/r$, ceci peut se réécrire pour tout $k$ de 1 à $K$ :

$$0 = \sum_{i,t} x_i^{(k)} (y_i - x_i\beta)\ w(y_i - x_i\beta)$$

Ces $K$ équations sont alors exactement semblables à celles obtenues par minimisation des carrés des résidus avec les poids individuels $w_i = w(y_i - x_i\beta)$. D'où l'idée de résoudre itérativement par les MCO pondérés : à l'étape $n$, $\beta^{(n)}$ est le résultat de l'estimation par les MCO pondérés par les poids individuels :

$$w_i^{(n)} = w\left(y_i - x_i\beta^{(n-1)}\right)$$

Si la suite $\beta^{(n)}$ converge vers une limite $\beta^{(\infty)}$, alors $\beta^{(\infty)}$ vérifie les conditions du premier ordre du problème initial et donc $\beta^{(\infty)}$ est un minimum local.

Selon la fonction objectif choisie, l'algorithme peut ne pas converger. L'algorithme peut aussi converger vers une solution qui ne soit pas le minimum global. L'algorithme doit donc partir d'un point correctement choisi et il faut vérifier que le processus a bien convergé. Le point de départ est obtenu par la méthode des MCO non-pondérés, c'est-à-dire avec $w_i^{(0)} = 1$.

# Références


[1] Fiche méthodologique : Enquête sur les investissements dans l'industrie, sur le site www.insee.fr. Sous la rubrique conjoncture/indicateurs de conjoncture/principaux indicateurs.

[2] **Ferrari N.**, Dossier "Prévoir l'investissement des entreprises ? Un indicateur des révisions d'anticipations dans l'enquête Investissement dans l'industrie" dans la Note de conjoncture de l'Insee de mars 2005. Disponible sur le site www.insee.fr.

[3] **Gourieroux C. et A. Monfort (1996)**, "Statistique et modèles économétriques", *vol. 1 , Économica (2ème édition).*

[4] **Hogg R. V. (1974)**, "Adaptative robust procedures : a partial review and some suggestions for future applications and theory", *Journal of the American Statistical Association n°69, 909 - 923.*

[5] **Hogg R. V., G. K. Bril, S. M. Han et L. Yul (1988)**, "An argument for adaptative robust estimation", *Probability and statistics Essays in Honor of Franklin A. Graybill.*

[6] **Huber P. J. (1964)**, "Robust estimation of a location parameter", *The Annals of Mathematical Statistics n°35, 73 - 101.*

[7] **Lecoutre J.-P. et P. Tassi (1987)**, "Statistique non paramétrique et robustesse", *Économica.*

[8] **Moberg T. F., J. S. Ramberg et R. H. Randles (1980)**, "An adaptive multiple regression procedure based on M-estimators", *Technometrics n°22, 212 - 224.*

[9] **Naboulet A. et S. Raspiller (2004)**, "Les déterminants de la décision d'investir : une approche par les perceptions subjectives des firmes", *Document de travail de l'Insee n°G2004/04.*

[10] **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 de l'Insee n°9605.*

[11] **Rosenwald F. (1994)**, "L'enquête sur l'investissement industriel", *Insee Méthodes n°45.*