ECONOMETRIE DE DONNEES DE PANEL

Cours Méthodologique EDOCIF

Examen Final
Juin 2002

Les documents de cours et les calcultarices sont autorisés

1 Exercice (8 points)

On considère un panel d'observations portant sur N individus et T périodes. Soit un modèle linéaire simple tel que $\forall\, i = 1, .., N, \forall\, t = 1, .., T$:

$$\underset{(T,1)}{y_i} = \underset{(T,1)}{e} \underset{(1,1)}{\alpha_i} + \underset{(T,K)}{X_i} \underset{(K,1)}{\beta} + \underset{(T,1)}{\varepsilon_i} \qquad \forall\, i = 1, .., N \tag{1}$$

où $\alpha_i \in \mathbb{R}$, $\beta' = (\beta_1 \beta_2 .... \beta_K) \in \mathbb{R}^K$ et où les résidus ε_i vérifient :

$$E(\varepsilon_i) = 0 \qquad E(\varepsilon_i \varepsilon_i') = V_i \tag{2}$$

On ne fait à ce stade de l'exercice aucune hypothèse sur la structure des matrices de variance covariance des résidus $V_i, \forall\, i = 1, .., N$.

1.1 Questions préliminaires

On suppose que la matrice V peut prendre les formes suivantes $\forall\, i = 1, .., N$:

$$V_i = \Omega_i \text{ non diagonale}$$
$$V_i = \Omega \text{ non diagonale}$$
$$V_i = \Omega_i \text{ diagonale}$$
$$V_i = \sigma^2 I_T$$
$$V_i = \sigma_i^2 I_T$$

Discutez suivant la forme des matrices V_i, les problèmes d'autocorrélation et d'héteroscédasticité qui se posent au niveau des résidus du modèle (1).

1.2 Estimateur Within

On suppose que les effets individuels α_i sont fixes. Soit Q l'opérateur Within défini par la relation

$$Q = I_T - \frac{1}{T} ee' \tag{3}$$

où I_T désigne la matrice identité de dimension (T, T) et e désigne un vecteur unitaire de dimension $(T, 1)$ avec $e'e = T$.

Question 1 : *Appliquez l'opérateur Within au modèle (1) et montrez que l'on obtient alors le modèle suivant*

$$\widetilde{y}_i = \widetilde{X}_i \beta + \widetilde{\varepsilon}_i \qquad \forall\, i = 1, .., N \tag{4}$$

où $\widetilde{y}_i$ et $\widetilde{X}_i$ désignent respectivement le vecteur de l'endogène centrée sur sa moyenne individuelle et la matrice des exogènes centrées sur leurs moyennes individuelles respectives.

Question 2 : *Déterminez la matrice de variance covariance* $\widetilde{V}_i = E\left(\widetilde{\varepsilon}_i\widetilde{\varepsilon}_i'\right)$ *des résidus* $\widetilde{\varepsilon}_i = Q\varepsilon_i$ *du modèle transformé (4). Montrez que*

$$\widetilde{V}_i = V_i \quad (5)$$

Question 3 : *D'après les résultats précédents, est ce que la transformation Within permet de corriger les problèmes d'hétéroscédasticité ou d'autocorrélation des résidus ? Quelles sont alors les solutions à envisager pour corriger de tels problèmes ?*

1.3 Estimateur Moindres Carrés Généralisés

On suppose à présent que les effets individuels du modèle (1) sont représentés par une v.a.r. α_i telle que $\forall\, i = 1, .., N$:

$$E(\alpha_i) = 0 \quad E(\alpha_i x_{i,t}) = 0 \quad (6)$$

$$E(\alpha_i\alpha_j) = \begin{cases} \sigma_\alpha^2 & i = j \\ 0 & \forall i \neq j \end{cases} \quad (7)$$

Question 1 : *Ecrivez le modèle (1) sous la forme d'un modèle à erreurs composées. Déterminez alors* Γ_i *la matrice de variance covariance des résidus de ce modèle en fonction de* σ_α^2, e *et* $V_i = E(\varepsilon_i\varepsilon_i')$*. Au regard de cette expression, quels sont les problèmes qui peuvent se poser ?*

Question 2 : On suppose à présent que ε_i satisfait la relation $V_i = E(\varepsilon_i\varepsilon_i') = \sigma_\varepsilon^2 I_{T.}$, $\forall\, i = 1, .., N$. *Commentez. En déduire la forme de* Γ_i*. Existe-t-il une corrélation inter-individuelle et/ou intra-individuelle des résidus ? Justifiez.*

Question 3 : On admet que l'inverse de la matrice Γ_i est alors définie par :

$$\Gamma^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[Q + \lambda\frac{1}{T}ee'\right] \quad (8)$$

$$\lambda = \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}\right) \quad (9)$$

Démontrez que l'on a les relations suivantes :

$$X_i'\Gamma^{-1}X_i = X_i'QX_i + \lambda T\ \overline{x}_i^2 \quad (10)$$

$$X_i'\Gamma^{-1}y_i = X_i'QX_i + \lambda T\ \overline{x}_i\overline{y}_i \quad (11)$$

avec $\overline{x}_i = (1/T)\sum_{t=1}^T x_{i,t}$ *et* $\overline{y}_i = (1/T)\sum_{t=1}^T y_{i,t}$.

Question 4 : On admet que l'estimateur $BLUE$ de β dans le modèle :

$$_i = X_i\beta + \mu_i \quad (12)$$

$$\mu_i = e\alpha_i + \varepsilon_i \quad (13)$$

$$E(\mu_i\mu_i') = \Gamma_i \quad (14)$$

est donné par l'estimateur des Moindres Carrés Généralisés. *Démontrez que cet estimateur peut s'écrire sous la forme suivante :*

$$\widehat{\beta}_{MCG} = \left[\sum_{i=1}^N X_i'QX_i + \lambda T\sum_{i=1}^N(\overline{x}_i)^2\right]^{-1}\left[\sum_{i=1}^N X_i'Qy_i + \lambda T\sum_{i=1}^N \overline{x}_i\overline{y}_i\right] \quad (15)$$

Commentez.

2 Problème (8 points)

Commentez l'article de Thornton (2001) concernant le lien entre les indicateurs des inégalités de revenu et la croissance du PIB réel par tête. Vous commencerez par décrire l'approche de l'auteur et insisterez sur les limites méthodologiques de cette étude et les réponses qu'il conviendrait d'y apporter ?

3 Exercice (4 points)

On considère un panel dont les données annuelles portent sur 17 pays[1] de l'OCDE et sont disponibles de 1951 à 1985 ($T = 35$). Soit $srt_{i,t}$ le nombre de jours chômés pour cause de grève, pour 1000 salariés du secteur industriel, du pays i observé à la date t. Nous cherchons à relier cette variable d'une part au taux de chômage de l'économie, noté $u_{i,t}$, et d'autre au niveau de l'inflation, notée $p_{i,t}$.

Question 1 : *Commentez rapidement le programme suivant et justifiez le choix de la forme estimée*[2]

```
load(file = "strikes.wks");
select i.ne.3;
? Construction des Variables
p2 = p**2;
u2 = u**2;
pu = p*u;
? Panel
panel (id = i, time = t, byid) srt p u p2 u2 pu;
```

Question 2 : *Analysez les résultats des fichiers de résultats TSP joints et dégagez le modèle adéquat qui doit être estimé dans ce cas ainsi que les résultats obtenus.*

[1] Le panel comporte initialement 18 pays, mais pour le pays 3, les données ne sont disponibles que jusqu'en 1980. C'est pourquoi afin de travailler sur un panel cylindré nous avons choisi de retirer ce pays de notre échantillon.

[2] On rappelle que l'instruction ** correspond au symbole puissance. Exemple U**2 équivaut à u^2.

Applied Economics Letters, 2001, **8**, 15–16

The Kuznets inverted-U hypothesis: panel data evidence from 96 countries

JOHN THORNTON

International Monetary Fund, 700 19th Street NW, Washington DC, USA
E-mail: jthornton@imf.org

Received 14 May 1999

Regression results from a panel data set of high-quality comparable data on Gini coefficients, income quintiles and real GDP per capita in 96 countries over the postwar period, suggest that the relation between income inequality and development corresponds to an inverted-U, as hypothesized by Kuznets.

I. INTRODUCTION

The Kuznets (1955, 1963) 'inverted-U' hypothesis suggests that, as a country develops, income inequality worsens initially but then improves as development proceeds. Kuznets argues that this is due to a shift of labour from low-productivity to high-productivity sectors in the early stage of development, which results in an increasing disparity in wages. Later, however, the high-productivity sector comes to dominate the economy, and wage inequality decreases. Aghion and Bolton (1992) and Galor and Tsiddon (1996) derive the Kuznets curve from the assumption that a (small) class of rich provides enough savings to begin the process of capital accumulation. The returns to their investment initially cause a widening of the distribution of income but, as aggregate income rises over time, the number of households that can afford to invest expands and the distribution narrows again.

The empirical evidence on the relationship between income redistribution and growth is still quite weak. Kuznets originally based his curve on observations of only three countries (Germany, the UK and the USA) and, although later studies extended the sample to as many as 60 countries, the empirical basis of the Kuznets curve remains in some dispute (see Anand and Kanbur, 1993). One fundamental problem has been the difficulty in constructing income distribution series that are comparable across countries. In this paper, the results of a test of the Kuznets hypothesis using a panel data set comprising income distribution and real GDP from 96 countries are reported. The results support the Kuznets hypothesis: income inequality appears to decline at higher income levels; moreover, a substantial part of the sample space is marked by an increasing tendency to greater income equality.

II. MODEL, DATA AND RESULTS

The basic form of the Kuznets hypothesis suggests a quadratic relation between income inequality and real GDP (the level of development), in which inequality increases with real GDP at early stages and, after reaching a peak, declines with economic growth. A common specification along these lines is:

$$INEQ_{it} = \alpha_0 + \beta_1 \ln Y_{it} + \Omega_1 (\ln Y)^2_{it} + \varepsilon_{it} \qquad (1)$$

where $INEQ_{it}$ is a measure of income inequality in country i and year t, $\ln Y$ denotes the natural logarithm of real gross domestic product (GDP) per capita to represent the level of economic development, and is a zero-mean, serially uncorrelated error term. The Kuznets hypothesis implies $\beta_1 > 0$, $\Omega_1 < 0$, provided that the entire 'inverted-U' is captured. The data on income distribution are from a recent high-quality compilation by Deininger and Squire (1996) which provides comparable series on the Gini index and quintile income shares for a large number of countries over much of the post World War II period; the panel is restricted to the subset of data which, in the view of those authors, is the highest-quality comparable subset. Data on real GDP per capita are taken from the January 1995 update (PWT 5.6) of Summers and Heston (1991),

Table 1. *Kuznets-type quadratic estimates from panel data for 96 countries*

Model observed	Constant	ln *Y*	(ln *Y*)^2	R^2/SEE	
(a) *GINI*	−27.1509	17.5852*	−1.1871*	0.0288	611
	(1.8162)	(4.4012)	(4.4653)	(9.2064)	
(a) *EQUAL*	0.5581*	−0.0781*	0.0049*	0.0127	547
	(5.1400)	(2.6857)	(2.5070)	(0.0659)	

Notes: *t*-statistics are in parenthesis below the parameter estimates; R^2 is adjusted for degrees of freedom; SEE is the standard error of the regression; the Gini coefficient (*GINI*) is measured in percentage points, the income share of the bottom 40 percent of the population (*EQUAL*) is in decimal points; ln *Y* is the logarithm of real GDP per capita in thousands of US dollars at 1985 international prices. * Implies statistically significant at the 95% level.

and the variable is measured in thousands of US dollars in international prices.

Table 1 presents estimates from two specifications of Equation 1. Panel (a) of the table reports the result using the Gini coefficient (*GINI*) as the measure of income *inequality*; in this case, the data set allows 611 observations from 96 developed and developing countries. Panel (b) reports the result using the sum of the first and second income quintiles (*EQUAL*) as the dependent variable, and for which the data set allow 547 observations from the same 96 countries; in this specification support for the Kuznets hypothesis implies $\beta_1 < 0$, $\Omega_1 > 0$. The results for the Gini coefficient reported support the hypothesis that the relation between income inequality and economic development takes the form of an inverted-U: the coefficients on ln *Y* and (ln *Y*)^2 are positive and negative, respectively, and are statistically significant at the 95 percent level. Thus, as real GDP per capita has increased, income inequality appear to have at first increased but then declined. The results using the first and second income quantiles as the dependent variable also support the Kuznets hypothesis: the coefficients on ln *Y* and (ln *Y*)^2 are negative and positive, respectively, and are statistically significant at the 95% level. A rough estimate of the sample turning point on the inverted-U can be derived by taking the exponent of $(-\beta_1/2\Omega_1)$ in Equation 1. Thus, the shift to greater income equality occurs when real GDP per capita is around \$1647 when the Gini coefficient variable and around \$2891 when the sum of the bottom two income quintiles is the dependent variable; both of these figures are well below the mean and median real GDP per capita for the sample (\$3644 and \$3019, respectively) suggesting that the 'turn' occurs early in the sample space, i.e. at a relatively low level of income.

III. CONCLUSION

This note reports results from a test of the relation between income inequality and real GDP growth per capita using a panel of high-quality, comparable data on Gini coefficients, income quintiles and real GDP per capita in 96 countries. The results suggest that the relation corresponds to an inverted-U, as hypothesized by Kuznets, and that the turning point on the inverted-U occurs at a relatively low level of income.

REFERENCES

Aghion, P. and Bolton, P. (1992) Distribution and growth in models of imperfect capital markets, *European Economic Review*, **36**, 603–11.

Anand, S. and Kanbur, S. (1993) Inequality and development: a critique, *Journal of Development Studies*, **41**, 19–43.

Deininger, K. and Squire, L. (1996) A new data set measuring income inequality, *World Bank Economic Review*, **10**, 565–91.

Galor, O. and Tsiddon, D. (1996) Income redistribution and growth: the Kuznets hypothesis revisited, *Economica*, **63**, S103–S117.

Kuznets, S. (1955) Economic growth and income inequality, *American Economic Review*, **45**, 1–28.

Kuznets, S. (1963) Quantitative aspects of the economic growth of nations: VIII. Distribution of income by size, *Economic development and Cultural Change*, **XI** (part II), 1–80.

Summers, R. and Heston, A. (1991) The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988, *Quarterly Journal of Economics*, **106**, 327–68.

```
                              TSP Version 4.3A
                         (06/07/95) DOS/Win      4MB
                    Copyright (C) 1995 TSP International
                             ALL RIGHTS RESERVED
                              06/12/02 3:00 PM
                In case of questions or problems, see your local TSP
                consultant or send a description of the problem and
the
                associated TSP output to:
                              TSP International
                         P.O. Box 61015, Station A
                            Palo Alto, CA 94306
                                    USA
          PROGRAM
 LINE
******************************************************************
 |      1  load(file="strikes.wks");
 |      2  select i.ne.3;
 |      3
 |      3  ? Construction des Series
 |      3  p2=p**2;
 |      4  u2=u**2;
 |      5  pu=p*u;
 |      6
 |      6  ? Panel
 |      6  panel (id=i,time=t,byid) srt p u p2 u2 pu ;
          EXECUTION

*********************************************************************
**********

 Current sample:  1 to 625

 Current sample:  1 to 70, 101 to 625

                       PANEL DATA ESTIMATION
                       =====================

 Balanced data:  NI=   17, T=   35, NOB=   595

 TOTAL (plain OLS) Estimates:

 Dependent variable: SRT

  Mean of dependent variable = 305.076        Std. error of
regression = 552.059
 Std. dev. of dependent var. = 571.637                        R-
squared = .075175
    Sum of squared residuals = .179509E+09          Adjusted R-
squared = .067324
       Variance of residuals = 304769.

            Estimated    Standard
 Variable  Coefficient     Error       t-statistic
 P         13.2165       14.6896       .899718
 U         94.3519       23.5889       3.99984
 P2        -.071900      .729535       -.098555
 U2        -6.98335      1.87748       -3.71953
 PU        1.24894       1.68431       .741517
 C         18.5695       70.4776       .263482

 F test of A,B=Ai,Bi:  F(96,493) = 2.2859,  P-value = [.0000]
 Critical F value for diffuse prior (Leamer, p.114) =   9.2602
```

BETWEEN (OLS on means) Estimates:

Dependent variable: SRT

Mean of dependent variable = 305.076 Std. error of regression = 171.959
Std. dev. of dependent var. = 278.196 R-squared = .737323
Sum of squared residuals = 325269. Adjusted R-squared = .617924
Variance of residuals = 29569.9

Variable	Estimated Coefficient	Standard Error	t-statistic
P	-146.890	123.786	-1.18665
U	184.484	105.350	1.75115
P2	12.4150	9.63187	1.28895
U2	-13.2877	12.0336	-1.10422
PU	.712552	19.5378	.036470
C	91.6256	400.425	.228821

WITHIN (fixed effects) Estimates:

Dependent variable: SRT

Sum of squared residuals = .146413E+09 R-squared = .245684
Variance of residuals = 255520. Adjusted R-squared = .218039
Std. error of regression = 505.490

Variable	Estimated Coefficient	Standard Error	t-statistic
P	33.4635	14.0852	2.37579
U	-13.6969	27.2537	-.502569
P2	-.962894	.682975	-1.40985
U2	-.713695	1.99589	-.357581
PU	-.100336	1.59034	-.063090

F test of Ai,B=Ai,Bi: F(80,493) = 1.1012, P-value = [.2702]
Critical F value for diffuse prior (Leamer, p.114) = 8.3854

F test of A,B=Ai,B: F(16,573) = 8.0952, P-value = [.0000]
Critical F value for diffuse prior (Leamer, p.114) = 6.7124

Variance Components (random effects) Estimates:

VWITH (variance of Uit) = 0.25552E+06
VBET (variance of Ai) = 49249.
(computed from small sample formula)
THETA (0=WITHIN, 1=TOTAL) = 0.12910

Dependent variable: SRT

Sum of squared residuals = .152469E+09 R-squared = .214483
Variance of residuals = 266089. Adjusted R-squared = .185694
Std. error of regression = 515.839

Variable	Estimated Coefficient	Standard Error	t-statistic
P	29.0100	13.9827	2.07471
U	13.6587	26.0568	.524188

```
P2        -.786655      .680600       -1.15583
U2        -2.26644      1.93938       -1.16864
PU        .145137       1.58349       .091656
C         173.510       89.0827       1.94774

Hausman test of H0:RE vs. FE:  CHISQ(5) = 17.949,  P-value = [.0030]


**********************************************************************
**********

END OF OUTPUT.

  MEMORY USAGE:    ITEM:    DATA ARRAY  TOTAL MEMORY
                  UNITS:  (4-BYTE WORDS) (MEGABYTES)
  MEMORY ALLOCATED        :    500000       4.0
  MEMORY ACTUALLY REQUIRED :     20119       2.2
  CURRENT VARIABLE STORAGE :     10096
```