

Test de la présence d'une
saisonnalité identifiable
dans le programme X-11-ARMMI

M. Morry et J. Lothian
Division de la désaisonnalisation
et de l'analyse des séries
chronologiques
Statistique Canada

78-10-002F

Octobre 1978.

TEST DE LA PRÉSENCE D'UNE SAISONNALITÉ IDENTIFIABLE
DANS LE PROGRAMME X-11-ARMMI

I. INTRODUCTION

La version de 1967 de la variante X-11 du programme de désaisonnalisation Census Method II contient un test F d'analyse de la variance qui permet de savoir si la série comporte un facteur saisonnier stable (J. Shiskin et al, 1967). Le test porte sur les rapports SI finals non modifiés (tableau D8). La valeur F calculée est comparée à 2.41 (la valeur F au niveau de 1 % pour une série mensuelle de 10 ans). Si elle dépasse 2.41, l'hypothèse de l'égalité des moyennes mensuelles est rejetée et le postulat de l'existence de différences significatives d'un mois à l'autre est accepté, ce que l'on interprète comme un signe de la présence d'un élément saisonnier stable dans la série.

Quoique ce test soit statistiquement valide dans la mesure où ses hypothèses de base ne sont pas violées, il a été constaté, avec les années, qu'en fait, l'ajustement final des séries dont la valeur F se révélait relativement faible mais quand même significative, était très pauvre. Cela nous a poussé à examiner les valeurs F et à mettre au point un test qui puisse nous dire si la série comporte une saisonnalité identifiable et qui prenne en compte tous les facteurs contribuant à fausser les estimations de la composante saisonnière.

II. ANALYSE DES VALEURS F DU TABLEAU D8 DU X-11

Nous avons d'abord choisi un échantillon de séries mensuelles de 10 ans parmi les séries chronologiques économiques désaisonnalisées à Statistique Canada. Les valeur F figurant au tableau D8 ont été notées et analysées.

Ensuite, nous avons désaisonnalisé un autre échantillon de 421 séries mensuelles et trimestrielles dont la longueur allait de cinq à 30 ans et nous avons examiné leurs valeurs F.

Au tableau I, nous présentons certaines statistiques récapitulatives sur les valeurs F tirées des deux échantillons.

TABLEAU I. Statistiques sur la valeur F du test de
saisonnalité stable

Statistiques	Échantillon de 118 séries mensuelles de 10 ans	Échantillon de 421 séries mensuelles et trimestrielles de longueurs variées
% des séries qui échouent au test F	1.7	1.9
% des séries dont la valeur F est inférieure à 10	14.4	7.0
Valeur moyenne des F calculés	101.3	200.8
Écart-type des F calculées	282.4	442.3
Valeur F maximum	2680.6	5565.7
Valeur F médiane	35.2	80.3
Asymétrie	7.3	7.0

Les chiffres du tableau nous montrent qu'en moyenne, les valeurs F de l'échantillon de séries mensuelles de 10 ans sont inférieures à celles de la classe, plus diverse, des séries de longueurs variées. Dans les deux cas, toutefois, on observe un regroupement de points à la gauche de la moyenne et des valeurs plus extrêmes à la droite, ce qui permet de conclure que les formes générales des deux distributions sont semblables.

Les annexes A, B, C et D contiennent des histogrammes des deux distributions de fréquences, selon des intervalles de 100 et de 10. Le mode des deux échantillons semble se situer dans l'intervalle (10-20).

De tous ces renseignements, il découle dès l'abord qu'il faut porter une attention toute spéciale aux séries dont la valeur F est inférieure à 100, puisque cette classe regroupe la majorité des cas.

Autre fait à retenir: quoique moins de 2 % des séries ne répondent pas aux critères du test F de saisonnalité stable, en pratique, le pourcentage des séries désaisonnalisées dont la qualité est jugée inacceptable est beaucoup plus élevé. À notre avis, il était préférable de fixer la borne supérieure de la valeur F à 10, vu le pourcentage de séries que cette modification toucherait.

Dans l'analyse qui suit, nous essaierons de définir la notion de saisonnalité identifiable et de cerner ses liens avec les valeurs F.

III. LA SAISONNALITÉ IDENTIFIABLE DANS LE CADRE DU X-11

- a) Mesures de la distorsion de l'estimation saisonnière lorsque la saisonnalité est stable.

Avant de continuer, il faut corriger une erreur mineure dans le test F du X-11. Nous savons déjà que les valeurs F calculées sont comparées à 2.41, la valeur F au niveau de 1 % pour une série mensuelle de 10 ans. Si

la différence entre la valeur F à 1 % d'une série mensuelle de 10 ans et celle d'une série mensuelle de 30 ans est négligeable, l'écart s'agrandit beaucoup lorsqu'on passe aux séries trimestrielles. Pour remédier à cette incohérence, nous avons inclus dans le nouveau X-11 une fonction génératrice de la distribution F, de façon que les valeurs calculées soient comparées aux valeurs F liées aux bons niveaux de probabilité et au bons degrés de liberté. Par ailleurs, l'emploi des valeurs F du tableau D.8 suscite d'autres complications.

L'étude de 70 séries simulées a montré qu'en raison de l'erreur introduite lors de la détermination de la tendance-cycle et, par conséquent, des rapports S-1, le X-11 a tendance à gonfler les valeurs F. En effet, les valeurs F calculées à partir du tableau des rapports S-I présentent un écart de 10 à 30 % par rapport à celles qui sont fondées sur les estimations du rapport S-I faites par le X-11. Le pourcentage d'accroissement dépend en très grande partie du mouvement saisonnier et de la forme de la tendance-cycle. Pour des série simulées composées d'éléments irréguliers seulement, l'augmentation des valeurs F est de 10 %, tandis que pour celles qui comportent des composantes saisonnières et tendanciennes, la hausse va jusqu'à 30 %.

Pour remédier à cette difficulté, nous avons abaissé le niveau de signification de 1 % à 0.1 %.

Nous sommes maintenant prêts à définir l'expression "saisonnalité identifiable". L'objectif de toute procédure de désaisonnalisation est de produire des estimations faibles de la série désaisonnalisée ou, ce qui revient au même, des coefficients saisonniers. Si l'erreur incorporée au coefficient saisonnier ne masque pas, du fait de son importance, le mouvement saisonnier sous-jacent, on dit que la saisonnalité est identifiable.

Pour mesurer la distorsion introduite dans l'estimation du coefficient saisonnier produite par le X-11, il faut étudier les coefficients de pondération appliqués aux observations pour arriver à ces estimations. Pour obtenir une première approximation, nous avons pris les coefficients de la moyenne mobile à sept termes (3 x 5), parce qu'ils jouent un très grand rôle dans la procédure d'estimation du facteur saisonnier. Si les mouvements irréguliers e_t de la série observée sont des variables aléatoires, indépendantes, de distribution identique et de variance σ^2 , la variance de la variable lissée s'écrit alors:

$$\text{Var} (\sum w_i e_t) = (\sum w_i^2) \text{Var} (e_t) = (\sum w_i^2) \sigma^2 \quad (3.1)$$

ou, si nous passons à l'écart-type:

$$\sigma(\sum w_i^2)^{1/2} \quad (3.2)$$

Les w_i sont les coefficients de lissage. En introduisant les coefficients centraux de la moyenne (3 x 5) dans (3.2), nous obtenons comme première approximation de l'erreur liée aux estimations du coefficient saisonnier, la valeur de 0.4055.

Les coefficients exacts employés dans l'estimation du coefficient saisonnier central sont fournis à l'annexe E. Si l'on a recours à la moyenne à 13 termes de Henderson pour faire l'estimation de la tendance-cycle, l'ensemble complet des coefficients compte alors 169 termes, mais comme les valeurs au-delà des points -77 et +77 sont presque nulles, elles sont tout simplement éliminées. En introduisant les coefficients de l'annexe C dans (3.2), nous obtenons la valeur .383 . Ce nouveau résultat correspond à l'écart-type de l'erreur liée à une estimation du coefficient saisonnier central, sans qu'il soit tenu compte de l'effet produit par l'élimination des valeurs extrêmes. Mais comme les coefficients terminaux de X-11 ne lissent pas autant que les centraux, les estimations des coefficients saisonniers du début et de la

fin de la série comportent des erreurs plus élevées. La valeur .383 constitue donc une espèce de borne inférieure pour l'écart-type de l'erreur liée aux estimations du coefficient saisonnier dans toute la série; si l'on considère plutôt la distorsion, cette valeur est équivalente à .305 (hypothèse de normalité). Au lieu de nous lancer dans toutes sortes de calculs mettant en jeu tous les ensembles possibles de coefficients terminaux, nous avons eu recours à la simulation.

Nous avons désaisonnalisé 50 séries simulées à l'aide du X-11. Dans tous les cas il s'agissait de séries mensuelles additives de 10 ans; les des aléas étaient assez variés. Nous avons comparé les estimations du coefficient saisonnier fournies par le X-11 aux coefficients saisonniers originaux et avons constaté que l'écart absolu moyen s'établissait à environ .37 (l'intervalle de variation allait de .39, pour les séries courtes, à .33 pour les longues).

Dans la suite, le chiffre .37 représentera la distorsion liée aux estimations du coefficient saisonnier lorsque les aléas originaux ont un écart-type égal à . Étudions maintenant la relation entre cette distorsion et la valeur F du test de saisonnalité stable. Comme nous l'avons déjà mentionné, le test en question porte sur les rapports S-1 finals non modifiés du tableau D8.

Le modèle qui sous-tend le test F est le suivant:

$$SI_{ij} = S_j + e_{ij} \quad (j=1, \dots, J \quad i=1, \dots, N) \quad (3.3)$$

S_j représente la composante saisonnière stable du mois J;

e_{ij} représente la composante irrégulière du mois j et de l'année i;

les e_{ij} sont des variables aléatoires, indépendantes, de distribution normale $(0, \sigma^2)$;

N est le nombre d'années;

J est égal à quatre pour les séries trimestrielles et à 12 pour les séries mensuelles.

Posons:

$$\bar{S}^2 = \frac{1}{J} \sum_{j=1}^J S_j^2 \quad (3.4)$$

L'espérance de la valeur F du tableau D8 devrait alors s'écrire

$$F_S \sim \frac{J}{J-1} \frac{\bar{S}^2}{\sigma^2} + 1 \quad (3.5)$$

Comme le test F se fonde sur le rapport (\bar{S}^2 / σ^2) , il nous donne aussi une bonne idée du pourcentage de fiabilité des estimations du coefficient saisonnier. Voici un exemple illustrant cette remarque.

Soit une série mensuelle de 10 ans (c'est-à-dire $N=10, J=12$) dont le mouvement saisonnier stable présente un écart-type de 25 et dont les mouvements irréguliers sont distribués selon une normale $(0, 5^2)$; alors, la valeur F de la série sera à peu près égale à:

$$F_S \sim \frac{J}{J-1} \frac{\bar{S}^2}{\sigma^2} + 1 = \frac{12 \times 10}{11} \frac{25^2}{5^2} + 1 = 273 \quad (3.6)$$

Soit, par ailleurs, une valeur F de 273 (S et σ étant inconnus); il est quand même possible de déduire la distorsion des estimations du coefficient saisonnier. Nous nous servons de la valeur empirique .37 calculée ci-dessus.

Le pourcentage de distorsion du coefficient saisonnier est égal à

$$\bar{D} = \frac{.37\sigma}{\bar{S}} \times 100\% \quad (3.7)$$

À partir de la formule (3.5) nous obtenons

$$\frac{\sigma}{\bar{S}} = \sqrt{\frac{N J}{(J-1) (F_S - 1)}} \quad (3.8)$$

donc

$$\bar{D} = 37 \sqrt{\frac{N J}{(J-1) (F_S - 1)}} \% \quad (3.9)$$

Par conséquent, si la valeur F d'une série mensuelle de 10 ans est de 273, l'espérance de la distorsion sera égale à $37 \times \frac{120}{11 \times 272}$, c'est-à-dire 7.4%. De la formule (3.9), on déduit facilement que le pourcentage de distorsion est inversement proportionnel à la racine carrée de la valeur F obtenue, et donc que plus la valeur F est petite, plus le pourcentage de distorsion des estimations saisonnières sera grand.

Examinons maintenant la distorsion introduite si la valeur F est de 2.41 (il s'agit là de l'ancienne limite utilisée dans le test de saisonnalité stable - c'est le F qu'on obtient pour une série mensuelle de 10 ans).

$$\begin{aligned} D &= 37 \sqrt{\frac{120}{11 \times (2.41 - 1.0)}} \% \\ &= 103\% \end{aligned}$$

De toute évidence, pareille distorsion est inacceptable.

Le tableau suivant présente les espérances des F pour certains niveaux de distorsion définis au préalable. Ces chiffres ont trait à des séries mensuelles de 10 ans.

TABLEAU II - Valeurs F pour un degré de distorsion donné

Pourcentage de distorsion	Valeur F
100%	2.5
75%	3.6
50%	7.0
40%	10.3
30%	17.6
20%	38.3
10%	150.3

Si l'on tient compte à la fois de ces résultats, de ceux du tableau I et du fait que la médiane des valeurs F pour les séries mensuelles de 10 ans est de 35, on découvre que dans 50% des séries mensuelles de 10 ans, la distorsion introduite dans les estimations saisonnières dépasse 20%.

S'il paraît naturel de vouloir désaisonnaliser toute série contenant un coefficient saisonnier significatif (ce que signalerait une valeur F significative), il faut néanmoins déterminer le niveau que doit atteindre cette valeur F pour que nous puissions juger fiables les estimations saisonnières. En d'autres termes, il faut déterminer le degré de distorsion acceptable.

À notre avis, il ne doit pas dépasser 50%; la valeur F correspondante, pour une série mensuelle de 10 ans, est sept. Selon ce critère, si le test de saisonnalité stable produit une valeur F inférieure à sept, la série testée ne contient pas de coefficient saisonnier "identifiable" puisque la distorsion introduite dans l'estimation du coefficient saisonnier dépasse 50%.

Jusqu'à présent, nous nous sommes attachés à étudier les renseignements que l'on peut tirer du test F lorsque l'hypothèse de la stabilité du coefficient saisonnier n'est pas violée. Nous avons constaté que la valeur F donne une bonne idée de l'importance de la distorsion globale introduite dans l'estimation de la composante saisonnière.

Dans la section suivante, nous étudierons les variations du F lorsque la saisonnalité comporte un élément évolutif.

III. b) Mesures de la distorsion des estimations saisonnières lorsque la saisonnalité est évolutive.

Si une série comporte un coefficient saisonnier évolutif, les rapports si sont définis par le modèle suivant:

$$SI_{ij} = S_j + S_{ij}^M + e_{ij} \quad (3.10)$$

Les symboles S_j , e_{ij} sont repris de la section ci-dessus S_{ij}^M correspond à l'ampleur du mouvement au cours de l'année i et du mois j , et il faut que:

$$\sum_i S_{ij}^M = \sum_j S_{ij}^M = 0 \quad (3.11)$$

Pour mieux cerner l'effet de l'évolution de la saisonnalité sur la valeur F , nous supposons que le mouvement est simplement linéaire:

$$SI_{ij} = S_j + b_j i + e_{ij} \quad (3.12)$$

S_j , b_j et e_{ij} sont indépendants les uns des autres, mais doivent satisfaire aux contraintes suivantes:

$$\sum_{j=1}^J S_j = 0; \quad J = 4 \text{ or } 12 \quad (3.13)$$

$$\sum_{j=1}^J b_j = 0 \quad (3.14)$$

Posons

$$\bar{b}^2 = \frac{1}{J} \sum_{j=1}^J b_j^2 \quad (3.15)$$

En calculant les espérances de la variation entre mois et de la variation résiduelle, nous pouvons exprimer F_S ainsi:

$$F_S \sim \frac{\frac{JN\bar{S}^2}{J-1} + \sigma^2}{\frac{N(N+1)\bar{b}^2}{J} + \sigma^2} \quad (3.16)$$

Cette formule nous montre bien que si la saisonnalité évolue, la valeur F diminue. Par conséquent, si l'on prend deux séries dont les aléas présentent la même variance, et les coefficients saisonniers, la même amplitude moyenne on obtiendra des valeurs F très différentes si dans l'une, la composante saisonnière est stable et dans l'autre, elle comporte un élément évolutif significatif.

À partir de séries simulées, nous avons obtenu des résultats qui concordent avec la formule (3.16). Le tableau suivant présente le genre de réduction que subit la valeur F moyenne lorsque la série testée comporte une composante saisonnière évolutive, que l'amplitude moyenne du coefficient saisonnier est égale à $\bar{S} = 18$ et $\sigma = 5$.

TABLEAU III - Variation de F_S lorsque la pente de la composante saisonnière moyenne s'accroît

\bar{b}	F_S
0.0	148.7
0.7	129.1
1.4	86.9
2.1	52.1
2.8	39.5

Il s'agissait ensuite d'étudier le genre de distorsion qu'introduit une saisonnalité évolutive.

Appliquons le filtre saisonnier 3×5 au rapport SI de (3.12). Les coefficients centraux de ce filtre reproduisent le mouvement linéaire $S_j + b_j i$ sans problème, donc l'erreur e_{ij} est la seule source de distorsion possible. Comme nous l'avons déjà signalé, lorsque les aléas originaux sont des variables normale, d'écart-type σ , l'emploi de la moyenne mobile centrale 3×5 réduit l'erreur à $.405\sigma$, c'est-à-dire, si l'on considère la distorsion, à $\frac{\sqrt{2}}{\pi} \times 0.405\sigma$, soit $.32\sigma$.

Malheureusement, les coefficients terminaux de la moyenne 3×5 ne reproduisent pas les mouvements linéaires, ce qui fausse davantage les estimations des coefficients saisonniers.

L'espérance de la distorsion est:

$$\text{Année T-2: } D = .16|\bar{b}| + .43 \sqrt{\frac{2}{\pi}} \sigma \quad (3.17)$$

$$\text{T-1: } D = .57|\bar{b}| + .47 \sqrt{\frac{2}{\pi}} \sigma \quad (3.18)$$

$$\text{T : } D = 1.30|\bar{b}| + .51 \sqrt{\frac{2}{\pi}} \sigma \quad (3.19)$$

$$\text{où } |\bar{b}| = \sum_{j=1}^J |b_j|.$$

C'est-à-dire que l'on trouve la plus forte distorsion dans la dernière année de la série. Ce raisonnement nous montre que dans une série de saisonnalité évolutive linéaire, les estimations courantes des coefficients saisonniers sont nettement plus déformées que celle des coefficients centraux.

Pour rattacher cette distorsion à la valeur F, nous avons ajusté un ensemble de séries simulées de saisonnalité évolutive linéaire au moyen du programme X-11 et totalisé les résultats:

TABLEAU IV - Variation de F_s et de la distorsion lorsque la pente de la composante saisonnière moyenne s'accroît

Pente moyenne absolue	Valeur F avant le X-11	Valeur F après le X-11	Distorsion globale	Distorsion sur la dernière année
0.00	46.8	56.1	18.0%	18.0%
0.62	45.8	55.5	18.2%	18.7%
1.24	39.2	48.6	19.0%	22.4%
1.86	30.7	39.0	20.8%	30.0%
2.48	24.8	32.1	22.0%	34.0%
3.10	20.1	25.5	25.3%	42.0%

Quoique ces valeurs F empiriques concordent avec la formule (3.16), celle-ci ne s'applique bien qu'aux séries dont le mouvement saisonnier est linéaire. Si la saisonnalité évolutive est d'un autre genre, le dénominateur de (3.16) comportera un terme de plus, ce qui réduira le F , certes, mais pas autant que ne le laisse croire cette formule. La baisse de la valeur F traduit en partie l'accroissement de la distorsion que cause le caractère évolutif de la saisonnalité. La distorsion introduite dans les années terminales de la série peut être supérieure ou inférieure aux chiffres fournis par les formules (3.17), (3.18) et (3.19). Tous dépend de la forme du mouvement saisonnier.

Le seul examen de la valeur F du test de saisonnalité stable ne suffit donc pas à l'utilisateur pour déterminer si le coefficient saisonnier est stable ou évolutif. Si l'on introduit cette valeur F dans la formule (3.9), on obtient une idée approximative de la distorsion globale. Des simulations nous ont montré que pour les séries mensuelles de 8 à 14 ans comportant un élément saisonnier évolutif linéaire, la formule (3.9) donne une bonne estimation de la distorsion globale. Malheureusement, si la série est considérablement plus longue ou plus courte que 10 ans, la formule (3.9) surestimerait (dans le premier cas) ou sous-estimerait (dans le deuxième) la distorsion globale.

Même dans le cas des séries de 8 à 14 ans, la distorsion obtenue par application à la formule (3.9) de la valeur F calculée n'est qu'une indication de sa valeur moyenne. Celle qui déforme les données de la dernière année (celle de l'estimation courante) est gravement sous-estimée dès que la saisonnalité évolue linéairement.

La valeur F fournie par le test de saisonnalité stable ne peut décrire que la distorsion globale; elle ne dit rien de sa distribution dans la série. Il faut savoir si la saisonnalité comporte un élément évolutif pour pouvoir déterminer si les estimations extrêmes présentent une erreur nettement plus élevée que les centrales.

C'est pourquoi nous avons incorporé au programme un test F de la saisonnalité évolutive.

IV. TEST F DE SAISONNALITÉ ÉVOLUTIVE

Pour mesurer l'importance du mouvement saisonnier, un test F de saisonnalité évolutive a été mis au point (Higginson, 1975) et incorporé au programme X-11. Il est fondé sur une analyse à deux facteurs de la variance et porte sur les rapports (différences) SI du tableau D8. Il vise à déterminer si la série comporte des coefficients saisonniers évolutifs caractérisés par une variation graduelle de l'amplitude saisonnière. Les valeurs du tableau D8 sont d'abord remplacées par leur écart-type par rapport à la moyenne du tableau. L'analyse à deux facteurs de la variance est fondée sur la ventilation de la variance total en une somme de trois facteurs, les variances "entre mois" et "entre années", et la variance résiduelle. L'hypothèse nulle est que la variation "entre années" est aléatoire et l'autre hypothèse, que la différence entre les moyennes annuelles a une signification et correspondrait en fait à un mouvement significatif de la saisonnalité d'une année à l'autre.

Au départ, le test était fait au niveau de 1 %, mais afin de réduire le risque de commettre une erreur de deuxième espèce (c'est-à-dire d'accepter que la série ne comporte aucune saisonnalité évolutive alors que c'est faux), nous avons relevé le seuil à 5 %. Pour plus de clarté, nous désignerons par F_M le test F de saisonnalité évolutive et par F_S , celui de saisonnalité stable.

Des simulations ont montré que la valeur F_M est liée à $|\bar{b}|$ et à σ , $|\bar{b}|$ correspondant à la pente absolue moyenne du mouvement linéaire de la composante saisonnière d'une année à l'autre.

Plus le mouvement des coefficients saisonniers est fort, plus la valeur F_M est élevée, tandis que σ varie inversement par rapport à F_M .

Le tableau suivant présente les valeurs F_M tirées de cinq ensembles de 20 séries simulées de $S=18$, de différents et de pentes absolues σ moyennes $|\bar{b}|$:

TABLEAU Va: Valeur F_M lorsque varie et que la pente moyenne augmente σ

$ \bar{b} $	Valeur F_M pour $\sigma=5$	Valeur F_M pour $\sigma=9$
0.00	1.10	1.00
0.62	3.21	1.78
1.26	7.90	3.49
1.26	13.30	5.91
2.48	18.10	8.59

Si l'on exprime Va en fonction de $\frac{|\bar{b}|}{\sigma}$, on obtient le tableau suivant:

TABLEAU Vb: F_M lorsque $\frac{|\bar{b}|}{\sigma}$ augmente

$\frac{ \bar{b} }{\sigma}$	F_M
.069	1.78
.124	3.21
.137	3.49
.206	5.91
.248	7.90
.275	8.59
.372	13.30
.496	18.10

Si l'on considère les couples $|\bar{b}|/\sigma$ et F_M on constate qu'il semble exister une relation linéaire entre $(|\bar{b}|/\sigma)^2$ et F_M .

Calculons maintenant la valeur F_M prévue pour le modèle décrit en (3.12):

$$SI_{ij} = S_j + b_j i + e_{ij}$$

Outre les hypothèses de la section IIb, nous supposons que la condition suivante est respectée:

Pour un j donnée, tous les SI_{ij} doivent se trouver du même côté de la moyenne du tableau.

Si nous prenons l'espérance de la variance entre années et de la variance résiduelle, nous obtenons:

$$F_M = \frac{\frac{|\bar{b}|^2 N J (N+1)}{12} + \sigma^2}{\frac{(N+1)N}{J+1} \frac{\sum (|b_j| - |\bar{b}|)^2}{12(J-1)} + \sigma^2} \quad (4.1)$$

Malheureusement, le dénominateur comprend le terme $\sum (|b_j| - |\bar{b}|)^2$, la variation de la pente absolue, qui peut changer considérablement d'une série à l'autre.

Il est possible d'exprimer cette variation sous la forme du produit d'une constante k par \bar{b}^2 , (la valeur de k dépend du genre de la pente). Nous réduisons donc F_M à:

$$F_M = \frac{K_1 \frac{|\bar{b}|^2}{\sigma^2} + 1}{K_2 \frac{|\bar{b}|^2}{\sigma^2} + 1} \quad (4.2)$$

Au moyen d'une approximation de premier ordre, nous pouvons écrire F_M sous une forme structurelle simple:

$$F_M \sim A + B \frac{|\bar{b}|^2}{\sigma^2} \quad (4.3)$$

ou, si nous l'exprimons en fonction de \bar{b}^2 .

$$F_M \sim A + C \frac{\bar{b}^2}{\sigma^2} \quad (4.4)$$

Ce dernier résultat est confirmé par le fait qu'une ligne ajuste bien le nuage de points $|\bar{b}|^2/\sigma^2$ et F_M tiré du tableau Vb.

Comme la distorsion s'accroît à mesure que $|\bar{b}|$ et σ augmentent, la forme structurelle de F_M nous montre bien qu'elle ne peut être entièrement décrite par F_M seulement.

Toutefois, en utilisant F_S et F_M ensemble, nous devrions obtenir assez d'information pour nous permettre d'estimer la distorsion à laquelle il faut nous attendre.

Dans la section suivante, nous étudierons le lien entre ces trois mesures et présenteront un test de saisonnalité identifiable fondé sur nos constatations.

V. TEST DE SAISONNALITÉ IDENTIFIABLE

Comme nous l'avons déjà mentionné, la saisonnalité d'une série chronologique est identifiable si le pourcentage de distorsion des estimations du coefficient saisonnier n'est pas trop élevé.

Comme les utilisateurs s'intéressent surtout aux estimations courantes, le test est fondé sur la distorsion des chiffres de la dernière année.

Selon la formule (3.19), si une série comporte un élément saisonnier évoluant de façon linéaire, l'espérance de la distorsion pour la dernière année sera:

$$D \sim 1.3 |\bar{b}| + .51 \frac{\sqrt{2}}{\pi} \sigma \quad (5.1)$$

ou, si nous l'exprimons en fonction de la variance,

$$V \sim (1.695 \frac{|\bar{b}|^2}{\sigma^2} + .263) \sigma^2 \quad (5.2)$$

Reprenons maintenant les formules établies pour F_S et F_M :

$$F_S = \frac{\frac{J}{J-1} \frac{\bar{S}^2}{\sigma^2} + 1}{\frac{N(N+1)}{J} \frac{\bar{b}^2}{\sigma^2} + 1} \quad (5.3)$$

$$F_M = A + C \frac{\bar{b}^2}{\sigma^2} \quad (5.4)$$

Comme nous nous intéressons au pourcentage de variance de l'erreur, nous divisons les deux membres de la formule de la variance par \bar{S}^2 pour obtenir:

$$\frac{V}{\bar{S}^2} = \frac{\sigma^2}{\bar{S}^2} (1.695 \frac{\bar{b}^2}{\sigma^2} + .263) \quad (5.5)$$

En exprimant $\frac{\bar{b}^2}{\sigma^2}$ et $\frac{\bar{S}^2}{\sigma^2}$ en fonction de F_S et de F_M , et en introduisant le résultat dans l'équation (5.5), nous obtenons l'équation suivante:

$$\frac{V}{\bar{S}^2} = \frac{K_1}{F_S} \times \frac{F_M + K_2}{F_M + K_3} \quad (5.6)$$

où K_1, K_2, K_3 sont des constantes.

Au moyen d'une approximation du premier ordre, il est possible d'écrire la formule (5.6) sous la forme structurelle suivante:

$$\frac{V}{\bar{S}^2} = \frac{C_1}{F_S} + \frac{C_2 F_M}{F_S} \quad (5.7)$$

ou, en termes de l'espérance en pourcentage de distorsion:

$$\bar{D} = \sqrt{\frac{C_1'}{F_S} + \frac{C_2' F_M}{F_S}} \times 100 \% \quad (5.8)$$

À notre avis, une distorsion de plus de 50 % est inacceptable, et une série dont le coefficient saisonnier présente ce genre de déformation devrait échouer au test de saisonnalité identifiable.

Empiriquement, nous avons découvert qu'en donnant à C_1 , la valeur de 3.5 et à C_2' , celle de 1.5, nous obtenions une bonne valeur-limite test. La statistique finale du test des saisonnalité identifiable prend donc la forme:

$$T = \sqrt{\frac{1}{2} \left(\frac{7}{F_S} + \frac{3F_M}{F_S} \right)} \quad (5.9)$$

Si la statistique T calculée à partir de F_S et de F_M dépasse 1, la saisonnalité de la série n'est pas identifiable.

Comme nous l'avons montré dans les sections précédentes, la présence d'une composante saisonnière évolutive dans la série modifie le résultat. Par conséquent, le test combiné de saisonnalité identifiable comporte deux branches correspondant aux deux cas possibles. Nous avons encore amélioré le test en y incorporant l'équivalent non paramétrique du test F, pour pouvoir résoudre le problème des séries dont les aléas n'ont pas une distribution normale. La version finale du test combiné prend donc la forme suivante:

(1) Si la valeur F_S du test de saisonnalité stable est inférieure à la valeur F au niveau de 0.1 %, l'hypothèse nulle, c'est-à-dire que la saisonnalité n'est pas identifiable, est acceptée; sinon, la série réussit le test de saisonnalité stable.

(2) Si la série réussit le test (1), mais que le test F_M de saisonnalité évolutive révèle qu'elle comporte une composante saisonnière évolutive significative au niveau de 5 %, alors la valeur obtenue est combinée à la valeur F_S tirée de (1) pour obtenir les statistiques $T_1 = \frac{7}{F_S}$ et $T_2 = \frac{3F_M}{F_S}$, puis on

fait la moyenne des deux T . Si le résultat est plus grand ou égal à 1, l'hypothèse nulle, à savoir que la saisonnalité n'est pas identifiable, est acceptée.

(3) Si le test F_M montre que la série ne comporte aucune composante saisonnière évolutive ou si l'hypothèse nulle du test fondé sur les statistiques T est rejetée, les valeurs T_1 et T_2 sont examinées; si l'une ou l'autre dépasse 1, le programme imprime un message signalant que la composante saisonnière de la série n'est probablement pas identifiable.

(4) Si la série réussit le test (3), mais que la valeur du khi-carré du test de Kruskal-Wallis est inférieure à la valeur du khi-carré à 0.1 %, le programme imprime un message indiquant que la saisonnalité de la série n'est probablement pas identifiable.

(5) Si la série comporte une composante saisonnière stable, qu'elle réussit (3) et que le test de Kruskal-Wallis signale aussi qu'elle comporte un coefficient saisonnier stable, le programme imprime un message annonçant qu'elle comporte une composante saisonnière identifiable.

Pour expliquer le test de façon plus visuelle, nous avons présenté à l'annexe F une table de décision.

Afin de mesurer l'effet de ce test, nous avons examiné 421 séries. Nous avons constaté que 21, soit 5 % d'entre elles, ne satisfaisaient pas aux critères du test de saisonnalité identifiable. En outre, dans 1.6 % des cas, le résultat a révélé que la série ne comportait probablement aucun coefficient saisonnier identifiable. Il s'agit là d'une hausse significative par rapport à la proportion des séries ne satisfaisant pas aux critères du test de saisonnalité stable (1.9 %). Toutefois, si l'on considère la piètre qualité de la désaisonnalisation des séries rejetées par notre test, cet accroissement et tout à fait justifiable.

Une autre étude a été menée sur un échantillon de 118 séries mensuelles de 10 ans. Cette fois, huit séries (6.8 %), ont été rejetées immédiatement et 5.9 % ont été classées dans la zone grise. De nouveau, ces chiffres ont été jugés raisonnables, compte tenu de la qualité globale de la désaisonnalisation des séries mensuelles de 10 ans.

Les annexes G et H présentent les nuages de points formés par les valeurs F_M et F_S ; les régions de rejet sont indiquées. Nous n'avons considéré que les valeurs F_S entre 0 et 100 puisque toutes les séries présentant des valeurs F_S supérieures à 100 ont réussi le test de saisonnalité identifiable.

Les valeurs T tirées de la formule (5.9) pour les deux échantillons ont été notées, et un graphe de leur distribution est présenté aux annexes I et J.

VI. CONCLUSION

Le test de saisonnalité identifiable tente d'utiliser des renseignements sur la valeur des coefficients saisonniers stables et évolutifs dans la série et de les exprimer en fonction de l'espérance de la distorsion des estimations courantes de ce coefficient.

Les formules établies pour F_S et F_M montrent qu'exception faite du mouvement saisonnier moyen et de la taille des aléas, la distorsion dépend de la forme de la saisonnalité, du nombre d'années et de l'intervalle d'échantillonnage de la série.

La statistique T présentée à la section V n'est qu'une formule empirique approximative: elle donne d'assez bons résultats pour les séries de 8 à 14 ans, mais est nettement moins utile pour les séries très courtes ou très longues.

Nous voulons essayer d'améliorer ce test en tenant compte de la longueur et d'autres facteurs influant sur le calcul de la distorsion.

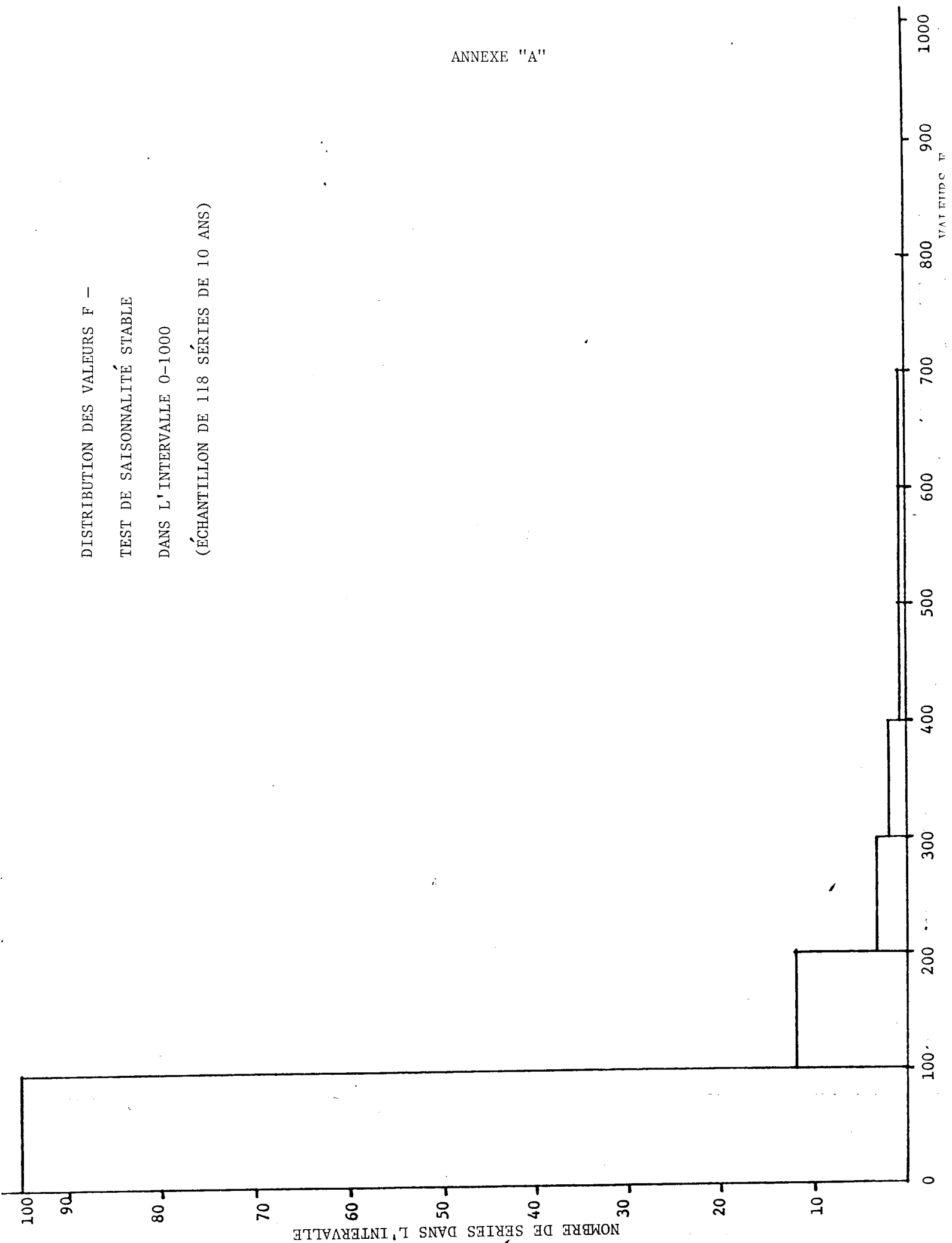
D'ici là, pour obtenir des résultats valides, il faudrait appliquer le test aux dix dernières années lorsque la série est assez longue.

BIBLIOGRAPHIE

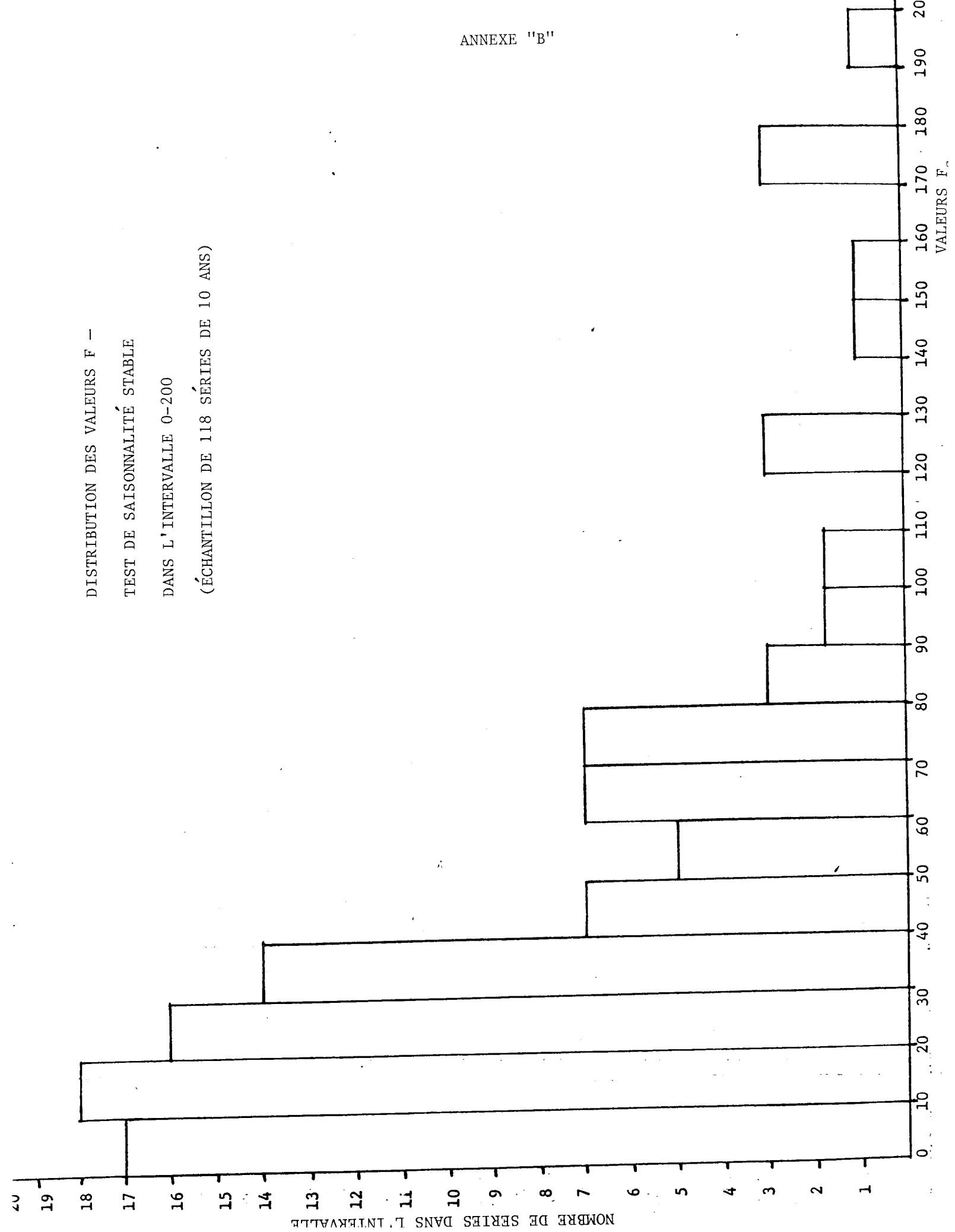
Higginson, J. (1975): "Test de F de la présence d'une saisonnalité mobile lors de l'utilisation de la variante X-11 de la Census Method II", document de recherche, Division de la désaisonnalisation et de l'analyse des séries chronologiques, Statistique Canada.

Shiskin, J., Young, A.H. and Musgrave, T.C. (1967): "The X-11 Variant of Census Method Seasonal Adjustment" Technical Paper No. 15, Bureau of the Census, U.S. Dept. of Commerce.

DISTRIBUTION DES VALEURS F -
TEST DE SAISONNALITÉ STABLE
DANS L'INTERVALLE 0-1000
(ÉCHANTILLON DE 118 SÉRIES DE 10 ANS)

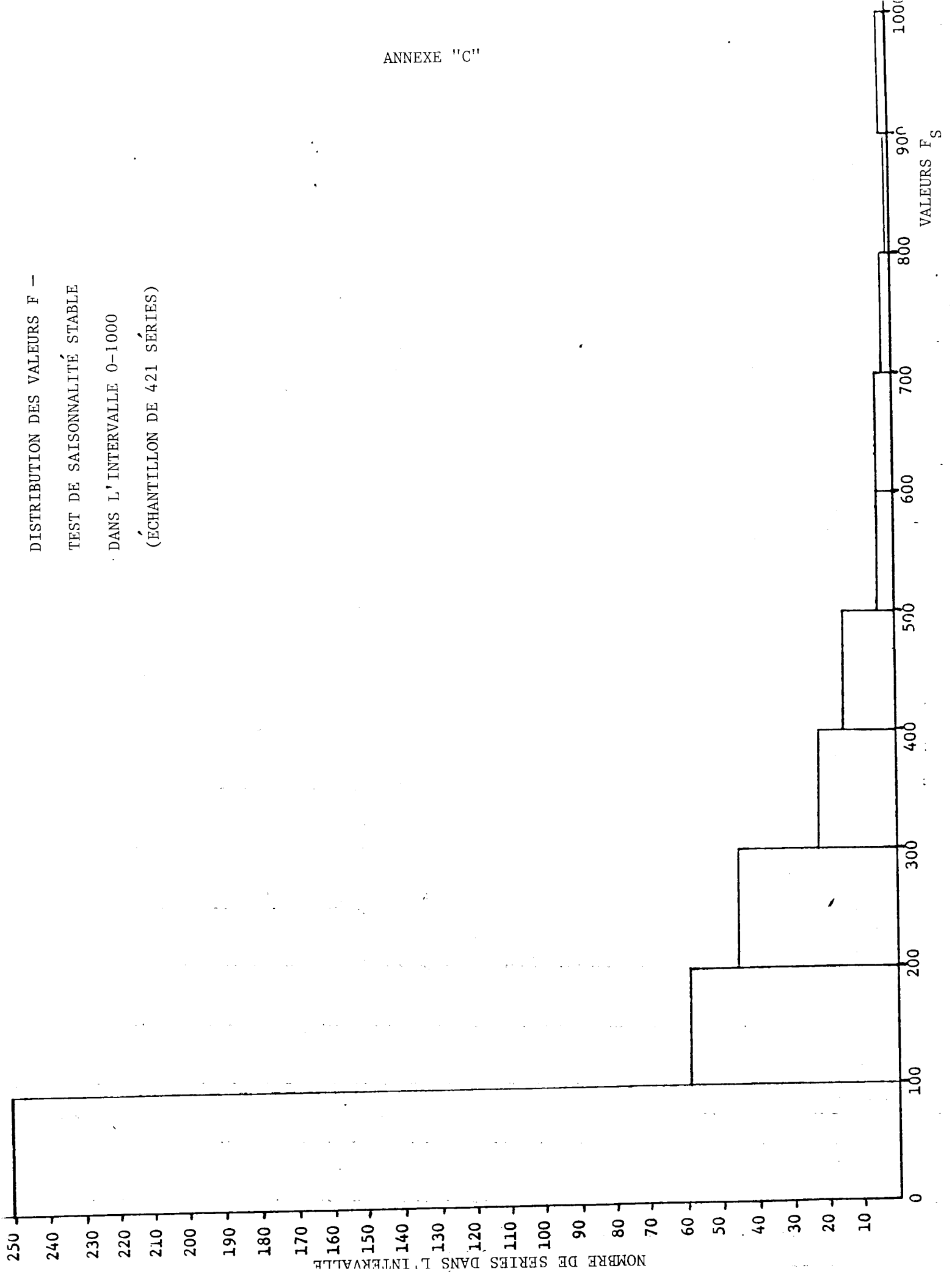


DISTRIBUTION DES VALEURS F -
 TEST DE SAISONNALITÉ STABLE
 DANS L'INTERVALLE 0-200
 (ÉCHANTILLON DE 118 SÉRIES DE 10 ANS)

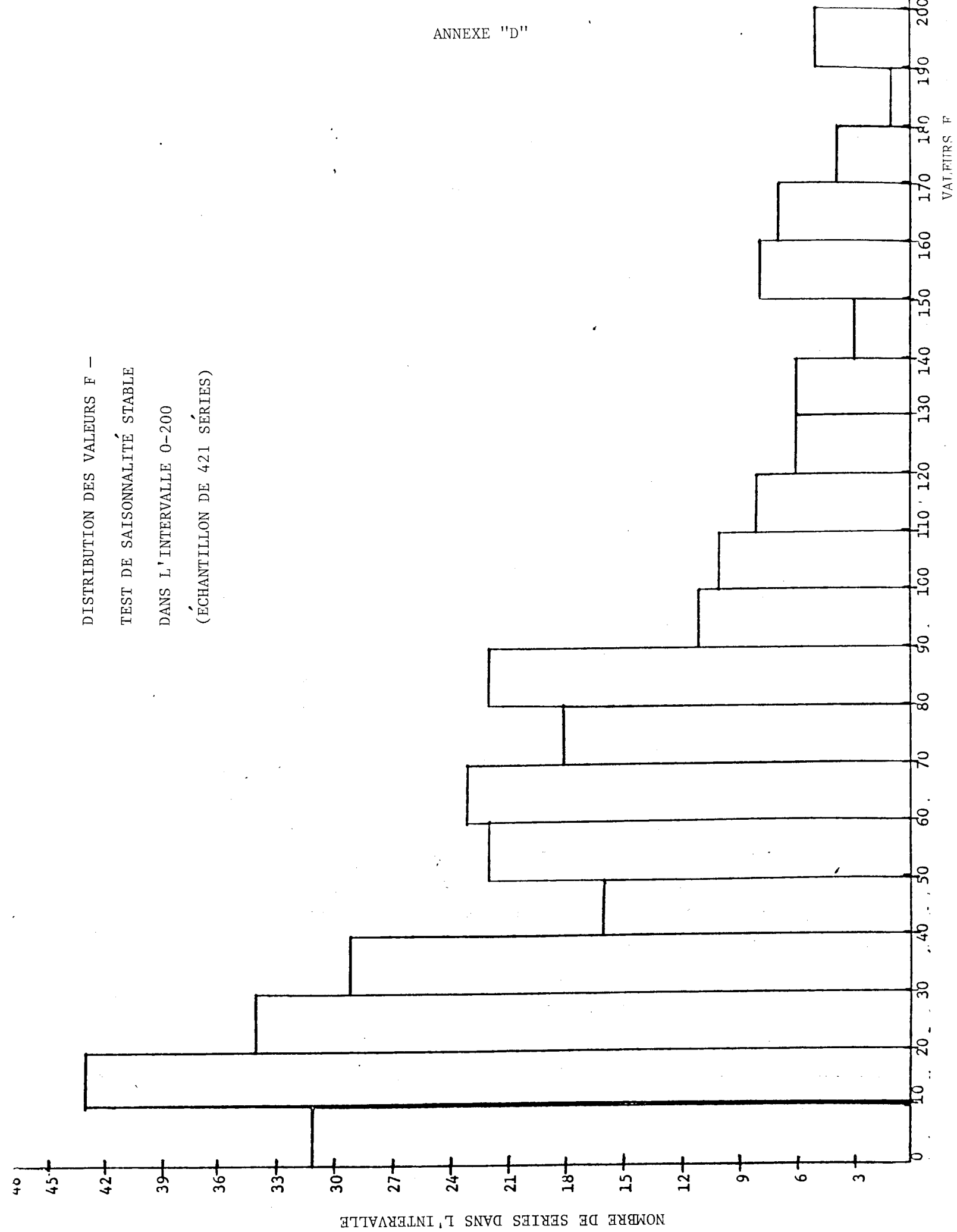


ANNEXE "C"

DISTRIBUTION DES VALEURS F -
 TEST DE SAISONNALITÉ STABLE
 DANS L'INTERVALLE 0-1000
 (ÉCHANTILLON DE 421 SÉRIES)



DISTRIBUTION DES VALEURS F —
 TEST DE SAISONNALITÉ STABLE
 DANS L'INTERVALLE 0-200
 (ÉCHANTILLON DE 421 SÉRIES)



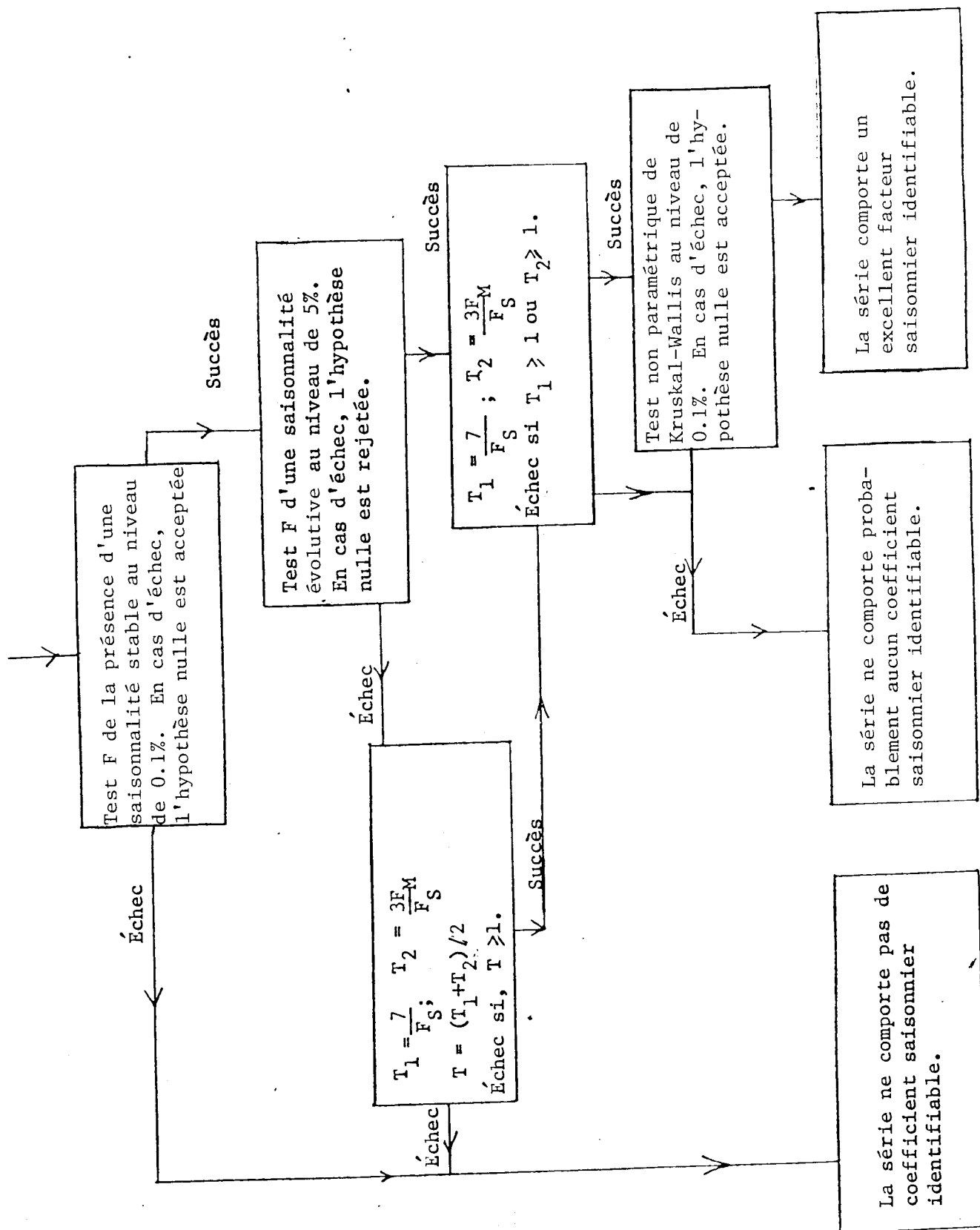
ANNEXE "E"

Coefficients de pondération de l'estimation centrale
du coefficient saisonnier

t	0.18094	t+26	-.01333	t+52	-.00326
t+1	-.01880	t+27	-.01321	t+53	-.00334
t+2	-.01802	t+28	-.01267	t+54	-.00226
t+3	-.01694	t+29	-.01186	t+55	-.00080
t+4	-.01583	t+30	-.00833	t+56	0.00007
t+5	-.01501	t+31	-.00481	t+57	0.00080
t+6	-.01395	t+32	-.00399	t+58	0.00126
t+7	-.01332	t+33	-.00346	t+59	0.00139
t+8	-.01364	t+34	-.00334	t+60	0.00120
t+9	-.01532	t+35	-.00365	t+61	0.00074
t+10	-.01780	t+36	0.06229	t+62	0.00010
t+11	-.02008	t+37	-.00543	t+63	-.00054
t+12	0.17867	t+38	-.00662	t+64	-.00090
t+13	-.02121	t+39	-.00766	t+65	-.00085
t+14	-.01988	t+40	-.00815	t+66	-.00045
t+15	-.01793	t+41	-.00795	t+67	0.0
t+16	-.01608	t+42	-.00504	t+68	0.00022
t+17	-.01491	t+43	-.00175	t+69	0.00030
t+18	-.01163	t+44	-.00057	t+70	0.00025
t+19	-.00871	t+45	0.00130	t+71	0.00013
t+20	-.00852	t+46	0.00326	t+72	0.00002
t+21	-.00901	t+47	0.00457	t+73	-.00003
t+22	-.01004	t+48	0.00465	t+74	-.00004
t+23	-.01124	t+49	0.00328	t+75	-.00004
t+24	0.12104	t+50	0.00088	t+76	-.00002
t+25	-.01302	t+51	-.00164	t+77	-.00001

Comme les coefficients sont symétriques, $w_{t+i} = w_{t-i}$ et, par conséquent, nous n'avons présenté que la moitié d'entre eux.

TEST À PLUSIEURS COMPOSANTES DE LA PRÉSENCE D'UNE
SAISONNALITÉ IDENTIFIABLE



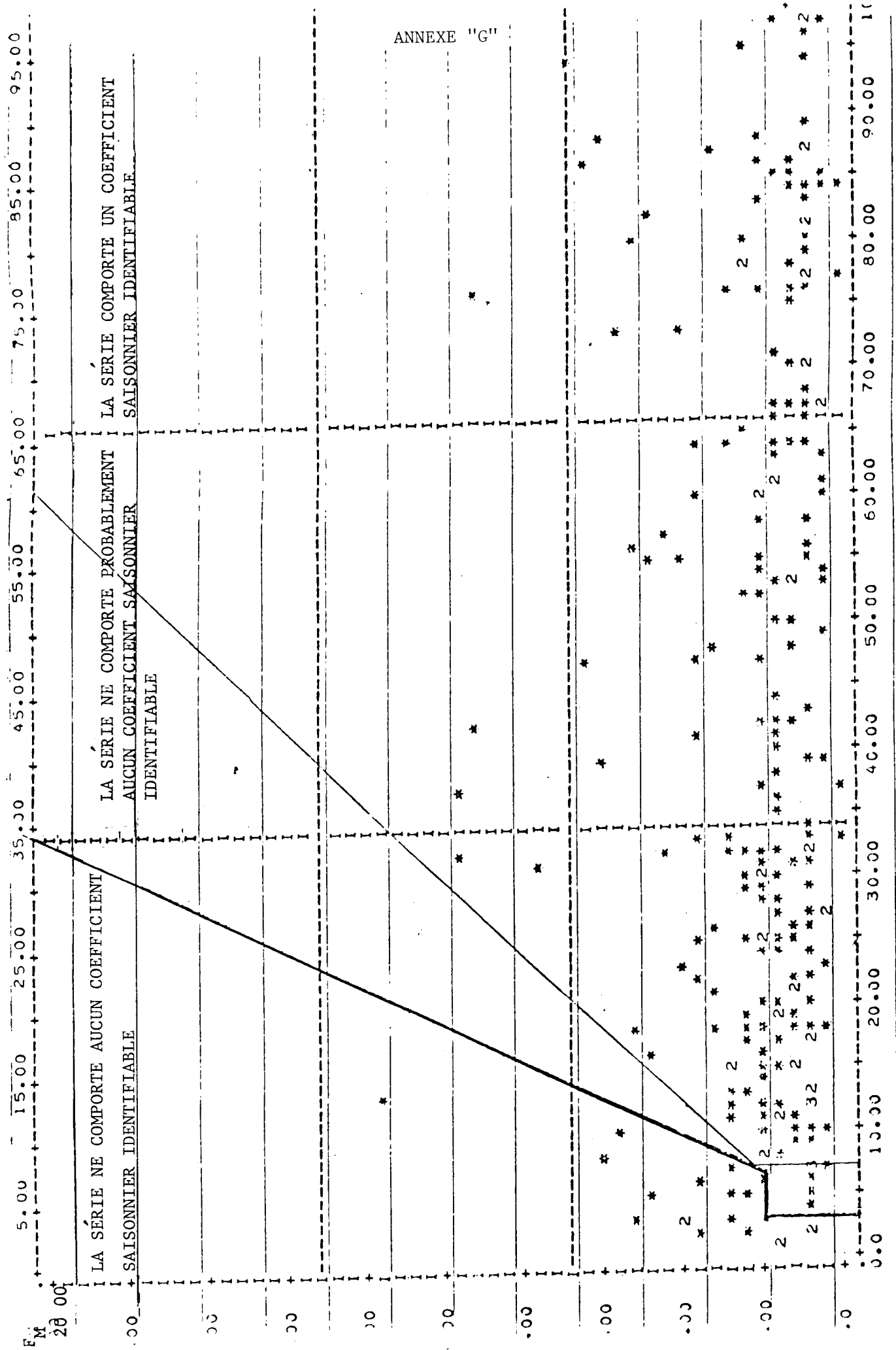


DIAGRAMME DES COUPLES (F_M , F_S) 0 - 100

(Echantillon de 421 séries de longueurs variées)

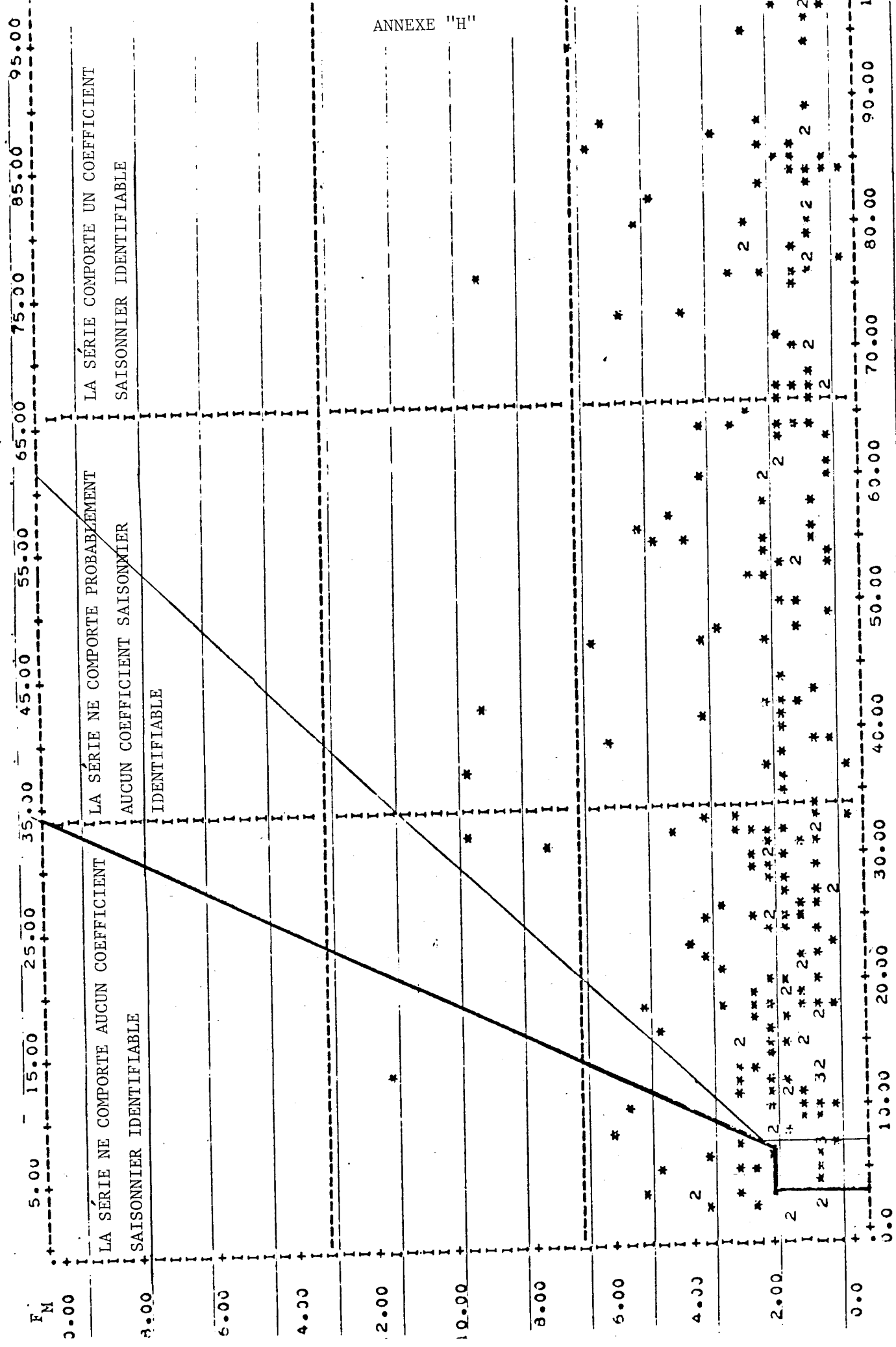
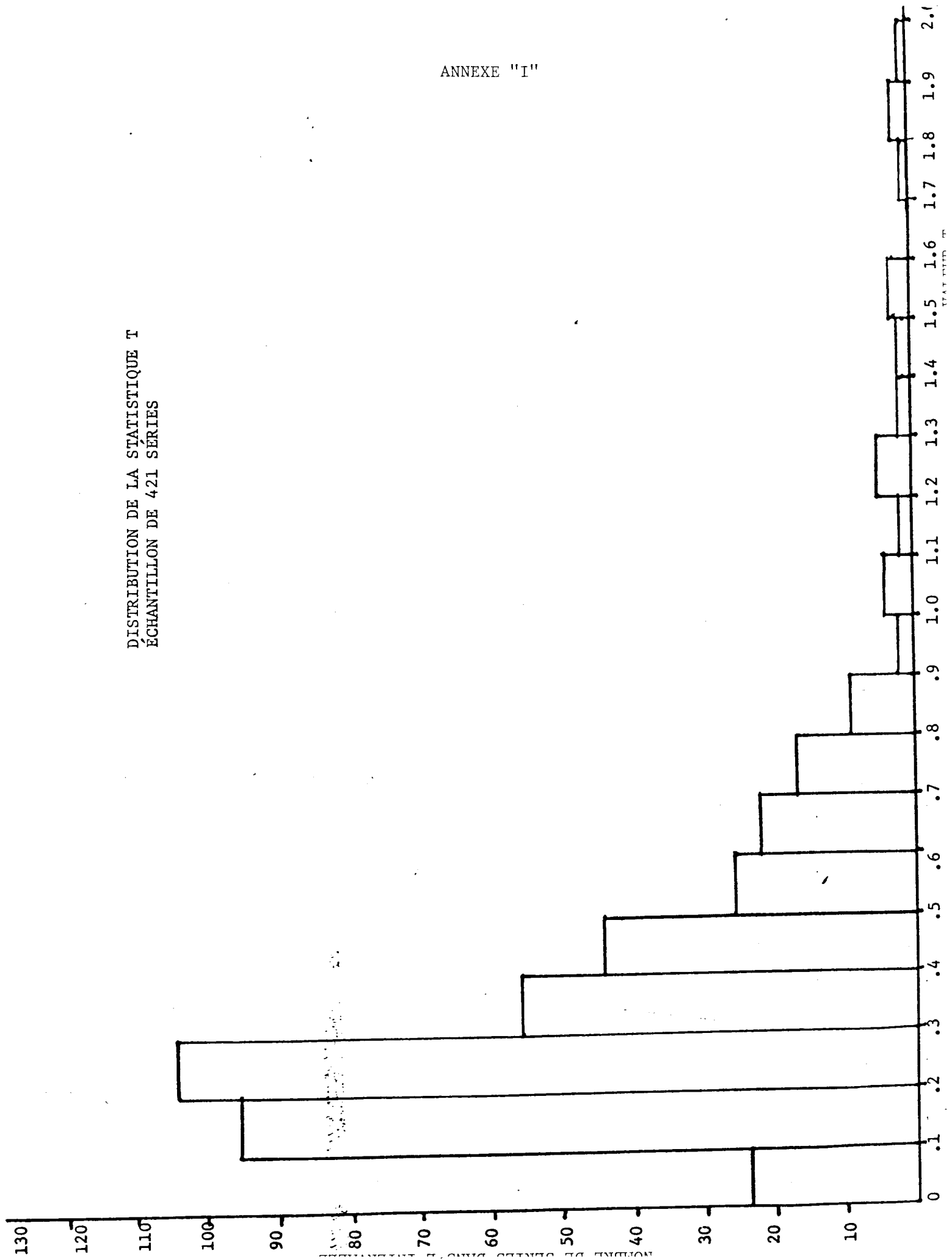


DIAGRAMME DES COUPLES (F_M , F_S) 0 - 100
(Échantillon de 118 séries de 10 ans)

ANNEXE "I"

DISTRIBUTION DE LA STATISTIQUE T
ÉCHANTILLON DE 421 SÉRIES



DISTRIBUTION DE LA STATISTIQUE T
ÉCHANTILLON DE 118 SÉRIES DE 10 ANS

