

No 11-522-XIF au catalogue

**La série des symposiums internationaux  
de Statistique Canada - Recueil**

**Symposium 2005 : Défis  
méthodologiques reliés aux  
besoins futurs d'information**



2005



Statistique  
Canada

Statistics  
Canada

Canada

## INFORMATION SUR LA VARIANCE À L'INTENTION DES UTILISATEURS DES DONNÉES

John Wood et Markus G. Šova<sup>1</sup>

### RÉSUMÉ

La façon habituelle de présenter l'information sur la variance aux utilisateurs des données consiste à publier des estimations de la variance ou des statistiques connexes. Cependant, des statistiques aussi simples ne permettent pas aux utilisateurs de bien comprendre comment ils devraient interpréter la variance et ce qu'ils pourraient faire pour en atténuer les effets ou pour tenir compte de son incidence dans leur interprétation des statistiques. L'article traite de certaines de ces questions générales. Nous y décrivons les effets de la variance sur l'interprétation des statistiques publiées, ainsi que la façon dont les utilisateurs et les producteurs des statistiques officielles pourraient faciliter l'analyse en présence de variance. Nous examinons aussi les sources éventuelles de variance et les meilleures mesures qui pourraient être prises afin de réduire cette dernière. Enfin, nous examinons ce qu'il en coûte financièrement aux producteurs et aux utilisateurs de données de réduire ou non la variance, ainsi que la façon dont les coûts associés à la production de statistiques plus exactes pourraient être compensés par les avantages financiers liés à une plus grande précision des données.

MOTS CLÉS : Analyse de variance; Analyse de rentabilité; Coûts de variance; Mesures de variance.

### 1. INTRODUCTION

Ces dernières années, de grands progrès ont été réalisés quant à la production et à la publication de mesures de la qualité des statistiques officielles. Bien que l'on ait défini de multiples attributs de la qualité des données, comme la pertinence, l'actualité et la cohérence, celui sur lequel nous nous concentrerons ici est l'exactitude. Plus précisément, notre discussion portera sur des questions concernant la variance d'échantillonnage et ne fera que brièvement allusion à d'autres aspects de l'exactitude, comme le biais et les erreurs non dues à l'échantillonnage. Notre propos n'est pas de nier l'importance de ces autres aspects de l'exactitude ou d'autres attributs de la qualité. Au contraire, les arguments que nous présentons pourraient et devraient être étendus à ces aspects, mais le faire ici compliquerait énormément l'exposé. Nous nous concentrons sur la variance en partie à cause de son rôle central dans la détermination de l'exactitude des statistiques publiées, mais aussi parce que cela nous permet de simplifier et de clarifier l'exposé et que certains exemples numériques illustratifs peuvent être obtenus facilement.

La question de savoir quels renseignements sur la variance il faut présenter aux utilisateurs des données et de quelle façon n'est pas aussi simple qu'il y paraît à première vue. L'approche classique de présentation de l'information sur la variance aux utilisateurs des données consiste à publier des estimations de la variance ou des statistiques connexes, comme les erreurs-types, les coefficients de variation, les limites de confiance ou de simples systèmes de classement. Cependant, ces simples statistiques ne permettent pas aux utilisateurs de bien comprendre comment ils devraient interpréter la variance ni ce qu'ils pourraient faire pour en atténuer les effets ou, du moins, pour tenir compte de son incidence sur l'interprétation des statistiques. Dans le présent article, nous abordons certaines de ces questions générales. À la section 2, nous décrivons l'information sur la variance habituellement fournie par les producteurs des données, à quels utilisateurs cette information pourrait être utile et quel usage il pourrait en être fait. À la section 3, nous utilisons un exemple précis afin d'examiner les effets de la variance sur l'interprétation des statistiques publiées et ce que les utilisateurs et les producteurs des statistiques officielles pourraient faire pour faciliter l'analyse en présence de variance. À la section 4, nous examinons les sources éventuelles de variance et envisageons quelles seraient les meilleures mesures à prendre pour réduire cette dernière en contrôlant la taille de l'échantillon, la répartition de l'échantillon et les taux de réponse. À la section 5, nous considérons l'utilisation

---

<sup>1</sup> John Wood, Office for National Statistics, Government Buildings, Cardiff Road, Newport, NP10 8XG, United Kingdom (courriel [john.wood@ons.gov.uk](mailto:john.wood@ons.gov.uk)); Markus Šova, Office for National Statistics, Government Buildings, Cardiff Road, Newport, NP10 8XG, United Kingdom (courriel [markus.sova@ons.gov.uk](mailto:markus.sova@ons.gov.uk))

éventuelle d'information sur la variance pour déterminer la répartition des ressources. À la section 6, nous examinons les coûts et les avantages financiers de la réduction de la variance et comment nous pourrions trouver un compromis entre les coûts associés à la production de statistiques plus exactes et les avantages financiers d'une plus grande précision. Enfin, à la section 7, nous résumons brièvement nos arguments en faveur d'autres moyens de présenter l'information sur la variance.

Tout au long de l'article, nous abordons ce sujet du point de vue des statistiques économiques produites par l'Office for National Statistics (ONS) du Royaume-Uni (R.-U.), mais nous espérons que les grands principes qui sous-tendent les exemples présentés seront applicables de façon plus générale, internationalement et dans tous les domaines de la statistique officielle. Notre but n'est pas de fournir des réponses catégoriques, mais de soulever des questions quant aux moyens les plus appropriés de présenter l'information sur la variance et de susciter un débat à ce sujet.

## 2. MESURES CLASSIQUES DE LA VARIANCE

Jusqu'à présent, les statistiques officielles publiées au Royaume-Uni ont inclus une grande variété de mesures classiques de la variance, comme les erreurs-types, les coefficients de variation, les intervalles de confiance, les systèmes de classement et l'identification de différences « statistiquement significatives ». Ces mesures sont toutes bien connues des statisticiens professionnels, mais ne sont pas jugées utiles, voire même intelligibles, par tous les utilisateurs. La gamme de connaissances des concepts statistiques que possèdent les utilisateurs des produits statistiques est très variée, allant de ceux qui les maîtrisent bien, comme les producteurs des statistiques, les décideurs, les universitaires et les chercheurs à ceux qui n'ont que peu de connaissances, voire aucune, des concepts statistiques, comme les gens d'affaires, les représentants des médias, les politiciens et le grand public, en passant par ceux qui ont une connaissance limitée, comme les organismes supranationaux tels que l'Union européenne, les ministères gouvernementaux et les organismes des administrations publiques locales. Il convient de souligner que nous incluons les producteurs de statistiques parmi les utilisateurs car, par exemple, les comptes nationaux sont produits d'après une gamme variée de statistiques officielles.

La grande diversité des utilisateurs, ayant des besoins différents et des niveaux différents de connaissances statistiques, signifie que l'information sur la variance doit être taillée sur mesure afin de répondre à ces différents besoins et niveaux de compréhension. Même aux utilisateurs qui comprennent les concepts statistiques, comme les responsables de l'élaboration des politiques et les universitaires, les mesures classiques de la variance pourraient ne pas fournir les renseignements les mieux adaptés à leurs besoins et objectifs particuliers. Nous présenterons des exemples de tels cas dans d'autres sections de l'article.

Par conséquent, nous suggérons que les statisticiens officiels doivent envisager d'utiliser une plus grande gamme de mesures possibles de la variance afin de fournir aux utilisateurs, des renseignements qui les aident à interpréter et à utiliser les statistiques. Cela pourrait nécessiter des mesures ayant un contenu analytique plus important que de simples énoncés de l'exactitude numérique.

Certaines mesures classiques visent déjà à fournir un certain contenu analytique. Ainsi, les systèmes de notation aident à faire disparaître la mystique statistique qui entoure les erreurs-types et les coefficients de variation en les convertissant en grandes catégories indiquant la qualité générale des statistiques. Ces systèmes sont plus facilement compris par les utilisateurs non statisticiens. Reste toutefois à savoir si les utilisateurs les jugent utiles, s'ils les utilisent intelligemment ou si les fourchettes présentées répondent à leurs besoins.

Certaines publications portant sur les statistiques sociales mentionnent explicitement les différences « statistiquement significatives ». Habituellement, elles ont trait à la comparaison de caractéristiques entre divers segments de la population et sont couvertes par une déclaration du genre « les différences mentionnées dans le texte sont statistiquement significatives sauf indication contraire ». Cette approche pourrait être adoptée plus fréquemment en statistique économique. Chaque mois, des statistiques sur l'inflation et la production économique sont publiées avec une précision du dixième de points de pourcentage. Cependant, leur exactitude ne correspond généralement pas à ce niveau de précision. Malheureusement, pour la plupart de ces statistiques, il n'existe aucune estimation de la variance permettant d'évaluer leur exactitude, quoique l'ONS soit en train de se pencher sur la question pour les statistiques du Royaume-Uni.

Néanmoins, nous soutenons que le contenu analytique des mesures classiques de la variance ne va pas suffisamment loin. Les exemples que nous donnons aux sections qui suivent illustrent d'autres mesures possibles de la variance qui, selon nous, pourraient fournir des renseignements d'un usage plus direct pour les utilisateurs des statistiques officielles.

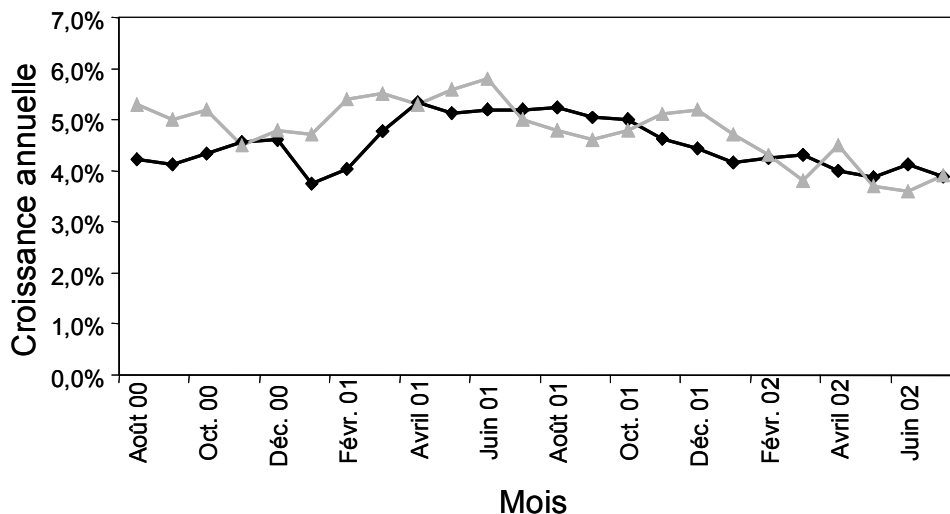
### 3. EFFETS DE LA VARIANCE

Nous commençons notre analyse par un examen de l'effet de la variance sur les statistiques publiées dans un cas précis. Il y a quelques années, l'ONS a réalisé un projet ayant pour but d'élaborer des estimateurs de la variance d'échantillonnage des mouvements de l'Indice des Gains Moyens (IGM) au Royaume Uni (Average Earnings Index AEI — voir Youll, 2002). Nous avons constaté que ces variances pouvaient être raisonnablement bien représentées par un modèle d'autocorrélation de premier ordre des taux de croissance mensuels. La variance ( $\sigma^2$ ) du taux de croissance mensuel de l'IGM est approximativement constante au cours du temps. Le coefficient de corrélation ( $\rho$ ) de la covariance entre les taux mensuels pour des mois adjacents est également approximativement constant et négatif. La covariance entre les taux de croissance mensuels pour des mois non adjacents est négligeable. Brièvement, nous pouvons représenter la variation mensuelle  $g_i$  de l'IGM pour le mois  $i$  par le processus AR(1) suivant :  $E[g_i]=\mu$ ;  $Var[g_i]=\sigma^2$ ;  $Cov[g_i, g_j]=\rho\sigma^2$  ( $|i-j|=1$ );  $Cov[g_i, g_j]=0$  ( $|i-j|>1$ ).

Pour l'Indice d'Économie Globale, excluant les primes de rendement,  $\sigma$  est égal à environ 0,15 point de pourcentage et  $\rho$ , à environ -0,3. Ces dernières années, la croissance des gains a été assez stable, soit environ 4 % par année, ou  $\mu=0,33$  % par mois. En utilisant ces paramètres, nous pouvons produire une réalisation aléatoire de ce modèle simple aux fins de comparaison avec la croissance réelle de l'indice.

La figure 1 qui suit montre, pour l'IGM, le taux d'inflation sur 12 mois (c.-à-d. l'accroissement de l'indice courant par rapport à l'indice 12 mois plus tôt), qui est la statistique intéressant le plus les utilisateurs, pour la période allant d'août 2000 à juillet 2002. Les données sont un peu dépassées, mais cela n'a aucune incidence sur le principe que nous voulons démontrer. L'une des courbes de la figure 1 représente l'inflation des gains mesurée par l'IGM et l'autre, une réalisation aléatoire du modèle d'autocorrélation décrit ci-dessus.

Figure 1. Croissance mensuelle estimée de l'IGM du R.-U. (pour l'ensemble de l'économie) comparativement à une réalisation d'un modèle AR (1).



À notre avis, même l'observateur ayant les connaissances statistiques les plus poussées aurait de la difficulté à déterminer, avec confiance, quelle courbe correspond à quoi. La leçon importante à tirer de cet exemple est que, dans certaines circonstances, il est presque impossible de faire la distinction entre les statistiques publiées et le bruit aléatoire autour d'un niveau de référence stable.

En octobre 2003, le gouverneur de la Banque d'Angleterre a prononcé un discours dans lequel il a fait allusion à la « confusion statistique régnant sur l'économie britannique » (King, 2003). Cette remarque a suscité un vif intérêt dans la presse financière et certaines critiques à peine voilées, des statisticiens officiels. Une autre déclaration, moins remarquée, mais extrêmement pertinente, figurait dans ce même discours, en fait dans le même paragraphe : « après une décennie de stabilité incomparable de la croissance et de l'inflation, il est temps de faire le point. » Cette « décennie de stabilité sans précédent » a créé exactement les conditions sous lesquelles nos statistiques ressemblent à un bruit aléatoire autour d'un niveau de référence stable. La question qui se pose donc est celle de savoir comment, dans de telles circonstances, nous devons communiquer aux utilisateurs de nos données que nos statistiques se situent dans le signal stable et non dans le bruit.

Pour étudier cette question, nous pouvons nous inspirer de Steel et Smith (2005), quoique l'analyse présentée ici soit nettement moins complexe que la leur. Au moyen de notre simple modèle d'autocorrélation, nous pouvons déterminer combien de temps il faudrait pour déceler un bond soudain faisant passer l'inflation des gains de 0,33 % par mois à 0,5 % par mois (c.-à-d. de 4 % à 6 % par année). Nous appuyons cette évaluation sur une simple règle décisionnelle : si le taux observé d'inflation sur 12 mois excède la limite supérieure de confiance à 95 % pour notre état stable de 4 % par année, nous inférons que l'inflation des gains a augmenté. Le tableau 1 montre, pour deux valeurs différentes de  $\sigma$ , les probabilités de violer cette limite après divers nombres de mois au taux d'inflation plus élevé.

Tableau 1: Probabilités (en %) de violer la limite supérieure de confiance à 95 % après  $n$  mois d'inflation annuelle 2 % plus élevée

$n$ : nombre de mois à un taux d'inflation annuelle 2 % plus élevé	$\sigma=0,15, \rho=-0,3$ (gains, primes non comprises)	$\sigma=0,30, \rho=-0,3$ (gains, y compris les primes)
1	12	8
3	42	18
6	89	42
9	100	69
12	100	89

Pour  $\sigma=0,15$ , correspondant aux gains, primes non comprises, nous devons attendre six mois avant d'atteindre une probabilité acceptablement élevée (89%) de déceler cette soudaine variation importante de l'inflation sous-jacente.

Pour  $\sigma=0,3$ , correspondant aux gains, y compris les primes, la situation est encore pire et nous devons attendre jusqu'à 12 mois avant d'être raisonnablement certain de déceler la variation.

Ici, le problème tient en partie à la mesure cible privilégiée : le taux d'inflation sur 12 mois ne convient pas bien au dépistage rapide des variations soudaines, récentes. L'exemple susmentionné suscite deux questions. Des mesures du type proposé, quantifiant l'effet de la variance sur la prise de décision, seraient-elles plus utiles pour les utilisateurs des données que de simples erreurs-types ou intervalles de confiance? Pouvons-nous nous appuyer sur notre connaissance de la structure de variance-covariance de nos statistiques pour proposer de meilleures mesures cibles, qui permettraient aux utilisateurs des données de repérer plus rapidement et plus facilement les points de renversement économique ou d'autres caractéristiques?

#### 4. SOURCES DE VARIANCE

L'une des raisons pour lesquelles l'ONS a entrepris de produire des estimateurs de la variance pour l'IGM était de pouvoir évaluer l'effet des changements d'échantillon sur l'exactitude de l'indice. Ces changements étaient les suivants : une nouvelle répartition de l'échantillon, un accroissement éventuel de la taille de l'échantillon et l'effet de l'accroissement des taux de réponse. Le tableau 2 donne les erreurs-types estimées pour chacun de ces changements éventuels comparativement à l'échantillon en vigueur au moment de l'étude.

Tableau 2 : Erreurs-types moyennes (d'août 2001 à juillet 2002) estimées (en points de pourcentage) des taux de croissance sur 12 mois de l'IGM

	Primes non comprises	Y compris les primes
Échantillon précédent	0,35	0,66
Échantillon reréparti	0,32	0,65
Échantillon 50 % plus grand	0,29	0,61
Échantillon 100 % plus grand	0,27	0,57
Échantillon à réponse totale	0,24	0,39

Si nous examinons les trois modifications éventuelles de l'échantillon l'une après l'autre, le tableau 2 donne les résultats suivants.

a. Répartition de l'échantillon

La nouvelle répartition proposée n'a produit qu'une légère réduction des erreurs-types estimées. Ce résultat est plutôt rassurant, parce qu'il indique que la répartition originale, qui était fondée sur des renseignements limités, était déjà proche de la situation optimale.

b. Taille de l'échantillon

L'accroissement de la taille de l'échantillon de 50 % réduit très peu les erreurs-types. Même le doublement de la taille de l'échantillon ne les réduit pas beaucoup (réduction n'excédant pas 15 % à 20 %).

c. Non-réponse

L'accroissement des taux de réponse est la meilleure source éventuelle d'une réduction appréciable de l'erreur-type. La plupart de l'avantage émanant d'un échantillon à réponse totale est attribuable aux plus grandes entreprises.

À l'époque, il était question que le Trésor du Royaume-Uni accorde plus de fonds afin d'augmenter la taille de l'échantillon, à condition que cela améliore l'exactitude des statistiques sur l'inflation des gains. Étant donné l'effet possible limité sur les erreurs-types, l'ONS n'a reçu aucuns fonds supplémentaires du Trésor. Cependant, il a mis en place un programme de suivi en vue d'accroître les taux de réponse ayant pour cible les plus grandes entreprises. La question de savoir si cette mesure a permis de réaliser la réduction prévue des erreurs-types est à l'étude.

Nous voulons en venir ici au fait que l'information sur les sources de variance peut aider les utilisateurs à prendre leurs propres décisions. Connaître l'effet qu'ont sur la variance les modifications des ressources affectées à une enquête peut aider les organismes qui financent l'enquête à décider du niveau de financement. Si l'on se penche de façon plus générale sur les fournisseurs de données d'enquête, l'information sur l'effet de la non-réponse peut-elle être utilisée afin d'encourager les participants aux enquêtes à répondre, particulièrement ceux ayant un effet important sur les estimations résultantes?

Nous pouvons aussi examiner les sources de variance de façon plus générale. L'ONS a achevé récemment un projet visant à estimer les erreurs-types des variations de l'indice de production (IdP) du Royaume-Uni, plus précisément pour le taux de croissance sur 12 mois de septembre 2003 à septembre 2004. Le processus d'estimation nous a permis d'extraire la contribution à la variance totale des quatre enquêtes intervenant dans l'établissement de l'indice, comme le montre le tableau 3.

Les très faibles contributions des mouvements des stocks, des indices des prix à l'exportation et même des indices des prix à la production sont assez étonnantes, mais sont en accord avec plusieurs contrôles de cohérence que nous avons appliqués afin de confirmer la validité des résultats. Cependant, ces résultats sont un exemple d'utilisation pratique d'une analyse de variance d'un aussi haut niveau : la dépendance extrême de l'exactitude de l'IdP à l'égard de l'enquête connexe sur les chiffres d'affaires oblige à se demander si les estimations du chiffre d'affaires calculées d'après cette enquête sont suffisamment exactes pour l'usage prévu. Cela, conséquemment, dépend de l'importance relative de l'IdP dans l'ensemble plus général des statistiques économiques de l'ONS.

Tableau 3: Contribution estimée à la variance de la croissance sur 12 mois (de septembre 2003 à septembre 2004) de l'indice de production du Royaume-Uni

Enquête source	Contribution à la variance ( $\times 10^4$ )	Contribution à la variance (%)
Chiffre d'affaires	0,591	93,8
Mouvements des stocks	0,021	3,3
Indices des prix à la production	0,009	1,4
Indices des prix à l'exportation	0,009	1,4

Malheureusement, l'information sur la variance n'est pas facile à obtenir pour cet ensemble plus général de statistiques. Nous pouvons, à la place, utiliser l'IdP comme exemple dans un microcosme, de l'utilisation possible de l'information sur la variance pour déterminer la répartition des ressources entre les diverses enquêtes. Nous discutons de ce point à la section suivante.

## 5. UTILISATIONS DE L'INFORMATION SUR LA VARIANCE

Les statisticiens ont l'habitude d'utiliser les renseignements qu'ils possèdent sur la structure probable de variance-covariance des données d'enquête pour essayer d'optimiser le plan d'échantillonnage et la répartition de l'échantillon. Ordinairement, nous considérons cet exercice comme une responsabilité purement professionnelle, mais il ne doit pas toujours en être ainsi. Les utilisateurs sont également concernés par la conception et par la répartition d'un échantillon, parce que les décisions à ce sujet influencent la portée et l'exactitude des données produites. Par exemple, lors d'un projet récent de répartition de l'échantillon utilisé pour établir les indices des prix à la production au Royaume-Uni, nous avons constaté qu'il serait possible d'accroître considérablement l'efficacité en regroupant un grand nombre d'indices de faible niveau en un plus petit nombre d'indices de niveau plus élevé. L'ONS n'a pas encore entrepris d'explorer cette réduction du nombre d'indices, à cause de l'effet probable sur les utilisateurs. Un remaniement et une réallocation aussi importante pourraient entraîner la suppression d'indices populaires, de sorte qu'une consultation approfondie des utilisateurs serait nécessaire. Dans un tel contexte, les coûts de la mise en œuvre d'une répartition inefficace de l'échantillon constitueraient-ils une mesure utile de variance pour les utilisateurs. Une économie de, disons, 0,5 m£ serait-elle suffisante pour justifier le retrait de certains indices populaires? Les utilisateurs seraient-ils disposés à payer une telle somme pour garder leurs indices préférés? La volonté de payer pour les inefficacités du plan d'échantillonnage et de la répartition de l'échantillon pourrait être un indice utile de l'importance effectivement accordée par les utilisateurs à ces indices favoris.

Les utilisateurs ont également un intérêt plus général pour les ressources affectées aux produits statistiques. Même s'ils s'attendent à ce que les bureaux gouvernementaux de la statistique utilisent efficacement l'argent des contribuables, ils pourraient aussi souhaiter qu'autant de ressources que possible soient affectées aux produits statistiques qui les intéressent le plus. Ils pourraient même demander aux bureaux de la statistique de justifier les cas où les produits statistiques qu'ils privilégient ne bénéficient pas d'un financement aussi important qu'ils le souhaiteraient.

Manifestement, déterminer la répartition optimale des ressources entre les produits statistiques serait fort compliqué, mais nous pouvons illustrer une approche possible en utilisant les résultats concernant la variance obtenue pour l'indice de production, dont nous avons discuté à la section précédente. Le tableau 4 donne le budget annuel courant affecté aux quatre enquêtes concernées et une répartition « optimale » théorique, calculée en supposant que la contribution à la variance est un versement proportionnel aux ressources affectées.

Tableau 4 : Coûts annuels (courants et « optimisés ») des enquêtes contribuant à l'indice de production du R.-U., avec l'effet correspondant sur l'erreur-type de la croissance de l'indice sur 12 mois.

Enquête source	Coûts annuels courants	Coûts annuels « optimisés »
Chiffre d'affaires	0,8m£	2,1m£
Mouvements des stocks	0,9m£	0,4m£
Indices des prix à la production	1,1m£	0,3m£
Indices des prix à l'exportation	0,2m£	0,2m£
Erreur-type (points de pourcentage)	0,79	0,56

Dans ces conditions, nous pourrions obtenir une réduction de près de 30 % de l'erreur-type de la croissance sur 12 mois de l'indice de production grâce à une répartition plus efficace des ressources entre les enquêtes sources. Ou bien, sous cette répartition « optimale » théorique, nous pourrions obtenir la même erreur-type au coût de 1,5 M£ seulement, soit une économie de 50 %.

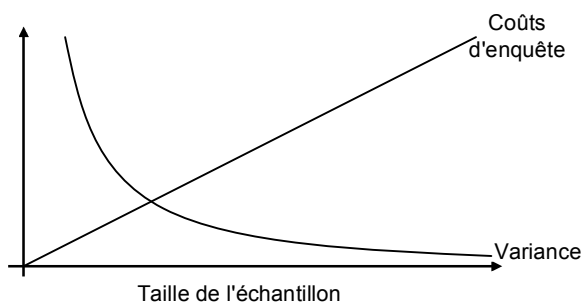
Naturellement, il s'agit d'une énorme sur-simplification. Les enquêtes sources sont utilisées dans d'autres contextes que l'IdP. Les indices des prix servent à la surveillance de l'inflation pour la gestion macroéconomique et pour les clauses de révision des prix dans les contrats juridiques. Les données sur les mouvements des stocks sont utilisées par d'autres secteurs du Système de comptabilité nationale, de sorte que leur effet n'est pas limité à l'IdP. Une analyse minutieuse devrait tenir compte de la contribution agrégée de chaque enquête à tous les produits, tâche écrasante dépassant le cadre du présent exposé.

Néanmoins, étant donné l'analyse qui précède, nous pouvons nous demander si les utilisateurs préféreraient une réduction de 30 % de l'erreur-type ou une économie de 1,5 m£ ou une combinaison de la réduction de l'erreur-type et de la réduction du coût. Pour y répondre, il faut pouvoir attribuer une valeur monétaire à la réduction de l'erreur-type. Nous abordons cette question à la section suivante.

## 6. COÛTS DE VARIANCE

À mesure qu'augmente la taille de l'échantillon d'une enquête, la variance des estimateurs diminue habituellement en proportion à peu près inverse, mais les coûts d'enquête augmentent presque linéairement. La figure 2 illustre schématiquement ces mouvements.

Figure 2. Illustration schématique de la relation entre la taille de l'échantillon et les coûts d'enquête ainsi que la variance.



Pour les petites tailles d'échantillon, nous pouvons obtenir une réduction appréciable de la variance pour un accroissement relativement faible du coût absolu mais, à mesure que la taille de l'échantillon augmente, la loi des rendements décroissants entre en jeu et nous devons affecter de plus en plus de ressources afin d'obtenir une réduction équivalente de la variance. La question est la suivante : Où nous arrêtons-nous? Quel est le compromis entre la variance et le coût? Combien vaut une réduction de 20 % de l'erreur-type?

Pour répondre à cette question, nous nous adonnons à une expérience de pensée (très simplifiée).

Le produit intérieur brut (PIB) annuel du R.-U. est d'environ 1 000 milliards de livres. Chaque mois, trimestre et année, l'ONS produit une vaste gamme de statistiques économiques et financières que les autorités pertinentes utilisent pour gérer l'économie. Nous nous concentrons sur l'un des mécanismes utilisés à cette fin : la gestion des taux d'intérêt.

Nous examinons le progrès des taux d'intérêt et du PIB au cours des six dernières années (voir BoE, 2005 et ONS, 2005a pour les sources des données), et dérivons d'après ceux-ci cette simple règle empirique : une variation du taux de base de la Banque d'Angleterre de 1 % entraîne une variation du PIB de 1 % en sens opposé environ un an plus tard. Il s'agit d'une règle très approximative qui pourrait sans aucun doute être améliorée, mais notre but ici est simplement d'examiner une approche possible, si bien qu'une exactitude parfaite n'est pas essentielle.



Supposons, maintenant, que nous ayons pu réduire la variance de nos statistiques de sorte que la Banque d'Angleterre puisse améliorer d'un mois la détermination du moment où il convient d'appliquer les changements d'intérêt. Normalement, la Banque fait varier le taux de base graduellement d'un quart de point de pourcentage à la fois. Avancer d'un mois une diminution appropriée entraînerait un bénéfice unique de 0,25 % du PIB d'un mois une année plus tard, donnant un bénéfice unique d'environ 200 millions de livres. Pour simplifier, nous supposons qu'un bénéfice semblable émanerait aussi d'une amélioration du moment où ont lieu les augmentations de taux d'intérêt, bien que nous reconnaissons qu'il serait plus difficile de le démontrer.

Ce bénéfice n'a lieu qu'une seule fois, au moment où nous procédons au passage des anciennes variances plus élevées aux nouvelles, plus faibles. Mais nous devons le comparer au coût annuel du maintien des variances réduites. L'investissement de ce bénéfice unique à un taux d'intérêt de 5 % produirait un bénéfice annuel équivalent de 10 millions de livres. Quel serait le coût annuel de la réalisation de ce bénéfice?

En utilisant les résultats du modèle de l'IGM décrits à la section 3, nous obtenons une indication (très approximative) que la réduction nécessaire de la variance pourrait être obtenue en doublant la taille de l'échantillon. Cela pourrait coûter environ 80 millions de livres, qui est le montant annuel affecté à l'heure actuelle par l'ONS à la production des statistiques économiques et financières (ONS, 2005b). De toute évidence, ce changement n'en vaut pas la peine. Par ailleurs, en nous souvenant des résultats obtenus à la section 3, il pourrait être justifié de consacrer plus de ressources à l'accroissement des taux de réponse, parce qu'il s'agit vraisemblablement d'un moyen plus rentable de réduire la variance.

Manifestement, ce genre d'analyse de rentabilité devrait être nettement plus élaborée avant qu'elle puisse fournir un moyen fiable de déterminer, pour la production de statistiques officielles, le montant des ressources financières qui devraient être accordées et comment elles devraient être réparties. Le but de cet exemple assez simple est de démontrer que les analyses de rentabilité de ce type peuvent constituer un outil utile pour communiquer l'information sur la variance aux utilisateurs, parce qu'elles traduisent des concepts statistiques abstraits que les utilisateurs pourraient avoir de la difficulté à comprendre ou à interpréter en termes financiers équivalents qui sont compris de tous.

Les analyses de rentabilité posent évidemment de nombreux problèmes. Elles sont compliquées à réaliser et souvent fondées sur des hypothèses douteuses. Les avantages émanant des améliorations apportées à une enquête peuvent être multiples, et veiller à ce qu'ils soient tous pris en considération et ne soient pas comptés deux fois est une tâche onéreuse. Inversement, un bénéfice basé sur l'amalgamation de plusieurs statistiques distinctes (par exemple, l'IdP, qui est fondé sur l'intégration de données provenant de quatre enquêtes) peut avoir des coûts multiples. Si l'on considère l'ensemble des enquêtes, s'assurer que tous les coûts soient pris en compte et rapprochés des bénéfices appropriés est un problème considérable et de nombreux travaux seront nécessaires pour faire progresser ce genre d'analyses.

Une complication supplémentaire est due au fait que les coûts et les bénéfices sont des phénomènes dynamiques qui doivent être suivis et examinés en permanence. Les instituts nationaux de statistique doivent répondre sans cesse à de nombreuses nouvelles demandes à l'aide de ressources limitées. Par exemple, au Royaume-Uni, les statistiques macroéconomiques sur le secteur des services sont de qualité assez médiocre, alors que le secteur des services représente environ 70 % du PIB. À la suite d'un examen récent (Allsopp, 2004), il a été recommandé d'instituer un programme d'amélioration de ces statistiques. Le même examen a également poussé à recommander de créer un programme en vue de mettre au point de meilleures statistiques régionales, quoiqu'il ne soit pas certain que la qualité des résultats justifie les ressources nécessaires et que la question n'ait pas été abordée durant l'examen. Évaluer les mérites relatifs de ces multiples demandes, en termes de coûts et de bénéfices, n'est pas une tâche facile.

Enfin, nous ne nous sommes pas penchés sur les questions du biais et des erreurs non dues à l'échantillonnage. Celles-ci doivent également être prises en considération lors de l'analyse des coûts et des avantages, même si l'évaluation de ces sources d'erreurs est très difficile. Néanmoins, nous soutenons que les statisticiens officiels doivent tenir compte des coûts et des avantages associés à la production des statistiques officielles, afin de justifier les ressources financières qui y sont consacrées et de permettre aux utilisateurs de comprendre la valeur de ces statistiques.

## 7. SOMMAIRE

Pour conclure, nous revenons à la question initiale : de quelle information sur la variance les utilisateurs ont-ils besoin? Nous avons soutenu que les mesures statistiques classiques, telles que les erreurs-types et les intervalles de confiance, ne permettent pas de répondre aux besoins des utilisateurs. À notre avis, les statisticiens officiels doivent penser plus créativement, afin de fournir des renseignements sur la variance ayant un intérêt direct et pratique pour les utilisateurs. Cela nécessitera vraisemblablement des mesures de la variance ayant un contenu plus analytique. Au niveau élémentaire, nous pourrions être plus directs, particulièrement en ce qui concerne les statistiques économiques, en indiquant aux utilisateurs quand les différences sont statistiquement significatives ou, fait tout aussi important, quand elles ne le sont pas. Mais nous devons également aller plus loin et concevoir des mesures qui répondent directement aux préoccupations des utilisateurs. Quel est l'effet de la variance sur les statistiques que nous publions? Les mesures cibles ayant la faveur des utilisateurs sont-elles sensées? Pouvons-nous utiliser l'information dont nous disposons sur la variance pour élaborer d'autres mesures qui sont des indicateurs plus exacts de ce que les utilisateurs veulent réellement savoir? Existe-t-il un moyen plus efficient ou plus efficace de répondre aux besoins des utilisateurs en répartissant les ressources différemment? Quel est le prix que les utilisateurs sont disposés à payer pour obtenir des statistiques plus exactes? Comment pouvons-nous évaluer l'avantage financier associé à la réduction de la variance?

Le présent article ne fait qu'effleurer ces questions profondes et difficiles. En dernière analyse, l'enjeu est de savoir dans quelle mesure nous pouvons y répondre.

## RÉFÉRENCES

- Allsopp, C. (2004), *Review of Statistics for Economic Policymaking – Final Report*, Norwich: HMSO, <http://www.hm-treasury.gov.uk/allsopp>
- BoE (2005), Repo Rate History, Bank of England, <http://213.225.136.206/mfsd/iadb/Repo.asp?Travel=NIx>
- King, M. (2003), Speech to the East Midlands Development Agency/Bank of England Dinner on 14<sup>th</sup> October, 2003, Bank of England, <http://www.bankofengland.co.uk/publications/speeches/2003/speech204.pdf>
- ONS (2005a), *GDP Growth*, Office for National Statistics, <http://www.statistics.gov.uk/CCI/nugget.asp?ID=192&Pos=&ColRank=1&Rank=374>
- ONS (2005b), *National Statistics Annual Report 2004/05*, Office for National Statistics, [http://www.statistics.gov.uk/about\\_ns/downloads/NSAnnualReport0405.pdf](http://www.statistics.gov.uk/about_ns/downloads/NSAnnualReport0405.pdf)
- Steel, D. et Smith, P. (2005), “La communication de variances servant à interpréter les changements et le points d'inflexion lors d'enquêtes à passages répétés”, *Recueil du Symposium 2005 de Statistique Canada*
- Youll, R. (2002), Quality of the estimates of earnings growth from the AEI, *Labour Market Trends*, Vol. 110, No. 4, April 2002, pp 207-213, Office for National Statistics