

LECTURE D'UNE SORTIE D'ORDINATEUR : LE MODÈLE

Notation :

PTOT : population totale d'un pays

PURB : population urbaine d'un pays

RURB : taux d'urbanisation d'un pays ($RURB = PURB / PTOT$)

GNPC : PIB per capita

Développement du modèle :

$$RURB = PURB / PTOT = A GNPC^c \quad (1)$$

$$PURB = A PTOT GNPC^c \quad (2)$$

$$PURB = A PTOT^b GNPC^c \quad (3)$$

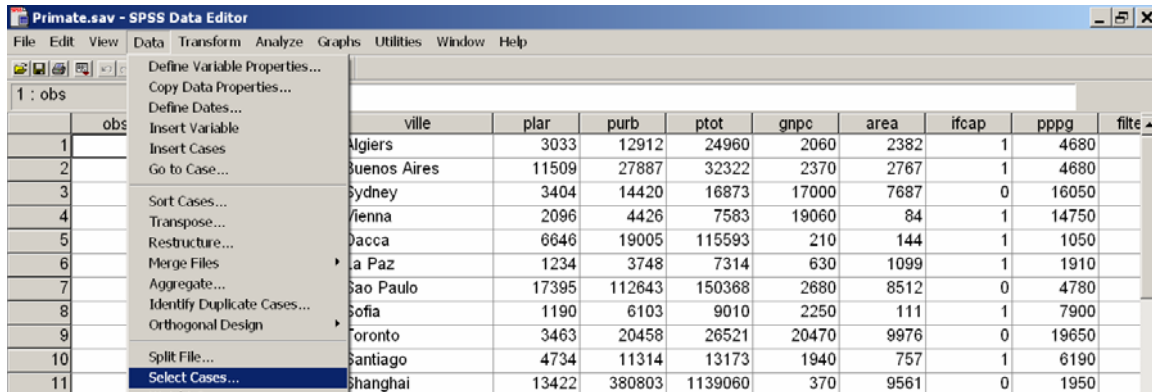
$$\ln PURB = a + b \ln PTOT + c \ln GNPC \quad (4)$$

$$\ln PURB = a + b \ln PTOT + c \ln GNPC + d (\ln GNPC)^2 \quad (5)$$

Source : Lemelin et Polèse (1995)

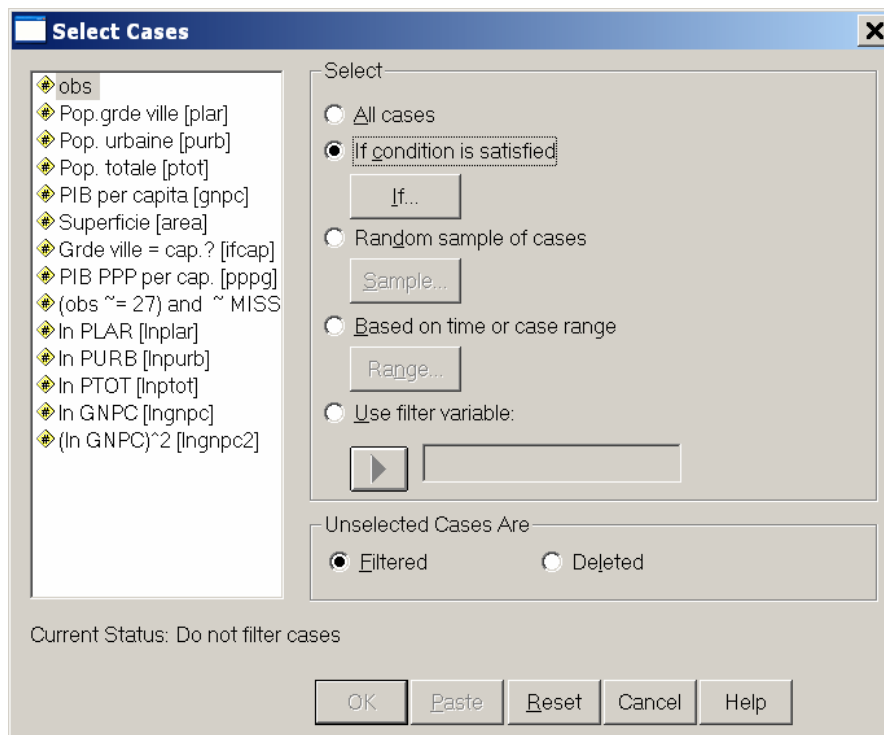
EXÉCUTION DE LA RÉGRESSION DANS SPSS

1. APPLICATION D'UN FILTRE POUR SUPPRIMER HONG KONG



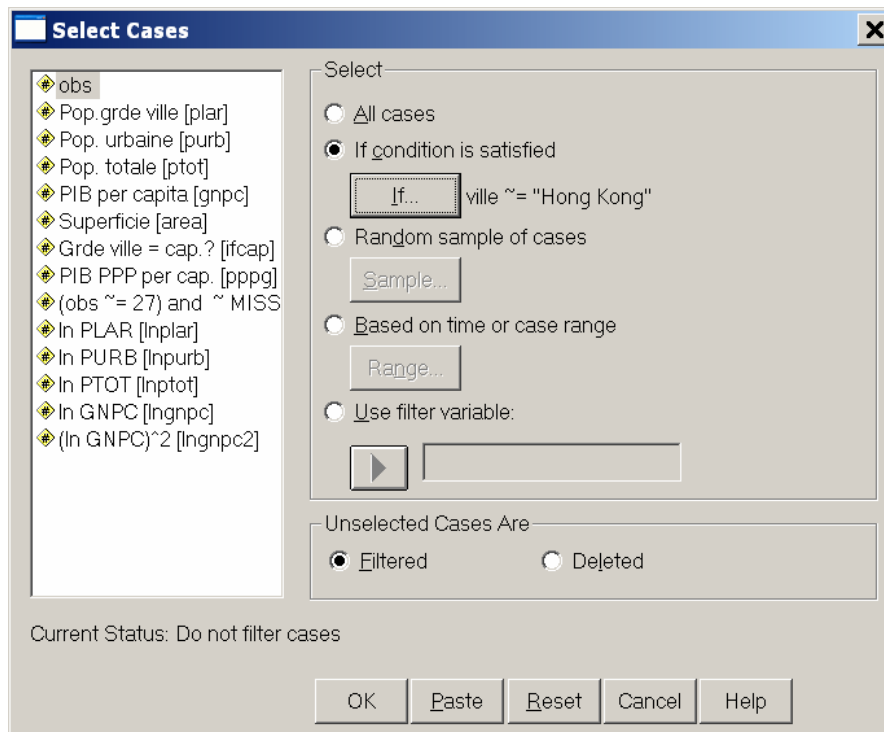
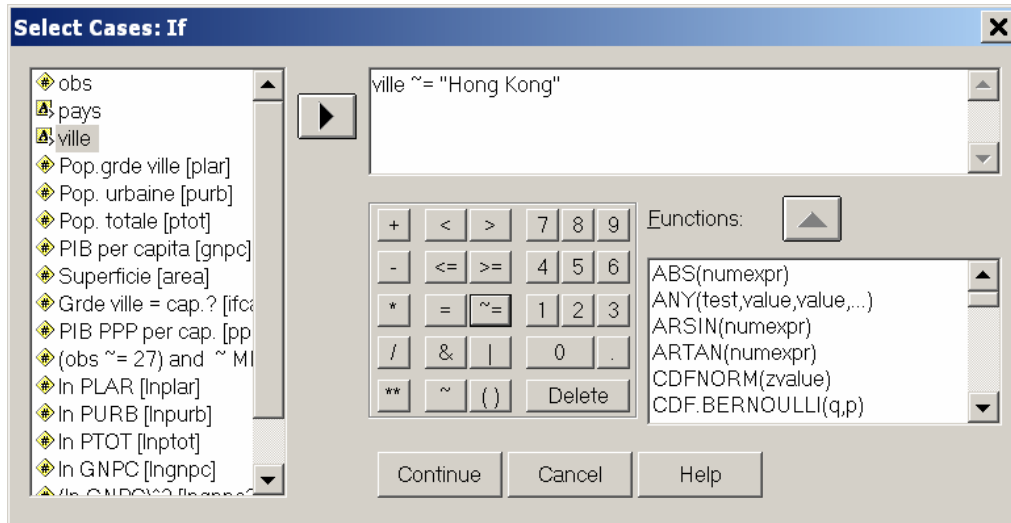
The screenshot shows the SPSS Data Editor window with the 'Select Cases...' option highlighted in the 'Data' menu. The data table below is a representation of the data shown in the editor.

obs	ville	plar	purb	ptot	gnpc	area	ifcap	pppg	filter
1	Algiers	3033	12912	24960	2060	2382	1	4680	
2	Buenos Aires	11509	27887	32322	2370	2767	1	4680	
3	Sydney	3404	14420	16873	17000	7687	0	16050	
4	Vienna	2096	4426	7583	19060	84	1	14750	
5	Dacca	6646	19005	115593	210	144	1	1050	
6	La Paz	1234	3748	7314	630	1099	1	1910	
7	Sao Paulo	17395	112643	150368	2680	8512	0	4780	
8	Sofia	1190	6103	9010	2250	111	1	7900	
9	Toronto	3463	20458	26521	20470	9976	0	19650	
10	Santiago	4734	11314	13173	1940	757	1	6190	
11	Shanghai	13422	380803	1139060	370	9561	0	1950	



EXÉCUTION DE LA RÉGRESSION DANS SPSS

1. APPLICATION D'UN FILTRE POUR SUPPRIMER HONG KONG (SUITE)



EXÉCUTION DE LA RÉGRESSION DANS SPSS

2. RÉGRESSION

obs	pays	plar	purb	ptot	gnpc	area	ifcap	pppg	filte
1	Algeria	3033	12912	24960	2060	2382	1	4680	
2	Argentina	11509	27887	32322	2370	2767	1	4680	
3	Australia	3404	14420	16873	17000	7687	0	16050	
4	Austria	26	7583	19060	84	1	14750		
5	Bangladesh	05	115593	210	144	1	1050		
6	Bolivia	48	7314	630	1099	1	1910		
7	Brazil	43	150368	2680	8512	0	4780		
8	Bulgaria	03	9010	2250	111	1	7900		
9	Canada	58	26521	20470	9976	0	19650		
10	Chile	14	13173	1940	757	1	6190		
11	China	03	1139060	370	9561	0	1950		
12	Colombia	79	32978	1260	1139	1	4950		
13	Costa Rica	20	3015	1900	51	1	4870		
14	Cote d'Ivoire	2168	4850	11997	750	322	1	1540	

Linear Regression

Dependent:

Block 1 of 1

Independent(s):

- ln PTOT [lnptot]
- ln GNPC [lngnpc]
- (ln GNPC)² [lngnpc2]

Method:

Selection Variable: Rule...

Case Labels:

WLS Weight:

Buttons: OK, Paste, Reset, Cancel, Help, Statistics..., Plots..., Save..., Options...

ANNOTATIONS DE LA SORTIE D'ORDINATEUR

1. La variable dépendante de la régression est $LPURB$, c'est-à-dire $\ln PURB$.
2. Ce tableau donne les résultats de l'estimation des paramètres a , b , c et d .
3. Les paramètres sont identifiés par le nom de la variable dont ils sont le coefficient (dans SAS, la variable `INTERCEP`, du mot anglais *intercept*, ordonnée à l'origine, désigne la constante du modèle).
4. Valeurs estimées des paramètres : \hat{a} , \hat{b} , \hat{c} et \hat{d} .
- 4b. Paramètres standardisés *Bêta* (dans SPSS seulement)

Les paramètres *Bêta* sont les valeurs que prendraient les coefficients si l'on remplaçait les variables du modèle par les variables standardisées z . Par exemple, on remplacerait

$$\ln PURB_i \text{ par } ZLPURB_i = \frac{\ln PURB_i - \overline{\ln PURB}}{s_{\ln PURB}}, \text{ avec } s_{\ln PURB} = \sqrt{\frac{1}{n-1} \sum_i (\ln PURB_i - \overline{\ln PURB})^2}$$

$$\ln PTOT_i \text{ par } ZLPTOT_i = \frac{\ln PTOT_i - \overline{\ln PTOT}}{s_{\ln PTOT}}, \text{ avec } s_{\ln PTOT} = \sqrt{\frac{1}{n-1} \sum_i (\ln PTOT_i - \overline{\ln PTOT})^2}$$

etc.

et le paramètre *Bêta* associé à $\ln PTOT$ est donné par $Beta_{\ln PTOT} = \hat{b} \frac{s_{\ln PTOT}}{s_{\ln PURB}}$

NOTE : Pour une raison que je ne comprends pas, les valeurs des écarts types produites par la procédure « Descriptives » dans SPSS ne sont pas assez précises pour vérifier le calcul qui vient d'être exposé. Mais la vérification se fait très bien dans Excel.

5. Erreur type (= écart type d'échantillonnage) des valeurs estimées des paramètres : $s_{\hat{a}}$, $s_{\hat{b}}$, $s_{\hat{c}}$ et $s_{\hat{d}}$
6. La valeur de la statistique t de Student pour l'hypothèse nulle que la valeur du paramètre est zéro; on a $t = \frac{\hat{a}}{s_{\hat{a}}}$, etc.
7. Probabilité critique correspondant à la valeur du t : c'est le seuil de signification qu'il faudrait choisir pour que la valeur critique correspondante soit égale à la valeur calculée du t ; c'est l'équivalent du résultat de la fonction `LOI.STUDENT` dans Excel (`TDIST` en anglais).

ANNOTATIONS DE LA SORTIE D'ORDINATEUR (SUITE)

8. Le tableau *Analysis of variance* (analyse de variance) détaille le calcul du R^2
9. Valeurs des sommes de carrés qui entrent dans le calcul du R^2 et qui servent aussi à construire les tests F de Fisher :
 - Ligne *Régression* dans SPSS ou *Model* dans SAS : SSM
= variabilité dont le modèle rend compte
 - Ligne *Résidu* dans SPSS ou *Error* dans SAS : SSR = variabilité résiduelle
 - Ligne « *C Total* » : SST = variabilité totale
10. Nombre de degrés de liberté (*ddl* ou *DF*, pour « degrees of freedom ») associé à chaque somme de carrés :
 - SSM : $k - 1$
 - SSR : $n - k$
 - SST : $n - 1$
11. Écarts quadratiques moyens (« Mean squares »); les moyennes sont calculées en divisant par le nombre de degrés de liberté correspondant :
 - Pour le modèle, $\frac{SSM}{k - 1}$
 - Pour les résidus, l'erreur quadratique moyenne (« Mean Square Error ») est donnée par

$$MSE = \frac{SSR}{n - k}$$

C'est la valeur estimée de la variance du terme aléatoire σ^2 .

12. Valeur de la statistique F de Fisher pour l'hypothèse nulle que **tous** les paramètres sont nuls, à l'exception de la constante. Cette statistique est calculée comme

$$F(k - 1; n - k) = \frac{SSM / (k - 1)}{SSR / (n - k)}$$

13. Probabilité critique correspondant à la valeur calculée de la statistique F .
14. Coefficient de détermination multiple :

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSR}{SST}$$

ANNOTATIONS DE LA SORTIE D'ORDINATEUR (SUITE)

15. Coefficient de détermination multiple ajusté :

$$\bar{R}^2 = 1 - \frac{SSR / (n - k)}{SST / (n - 1)}$$

16. Évaluation de la performance prédictive du modèle :

- Racine carrée de l'erreur quadratique moyenne (« Root MSE ») = \sqrt{MSE}
- Valeur moyenne de la variable dépendante = m_y (dans SAS seulement)
- « Coefficient de variation » (C.V.); en fait, il s'agit du coefficient de variation **en pourcentage**, c'est-à-dire de l'erreur moyenne de prévision du modèle = $100 \times \frac{\sqrt{MSE}}{m_y}$

(dans SAS seulement)

17. Test de l'hypothèse que le coefficient de *PTOT* est égal à 1. ce test est effectué avec la variable-test *F* de Fisher. Puisque le test *F* et le test *t* de Student sont basés sur le même modèle d'échantillonnage (modèle classique de la régression linéaire normale), le résultat est rigoureusement le même : la probabilité critique est égale à 0,3144 (voir 3-2.1.1).

SORTIE SPSS

Variables introduites/éliminées^b

Modèle	Variabiles introduites	Variabiles éliminées	Méthode
1	(ln GNPC) ² , ln PTOT ^a , ln GNPC	.	Introduire

1

a. Toutes variables requises introduites

b. Variable dépendante : ln PURB

Récapitulatif du modèle

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,976 ^a	,953	,951	,26876

a. Valeurs prédites : (constantes), (ln GNPC)², ln PTOT, ln GNPC

ANOVA^b

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	88,819	3	29,606	409,878	,000 ^a
	Résidu	4,334	60	,072		
	Total	93,152	63			

a. Valeurs prédites : (constantes), (ln GNPC)², ln PTOT, ln GNPC

b. Variable dépendante : ln PURB

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés		t	Signification
		B	Erreur standard	Bêta			
1	(constante)	-4,603	,877			-5,250	,000
	ln PTOT	,972	,028	,993		34,787	,000
	ln GNPC	,916	,208	1,147		4,412	,000
	(ln GNPC) ²	-,045	,013	-,873		-3,366	,001

a. Variable dépendante : ln PURB

SORTIE SAS

PRIMATE moins HK ; UN POP DIV et GNPC 15
18:19 Tuesday, August 29, 1995

1 Model: EQPURB
Dependent Variable: LPURB

9 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	3	88.81851	29.60617	409.878	0.0001
Error	60	4.33390	0.07223		
C Total	63	93.15241			

16 Root MSE 0.26876
Dep Mean 9.47138
C.V. 2.83760

11 R-square 0.9535
Adj R-sq 0.9511

12

13

2

3

4

5

6

7

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-4.602824	0.87667347	-5.250	0.0001
LPTOT	1	0.971663	0.02793210	34.787	0.0001
LGNPC	1	0.915958	0.20762625	4.412	0.0001
SQLGNPC	1	-0.045267	0.01344728	-3.366	0.0013

17

PRIMATE moins HK ; UN POP DIV et GNPC 16
18:19 Tuesday, August 29, 1995

Dependent Variable: LPURB

Test: PTOTEG1	Numerator:	0.0743	DF:	1	F value:	1.0292
	Denominator:	0.072232	DF:	60	Prob>F:	0.3144

Source : IREGSCI2.LST

PRIMATE moins HK ; UN POP DIV et GNPC 15
 18:19 Tuesday, August 29, 1995

Model: EQPURB
 Dependent Variable: LPURB

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	88.81851	29.60617	409.878	0.0001
Error	60	4.33390	0.07223		
C Total	63	93.15241			
Root MSE	0.26876	R-square	0.9535		
Dep Mean	9.47138	Adj R-sq	0.9511		
C.V.	2.83760				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-4.602824	0.87667347	-5.250	0.0001
LPTOT	1	0.971663	0.02793210	34.787	0.0001
LGNPC	1	0.915958	0.20762625	4.412	0.0001
SQLGNPC	1	-0.045267	0.01344728	-3.366	0.0013

PRIMATE moins HK ; UN POP DIV et GNPC 16
 18:19 Tuesday, August 29, 1995

Dependent Variable: LPURB

Test: PTOTEG1 Numerator: 0.0743 DF: 1 F value: 1.0292
 Denominator: 0.072232 DF: 60 Prob>F: 0.3144

Source : IREGSCI2.LST

PRIMATE moins HK ; UN POP DIV et GNPC 17
 18:19 Tuesday, August 29, 1995

Model: EQPURB1
 Dependent Variable: LPURB

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	87.99999	43.99999	520.920	0.0001
Error	61	5.15242	0.08447		
C Total	63	93.15241			
Root MSE	0.29063	R-square	0.9447		
Dep Mean	9.47138	Adj R-sq	0.9429		
C.V.	3.06851				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.901520	0.38173324	-4.981	0.0001
LPTOT	1	0.957929	0.02988116	32.058	0.0001
LGNPC	1	0.221163	0.02438139	9.071	0.0001

PRIMATE moins HK ; UN POP DIV et GNPC 18
 18:19 Tuesday, August 29, 1995

Dependent Variable: LPURB

Test: PTOTEG1 Numerator: 0.1674 DF: 1 F value: 1.9823
 Denominator: 0.084466 DF: 61 Prob>F: 0.1642

Source : IREGSCI2.LST

PRIMATE moins HK ; UN POP DIV et GNPC 24
 18:19 Tuesday, August 29, 1995

Model: EQPLAR1C
 Dependent Variable: LPLAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	37.70040	37.70040	222.803	0.0001
Error	62	10.49099	0.16921		
C Total	63	48.19138			
Root MSE	0.41135	R-square	0.7823		
Dep Mean	8.09242	Adj R-sq	0.7788		
C.V.	5.08316				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	2.066973	0.40693370	5.079	0.0001
LPURB	1	0.636174	0.04262018	14.927	0.0001

Source : IREGSCI2.LST

QUELLE EST LA FORME DE LA RELATION ENTRE LA POPULATION URBAINE ET LE PIB PER CAPITA ?

Équation linéaire

$$\ln PURB = a + b \ln PTOT + c \ln GNPC \quad (4)$$

avec

INTERCEP	1	-1.901520
LPTOT	1	0.957929
LGNPC	1	0.221163

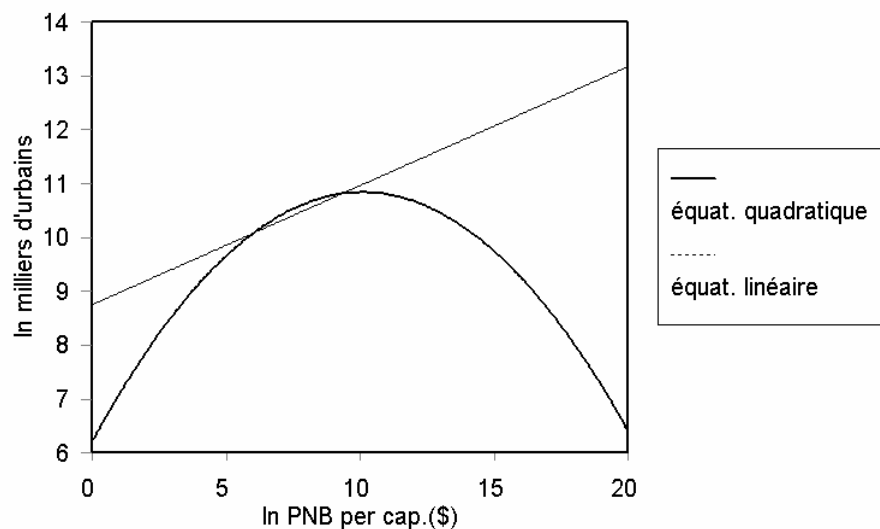
Équation quadratique

$$\ln PURB = a + b \ln PTOT + c \ln GNPC + d (\ln GNPC)^2 \quad (5)$$

avec

CONSTANTE	-4.602824
LPTOT	0.971663
LGNPC	0.915958
SQLGNPC	-0.045267

Population urbaine calculée pour une pop. totale de 67,6 millions



Repères

18. $0 = \ln(1)$: le point zéro sur l'axe horizontal représente un PIB per capita de 1 \$

19. $10 = \ln(22026)$: le maximum de la courbe se situe un peu au-delà, à 26 312 \$

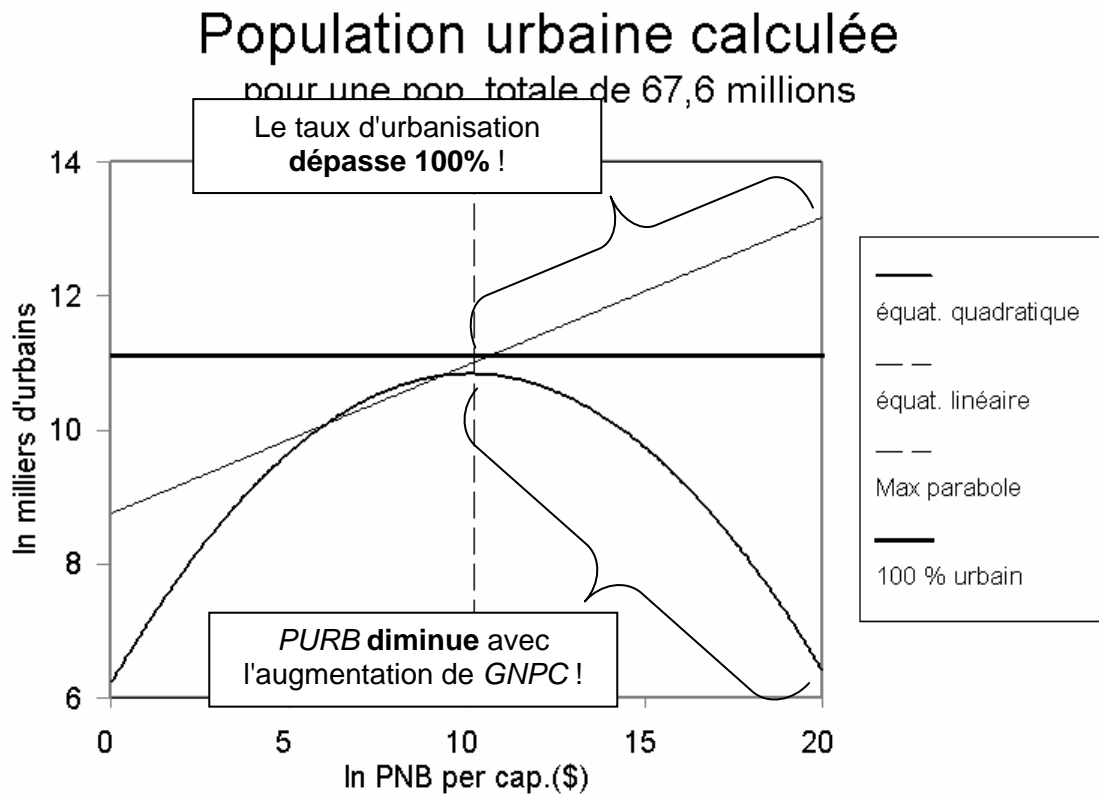
20. $6 = \ln(403\,429)$: pour un PIB per capita de 1\$, le modèle prédit une population urbaine d'environ 494 610; avec une population totale de 67,6 millions, cela donne un taux d'urbanisation de

$$\frac{PURB}{PTOT} = \frac{494\,610}{67\,600\,000} \approx \frac{7}{1000}$$

UN MODÈLE DESCRIPTIF QUI A SES LIMITES !

Il y a des parties des courbes qui ne sont pas pertinentes à la représentation du phénomène :

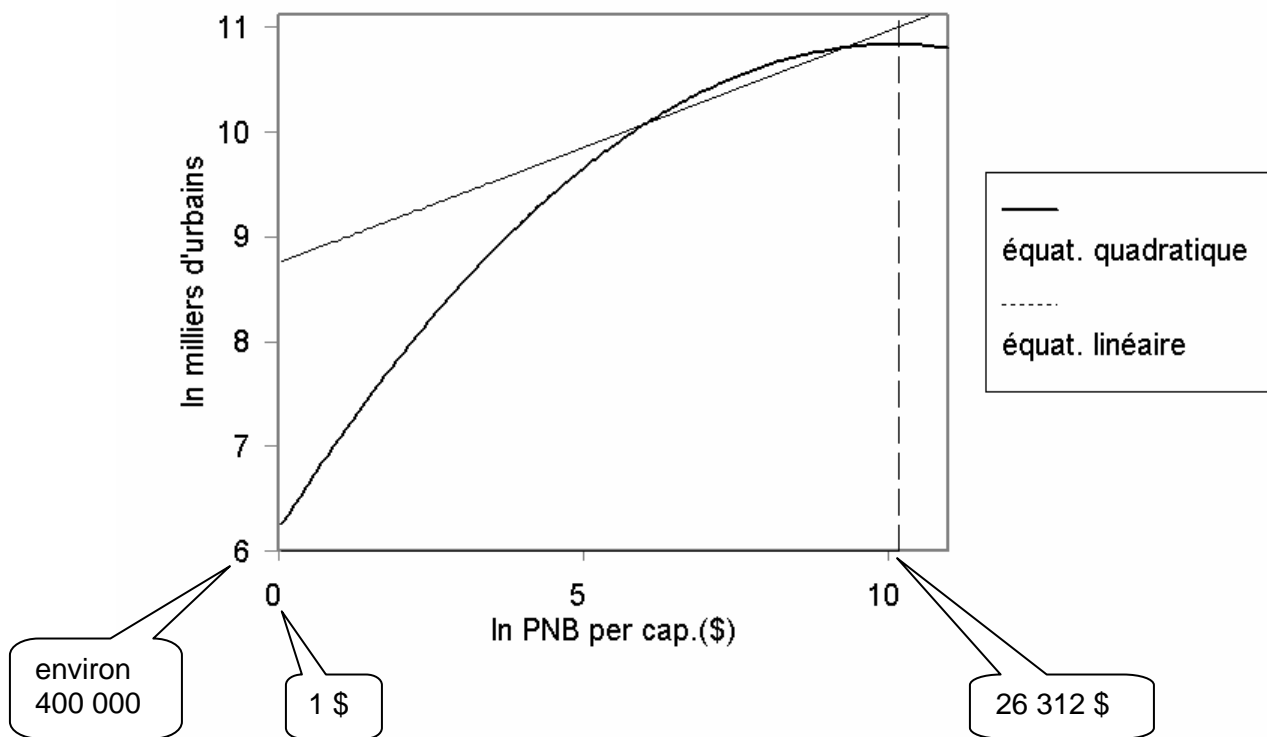
les modèles ne sont pas valides pour toutes les valeurs de la variable exogène.



EXAMEN DE LA ZONE PERTINENTE DU GRAPHIQUE DE LA RELATION ENTRE LA POPULATION URBAINE ET LE PIB PER CAPITA

Nous retenons seulement la partie pertinente des courbes.

Population urbaine calculée pour une pop. totale de 67,6 millions

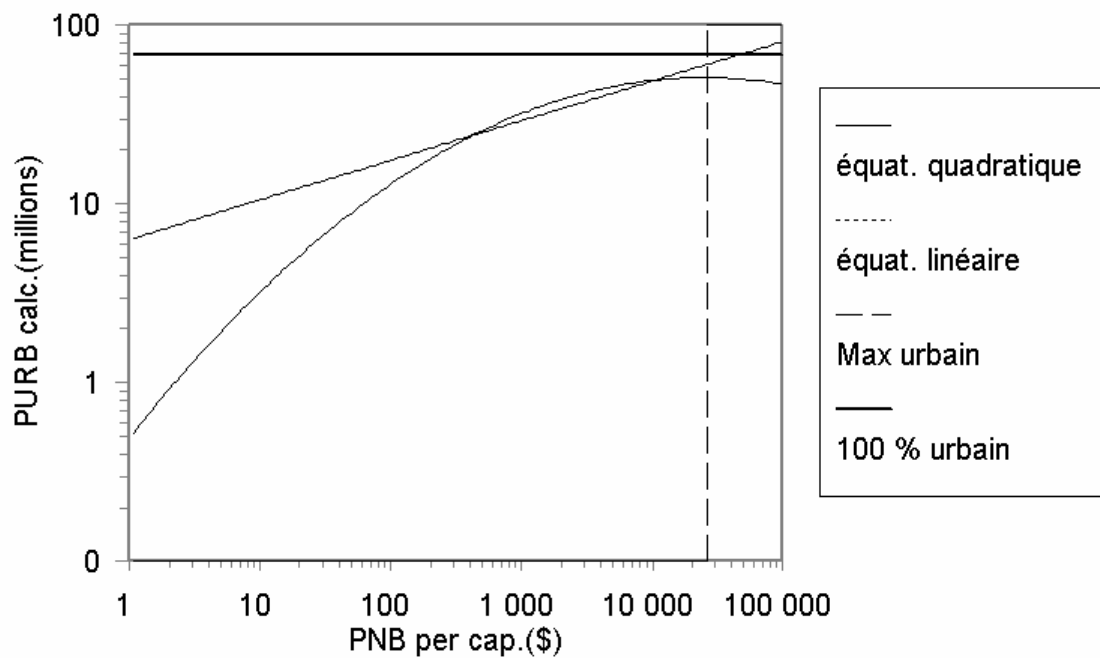


GRAPHIQUE DE LA RELATION ENTRE LA POPULATION URBAINE ET LE PIB PER CAPITA AVEC ÉCHELLES LOGARITHMES

Mêmes courbes, mais deux modifications qui s'annulent :

- Les logarithmes des variables sont remplacés par les variables elles-mêmes
- L'échelle des axes est remplacée par une échelle logarithmique

Population urbaine calculée pour une pop. totale de 67,6 millions

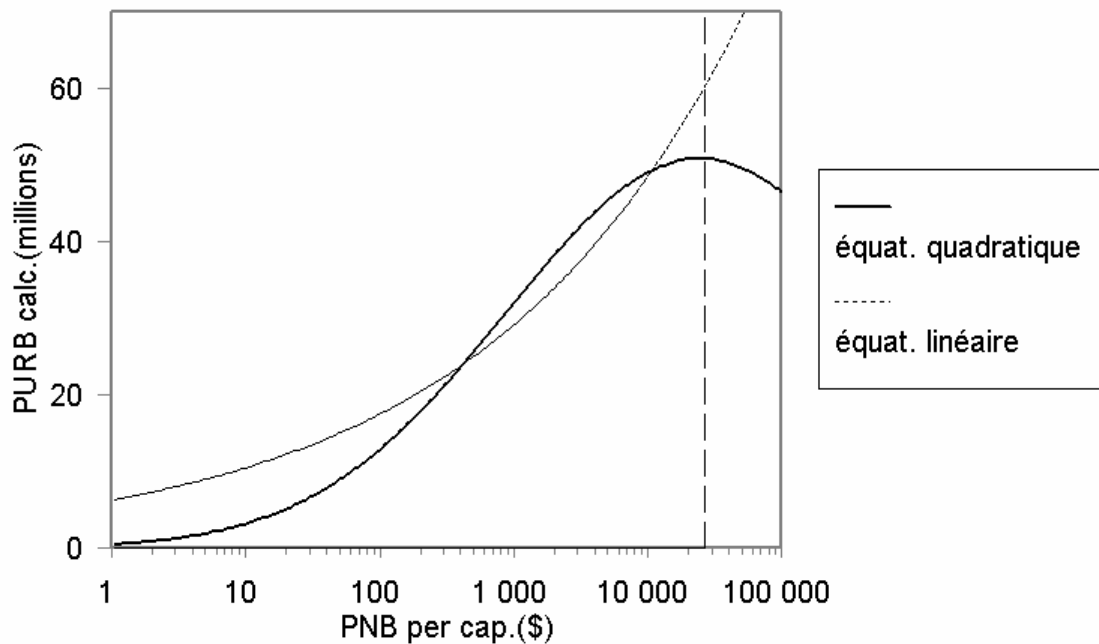


GRAPHIQUE SEMI-LOGARITHME DE LA RELATION ENTRE LA POPULATION URBAINE ET LE PIB PER CAPITA

Graphique semi-logarithmique

- L'axe des ordonnées est une échelle normale
- L'axe des abscisses est une échelle logarithmique

Population urbaine calculée pour une pop. totale de 67,6 millions

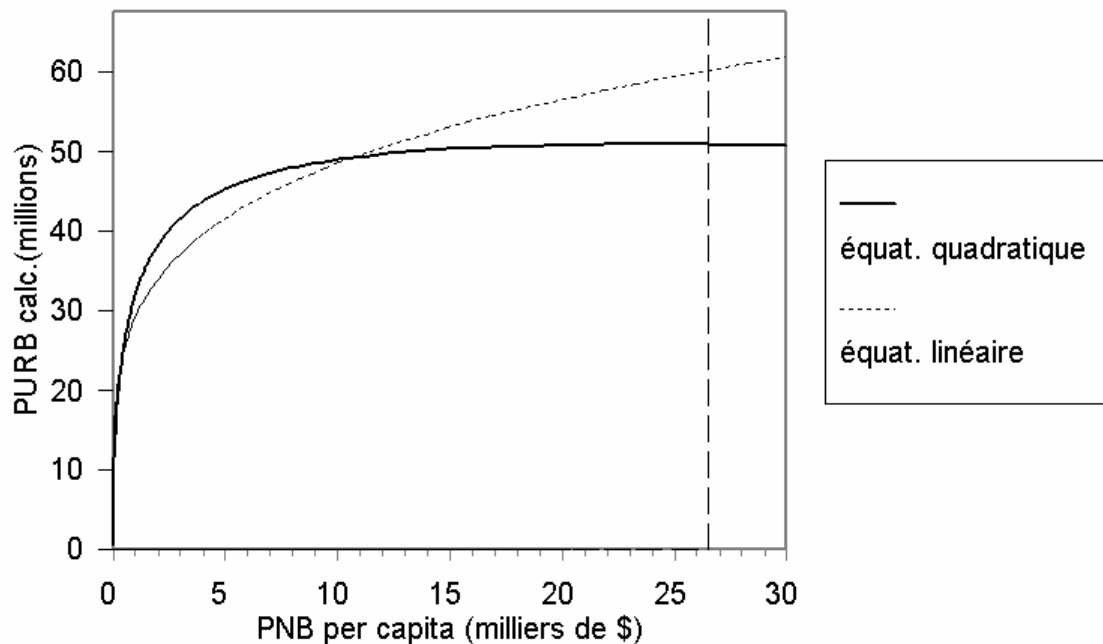


GRAPHIQUE NORMAL DE LA RELATION ENTRE LA POPULATION URBAINE ET LE PIB PER CAPITA

Graphique normal

- L'axe des ordonnées est une échelle normale
- L'axe des abscisses est une échelle normale

Population urbaine calculée pour une pop. totale de 67,6 millions

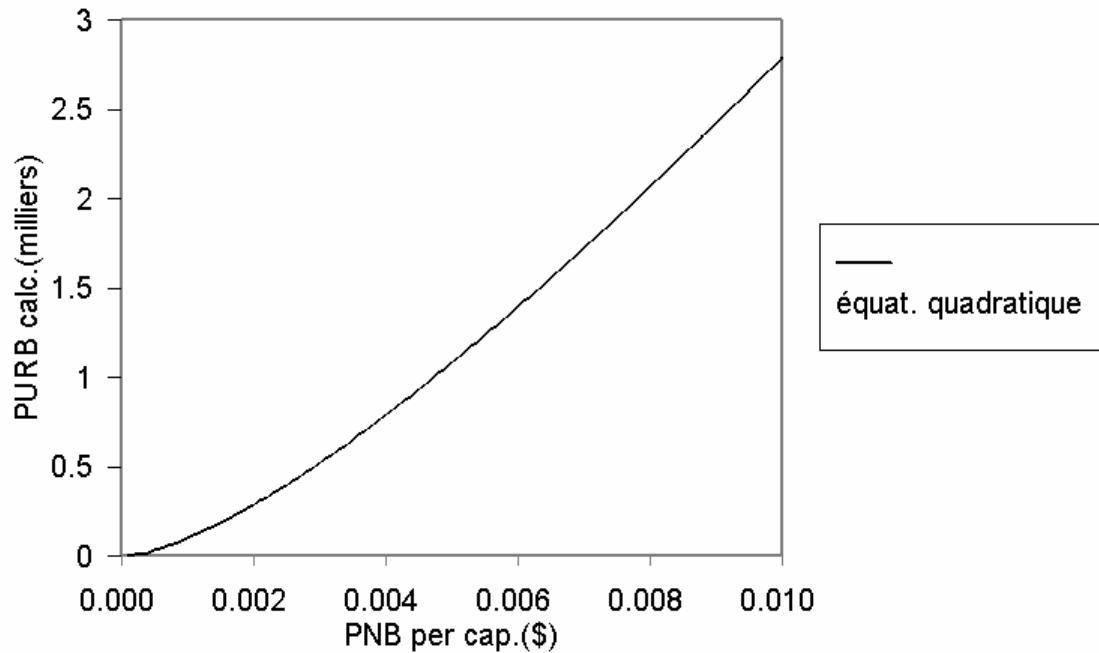


Où est passée la forme en « S » ?

OÙ EST PASSÉE LA FORME EN « S » ?

On retrouve la forme en « S » avec un microscope !

Population urbaine calculée pour une pop. totale de 67,6 millions



TAUX D'URBANISATION OBSERVÉ ET CALCULÉ SELON LE MODÈLE

Calcul du taux d'urbanisation selon le modèle

Le taux d'urbanisation calculé est dérivé de l'équation

$$\ln PURB = a + b \ln PTOT + c \ln GNPC + d (\ln GNPC)^2$$

qui donne

$$\ln \left(\frac{PURB}{PTOT} \right) = \ln PURB - \ln PTOT = a + (b - 1) \ln PTOT + c \ln GNPC + d (\ln GNPC)^2$$

c'est-à-dire

$$\left(\frac{PURB}{PTOT} \right) = \exp \left[a + (b - 1) \ln PTOT + c \ln GNPC + d (\ln GNPC)^2 \right]$$

Le taux d'urbanisation calculé n'est cependant pas le taux prédit par le modèle : pour avoir une courbe lisse dans le graphique en deux dimensions, on a fixé $PTOT$ à 67,626 millions pour tous les pays. Cela revient à ajuster la constante pour imposer la restriction $b = 1$.

Urbanisation observée/calculée

