UNIVERSITÀ POLITECNICA DELLE MARCHE
DIPARTIMENTO DI ECONOMIA

# INTERVAL REGRESSION MODELS WITH ENDOGENOUS EXPLANATORY VARIABLES

Giulia Bettin    Riccardo (Jack) Lucchetti

QUADERNI DI RICERCA n. 339

April 2010

**Abstract**

We consider the estimation of linear models where the dependent variable is observed by intervals and some continuous regressors may be endogenous. Our approach is fully parametric and two estimators are proposed: a two-step estimator and a limited-information maximum-likelihood estimator.

The results can be summarised as follows: the two-step estimator may offer some computational advantages over the LIML (Limited Information Maximum Likelihood) estimator, and a Monte Carlo experiment suggests that its relative efficiency is rather satisfactory. The LIML estimator, however, is probably simpler to implement and has the advantage of providing a framework in which several testing procedures are more straightforward to perform. The application of TSLS (Two-Stage Least Squares) to a proxy of the dependent variable built by taking midpoints, on the other hand, leads to inconsistent estimates.

An example application is also included, which uses Australian data on migrants' remittances.

| | |
|---|---|
| **JEL Class.:** | C24, C25, F24, F22 |
| **Keywords:** | Interval models, Instrumental variables, Migration, Remittances |

**Indirizzo:** Giulia Bettin: Dipartimento di Studi Economici – Università di Napoli "Parthenope" – Napoli, Italy `g.bettin@univpm.it`
Riccardo (Jack) Lucchetti: Dipartimento di Economia - Università Politecnica delle Marche – Ancona, Italy `r.lucchetti@univpm.it`

# Interval Regression Models with Endogenous Explanatory Variables*

*Giulia Bettin*  *Riccardo (Jack) Lucchetti*

## 1 Introduction

Interval data are very common in survey databases. In most cases, when it comes to recording economic and financial data, such as income, wealth and so forth, questionnaires provide an array of categories and the respondent is asked to state which category she belongs to.

Using this kind of data as the dependent variable poses a partial censoring problem, since the dependent variable of interest $y_i^*$ is unobserved; what is observed is an interval that contains it:

$$m_i \leq y_i^* \leq M_i$$

where the interval may be left- or right-unbounded.

While the idea of converting the intervals into a pseudo-continuous variable by taking midpoints may seem attractive to many a practitioner, it is in general a bad idea, as the resulting estimators possess no desirable properties. In Stewart (1983) this issue is comprehensively analysed and the estimation of interval models by maximum likelihood is advocated. Stewart's procedure can be briefly described as follows: the data generating process is assumed to be

$$y_i^* = X_i'\beta + \varepsilon_i \tag{1}$$

where $y_i^*$ is unobservable *per se*; however, once a distributional hypothesis for $\varepsilon_i$ is made, estimation becomes a simple application of maximum likelihood techniques. Under normality, the log-likelihood for one observation is

$$\ell_i(\beta, \sigma) = \ln P(m_i < y_i^* \leq M_i) = \ln \left[ \Phi\left(\frac{M_i - X_i'\beta}{\sigma}\right) - \Phi\left(\frac{m_i - X_i'\beta}{\sigma}\right) \right] \tag{2}$$

*We would like to thank the Commonwealth Department of Immigration and Multicultural and Indigenous Affairs for providing us with the dataset. Thanks are also due to Massimiliano Bratti, Luca Fanelli, Chiara Monfardini, Giulio Palomba, Sergio Pastorello and the participants to the* 1st Gretl Conference*, Bilbao, 2009 for useful comments and suggestions.*

and the total log-likelihood can be maximised by standard numerical methods, which are, in most cases, very effective. The above procedure is implemented natively in several econometric packages, among which Gretl, Limdep, Stata and TSP.

The extension of this model to the general case of endogenous regressors was considered in a non-parametric setting by (Hong and Tamer, 2003), building on the same techniques as in (Manski and Tamer, 2002). However, these techniques seem rare in the applied literature, arguably because of their complexity[1]. For example, in a recent paper (Neumann, Olitsky, and Robbins, 2009) a Mincer wage equation was estimated via an interval model. Regrettably, the lack of a viable estimating procedure prevented the authors from taking into account the possible endogeneity of education, as customary in this type of literature.

If we confine our attention to the case when the endogenous regressors are continuous, estimation can be performed by much simpler methods: equation (1) can be generalised to

$$\begin{aligned} y_i^* &= Y_i'\beta + X_i'\gamma + \varepsilon_i && (3) \\ Y_i &= \Pi_1 X_i + \Pi_2 Z_i + u_i = \Pi W_i + u_i && (4) \\ \begin{bmatrix} \varepsilon_i \\ u_i \end{bmatrix} &\sim N\left(0, \begin{bmatrix} \sigma_\varepsilon^2 & \theta' \\ \theta & \Sigma \end{bmatrix}\right) && (5) \end{aligned}$$

where $Y_i$ is a vector of $m$ explanatory variables and $\theta$, the covariance between $\varepsilon_i$ and $u_i$, may be nonzero. In this case, $Y_i$ becomes endogenous and ordinary interval regression does not provide consistent estimates of $\beta$ and $\gamma$.

In the next section, we will tackle estimation and the related computational issues. In section 3 the performance of several estimators will be compared by means of a Monte Carlo experiment, while section 4 contains an example application on real data.

# 2 Estimation

A simple two-step estimator can be defined in the same vein as the two-step estimator for probit models with endogenous regressors by Rivers and Vuong (1988).

In practice, two-step estimation may be carried out as follows[2]:

1. Perform the estimation of $\Pi$ and $\Sigma$ by running first-stage OLS of $Y_i$ on $W_i$ and collect the residuals $\hat{u}_i$.

[1] See Chesher (2007) for a comprehensive analysis of identification problems. See also Terza, Kenkel, Lin, and Sakata (2008).

[2] We assume that the instruments $Z_i$ satisfy the order and rank identification conditions.

2. Estimate the structural parameters by running an ordinary interval regression with $Y_i$, $X_i$ and $\hat{u}_i$ as explanatory variables.

The above procedure is very easy to perform if an interval regression routine is available.

As for the asymptotic properties of the resulting estimator, its consistency follows quite trivially from the consistency of $\hat{\Pi}$ and the clear fulfillment of the identification condition stated in (Wooldridge, 2002, p. 354). However, estimation of the asymptotic covariance matrix for the structural parameters should take into account the fact that the second step uses the generated regressors $\hat{u}_i$. The covariance matrix for the parameter must therefore be computed through the logic explained, among others, in Wooldridge, section 12.5.2. Details are given in section B in the appendix.

The estimation problem can also be tackled by maximum likelihood. The loglikelihood can be split into a conditional component and a marginal component as follows:

$$\ell_i(\Psi) = \ell_i^C + \ell_i^M \tag{6}$$

where $\Psi$ is a vector containing all the parameters. Since $y_i^*$ is imperfectly observed, the conditional component is

$$\ell_i^C(\Psi^*) = \ln P(m_i < y_i^* \leq M_i | u_i) = \ln\left[\Phi(E_i) - \Phi(e_i)\right] \tag{7}$$

where

$$\begin{aligned} E_i &= \frac{M_i - \hat{y}_i}{\sigma} && (8)\\ e_i &= \frac{m_i - \hat{y}_i}{\sigma} && (9)\\ \hat{y}_i &= Y_i'\beta + X_i'\gamma + u_i'\delta && (10) \end{aligned}$$

The marginal component $\ell_i^M$ is just an ordinary multivariate Gaussian loglikelihood:

$$\ell_i^M = \ln f(u_i; \Psi^*) = -1/2\left[m\ln(2\pi) + \ln|\Sigma| + (Y_i - \Pi W_i)'\Sigma^{-1}(Y_i - \Pi W_i)\right] \tag{11}$$

Estimation can be routinely carried out via standard numerical optimisation algorithms, such as BFGS or similar[3]. However, it is interesting to note that

[3] A recent paper by (Kawakatsu and Largey, 2009) advocates the usage of the EM algorithm for dealing with numerical problems in a closely related case (the ordered probit model), but we found it unnecessary in our case, provided that the anaytical score is used. Details on the computation of the analytical score are provided in section A in the appendix.

point estimation is considerably simpler in the just-identified case than in the over-identified case, since it coincides with the two-step estimator. The proof can be obtained by a reasoning similar to the one presented in (Rivers and Vuong, 1988, p. 356), by noting that in the exactly identified case the model can be reparametrised in such a way that the parameters of the two components of the loglikelihood are variation free, so they can be maximised separately. In the over-identified case, the two-step estimator is still consistent, which makes it a natural choice for initialising the ML numerical search procedure.

In this context, several hypothesis tests are likely to be of interest: the most obvious one is an exogeneity test, which is constructed by testing for $\delta = 0$ by means of a Wald test. This test can be carried out equivalently for the two-step and the LIML estimator.

However, other hypotheses of interest are quite natural to test in a likelihood framework: first, a LR test for overidentifying restrictions may be simply computed via the difference $\ell^C$ and that for an interval regression of $(m_i, M_i)$ on $W_i$ and $\hat{u}_i$, which would be the unrestricted log-likelihood. Additionally, the per-observation contribution to the log-likelihood can be used for performing test against non-nested competing models, such as Vuong's (1989).

# 3 Does it really matter?

From the viewpoint of an applied economist, the methods outlined in the previous section may seem overkill. After all, how much inaccuracy do we introduce in the data by choosing the interval midpoint? In fact, a procedure that is commonly used is to approximate $y_i^*$ by

$$\tilde{y}_i = \frac{M_i + m_i}{2}$$

and assume that $\tilde{y}_i$ can be used as a proxy for $y_i^*$: an additional source of error in the model (most likely heteroskedastic), that could be accommodated via robust estimation of the parameters covariance matrix. Hence, running TSLS on $\tilde{y}_i$ may look as a sensible and inexpensive procedure.

Trivially, a first problem that arises with this method is that it does not provide an obvious indication on how to treat unbounded observations (that is, when $m_i = -\infty$ or $M_i = \infty$). A more serious problem, however, is that that the above procedure leads to substantial inference errors. The analytical explanation is obvious after rearranging equation (3) as

$$\tilde{y}_i = Y_i'\beta + X_i'\gamma + (\varepsilon_i + \xi_i), \tag{12}$$

where $\xi_i$ is defined as $\tilde{y}_i - y_i^*$. If the interval $(m_i, M_i)$ is "small", then $V(\xi_i)$ should be negligible[4] compared to $V(\varepsilon_i)$.

However, even if the basic instrument validity condition $E(\varepsilon_i|W_i) = 0$ holds, there is no reason why the midpoint rule should guarantee $E(\xi_i|W_i) = 0$. This can be proven by a simple extension to the IV case of the line of reasoning in (Stewart, 1983). As a consequence, the TSLS estimator converges in probability to a vector that differs from the true values of $\beta$ and $\gamma$.

To explore the consequences of the above, in a seemingly harmless case, we ran a Monte Carlo experiment. Its setup is:

$$
\begin{aligned}
y_i^* &= \gamma_0 + Y_i\beta + X_i\gamma_1 + \varepsilon_i \\
& \quad \gamma_0 = \beta = \gamma_1 = 1 \\
Y_i &= 1 + X_i + \sum_{j=1}^{5} Z_{ji} + u_i \\
\begin{bmatrix} \varepsilon_i \\ u_i \end{bmatrix} &\sim N\left(0, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)
\end{aligned}
$$

and the cutpoints are the vector $[-2, 0, 1, 2, 5]$, so that, for example, if $y_i^* = 3$, then $m_i = 2$ and $M_i = 5$. The variables $X_i$ and $Z_{ji}$ are independent $N(0,1)$. A "naive" proxy for $y_i^*$ was constructed via the midpoint rule, as

$$
\tilde{y}_i = \begin{cases} -4 & \text{for } \; y_i^* < -2 \\ \frac{M_i + m_i}{2} & \text{for } \; -2 < y_i^* < 5 \\ 10 & \text{for } \; y_i^* > 5 \end{cases}
$$

The above DGP was simulated with sample sizes of 100, 400 and 1600 observations; 40000 simulations were run in which the midpoint estimator was compared to the the maximum likelihood estimator and the two-step estimator.

The results are summarised in table 1 which is organised as follows: the first three lines report the mean bias, the median bias and the RMSE. The next two lines report the mean of the estimated standard errors[5] and the ex-post dispersion of the parameters, namely the standard error of the estimates across the 40000 replications. Note that estimated and Monte Carlo standard errors should roughly match, if inference is to be at all credible. The last row shows the frequency of rejection of the hypothesis that the corresponding parameter equals its true value at 95%.

The evidence from the Monte Carlo experiment is:

[4] By construction, the support of $\xi_i$ is a finite interval, whose length goes to 0 as $M_i - m_i \to 0$.

[5] For the midpoint estimator robust standard errors are used in the TSLS stage, while the robust "sandwich" estimator is used for LIML: see for instance (Davidson and MacKinnon, 1999, chap. 10).

Table 1: Monte Carlo experiment

| | Sample size = 100 | | | Sample size = 400 | | | Sample size = 1600 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\gamma_0$ | $\beta$ | $\gamma_1$ | $\gamma_0$ | $\beta$ | $\gamma_1$ | $\gamma_0$ | $\beta$ | $\gamma_1$ |
| | | | | | Midpoint | | | | |
| mean bias | 0.2220 | 0.3304 | 0.1866 | 0.2051 | 0.3500 | 0.1965 | 0.2012 | 0.3542 | 0.1986 |
| median bias | 0.2249 | 0.3235 | 0.1806 | 0.2061 | 0.3478 | 0.1950 | 0.2016 | 0.3535 | 0.1980 |
| RMSE | 0.2594 | 0.3913 | 0.2981 | 0.2159 | 0.3652 | 0.2282 | 0.2040 | 0.3580 | 0.2069 |
| estimated se | 0.1301 | 0.2047 | 0.2262 | 0.0665 | 0.1036 | 0.1147 | 0.0334 | 0.0520 | 0.0576 |
| MC se | 0.1343 | 0.2096 | 0.2325 | 0.0672 | 0.1044 | 0.1159 | 0.0336 | 0.0521 | 0.0580 |
| rejections | 43.470 | 34.358 | 12.338 | 85.047 | 94.040 | 39.612 | 99.985 | 100.000 | 93.637 |
| | | | | | LIML | | | | |
| mean bias | -0.0041 | 0.0051 | 0.0081 | -0.0012 | 0.0016 | 0.0018 | -0.0002 | 0.0002 | 0.0004 |
| median bias | -0.0012 | 0.0008 | 0.0012 | -0.0004 | 0.0004 | 0.0001 | -0.0000 | -0.0002 | -0.0001 |
| RMSE | 0.0985 | 0.1548 | 0.1617 | 0.0473 | 0.0749 | 0.0772 | 0.0235 | 0.0371 | 0.0388 |
| estimated se | 0.0947 | 0.1487 | 0.1531 | 0.0467 | 0.0741 | 0.0767 | 0.0232 | 0.0370 | 0.0384 |
| MC se | 0.0984 | 0.1547 | 0.1615 | 0.0473 | 0.0749 | 0.0772 | 0.0235 | 0.0371 | 0.0388 |
| rejections | 5.940 | 6.098 | 6.652 | 5.237 | 5.290 | 5.197 | 5.203 | 5.112 | 5.272 |
| | | | | | Two-step | | | | |
| mean bias | 0.0172 | -0.0164 | -0.0133 | 0.0039 | -0.0035 | -0.0033 | 0.0010 | -0.0011 | -0.0009 |
| median bias | 0.0196 | -0.0197 | -0.0194 | 0.0045 | -0.0046 | -0.0050 | 0.0012 | -0.0014 | -0.0014 |
| RMSE | 0.0953 | 0.1509 | 0.1575 | 0.0469 | 0.0745 | 0.0768 | 0.0234 | 0.0370 | 0.0388 |
| estimated se | 0.0917 | 0.1455 | 0.1503 | 0.0463 | 0.0738 | 0.0764 | 0.0232 | 0.0370 | 0.0384 |
| MC se | 0.0938 | 0.1500 | 0.1569 | 0.0468 | 0.0744 | 0.0767 | 0.0234 | 0.0370 | 0.0388 |
| rejections | 6.713 | 6.495 | 7.085 | 5.375 | 5.390 | 5.322 | 5.200 | 5.130 | 5.270 |

1. The LIML and the 2-step estimators are both consistent and the actual size of $t$-test is reasonably accurate even for small samples, although in this respect LIML seems to perform slightly better.

2. The relative efficiency of these two estimators is comparable in the RMSE metric. However, it should be noted that this may be a misleading comparison, since in the linear case the LIML estimator is known to possess no finite moments (see Mariano and Sawa (1972)) and it is not unlikely that a similar property carries over to the present case. On the contrary, the median bias seems to be lower for the LIML estimator than for the 2-step estimator.

3. The application of TSLS to the naïve midpoint dependent variable proxy leads to seriously inconsistent estimates and substantial inference errors.

# 4 Example: the Analysis of Immigrants' Remittance Behaviour

An empirical application of the estimators presented above deals with the determinants of remittance behaviour by immigrants and replicates the analysis put forward in Bettin, Lucchetti, and Zazzaro (2009).

Economic literature has long ago turned its attention to the analysis of the decision to remit at the microeconomic level, regarding it as a function of migrants' characteristics and of the household's welfare in the country of origin. Since the pioneering work of (Lucas and Stark, 1985) on Botswana, many attemps have been made to identify the main motivations to remit: altruism, inheritance, exchange, self-insurance and so forth (see (Rapoport and Docquier, 2005)).

The most crucial aspect in this context is that the empirical literature dealing with the topic usually treats migrant's income as an exogenous determinant of remittance behaviour. Yet, the need of sending money back home can affect working, consumption and possibly investment decisions. The amount of money to remit is determined in the broader context of the household's strategies. Hence, while estimating a remittance equation that detects the main determinants of remittance behaviour, endogeneity and reverse causality relationships between remittances, income, consumption and saving need to be addressed..

Since data for microeconomic analyses on remittance behaviour are typically taken from household surveys, the amount of remittances is often designed in the questionnaires as a discrete ordered variable, with mutually exclusive intervals.

The dataset used in this empirical application is the first cohort of the Longitudinal Survey of Immigrants to Australia (LSIA), selected from visaed immigrants aged 15 years and over, who came to Australia between 1993 and 1995.

Table 2: Remittances by immigrants: frequency distribution

| Amount | Abs. Freq. | Cumul. Freq. | % | Cumul. % |
|---|---|---|---|---|
| 1-1000 AUS$ | 1584 | 1584 | 64% | 64% |
| 1001-5000 AUS$ | 728 | 2312 | 29.4% | 93.6% |
| more than 5000 AUS$ | 164 | 2476 | 6.6% | 100% |

Individuals were interviewed three times: the first time five or six months after arrival, the second time one year later and the third one two years later[6]. Unlike most of the previous microeconomic studies on the topic, data here do not concern a specific ethnic group, but people from 125 different countries (both developed and developing).

When answering the question on remittance, immigrants could choose between six intervals: 1-1000 AUS $, 1001-5000, 5001-10000, 10001-20000, 20001-50000, more than 50000 AUS $. Since observations concentrate mainly in the first two intervals, the top four are compacted to a single one going from 5001 AUS$ upwards. Table 2 shows the frequency distribution for the remittance variable used in the estimation. The dependent variable in our model is therefore:

$$r_i = \begin{cases} 1 & \text{for} \quad 1 < R_i < 1000 \\ 2 & \text{for} \quad 1001 < R_i < 5000 \\ 3 & \text{for} \quad R_i > 5001 \end{cases}$$

where $R_i$ represents the real amount of money remitted.

The remittance equation that we estimate can be written as:

$$r_i = \alpha_1^* y_i + \alpha_2^* c_i + \alpha_3^* X_i + u_i \tag{13}$$

where $y_i$ is the yearly income of the migrants' household, $c_i$ total yearly consumption and $X_i$ a set of exogenous controls. $r_i$, $y_i$ and $c_i$ are all in natural logarithms.

$X_i$ includes two sets of control variables that can influence remittance behaviour. The first one refers to immigrants' individual characteristics: the age and its square, gender, a dummy for the presence of close relatives in the country of origin, another dummy for the intention to return to the home country and the time since the arrival in Australia. Dummies for education level are added as a further control. Educational attainment is divided into five levels, from upper tertiary to primary education[7].

[6]The time between interviews may vary substantially between households; this problem, together with considerable sample attrition, led us to ignore the "panel" aspect of our dataset and use all data as pooled data.

[7]Our group of reference is the highest level of education.

The second set of control variables includes characteristics of the origin countries: the level of per capita GDP[8], as a general measure of the level of development and the distance between Australia and the country of origin, to proxy for transaction costs for money transfers[9].

The problem of endogeneity of income and consumption is addressed using a set $Z_i$ of six instruments. The first two describe the migrant's knowledge of the English language: one dummy for English being the language the immigrant speaks best and another one which equals 1 if the immigrant declared a good knowledge of English[10]. The third instrument is a dummy variable stating if the immigrant lives in a urban or a rural environment. Dummies for the presence of partner and children in migrant's household are also employed, together with the number of migrant's household members (expressed in natural logarithm).

In order to focus on IV techniques for interval models, we chose to ignore problems that arise when the sample is made up of people who remit and people who do not, so that the dependent variable is truncated below a zero threshold[11].

Results are reported in Table 3. With non-IV interval estimates, both income and consumption appear statistically significant with a positive sign. The result is expected for income, and in line with the previous findings[12]; on the other hand, the only way to make sense of a positive effect of consumption on remittances is to consider this specification as a conditional mean, in which the coefficients do not have a behavioural interpretation, rather than a proper structural form, in which they do. Consumption may well be an excellent predictor of remittances, despite not causing them. This is likely to be true, for example, in a permanent-income setting, in which consumption follows a smoother time path than current income due to forward-looking behaviour. In this case, failing to take endogeneity into account makes a behavioural interpretation of the estimated coefficients impossible.

In fact, the Wald test for exogeneity rejects strongly the null hypothesis for income and consumption. Moreover, the result from a LR test of over-identifying

[8]Data are from the World Development Indicators database; the variable is the log of the mean per capita GDP over the period 1992-2000.

[9]Geographical distance could also represent the strength of family relationship with those left behind. The source of the data employed here is the CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) dataset on bilateral distances.

[10]Although it may be argued that the two instruments are highly correlated, we use them both to capture different aspects: on one hand, the direct effect of the English knowledge on migrants' job performance, and then income; on the other hand, their degree of integration in the host country, trying to take into account that if migrants rigidly stick to their native language and do not improve their use of English, this could be read as a difficulty (or even a choice not) to integrate in the host society.

[11]For a thorough analysis of this problem, see Bettin, Lucchetti, and Zazzaro (2009).

[12]Among others, see Brown (1997) and Clark and Drinkwater (2007).

Table 3: Estimates for the Australian remittances data

| | Non-IV | | Midpoint | | Two-step | | LIML | |
|---|---|---|---|---|---|---|---|---|
| const | 1.627 | | 34.226 | *** | 15.328 | *** | 16.106 | *** |
| male | 0.299 | *** | 0.713 | ** | 0.259 | ** | 0.255 | ** |
| age | 0.025 | | 0.250 | ** | 0.099 | ** | 0.103 | ** |
| age2 | 0.000 | | -0.003 | * | -0.001 | ** | -0.001 | * |
| time in AUS | 0.439 | *** | 1.438 | *** | 0.451 | *** | 0.450 | *** |
| back home | 0.412 | ** | 1.445 | ** | 0.461 | ** | 0.465 | ** |
| relatives overseas | 0.048 | | 0.007 | | -0.019 | | -0.029 | |
| qualifications_2 | 0.177 | | 0.533 | | 0.149 | | 0.149 | |
| qualifications_3 | -0.151 | | -0.697 | | -0.210 | | -0.207 | |
| qualifications_4 | -0.408 | ** | -0.390 | | -0.150 | | -0.118 | |
| qualifications_5 | -0.633 | *** | -1.645 | *** | -0.608 | *** | -0.601 | *** |
| per capita GDP | 0.188 | *** | 0.948 | *** | 0.283 | *** | 0.287 | *** |
| distance | -0.408 | *** | -1.286 | *** | -0.412 | *** | -0.411 | *** |
| income | 0.217 | ** | 3.050 | *** | 0.993 | *** | 1.083 | *** |
| consumption | 0.351 | ** | -6.725 | *** | -2.087 | *** | -2.275 | *** |
| N | 1135 | | 1130 | | 1130 | | 1130 | |
| sigma | 1.18 | | 5.23 | | 1.15 | | 1.14 | |
| Wald test | | | | | 25.825 | | 26.139 | |
| Wald test p-value | | | | | 2.47e-6 | | 2.11e-6 | |

restrictions yields a test statistic of 6.051 (p-value: 0.19), which confirms the validity of the set of instruments we chose to address the issue.

Two-step estimates (column 3) are only slightly different from LIML estimates (column 4). Income and consumption are in both cases statistically significant at 1%, with the signs predicted by the theory. Elasticity of remittances to income is not significantly different from 1, while elasticity to consumption is slightly larger than -2. It is noteworthy that this result matches the estimates presented in Bettin, Lucchetti, and Zazzaro (2009) very closely, despite the difference in the estimation techniques used.

The fact that the coefficient on income is never statistically different from 1 gives some support to exchange-driven models of remittances, in which their elasticty to income is usually not bigger than one and remittances therefore can be considered as a normal good. On the other hand, pure altruistic models typically predict a higher elasticity to income, meaning that remittances act as a superior good.

It is also interesting to compare the income and consumption elasticities to the results from the midpoint estimates. In column 2, the sign on the elasticity to consumption is negative as we espect when using IV techniques but the coefficients are much bigger (in modulus) both for income (3.050) and for consumption (-6.725) and hardly justifiable from an economic point of view. Keeping in mind also the outcomes of the Monte Carlo experiment presented in Section 3 it should be then rather clear that this empirical procedure may produce unreliable results.[13]

Among individual characteristics, gender differences result statistically significant: other things being equal, male migrants remit on average 25% more than female (Aggarwal and Horowitz, 2002, Amuedo-Dorantes and Pozo, 2006). The relationship between the age of the immigrant and the amount remitted seems to be nonlinear, with a positive coefficient on the variable and a negative one on its squared term[14]. The desire to return to the country of origin predictably affects the amount remitted in a significant way, with potential returnees remitting around 46% more, and the same positive effect is associated to the length of the period spent in Australia.

As far as the immigrants' education is concerned, what emerges is that, even after controlling for the level of income, more educated migrants are likely to remit higher amounts than the less educated as the repayment motivation would suggest (Funkhouser, 1995, Hoddinott, 1994).

Surprisingly, per capita GDP of immigrants' country of origin turns out to be significant with a positive sign. Immigrants coming from richer countries seem to remit more. However, if we consider that this variable might act as a rough proxy of the economic conditions of the recipients, the result is consistent with the exchange bargaining-type hypothesis of Cox, Eser, and Jimenez (1998), since a higher income for recipients might be associated with a stronger bargaining power with respect to the migrant and hence a more effective enforcement of the contract[15]. Finally, the distance from the country of origin plays a significant role in decreasing the amount of remittances. The intensity of the altruistic feelings is weakened as distance increases and all kinds of exchange become more and more difficult.

# 5 Conclusions

We argue that estimation of models in which the dependent variable is observed by intervals and explanatory variables may be endogenous ought to be conducted via appropriate methods, "sensible" alternatives being inefficient at best and plain

[13]It should be noted, however, that the midpoint estimates are roughly proportional to the estimates obtained by the two consistent estimation methods. It may be conjectured that the bias in the midpoint estimator is essentially a consequence of its failure to properly estimate the scale of the latent variable. This, in turn may be due to the fact that the imputed value for the unbounded category is essentially arbitrary. However, in our dataset the observations falling into the top category comprise only a small share (see Table 2), and it is quite unlikely that they exert such a large influence on the final outcome.

[14]Hoddinott (1994) and Clark and Drinkwater (2007) also found an inverted U shaped relationship between migrants' age and remittances.

[15]A similar result emerges also in Bettin, Lucchetti, and Zazzaro (2009).

wrong at worst.

These appropriate methods include maximum likelihood and 2-step estimation: as Monte Carlo evidence shows, they both lead to consistent estimates, but thanks to the ease of implementation and other properties, we find the LIML estimator to be slightly preferable. The two-step estimator, however, offers comparable performance and is far less demanding in terms of CPU usage, although more difficult to implement. The most interesting result to applied economists is that from our experiment the commonly used midpoint technique clearly results in inefficient and biased estimates.

An example with Australian remittances data shows that our procedure is effective. Endogeneity of income and consumption in the context of immigrants' remittance behaviour does matter. Failing to account for these endogeneity effects, that we showed to be altogether highly significant, will lead to incorrect estimates. A similar *caveat* applies to the midpoint IV technique, which appears to correct the estimates for endogeneity in the right direction, but grossly overestimates their absolute size.

# References


AGGARWAL, R., AND A. W. HOROWITZ (2002): "Are international remittances altruism or insurance? Evidence from Guyana using multiple-migrant households," *World Development*, 30(11), 2033–2044.

AMUEDO-DORANTES, C., AND S. POZO (2006): "Remittances as insurance: evidence from Mexican immigrants," *Journal of Population Economics*, 19(2), 227–254.

BETTIN, G., R. LUCCHETTI, AND A. ZAZZARO (2009): "Income, consumption and remittances: evidence from immigrants to Australia," Mo.Fi.R. Working Papers 34, Money and Finance Research group (Mo.Fi.R.) - Univ. Politecnica Marche - Dept. Economics.

BROWN, R. P. (1997): "Estimating remittance functions for Pacific island migrants," *World Development*, 25(4), 613–626.

CHESHER, A. (2007): "Endogeneity and discrete outcomes," Cemmap working papers, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

CLARK, K., AND S. DRINKWATER (2007): "An investigation of household remittance behaviour; evidence from the United Kingdom," *The Manchester School*, 75(6), 717–741.

COX, D., Z. ESER, AND E. JIMENEZ (1998): "Motives for private transfers over the life cycle: an analytical framework and evidence for Peru," *Journal of Development Economics*, 55(1), 57–80.

DAVIDSON, R., AND J. G. MACKINNON (1999): *Econometric theory and methods*. Oxford University Press, Oxford.

FUNKHOUSER, E. (1995): "Remittances from international migration: a comparison of El Salvador and Nicaragua," *The Review of Economics and Statistics*, 77(1), 137–146.

HODDINOTT, J. (1994): "A model of migration and remittances applied to Western Kenya," *Oxford Economic Papers*, 46, 459–476.

HONG, H., AND E. TAMER (2003): "Inference in censored models with endogenous regressors," *Econometrica*, 71(3), 905–932.

KAWAKATSU, H., AND A. G. LARGEY (2009): "EM Algorithms for ordered probit models with endogenous regressors," *Econometrics Journal*, 12(1), 164–186.

LUCAS, R. E., AND O. STARK (1985): "Motivations to remit: evidence from Botswana," *Journal of Political Economy*, 93, 901–918.

MAGNUS, J. (1988): *Linear structures*. Charles Griffin & Co.

MANSKI, C. F., AND E. TAMER (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519–546.

MARIANO, R. S., AND T. SAWA (1972): "The exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables," *Journal of the American Statistical Association*, 67, 159–163.

NEUMANN, G., N. OLITSKY, AND S. ROBBINS (2009): "Job congruence, academic achievement, and earnings," *Labour Economics*, 16(5), 503–509.

RAPOPORT, H., AND F. DOCQUIER (2005): "The economics of migrants' remittances," IZA Discussion Papers 1531, Institute for the Study of Labor (IZA).

RIVERS, D., AND Q. H. VUONG (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

STEWART, M. B. (1983): “On least squares estimation when the dependent variable is grouped,” *Review of Economic Studies*, 50(3), 737–753.

TERZA, J. V., D. S. KENKEL, T.-F. LIN, AND S. SAKATA (2008): “Care-giver advice as a preventive measure for drinking during pregnancy: zeros, categorical outcome responses, and endogeneity,” *Health Economics*, 17(1), 41–54.

VUONG, Q. H. (1989): “Likelihood ratio tests for model selection and non-nested hypotheses,” *Econometrica*, 57(2), 307–33.

WOOLDRIDGE, J. M. (2002): *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, Mass.

# A Analytical score for the ML estimator

In the maximum likelihood estimation, the analytical score can be computed separately for the conditional component $\ell_i^C$ and the marginal component $\ell_i^M$.

In order to optimise the numerical search for the ML estimator, it is useful to reparametrise equations (2) and (11) by re-expressing the covariance matrix of $[\varepsilon_i, u_i]'$ as

$$\begin{bmatrix} \exp(\kappa)^2 + \psi'\psi & \psi' C^{-1} \\ (C')^{-1}\psi & (CC')^{-1} \end{bmatrix}$$

where $C$ is the Cholesky factorisation of $\Sigma^{-1}$, $\psi \equiv C'\theta$ and $\kappa \equiv \ln\sigma$. Upon defining $c \equiv \operatorname{vech} C$, this parametrisation ensures that the covariance matrix above is symmetric and positive definite for any real vector $[\kappa, \psi', c']$.

With the single exception of $\kappa$, $\ell_i^C$ depends on $\Psi$ only through $\hat{y}_i$. Hence, it is useful to define[16]

$$\mu_i \equiv \frac{\partial \ell_i^C}{\partial \hat{y}_i} = \frac{1}{\sigma} \frac{\varphi(e_i) - \varphi(E_i)}{\Phi(E_i) - \Phi(e_i)} \tag{14}$$

and consider that

$$\hat{u}_t = Y_i - \Pi W_i = Y_i - \left(W_i' \otimes I_m\right)\pi,$$

[16]It may be interesting to note that $\mu_i$ equals the conditional mean of $\varepsilon_i$ divided by its standard deviation.

so that

$$
\begin{aligned}
\frac{\partial \ell_i^C}{\partial \beta} &= \frac{\partial \ell_i^C}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \beta} = \mu_i Y_i' \\
\frac{\partial \ell_i^C}{\partial \gamma} &= \frac{\partial \ell_i^C}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \gamma} = \mu_i X_i' \\
\frac{\partial \ell_i^C}{\partial \pi} &= \frac{\partial \ell_i^C}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \pi} = \mu_i \delta' \frac{\partial \hat{u}_i}{\partial \pi} = -\mu_i \delta' \left(W_i' \otimes I_m\right) = -\mu_i \operatorname{vec}(\delta W_i')' \\
\frac{\partial \ell_i^C}{\partial \psi} &= \frac{\partial \ell_i^C}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \psi} = \mu_i \hat{u}_i' \frac{\partial \delta}{\partial \psi} = \mu_i \hat{u}_i' C = \mu_i \omega_i' \\
\frac{\partial \ell_i^C}{\partial c} &= \frac{\partial \ell_i^C}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial c} = \mu_i \hat{u}_i' \frac{\partial \delta}{\partial c} = \mu_i \hat{u}_i' \left(\psi' \otimes I_m\right) L_m' = \mu_i \operatorname{vech}(\hat{u}_i \psi')'
\end{aligned}
$$

where $L_m$ is a selection matrix[17] such that $\operatorname{vech}(A) = L_m \operatorname{vec}(A)$.

As for $\frac{\partial \ell_i^C}{\partial \kappa}$, we have

$$
\frac{\partial \ell_i^C}{\partial \kappa} = \frac{\varphi(E_i)\frac{\partial E_i}{\partial \kappa} - \varphi(e_i)\frac{\partial e_i}{\partial \kappa}}{\Phi(E_i) - \Phi(e_i)} = \frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)}
$$

The evaluation of the score for the marginal component of the log-likelihood is considerably simpler, since $\ell_i^M$ only depends on $\pi$ (via $\omega_i$) and $c$. We have

$$
\begin{aligned}
\frac{\partial \ell_i^M}{\partial \pi} &= -\omega_i' \frac{\partial \omega_i}{\partial \pi} = -\omega_i' C \frac{\partial \hat{u}_i}{\partial \pi} = \omega_i' C \left(W_i' \otimes I_m\right) = W_i' \otimes \omega_i' C \\
\frac{\partial \ell_i^M}{\partial c} &= \frac{\partial}{\partial c} \sum_{i=1}^{m} \ln C_{ii} - \omega_i' \frac{\partial \omega_i}{\partial c} = \tau' - \omega_i' \left(\hat{u}_i' \otimes I_m\right) L_m' = \tau' - \operatorname{vech}(\omega_i \hat{u}_i')'
\end{aligned}
$$

where $\tau$ is the result of applying the vech operator to a diagonal matrix containing the inverses of the diagonal of $C$.

# B Computation of the covariance matrix for the 2-stage estimator

Here, we consider the model as spelled out in equations 3 and 4, with the only exception that we consider the set of exogenous explanatory variables $X_i$ empty, to keep notation as simple as possible with no loss of generality.

For the two step estimator the second-step objective function is the ordinary interval regression log-likelihood. Denote $\theta$ as the vector of its arguments; clearly,

[17] Technically, it is called the elimination matrix. See (Magnus, 1988), chapter 5.

the second-step loglikelihood also depends on the first-step parameters $\hat{\pi}$ through the first-step residuals.

We need do compute $V(\tilde{\theta})$ through the logic explained, among others, in Wooldridge, section 12.5.2. In short, the idea is based on the first-order expansion

$$s(\tilde{\theta}, \hat{\pi}) = 0 = s(\theta_0, \pi_0) + \ddot{H}(\tilde{\theta} - \theta_0) + \ddot{F}(\hat{\pi} - \pi_0),$$

where $s(\cdot)$ is the total score for the second step, $H$ is the Hessian of the second-step loglikelihood evaluated somewhere between $\tilde{\theta}$ and $\theta_0$, while $F$ is $\frac{\partial s(\theta,\pi)}{\partial \pi}$ evaluated somewhere between $\hat{\pi}$ and $\pi$. This leads to

$$\frac{1}{\sqrt{n}} s(\theta_0, \pi_0) = -\frac{1}{\sqrt{n}} \left[ \ddot{H}(\tilde{\theta} - \theta_0) + \ddot{F}(\hat{\pi} - \pi_0) \right]$$

which is an $O_p(1)$ quantity by the customary score properties. By consistency of $\tilde{\theta}$ and $\hat{\pi}$, $\ddot{H}$ and $\ddot{F}$ can be replaced by consistent estimators $\bar{H}$ and $\bar{F}$ to yield

$$\sqrt{n}(\tilde{\theta} - \theta_0) = -\bar{H}^{-1} \left[ \frac{1}{\sqrt{n}} s(\tilde{\theta}, \hat{\pi}) - \sqrt{n} \bar{F}(\hat{\pi} - \pi_0) \right] + o_p(1)$$

Since $\hat{\pi}$ is simply an OLS estimate, from the theory of extremum estimators we have

$$\sqrt{n}(\hat{\pi} - \pi_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_i + o_p(1),$$

where $r_i = \left[ \left( \frac{W'W}{n} \right)^{-1} W_i \otimes \hat{u}_i \right]$. Hence,

$$\sqrt{n}(\tilde{\theta} - \theta_0) = -\frac{1}{\sqrt{n}} \bar{H}^{-1} \left[ \sum_{i=1}^{n} s_i(\tilde{\theta}, \hat{\pi}) + \bar{F} \sum_{i=1}^{n} r_i \right] + o_p(1)$$

so a consistent estimator of the asymptotic covariance matrix for $\tilde{\theta}$ can be obtained as

$$\widehat{V}(\tilde{\theta}) = \bar{H}^{-1} \left( \sum_{i=1}^{n} g_i g_i' \right) \bar{H}^{-1},$$

where

$$g_i = s_i(\tilde{\theta}, \hat{\pi}) + \bar{F} r_i.$$

In the next subsections, analytical expressions for $H_i$ and $F_i$ will be derived, so estimation of $\bar{F}$ and $\bar{H}$ is simply a matter of computing sample moments.

## B.1 Preliminary definitions

Upon defining

$$\hat{u}_i = Y_i - \Pi W_i = Y_i - (W_i' \otimes I_m)\pi$$

and $\hat{y}_i$, $E_i$ and $e_i$ as in equations (10–9), we have the following partial derivatives:

$$\frac{\partial \hat{u}_i}{\partial \pi} = -(W_i' \otimes I_m)$$

$$\frac{\partial \hat{y}_i}{\partial \beta} = Y_i' \qquad \frac{\partial \hat{y}_i}{\partial \gamma} = \hat{u}_i' \qquad \frac{\partial \hat{y}_i}{\partial \hat{u}_i} = \gamma'$$

$$\frac{\partial E_i}{\partial \hat{y}_i} = \frac{\partial e_i}{\partial \hat{y}_i} = -\frac{1}{\sigma}$$

$$\frac{\partial E_i}{\partial \beta} = \frac{\partial E_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \beta} = -\frac{1}{\sigma}Y_i' \qquad \frac{\partial e_i}{\partial \beta} = \frac{\partial e_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \beta} = -\frac{1}{\sigma}Y_i'$$

$$\frac{\partial E_i}{\partial \gamma} = \frac{\partial E_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \gamma} = -\frac{1}{\sigma}\hat{u}_i' \qquad \frac{\partial e_i}{\partial \gamma} = \frac{\partial e_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \gamma} = -\frac{1}{\sigma}\hat{u}_i'$$

$$\frac{\partial E_i}{\partial \sigma} = -\frac{E_i}{\sigma} \qquad \frac{\partial e_i}{\partial \sigma} = -\frac{e_i}{\sigma}$$

The second-stage loglikelihood is:

$$\ell_i = \ln\left[\Phi(E_i) - \Phi(e_i)\right]$$

## B.2 The score

For a start, define $\mu_i$ as

$$\frac{\partial \ell_i}{\partial \hat{y}_i} = \frac{1}{\Phi(E_i) - \Phi(e_i)}\left[\varphi(E_i)\frac{\partial E_i}{\partial \hat{y}_i} - \varphi(e_i)\frac{\partial e_i}{\partial \hat{y}_i}\right] = \frac{1}{\sigma}\frac{\varphi(e_i) - \varphi(E_i)}{\Phi(E_i) - \Phi(e_i)} \equiv \mu_i;$$

note that $\sigma \cdot \mu_i$ can be interpreted as $E(z|e_i < z < E_i)$ where $z \sim N(0,\sigma)$. Hence, $e_i < \sigma \cdot \mu_i < E_i$.

By a reasoning akin to section A, the score is:

$$s_i^\beta = \frac{\partial \ell_i}{\partial \beta} = \frac{\partial \ell_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \beta} = \mu_i Y_i' \tag{15}$$

$$s_i^\gamma = \frac{\partial \ell_i}{\partial \gamma} = \frac{\partial \ell_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial \gamma} = \mu_i \hat{u}_i' \tag{16}$$

$$\begin{aligned} s_i^\sigma = \frac{\partial \ell_i}{\partial \sigma} &= \frac{1}{\Phi(E_i) - \Phi(e_i)}\left[\varphi(E_i)\frac{\partial E_i}{\partial \sigma} - \varphi(e_i)\frac{\partial e_i}{\partial \sigma}\right] = \\ &= \frac{1}{\sigma}\frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)} \equiv \lambda_i \end{aligned} \tag{17}$$

## B.3 The Hessian

Start from

$$H_i^{\beta\beta} = \frac{\partial s_i^{\beta}}{\partial \beta} = Y_i \frac{\partial \mu_i}{\partial \beta} = Y_i \left[ \frac{\partial \mu_i}{\partial E_i} \frac{\partial E_i}{\partial \beta} + \frac{\partial \mu_i}{\partial e_i} \frac{\partial e_i}{\partial \beta} \right]$$

and

$$\frac{\partial \mu_i}{\partial \beta} = -\frac{1}{\sigma} \left( \frac{\partial \mu_i}{\partial E_i} + \frac{\partial \mu_i}{\partial e_i} \right) Y_i';$$

note that

$$\begin{aligned} \frac{\partial \mu_i}{\partial E_i} &= \frac{1}{\sigma} \frac{1}{[\Phi(E_i) - \Phi(e_i)]^2} \left[ \varphi(E_i) E_i [\Phi(E_i) - \Phi(e_i)] - \varphi(E_i) [\varphi(e_i) - \varphi(E_i)] \right] = \\ &= \frac{1}{\sigma} \frac{\varphi(E_i) E_i - \varphi(E_i) \mu_i}{\Phi(E_i) - \Phi(e_i)} \qquad (18) \\ \frac{\partial \mu_i}{\partial e_i} &= \frac{1}{\sigma} \frac{1}{[\Phi(E_i) - \Phi(e_i)]^2} \left[ -\varphi(e_i) e_i [\Phi(E_i) - \Phi(e_i)] + \varphi(e_i) [\varphi(e_i) - \varphi(E_i)] \right] = \\ &= -\frac{1}{\sigma} \frac{\varphi(e_i) e_i - \varphi(e_i) \mu_i}{\Phi(E_i) - \Phi(e_i)} \qquad (19) \end{aligned}$$

Observe that (18) can also be written as $\frac{1}{\sigma} \frac{\varphi(E_i)}{\Phi(E_i) - \Phi(e_i)} (E_i - \mu_i)$ (should be positive).

Using (18) and (19) yields

$$\frac{\partial \mu_i}{\partial E_i} + \frac{\partial \mu_i}{\partial e_i} = \frac{1}{\sigma} \left\{ \frac{\varphi(E_i) E_i - \varphi(e_i) e_i}{\Phi(E_i) - \Phi(e_i)} + \frac{[\varphi(e_i) - \varphi(E_i)]^2}{[\Phi(E_i) - \Phi(e_i)]^2} \right\} = \sigma \mu_i^2 - \lambda_i \qquad (20)$$

So in the end

$$H_i^{\beta\beta} = -\left( \mu_i^2 - \frac{\lambda_i}{\sigma} \right) Y_i Y_i'$$

By a similar reasoning

$$H_i^{\beta\gamma} = \frac{\partial s_i^{\beta}}{\partial \gamma} = Y_i \frac{\partial \mu_i}{\partial \gamma} = Y_i \left[ \frac{\partial \mu_i}{\partial E_i} \frac{\partial E_i}{\partial \gamma} + \frac{\partial \mu_i}{\partial e_i} \frac{\partial e_i}{\partial \gamma} \right]$$

so

$$\frac{\partial \mu_i}{\partial \gamma} = -\frac{1}{\sigma} \left( \frac{\partial \mu_i}{\partial E_i} + \frac{\partial \mu_i}{\partial e_i} \right) \hat{u}_i'$$

and

$$H_i^{\beta\gamma} = -\left( \mu_i^2 - \frac{\lambda_i}{\sigma} \right) Y_i \hat{u}_i'.$$

So by analogy,

$$H_i^{\gamma\gamma} = -\left(\mu_i^2 - \frac{\lambda_i}{\sigma}\right)\hat{u}_i'\hat{u}_i'.$$

It only remains to evaluate the elements pertaining to $\sigma$:

$$H_i^{\beta\sigma} = \frac{\partial \lambda_i}{\partial \beta} = \frac{\partial \lambda}{\partial E_i}\frac{\partial E_i}{\partial \beta} + \frac{\partial \lambda}{\partial e_i}\frac{\partial e_i}{\partial \beta}$$

but since $\frac{\partial E_i}{\partial \beta} = \frac{\partial e_i}{\partial \beta} = -\frac{1}{\sigma}Y_i'$, the expression above becomes

$$H_i^{\beta\sigma} = -\frac{1}{\sigma}\left[\frac{\partial \lambda_i}{\partial E_i} + \frac{\partial \lambda_i}{\partial e_i}\right]Y_i'$$

so we only need the derivatives of $\lambda_i$ wrt $E_i$ and $e_i$ and their sum:

$$\begin{aligned}\frac{\partial \lambda_i}{\partial E_i} =& \frac{1}{\sigma}\frac{1}{[\Phi(E_i)-\Phi(e_i)]^2}\times \\ &\times\left[\varphi(E_i)(E_i^2-1)[\Phi(E_i)-\Phi(e_i)] - \varphi(E_i)[\varphi(e_i)e_i - \varphi(E_i)E_i]\right] \quad (21)\\ \frac{\partial \lambda_i}{\partial e_i} =& \frac{1}{\sigma}\frac{1}{[\Phi(E_i)-\Phi(e_i)]^2}\times \\ &\times\left[-\varphi(e_i)(e_i^2-1)[\Phi(E_i)-\Phi(e_i)] + \varphi(e_i)[\varphi(e_i)e_i - \varphi(E_i)E_i]\right] \quad (22)\end{aligned}$$

from which

$$\begin{aligned}\frac{\partial \lambda_i}{\partial E_i} + \frac{\partial \lambda_i}{\partial e_i} = \frac{1}{\sigma}\Bigg\{&\frac{\varphi(E_i)(E_i^2-1) - \varphi(e_i)(e_i^2-1)}{\Phi(E_i)-\Phi(e_i)} + \\ &+\frac{[\varphi(e_i)-\varphi(E_i)][\varphi(e_i)e_i - \varphi(E_i)E_i]}{[\Phi(E_i)-\Phi(e_i)]^2}\Bigg\} = \\ =& \sigma\mu_i\lambda_i - \nu_i \quad (23)\end{aligned}$$

where

$$\nu_i \equiv \frac{1}{\sigma}\frac{\varphi(e_i)(e_i^2-1) - \varphi(E_i)(E_i^2-1)}{\Phi(E_i)-\Phi(e_i)};$$

as a consequence,

$$H_i^{\beta\sigma} = \left[\frac{\nu_i}{\sigma} - \mu_i\lambda_i\right]Y_i'; \quad (24)$$

a parallel line of reasoning yields

$$H_i^{\gamma\sigma} = \left[\frac{\nu_i}{\sigma} - \mu_i\lambda_i\right]\hat{u}_i'. \quad (25)$$

The last element of the Hessian is a bit more cumbersome:

$$H_i^{\sigma\sigma} = -\frac{\lambda_i}{\sigma} + \frac{1}{\sigma}\frac{\partial}{\partial\sigma}\frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)};$$

by using the chain rule again,

$$\begin{aligned}
\frac{\partial}{\partial\sigma}\frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)} &= \frac{\partial \frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)}}{\partial E_i}\frac{\partial E_i}{\partial\sigma} + \frac{\partial \frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)}}{\partial e_i}\frac{\partial e_i}{\partial\sigma} = \\
&= -\frac{\partial \frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)}}{\partial E_i}\frac{E_i}{\sigma} - \frac{\partial \frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)}}{\partial e_i}\frac{e_i}{\sigma} = \\
&= -\frac{(\varphi(E_i)(E_i^2-1))[\Phi(E_i) - \Phi(e_i)] - \varphi(E_i)\left[\varphi(e_i)e_i - \varphi(E_i)E_i\right]}{[\Phi(E_i) - \Phi(e_i)]^2}\frac{E_i}{\sigma} + \\
&\quad - \frac{(-\varphi(e_i)(e_i^2-1))[\Phi(E_i) - \Phi(e_i)] + \varphi(e_i)\left[\varphi(e_i)e_i - \varphi(E_i)E_i\right]}{[\Phi(E_i) - \Phi(e_i)]^2}\frac{e_i}{\sigma}
\end{aligned}$$

which can be re-arranged as

$$\begin{aligned}
\frac{\partial}{\partial\sigma}\frac{\varphi(e_i)e_i - \varphi(E_i)E_i}{\Phi(E_i) - \Phi(e_i)} &= -\frac{1}{\sigma}\left\{\frac{\varphi(E_i)(E_i^3 - E_i)}{\Phi(E_i) - \Phi(e_i)} - \frac{\varphi(E_i)E_i\sigma\lambda_i}{\Phi(E_i) - \Phi(e_i)} + \right. \\
&\quad \left. - \frac{\varphi(e_i)(e_i^3 - e_i)}{\Phi(E_i) - \Phi(e_i)} + \frac{\varphi(e_i)e_i\sigma\lambda_i}{\Phi(E_i) - \Phi(e_i)}\right\} = \\
&= -\frac{1}{\sigma}\frac{\varphi(E_i)(E_i^3 - E_i) - \varphi(e_i)(e_i^3 - e_i)}{\Phi(E_i) - \Phi(e_i)} - \sigma\lambda_i^2
\end{aligned}$$

So finally

$$H_i^{\sigma\sigma} = -\frac{\lambda_i}{\sigma} - \frac{1}{\sigma^2}\frac{\varphi(E_i)(E_i^3 - E_i) - \varphi(e_i)(e_i^3 - e_i)}{\Phi(E_i) - \Phi(e_i)} - \lambda_i^2 \qquad (26)$$

## B.4 The $F$ matrix

By using the results of the previous section, the calculation of $F$ is relatively simple: the first results we need are

$$\begin{aligned}
\frac{\partial \hat{y}_i}{\partial \pi} &= \gamma' \frac{\partial \hat{u}_i}{\partial \pi} = -\gamma'(W_i' \otimes I_m) = -[\mathrm{vec}(\gamma W_i')]' = -(W_i' \otimes \gamma') \\
\frac{\partial E_i}{\partial \pi} &= \frac{\partial e_i}{\partial \pi} = -\frac{1}{\sigma}\frac{\partial \hat{y}_i}{\partial \pi} \\
\frac{\partial \mu_i}{\partial \pi} &= \frac{\partial \mu_i}{\partial E_i}\frac{\partial E_i}{\partial \pi} + \frac{\partial \mu_i}{\partial e_i}\frac{\partial e_i}{\partial \pi}.
\end{aligned}$$

so

$$\frac{\partial \mu_i}{\partial \pi} = \frac{1}{\sigma}\left(\frac{\partial \mu_i}{\partial E_i} + \frac{\partial \mu_i}{\partial e_i}\right)\gamma'(W_i' \otimes I_m)$$

and using (20) again we have

$$\frac{\partial \mu_i}{\partial \pi} = \left(\mu_i^2 - \frac{\lambda_i}{\sigma}\right)\gamma'(W_i' \otimes I_m)$$

Moreover,

$$\frac{\partial \lambda_i}{\partial \pi} = \frac{\partial \lambda_i}{\partial E_i}\frac{\partial E_i}{\partial \pi} + \frac{\partial \lambda_i}{\partial e_i}\frac{\partial e_i}{\partial \pi} = -\frac{1}{\sigma}\left[\frac{\partial \lambda_i}{\partial E_i} + \frac{\partial \lambda_i}{\partial e_i}\right]\gamma'(W_i' \otimes I_m)$$

so using (23) again, we have

$$\frac{\partial \lambda_i}{\partial \pi} = \left(\frac{\mu_i}{\sigma} - \mu_i\lambda_i\right)\gamma'(W_i' \otimes I_m)$$

In order to evaluate $F$, we need the derivatives of the score (15)–(17) with respect to $\pi$. Hence:

$$\frac{\partial s_i^\beta}{\partial \pi} = Y_i \frac{\partial \mu_i}{\partial \pi} = Y_i\left(\mu_i^2 - \frac{\lambda_i}{\sigma}\right)(W_i' \otimes \gamma') \tag{27}$$

$$\begin{aligned}\frac{\partial s_i^\gamma}{\partial \pi} &= \mu_i \frac{\partial \hat{u}_i}{\partial \pi} + \hat{u}_i \frac{\partial \mu_i}{\partial \pi} = -\mu_i(W_i' \otimes I_m) + \hat{u}_i\left(\mu_i^2 - \frac{\lambda_i}{\sigma}\right)\gamma'(W_i' \otimes I_m) \\ &= \hat{u}_i\left(\mu_i^2 - \frac{\lambda_i}{\sigma}\right)(W_i' \otimes \gamma') - \mu_i(W_i' \otimes I_m)\end{aligned} \tag{28}$$

$$\frac{\partial s_i^\sigma}{\partial \pi} = \left(\frac{\mu_i}{\sigma} - \mu_i\lambda_i\right)(W_i' \otimes \gamma') \tag{29}$$