

Les managers connaissent-ils leurs entreprises ?

Les leçons de l'enquête *REPONSE*

Philippe Askenazy* et Julien Grenet**

Cet article explore les écueils potentiels liés à l'utilisation des enquêtes sur les pratiques organisationnelles, les choix technologiques des entreprises et les relations professionnelles. Une méthodologie générique permet d'estimer, sous certaines hypothèses, le niveau de véracité ou de cohérence des déclarations des dirigeants enquêtés travaillant dans différents établissements d'une même entreprise. Appliquée à l'enquête française *REPONSE* pour les aspects généraux de l'entreprise (taille, appartenance à un groupe, structure de l'actionnariat, etc.), cette méthodologie suggère que les dirigeants donnent en général des réponses plutôt cohérentes d'une question prise isolément à une autre, que les réponses sont d'autant plus précises que les questions sont simples et qu'elles relèvent du champ de compétence du dirigeant interrogé. Cependant les réponses aux questions portant sur les relations sociales et la représentation des syndicats dans l'entreprise sont moins fiables. La prise en compte des erreurs commises par les dirigeants conduit à réviser sensiblement à la fois le niveau et l'évolution d'un certain nombre de variables, en particulier le poids des différents syndicats de salariés dans les entreprises.

* Philippe Askenazy est directeur de recherche CNRS à l'École d'Économie de Paris et membre du Cepremap.

** Julien Grenet est chercheur au Centre for Economic Performance de la London School of Economics and Political Science.

Les auteurs tiennent à remercier vivement deux rapporteurs anonymes pour leurs très stimulantes remarques et suggestions. Ils remercient également Loup Wolff de la Dares pour son aide sur les données de pondération.

Une très importante littérature en économie s'est construite depuis une quinzaine d'années autour de la thématique des changements technologiques et organisationnels que connaissent les entreprises (Aghion *et al.*, 1999). La plupart des travaux confirment les toutes premières études et les analyses issues du management : même si on ne peut tracer un modèle unique, les entreprises tendent à adopter simultanément les technologies de l'information et de la communication et des pratiques de travail flexibles. Ces pratiques recouvrent le développement de la polyvalence, de l'autonomie des salariés, du juste à temps ou des démarches qualité. Un faisceau d'indices souligne que les performances des entreprises dépendraient des complémentarités entre ces changements (Bresnahan *et al.*, 2002).

À l'appui de ces avancées, des enquêtes de plus en plus larges ont été réalisées auprès d'entreprises. Les premières se sont limitées à des secteurs particuliers comme la sidérurgie (Ichniowski *et al.*, 1997) ou le textile (Dunlop et Weil, 1996). Les suivantes ont porté sur des vagues d'échantillons représentatifs, comme les enquêtes NES (*National Employment Survey*) (Black et Lynch, 2001) et NOS (*National Organizations Survey*) aux États-Unis (Osterman, 1994), WIRS-WERS (*Workplace Industrial Relations Survey/Workplace Employment Relations Survey*) en Grande-Bretagne (Addison *et al.*, 2004) ou COI (*Changement Organisationnel et Informatisation*) (Aubert *et al.*, 2006) et REPONSE en France (Caroli et van Reenen, 2001). Les questionnaires sont également devenus plus riches et homogènes, tenant compte à la fois de l'émergence de nouvelles pratiques dans les entreprises et de la qualité des réponses obtenues dans les premières vagues. Plus récemment, des enquêtes internationales ont été réalisées, la plus complète à ce jour ayant été menée par Bloom et van Reenen (2007). Il s'agissait pour ces auteurs de comprendre comment les pratiques managériales varient d'un pays à un autre, afin d'identifier une cause possible des écarts de performance macroéconomique observés entre les pays de l'OCDE.

Mais dès le début de l'utilisation de ces enquêtes, un certain nombre d'auteurs ont souligné l'existence de failles potentielles (Huselid et Becker, 1996) : les réponses des dirigeants d'entreprises ou des cadres peuvent être biaisées et on ne peut exclure que des erreurs significatives soient commises, en particulier lorsque les questions sont complexes ou renvoient à des concepts managériaux ou technologiques

généraux. Les sources de biais sont multiples : les dirigeants peuvent ne pas connaître la bonne réponse et répondre au hasard ; ils peuvent avoir une mauvaise connaissance du terrain et se contenter de réciter l'enseignement qu'ils ont reçu en école de management. Ils peuvent aussi vouloir paraître en pointe dans l'adoption de pratiques managériales et organisationnelles innovantes, en déclarant justement mettre en œuvre ces nouvelles pratiques. Si tel était le cas, la validation par l'enquête des prédictions des manuels de management pourrait relever davantage de la fiction auto-réalisatrice que de la réalité.

Conscients de ces obstacles et bénéficiant de financements plus importants, les travaux les plus prudents ont réalisé des enquêtes complémentaires sur des échantillons très réduits afin de vérifier l'existence d'une corrélation positive entre les réponses des dirigeants d'entreprises entre les deux vagues. Si les résultats sont plutôt rassurants (Bloom et van Reenen, 2007), ils ne permettent ni de quantifier précisément l'exactitude des réponses fournies par les dirigeants interrogés, ni d'évaluer les conséquences de la non-prise en compte des erreurs éventuellement commises par ces derniers.

Cet article présente une exploration systématique des écueils potentiels liés à l'utilisation des enquêtes sur les pratiques organisationnelles en tirant parti de la présence de grappes d'établissements appartenant aux mêmes entreprises dans l'échantillon du volet « Représentants de la direction » de l'enquête française REPONSE. Cette enquête, menée tous les sept ans auprès d'environ 3 000 représentants de la direction d'entreprise interrogés au niveau des établissements, contient un grand nombre de questions portant sur l'organisation du travail, les technologies et les relations professionnelles. Les entretiens sont réalisés en face à face et la position professionnelle des enquêtés est donnée en clair.

Pour les besoins de cette étude, nous développons une méthodologie générique pour estimer, sous certaines hypothèses, le niveau de véracité et de cohérence des réponses fournies par les individus aux questions portant sur les caractéristiques de leur groupe, en extrayant l'information issue du rapprochement des réponses de membres d'un même groupe. Cette approche permet d'estimer séparément le taux de prévalence réel d'une caractéristique donnée parmi l'ensemble des groupes et la probabilité que les individus interrogés fournissent une réponse erronée à la question portant sur cette caractéristique.

Cette méthodologie est notamment adaptée aux situations dans lesquelles les informations sur le groupe ne peuvent être connues que de manière « indirecte », par l'intermédiaire des réponses fournies par les membres de ce groupe, soit parce qu'il n'existe pas de représentant clairement identifiable du groupe (comme dans le cas des réseaux sociaux), soit que le plan de sondage ne le sélectionne pas systématiquement.

Le volet « Représentants de la direction » de l'enquête *REPONSE* présente un intérêt triple du point de vue de l'application de la méthodologie exposée dans cette étude.

Le premier intérêt de cette source tient au fait qu'elle satisfait aux critères requis pour évaluer le degré de fiabilité des réponses fournies par les individus sur les caractéristiques de leur groupe. L'enquête *REPONSE* permet en effet de travailler à partir d'un échantillon conséquent de cadres dirigeants appartenant aux mêmes entreprises. Le tirage étant effectué sur des établissements et non sur des entreprises, l'échantillon principal de chaque vague d'enquête contient un nombre important d'établissements faisant partie de la même entreprise, si bien que 8 % environ des entreprises présentes dans *REPONSE* disposent de deux établissements, ou plus, enquêtés. Munis de ce sous-échantillon, il est possible de rapprocher les réponses fournies par un représentant de la direction interrogé dans un établissement particulier de celles de son, ou de ses, homologue(s) au sein de la même entreprise. L'enquête *REPONSE* présente également la particularité de n'approcher la dimension « entreprise » qu'à travers les réponses fournies au niveau des établissements, alors qu'il aurait pu sembler plus direct d'étudier cette dimension en se limitant à l'établissement-siège de chaque entreprise. Le choix finalement retenu s'explique par la nécessité de privilégier la dimension « établissement » dans l'analyse des pratiques organisationnelles, en travaillant à partir d'un échantillon représentatif d'établissements, ce qui exclut l'inclusion systématique du siège des entreprises.

Le second intérêt de l'application de notre méthodologie à l'enquête *REPONSE* tient au fait qu'elle permet d'évaluer le degré de connaissance que les représentants de la direction ont de leur entreprise. L'approche indirecte de la dimension « entreprise », qui est privilégiée par l'enquête *REPONSE*, pourrait laisser croire qu'il s'agit là d'un aspect relativement secondaire du point de vue des utilisateurs de l'enquête. Tel n'est pas le cas en réalité. En pratique, on constate en effet que les réponses aux

questions portant sur les caractéristiques des entreprises sont abondamment utilisées dans la littérature. Les réponses fournies par les représentants de la direction aux questions portant sur les relations sociales dans leur entreprise constituent notamment l'une des principales sources d'information sur la présence syndicale dans les entreprises (Pignoni et Tenret, 2007). Par ailleurs, un certain nombre de pratiques managériales étudiées dans la littérature consacrée au changement organisationnel sont définies au niveau de l'entreprise, qu'il s'agisse de l'adoption d'une démarche de qualité totale ou de l'utilisation de progiciels de gestion intégrée. Mesurer la fiabilité des réponses fournies par les représentants de la direction à ces questions représente par conséquent un enjeu important lorsqu'on s'intéresse aux transformations des modes d'organisation du travail.

L'enquête *REPONSE* présente enfin l'avantage de pouvoir se prêter à des analyses longitudinales. Les deux vagues d'enquêtes réalisées en 1998 et 2004, qui contiennent un grand nombre de questions communes, permettent de mesurer l'évolution de la fiabilité des réponses fournies par les représentants de la direction et d'en déduire l'évolution réelle de certaines caractéristiques de leurs entreprises.

L'application de notre méthodologie à l'enquête *REPONSE* pour les aspects généraux de l'entreprise (taille, appartenance à un groupe, structure de l'actionnariat, etc.) révèle que les dirigeants donnent en général des réponses plutôt cohérentes d'une question prise isolément à une autre. La cohérence est d'autant plus forte que la question est simple, comme « *votre entreprise est-elle cotée en Bourse ?* ». Par ailleurs, les erreurs ne sont pas significativement corrélées d'une question à une autre, ce qui suggère que les dirigeants qui se trompent sur un aspect de l'entreprise ne se trompent pas systématiquement sur les autres aspects.

Seules les questions portant sur les relations sociales et la présence syndicale dans l'entreprise présentent des niveaux d'incohérence marqués, soit que les dirigeants méconnaissent cet aspect de leur entreprise, soit qu'ils répugnent à reconnaître la présence de délégués syndicaux. Par ailleurs, la mauvaise connaissance des relations sociales dans l'entreprise est corrélée à des caractéristiques telles que la fonction du représentant de la direction interrogé, la durée de l'entretien ou encore l'année d'enquête. Enfin, contrairement aux questions portant sur les aspects généraux de l'entreprise, les erreurs

portant sur la présence syndicale sont significativement corrélées d'une question à l'autre.

Pour faibles qu'elles soient, les erreurs commises par les dirigeants sont susceptibles de biaiser sensiblement l'estimation, d'une part, du taux de prévalence de certaines caractéristiques des entreprises (actionnariat salarié, adoption d'une norme ISO, poids relatif des différents syndicats) et, d'autre part, leur évolution dans le temps. Aussi proposons-nous un certain nombre d'améliorations méthodologiques susceptibles d'accroître la qualité des réponses recueillies dans le cadre de l'enquête *REPONSE* et de limiter les biais potentiellement associés à l'utilisation des variables de cette enquête.

Développement de la méthodologie générique

Une méthodologie générique permet d'estimer séparément le taux de prévalence réel d'une caractéristique binaire d'un groupe et la probabilité que les individus interrogés sur cette caractéristique se trompent, à partir des réponses fournies par des binômes ou triplets d'individus appartenant à un même groupe.

Le problème qui justifie l'utilisation de la méthodologie développée dans cet article est relativement simple à comprendre et peut être exposé au moyen d'un exemple volontairement trivial. Supposons que l'on observe cinq ménages formés d'un binôme de conjoints et que l'on demande à chaque membre du binôme d'indiquer si la valeur totale du patrimoine du ménage est supérieure ou inférieure à 100 000 euros. En notant « 1 » si la personne interrogée déclare que le patrimoine du ménage est supérieur à cette somme et « 0 » sinon, les réponses apportées par les cinq binômes sont :

(1,1) (1,1) (1,1) (0,0) (1,0).

Dans cet exemple, les quatre premiers binômes sont « cohérents » au sens où les conjoints fournissent la même réponse, alors que le cinquième binôme est « incohérent » car ses deux membres donnent des réponses différentes.

Deux méthodes distinctes pourraient *a priori* être utilisées pour calculer la proportion de binômes dont le patrimoine excède 100 000 euros : soit se limiter aux binômes « cohérents », ce qui donnerait une proportion de 75 %, soit calculer la moyenne de l'ensemble des réponses, ce qui

dans le cas présent conduirait à une proportion de 70 %.

Le problème est qu'en général, ces approches conduisent à des estimations biaisées du taux de prévalence d'une caractéristique donnée, parce qu'elles ne tiennent pas compte des erreurs qui affectent les réponses des individus. La méthode développée *infra* consiste à extraire l'information issue du rapprochement des réponses fournies par les individus sur une caractéristique de leur groupe pour estimer, sous certaines hypothèses, la valeur réelle de la prévalence de cette caractéristique et la proportion de réponses erronées.

Une méthode applicable sous deux conditions

Notre méthodologie n'est applicable que si les conditions suivantes sont réunies :

- au sein de chaque groupe, les mêmes questions relatives aux caractéristiques du groupe sont posées à plusieurs individus ;
- les caractéristiques sur lesquels ils sont interrogés sont binaires.

La première condition comporte deux implications. D'une part, elle exclut les situations où les questions relatives au groupe ne sont posées qu'à un seul individu. En pratique, les enquêtes ménages sont généralement conduites selon ce principe, dans la mesure où un certain nombre de questions portant sur les caractéristiques du ménage (en particulier celles qui concernent les enfants) ne sont posées qu'à la personne de référence. D'autre part, cette condition implique que les paramètres que l'on cherche à estimer ne peuvent être identifiés que pour le sous-échantillon des groupes comportant au moins deux individus, ce qui interdit leur généralisation au champ des groupes formés d'un seul individu.

La seconde condition limite le champ d'application de cette méthodologie aux questions qui peuvent être codées sous formes d'indicateurs. Cette restriction n'est pas nécessairement très contraignante, dans la mesure où de nombreuses variables peuvent être recodées de manière à ne prendre que deux modalités, qu'il s'agisse des variables continues (comme le salaire) ou des variables catégorielles qui sont naturellement ordonnées (comme le niveau de diplôme). Dans le cas des variables catégorielles non ordonnées, notre méthodologie ne peut

être appliquée que pour modéliser la probabilité d'occurrence d'une modalité par opposition à toutes les autres.

Si les deux conditions citées *supra* sont vérifiées, on observe les réponses de J individus appartenant à I groupes distincts, à une série de K questions à modalités binaires portant sur les caractéristiques de leur groupe. Chaque question ou caractéristique est « indiquée » par la lettre k ($k \in \{1, \dots, K\}$). Les individus sont ici les représentants de la direction d'entreprises interrogés au niveau des établissements et les groupes sont les entreprises auxquelles appartiennent ces établissements. Chaque dirigeant d'établissement est interrogé sur une caractéristique k de son entreprise (par exemple, « l'entreprise à laquelle appartient votre établissement fait-elle partie d'un groupe ? »).

On cherche à modéliser les réponses des dirigeants d'établissement aux questions portant sur les caractéristiques de leur entreprise. La réponse du dirigeant d'établissement j de l'entreprise i à la question k , notée y_{ijk} est supposée ne pouvoir prendre que deux modalités : y_{ijk} vaut 1 lorsque l'individu affirme que la caractéristique est présente et 0 dans le cas contraire. Deux situations sont envisageables : soit l'individu donne la réponse exacte, soit il fournit une réponse erronée (intentionnellement ou non, un aspect sur lequel nous reviendrons plus loin). En notant θ_{ik} ($\theta_{ik} \in \{0, 1\}$) la vraie valeur de la caractéristique k et λ_{ijk} ($\lambda_{ijk} \in \{0, 1\}$) la variable indiquant si le dirigeant d'établissement fournit une réponse exacte ($\lambda_{ijk} = 0$) ou erronée ($\lambda_{ijk} = 1$), sa réponse y_{ijk} peut s'écrire sous la forme :

$$y_{ijk} = \theta_{ik}(1 - \lambda_{ijk}) + (1 - \theta_{ik}) \lambda_{ijk} \quad \forall i, j, k$$

où y_{ijk} , θ_{ik} et λ_{ijk} sont des variables aléatoires binaires. Le membre de droite de l'équation prend la valeur θ_{ik} lorsque le dirigeant d'établissement donne une réponse exacte ($\lambda_{ijk} = 0$) et $1 - \theta_{ik}$ dans le cas contraire ($\lambda_{ijk} = 1$).

Sans perte de généralité, on suppose que la variable θ_{ik} , qui vaut 1 lorsque la caractéristique k est effectivement présente dans l'entreprise i , suit une loi de Bernoulli $B(p_k)$ où p_k désigne la probabilité d'observer la caractéristique k dans la population des entreprises :

$$\theta_{ik} \sim B(p_k)$$

Pour que le modèle puisse être estimé, il faut non seulement supposer que la variable θ_{ik} prend effectivement la même valeur dans tous les établissements d'une même entreprise i , mais aussi faire un certain nombre d'hypothèses probabilistes sur la distribution jointe des variables θ_{ik} (présence de la caractéristique k) et λ_{ijk} (réponse erronée).

Le cas simple : taux d'erreur homogène

On commence par supposer que tous les dirigeants d'établissement ont la même probabilité de fournir une réponse erronée, si bien que λ_{ijk} suit une loi de Bernoulli $B(\varphi_k)$ où φ_k désigne le taux d'erreur associé à la question k , le paramètre φ_k ne dépendant pas de l'entreprise i .

$$H1 : \lambda_{ijk} \sim B(\varphi_k)$$

Cette hypothèse est relativement forte, dans la mesure où elle implique que :

- le taux d'erreur est identique d'une entreprise à l'autre. Cela exclut, par exemple, que les dirigeants des établissements appartenant à des entreprises qui n'ont pas adopté les pratiques managériales les plus innovantes se trompent autant sur l'existence de ces pratiques que ceux qui appartiennent à des entreprises ayant effectivement adopté ces pratiques. Formellement, cette condition s'écrit :

$$E(\lambda_{ijk} | \theta_{ik} = 1) = E(\lambda_{ijk} | \theta_{ik} = 0) ;$$

- le taux d'erreur est le même quel que soit le dirigeant d'établissement considéré au sein d'une même entreprise, ce qui exclut la possibilité qu'il puisse exister des dirigeants se trompant plus souvent que d'autres.

Biais associé à l'estimateur naïf

Sous l'hypothèse H1, la réponse de chaque dirigeant y_{ijk} suit une loi de Bernoulli $B(q_k)$, où q_k est par définition égal à :

$$\begin{aligned} q_k &= P(y_{ijk} = 1) \\ &= p_k(1 - \varphi_k) + (1 - p_k)\varphi_k \\ &= p_k + (1 - 2p_k)\varphi_k \end{aligned}$$

En présence d'un taux d'erreur homogène, la moyenne empirique des réponses $\overline{y_{ijk}}$ (pondérée par l'inverse du nombre d'établissements de chaque entreprise, qu'on note N_i) est un estima-

teur biaisé du taux de prévalence p_k de la caractéristique k dans la population des entreprises. D'après ce qui précède, on a en effet :

$$E(\overline{y_{ijk}}) - p_k = E\left(\frac{1}{I} \sum_{i=1}^I \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ijk}\right) - p_k = q_k - p_k = \varphi_k(1 - 2p_k) \neq 0$$

Ainsi, quel que soit le taux d'erreur des réponses fournies par les dirigeants d'établissement, si $p_k > 0,5$ (c'est-à-dire si plus de la moitié des entreprises possèdent la caractéristique), la moyenne empirique totale des réponses sous-estime la prévalence de la caractéristique dans la population des entreprises. Au contraire, elle la surestime si $p_k < 0,5$. Pour un taux de prévalence réel p_k donné, le biais est d'autant plus fort que le taux d'erreur φ_k est important.

Identification des paramètres à partir des binômes de dirigeants

Les paramètres dont on cherche à estimer les valeurs sont q_k , le taux de prévalence naïf de la caractéristique k , p_k , son taux de prévalence réel dans la population des entreprises et φ_k , le taux d'erreur affectant les réponses des représentants de la direction.

Pour estimer ces différents paramètres, on suppose qu'on peut observer M binômes de cadres dirigeants appartenant à M entreprises distinctes (donc un binôme par entreprise) et interrogés sur la caractéristique k de leur entreprise. Au sein de chaque binôme m , les dirigeants sont indicés par la lettre j ($j = 1, 2$). Dans cette situation, le paramètre q_k peut être aisément estimé en calculant la moyenne empirique totale des réponses y_{mjk} à la question k :

$$\hat{q}_k = \overline{y_{mjk}} = \frac{1}{2M} \sum_{m=1}^M \sum_{j=1}^2 y_{mjk}$$

La moyenne empirique des réponses des représentants de la direction est un estimateur sans biais convergent du paramètre q_k .

Les paramètres p_k et φ_k peuvent quant à eux être estimés séparément en utilisant la cohérence des réponses de ces binômes de dirigeants. En notant y_{m1k} (resp. y_{m2k}) la réponse du premier (resp. du second) dirigeant interrogé sur la caractéristique k de l'entreprise m à laquelle appartient son établissement, trois événements sont possibles :

- les deux dirigeants répondent « 1 » ;
- les deux dirigeants répondent « 0 » ;

- les deux dirigeants donnent des réponses différentes ($y_{m1k} \neq y_{m2k}$).

Dans ce dernier cas, l'un des deux dirigeants a nécessairement donné une réponse inexacte.

Sous l'hypothèse H1, les probabilités associées à ces trois événements distincts, que l'on note respectivement α , β et γ se calculent facilement (1) (on omet l'indice k) :

$$(S) \begin{cases} \alpha = P\left(\sum_{j=1}^2 y_{mjk} = 2\right) = p(1-\varphi)^2 + (1-p)(\varphi)^2 \\ \beta = P\left(\sum_{j=1}^2 y_{mjk} = 0\right) = p(\varphi)^2 + (1-p)(1-\varphi)^2 \\ \gamma = P\left(\sum_{j=1}^2 y_{mjk} = 1\right) = 1 - \alpha - \beta = 2\varphi(1-\varphi) \end{cases}$$

Le système (S) se ramène à un système non linéaire de deux équations à deux inconnues p et φ , comportant deux solutions distinctes. Sous l'hypothèse que le taux d'erreur des dirigeants est strictement inférieur à 0,5 (ce qui signifie que les dirigeants se trompent moins d'une fois sur deux), ce système présente comme unique solution :

$$\begin{cases} \varphi = f(\gamma) = \frac{1}{2}(1 - \sqrt{1 - 2\gamma}) \\ p = g(\alpha, \gamma) = \frac{\alpha - (\varphi)^2}{1 - 2\varphi} = \frac{\alpha - \frac{1}{4}(1 - \sqrt{1 - 2\gamma})^2}{\sqrt{1 - 2\gamma}} \end{cases}$$

Estimation des paramètres par la méthode des moments

Le taux de prévalence p_k de la caractéristique k et le taux d'erreur φ_k des dirigeants d'établissement dépend de deux paramètres α et γ qui ont une contrepartie empirique :

- $\hat{\alpha}_k$: la proportion de binômes de dirigeants qui ont répondu « 1 » à la question k , soit $\frac{1}{M} \sum_{m=1}^M 1_{(y_{m1} + y_{m2} = 2)}$;

1. On notera au passage que l'estimateur du taux de prévalence de la caractéristique k qui consiste à calculer la proportion de binômes répondant « 1 » parmi l'ensemble des binômes qui fournissent des réponses cohérentes est égal à $\frac{\alpha}{\alpha + \beta}$. Il est aisé de montrer que cet estimateur est biaisé dès lors que le taux d'erreur est non nul et que le taux de prévalence réel de la caractéristique diffère de 0,5.

- $\hat{\gamma}_k$: la proportion de binômes de dirigeants qui ont donné des réponses différentes à la question k , soit $\frac{1}{M} \sum_{m=1}^M 1_{(y_{m1}+y_{m2}=1)}$.

Dans la mesure où les binômes de dirigeants sont deux à deux indépendants (puisque chaque binôme appartient à une entreprise distincte) et que les événements « les deux dirigeants répondent 1 » et « les deux dirigeants donnent des réponses différentes » sont équidistribués d'un binôme de dirigeants à l'autre, les estimateurs $\hat{\alpha}_k$ et $\hat{\gamma}_k$ convergent en probabilité vers la vraie valeur des paramètres α et γ .

Sous l'hypothèse que $\varphi < 0,5$, on peut estimer les paramètres φ et p en remplaçant α et γ par leurs contreparties empiriques $\hat{\alpha}_M$ et $\hat{\gamma}_M$ dans les fonctions f et g , soit :

$$\begin{cases} \hat{\varphi}_M = f(\hat{\gamma}_M) \\ \hat{p}_M = g(\hat{\alpha}_M, \hat{\gamma}_M) \end{cases}$$

La fonction f est continue au point γ et, sous l'hypothèse que $\varphi < 0,5$, la fonction g est continue au point (α, γ) . Par ailleurs, les estimateurs $\hat{\alpha}_M$ et $\hat{\gamma}_M$ convergent en probabilité vers α et γ . Par conséquent, en vertu du théorème de Slutsky, les estimateurs $\hat{\varphi}_M$ et \hat{p}_M convergent en probabilité vers le vrai taux de prévalence p et le vrai taux d'erreur φ :

$$\text{plim } \hat{\varphi}_M = \varphi \text{ et } \text{plim } \hat{p}_M = p$$

La méthode peut être appliquée à l'exemple introductif où l'on interrogeait cinq binômes sur la valeur de leur patrimoine. Sous l'hypothèse que tous les membres des différents binômes ont la même probabilité de fournir une réponse erronée, l'application des formules exposées ci-dessus permet d'estimer la proportion réelle de ménages disposant d'un patrimoine supérieur à 100 000 euros ainsi que la probabilité qu'un individu donne une réponse inexacte. Les contreparties empiriques des paramètres α (proportion de ménages qui donnent la même réponse) et γ (proportion de ménages qui donnent des réponses différentes) sont égales respectivement 0,6 et 0,2. On a donc :

$$\begin{cases} \hat{\varphi} = f(\hat{\gamma}) = 0,11 \\ \hat{p} = g(\hat{\alpha}, \hat{\gamma}) = 0,76 \end{cases}$$

On estime donc que 76 % des ménages ont un patrimoine supérieur à 100 000 euros et que 11 % des individus interrogés fournissent une réponse inexacte.

Extension de la méthode à des formes d'erreur plus complexes en utilisant des triplets de dirigeants

L'hypothèse d'homogénéité des taux d'erreurs φ_k d'un représentant de la direction à l'autre est sans doute trop restrictive : elle exclut par exemple la possibilité que les dirigeants dont l'entreprise possède une caractéristique donnée se trompent plus (ou moins) souvent que ceux dont l'entreprise ne possède pas cette caractéristique. Or, on ne peut exclure *a priori* que la probabilité de commettre volontairement ou involontairement une erreur diffère selon que la caractéristique est présente ou non. Ainsi, les dirigeants dont l'entreprise appartient à un groupe pourraient se tromper moins souvent sur l'appartenance de leur entreprise à un groupe que ceux dont l'entreprise n'appartient pas à un groupe, parce que les indices de cette appartenance se manifestent régulièrement dans le cadre de l'activité d'une l'entreprise. Dans ce cas, le taux d'erreur des premiers (φ_1) serait plus faible que celui des seconds (φ_0). De même, on peut imaginer qu'aux questions portant sur des pratiques managériales connotées positivement, un représentant d'une entreprise ne les ayant pas adoptées sera tenté d'« embellir » la situation, tandis qu'un dirigeant dont l'entreprise a adopté ces pratiques ne sera pas tenté de l'« enlaidir ».

Si tel est le cas, on ne peut plus faire l'hypothèse d'indépendance des variables λ_{ijk} et θ_{ik} , la probabilité pour un représentant de la direction de répondre de manière erronée à la question k pouvant être corrélée à la valeur de la caractéristique k dans l'entreprise à laquelle il appartient. En présence de ce type d'erreurs, l'identification des paramètres du modèle ne peut plus être réalisée à partir de binômes de dirigeants. Elle peut en revanche être obtenue si l'on dispose de triplets d'établissements appartenant à la même entreprise. La méthodologie peut être étendue à trois formes d'erreurs (cf. encadré 1).

Erreurs homogènes

On commence par analyser le cas simple où tous les dirigeants d'établissement ont la même probabilité de fournir une réponse erronée. Les développements précédents ont montré que sous cette hypothèse, l'identification du taux de prévalence et du taux d'erreur associés à une question portant sur une caractéristique de l'entreprise pouvait être réalisée à partir de binômes de dirigeants d'établissement. L'estimation des paramètres du modèle peut également être obtenue en utilisant des triplets de dirigeants (cf. encadré 1).

IDENTIFICATION DES PARAMÈTRES À PARTIR DES TRIPLETS DE REPRÉSENTANTS DE LA DIRECTION

Le modèle décrit dans le corps de l'article repose sur une hypothèse relativement forte : la probabilité qu'un représentant de la direction fournisse une réponse erronée à une question portant sur une caractéristique de son entreprise est supposée constante d'un individu à l'autre. La prise en compte d'hypothèses plus réalistes concernant la forme des erreurs affectant les réponses des dirigeants d'établissement nécessite de travailler non plus à partir de couples, mais à partir de triplets de représentants appartenant à la même entreprise. Cet encadré présente la manière dont ces triplets de représentants peuvent être utilisés pour estimer séparément le taux de prévalence réel des caractéristiques des entreprises et la probabilité que les réponses fournies soient erronées lorsque les erreurs sont respectivement homogènes, asymétriques et hétérogènes.

Le modèle à erreurs homogènes

La méthode exposée dans la première partie de l'article repose sur l'hypothèse que tous les représentants de la direction ont la même probabilité de donner une réponse inexacte et consiste à identifier séparément le taux de prévalence de chaque caractéristique des entreprises et la probabilité de réponse erronée. Autrement dit, la variable indiquant si le dirigeant se trompe ou non, λ_{ijk} , suit une loi de Bernoulli $B(\varphi_k)$, où φ_k désigne le taux d'erreur associé à la question k , la probabilité φ_k ne dépendant pas de l'entreprise i .

$$H1 : \lambda_{ijk} \sim B(\varphi_k)$$

La méthode présentée dans le corps de l'article pour estimer la valeur des paramètres p_k et φ_k , qui consiste à utiliser les réponses de couples de dirigeants appartenant à la même entreprise, peut être étendue au cas où l'on dispose de triplets de dirigeants.

Supposons que l'on observe m triplets de dirigeants. Dans ce cas, trois événements sont possibles :

- les trois dirigeants répondent « 1 » ;
- les trois dirigeants répondent « 0 » ;
- un dirigeant répond différemment des deux autres.

Sous l'hypothèse H1, les probabilités associées à ces trois événements s'écrivent (en omettant l'indice k) :

$$(S') \begin{cases} \alpha = P\left(\sum_{j=1}^3 Y_{mj} = 3\right) = p(1-\varphi)^3 + (1-p)(\varphi)^3 \\ \beta = P\left(\sum_{j=1}^3 Y_{mj} = 0\right) = p(\varphi)^3 + (1-p)(1-\varphi)^3 \\ \gamma = P\left(1 \leq \sum_{j=1}^3 Y_{mj} \leq 2\right) = 1 - \alpha - \beta = 3\varphi(1-\varphi) \end{cases}$$

Le système (S') se ramène à un système non linéaire de deux équations à deux inconnues p et φ , comportant deux solutions distinctes. Sous l'hypothèse que le taux d'erreur des représentants de la direction est strictement inférieur à 0,5, ce système admet une solution unique :

$$\begin{cases} \varphi = f(\gamma) = \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{3}\gamma}\right) \\ p = g(\alpha, \gamma) = \frac{\alpha - \varphi^3}{(1-\varphi)^3 - \varphi^3} = \\ = \frac{\alpha - \frac{1}{8}\left(1 - \sqrt{1 - \frac{4}{3}\gamma}\right)^3}{\left[1 - \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{3}\gamma}\right)\right]^3 - \frac{1}{8}\left[1 - \sqrt{1 - \frac{4}{3}\gamma}\right]^3} \end{cases}$$

Comme dans le cas des couples de dirigeants, les paramètres φ et p peuvent être estimés de manière convergente à partir des triplets, en remplaçant α et γ par leurs contreparties empiriques $\hat{\alpha}_M$ et $\hat{\gamma}_M$ dans les deux équations.

Le modèle à erreurs asymétriques

Dans le modèle à taux d'erreur asymétrique, on ne fait plus l'hypothèse d'indépendance des variables λ_{ijk} et θ_{ik} , la probabilité pour un dirigeant de répondre de manière erronée à la question k pouvant dorénavant être corrélée à la valeur de cette caractéristique dans l'entreprise à laquelle appartient son établissement. On choisit donc de remplacer l'hypothèse H1 par l'hypothèse H2 qui stipule que conditionnellement à la valeur θ_{ik} de la caractéristique k dans l'entreprise i , l'erreur du dirigeant j suit une loi de Bernoulli $B(\varphi_k(\theta_{ik}))$:

$$H2 : \lambda_{ijk} \sim B(\varphi_k(\theta_{ik}))$$

Dans la suite du texte, on notera $\varphi_{k1} = \varphi_k(1)$, la probabilité d'erreur des dirigeants dont l'entreprise possède la caractéristique k et $\varphi_{k0} = \varphi_k(0)$, la probabilité d'erreur des dirigeants dont l'entreprise ne possède pas cette caractéristique.

En présence d'un taux d'erreur asymétrique, la moyenne empirique des réponses \bar{y}_{ijk} (pondérée par l'inverse du nombre d'établissements de chaque entreprise) est un indicateur biaisé du taux de prévalence p_k de la caractéristique k dans la population des entreprises, puisque :

$$E(\bar{y}_{ijk}) - p_k = q_k - p_k = \varphi_{k0} - p_k(\varphi_{k0} + \varphi_{k1}) \neq 0$$

L'estimateur naïf du taux de prévalence de la caractéristique k tendra donc à surestimer le taux de prévalence réel si et seulement si :

$$\frac{\varphi_{k0}}{\varphi_{k1}} > \frac{p_k}{1-p_k}$$

ce qui est vrai en particulier lorsque $p_k < 0,5$ et $\varphi_{k0} > \varphi_{k1}$. Autrement dit, lorsque moins de la moitié des entreprises possède la caractéristique et que les dirigeants de celles qui ne la possèdent pas se trompent plus souvent que les dirigeants de celles qui en disposent, l'estimateur naïf surestimerait le taux de prévalence réel. Inversement, l'estimateur naïf du taux de prévalence

Encadré 1 (suite)

de la caractéristique k tendra à sous-estimer le taux de prévalence réel si et seulement si :

$$\frac{\varphi_{k0}}{\varphi_{k1}} < \frac{p_k}{1-p_k}$$

Outre le taux de prévalence naïf (estimé à partir de la moyenne empirique des réponses y_{ijk}), trois paramètres doivent être estimés pour chaque question k : p_k , φ_{k0} et φ_{k1} . L'utilisation des couples de dirigeants n'est pas suffisante pour les identifier séparément (sauf dans le cas où $\varphi_{k0} = \varphi_{k1}$). Cela devient possible en revanche lorsqu'on s'intéresse aux réponses données par les triplets de dirigeants.

Quatre événements peuvent être distingués :

- les trois dirigeants répondent « 1 » ;
- les trois dirigeants répondent « 0 » ;
- deux dirigeants répondent « 1 » et le troisième répond « 0 » ;
- deux dirigeants répondent « 0 » et le troisième répond « 1 ».

Sous l'hypothèse H2, les probabilités associées à ces quatre événements s'écrivent (en omettant l'indice k) :

$$(S^n) \begin{cases} \alpha = P\left(\sum_{j=1}^3 y_{mj} = 3\right) = p(1-\varphi_1)^3 + (1-p)(\varphi_0)^3 \\ \beta = P\left(\sum_{j=1}^3 y_{mj} = 0\right) = p(\varphi_1)^3 + (1-p)(1-\varphi_0)^3 \\ \delta = P\left(\sum_{j=1}^3 y_{mj} = 2\right) = 3\left[p(1-\varphi_1)^2\varphi_1 + (1-p)(\varphi_0)^2(1-\varphi_0)\right] \\ \sigma = P\left(\sum_{j=1}^3 y_{mj} = 1\right) = 3\left[p(1-\varphi_1)(\varphi_1)^2 + (1-p)\varphi_0(1-\varphi_0)^2\right] \end{cases}$$

Le système (S^n) se ramène à un système non linéaire de trois équations à trois inconnues p , φ_0 et φ_1 .

Ce système admet deux solutions distinctes. Sous l'hypothèse que $\varphi_0 < 0,5$ et que $\varphi_1 < 0,5$, il présente une solution unique (on peut en effet montrer que $\forall \varphi_0 \in [0; 0,5[$, $\forall \varphi_1 \in [0; 0,5[$, on a : $K \neq 0$ et

$$(\varphi_0)^3 \neq (1-\varphi_1)^3 :$$

$$\begin{cases} \varphi_0 = f(\alpha, \beta, \delta) = \frac{1}{2K} \left(L + \sqrt{L^2 - 4KM} \right) \\ \varphi_1 = g(\alpha, \beta, \delta) = 1 - \frac{1}{2K} \left(L + \sqrt{L^2 - 4KM} \right) \\ p = h(\alpha, \beta, \delta) = \frac{(\varphi_0)^3 - \alpha}{(\varphi_0)^3 - (1-\varphi_1)^3} \end{cases}$$

où :

$$\begin{cases} K = 1 + \beta^2 + \delta(\delta - 1) - 2\beta(\delta + 1) + \alpha(4\alpha - 4\beta + 4\delta - 5) \\ L = \alpha(6\alpha - 3\beta + 5\delta - 6) + \delta(\delta - \beta + 1) \\ M = 3\alpha(\alpha + \beta + \delta - 1) + \delta^2 \end{cases}$$

Les fonctions f , g et h étant continues au point (α, β, δ) , les paramètres φ_0 , φ_1 et p peuvent être estimés de manière convergente en remplaçant α , β et δ par leurs contreparties empiriques $\hat{\alpha}_M$, $\hat{\beta}_M$, et $\hat{\delta}_M$ dans les trois équations du système.

Le modèle à erreurs hétérogènes

Dans le modèle à taux d'erreur hétérogène, on fait l'hypothèse que conditionnellement à la valeur θ_{ik} prise

par la caractéristique k dans l'entreprise i , la probabilité individuelle de donner une réponse erronée à une question k dévie aléatoirement de la moyenne φ_k , la déviation ε_{ijk} étant propre à chaque dirigeant d'établissement, de moyenne nulle et indépendante de θ_{ik} :

$$H3 : \lambda_{ijk} \sim B(\varphi_k(\theta_{ik}) + \varepsilon_{ijk}) \text{ avec } \varepsilon_{ijk} \perp \theta_{ik} \text{ et } E(\varepsilon_{ijk}) = 0$$

Sous cette hypothèse, la réponse y_{ijk} de chaque dirigeant suit une loi de Bernoulli $B(q_k)$, où la probabilité q_k est égale à :

$$q_k = E(y_{ijk}) = E_\varepsilon \left[E_y(y_{ijk} | \varepsilon_{ijk}) \right]$$

Or par définition :

$$\begin{aligned} E_y(y_{ijk} | \varepsilon_{ijk}) &= E_\theta \left[E_y(y_{ijk} | \varepsilon_{ijk}, \theta_{ik}) \right] \\ &= P(y_{ijk} = 1 | \varepsilon_{ijk}, \theta_{ik} = 1)P(\theta_{ik} = 1 | \varepsilon_{ijk}) + \\ &\quad P(y_{ijk} = 1 | \varepsilon_{ijk}, \theta_{ik} = 0)P(\theta_{ik} = 0 | \varepsilon_{ijk}) \\ &= P(y_{ijk} = 1 | \varepsilon_{ijk}, \theta_{ik} = 1)P(\theta_{ik} = 1) + \\ &\quad P(y_{ijk} = 1 | \varepsilon_{ijk}, \theta_{ik} = 0)P(\theta_{ik} = 0) \\ &= \left[1 - P(\lambda_{ijk} = 1 | \varepsilon_{ijk}, \theta_{ik} = 1) \right] p_k + P(\lambda_{ijk} = \\ &\quad 1 | \varepsilon_{ijk}, \theta_{ik} = 0)(1 - p_k) \\ &= p_k(1 - \varphi_{k1} - \varepsilon_{ijk}) + (1 - p_k)(\varphi_{k0} + \varepsilon_{ijk}) \end{aligned}$$

On en déduit que :

$$\begin{aligned} q_k &= E_\varepsilon \left[E_y(y_{ijk} | \varepsilon_{ijk}) \right] \\ &= E_\varepsilon \left[p_k(1 - \varphi_{k1} - \varepsilon_{ijk}) + (1 - p_k)(\varphi_{k0} + \varepsilon_{ijk}) \right] \\ &= p_k \left[(1 - \varphi_{k1} - E_\varepsilon(\varepsilon_{ijk})) \right] + (1 - p_k) \left[(\varphi_{k0} + E_\varepsilon(\varepsilon_{ijk})) \right] \\ &= p_k(1 - \varphi_{k1}) + (1 - p_k)\varphi_{k0} \end{aligned}$$

puisque, par hypothèse, $E(\varepsilon_{ijk}) = 0$.

Les réponses des dirigeants d'entreprises suivent donc la même distribution que dans le cas des taux d'erreurs asymétriques, si bien que l'estimateur naïf du taux de prévalence de la caractéristique k est affecté du même biais que celui mis en évidence dans le cas du modèle à erreurs asymétriques.

Dans la mesure où les déviations individuelles ε_{ijk} sont indépendantes de la valeur θ_{ik} de la caractéristique k de l'entreprise i , on peut aisément montrer que la stratégie d'identification exposée dans la section précédente est toujours valable pour estimer de manière convergente les paramètres p_k , φ_{k0} et φ_{k1} . Comme dans le cas des taux d'erreurs asymétriques, les estimateurs \hat{p}_k et $\hat{\varphi}_{k0}$ et $\hat{\varphi}_{k1}$ sont la solution du système non linéaire (S^n) , dont les paramètres α , β et δ peuvent être calculés en utilisant les réponses des triplets de dirigeants.

On notera que dans le cas où les taux d'erreurs sont hétérogènes mais symétriques (soit $\varphi_{k0} = \varphi_{k1} = \varphi_k$), les paramètres p_k et φ_k peuvent être estimés de manière convergente en utilisant la méthode présentée dans le cas du modèle à erreurs homogènes.

Erreurs asymétriques

Nous nous intéressons ensuite au cas des erreurs asymétriques, qui signifient que la probabilité pour un dirigeant de répondre de manière erronée à une question portant sur une caractéristique k dépend de la valeur de cette caractéristique dans son entreprise. En vertu de cette hypothèse, la fiabilité des réponses pourra varier selon que l'entreprise possède ou non la caractéristique. Formellement, cela revient à remplacer l'hypothèse H1 par l'hypothèse H2 qui stipule que conditionnellement à la valeur θ_{ik} de la caractéristique k dans l'entreprise i , l'erreur du dirigeant j suit une loi de Bernoulli $B(\varphi_k(\theta_{ik}))$:

$$H2 : \lambda_{ijk} \sim B(\varphi_k(\theta_{ik}))$$

Dans la suite du texte, $\varphi_{k1} = \varphi_k(1)$ exprime la probabilité d'erreur des dirigeants dont l'entreprise possède la caractéristique k et $\varphi_{k0} = \varphi_k(0)$, la probabilité d'erreur des dirigeants dont l'entreprise ne possède pas cette caractéristique. Les triplets de représentants de la direction peuvent être utilisés pour estimer séparément le taux de prévalence réel p_k de chaque caractéristique, le taux d'erreur φ_{k1} des dirigeants dont l'entreprise possède cette caractéristique et le taux d'erreur φ_{k0} de ceux dont l'entreprise ne la possède pas (cf. encadré 1).

Erreurs hétérogènes

Le modèle asymétrique impose que la probabilité de se tromper est la même pour tous les représentants de la direction d'une même entreprise, indépendamment de l'établissement où ils exercent leur activité. Cette hypothèse peut paraître restrictive, dans la mesure où on ne peut exclure que certains dirigeants connaissent moins bien leur entreprise que d'autres, ou encore qu'ils aient tendance à se montrer systématiquement plus « vantards » (ou pessimistes) que leurs homologues, en donnant de leur entreprise un profil excessivement flatteur (ou dévalué). Le fait que les individus interrogés dans le cadre de l'enquête *REPONSE* exercent des fonctions variées (direction d'établissement, direction des ressources humaines, etc.) au sein de la direction de leur entreprise rend cette hypothèse plausible dans le contexte empirique que nous étudions. Cependant la prise en compte d'erreurs hétérogènes n'invalide pas l'estimateur proposé ci-dessus dans le cas des erreurs asymétriques (cf. encadré 1).

L'hypothèse d'hétérogénéité des taux d'erreurs revient à relâcher l'hypothèse H2 en supposant

que conditionnellement à la valeur θ_{ik} prise par la caractéristique k dans l'entreprise i , la probabilité individuelle de donner une réponse erronée à une question k dévie aléatoirement de la moyenne φ_k , la déviation ε_{ijk} étant propre à chaque dirigeant, de moyenne nulle et indépendante de θ_{ik} :

$$H3 : \lambda_{ijk} \sim B(\varphi_k(\theta_{ik}) + \varepsilon_{ijk}) \text{ avec } \varepsilon_{ijk} \perp \theta_{ik} \text{ et } E(\varepsilon_{ijk}) = 0$$

L'hypothèse d'indépendance entre les variables ε_{ijk} et θ_{ik} n'impose pas en réalité de restriction sur la forme de l'hétérogénéité des taux d'erreurs, dans la mesure où la dépendance entre λ_{ijk} et θ_{ik} est entièrement capturée par l'asymétrie des taux d'erreurs φ_{k0} et φ_{k1} .

La probabilité d'erreur d'un dirigeant d'établissement dont l'entreprise possède effectivement la caractéristique k s'écrit $\varphi_{k1} + \varepsilon_{ijk}$; celle d'un dirigeant dont l'entreprise ne possède pas la caractéristique, $\varphi_{k0} + \varepsilon_{ijk}$. L'hypothèse d'hétérogénéité des taux d'erreur permet de rendre compte de l'existence de dirigeants moins bien informés ou plus « vantards » et qui par conséquent se « trompent » davantage en moyenne que leurs homologues de la même entreprise. Formellement, une telle situation se traduirait par une corrélation positive entre les déviations individuelles ε_{ijk} d'une question à l'autre : $\text{Cov}(\varepsilon_{ijk}, \varepsilon_{ijl}) > 0$.

Sous l'hypothèse H3, les réponses des dirigeants d'entreprises suivent la même distribution que dans le cas des taux d'erreurs asymétriques. Dans la mesure où les déviations individuelles ε_{ijk} sont indépendantes de la valeur θ_{ik} de la caractéristique k de l'entreprise i , la stratégie d'identification exposée dans le cas des erreurs asymétriques est toujours valable pour estimer de manière convergente les paramètres p_k , φ_{k0} et φ_{k1} (cf. encadré 1).

Les estimateurs du taux de prévalence et des taux d'erreur que nous proposons ne sont qu'asymptotiquement sans biais. Pour de petits échantillons tels que ceux utilisés dans la suite de cette étude, il est possible que ces estimateurs ne puissent être calculés ou ne donnent pas systématiquement des valeurs admissibles pour les paramètres (on peut par exemple avoir $\hat{p} > 1$).

Pour traiter ce problème, nous utilisons la méthode du *bootstrap* en appliquant nos formules à 1 000 sous-échantillons issus de tirages

avec remise. Nous calculons ensuite la valeur moyenne des paramètres estimés, ainsi que les intervalles de confiance non paramétriques au seuil de 5 %, en ne retenant que les répliques qui donnent des valeurs admissibles pour les paramètres (soit $\hat{p} \in [0; 1]$, $\hat{\varphi} \in [0; 0,5[$, $\hat{\varphi}_0 \in [0; 0,5[$ et $\hat{\varphi}_1 \in [0; 0,5[$).

Les propriétés de cette procédure sont analysées grâce à une série de simulations de Monte-Carlo, en faisant varier la taille de l'échantillon et en calculant nos estimateurs par *bootstrap*. Les résultats de cet exercice (non reportés ici) montrent que ces estimateurs se comportent bien et que les biais éventuels sont négligeables lorsque la taille de l'échantillon dépasse 100 binômes ou triplets et que les taux d'erreurs ne sont pas trop proches de 0,5.

Les conditions de validité de cette procédure seront vérifiées pour la plupart des estimations réalisées dans cette étude, la taille des échantillons étant toujours supérieure à 100 binômes ou triplets, les taux d'erreur estimés ne dépassant que rarement 0,3 et la quasi-totalité des répliques donnant des valeurs admissibles des paramètres. Nous pouvons donc nous montrer raisonnablement confiants quant à la pertinence de nos estimations.

Une application à l'enquête *REPONSE*

Analogue à l'enquête britannique *WIRSWERS*, l'enquête *Relations professionnelles et négociations d'entreprises (REPONSE)* est réalisée à intervalles de sept ans par la Dares (ministère de l'Économie, de l'Industrie et de l'Emploi et ministère du Travail, des Relations sociales, de la Famille, de la Solidarité et de la Ville) et l'Insee auprès d'environ 3 000 établissements d'au moins 20 salariés appartenant au secteur marchand non agricole. Cette enquête est largement utilisée tant par la Dares que par les chercheurs pour le suivi des relations professionnelles et des modes d'organisation du travail (cf. Askenazy *et al.* (2006) pour la vague 1998, Behaghel *et al.* (2007) pour la vague 2004).

L'enquête *REPONSE* est conduite au début de l'année $t+1$ (soit 1999 et 2005) et porte sur l'année t (c'est-à-dire 1998 et 2004). Elle permet de croiser les points de vue des différents acteurs de l'entreprise : représentants de la direction

(président, directeur général, directeur d'usine, directeur du personnel, etc.), représentants du personnel (lorsqu'il y en a) et salariés. Ce travail utilise exclusivement la partie de l'enquête qui porte sur les représentants de la direction interrogés au niveau des établissements de l'entreprise (avec un maximum d'un représentant par établissement). Les représentants de la direction enquêtés occupent des fonctions variées qui vont de la direction de l'entreprise à la gestion des ressources humaines en passant par la direction d'établissement (2). À quelques exceptions près, ils se revendiquent bien comme responsables au sein de leurs établissements et occupent la plupart du temps un poste de représentant de la direction chargé des relations avec les salariés, cette catégorie représentant la cible de l'enquête.

Les questions sont posées aux représentants de la direction au cours d'un entretien en face à face de 75 minutes en moyenne. Elles sont variées et portent aussi bien sur la situation économique de leur entreprise ou de leur établissement que sur les pratiques managériales, la description des instances représentatives du personnel, les négociations salariales ou encore le climat social qui prévaut au sein de leur établissement. Le taux d'acceptation est relativement important pour ce type d'enquêtes puisqu'il est de l'ordre de 60 % : ainsi, pour 4 713 établissements sélectionnés aléatoirement, identifiés et dans le champ de l'enquête, 2 930 ont répondu pour la vague 2004/2005.

Le plan de sondage de l'enquête *REPONSE* consiste en un tirage aléatoire des établissements dans le répertoire SIRENE, réalisé selon une allocation proportionnelle à l'effectif salarié par strate de taille (en cinq tranches (3)) et, à tranche de taille donnée, au nombre d'établissements par strate de secteur (en 15 postes). Le calcul de la pondération « établissement » fait intervenir deux composantes : d'une part, le calcul des poids de tirage, effectué en rapportant, pour chaque taille et secteur croisés, le nombre d'établissements dans le champ de l'enquête au nombre d'établissements dans le répertoire SIRENE au moment de l'enquête ; d'autre part, une correction de la non-réponse qui conjugue

2. Les fonctions exercées par les représentants de la direction interrogés dans le cadre de l'enquête sont : président, directeur général, secrétaire général, directeur général adjoint, directeur administratif, directeur financier, directeur d'établissement, directeur d'usine, directeur d'agence, directeur ou responsable des ressources humaines, directeur du personnel, chef du personnel, directeur des relations sociales.

3. De 20 à 49 salariés, de 50 à 99 salariés, de 100 à 249 salariés, de 250 à 499 salariés, 500 salariés ou plus.

deux méthodes : le calage sur marge et l'estimation de la probabilité de réponse selon un modèle de régression logistique.

Le calcul de la pondération « entreprise »

La conséquence de ce plan de sondage est que l'échantillon pondéré issu de l'enquête *REPONSE* est représentatif de la population des établissements mais pas de la population des entreprises, puisqu'une entreprise a une probabilité d'autant plus forte d'y figurer qu'elle contient un grand nombre d'établissements. Un peu moins de la moitié des entreprises représentées dans l'échantillon de l'enquête *REPONSE* (43 % en 1998 et 45 % en 2004) ont plus d'un établissement d'au moins 20 salariés ; cette proportion n'est que de 15 % dans la population des entreprises entrant dans le champ de l'enquête (cf. tableau 1).

Pour tenir compte de cette hétérogénéité dans la probabilité d'inclusion des entreprises, nous avons attribué à chaque entreprise de l'échantillon un poids de sondage calculé selon la méthode du partage des poids mise au point par Lavallée (1995, 2002) (cf. encadré 2). Le poids attribué à chaque entreprise est d'autant plus élevé que celle-ci contient peu d'établissements et que les poids de sondage des établissements qui ont été sélectionnés au sein de cette entreprise sont importants. Dans la mesure où nos estimations visent à nous renseigner sur les caractéristiques des entreprises à partir de binômes ou de triplets de dirigeants d'établissement, l'ensemble des estimations réalisées dans la suite de cet article font intervenir ces poids de sondage (4).

La répartition obtenue est relativement proche de la structure réelle des entreprises dans la population : la proportion d'entreprises ayant plus de deux établissements d'au moins 20 salariés est de l'ordre de 10 à 15 %.

Construction de l'échantillon d'étude

Pour mettre en œuvre l'approche méthodologique décrite plus haut, nous avons restreint l'échantillon de travail aux entreprises dont au moins deux établissements ont été sélectionnés dans l'enquête *REPONSE*. Ces dernières peuvent être identifiées grâce aux premiers chiffres de l'identifiant SIRET de chaque établissement, qui sont communs à tous les établissements d'une même entreprise (5). Nous avons alors généré un identifiant aléatoire « entreprise » pour chaque établissement, qui a été apparié à la version anonyme de *REPONSE* (grâce à l'identifiant « établissement ») ; il a ainsi été possible de mobiliser, en assurant l'anonymat, des informations concernant le statut du responsable enquêté.

Parmi les 2 589 entreprises présentes dans l'enquête *REPONSE* 1998, 191 sont représentées par au moins deux établissements distincts ; en 2004, c'est le cas de 205 entreprises sur 2 476 (cf. tableau 2). En général, on dispose pour ces entreprises de deux établissements exactement,

4. Les poids de sondage « entreprise » étant calculés à partir des poids de sondage « établissement » présents dans les fichiers de l'enquête, ils tiennent également compte de la non-réponse.
5. L'échantillon de départ ayant été sélectionné par tirage aléatoire au sein de la population des établissements, l'établissement-siège d'une entreprise donnée ne figure pas systématiquement parmi les établissements initialement présents dans l'enquête *REPONSE*.

Tableau 1
Comparaison de la structure par établissements des entreprises présentes dans les échantillons des enquêtes *REPONSE* 1998 et 2004 et dans la population de référence

Nombre d'établissements de 20 salariés et plus qui composent réellement l'entreprise	REPONSE 1998			REPONSE 2004		
	Nombre d'entreprises	% dans l'échantillon	% estimé dans la population	Nombre d'entreprises	% dans l'échantillon	% estimé dans la population
Un seul établissement de 20 salariés et plus	1 470	56,8	85,1	1 362	55,0	88,7
Plusieurs établissements de 20 salariés et plus	1 119	43,2	14,9	1 114	45,0	11,4
Total	2 589	100	100	2 476	100	100

Lecture : parmi les 2 589 entreprises présentes dans l'enquête *REPONSE* 1998, 1 470 ne comportent qu'un seul établissement de 20 salariés et plus (soit 56,8 % de l'échantillon) et 1 119 en comportent plus d'un (soit 43,2 % de l'échantillon). Dans la population globale des entreprises, les proportions estimées sont respectivement de 85,1 et 14,9 %.

Pour estimer la répartition des entreprises dans la population de référence (et non dans l'échantillon, qui surpondère les entreprises multi-établissements), on utilise les poids de sondage « entreprise » dont le calcul est présenté dans l'encadré 2.

Champ : entreprises ayant au moins un établissement de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes *REPONSE* 1998 et 2004.

Encadré 2

CALCUL DES PONDÉRATIONS « ENTREPRISE »

Le plan de sondage du volet « représentants de la direction » de l'enquête *REPONSE* correspond à un tirage aléatoire réalisé selon une allocation proportionnelle à l'effectif salarié par strate de taille (en cinq tranches) et, à tranche de taille donnée, au nombre d'établissements par strate de secteur (en 15 postes).

Une pondération a été attribuée à chaque établissement pour tenir compte du poids de tirage et de la non-réponse, qui a été corrigée en utilisant deux méthodes distinctes : le calage sur marges, le plus souvent selon la méthode du *Logit*, et l'estimation de la probabilité de réponse selon un modèle de régression logistique.

Dans la mesure où le plan de sondage a porté sur les établissements, l'échantillon pondéré est représentatif de la population des établissements, mais pas de la population des entreprises. En effet, une entreprise a d'autant plus de chances de figurer dans l'échantillon de *REPONSE* qu'elle contient de nombreux établissements et compte un nombre important de salariés. Dans la mesure où les calculs réalisés dans notre étude visent à nous renseigner sur les caractéristiques des entreprises à partir de binômes ou de triplets de représentants de la direction, il est nécessaire de pondérer nos estimateurs par la probabilité d'inclusion des entreprises utilisées.

Ces probabilités peuvent être estimées à partir de la méthode dite du « partage des poids », mise au point par Lavallée (1995, 2002). Cette méthode s'applique lorsque, au lieu de disposer d'une base de sondage contenant les unités de collecte souhaitées, on a accès à une base de sondage contenant des unités liées d'une certaine façon à la liste d'unités de collecte. La difficulté consiste alors à attribuer un poids d'estimation (ou une probabilité de sélection) aux unités étudiées de la population cible. On donne à cette approche le nom de sondage indirect (pour une présentation de la méthode de partage des poids et de ses applications, voir Ardilly (2006), section II.4.5).

Pour les besoins de notre étude, on cherche à estimer le poids de sondage des entreprises dont les établissements ont été sélectionnés dans l'échantillon

de *REPONSE*. En indexant par la lettre j les établissements de l'entreprise i et en notant N_i le nombre total d'établissements de 20 salariés ou plus de l'entreprise i , n_i le nombre d'établissements de l'entreprise i qui figurent dans l'échantillon de l'enquête *REPONSE* ($n_i \leq N_i$) et w_{ij} le poids de sondage attribué à l'établissement j de l'entreprise i , le poids de sondage w_i de l'entreprise i peut être estimé au moyen de la formule suivante :

$$\hat{w}_i = \frac{1}{N_i} \sum_{j=1}^{n_i} w_{ij}$$

Sous l'hypothèse qu'il n'existe aucune entreprise sans établissement (ce qui est bien le cas), l'utilisation des pondérations \hat{w}_i permet d'estimer sans biais asymptotique la moyenne \bar{Y} d'une caractéristique Y au sein d'une population d'entreprises à partir des observations y_i fournies par les I entreprises de l'échantillon :

$$\hat{\bar{Y}} = \frac{1}{\sum_{i=1}^I \hat{w}_i} \sum_{i=1}^I \hat{w}_i y_i$$

Le poids de sondage \hat{w}_i de l'entreprise i est d'autant plus important que cette entreprise contient peu d'établissements et que la somme des poids de sondage des établissements *effectivement* sélectionnés au sein de cette entreprise est élevée.

Les paramètres n_i (nombre d'établissements de l'entreprise i qui figurent dans l'enquête *REPONSE*) et w_{ij} (poids de sondage de chaque établissement) sont fournis par les données de l'enquête. Le paramètre N_i (nombre total d'établissements de l'entreprise i de 20 salariés ou plus au moment de l'enquête) ne peut en revanche être connu pour chaque entreprise qu'en remontant à la source de tirage SIRENE. La Dares (ministère de l'Économie, des Finances et de l'Emploi et ministère du Travail, des Relations sociales et de la Solidarité) nous a transmis le fichier indiquant le nombre d'établissement de 20 salariés ou plus des entreprises présentes dans les enquêtes *REPONSE* 1998 et 2004.

mais on trouve jusqu'à 37 établissements d'une même enseigne de distribution dans l'échantillon de l'enquête.

Pour les besoins de notre étude, nous avons extrait deux sous-échantillons distincts de la base des entreprises représentées par au moins deux établissements :

- l'échantillon A est une base de binômes de dirigeants d'établissement, qui regroupe l'en-

semble des binômes formés à partir d'entreprises représentées par deux établissements exactement, ainsi que tous les binômes pouvant être obtenus par tirage aléatoire sans remise parmi les entreprises représentées par trois établissements ou plus, dans la limite d'un binôme par entreprise. Cette base contient 191 binômes de dirigeants d'établissement (et autant d'entreprises distinctes) pour l'année 1998 et 205 binômes pour l'année 2004 ;

- l'échantillon B est une base de triplets de dirigeants d'établissement, construite selon le même principe. Cette base contient 64 triplets pour l'année 1998 et 75 triplets pour l'année 2004.

Pour étudier la qualité des réponses fournies par les représentants de la direction, nous avons retenu des questions portant sur les aspects généraux de l'entreprise à laquelle se rattache l'établissement du dirigeant interrogé. Ces questions peuvent être regroupées selon leur degré de généralité : les aspects généraux de l'entreprise et des aspects plus spécifiques (cf. encadré 3).

Représentativité des échantillons et statistiques descriptives

La construction des deux échantillons de travail n'est pas neutre du point de vue de l'interprétation de nos résultats. En effet, si l'échantillon formé de l'ensemble des entreprises présentes dans les vagues 1998 et 2004 de l'enquête *REPONSE* peut être considéré, une fois muni de la pondération « entreprise » (cf. encadré 2), représentatif de la population des entreprises (limitées aux établissements d'au moins 20 salariés), ce n'est pas nécessairement le cas pour les échantillons formés de binômes (échantillon A) ou de triplets (échantillon B) de représentants de la direction d'une même entreprise. En effet, ces échantillons ne peuvent pas par définition contenir d'entreprises mono-établissement. L'échantillon A doit donc être considéré comme représentatif de la population des entreprises formées d'au moins deux établissements de 20 salariés et plus et l'échantillon B comme représentatif de la population des entreprises formées d'au moins trois établissements de 20 salariés et plus.

La répartition sectorielle des entreprises des échantillons A et B diffère sensiblement de celle des entreprises entrant dans le champ de l'enquête *REPONSE* (cf. tableau 3). Les entreprises pour lesquelles on dispose d'un binôme ou d'un triplet de représentants de la direction appartient, par exemple, plus souvent au secteur industriel (hors industries agricoles et alimentaires) et moins souvent au secteur de la construction.

Il est possible de comparer les caractéristiques binaires des entreprises représentées dans les différents échantillons, telles qu'elles sont décrites par les représentants de la direction interrogés au niveau des établissements (cf. tableau 4). Les moyennes sont calculées à l'issue d'une procédure en deux étapes : on commence par calculer la réponse moyenne fournie par les représentants de la direction d'une même entreprise avant de calculer la moyenne générale pondérée par le poids de sondage propre à chaque entreprise. Les caractéristiques des entreprises multi-établissements diffèrent sensiblement de celles de l'ensemble des entreprises : les premières sont plus grandes, appartiennent plus souvent à un groupe, sont plus souvent cotées en Bourse et sont moins souvent aux mains de familles que les secondes ; elles sont également plus souvent concernées par l'utilisation d'une démarche de qualité totale et l'adoption d'une norme ISO ; enfin, la représentation syndicale y est beaucoup plus forte.

En revanche, les caractéristiques déclarées des entreprises pour lesquelles on dispose de binômes ou de triplets de représentants de la direction apparaissent très proches de celles des entreprises multi-établissements. Par construction, les différences sont infimes lorsqu'on considère l'échantillon formé des binômes de représentants (les deux échantillons pondérés

Tableau 2
Nombre d'établissements effectivement sélectionnés parmi les entreprises présentes dans les échantillons des enquêtes *REPONSE* 1998 et 2004

Nombre d'établissements de l'entreprise présents dans l'échantillon	<i>REPONSE</i> 1998		<i>REPONSE</i> 2004	
	Nombre d'entreprises	% de l'échantillon	Nombre d'entreprises	% de l'échantillon
1	2 398	92,6	2 271	91,7
2	127	4,9	130	5,3
3	26	1,0	38	1,5
4 ou plus	38	1,5	37	1,5
Total	2 589	100	2476	100

Lecture : parmi les 2 589 entreprises présentes dans l'enquête *REPONSE* 1998, 2 398 n'ont qu'un seul établissement figurant dans l'échantillon, ce qui représente 92,6 % de l'échantillon.

Champ : entreprises ayant au moins un établissement de 20 salariés et plus du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes *REPONSE* 1998 et 2004.

Encadré 3

SÉLECTION ET TRANSFORMATION DES VARIABLES D'INTÉRÊT

Pour étudier la qualité des réponses fournies par les représentants de la direction, nous avons retenu des questions portant sur les caractéristiques de l'entreprise à laquelle se rattache l'établissement du dirigeant interrogé. Ces questions peuvent être regroupées selon leur degré de généralité.

Les questions concernant les aspects généraux de l'entreprise sont au nombre de sept :

- l'entreprise comporte-t-elle plusieurs établissements (MULTI) ?
- Compte-t-elle plus de 1 000 salariés (SALENT) ?
- Appartient-elle à un groupe (LIENS) ?
- Est-elle cotée en Bourse (BOURSE) ?
- Les salariés détiennent-ils une part de l'entreprise (ACTIONS) ?
- Ont-ils accès à un plan d'épargne entreprise ou inter-entreprise (PEE*) ?
- Ont-ils accès à un plan d'épargne-retraite (PER*) ?

Les variables marquées d'un * ne sont disponibles que pour l'enquête *REPONSE* 2004.

Les questions portant sur des aspects plus spécifiques de l'entreprise sont quant à elles au nombre de onze :

- la principale catégorie d'actionnaires est-elle une famille (K1) ?
- L'entreprise a-t-elle connu un changement de propriétaire au cours des trois dernières années (CHPROP) ?
- Des représentants des salariés siègent-ils au CA/CS de l'entreprise (RPCA) ?
- L'entreprise est-elle affiliée à une fédération patronale (FEDPAT) ?
- Y a-t-il des délégués syndicaux dans l'entreprise (DSENTR) ?
- Y a-t-il des délégués CGT (CGTPE), CFDT (CFDTPE), CFTC (CFTCPE), FO (FOPE), SUD (SUDPE*), UNSA

(UNSAPE*) dans l'entreprise ?

Nous retenons également trois questions portant sur des technologies ou méthodes d'organisation posées pour l'établissement mais dont l'usage est en principe commun à l'entreprise :

- une démarche de qualité totale est-elle utilisée dans votre établissement (QUALTOT) ?
- Si oui s'agit-il une normalisation ISO (ISO) ?
- Un progiciel de gestion intégrée est-il utilisé dans votre établissement (PGIERP*) ?

Telles qu'elles se présentent dans l'enquête *REPONSE*, ces questions sont à modalités multiples. Pour les besoins de l'analyse, nous avons dû regrouper ces catégories afin de former des indicatrices prenant les valeurs 0 ou 1 (cf. annexe). Il faut noter que les indicatrices ainsi construites ne sont pas disponibles pour tous les binômes ou triplets de dirigeants. Il y a à cela deux raisons : d'une part, un certain nombre de questions n'ont été posées que pour des entreprises présentant des caractéristiques particulières (la question « des représentants des salariés siègent-ils au CA/CS de l'entreprise ? » n'est par exemple posée qu'aux dirigeants qui ont déclaré que les salariés détiennent une part du capital de l'entreprise) ; d'autre part, certains dirigeants n'ont pas souhaité répondre ou ont déclaré ne pas connaître la réponse à la question. Nos estimations ne sont réalisées que sur les binômes ou triplets de dirigeants dont les réponses sont renseignées pour chacun des membres du binôme ou du triplet. Par conséquent, les échantillons utilisés sont susceptibles de varier légèrement d'une question à une autre. Le tableau présenté en annexe montre cependant que ces variations sont limitées et qu'elles sont dues pour l'essentiel au fait que certaines questions peuvent être sans objet compte tenu du profil de l'entreprise. L'existence de valeurs manquantes ne nous paraît donc pas représenter une menace importante pour la fiabilité de nos estimations.

Tableau 3
Répartition sectorielle des entreprises présentes dans les enquêtes *REPONSE* 1998 et 2004 selon le nombre d'établissements de la même entreprise figurant dans l'échantillon

En %

Secteur de l'entreprise	Ensemble des entreprises de l'échantillon	Entreprises représentées par au moins deux établissements	Entreprises représentées par au moins trois établissements
Industries agricoles et alimentaires (EB)	3,8	7,9	3,5
Industries hors IAA (EC à EG)	25,7	35,6	34,0
Construction (EH)	10,2	4,4	1,4
Commerce (EJ)	18,6	11,1	19,7
Transports (EK)	6,7	7,1	4,4
Activités financières (EL)	2,3	3,5	0,4
Activités immobilières (EM)	1,3	0,0	0,0
Services aux entreprises (EN)	13,6	14,4	11,5
Autres services (EP à ER)	17,7	15,9	25,2
Nombre d'entreprises	5 065	396	139

Lecture : parmi les 5 065 entreprises présentes dans les enquêtes REPONSE 1998 et 2004, 18,6 % appartiennent au secteur du commerce. Dans ce secteur, la proportion est de 11,1 % d'entreprises représentées par au moins deux établissements et de 19,7 % pour les entreprises représentées par au moins trois établissements.

La répartition sectorielle des entreprises est calculée en appliquant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises ayant au moins un établissement de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

Tableau 4

Comparaison des caractéristiques des entreprises telles qu'elles sont décrites par les représentants de la direction interrogés au niveau des établissements, pour différents échantillons

En %

Variable	Échantillons			
	Ensemble des entreprises figurant dans les enquêtes REPONSE 1998 et 2004	Ensemble des entreprises dont plusieurs établissements ont été sélectionnés	Entreprises pour lesquelles ont dispose d'un binôme de représentants de la direction	Entreprises pour lesquelles ont dispose d'un triplet de représentants de la direction
MULTI	34	98	98	100
Entreprise multi-établissements	(0,007)	(0,005)	(0,005)	(0,003)
SALENT	23	60	62	99
Entreprise de + 1000 salariés	(0,008)	(0,025)	(0,025)	(0,005)
LIENS	34	66	67	64
Entreprise appartient à un groupe	(0,007)	(0,021)	(0,022)	(0,036)
BOURSE	15	43	44	49
Entreprise cotée en Bourse	(0,005)	(0,024)	(0,024)	(0,041)
K1	46	23	24	20
Famille, principal actionnaire	(0,008)	(0,020)	(0,023)	(0,039)
CHPROP	13	12	12	9
Changement de propriétaire	(0,005)	(0,015)	(0,015)	(0,021)
QUALTOT	49	70	72	69
Démarche de qualité totale	(0,007)	(0,019)	(0,019)	(0,025)
ISO	25	56	59	48
Adoption de norme ISO	(0,006)	(0,021)	(0,021)	(0,033)
PGIERP*	31	46	46	63
Progiciel de gestion intégrée	(0,009)	(0,027)	(0,028)	(0,035)
ACTIONS	8	22	19	26
Salariés actionnaires	(0,004)	(0,019)	(0,020)	(0,038)
PEE*	28	50	51	51
Plan d'épargne entreprise	(0,009)	(0,033)	(0,033)	(0,056)
PER*	8	12	11	19
Plan d'épargne retraite	(0,005)	(0,017)	(0,018)	(0,027)
RPCA	40	68	71	83
Représentant des salariés au CA	(0,016)	(0,030)	(0,041)	(0,037)
FEDPAT	58	73	72	68
Entreprise dans fédération patronale	(0,007)	(0,019)	(0,021)	(0,030)
DSENT	38	95	95	99
Délégué syndical dans l'entreprise	(0,007)	(0,009)	(0,009)	(0,004)
CGTPE	19	80	82	90
Délégué CGT dans l'entreprise	(0,006)	(0,016)	(0,016)	(0,020)
CFDTPE	20	75	76	91
Délégué CFDT dans l'entreprise	(0,006)	(0,020)	(0,020)	(0,021)
CGCPE	8	42	44	63
Délégué CGC dans l'entreprise	(0,004)	(0,022)	(0,023)	(0,031)
CFTCPE	8	36	36	57
Délégué CFTC dans l'entreprise	(0,004)	(0,020)	(0,021)	(0,031)
FOPE	13	59	62	87
Délégué FO dans l'entreprise	(0,005)	(0,021)	(0,022)	(0,023)
SUDPE*	1	4	4	11
Délégué SUD dans l'entreprise	(0,002)	(0,012)	(0,013)	(0,034)
UNSAPE*	1	6	7	10
Délégué UNSA dans l'entreprise	(0,002)	(0,013)	(0,015)	(0,029)
Nombre d'entreprises	5 065	396	396	139
Nombre d'établissements	5 908	1 239	792	417

Lecture : d'après les réponses fournies par les représentants de la direction, 15 % des entreprises de l'échantillon complet de l'enquête REPONSE sont cotées en Bourse. Cette proportion est de 43 % pour les entreprises dont plusieurs établissements ont été sélectionnés. Lorsqu'on se limite à un binôme de représentants de la direction par entreprise, cette proportion est de 44 %. Elle est de 49 % lorsqu'on se limite à un triplet de représentants de la direction par entreprise.

Toutes les variables sont à modalités binaires et celles qui sont marquées d'un astérisque ne sont disponibles qu'en 2004. Les moyennes sont calculées au niveau de chaque entreprise (à partir des réponses fournies par les différents représentants de la direction interrogés au niveau des établissements) puis sont pondérées en utilisant la pondération « entreprise » (cf. encadré 2). Les écarts-types des moyennes sont indiqués entre parenthèses.

Champ : entreprises ayant au moins un établissement de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

étant représentatifs du champ des entreprises possédant plus de deux établissements de 20 salariés et plus). Elles sont en revanche plus prononcées pour les triplets : les entreprises de cet échantillon sont plus grandes, davantage cotées en Bourse et ont un actionnariat salarié plus développé que les entreprises multi-établissements. Par ailleurs, la représentation des syndicats où les cadres sont majoritaires (CGC, CFDT, CFTC) y semble plus développée.

Ces différentes comparaisons indiquent qu'en toute rigueur, la validité des conclusions de notre analyse ne s'étend qu'aux entreprises multi-établissements et que nos résultats ne peuvent être immédiatement généralisés au champ des entreprises mono-établissement (6).

Les facteurs qui influent sur la qualité des réponses

Les réponses des binômes et triplets de représentants de la direction d'une même entreprise peuvent être utilisées pour estimer les taux d'erreur et le taux de prévalence réels des caractéristiques des entreprises, selon la méthode exposée plus haut.

L'importance des erreurs affectant les réponses varie selon le degré de « difficulté » de la question

Le modèle à taux d'erreur homogène fournit un cadre théorique simple pour évaluer l'importance des erreurs commises par les représentants de la direction interrogés sur les caractéristiques de leur entreprise.

L'estimation des paramètres a été faite à partir de l'échantillon des 396 binômes de dirigeants interrogés soit en 1998, soit en 2004 (cf. tableau 5).

Le premier enseignement qui peut être tiré de ces estimations est que, toutes questions confondues, le taux d'erreur moyen est relativement faible, mais non négligeable puisqu'il se situe généralement entre 10 et 20 %. Le second enseignement est que ce taux d'erreur varie beaucoup d'une question à l'autre : de 1 % pour la question portant sur le fait que l'entreprise comporte plus de 1 000 salariés (SALENT) à 25 % pour la question portant sur l'adoption d'un progiciel de gestion intégrée (PGIERP).

Dans le détail, la probabilité d'erreur apparaît logiquement faible (moins de 10 %) pour les questions *a priori* les plus faciles et dont l'interprétation est claire : entreprise multi-établissements (MULTI), de plus de 1 000 salariés (SALENT), appartenant à un groupe (LIENS), cotée en Bourse (BOURSE), changement de propriétaire (CHPROP), actionnariat salarié (ACTIONS), adoption d'un plan d'épargne entreprise (PEE) ou d'un plan d'épargne retraite (PER), présence d'un délégué syndical dans l'entreprise (DSENT).

Les taux d'erreur sont en revanche plus élevés lorsque les représentants de la direction sont interrogés sur des aspects plus spécifiques de leur entreprise : ils se situent alors entre 10 et 25 % (famille « principal actionnaire », démarche de qualité totale, adoption d'une norme ISO, utilisation d'un progiciel de gestion intégrée). Ces taux d'erreurs plus importants pourraient également traduire la nature équivoque de certaines questions portant sur des concepts managériaux ou technologiques, dont l'interprétation peut s'éloigner des canons de la littérature de gestion, notamment pour les moins grandes entreprises (7).

Les taux d'erreur sont du même ordre de grandeur lorsque les dirigeants d'établissement sont interrogés sur les différents aspects du dialogue social dans leur entreprise : présence de représentants des salariés au conseil d'administration de l'entreprise (RPCA), affiliation de l'entreprise à une fédération patronale (FEDPAT) et présence de représentants des principales centrales syndicales (CGT, CFDT, CGC, CFTC, FO). La plupart des réponses à ces questions présentent des taux d'erreur compris entre 10 et 20 %.

Ces différentes conclusions ne sont guère affectées lorsqu'on estime les paramètres du modèle à taux d'erreur homogène non plus à partir de l'échantillon des binômes de représentants de la direction, mais à partir de l'échantillon des triplets (cf. tableau 6). Bien que moins préci-

6. Le fait de se limiter aux entreprises multi-établissements conduit sans doute à une vision plus pessimiste de la qualité des réponses fournies par les représentants de la direction que si l'estimation avait pu être généralisée aux entreprises mono-établissement. En effet, le siège de ces dernières se confond par définition avec l'établissement enquêté, ce qui doit tendre à améliorer la qualité des réponses aux questions portant sur l'entreprise.

7. On retrouverait ici une problématique classique des enquêtes (voir par exemple Desrosières et Thévenot, 1979) dont l'analyse, bien que dépassant le cadre de cet article, pourrait être approfondie en rapprochant nos résultats de ceux des post-enquêtes conduites après les vagues REPONSE pour le compte de la Dares.

Tableau 5

Modèle à taux d'erreur homogène : estimation du taux de prévalence réel des caractéristiques des entreprises et du taux d'erreur affectant les réponses de 396 binômes de représentants de la direction d'une même entreprise

En %

Variable	Taux de prévalence naïf (q)	Taux de prévalence estimé (p)	Taux d'erreur estimé (φ)	% de binômes utilisés	% de réponses différentes
MULTI Entreprise multi-établissements	98 (0,005)	100 [1,00 ; 1,00]	2 [0,00 ; 0,05]	100	4
SALENT Entreprise de + 1000 salariés	62 (0,025)	62 [0,51 ; 0,74]	1 [0,00 ; 0,01]	92	1
LIENS Entreprise appartient à un groupe	67 (0,022)	70 [0,57 ; 0,83]	9 [0,06 ; 0,13]	97	17
BOURSE Entreprise cotée en Bourse	44 (0,024)	43 [0,34 ; 0,54]	5 [0,02 ; 0,08]	97	9
K1 Famille, principal actionnaire	24 (0,023)	15 [0,06 ; 0,27]	12 [0,05 ; 0,21]	62	21
CHPROP Changement de propriétaire	12 (0,015)	9 [0,04 ; 0,15]	4 [0,02 ; 0,07]	99	8
QUALTOT Démarche de qualité totale	72 (0,019)	81 [0,70 ; 0,90]	14 [0,09 ; 0,21]	97	24
ISO Adoption de norme ISO	59 (0,021)	63 [0,52 ; 0,73]	16 [0,10 ; 0,25]	97	27
PGIERP* Progiciel de gestion intégrée	46 (0,028)	42 [0,14 ; 0,69]	25 [0,13 ; 0,43]	94	39
ACTIONS Salariés actionnaires	19 (0,020)	14 [0,08 ; 0,22]	7 [0,05 ; 0,11]	75	13
PEE* Plan d'épargne entreprise	51 (0,033)	53 [0,34 ; 0,72]	7 [0,03 ; 0,12]	96	13
PER* Plan d'épargne retraite	11 (0,018)	4 [0,00 ; 0,10]	8 [0,03 ; 0,14]	93	14
RPCA Représentant des salariés au CA	71 (0,041)	77 [0,51 ; 0,98]	10 [0,03 ; 0,19]	25	18
FEDPAT Entreprise dans fédération patronale	72 (0,021)	83 [0,72 ; 0,94]	17 [0,09 ; 0,28]	79	28
DSENT Délégué syndical dans l'entreprise	95 (0,009)	98 [0,93 ; 1,00]	3 [0,00 ; 0,06]	99	5
CGTPE Délégué CGT dans l'entreprise	82 (0,016)	93 [0,86 ; 0,99]	12 [0,08 ; 0,18]	90	22
CFDTPE Délégué CFDT dans l'entreprise	76 (0,020)	82 [0,70 ; 0,91]	8 [0,05 ; 0,12]	90	15
CGCPE Délégué CGC dans l'entreprise	44 (0,023)	42 [0,31 ; 0,54]	13 [0,08 ; 0,21]	90	23
CFTCPE Délégué CFTC dans l'entreprise	36 (0,021)	26 [0,17 ; 0,37]	20 [0,13 ; 0,30]	90	32
FOPE Délégué FO dans l'entreprise	62 (0,022)	68 [0,54 ; 0,84]	15 [0,09 ; 0,21]	90	25
SUDPE* Délégué FO dans l'entreprise	4 (0,013)	3 [0,01 ; 0,06]	1 [0,00 ; 0,02]	91	2
UNSAPE* Délégué FO dans l'entreprise	7 (0,015)	3 [0,01 ; 0,06]	4 [0,01 ; 0,11]	91	8

Lecture : lorsqu'on ne tient pas compte des erreurs pouvant affecter la qualité des réponses fournies par les représentants de la direction des entreprises, on trouve que 59 % des entreprises de l'échantillon ont adopté une norme ISO. La prise en compte de ces erreurs (supposées homogènes) à travers la comparaison des réponses données par les binômes de représentants de la direction d'une même entreprise permet d'estimer un taux de prévalence réel de 63 % et un taux d'erreur de 16 %. Ce calcul est réalisé en utilisant 97 % des binômes disponibles. 27 % d'entre eux ont donné des réponses différentes à la question portant sur l'adoption d'une norme ISO.

Toutes les variables sont à modalités binaires et celles qui sont marquées d'un astérisque ne sont disponibles qu'en 2004. L'écart-type du taux de prévalence naïf est indiqué entre parenthèses. Les bornes des intervalles de confiance non paramétriques au seuil de 5 % pour le taux de prévalence réel et le taux d'erreur, calculés par bootstrap à l'issue de 1 000 réplifications, sont indiquées entre crochets. Tous les calculs sont effectués en utilisant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises formées d'au moins deux établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

ses du fait de la taille réduite de l'échantillon (139 triplets au total), les estimations du taux d'erreur et du taux de prévalence réel des caractéristiques sont très proches de celles obtenues à partir de l'échantillon des binômes. Les taux d'erreur sont relativement faibles pour les questions portant sur les caractéristiques générales de l'entreprise (généralement moins de 10 % d'erreurs) ; ils sont plus élevés en revanche pour les questions plus spécifiques et pour celles portant sur les instances du dialogue social dans l'entreprise (taux d'erreur le plus souvent compris entre 10 et 25 %).

La probabilité de donner une réponse inexacte dépend de la présence ou non de la caractéristique dans l'entreprise

Les estimations discutées précédemment reposent sur l'hypothèse d'homogénéité des taux d'erreurs d'un représentant de la direction à un autre. Or il est probable que dans la réalité, ces taux diffèrent sensiblement selon que l'entreprise possède ou non une caractéristique donnée. Cet aspect est analysé en estimant le modèle à taux d'erreur asymétrique (cf. encadré 1).

Si le taux d'erreur en l'absence de caractéristique n'a pas pu être estimé dans un certain nombre de cas (entreprise multi-établissements, entreprise de plus de 1 000 salariés, présence d'un délégué syndical dans l'entreprise, présence d'un délégué CGT), c'est parce que la quasi-totalité des dirigeants interrogés ont déclaré que leur entreprise possédait la caractéristique.

Le principal résultat (cf. tableau 7) réside dans la grande variabilité des écarts entre les taux d'erreurs φ_0 (cas où l'entreprise ne possède pas la caractéristique) et φ_1 (cas où l'entreprise possède effectivement la caractéristique). Si l'hypothèse de symétrie des taux d'erreur (les dirigeants des deux types d'entreprises ont la même probabilité de se tromper) ne peut être rejetée statistiquement dans la plupart des cas, les taux d'erreur moyens estimés sont néanmoins souvent assez différents selon que l'entreprise possède ou non la caractéristique en question. Les différences entre les deux types de taux d'erreurs φ_0 et φ_1 sont compatibles avec les prédictions d'un modèle dans lequel prédominent les erreurs « involontaires », où les dirigeants qui ne connaissent pas la bonne réponse suivent la stratégie rationnelle consistant à répondre « oui » avec la même probabilité que le taux de

prévalence de la caractéristique dans la population des entreprises multi-établissements (cf. encadré 4). On constate en effet que le taux d'erreur φ_0 est supérieur (respectivement inférieur) au taux φ_1 lorsque le taux de prévalence réel p est supérieur (respectivement inférieur) à 0,5 (conformément à l'analyse développée dans l'encadré 4). En supposant que les erreurs intentionnelles sont négligeables, les résultats suggèrent que si les représentants de la direction se trompent rarement lorsqu'ils sont interrogés sur les caractéristiques générales de leur entreprise (cotation en Bourse, famille principal actionnaire, changement récent de propriétaire), ils sont moins assurés lorsqu'il s'agit de répondre aux questions portant sur les relations sociales. La proportion de dirigeants qui ne connaissent pas la bonne réponse varie en effet le plus souvent entre 2 et 33 % lorsqu'ils sont interrogés sur la présence de syndicats effectivement présents dans leur entreprise.

La qualité des réponses est liée à la fonction occupée par la personne interrogée...

L'enquête *REPONSE* contient des variables qui permettent d'étudier d'éventuelles corrélations entre la qualité des réponses fournies par les dirigeants d'établissement et la fonction qu'ils occupent, la durée consacrée à l'entretien ou encore l'année d'enquête. S'il faut se garder de donner une interprétation exagérément « causale » à ces corrélations, leur signe et leur taille peuvent néanmoins permettre de mieux caractériser les erreurs commises par les cadres dirigeants enquêtés.

Il paraît raisonnable de penser que la cohérence des réponses fournies par les représentants de la direction est susceptible de varier en fonction de la position qu'ils occupent au sein de l'entreprise. On peut notamment supposer que les responsables des ressources humaines ont une meilleure connaissance des relations sociales dans leur entreprise que les autres catégories de représentants de la direction (directeurs administratifs ou financiers, directeurs d'établissements, etc.). Pour tester cette hypothèse, on a divisé l'échantillon des binômes de représentants de la direction des entreprises en deux sous-groupes : le premier groupe est constitué des binômes formés de deux directeurs des ressources humaines (155 au total) ; le second rassemble les binômes dont l'un au moins des deux membres n'est pas directeur des ressources humaines (241 en tout).

Tableau 6

Modèle à taux d'erreur homogène : estimation du taux de prévalence réel des caractéristiques des entreprises et du taux d'erreur affectant les réponses de 139 triplets de représentants de la direction d'une même entreprise

En %

Variable	Taux de prévalence naïf (q)	Taux de prévalence estimé (p)	Taux d'erreur estimé (φ)	% de triplets utilisés	% de réponses différentes
MULTI	100	100	1	100	1
Entreprise multi-établissements	(0,003)	[1,00 ; 1,00]	[0,00 ; 0,01]		
SALENT	99	100	1	93	3
Entreprise de + 1000 salariés	(0,005)	[1,00 ; 1,00]	[0,00 ; 0,03]		
LIENS	80	96	14	94	36
Entreprise appartient à un groupe	(0,025)	[0,91 ; 1,00]	[0,09 ; 0,20]		
BOURSE	49	52	4	98	11
Entreprise cotée en Bourse	(0,041)	[0,31 ; 0,73]	[0,02 ; 0,07]		
K1	18	19	5	56	12
Famille principal actionnaire	(0,039)	[0,06 ; 0,35]	[0,02 ; 0,11]		
CHPROP	9	6	4	98	11
Changement de propriétaire	(0,021)	[0,02 ; 0,13]	[0,02 ; 0,06]		
QUALTOT	69	89	24	96	57
Démarche de qualité totale	(0,025)	[0,79 ; 0,98]	[0,15 ; 0,35]		
ISO	48	49	19	96	47
Adoption de norme ISO	(0,033)	[0,35 ; 0,64]	[0,10 ; 0,33]		
PGIERP*	64	69	26	92	64
Progiciel de gestion intégrée	(0,036)	[0,43 ; 0,92]	[0,12 ; 0,43]		
ACTIONS	25	20	9	65	23
Salariés actionnaires	(0,039)	[0,08 ; 0,39]	[0,04 ; 0,16]		
PEE*	51	57	5	93	12
Plan d'épargne entreprise	(0,057)	[0,25 ; 0,93]	[0,01 ; 0,10]		
PER*	19	3	13	91	47
Plan d'épargne retraite	(0,027)	[0,00 ; 0,08]	[0,03 ; 0,26]		
RPCA	82	-	-	30	38
Représentant des salariés au CA	(0,039)	-	-		
FEDPAT	66	94	22	69	57
Entreprise dans fédération patronale	(0,032)	[0,85 ; 0,99]	[0,11 ; 0,32]		
DSENT	99	100	1	99	2
Délégué syndical dans l'entreprise	(0,004)	[1,00 ; 1,00]	[0,00 ; 0,02]		
CGTPE	92	100	3	87	20
Délégué CGT dans l'entreprise	(0,016)	[1,00 ; 1,00]	[0,02 ; 0,03]		
CFDTPE	93	98	5	87	14
Délégué CFDT dans l'entreprise	(0,017)	[0,96 ; 1,00]	[0,02 ; 0,09]		
CGCPE	63	83	21	87	52
Délégué CGC dans l'entreprise	(0,031)	[0,67 ; 0,97]	[0,11 ; 0,33]		
CFTCPE	58	74	23	87	54
Délégué CFTC dans l'entreprise	(0,032)	[0,56 ; 0,92]	[0,12 ; 0,37]		
FOPE	89	96	8	87	22
Délégué FO dans l'entreprise	(0,021)	[0,90 ; 1,00]	[0,04 ; 0,14]		
SUDPE*	11	9	4	91	9
Délégué SUD dans l'entreprise	(0,034)	[0,02 ; 0,23]	[0,01 ; 0,09]		
UNSAPE*	10	3	5	91	13
Délégué UNSA dans l'entreprise	(0,029)	[0,00 ; 0,10]	[0,02 ; 0,12]		

Lecture : lorsqu'on ne tient pas compte des erreurs pouvant affecter la qualité des réponses fournies par les représentants de la direction des entreprises, on trouve que 80 % des entreprises de l'échantillon appartiennent à un groupe. La prise en compte de ces erreurs (supposées homogènes) à travers la comparaison des réponses données par les triplets de représentants de la direction d'une même entreprise permet d'estimer un taux de prévalence réel de 96 % et un taux d'erreur de 14 %. Ce calcul est réalisé en utilisant 94 % des triplets disponibles. Dans 36 % des cas, les trois membres du triplet n'ont pas donné la même réponse à la question portant sur l'appartenance de leur entreprise à un groupe.

Toutes les variables sont à modalités binaires et celles qui sont marquées d'un astérisque ne sont disponibles qu'en 2004. L'écart-type du taux de prévalence naïf est indiqué entre parenthèses. Les bornes des intervalles de confiance non paramétriques au seuil de 5 % pour le taux de prévalence réel et le taux d'erreur, calculés par bootstrap à l'issue de 1 000 répliques, sont indiquées entre crochets. Tous les calculs sont effectués en utilisant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises formées d'au moins trois établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

Tableau 7

Modèle à taux d'erreur asymétrique : estimation du taux de prévalence réel des caractéristiques des entreprises et du taux d'erreur affectant les réponses de 139 triplets de représentants de la direction d'une même entreprise

En %

Variable	Taux de prévalence naïf (q)	Taux de prévalence estimé (p)	Taux d'erreur si caractéristique absente (ψ_0)	Taux d'erreur si caractéristique présente (ψ_1)	% de triplets utilisés	% de réponses différentes
MULTI Entreprise multi-établissements	100 (0,003)	100 [1,00 ; 1,00]	- -	1 [0,00 ; 0,01]	100	1
SALENT Entreprise de + 1000 salariés	99 (0,005)	100 [1,00 ; 1,00]	- -	1 [0,00 ; 0,03]	93	3
LIENS Entreprise appartient à un groupe	80 (0,025)	76 [0,57 ; 0,93]	42 [0,23 ; 0,50]	7 [0,01 ; 0,14]	94	36
BOURSE Entreprise cotée en Bourse	49 (0,041)	50 [0,32 ; 0,68]	5 [0,01 ; 0,15]	4 [0,01 ; 0,07]	98	11
K1 Famille principal actionnaire	18 (0,039)	20 [0,08 ; 0,35]	5 [0,01 ; 0,12]	9 [0,02 ; 0,25]	56	12
CHPROP Changement de propriétaire	9 (0,021)	7 [0,02 ; 0,13]	3 [0,01 ; 0,06]	9 [0,01 ; 0,28]	98	11
QUALTOT Démarche de qualité totale	69 (0,025)	71 [0,39 ; 0,96]	33 [0,03 ; 0,49]	16 [0,01 ; 0,30]	96	57
ISO Adoption de norme ISO	48 (0,033)	46 [0,22 ; 0,70]	17 [0,02 ; 0,33]	13 [0,01 ; 0,27]	96	47
PGIERP* Progiciel de gestion intégrée	64 (0,036)	70 [0,46 ; 0,87]	15 [0,01 ; 0,35]	18 [0,07 ; 0,30]	92	64
ACTIONS Salariés actionnaires	25 (0,039)	29 [0,14 ; 0,50]	5 [0,01 ; 0,15]	20 [0,08 ; 0,35]	65	23
PEE* Plan d'épargne entreprise	51 (0,057)	57 [0,26 ; 0,90]	7 [0,00 ; 0,31]	5 [0,01 ; 0,10]	93	12
PER* Plan d'épargne retraite	19 (0,027)	11 [0,04 ; 0,23]	4 [0,00 ; 0,08]	29 [0,06 ; 0,49]	91	47
RPCA Représentant des salariés au CA	82 (0,039)	- -	- -	- -	30	38
FEDPAT Entreprise dans fédération patronale	66 (0,032)	83 [0,61 ; 0,96]	29 [0,05 ; 0,49]	14 [0,05 ; 0,23]	69	57
DSENT Délégué syndical dans l'entreprise	99 (0,004)	100 [1,00 ; 1,00]	- -	1 [0,00 ; 0,02]	99	2
CGTPE Délégué CGT dans l'entreprise	92 (0,016)	100 [1,00 ; 1,00]	- -	3 [0,02 ; 0,03]	87	20
CFDTPE Délégué CFDT dans l'entreprise	93 (0,017)	93 [0,83 ; 1,00]	33 [0,05 ; 0,49]	3 [0,01 ; 0,07]	87	14
CGCPE Délégué CGC dans l'entreprise	63 (0,031)	73 [0,33 ; 0,89]	22 [0,03 ; 0,45]	13 [0,02 ; 0,22]	87	52
CFTCPE Délégué CFTC dans l'entreprise	58 (0,032)	63 [0,26 ; 0,83]	21 [0,03 ; 0,43]	13 [0,01 ; 0,24]	87	54
FOPE Délégué FO dans l'entreprise	89 (0,021)	94 [0,85 ; 0,98]	19 [0,01 ; 0,45]	7 [0,03 ; 0,12]	87	22
SUDPE* Délégué SUD dans l'entreprise	11 (0,034)	11 [0,03 ; 0,24]	3 [0,00 ; 0,09]	11 [0,01 ; 0,30]	91	9
UNSAPE* Délégué UNSA dans l'entreprise	10 (0,029)	13 [0,03 ; 0,30]	2 [0,00 ; 0,05]	32 [0,12 ; 0,48]	91	13

Lecture : lorsqu'on ne tient pas compte des erreurs pouvant affecter la qualité des réponses fournies par les représentants de la direction des entreprises, on trouve que 80 % des entreprises de l'échantillon appartiennent à un groupe. La prise en compte de ces erreurs (supposées asymétriques) à travers la comparaison des réponses données par les triplets de représentants de la direction d'une même entreprise permet d'estimer un taux de prévalence réel de 76 %, un taux d'erreur de 42 % lorsque l'entreprise du représentant ne possède pas la caractéristique et de 7 % lorsqu'elle la possède. Ce calcul est réalisé en utilisant 94 % des triplets disponibles. Dans 36 % des cas, les trois membres du triplet n'ont pas donné la même réponse à la question portant sur l'appartenance de leur entreprise à un groupe.

Toutes les variables sont à modalités binaires et celles qui sont marquées d'un astérisque ne sont disponibles qu'en 2004. L'écart-type du taux de prévalence naïf est indiqué entre parenthèses. Les bornes des intervalles de confiance non paramétriques au seuil de 5 % pour le taux de prévalence réel et le taux d'erreur, calculés par bootstrap à l'issue de 1 000 réplifications, sont indiquées entre crochets. Tous les calculs sont effectués en utilisant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises formées d'au moins trois établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

Si les écarts séparant les taux de cohérence des réponses selon le type de binôme sont rarement significatifs au seuil de 5 % (en raison de la taille modeste des échantillons), leur comparaison (cf. tableau 8) suggère que les directeurs des ressources humaines ont une meilleure connaissance des relations sociales que les autres catégories de représentants de la direction. En effet, pour les questions portant sur les instances du dialogue social et sur la représentation syndicale dans l'entreprise, le taux de cohérence des

réponses qu'ils fournissent est toujours plus élevé que celui des réponses données par les autres binômes. À l'inverse, il semblerait que les responsables des ressources humaines aient une connaissance un peu moins bonne des caractéristiques générales de leur entreprise (taille de l'entreprise, cotation en Bourse, actionnariat salarié, changement de propriétaire, démarche de qualité totale, adoption d'une norme ISO...), ce qui paraît logique compte tenu de leur domaine de compétence.

Encadré 4

ERREURS VOLONTAIRES OU INVOLONTAIRES ?

L'asymétrie des taux d'erreurs affectant les réponses des représentants de la direction d'entreprises recouvre deux phénomènes conceptuellement distincts : les erreurs *volontaires* et les erreurs *involontaires*. Les premières sont commises par les dirigeants qui décident de mentir sciemment, les secondes par les dirigeants qui ne connaissent pas la bonne réponse et répondent « au hasard ».

Dans quelle mesure ces deux types d'erreurs vont-elles affecter la probabilité de fournir une réponse erronée ? Considérons dans un premier temps les dirigeants d'établissement interrogés sur une caractéristique effectivement présente dans leur entreprise (soit $\theta_k = 1$). En notant η_1 la proportion de dirigeants qui ne connaissent pas la bonne réponse, ρ_1 la probabilité pour ces dirigeants de donner une mauvaise réponse quand ils ne savent pas et μ_1 la probabilité de mentir sciemment lorsqu'ils connaissent la bonne réponse, le taux d'erreur des dirigeants dont l'entreprise possède la caractéristique k s'écrit (en omettant l'indice k) :

$$\varphi_1 = \underbrace{\eta_1 \rho_1}_{\text{erreurs « involontaires »}} + \underbrace{(1 - \eta_1) \mu_1}_{\text{erreurs « volontaires »}} \quad (1)$$

De même, le taux d'erreur des dirigeants dont l'entreprise ne possède pas la caractéristique k s'écrit :

$$\varphi_0 = \underbrace{\eta_0 \rho_0}_{\text{erreurs « involontaires »}} + \underbrace{(1 - \eta_0) \mu_0}_{\text{erreurs « volontaires »}} \quad (2)$$

La comparaison des équations (1) et (2) montre que l'asymétrie des taux d'erreurs peut résulter aussi bien de taux d'erreurs volontaires différents que de taux d'erreurs involontaires différents.

Il est probable que les taux d'erreurs *volontaires* $\eta_1 \rho_1$ et $\eta_0 \rho_0$ soient peu différents pour les questions portant sur des caractéristiques sans connotation particulière (entreprise multi-établissements, par exemple). L'écart entre ces taux est en revanche potentiellement important pour les caractéristiques positivement connues (l'adoption d'une démarche de qualité totale, par exemple) : on peut en effet imaginer que les dirigeants

dont l'entreprise ne possède pas ce type de caractéristique auront tendance à embellir la réalité et donc à mentir plus souvent que les dirigeants dont l'entreprise possède effectivement cette caractéristique, si bien que $\mu_0 > \mu_1$.

Bien que cela puisse paraître relativement contre-intuitif à première vue, il est également vraisemblable que les taux d'erreurs *involontaires* soient en général différents selon que l'entreprise possède ou non la caractéristique. Un exemple trivial permet de comprendre la raison de ce phénomène. Supposons que la proportion de dirigeants ne connaissant pas la bonne réponse soit identique pour toutes les entreprises ($\eta_0 = \eta_1 = \eta$) et que lorsqu'ils ne savent pas si leur entreprise possède ou non la caractéristique, ces dirigeants adoptent la stratégie suivante : répondre « oui » avec la même probabilité que celle d'observer la caractéristique parmi l'ensemble des entreprises, cette probabilité (égale par définition à p) étant supposée connue. Une telle stratégie est rationnelle au sens où elle permet de minimiser l'espérance d'erreur dans la réponse. Si les dirigeants qui ne connaissent pas la bonne réponse adoptent cette stratégie, ils se tromperont avec la probabilité $\rho_1 = 1 - p$ si leur entreprise ne possède pas la caractéristique et avec la probabilité $\rho_0 = p$ si, au contraire, elle la possède effectivement. Les taux d'erreur des représentants de la direction s'écrivent alors :

$$\begin{cases} \varphi_1 = \eta(1 - p) \\ \varphi_0 = \eta p \end{cases}$$

Dans cette situation et à moins que p ne soit égal à 0,5, les taux d'erreurs involontaires seront en général différents, les dirigeants se trompant le plus souvent étant ceux qui appartiennent aux entreprises « minoritaires » par rapport à la possession de la caractéristique. Si $p < 0,5$ (resp. $p > 0,5$), on aura $\varphi_0 > \varphi_1$ (resp. $\varphi_0 < \varphi_1$).

Ce petit exemple indique que même si les dirigeants ne commettent que des erreurs involontaires, les taux d'erreurs n'ont pas de raison d'être égaux pour les deux types d'entreprises.

Tableau 8

Cohérence des réponses fournies par les binômes de représentants de la direction d'une même entreprise en fonction de la fonction occupée, de l'année d'enquête et de la durée de l'entretien

En %

Variable	Taux de cohérence des réponses lorsque les deux membres du binôme sont directeurs des ressources humaines	Taux de cohérence des réponses lorsque l'un au moins des membres du binôme n'est pas directeur des ressources humaines	Taux de cohérence des réponses en 1998	Taux de cohérence des réponses en 2004	Corrélation entre le taux de cohérence des réponses et la durée totale du thème correspondant (dont la moyenne est calculée au niveau du binôme de représentants de la direction)	
MULTI Entreprise multi-établissements	97 (0,01)	95 (0,01)	100 (0,00)	92 (0,02)	Thème 0 (11 items) (moyenne : 9,61 min)	0,05 [0,481]
SALENT Entreprise de + 1000 salariés	97 (0,01)	99 (0,01)	100 (0,00)	97 (0,01)		0,08 [0,284]
LIENS Entreprise appartient à un groupe	82 (0,03)	84 (0,02)	77 (0,03)	89 (0,02)		- 0,02 [0,816]
BOURSE Entreprise cotée en Bourse	90 (0,02)	92 (0,02)	88 (0,02)	94 (0,02)		0,02 [0,804]
K1 Famille principal actionnaire	85 (0,03)	77 (0,04)	77 (0,05)	79 (0,03)		0,21*** [0,005]
CHPROP Changement de propriétaire	80 (0,03)	95 (0,01)	87 (0,02)	96 (0,01)	Thème 5 (18 items) (moyenne : 12,03 min)	- 0,07 [0,348]
QUALTOT Démarche de qualité totale	75 (0,04)	76 (0,03)	74 (0,03)	77 (0,03)		0,14** [0,049]
ISO Adoption de norme ISO	71 (0,04)	73 (0,03)	73 (0,03)	73 (0,03)		0,13* [0,070]
ACTIONS Salariés actionnaires	78 (0,04)	89 (0,02)	80 (0,03)	93 (0,02)	Thème 6 (17 items) (moyenne : 17,54 min)	0,00 [0,980]
RPCA Représentant des salariés au CA	83 (0,05)	82 (0,06)	77 (0,06)	87 (0,05)		0,37*** [0,006]
FEDPAT Entreprise dans fédération patronale	81 (0,03)	69 (0,04)	73 (0,04)	72 (0,04)	Thème 1 (11 items) (moyenne : 5,40 min)	0,11 [0,173]
DSENT Délégué syndical dans l'entreprise	100 (0,00)	93 (0,02)	100 (0,00)	90 (0,02)	Thème 2 (19 items) (moyenne : 11,89 min)	0,32*** [0,000]
CGTPE Délégué CGT dans l'entreprise	85 (0,03)	76 (0,03)	72 (0,03)	84 (0,03)		0,21*** [0,004]
CFDTP Délégué CFDT dans l'entreprise	90 (0,02)	84 (0,03)	79 (0,03)	91 (0,02)		0,03 [0,636]
CGCPE Délégué CGC dans l'entreprise	83 (0,03)	76 (0,03)	81 (0,03)	74 (0,03)		0,31*** [0,000]
CFTCPE Délégué CFTC dans l'entreprise	69 (0,04)	68 (0,03)	71 (0,04)	66 (0,03)		0,29*** [0,000]
FOPE Délégué FO dans l'entreprise	79 (0,03)	74 (0,03)	67 (0,04)	82 (0,03)		0,12 [0,114]
Année(s)	1998/2004	1998/2004	1998	2004		2004
Nombre d'entreprises	155	241	191	205	396	

Lecture : la proportion de binômes de représentants de la direction qui donnent la même réponse lorsqu'on les interroge sur l'appartenance de leur entreprise à un groupe est de 82 % lorsque les deux membres sont directeurs des ressources humaines et de 84 % dans le cas inverse. En 1998 (resp. 2004), 77 % (resp. 89 %) des binômes de représentants de la direction d'une même entreprise donnent la même réponse lorsqu'on les interroge sur l'appartenance de leur entreprise à un groupe. La question portant sur l'appartenance à un groupe se rattache au thème 0 qui comprend 11 items et dont la durée moyenne est de 9,61 minutes. Le coefficient de corrélation entre la cohérence des réponses fournies quant à l'appartenance de l'entreprise à un groupe et la durée du thème 0 est de - 0,02.

Les écarts-types des taux de cohérences moyens sont indiqués entre parenthèses. Le seuil de significativité atteint par les coefficients de corrélation est indiqué entre crochets. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

Tous les calculs sont effectués en utilisant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises formées d'au moins deux établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

...ainsi qu'à la durée de l'entretien

Le temps consacré par les représentants de la direction à répondre aux questions posées dans le cadre de l'enquête *REPONSE* constitue un second paramètre susceptible d'être corrélé à la qualité de leurs réponses. On peut supposer que les représentants les plus consciencieux ont dédié davantage de temps à comprendre et à répondre aux questions que les autres. Lors de l'enquête *REPONSE 2004*, il était demandé aux enquêteurs d'indiquer, pour chaque représentant interrogé, la durée (en secondes) consacrée à chacun des dix thèmes de l'enquête, regroupant eux-mêmes entre 2 et 24 items, éventuellement subdivisés en sous-items.

La durée moyenne d'un thème varie entre 5,4 minutes pour le thème 1 et 17,5 minutes pour le thème 6 (cf. tableau 8). Les coefficients de corrélation entre la durée moyenne du thème et la cohérence des réponses fournies par les binômes de représentants de la direction d'une même entreprise sont généralement positifs et significatifs pour un certain nombre de questions portant sur la représentation syndicale dans l'entreprise (représentants du personnel au CA, présence d'un délégué syndical, présence de délégués CGT, CGC ou CFDC). Ces résultats semblent donc indiquer que lorsqu'il s'agit d'évoquer les relations sociales dans l'entreprise, les représentants de la direction qui passent davantage de temps à répondre aux questions donnent des réponses de meilleure qualité. On ne peut néanmoins exclure une relation inverse : ceux qui ont une meilleure connaissance de leur entreprise sont plus disert.

La qualité des réponses est meilleure en 2004 qu'en 1998

Le fait de disposer de deux vagues d'enquête successives permet de tester la cohérence des estimations d'une année à l'autre et d'examiner la stabilité des taux d'erreurs au cours du temps (cf. tableau 8).

Le principal enseignement de la comparaison des taux de cohérence est que la qualité des réponses est en général meilleure en 2004 qu'en 1998, en particulier pour les questions portant sur la présence des trois plus importantes centrales syndicales (CGT, CFDT, FO) dans l'entreprise : l'augmentation des taux de cohérence associés à ces questions est de l'ordre de 15 %. Ce phénomène semble traduire une meilleure connaissance par les dirigeants d'établissement

des réalités sociales de leur entreprise. Faut-il y voir l'effet des négociations syndicales qui ont accompagné la mise en place des 35 heures dans les entreprises dans le cadre des lois Aubry I (1998) et Aubry II (2000) ? Si les données disponibles ne permettent pas de tester cette hypothèse (8), il paraît néanmoins raisonnable de penser que le fait de conclure des accords de réduction du temps de travail a pu contribuer à cette évolution. L'amélioration du dispositif d'enquête peut également avoir renforcé sa pertinence.

Les erreurs paraissent peu corrélées entre elles, sauf pour les questions portant sur les relations sociales...

Les binômes de dirigeants classés comme « incohérents » pour une question ont-ils aussi tendance à l'être pour d'autres questions ? Pour étudier cet aspect, on calcule, pour l'ensemble des binômes interrogés lors des enquêtes *REPONSE 1998* et *2004*, la matrice de corrélation (9) de la cohérence des réponses aux différentes questions posées sur l'entreprise : une corrélation positive et significative pour deux questions indique que les dirigeants cohérents pour la première question ont également tendance à l'être pour la seconde (cf. tableau 9). En général, la cohérence des réponses fournies par les binômes de représentants n'est pas fortement corrélée (positivement ou négativement) d'une question à l'autre, ce qui suggère que si les dirigeants se trompent, ce n'est pas en raison d'une mauvaise connaissance générale de leur entreprise. Néanmoins, les réponses incohérentes aux questions relatives à la présence syndicale dans l'entreprise sont souvent positivement et significativement corrélées entre elles, ce qui laisse penser que les dirigeants d'établissement qui se trompent sur la présence d'un délégué d'une centrale syndicale particulière, tendent à se tromper également sur les autres composantes de la présence syndicale dans leur entreprise. Ainsi, la maîtrise imparfaite que les représentants de la direction ont des réalités sociales de leur entreprise ne semble pas traduire une méconnaissance des autres dimensions de celle-ci.

8. Les entreprises de l'échantillon utilisé ayant au moins deux établissements, elles sont pour la plupart de taille importante et sont donc passées par des négociations 35 heures.

9. Les variables considérées dans notre étude étant dichotomiques, la mesure d'association entre deux variables binaires que

nous calculons est le coefficient $\rho = \frac{n_{00}n_{11} - n_{01}n_{10}}{\sqrt{n_{00}n_{11}n_{01}n_{10}}}$, où les n_{ij}

indiquent la fréquence avec laquelle on observe que la première variable vaut i et la seconde vaut j et n_i (resp. n_j) le nombre de fois où la première variable vaut i (resp. où la seconde variable vaut j).

Tableau 9

Matrice de corrélation de la cohérence des réponses fournies par l'ensemble des 396 binômes de représentants de la direction d'une même entreprise

	MULTI	SALENTN	LIENS	BOURSE	CHPROP	ISO	DSEINTR	FEDPAT	CGTPE	CFDTPE	CGCPE	CFTCPE	FOPE
MULTI	1,00*** [0,000]												
Ent. Multi-établissements													
SALENTN	1,00*** [0,000]												
Ent. de + 1 000 salariés													
LIENS	0,07* [0,070]	1,00*** [0,000]											
Ent. dans un groupe													
BOURSE	- 0,05 [0,143]	- 0,03 [0,389]	0,24*** [0,000]	1,00*** [0,000]									
Ent. cotée en Bourse													
CHPROP	0,01 [0,730]	- 0,03 [0,351]	0,04 [0,230]	- 0,04 [0,221]	1,00*** [0,000]								
Chgt de propriétaire													
ISO	- 0,05 [0,213]	- 0,07* [0,079]	0,05 [0,176]	- 0,05 [0,188]	0,17*** [0,000]	1,00*** [0,000]							
Adoption de norme ISO													
DSEINTR	- 0,14*** [0,000]	0,00 [0,942]	- 0,11*** [0,005]	0,00 [0,957]	- 0,11*** [0,005]	0,14*** [0,000]	1,00*** [0,000]						
Déleg. synd. dans l'ent.													
FEDPAT	0,49*** [0,000]	- 0,00 [0,920]	- 0,08** [0,029]	- 0,08** [0,038]	- 0,07* [0,054]	0,14*** [0,000]	0,15*** [0,000]	1,00*** [0,000]					
Ent. dans féd. patronale													
CGTPE	0,20*** [0,000]	- 0,05 [0,224]	0,19*** [0,000]	0,21*** [0,000]	0,04 [0,261]	0,02 [0,606]	- 0,07 [0,113]	0,33*** [0,000]	1,00*** [0,000]				
Délégué CGT dans l'ent.													
CFDTPE	- 0,06 [0,113]	- 0,04 [0,330]	- 0,02 [0,556]	0,12*** [0,001]	- 0,05 [0,190]	- 0,06* [0,099]	0,04 [0,287]	- 0,05 [0,146]	0,10*** [0,008]	1,00*** [0,000]			
Délégué CFDT dans l'ent.													
CGCPE	- 0,10** [0,010]	- 0,05 [0,177]	- 0,03 [0,379]	0,09** [0,017]	- 0,05 [0,149]	0,23*** [0,000]	0,24*** [0,000]	- 0,08** [0,034]	0,10*** [0,005]	0,18*** [0,000]	1,00*** [0,000]		
Délégué CGC dans l'ent.													
CFTCPE	- 0,14*** [0,000]	- 0,06* [0,095]	- 0,01 [0,708]	0,05 [0,209]	- 0,07* [0,073]	0,18*** [0,000]	0,13*** [0,001]	- 0,11*** [0,005]	0,01 [0,726]	0,24*** [0,000]	0,42*** [0,000]	1,00*** [0,000]	
Délégué CFTC dans l'ent.													
FOPE	- 0,01 [0,813]	- 0,02 [0,635]	0,16*** [0,000]	0,14*** [0,000]	0,15*** [0,000]	0,09** [0,021]	- 0,09** [0,036]	- 0,10*** [0,008]	0,28*** [0,000]	0,23*** [0,000]	- 0,00 [0,990]	0,03 [0,500]	1,00*** [0,000]
Délégué FO dans l'ent.													

Lecture : les binômes de représentants de la direction d'une même entreprise qui donnent des réponses cohérentes lorsqu'on les interroge sur la présence d'un délégué CFDT dans leur entreprise ont également tendance à être cohérents lorsqu'on les interroge sur la présence d'un délégué CFTC. Le coefficient de corrélation de la cohérence des réponses à ces deux questions est égal à 0,24 et est significatif au seuil de 1 %. La mesure d'association entre deux variables binaires est le coefficient $\rho = \frac{n_{00}n_{11} - n_{01}n_{10}}{\sqrt{n_{00}n_{11}n_{01}n_{10}}}$, où les n_{ij} indiquent la fréquence avec laquelle on observe que la première variable vaut i et la seconde vaut j et n_j (resp. n_i) le nombre de fois où la première variable vaut i (resp. où la seconde variable vaut j). Les corrélations sont estimées en utilisant la pondération « entreprise » (cf. encadré 2). Le seuil de significativité atteint par les coefficients de corrélation est indiqué entre crochets. * : significatif au seuil de 10 % ; ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

Champ : entreprises formées d'au moins deux établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

...pour autant, leur prise en compte conduit à modifier l'estimation du niveau et de l'évolution de la prévalence de certaines caractéristiques des entreprises

Si nos résultats suggèrent que les erreurs commises par les dirigeants d'établissement sont en grande partie involontaires et non corrélées entre elles, leur prise en compte conduit néanmoins à corriger assez sensiblement le taux de prévalence de certaines caractéristiques des entreprises, tant en niveau qu'en évolution.

Les estimations des taux de prévalence réalisées sous l'hypothèse d'homogénéité des taux d'erreurs s'écartent souvent des estimations naïves. Que ce soit pour les binômes (cf. tableau 5) ou pour les triplets (cf. tableau 6) de représentants de la direction, l'écart moyen entre le taux de prévalence réel estimé et le taux de prévalence naïf s'élève (en valeur absolue) à environ 5 points de pourcentage. Comme expliqué précédemment, cet écart est généralement positif lorsque le taux de prévalence est inférieur à 0,5 et négatif sinon. La prise en compte des erreurs dans les réponses des dirigeants pourrait conduire à réévaluer sensiblement la présence de certains syndicats dans les entreprises.

La prise en compte d'erreurs asymétriques (cf. tableau 7) conduit à des estimations des taux de prévalence réels des caractéristiques sensiblement différentes de celles obtenues dans le cadre du modèle à taux d'erreur homogène : le taux de prévalence apparaît tantôt sous évalué (utilisation d'un progiciel de gestion intégrée), tantôt sur-évalué (appartenance à un groupe) par l'approche naïve. En particulier, ignorer la possibilité que les représentants de la direction se trompent dans leurs réponses conduit généralement à sous-estimer la présence syndicale dans les entreprises : le taux naïf de présence de délégués syndicaux CGT (92 %) ou FO (89 %) dans ces entreprises se situe par exemple nettement en deçà du taux réel estimé (respectivement 100 % et 94 %).

En ignorant le fait que les cadres dirigeants puissent donner des réponses erronées aux questions portant sur les caractéristiques de leur entreprise, on risque donc de produire des estimations inexactes de ces caractéristiques. De manière peut-être plus problématique encore, la variation des taux d'erreurs d'une enquête à l'autre a pour conséquence de fausser l'estimation d'un certain nombre d'évolutions. En effet, si les taux d'erreurs ne sont pas constants d'une enquête à une autre, alors l'évolution « naïve »

des taux de prévalence différera en général de l'évolution réelle (cf. tableau 10). Si l'on s'intéresse, par exemple, aux caractéristiques des entreprises étudiées jusqu'à présent, on constate que les évolutions naïves contrastent souvent avec les évolutions réelles estimées sous l'hypothèse d'erreurs asymétriques. D'après nos calculs, une approche naïve conduirait par exemple à sous-estimer, dans la population des entreprises multi-établissements, l'augmentation de la fraction d'entre elles qui pratiquent l'actionnariat salarié et, inversement, à surestimer la diminution de la proportion de celles qui ont adopté une démarche de qualité totale. L'estimation de l'évolution de la représentation syndicale dans les entreprises est également faussée par l'approche naïve : cela conduit par exemple à relativiser l'accroissement de la présence de certains syndicats (CGT, FO) et, au contraire, à réévaluer l'augmentation de la présence d'autres centrales syndicales (CFDT, CGC, CFTC) (cf. tableau 10).

* *
*

Pris globalement, nos résultats sont plutôt rassurants : ils indiquent que les représentants de la direction interrogés dans le cadre de l'enquête REPONSE ont une assez bonne connaissance de leur entreprise et que les erreurs éventuellement commises ne paraissent pas intentionnelles, dans la plupart des cas. Logiquement, leurs réponses sont d'autant plus précises que la question est simple et qu'elle relève du champ de compétence du dirigeant interrogé. Il reste que l'analyse révèle qu'ils maîtrisent moins les réalités sociales de leur entreprise et que ce phénomène tend à fausser de manière très sensible l'estimation de la présence syndicale et de son évolution dans le temps.

D'un point de vue méthodologique, un certain nombre d'améliorations pourraient être apportées au pilotage des enquêtes sur les pratiques organisationnelles et technologiques des entreprises. Les résultats de la comparaison des taux d'erreurs selon la fonction occupée par le représentant de la direction interrogé suggèrent que ces enquêtes gagneraient à être administrées différemment. Plutôt que de poser toutes les questions à un seul interlocuteur par établissement, il pourrait être plus efficace de s'adresser à plusieurs représentants de la direction dans les établissements où cela est possible (typiquement un directeur d'établissement et un responsable des ressources humaines) et de les interroger sur les aspects de leur entreprise ou de leur établisse-

ment qui relèvent de leur champ de compétence. À ce titre, un sous-ensemble de questions communes pourrait être soumis aux différents représentants de la direction afin d'estimer la fiabilité de leurs réponses et de détecter d'éventuels biais susceptibles d'affecter l'estimation des modè-

les organisationnels et technologiques qui se fondent sur ce type d'enquêtes. Par ailleurs la qualité des réponses fournies est susceptible de varier en fonction de la durée consacrée à chaque thème, ce qui plaide au minimum en faveur d'une plus forte sensibilisation des enquêteurs à

Tableau 10
Évolutions naïves et corrigées d'un certain nombre de caractéristiques des entreprises entre 1998 et 2004

Variable	Évolution « naïve » du taux de prévalence de la caractéristique			Évolution corrigée du taux de prévalence de la caractéristique
	Ensemble des entreprises de l'échantillon	Entreprises de l'échantillon dont plusieurs établissements ont été sélectionnés	Entreprises pour lesquelles ont dispose d'un triplet de représentants de la direction	Entreprises pour lesquelles ont dispose d'un triplet de représentants de la direction
LIENS Entreprise appartient à un groupe	+5	- 8	+8	+16
BOURSE Entreprise cotée en Bourse	+7	- 30	- 20	- 8
K1 Famille principal actionnaire	- 11	+5	+58	+15
CHPROP Changement de propriétaire	+1	- 41	- 64	- 60
QUALTOT Démarche de qualité totale	- 3	- 8	+6	+3
ISO Adoption de norme ISO	- 6	- 12	- 2	+11
ACTIONS Salariés actionnaires	+0	- 42	- 18	+23
RPCA Représentant des salariés au CA	- 1	+0	- 2	+0
FEDPAT Entreprise dans fédération patronale	- 10	+14	- 20	- 9
DSENR Délégué syndical dans l'entreprise	- 2	- 8	- 1	+0
CGTPE Délégué CGT dans l'entreprise	- 8	+0	+18	+6
CFDTPE Délégué CFDT dans l'entreprise	+1	- 1	+11	+15
CGCPE Délégué CGC dans l'entreprise	- 12	- 25	- 3	+32
CFTCPE Délégué CFTC dans l'entreprise	+14	+4	+11	+54
FOPE Délégué FO dans l'entreprise	- 5	- 13	+16	+9
Nombre d'entreprises en 1998	2 589	191	64	64
Nombre d'entreprises en 2004	2 476	205	75	75

Lecture : lorsqu'on ne tient pas compte des erreurs potentiellement commises par les représentants de la direction des entreprises interrogés dans le cadre de l'enquête REPONSE, on trouve que la proportion d'entreprises ayant un délégué CGC a diminué de 3 % pour les entreprises de trois établissements ou plus. La prise en compte de ces erreurs (supposées asymétriques) conduit à estimer que l'évolution réelle aurait été une augmentation de 32 % pour cette catégorie d'entreprises. À titre de comparaison, les deux premières colonnes indiquent que l'évolution « naïve » de la proportion d'entreprise ayant un délégué CGC est de - 12 % pour l'ensemble des entreprises et de - 25 % pour les entreprises multi-établissements.

Tous les calculs sont effectués en utilisant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises formées d'établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.

cet aspect, ainsi qu'à la nécessité d'explicitier les questions renvoyant aux concepts managériaux dont l'enquête cherche à cerner la prévalence. D'une manière générale, les conclusions de

ce travail plaident pour une meilleure prise en compte des caractéristiques des répondants dans l'exploitation tant quantitative que qualitative des enquêtes menées auprès des entreprises. □

BIBLIOGRAPHIE

Addison K. et Belfield C. (2004), « Unions and Employment Growth: The One Constant ? », *Industrial Relations*, vol. 43, n° 2, pp. 305–323.

Aghion P., Caroli E. et Garcia Peñalosa C. (1999), « Inequality and Income Growth: The Perspective of the New Growth Theories », *Journal of Economic Literature*, vol. 37, pp. 1615–1660.

Ardilly P. (2006), *Les techniques de sondage*, Éditions Technip, Paris.

Askenazy P., Thoenig M. et Thesmar D. (2006), « On the Relation Between Organizational Practices and New Technologies: the Role of Time Based Competition », *Economic Journal*, vol. 116, pp. 128–154.

Aubert P., Caroli E. et Roger M. (2006), « New technologies, Organisation and Age: Firm-Level Evidence », *Economic Journal*, vol. 16, pp. 73–93.

Behaghel L., Caroli E. et Walkowiak E. (2007), « Innovation and Skill Upgrading: The role of External vs. Internal Labour Markets », document de travail PSE, n° 2007-04.

Black S. et Lynch L. (2001), « How to Compete: The Impact of Workplace Practices and Information Technology », *Review of Economics and Statistics*, vol. 83, n° 3, pp. 434–445.

Bloom N. et van Reenen J. (2007), « Measuring and Explaining Management Practices Across Firms and Countries », *Quarterly Journal of Economics*, vol. 122, n° 4, pp. 1351–1408.

Bresnahan T., Brynjolfsson E. et Hitt L. (2002), « Information Technology, Workplace Organization, And The Demand For Skilled Labor: Firm-Level Evidence », *Quarterly Journal of Economics*, vol. 117, n° 1, p. 339–376.

Caroli E. et Van Reenen J. (2001), « Skill Biased Organizational Change? Evidence from

a Panel of British and French Establishments », *Quarterly Journal of Economics*, vol. 116, n° 4, pp. 1449–1492.

Desrosières A. et Thévenot L. (1979), « Les mots et les chiffres : les nomenclatures socio-professionnelles », *Économie et statistique*, n° 110, avril, p. 49-65.

Dunlop J. et Weil D. (1996), « Diffusion and Performance of Modular Production in the U.S. Apparel Industry », *Industrial Relations*, vol. 35, n° 3, pp. 334–355.

Huselid M. et Becker B. (1996), « Methodological Issues in Cross-sectional and Panel Estimates of the Human Resource-firm Performance Link », *Industrial Relations*, vol. 35, n° 3, pp. 400–422.

Ichniowski C., Shaw K. et Prennushi G. (1997), « The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines », *American Economic Review*, vol. 87, n° 3, pp. 291–313.

Lavallée P. (1995), « Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à ménages à l'aide de la méthode du partage des poids », *Techniques d'enquêtes*, vol. 21, n° 1, pp. 27–35.

Lavallée P. (2002), *Le Sondage Indirect ou la Méthode généralisée du partage des poids*, Éditions de l'Université de Bruxelles, Bruxelles.

Pignoni M.-T. et Tenret E. (2007), « Présence syndicale : des implantations en croissance, une confiance des salariés qui ne débouche pas sur des adhésions », *Premières informations, Premières synthèses*, n° 14.2, avril, Dares.

Osterman P. (1994), « How Common is Workplace Transformation and Who Adopts it? », *Industrial and Labor Relations Review*, vol. 47, n° 2, pp. 173–188.

TRANSFORMATION DES VARIABLES À MODALITES MULTIPLES EN VARIABLES BINAIRES

Création de variables binaires

Les variables de l'enquête *REPONSE* que nous avons sélectionnées pour notre étude sont généralement des variables catégorielles à modalités multiples. La question portant sur l'appartenance d'une entreprise à un groupe (variable LIENS) comporte par exemple cinq modalités.

L'entreprise :

- exerce en franchise,
- est liée à un groupe d'entreprise (GIE, association d'entreprises, adhérent d'une centrale d'achat...),
- appartient à un groupe,
- est totalement indépendante,
- ne sait pas.

Pour les besoins de l'analyse, nous avons dû procéder au regroupement de ces catégories afin de former des indicatrices prenant les valeurs 0 ou 1. Plusieurs cas de figure peuvent se présenter.

Le plus souvent, le codage des variables brutes présente trois modalités : « oui », « non » et « ne sait pas ». Dans ce cas, nous avons choisi d'attribuer la valeur 1 lorsque la réponse est « oui », la valeur 0 lorsque la réponse est « non » et une valeur manquante lorsque le représentant de la direction ne sait pas.

Certaines variables brutes utilisées dans notre étude présentent plus de trois modalités (en plus de la catégorie « ne sait pas ») qui sont généralement ordonnées. Par exemple, à la question « *Au 31 décembre de l'année passée, quel était le nombre total de salariés dans l'entreprise, en France ?* », les dirigeants ont le choix entre 9 réponses possibles : « *moins de 50 salariés* », « *de 50 à 99 salariés* », « *de 100 à 199 salariés* », « *de 200 à 499 salariés* », « *de 500 à 999 salariés* », « *de 1 000 à 4 999 salariés* », « *de 5 000 à 9 999 salariés* », « *10 000 salariés et plus* » et « *ne sait pas* ». Dans ce cas, nous avons regroupé les modalités en deux catégories (0 et 1) et attribué une valeur manquante lorsque le dirigeant ne sait pas. Dans le cas de la variable indiquant le nombre total de salariés de l'entreprise, nous avons ainsi créé une variable binaire indiquant si l'entreprise compte plus de 1 000 salariés (SALENT). Cette indicatrice vaut 1 pour les quatre premières modalités de la variable brute, 0 pour les trois modalités suivantes et est manquante pour la dernière modalité.

Dans certains cas, une question n'est posée au dirigeant d'établissement que s'il a répondu « oui » à une ou plusieurs autres questions : c'est le cas notamment pour la question portant sur l'adoption d'une norme ISO, qui n'est posée que s'il a déclaré que son entreprise a adopté une démarche de qualité totale (QUALTOT). Pour ces variables, nous avons choisi de définir l'indicatrice

même si la réponse à la question « conditionnante » est « non », en attribuant la valeur 0 à la variable d'intérêt dans ce cas. Ainsi, dans le cas des questions à « tiroirs », le taux de prévalence d'une caractéristique est calculé par rapport à l'ensemble des entreprises : le taux de prévalence de la norme ISO doit donc être interprété comme la proportion d'établissements ayant adopté cette norme parmi l'ensemble des entreprises, et non parmi les seules entreprises concernées par une démarche de qualité totale.

Valeurs manquantes

Le tableau présenté ci-dessous indique, pour chacun des quatre sous-échantillons utilisés (binômes et triplets de représentants de la direction en 1998 et 2004) et pour chaque question considérée isolément, la proportion d'entreprises dont les représentants se sont vus poser la question et la proportion d'entreprises dont les représentants ont déclaré ne pas connaître la réponse ou n'ont pas souhaité répondre. On a calculé ces proportions au niveau de chaque entreprise avant d'estimer la moyenne générale en utilisant la pondération « entreprise » (cf. encadré 2).

Ce tableau fait apparaître que la plupart des questions étudiées dans cet article ont été posées à tous les représentants de la direction. Les principales exceptions concernent : la question portant sur la présence d'une famille comme principale catégorie d'actionnaires de l'entreprise (K1), qui n'est posée en 1998 que lorsque le représentant de la direction a déclaré connaître la répartition de la détention du capital de son entreprise ; la question portant sur l'utilisation d'une norme ISO, qui n'a été posée que si le représentant a déclaré que son entreprise a adopté une démarche de qualité totale ; enfin, la question portant sur la présence d'un représentant des salariés au conseil d'administration de l'entreprise (RPCA), qui n'a été posée que si le représentant a déclaré que les salariés détenaient une part du capital de l'entreprise.

Les colonnes 3, 4, 7 et 8 de ce tableau indiquent que la proportion de dirigeants qui ont déclaré ne pas connaître la réponse à la question ou n'ont pas souhaité répondre est généralement faible (moins de 10 %), y compris pour les questions portant sur la présence syndicale dans l'entreprise. Seule les questions portant sur l'adoption d'une norme ISO et l'appartenance de l'entreprise à une fédération patronale recueillent des taux de non-réponse élevés (de 10 à 32 %).

Ces résultats suggèrent que d'une question à l'autre (sauf pour les variables ISO et RPCA), les estimations réalisées dans le cadre de cette étude portent sur des échantillons d'entreprises très proches et que les biais induits par la non-réponse de certains représentants de la direction sont *a priori* négligeables.

Statistiques descriptives sur le taux de renseignement des items des enquêtes REPONSE 1998 et 2004 pour les échantillons composés d'entreprises représentées par un binôme ou un triplet de dirigeants d'établissements

En %

Variable	Entreprises représentées par un binôme de représentants de la direction				Entreprises représentées par un triplet de représentants de la direction			
	Proportion de représentants de la direction auxquels la question à été posée		Proportion de représentants de la direction qui ont répondu « ne sait pas » ou n'ont pas souhaité répondre à la question		Proportion de représentants de la direction auxquels la question à été posée		Proportion de représentants de la direction qui ont répondu « ne sait pas » ou n'ont pas souhaité répondre à la question	
	1998	2004	1998	2004	1998	2004	1998	2004
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MULTI	100	100	0	0	100	100	0	0
SALENTR	100	96	1	7	100	99	0	2
LIENS	100	100	1	4	100	100	1	12
BOURSE	100	100	4	1	100	100	2	0
K1	57	100	0	5	58	100	0	3
CHPROP	100	100	0	0	100	100	0	0
QUALTOT	100	100	0	3	100	100	1	1
ISO	72	69	14	15	68	70	32	31
PGIERP*		100		2		100		1
ACTIONS	91	85	0	3	84	91	2	2
PEE*		100		1		100		0
PER*		100		3		100		2
RPCA	32	28	5	8	44	31	2	15
FEDPAT	100	100	12	13	100	100	12	10
DSENTR	100	100	0	1	100	100	0	0
CGTPE, CFDTPE, CGCPE, CFTCPE, FOPE, SUDPE*, UNSAPE*	93	92	3	11	96	99	6	4
Nombre d'entreprises	155	241	155	241	64	75	64	75

Lecture : au sein de l'échantillon des entreprises pour lesquelles on dispose d'un binôme de représentants de la direction en 1998, 72 % des individus enquêtés se sont vus poser la question de l'adoption par leur entreprise d'une norme ISO et 14 % de ceux qui ont été interrogés ont déclaré ne pas connaître la réponse ou n'ont pas souhaité répondre.

Toutes les variables sont à modalités binaires et celles qui sont marquées d'un astérisque ne sont disponibles qu'en 2004. Tous les calculs sont effectués en utilisant la pondération « entreprise » (cf. encadré 2).

Champ : entreprises formées d'au moins deux établissements de 20 salariés et plus, du secteur marchand non agricole.

Source : volet « Représentants de la direction » des enquêtes REPONSE 1998 et 2004.