

Rénovation de l'enquête CAMME

De l'échantillonnage au calcul de précision, en passant
par l'allocation optimale entre les enquêteurs

JMS

14 mars 2005





1. Introduction.

Contexte de l'enquête CAMME et de sa refonte

JMS

14 mars 2005





L'enquête CAMME

- Enquête mensuelle de conjoncture auprès des ménages
 - Chaque mois, 1.100 ménages rejoignent le panel
 - interrogés trois mois de suite par téléphone
 - Par un réseau d'enquêteurs, dans 8 DR de gestion
- Connaître l'OPINION des ménages sur
 - leur environnement économique général
 - leur environnement économique personnel
- Etudier et anticiper les comportements
 - en matière de consommation et d'épargne

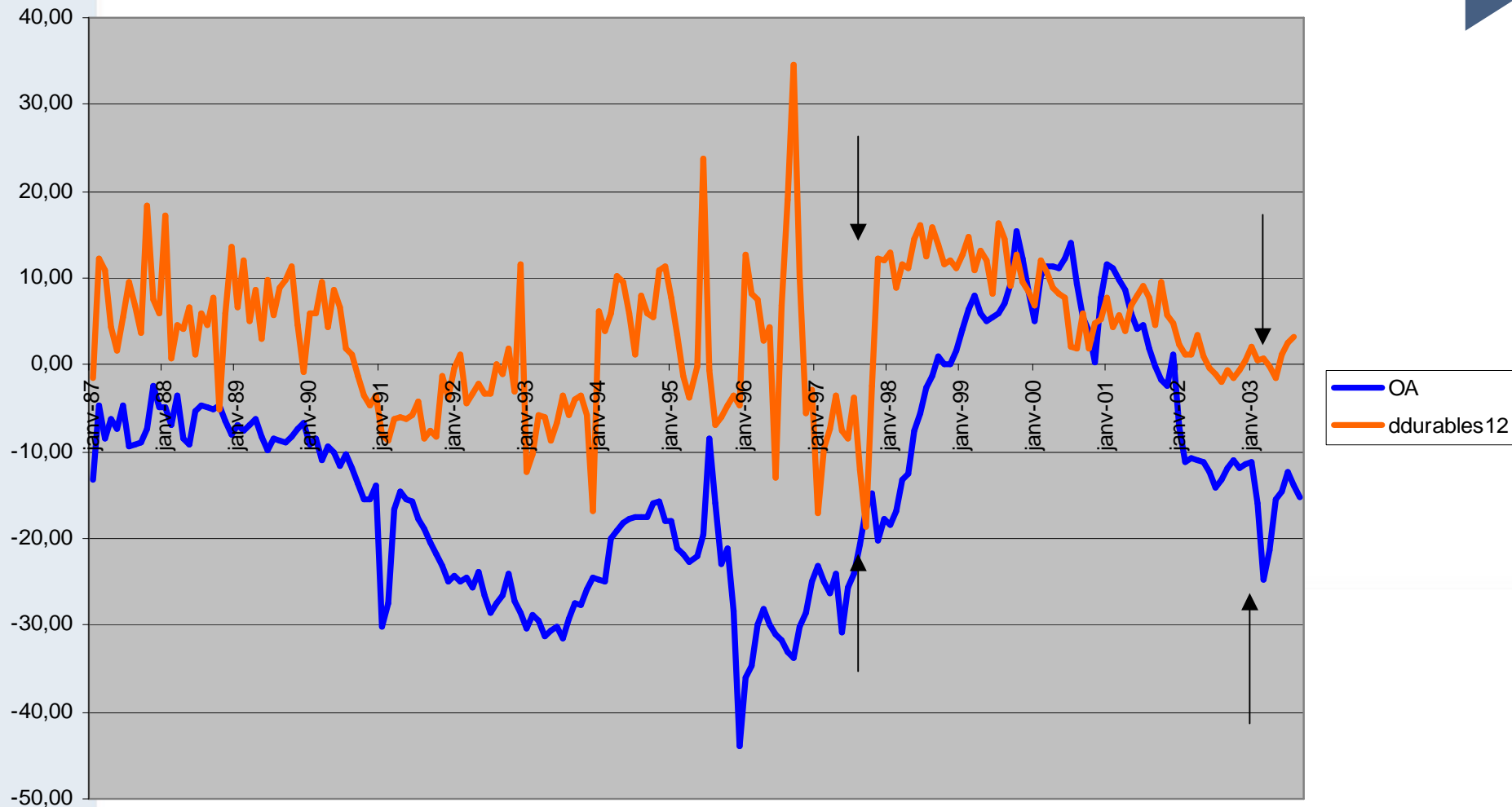


L'indicateur résumé : le « moral des Français »

- › C'est plus l'ÉVOLUTION des indicateurs que leur niveau qui entre en ligne de compte
 - ➔ calcul des Soldes d'opinion

- › L'indicateur résumé = Moyenne arithmétique des soldes sur :
 - le niveau de vie passé,
 - le niveau de vie futur,
 - la situation financière passée,
 - la situation financière future,
 - et l'opportunité d'acheter

Interprétation des indicateurs





Publication : l' « Info Rapide »

- > publiée au début de chaque mois
 - ➔ très rapidement diffusée et commentée
- > Revue de presse – résultats de février 2005 :
 - AFP Général, 28 février – 08h52 : *Le moral des ménages en légère amélioration en février*
 - AFP Economie – 10h01 : *Commentaires sur la confiance des ménages en France*
 - AFP, commentaire des analystes – 13h11 : *Embellie surprise du moral des ménages, malgré le chômage*
 - AFP World news – 14h33 : *French consumer resists unemployment rise but still weak...*

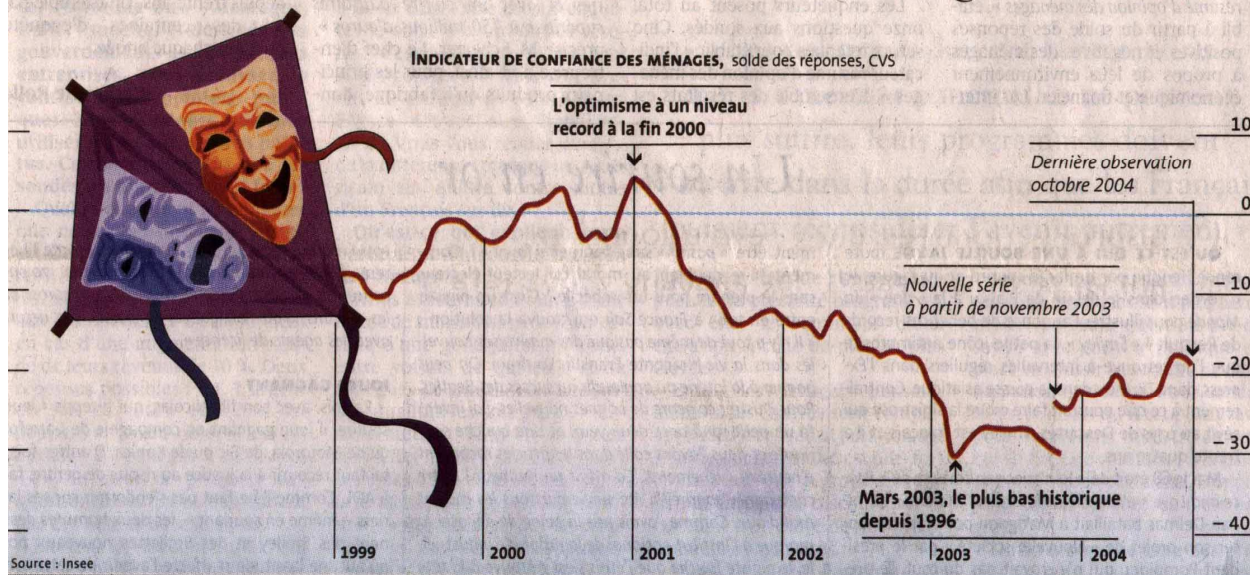
Le moral ne se décrète pas, il se mesure

DE L'INSEE À LA BANQUE DE FRANCE
EN PASSANT PAR EUROSTAT, LES INDICATEURS
DE CONFIANCE SONT MÉDIOCRES,
MAIS QUE REFLÈTENT-ILS RÉELLEMENT ?

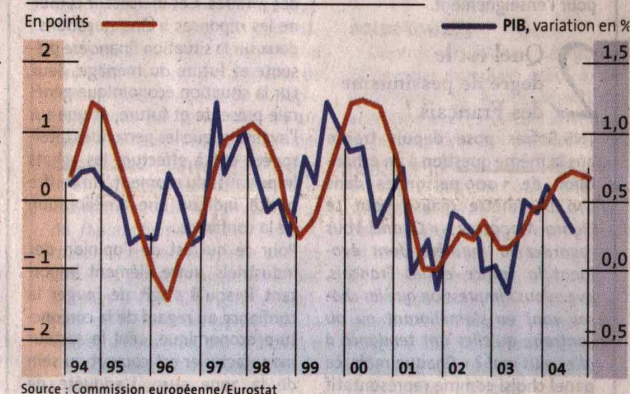
À LIRE DANS LE DOSSIER

- > Quand l'Insee sonde les cœurs et les portefeuilles p. II
- > Un sourire en or p. II
- > Les économistes préfèrent scruter les comportements réels p. II
- > Pierre Giacometti, codirecteur général d'Ipsos France : « Les Français transmettent leur inquiétude d'une génération à l'autre » p. III
- > Aux Etats-Unis, un indice attendu le dernier mardi de chaque mois p. III

INDICATEUR DE CONFIANCE DES MÉNAGES, solde des réponses, CVS



INDICE DU CLIMAT DES AFFAIRES DANS LA ZONE EURO



UNE MOROSITÉ ÉGALEMENT EUROPÉENNE

LE BLUES DE L'HEXAGONE



La Refonte : les raisons

- › Nouvelles attentes de la Commission Européenne :
 - Nouvelles questions
 - Nouveaux tableaux (forme, redressements)
 - Application plus stricte et harmonisation
- › Signes de fragilité de l'ancienne chaîne de traitement
- › Moderniser et adapter le tirage des échantillons
- › Optimiser l'allocation entre enquêteurs des échantillons
- › Intégrer des calculs de précision



2. L'échantillonnage.

JMS

14 mars 2005





Une base de sondage incomplète.

- › Ne couvre que les abonnés au téléphone fixe
- › Ne couvre pas les abonnés inscrits sur liste d'opposition.
→ 25 % environ des titulaires d'une ligne fixe...
- › L'unité statistique est la ligne et non le logement.
Le répondant est le titulaire de la ligne ou son conjoint.
- › Au total, environ 18 millions d'abonnés « joignables »
(particuliers seulement, hors doublons)



Objectif : améliorer et maîtriser l'échantillonnage

Redresser les défauts de structure dus aux listes d'opposition.

- Améliorer la précision au moyen d'une stratification :
 - Par type de commune (rurales,... ,urbaines)
 - Par type d'habitat (individuel / collectif)
- Complétée d'une stratification implicite par tri du fichier :
 - Critère géographique
 - Code typologique reflétant le niveau de revenu



Les principes d'estimation.

- Chaque unité i va être tirée avec une probabilité d'inclusion Π_i
- Un estimateur sans biais du total d'une variable d'intérêt Y sera donné par :

$$\hat{T} = \sum_{i \in S} \omega_i Y_i = \sum_{i \in S} \frac{1}{\Pi_i (1 - \tau_i)} Y_i = \sum_{i \in P_1} \frac{1}{\Pi_i (1 - \tau_i)} \mathbf{1}_{i \in S} Y_i$$

- c'est-à-dire qu'il y a dilatation de la variable d'intérêt pour extrapoler à l'ensemble des lignes...



- ... sur la base d'une ***hypothèse d'homogénéité des comportements*** entre les joignables et les non joignables.
- τ_i est le taux de pénétration de la liste rouge pour l'unité i.
- Chaque unité a donc le poids

$$\omega_i = \frac{1}{\prod_i (1 - \tau_i)}$$



Comment choisir les probabilités de sélection ?

- On impose l'équipondération de toutes les unités tirées, quelle que soit la strate.
- On en déduit l'allocation par strate de l'échantillon :

$$n_h = n \frac{\sum_{i \in P_{1,h}} \frac{1}{(1-\tau_i)}}{\sum_k \sum_{j \in P_{1,k}} \frac{1}{(1-\tau_j)}}$$



> ... et les probabilités d'inclusion :

$$\Pi_i = n \frac{1}{1 - \tau_i} \frac{1}{\sum_k \sum_{j \in P_{1,k}} \frac{1}{1 - \tau_j}}$$



- › Ces formules se simplifient lorsqu'on admet que le taux de pénétration de la liste rouge est constant sur des zones géographiques données (idéalement IRIS, en pratique commune) :

$$\Pi_i = \frac{n}{N} \frac{1}{1 - \tau_i}$$

avec une allocation :

$$n_h = n \frac{N_h}{N}$$



Répartition de l'échantillon mensuel entrant

TUAG →	0	1	3	5	7	8	TOTAL
Type d'habitat ↓							
Individuel	236	100	80	72	98	39	625
Collectif	22	25	49	74	157	148	475
TOTAL	258	125	129	146	255	187	1100



3. ALLOCAMME : *allocation optimale de l'échantillon entre les enquêteurs.*

JMS

14 mars 2005





Objectif : allocation « harmonieuse »

- › Répartir l'échantillon mensuel entrant entre les différents enquêteurs...
 - › ... sous contrainte de charge par enquêteur
 - › ... en favorisant la proximité entre enquêteurs et enquêtés
- construction d'un algorithme d'optimisation.



Formalisation de la méthode.

- > *Notations* :
 - s = échantillon (ou, plus généralement, population de référence à répartir) de taille n .
 - E = ensemble des enquêteurs, de taille K .
- > Pour chaque enquêteur $e_k \in E$, on définit une charge utile C_k

- > On suppose $\sum_{k=1}^K C_k = n$



Expression du problème :

- Chercher une partition d'un ensemble s en K classes notées A_k
- Chaque classe représentant le sous-ensemble des individus affectés à l'enquêteur k , telles que :

$$\forall k \in \{1, 2, \dots, K\} : \text{card}(A_k) = C_k$$

- et minimisant la fonction :
- $$\sum_{k=1}^K \sum_{i \in A_k} d(i, e_k)$$



L'algorithme proposé :

- Étape initiale :
 - attribuer chaque n° tiré à l'enquêteur le plus proche sans tenir compte des contraintes de charge

- Étape courante :
 - réaliser des transferts de n°s des enquêteurs en surcharge vers d'autres enquêteurs...

 - ... en essayant de détériorer le moins possible le critère à optimiser.



Exemple

	ENQUÊTEURS		
UNITES ECHANTILLONS	1	2	3
1	10	20	90
2	10	40	100
3	70	10	60
4	30	30	10
Charge utile	1	1	2



Etape initiale

UNITES ECHANTILLONS	ENQUÊTEURS		
	1	2	3
1	10	20	90
2	10	40	100
3	70	10	60
4	30	30	10
Charge utile	1	1	2

→ Enquêteur 1 en surcharge (charge utile = 1)



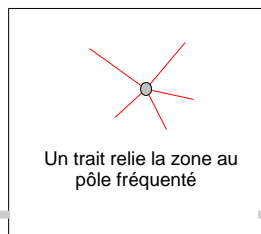
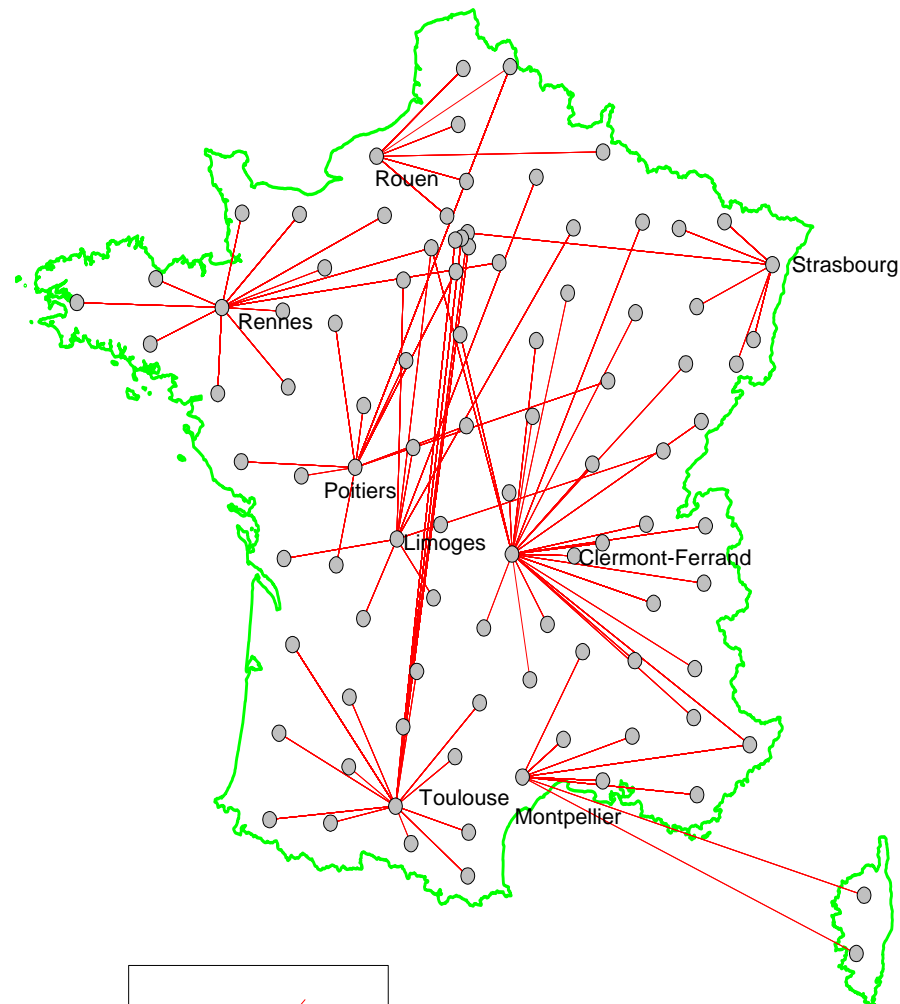
Etape de transfert

	ENQUÊTEURS		
UNITES ECHANTILLONS	1	2	3
1	10	20	90
2	10	40	100
3	70	10	60
4	30	30	10
Charge utile	1	1	2

→ Enquêteur 2 en surcharge temporaire

→ Valeur finale du critère = 100

Application à CAMME





4. Traitements statistiques aval

JMS

14 mars 2005





Révision et amélioration de l'aval

- › Une logique de simplicité des traitements mis en œuvre :
 - imputation de la Non Réponse Partielle ;
 - calages sur marge → pour les besoins de la Commission

- › Priorité à la robustesse pour les modèles d'imputation

- répondre aux contraintes fortes imposées en termes de délai d'élaboration des résultats.



5. Calculs de précision

JMS

14 mars 2005





L'estimation d'un solde

› L'estimateur d'un solde

$\hat{Y}_{m,j}$ = différence entre les pourcentages
de réponses positives et négatives
(mois m , item Y_j)

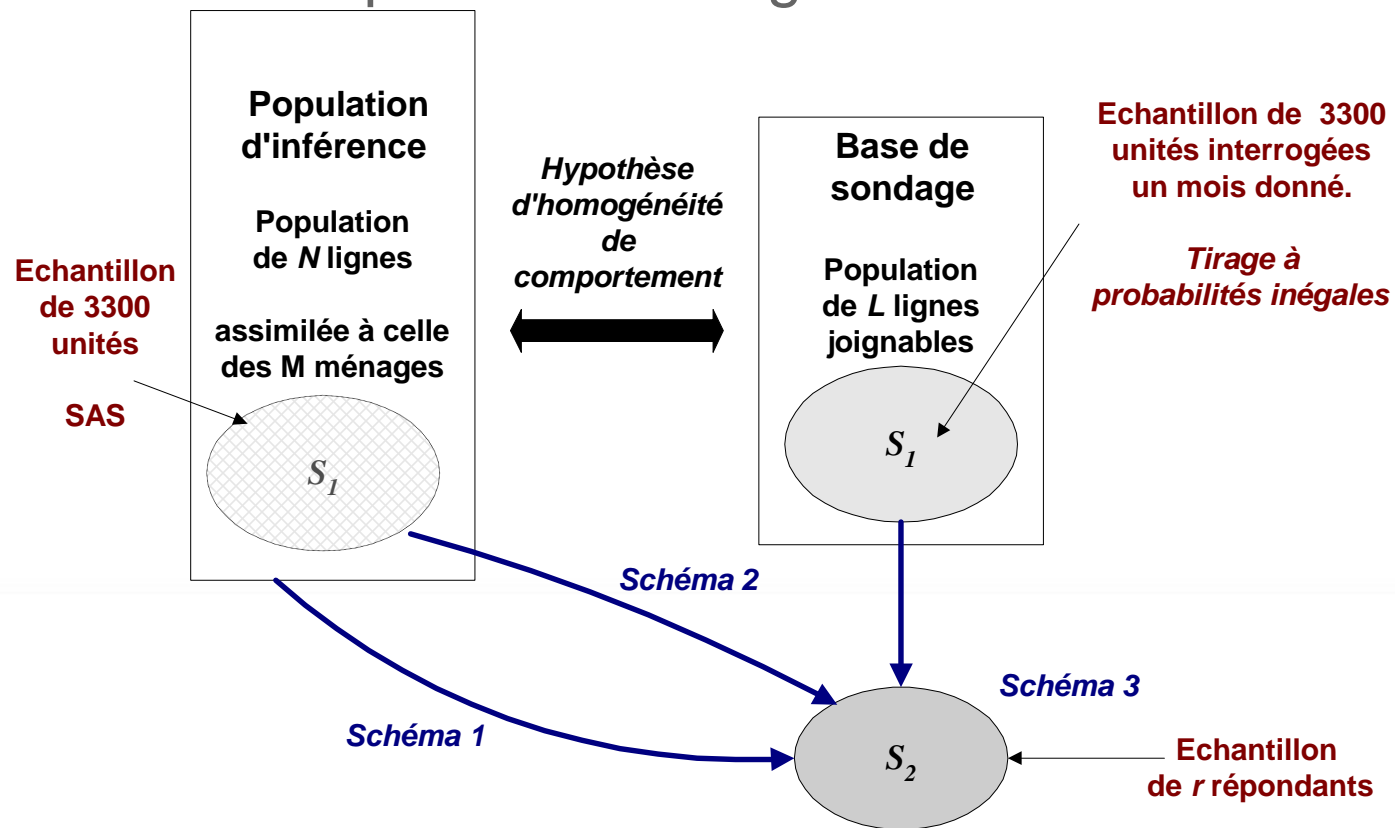
› L'indicateur résumé : le « moral des Français »

$$\hat{Y}_m = \frac{1}{5} \sum_{j=1}^5 \hat{Y}_{m,j}$$



Une démarche progressive

- Précision du niveau d'un solde
- Modélisation du plan de sondage





Précision d'une évolution

- **Sur une longue période**, les échantillons des mois m et m' sont disjoints

$$\hat{V}ar\left(\hat{Y}_{m'} - \hat{Y}_m\right) = \hat{V}ar\left(\hat{Y}_{m'}\right) + \hat{V}ar\left(\hat{Y}_m\right)$$

- **D'un mois sur l'autre où à un mois d'intervalle**, les échantillons des mois m et m' se chevauchent

Corrélation temporelle des opinions

→ Gain de précision

$$Var\left(\hat{Y}_{m'} - \hat{Y}_m\right) = Var\left(\hat{Y}_m\right) + Var\left(\hat{Y}_{m'}\right) - 2Cov\left(\hat{Y}_m, \hat{Y}_{m'}\right)$$



Précision d'une évolution (suite)

- Estimation de la précision avec un SAS

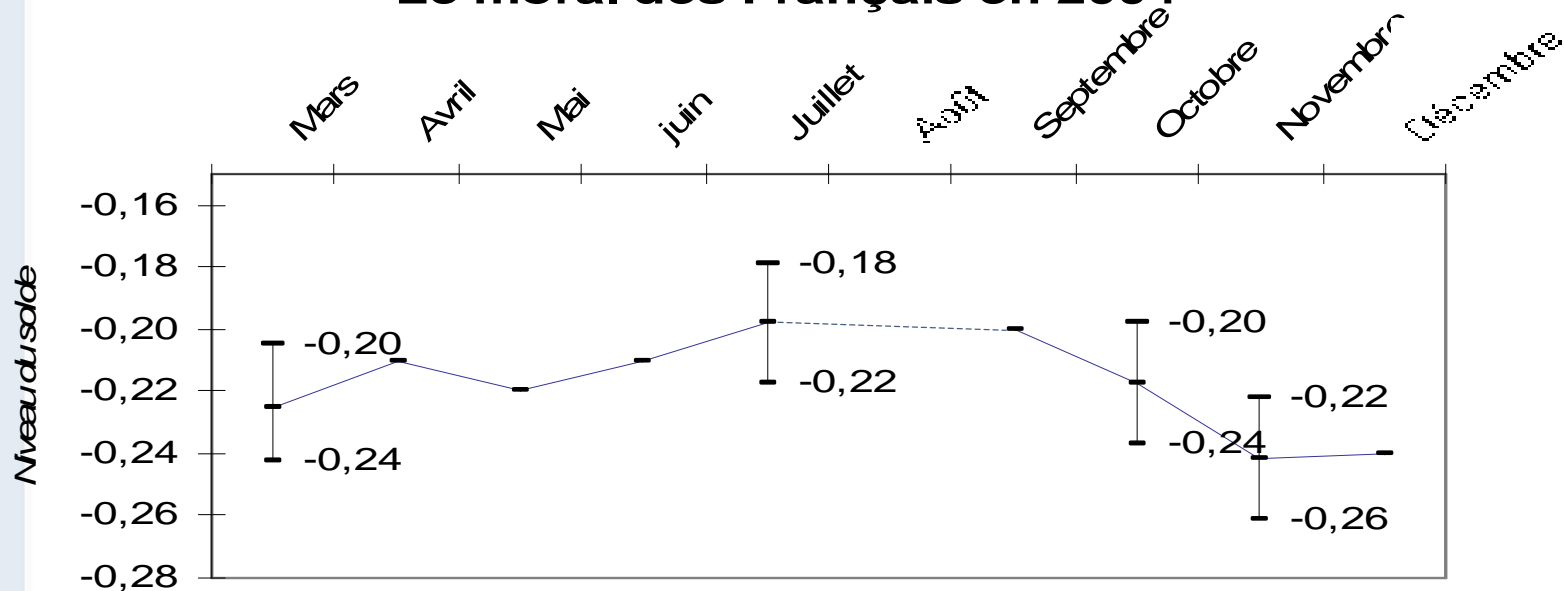
$$\hat{V}ar_{sas}\left(\hat{Y}_{m'} - \hat{Y}_m\right) \cong \hat{s}_y^2 \left(\frac{1}{r_m} + \frac{1}{r_{m'}} - 2\theta \frac{\hat{\rho}}{r_c} \right)$$

- Application de l'effet de sondage obtenu pour le niveau du solde

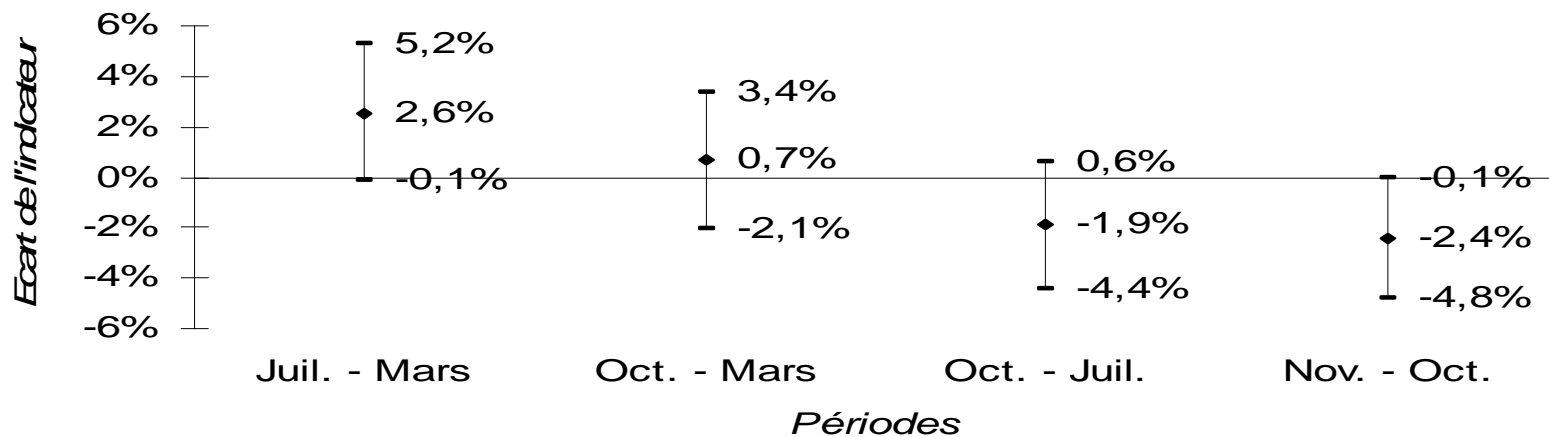
$$\hat{V}ar\left(\hat{Y}_{m'} - \hat{Y}_m\right) = Deff \times \hat{V}ar_{SAS}\left(\hat{Y}_{m'} - \hat{Y}_m\right)$$

Résultats

Le moral des Français en 2004



L'évolution du moral des Français





6. Conclusions et perspectives d'évolution

JMS

14 mars 2005





Une refonte fructueuse :

- › Une enquête simple... mais des travaux et des moyens conséquents.
- › Des investissements méthodologiques utiles...
- › ...permettant une meilleure maîtrise des maillons du processus de production
- › ... et une plus grande rigueur.
- › Une amélioration des résultats de collecte.



Des perspectives d'amélioration à la marge :

- Finaliser les calculs de précision sur des indicateurs utilisant les poids de calage
- Mettre à disposition un module « clefs en main » de calcul de la précision
- Améliorer la prise en compte du taux de pénétration de la liste rouge à un niveau fin.
- Améliorer l'imputation des revenus.
- Automatiser complètement la production des traitements statistiques



Conduire des études sur les résultats :

- Analyser les facteurs explicatifs de l'opinion.
- Etudier les déterminants des changements d'opinion...
- ... et, notamment, l'influence du répondant.
- Définir les indicateurs synthétiques les plus porteurs d'information.
- Vérifier, au moyen d'opérations méthodologiques, l'hypothèse d'indépendance du « comportement téléphonique » et de l'opinion.