**Comments on "Economic Integration: The Mature Portfolio Criterion", by Dimitris Christelis, Dimitris Georgarakos, and Michael Haliassos**

Raquel Carrasco

Madrid, April 24, 2007

# 1. Introduction

- This paper documents international differences in asset holdings among older households in the US, UK, and 11 European countries using internationally comparable household level data.
- The literature on international integration has mainly focused on features like flows of goods and services, prices, and consumption behavior.
- This paper introduces a different approach to analyzing international integration based on mature portfolios:
    - In a fully integrated world all households would experience the same prospects.
    - After controlling for "everything", remaining international differences in portfolios among households reflect departures from full integration.

# 2. Summary of the framework

- Let $y$ be the level of asset holdings.

- Variable $y$ takes on the value zero with positive probability and is a continuous variable over positive values (censored variable, corner solution variable).

- We could be interested in features of the distribution of $y$ given $x$, such as $E(y \mid x)$, $Q_\theta(y \mid x)$, or $\Pr(y = 0 \mid x)$.

- Consider the model:

$$y_{ij} = \alpha + x_i'(\beta + \sum_{j=1}^{J-1} d_j \delta_j) + \sum_{j=1}^{J-1} d_j \gamma_j + v_{ij} \quad i = 1, ..., N \quad j = 1, ..., J,$$

where $x_i$ is a vector of individual characteristics $(x_{i1}, ..., x_{ik})$, $d_j$ is a country dummy, and $v_{ij} \mid x_i, d_j \sim N(0, 1)$.

# 2.1. Asset participation

- Define the binary variable $w_{ij} = 1$ if $y_{ij} > 0$, $w_{ij} = 0$ if $y_{ij} = 0$. Then $w$ follows a probit model.

- Marginal effects of variable $x_{il}$:

$$\frac{\partial \Pr(w_{ij} = 1 \mid x_i, d_j)}{\partial x_{il}} = (\beta_l + \sum_{j=1}^{J-1} d_j \delta_j)\phi(\alpha + \sum_{j=1}^{J-1} d_j \gamma_j + x_i'(\beta + \sum_{j=1}^{J-1} d_j \delta_j)),$$

where $\phi(.)$ is the standard normal density function.

  - Specifically, for country $j = s$ ($d_s = 1$):

$$\frac{\partial \Pr(w_{is} = 1 \mid x_i, d_s = 1)}{\partial x_{il}} = (\beta_l + \delta_s)\phi(\alpha + \gamma_s + x_i'(\beta + \delta_s)).$$

- Country effects (relative to US):

$\Pr(w_{ij} = 1 \mid x_i, d_{US} = 1) - \Pr(w_{ij} = 1 \mid x_i, d_s = 1) = \Phi(\alpha + \gamma_{US} + x_i'(\beta + \delta_{US})) - \Phi(\alpha + \gamma_s + x_i'(\beta + \delta_s))$,

where $\Phi(.)$ is the cdf of the standard normal.

# 2.2. Asset holdings

- Interest could be in $E(y \mid x)$, $E(y \mid x, y > 0)$, or $Q_\theta(y \mid x)$, $Q_\theta(y \mid x, y > 0)$.

- Specifically, in this paper the interest is in analyzing to what extent observed international differences in asset holdings at different points of the distribution arise from:
  - different observable characteristics of asset holders across countries (“characteristics”' effect), or
  - different “prices” of characteristics (“coefficients”' effect).

- Approach: Quantile Regressions (QR) for asset holdings for the sample of owners.
  - generalization of the Oaxaca-Blinder decomposition to a QR framework (Machado and Mata, 2005).
  - decomposition based on the construction of a counterfactual distribution of the asset holdings for European countries

# 3. Comments

## 3.1. Selectivity issues

- No discussion of the potential selection problem arising from the fact that the analysis is conditional on participation.
- If the outcome variable is "censored" it might yield inconsistent parameter estimates and misleading decomposition results.
- Large proportion of zeros for some countries (Table 1).

Table1. Household Risky Financial Assets*, PPP-adjusted (euros)

| Countries | $N$ | % Participation | Mean | Conditional Mean |
|---|---|---|---|---|
| Sweden | 2,905 | 76.8 | 33,691 | 43,869 |
| Denmark | 1,539 | 59.8 | 15,232 | 25,453 |
| Germany | 2,826 | 28.4 | 7,534 | 26,483 |
| Netherlands | 2,705 | 25.8 | 11,606 | 44,852 |
| France | 1,542 | 45.6 | 11,414 | 25,036 |
| Switzerland | 909 | 36.1 | 41,626 | 115,360 |
| Austria | 1,885 | 9.2 | 2,719 | 29,288 |
| Italy | 2,425 | 9.4 | 3,235 | 34,109 |
| Spain | 2,218 | 10.5 | 2,145 | 20,340 |
| Greece | 1,884 | 11.1 | 2,146 | 19,352 |

Source: SHARE

*Sum of direct stockholding and imputed share of mutual funds and retirement accounts invested in stocks.

- Censored regression model:

$$y_i = x_i'\beta + u_i, \quad E(u_i|x_i) = 0,$$

and $u_i \sim N(0, \sigma^2)$.

- The expectation of $y$ given $x$ consists of the conditional expectation of $y$ weighted with the probability of being censored:

$$\begin{aligned} E(y_i|x_i) &= E(y_i|x_i, y_i > 0) \Pr(y_i > 0) \\ &= x_i'\beta\Phi(x_i'\beta/\sigma) + \sigma\phi(x_i'\beta/\sigma) \neq x_i'\beta. \end{aligned}$$

- In contrast to the linear regression model, the conditional expectation $E(y_i|x_i)$ in the Tobit model depends on he variance of the error term $\sigma^2$.

- Truncated model:

$$E(y_i|x_i, y_i > 0) = x_i'\beta + E(u_i|x_i, u_i > -x_i'\beta).$$

- Again, the problem is that although $E(u_i|x_i) = 0$, in general $E(u_i|x_i, u_i > -x_i'\beta) \neq 0$.

- Specifically, under the assumption of normality:

$$E(y_i|x_i, y_i > 0) = x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{[1 - \Phi(x_i'\beta/\sigma)]},$$

  where $\lambda(z) = \phi(z)/[1 - \Phi(z)]$.

- There is little consensus regarding the most appropriate correction procedure for selectivity bias in quantile regression models.
  - Buchinsky (1998) approximates the selection term by a higher order series expansion based on the inverse Mill ratio.
  - Bauer and Sinning (2005) present a decomposition method for Tobit-models.

- Example: Differences in the estimated coefficients from:
  - $E(y_i|x_i, y_i > 0) = x_i'\gamma$
  - $E(y_i|x_i, y_i > 0) = x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{[1-\Phi(x_i'\beta/\sigma)]}$
  - $E(y_i|x_i) = x_i'\delta$
  - $E(y_i|x_i) = x_i'\beta\Phi(x_i'\beta/\sigma) + \sigma\phi(x_i'\beta/\sigma)$

- It seems worthwhile to estimate these type of models (Table 2) and to determine the contribution of characteristics, coefficients and selectivity in the gap.

Table 2. Pool estimates

| Variables | OLS$\mid y > 0$ | Truncated | OLS | Tobit | Generalized Select. |
|---|---|---|---|---|---|
| Educ2 | .2983*** | .2886*** | .4152*** | 1.4593*** | |
| Educ3 | .5163*** | .4991*** | .9035*** | 2.8674*** | |
| Educ4 | .7702*** | .7570*** | 1.9549*** | 5.0727*** | |
| Bequest motive | -.0051*** | -.0052*** | -.0074*** | -.0206*** | -.0033*** |
| Head working | -.0873 | -.0957 | .6553*** | 1.6516*** | -.2442*** |
| Head retired | -.2936*** | -.3115*** | -.2048*** | .2374 | -.1058*** |
| Bad health | -.3870*** | -.4035*** | -.5108*** | -2.4267*** | -.1462 |
| Recall ability | .0327** | .0317** | .1569*** | .5351*** | -.0101 |
| $log$Income | 0.2460*** | 0.2317*** | 0.6034*** | 1.9268*** | 0.1187** |
| Age | .1415*** | .1266*** | .1197*** | .2237** | .1125*** |
| Age$^2$ | -.0009*** | -.0007 *** | -.0008*** | -.0020** | -.0006*** |
| Male | .0684* | .0663* | .0952* | .4597*** | .0298 |
| Hhold. size | -.0146 | -.0191 | -.0774** | -.2599** | .0169 |
| Single | -.3649*** | -.3765*** | -1.0161*** | -3.270*** | -.0133 |
| Provides help | .0344 | .0299 | .6690*** | 1.5404*** | -.1262** |
| Voluntary actv. | .0636 | .0800 | .3984*** | .9196*** | -.0084 |
| Constant | 3.5452*** | 4.0248*** | 3.4277*** | -17.582*** | 6.8897*** |
| sigma | | 1.5450*** | | 8.6354*** | -1.3343*** |
| $N$ | 6,535 | | 20,835 | | |

## 3.2. Decomposition

- The hypothesis is that differences in the levels of assets of US and the European countries occur both because of differences in their average characteristics and differences in the way the market evaluates these characteristics.

- The decomposition method used is a very popular descriptive tool. Typically, one uses the separately estimated equations for two groups to decompose the difference into an unexplained and an explained part.

- The simplest decomposition procedure is to adopt one of the estimated structure as the one that would prevail under full integration. But alternative decompositions are possible.

- Mean regression decompositions:

$$\overline{y}_{US} - \overline{y}_{EU} = \overline{x}'_{US}\widehat{\beta}_{US} - \overline{x}'_{EU}\widehat{\beta}_{EU}. \tag{1}$$

  - QR: $Q_\theta(y_{US}) - Q_\theta(y_{EU}) = \overline{x}'_{\theta,US}\widehat{\beta}_{\theta,US} - \overline{x}'_{\theta,EU}\widehat{\beta}_{\theta,EU} +$ residual

- The terms on the right-hand side of (1) can be decomposed into either

$$\overline{y}_{US} - \overline{y}_{EU} = (\overline{x}_{US} - \overline{x}_{EU})'\widehat{\beta}_{EU} + \overline{x}'_{US}(\widehat{\beta}_{US} - \widehat{\beta}_{EU}), \tag{2}$$

or

$$\overline{y}_{US} - \overline{y}_{EU} = (\overline{x}_{US} - \overline{x}_{EU})'\widehat{\beta}_{US} + \overline{x}'_{EU}(\widehat{\beta}_{US} - \widehat{\beta}_{EU}). \tag{3}$$

- The formulations in (2) and (3) correspond to two different assumptions. For instance, if you use (3) is because you believe the assets structure that would prevail under international integration was more likely to be close to the US function that to the function of Europe.

- Because of the familiar index number problem these choices do not imply the same estimates of both effects.

- A further useful exercise would be to obtain estimates from both formulations, using them to establish the range within which the "true" values of the components presumably would fall.

- Another possibility is to derive this type of decomposition:

$$\overline{y}_{US} - \overline{y}_{EU} = (\overline{x}_{US} - \overline{x}_{EU})'\widehat{\beta}^{*} + \overline{x}'_{US}(\widehat{\beta}_{US} - \widehat{\beta}^{*}) + \overline{x}'_{EU}(\widehat{\beta}^{*} - \widehat{\beta}_{EU}), \tag{4}$$

  where $\widehat{\beta}^{*}$ is the asset structure under international integration. It can be estimated from a pooled sample of the two groups.

- Thus, in this decomposition the discrimination component is made up of two elements, one representing the amount by which US characteristics are overvalued and the other the amount by which Europe characteristics are undervalued.

- Tables 3 and 4: with only equations (2) and (3) to capture the values of the characteristic and coefficient components we would have a very wide range to deal with.

Table 3. Mean Values and Regression Coefficients

| | Spain | | Rest | | Pool | |
|---|---|---|---|---|---|---|
| Variables | Mean | Coeff. | Mean | Coef. | Mean | Coef. |
| Educ2 | 0.2307 | 0.1656 | 0.1706 | 0.2480*** | 0.1727 | 0.2450*** |
| Educ3 | 0.1495 | 0.1546 | 0.3036 | 0.2898*** | 0.2980 | 0.2936*** |
| Educ4 | 0.2264 | 0.5582* | 0.3437 | 0.5382*** | 0.3395 | 0.5439*** |
| Bequest motive | 10.170 | 0.0023 | 9.2420 | -0.0048*** | 9.2752 | -0.0045*** |
| Head working | 0.4145 | -0.2942 | 0.4321 | -0.0580 | 0.4315 | -0.0661 |
| Head retired | 0.2649 | 0.1011 | 0.43183 | -0.3343*** | 0.4258 | -0.3178*** |
| Bad health | 0.0256 | 0.1227 | 0.0285 | -0.3336*** | 0.0284 | -0.3283*** |
| Recall ability | 4.3119 | -0.0016 | 5.5169 | 0.0376*** | 5.4737 | 0.0370*** |
| logIncome | 10.530 | 0.2507** | 10.7742 | 0.2641*** | 0.9507 | 0.2627*** |
| Age | 61.380 | 0.0622 | 62.413 | 0.1555*** | 62.376 | 0.1490*** |
| Age$^2$ | 3846 | -0.0004 | 3976 | -0.0009*** | 3972 | -0.0009*** |
| Male | 0.4786 | -0.0174 | 0.4978 | 0.0550 | 0.4971 | 0.0546 |
| Hhold. size | 2.8418 | -0.1346 | 2.1391 | -0.0406 | 0.4971 | -0.0555* |
| Single | 0.1581 | -0.5146* | 0.1693 | -0.2693*** | 0.1689 | -0.2935*** |
| Provides help | 0.1367 | 0.3706 | 0.3640 | 0.0668 | 0.3559 | 0.0761* |
| Voluntary actv. | 0.0598 | 0.5658 | 0.1890 | 0.1126** | 0.1843 | 0.1229** |
| Constant | 1 | 4.4891 | 1 | 0.4180 | 1 | 0.6921 |
| $N$ | 233 | | 6,287 | | 6,520 | |

Table 4. Decomposition of differentials
(percentage in parantheses)

| Nondiscriminatory Decomposition | Charact. effect | Coefficient effect (Spain's Advantage) | Coefficient effect (Rest's Disadvantage) |
|---|---|---|---|
| (4) | -0.2532 | -0.1426 | -0.0125 |
| | (62 %) | (34.9 %) | (3.1 %) |
| | Charact. effect | | Coefficient effect |
| Decomposition (2) | -0.3984 | | -0.001 |
| | (97.6 %) | | (2.4 %) |
| Decomposition (3) | -0.2383 | | -0.1701 |
| | (58.3 %) | | (41.6 %) |

# 3.3. Decomposition in QR

- In the context of quantile regression models, the decomposition of the gap at different quantiles is not entirely straightforward.
- While the decomposition of the gap at the mean is exact, this property is lost when applied to the gap at quantile $\theta$.
- In the case of the mean, a lineal specification implies that:

$$y_i = x_i'\beta + u_i \ \rightarrow \ E(y_i \mid x_j) = x_i'\beta,$$

since $E(u_i|x_i) = 0$.

- Thus, the properties of OLS estimators ensure that the predicted $y$ evaluated at the vector of mean characteristics of the sample is exactly the average $y$:

$$E(y_i) = E(x_i')\beta.$$

- Therefore we can obtain an exact decomposition of the average gap between both groups.

- However, in the context of quantile regression:
$$y_i = x_i' \beta_\theta + u_{\theta i}, \tag{5}$$
where $Q_\theta(y_i \mid x_i) = x_i' \beta_\theta$ and $Q_\theta(u_{\theta i} \mid x_i) = 0$.

- The expectation of (5) conditional on $y_i$ being equal to its unconditional quantile of order $\theta$, yields
$$Q_\theta(y_i) = E[x_i' \mid y_i = Q_\theta(y_i)]\beta_\theta + E[u_{\theta i} \mid y_i = Q_\theta(y_i)]. \tag{6}$$

- In this case $E[u_{\theta i} \mid y_i = Q_\theta(y_i)] \neq 0$.

- This latter term does not appear in the man regression and thus, in a quantile regression decomposition, there will be some part of the gap left unassigned to either the "characteristics" or "coefficients" components.

- It could be useful to comment on the size of this unexplained component.

## 3.4. A final comment

- Problem of omitted variables. Since the "coefficient" term is a residual, for it to be an accurate measure of lack of international integration all of the factors that determine the level of asset holdings must be properly accounted for.

- If they are not, then the the "coefficient" term will reflect these omitted influences as well, and will therefore either over- or underestimate the extent of lack of economic integration.

- This is a long-standing problem and not very much can be done except to recognize the role it plays on the interpretation of results.