**Modeling and predicting IBNR reserve:
extended chain ladder and heteroscedastic regression
analysis**

Leonardo Costa

Institute of Applied Social Sciences
Alfenas Federal University (UNIFAL)

Adrian Pizzinga
(corresponding author)

Department of Statistics
Institute of Mathematics and Statistics
Fluminense Federal University (UFF)

Address:

Rua Mario Santos Braga s/n
Instituto de Matemática e Estatística – 7º andar – Centro
24020-140 – Niterói – RJ

Phone number:

55-21-2629-2099

E-mails:

adrianhpster@gmail.com

adrianhp@est.uff.br

Rodrigo Atherino

JGP Global Gestao de Recursos

6.1 FIRST EXAMPLE: T1 TRIANGLE (EXTENDED VERSION)

The first runoff triangle considered here (T1 triangle, from now on) had been previously studied by Taylor & Ashe (1983) and Verrall (1989). It is presented in Table 8. The data unit is millions of dollars.

Accident year	Development									
	0	1	2	3	4	5	6	7	8	9
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
3	290507	1001799	926219	1016654	750816	146923	495992	280405		
4	310608	1108250	776189	1562400	272482	352053	206286			
5	443160	693190	991983	769488	504851	470639				
6	396132	937085	847498	805037	705960					
7	440832	847631	1131398	1063269						
8	359480	1061648	1443370							
9	376686	986608								
10	344014									

Table 8. T1 triangle: Values of IBNR payments (in millions of dollars).

The scatter plots in Figure 3 suggest that the column effect shows an increasing pattern from the first until the third, maybe fourth development year, and then begins to decay until the years near the tail. Also, some degree of heteroscedasticity is noticeable, as there is less variability along the first column and those situated near the tail. The row effect seems to be present somehow, but its significance need be tested.

Figure 3. Scatter plot of T1 triangle data against column and row effects.

The application of the chain ladder consists basically on direct use of expressions discussed in section 4. The resulting goodness-of-fit statistics are disposed in Table 9. The in-sample measures do not consider the very first column of the T1 triangle in their computations, as the chain ladder does not allow in-sample estimation of the corresponding data, and the out-of-sample measures use the diagonal corresponding to the last calendar year of the triangle, except for its extremes $C_{10,0} = 344014$ and $C_{1,9} = 67948$.

Table 9. Analytical goodness-of-fit measures regarding the chain ladder for the T1 triangle¹.

Measure	Observed value
In-sample Pseudo R ²	0.708
In-sample MAPE (%)	33.2783
In sample MSE	91,548,197,330.306
Out-of-sample Pseudo R ²	0.797
Out-of-sample MAPE (%)	32.844
Out-of-sample MSE	43,010,951,802,505

¹The MSE are *sample* mean squared errors and should not be confused with the theoretical expressions of section 4.

The estimated reserves and corresponding conditional MSE are in Table 10. The chain ladder is estimating, for instance, that the insurance company has to pay US\$ 5,433,718.81 for accidents occurred in the second year of origin. Likewise, the amount to be paid in the next calendar year is being estimated, by eq.(7) of section 4, to be US\$ 9,509,080.76.

Table 10. Estimated IBNR reserves from the chain ladder and corresponding conditional mean square errors for the T1 triangle.

Accident year	Estimated Reserve	Conditional MSE
2	5,433,718.81	5,705,542,382.23
3	5,754,659.78	14,810,539,906.56
4	5,915,275.15	17,835,296,140.84
5	5,758,477.72	68,333,331,757.95
6	6,441,614.59	168,928,976,684.39
7	7,739,823.08	311,717,714,006.48
8	10,586,935.76	766,198,253,108.93
9	5,642,266.26	943,341,726,628.80
10	9,822,972.64	1,858,191,313,379.91
Calendar year Total	9,509,080.76	372,143,030,771.86
	66,962,558.31	5,988,273,257,923.43

Now, the focus is on the regression analysis conducted with the T1 triangle in order to estimate IBNR reserves. The very first step is to estimate model (9) using OLS. Since there is no additional clue about the column and row effects in the tail, it has been supposed that the unobserved values to be estimated on the first year of the tail on the left of the T1 triangle has the same column effect of the last development year right before this “eastern” tail; and, for the value of the first accident year with no delays ($j = 0$) of the tail on the bottom of the T1 triangle, the row effect of the last accident year is assumed. The remaining cells of the tail effects (see Table 3) are assumed null.

Figure 4 indicates heteroscedasticity for the residuals that seems to be closely related to the columns of the T1 triangle. Following this, Breush-Pagan heteroscedasticity tests (cf. Wooldridge, 2000, ch.8; and Johnston & DiNardo, 1997, ch.6) have been applied using column aggregations as explanatory variables. The first test considered two groups (columns 1-3 and 4-10) and, the second, three groups (columns 1-3, 4-6 and 7-10). As can be seen in the third column of Table 11, which also shows in-sample and out-of-sample goodness-of-fit measures, the null hypothesis of homocedasticity is being reject for at least one of the tests at a 5% significance level.

Figure 4. Scatter plot of OLS residuals against row and column factors for the T1 triangle.

The heteroscedastic model suggested in Corollaries 1(a) and 2 was estimated by feasible GLS. The feasible covariance matrix was diagonal with elements given by the mean of the squared residuals coming from a former OLS estimation, adapting the formula for $\hat{\sigma}_i^2$ in Theorem 4 to the runoff triangle framework. This model fitted the data very adequately (in terms of basic assumptions), since Figures 5 clearly indicates that residual variance estimation has been quite appropriately done – indeed: one can hardly see heteroscedastic patterns in the standardized residuals – and Figure 6 provides no significant evidence against the normality assumption. Moreover, the second column of Table 11 has information that supports predictive capability – see the first six lines of that table – and model basic assumptions: (i) the Breush-Pagan tests and mean and variance for the standardized residuals reiterate graphical conclusion from Figure 5; and (ii) three normality tests agree with Figure 6.

One can also see that OLS and feasible GLS estimations are at least slightly superior to the chain ladder in terms of IBNR estimation (compare Table 9 with the first six lines of Table 11). Also, one can conclude that the OLS results compare to feasible GLS from this same predictive ability standpoint. However, the latter have shown to be most adequate in terms of basic assumptions than the former, something in tune with statistical inference procedures, such as significance tests for row and column effects. The results of F tests are displayed in Table 11. Corroborating previous insights given by Figure 3, both column and row effects are statistically significant even at a 1% significance level. Additionally, Akaike and Bayesian information criteria suggest superiority of the heteroscedastic model as compared with the homocedastic one (see last two rows of Table 11).

Figure 5. Scatter plot of feasible GLS standardized residuals against row and column factors for the T1 triangle.

Figure 6. Normal Q-Q plot with the feasible GLS standardized residuals for the T1 triangle.

Table 11. Analytical information about the least squares estimation for the T1 triangle (p-values in parentheses).

Criterion	Results from feasible GLS	Results from OLS
In-sample Pseudo R ²	0.791	0.795
In-sample MAPE (%)	22.974	23.842
In sample MSE	25,559,509,634.744	25,087,615,884.871
Out-of-sample Pseudo R ²	0.805	0.768
Out-of-sample MAPE (%)	32.139	31.344
Out-of-sample MSE	39,711,816,450.333	48,484,637,352.882
Breush-Pagan test (2 groups)	0.023 (0.878)	2.795 (0.100)
Breush-Pagan test (3 groups)	0.060 (0.942)	4.566 (0.015)
Standardized residual mean	1.0×10^{-16}	1.78×10^{-12}
Standardized residual variance	1.452	-
Anderson-Darling normality test	0.316 (0.533)	-
Jarque-Bera normality test	1.644 (0.440)	-
Shapiro-Wilk normality test	0.974 (0.262)	-
F test – row effect	10.244 (0.000)	-
F test – column effect	31.341 (0.000)	-
AIC	1,474.648	1,512.084
BIC	1,514.800	1,552.231

Table 12 offers complimentary information regarding serial dependence/correlation investigation under four different data ordering of the runoff triangle: row-wise, column-wise, diagonal-wise “from the bottom to the top” (*diagonal 1*, hereafter), and diagonal-wise “from the top to the bottom” (*diagonal 2*, hereafter). This table reports Ljung-Box tests for the original scale standardized residuals and for their squares and, as can be seen, there is no statistical evidence from the data suggesting serial dependence under the first three ordering here considered; and, albeit the Ljung-Box test for residuals in their original scale is slightly significant at 10% significance level under the diagonal 2 ordering, this might have been a result from the very dependence of the residuals rather than a symptom of possible misbehavior of the actual unobserved error terms.

Table 12. Serial dependence tests with the feasible GLS standardized residuals under four different data ordering for the T1 triangle (p-values in parentheses).

Measure	Ordering			
	Row-wise	Column-wise	Diagonal 1	Diagonal 2
Ljung-Box test (levels)	9.327 (0.860)	22.870 (0.087)	18.762 (0.225)	27.115 (0.028)
Ljung-Box test (squares)	5.132 (0.991)	9.711 (0.838)	10.288 (0.801)	9.016 (0.877)

Finally, Table 13 shows the predicted IBNR reserves using OLS and feasible GLS expressions of section 5. The first point worth noticing is that the mean squared errors corresponding to the OLS estimation and using the White matrix (cf. Theorem 4(e)) are

not always larger than their counterparts corresponding to the feasible GLS estimation (cf. Theorem 4(d)), as one can easily see by comparing the third and fifth columns of Table 13. Even though contradicting the fact that, theoretically, the GLS are minimum variance IBNR estimators, the natural explanation is that the feasible expressions for such accuracy measures are subject to sample variation. Secondly, the reader is advised to note that the column corresponding to the mean squared error of the OLS estimation without any heteroscedasticity correction device (fourth column) is clearly wrong and should neither be interpreted nor used – recall: the homocedastic model has been already discarded by the data. The reason for showing this misleading information here is to illustrate how determinant the White correction would be under heteroscedastic scenarios. Focusing a bit more in this point, Greene (2003), ch. 11, and Jonhston & DiNardo (1997), ch. 6, agree that the White covariance matrix's good asymptotic properties might not work properly with small data sets. To alleviate such undesirable finite sample effects, corrections for the White matrix suggested by the latter authors have been considered in this paper and might provide more confidence to the mean squared error calculations. Each of these corrections displayed in the last three columns of Table 13 simply consists of multiplying the White matrix by some inflating constant. For instance, the seventh column of Table 8 gives the feasible OLS mean squared errors using a White matrix previously multiplied by $\frac{n}{n-p}$, where n is the number of observation in the runoff triangle and p is the number of coefficients used in the model. By their turn, the last two columns used $h_{ii} = x_i'(X'X)^{-1}x_i$ and h_{ii}^2 as inflating constants, respectively.

Table 13. Estimated IBNR reserves from the regression model
and corresponding mean squared errors for the T1 triangle.

Accident year	Estimated reserves GLS	Estimated reserves OLS	Mean squared errors					
			GLS	OLS	White	White (n-p)	White (h_{ii})	White (h_{ii}^2)
1	67948.00	67948.00	466084549.93	75262847654.61	233042274.97	233042274.97	233042274.97	233042274.97
2	463929.96	470467.10	2376717412.06	259238697477.00	19782139356.35	29976723837.51	26987602784.33	21334310751.25
3	859896.39	872751.20	5467322959.06	354101245041.67	47418888698.82	72076540798.95	65593399909.20	51430413859.98
4	974523.17	1133913.30	60898788825.17	462515585115.58	209780234768.21	319978579137.97	289532820377.57	227140446213.91
5	1225265.26	1235873.20	129056771170.34	598033510207.97	158118385827.42	232147157947.78	213052407334.15	170269379844.29
6	1925644.15	1804699.40	232930850036.82	780169601532.13	238879952554.19	334088903211.66	315350726778.20	256576143446.91
7	2728648.36	2955041.00	348383016219.32	1045784734713.21	298011586569.90	409042275634.08	388965810175.98	319373742637.66
8	4664073.76	5231739.80	521093464060.46	1479442095008.84	1155785822682.66	1679127159435.17	1826337098355.66	1349542755309.84
9	4775343.11	4889991.20	722111343898.07	2333205023090.88	345691675812.66	430207960561.85	443004139996.23	374007277019.31
10	5117725.24	5226297.70	1407858013639.93	4862872958148.74	397888856235.93	504233341817.04	487776147414.99	419807250680.67
Calendar year	5857001.22	5968842.20	270431594102.66	743967275823.98	272440152896.70	307111946879.92	305003261712.93	280742580977.62
Total	23147011.41	24232735.90	5536158851171.19	31361013166558.60	6240359332563.92	9152378062267.45	8747155206895.64	6856322558098.04

As regards the IBNR reserves themselves, observe that the estimates from OLS and feasible GLS are both smaller than those obtained by the chain ladder (compare Tables 10 and 13). Since previous analysis indicates moderate evidence against the chain ladder modeling framework as the “true” IBNR data generator, in a hypothesized situation where the chain ladder has been used for estimating IBNR reserves with the T1 triangle, the insurance company could have probably overtaxed their clients as a consequence from overestimation of IBNR reserves – this might have limited the insurance company to be less competitive in the insurance market.

APPENDIX 1: PROOF OF LEMMA 1

Take an arbitrary $A \times B \in \underline{B}^{n_i} \times \underline{B}^{n_k}$.

Then, $((Y_i, Y_k) \in A \times B) = (Y_i \in A, Y_k \in B) = (Y_i \in A) \cap (Y_k \in B)$. It follows from the very definition of conditional covariance and from the identity above that

$$\begin{aligned}
 & E\left(\text{Cov}(C_{i,J-i+1}, C_{k,J-k+1} \mid \mathfrak{T}_{i,k}) I_{((Y_i, Y_k) \in A \times B)}\right) \\
 &= E\left\{E\left[\left(C_{i,J-i+1} - E(C_{i,J-i+1} \mid \mathfrak{T}_{i,k})\right)\left(C_{k,J-k+1} - E(C_{k,J-k+1} \mid \mathfrak{T}_{i,k})\right) \mid \mathfrak{T}_{i,k}\right] I_{(Y_i \in A) \cap (Y_k \in B)}\right\} \quad (\clubsuit) \\
 &= E\left[\left(C_{i,J-i+1} - E(C_{i,J-i+1} \mid \mathfrak{T}_i)\right)\left(C_{k,J-k+1} - E(C_{k,J-k+1} \mid \mathfrak{T}_k)\right) I_{(Y_i \in A)} I_{(Y_k \in B)}\right] \\
 &= E\left[\left(C_{i,J-i+1} - E(C_{i,J-i+1} \mid \mathfrak{T}_i)\right) I_{(Y_i \in A)}\right] E\left[\left(C_{k,J-k+1} - E(C_{k,J-k+1} \mid \mathfrak{T}_k)\right) I_{(Y_k \in B)}\right] = 0.
 \end{aligned}$$

The second identity in $(\clubsuit)$ follows from the definition of conditional expectation given $\mathfrak{T}_{i,k}$ and from Shao (2003), Proposition 1.11 (which is applicable, because the rows of the runoff triangle are independent – cf. M2), and the third identity is a consequence of independence between the rows of the runoff triangle (cf. M2 again), noting that $E(C_{i,J-i+1} \mid \mathfrak{T}_i)$ and $I_{(Y_i \in A)}$ are $\mathfrak{T}_i$ -measurable, like $E(C_{k,J-k+1} \mid \mathfrak{T}_k)$ and $I_{(Y_k \in B)}$ are $\mathfrak{T}_k$ -measurable.

On the other hand, $\underline{B}^{n_i} \times \underline{B}^{n_k}$ is a π -system on $\mathfrak{R}^{n_i+n_k}$, which is sufficient to $(Y_i, Y_k)^{-1}(\underline{B}^{n_i} \times \underline{B}^{n_k})$ be a π -system itself on the primitive sample space Ω . Besides,

$$\sigma(Y_i, Y_k) = (Y_i, Y_k)^{-1}(\underline{B}^{n_i+n_k}) = (Y_i, Y_k)^{-1}(\sigma(\underline{B}^{n_i} \times \underline{B}^{n_k})) = \sigma((Y_i, Y_k)^{-1}(\underline{B}^{n_i} \times \underline{B}^{n_k})),$$

where the first and second identities are guaranteed by Billingsley (1995), Theorem 20.1 and Example 18.1 respectively, and the third identity follows from Shao (2003), ch.1, Exercise 13. Combining these latter facts with $(\clubsuit)$, it follows from Billingsley (1995), Theorem 16.10, that eq.(6) is valid P -almost surely. *Q.E.D.*

APPENDIX 2: PROOF OF THEOREM 1

From the definition of Y^d , it follows that

$$\begin{aligned} E(Y^d | \mathfrak{I}_\Delta) &= \sum_{k=2}^J E(C_{k,J-k+1} | \mathfrak{I}_\Delta) = \sum_{k=2}^J E(C_{k,J-k+1} | \mathfrak{I}_k) = \sum_{k=2}^J E(D_{k,J-k+1} - D_{k,J-k} | \mathfrak{I}_k) \\ &= \sum_{k=2}^J (D_{k,J-k+1} f_{J-k} - D_{k,J-k}), \end{aligned}$$

where the second identity comes from Shao (2003), Proposition 1.11 (the rows of the runoff triangle are independent – cf. M2) and the last identity is a consequence of M1 and of $D_{k,J-k}$ being $\mathfrak{I}_k$ -measurable. This proves formula (i). Regarding (ii), it follows from the definition of conditional mean squared error given in the beginning of section 4 that

$$\begin{aligned} MSE(\hat{Y}^d | \mathfrak{I}_\Delta) &= Var(Y^d | \mathfrak{I}_\Delta) + (E(Y^d | \mathfrak{I}_\Delta) - \hat{Y}^d)^2 \\ &= \sum_{k=2}^J Var(C_{k,J-k+1} | \mathfrak{I}_\Delta) + 2 \sum_{i < k} Cov(C_{i,J-i+1}, C_{k,J-k+1} | \mathfrak{I}_\Delta) \\ &= \sum_{k=2}^J Var(C_{k,J-k+1} | \mathfrak{I}_k) + 2 \sum_{i < k} Cov(C_{i,J-i+1}, C_{k,J-k+1} | \mathfrak{I}_{i,k}) \\ &= \sum_{k=2}^J Var(D_{k,J-k+1} - D_{k,J-k} | \mathfrak{I}_k) = \sum_{k=2}^J D_{k,J-k} \alpha_{J-k}^{(cl)2}, \end{aligned}$$

where the third identity is another application of Proposition 1.11 of Shao (2003), the fourth identity is guaranteed by Lemma 1, and the last identity follows from M3 and the $\mathfrak{I}_k$ -measurability of $D_{k,J-k}$. *Q.E.D.*

APPENDIX 3: PROOF OF THEOREM 2

Firstly, observe that

$$E(Y^s - \hat{Y}^s) = E\left(\sum_{j=1}^h Y_{i_j} - \sum_{j=1}^h \hat{Y}_{i_j}\right) = \sum_{j=1}^h E(Y_{i_j}) - \sum_{j=1}^h E(\hat{Y}_{i_j}) = 0 \quad (\text{I})$$

Therefore, from (I),

$$\text{Var}(Y^s - \hat{Y}^s) = E\left[(Y^s - \hat{Y}^s)^2\right] = EQM(\hat{Y}^s).$$

In view of this, it follows that

$$\begin{aligned} MSE(\hat{Y}^s) &= \text{Var}\left(\sum_{j=1}^h Y_{i_j} - \sum_{j=1}^h \hat{Y}_{i_j}\right) = \text{Var}\left(\sum_{j=1}^h (x_{i_j}' \beta + \varepsilon_{i_j}) - \sum_{j=1}^h x_{i_j}' \hat{\beta}\right) \\ &= \text{Var}\left(\sum_{j=1}^h x_{i_j}' (\beta - \hat{\beta}) + \Pi' \varepsilon\right) \\ &= \text{Var}\left(\sum_{j=1}^h x_{i_j}' (\beta - \hat{\beta})\right) + \text{Var}(\Pi' \varepsilon) + 2\text{Cov}\left(\sum_{j=1}^h x_{i_j}' (\beta - \hat{\beta}), \Pi' \varepsilon\right). \end{aligned} \quad (\text{II})$$

Solving (II) by parts:

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^h x_{i_j}' (\beta - \hat{\beta})\right) &= \sum_{j=1}^h x_{i_j}' \text{Var}(\beta - \hat{\beta}) \sum_{j=1}^h x_{i_j} = \sum_{j=1}^h x_{i_j}' \text{Var}(\hat{\beta}) \sum_{j=1}^h x_{i_j} \\ &= \sum_{j=1}^h x_{i_j}' (X_o' \Sigma_o^{-1} X_o)^{-1} \sum_{j=1}^h x_{i_j}, \end{aligned}$$

$$\text{Var}(\Pi' \varepsilon) = \Pi' \text{Var}(\varepsilon) \Pi = \Pi' \Sigma \Pi,$$

$$\begin{aligned} \text{Cov}\left(\sum_{j=1}^h x_{i_j}' (\beta - \hat{\beta}), \Pi' \varepsilon\right) &= \text{Cov}\left(\sum_{j=1}^h x_{i_j}' \beta - \sum_{j=1}^h x_{i_j}' \hat{\beta}, \Pi' \varepsilon\right) \\ &= \text{Cov}\left(-\sum_{j=1}^h x_{i_j}' \hat{\beta}, \Pi' \varepsilon\right) = -\sum_{j=1}^h x_{i_j}' \text{Cov}(\hat{\beta}, \varepsilon) \Pi \\ &= -\sum_{j=1}^h x_{i_j}' \text{Cov}\left((X_o' \Sigma_o^{-1} X_o)^{-1} X_o \Sigma_o^{-1} Y_o, \varepsilon\right) \Pi \\ &= -\sum_{j=1}^h x_{i_j}' (X_o' \Sigma_o^{-1} X_o)^{-1} X_o \Sigma_o^{-1} \text{Cov}(Y_o, \varepsilon) \Pi \\ &= -\sum_{j=1}^h x_{i_j}' (X_o' \Sigma_o^{-1} X_o)^{-1} X_o \Sigma_o^{-1} \text{Cov}(X_o \beta + \varepsilon_o, \varepsilon) \Pi \\ &= -\sum_{j=1}^h x_{i_j}' (X_o' \Sigma_o^{-1} X_o)^{-1} X_o \Sigma_o^{-1} \text{Cov}(\varepsilon_o, \varepsilon) \Pi, \\ &= -\sum_{j=1}^h x_{i_j}' (X_o' \Sigma_o^{-1} X_o)^{-1} X_o \Sigma_o^{-1} [\Sigma_o \quad \Sigma_{oA}] \Pi. \end{aligned} \quad Q.E.D.$$

APPENDIX 4: PROOF OF THEOREM 4

a) Take an arbitrary $j \in \{1, 2, \dots, h\}$. From H1, and defining $\bar{\sigma}^2 \equiv (\bar{\sigma}_1^2, \dots, \bar{\sigma}_K^2)'$, it follows that $\sigma_{i_j}^2 = z_{i_j}' \bar{\sigma}^2$, where z_{i_j} is a K -dimension vector whose k th entry is equal to 1 if $\sigma_{i_j}^2 = \bar{\sigma}_k^2$, and the remaining entries are zero. Note that $\hat{\sigma}_i^2$ is the OLS estimator of $\bar{\sigma}_i^2$ from the regression model with the squared $\tilde{\varepsilon}_o \equiv Y_o - X_o \tilde{\beta}_{MQO}$ as the dependent variable and $Z_i = [z_{i1}, \dots, z_{iK}]$ as the independent variables, the latter being formed with dummy variables z_k corresponding to I_k , $k = 1, 2, \dots, K$. Since $\hat{\sigma}_i^2 \xrightarrow{P} \bar{\sigma}_i^2$ for each i (cf. Amemiya, 1977; or Hayashi, 2000, pg. 170) and since $\hat{\sigma}_{i_j} = \hat{\sigma}_i$ for some i (cf. H1), then the result follows from the fact that convergence in probability is preserved by sums.

b) Since $\sum_{j=1}^k x_{i_j}$ does not depend on n_o , it is sufficient to show that

$$\frac{X_o' \hat{\Sigma}_o^{-1} X_o - X_o' \Sigma_o^{-1} X_o}{n_o} \xrightarrow{P} 0. \quad (\text{III})$$

Implementing: From H1, the block corresponding to Y_o of the regression model considered in Theorems 1 and 2 can be re-ordered as

$$\begin{bmatrix} \hat{Y}_{O,1m_1 \times 1} \\ \hat{Y}_{O,2m_2 \times 1} \\ \vdots \\ \hat{Y}_{O,Km_K \times 1} \end{bmatrix} = \begin{bmatrix} \hat{X}_{O,1m_1 \times p} \\ \hat{X}_{O,2m_2 \times p} \\ \vdots \\ \hat{X}_{O,Km_K \times p} \end{bmatrix} \beta + \begin{bmatrix} \hat{\varepsilon}_{O,1m_1 \times 1} \\ \hat{\varepsilon}_{O,2m_2 \times 1} \\ \vdots \\ \hat{\varepsilon}_{O,Km_K \times 1} \end{bmatrix}, \quad \begin{bmatrix} \hat{\varepsilon}_{O,1m_1 \times 1} \\ \hat{\varepsilon}_{O,2m_2 \times 1} \\ \vdots \\ \hat{\varepsilon}_{O,Km_K \times 1} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{\sigma}_1^2 I_{m_1} & 0 & \dots & 0 \\ 0 & \bar{\sigma}_2^2 I_{m_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \bar{\sigma}_K^2 I_{m_K} \end{bmatrix} \right). \quad (\text{IV})$$

Using (IV), it follows that

$$\begin{aligned} X_o' \Sigma_o^{-1} X_o &= \sum_{l=1}^{n_o} \sigma_l^{-2} x_l x_l' = \sum_{l_1=1}^{m_1} \bar{\sigma}_{l_1}^{-2} \hat{x}_{O,l_1} \hat{x}_{O,l_1}' + \dots + \sum_{l_K=1}^{m_K} \bar{\sigma}_{l_K}^{-2} \hat{x}_{O,l_K} \hat{x}_{O,l_K}' \\ &= \bar{\sigma}_1^{-2} \sum_{l_1=1}^{m_1} \hat{x}_{O,l_1} \hat{x}_{O,l_1}' + \dots + \bar{\sigma}_K^{-2} \sum_{l_K=1}^{m_K} \hat{x}_{O,l_K} \hat{x}_{O,l_K}' \\ &= \bar{\sigma}_1^{-2} \hat{X}_{O,1}' \hat{X}_{O,1} + \dots + \bar{\sigma}_K^{-2} \hat{X}_{O,K}' \hat{X}_{O,K} = \sum_{k=1}^K \bar{\sigma}_k^{-2} \hat{X}_{O,k}' \hat{X}_{O,k}. \end{aligned} \quad (\text{V})$$

Analogously,

$$X_o' \hat{\Sigma}_o^{-1} X_o = \sum_{k=1}^K \hat{\sigma}_k^{-2} \hat{X}_{o,k}' \hat{X}_{o,k}. \quad (\text{VI})$$

From (V) and (VI),

$$\begin{aligned} \frac{1}{n_o} (X_o' \hat{\Sigma}_o^{-1} X_o - X_o' \Sigma_o^{-1} X_o) &= \frac{1}{n_o} \sum_{k=1}^K (\hat{\sigma}_k^{-2} - \bar{\sigma}_k^{-2}) \hat{X}_{o,k}' \hat{X}_{o,k} \\ &= \sum_{k=1}^K (\hat{\sigma}_k^{-2} - \bar{\sigma}_k^{-2}) \frac{\hat{X}_{o,k}' \hat{X}_{o,k}}{n_o} \xrightarrow{P} 0, \end{aligned}$$

where the convergence above follows from the fact that, given any $k \in \{1, \dots, K\}$,

$\hat{\sigma}_k^{-2} \xrightarrow{P} \bar{\sigma}_k^{-2}$ (cf. item (a) and the fact that convergence in probability is preserved under continuous functions) and from $\hat{X}_{o,k}' \hat{X}_{o,k} = O(n_o)$ (indeed:

$$0 < \frac{\hat{X}_{o,k}' \hat{X}_{o,k}}{n_o} \leq \frac{\sum_{k=1}^K \hat{X}_{o,k}' \hat{X}_{o,k}}{n_o} = \frac{X_o' X_o}{n_o} \rightarrow Q \text{ whenever } n_o \rightarrow \infty; \text{ see H2}).$$

c) Since X_o only has non-stochastic and uniformly limited regressors, and under the validity of H3, it follows (cf. Woodridge, 2000, ch. 8, pg. 250; and Greene, 2003, ch.10, pg. 199 *apud* White, 1980) that:

$$n_o \left((X_o' X_o)^{-1} (X_o' \tilde{\Sigma}_o X_o) (X_o' X_o)^{-1} - (X_o' X_o)^{-1} (X_o' \Sigma_o X_o) (X_o' X_o)^{-1} \right) \xrightarrow{P} 0.$$

Consequently, the result follows trivially from $\sum_{j=1}^k x_{i_j}$ not depend on n_o .

d) The additional normality hypothesis leads to $\hat{\beta}_{MQG} \sim N(\beta, (X_o' \Sigma_o^{-1} X_o)^{-1})$. Therefore, using Theorem 2, $Y^s - \hat{Y}^s \sim N(0, MSE(\hat{Y}^s))$ for each n_o , or equivalently,

$$\frac{Y^s - \hat{Y}^s}{\sqrt{MSE(\hat{Y}^s)}} \sim N(0, 1), \quad (\text{VII})$$

for each n_o .

Besides, from the re-ordering formerly proposed in the proof of item (b),

$$\begin{aligned}
X_o' \Sigma_o^{-1} \varepsilon_o &= \sum_{l=1}^{n_o} \sigma_l^{-2} x_l \varepsilon_l = \sum_{l_1=1}^{m_1} \hat{\sigma}_{l_1}^{-2} \hat{x}_{o,l_1} \hat{\varepsilon}_{o,l_1} + \dots + \sum_{l_K=1}^{m_K} \hat{\sigma}_{l_K}^{-2} \hat{x}_{o,l_K} \hat{\varepsilon}_{o,l_K} \\
&= \bar{\sigma}_1^{-2} \sum_{l_1=1}^{m_1} \hat{x}_{o,l_1} \hat{\varepsilon}_{o,l_1} + \dots + \bar{\sigma}_K^{-2} \sum_{l_K=1}^{m_K} \hat{x}_{o,l_K} \hat{\varepsilon}_{o,l_K} \\
&= \bar{\sigma}_1^{-2} \hat{X}_{o,1}' \hat{\varepsilon}_{o,1} + \dots + \bar{\sigma}_K^{-2} \hat{X}_{o,K}' \hat{\varepsilon}_{o,K} = \sum_{k=1}^K \bar{\sigma}_k^{-2} \hat{X}_{o,k}' \hat{\varepsilon}_{o,k} ; \tag{VIII}
\end{aligned}$$

Analogously,

$$X_o' \hat{\Sigma}_o^{-1} \varepsilon_o = \sum_{k=1}^K \hat{\sigma}_k^{-2} \hat{X}_{o,k}' \hat{\varepsilon}_{o,k} \tag{IX}$$

Combining (VIII) and (IX),

$$\begin{aligned}
\frac{1}{n_o} (X_o' \hat{\Sigma}_o^{-1} \varepsilon_o - X_o' \Sigma_o^{-1} \varepsilon_o) &= \frac{1}{n_o} \sum_{k=1}^K (\hat{\sigma}_k^{-2} - \bar{\sigma}_k^{-2}) \hat{X}_{o,k}' \hat{\varepsilon}_{o,k} \\
&= \sum_{k=1}^K (\hat{\sigma}_k^{-2} - \bar{\sigma}_k^{-2}) \frac{\hat{X}_{o,k}' \varepsilon_{o,k}}{n_o} \xrightarrow{P} 0, \tag{X}
\end{aligned}$$

since, for each $k = 1, 2, \dots, K$, $\hat{\sigma}_k^{-2} \xrightarrow{P} \bar{\sigma}_k^{-2}$ and $\hat{X}_{o,k}' \hat{\varepsilon}_{o,k} = o_p(n_o)$ (recall that

$$E\left(\frac{1}{n_o} \hat{X}_{o,k}' \hat{\varepsilon}_{o,k}\right) = 0 \text{ and}$$

$$\begin{aligned}
0 &\leq \text{Var}\left(\frac{1}{n_o} \hat{X}_{o,k}' \hat{\varepsilon}_{o,k}\right) \leq \text{Var}\left(\frac{1}{n_o} \sum_{k=1}^K X_{o,k}' \hat{\varepsilon}_{o,k}\right) = \text{Var}\left(\frac{1}{n_o} X_o' \varepsilon_o\right) \\
&= \frac{1}{n_o^2} X_o' \Sigma_o^{-1} X_o = \frac{1}{n_o} \left(\frac{1}{n_o} X_o' \Sigma_o^{-1} X_o\right) \longrightarrow 0 \tilde{Q} = 0 \text{ whenever } n_o \longrightarrow \infty ; \text{ see H2).}
\end{aligned}$$

In view of (III) and (X), it follows from Greene (2003) sec.10.5.2 or Davidson & Mackinnon (2004) sec.7.4 that $\hat{\beta}_{FGLS} - \hat{\beta}_{GLS} \xrightarrow{P} 0$.

Now, write

$$\begin{aligned}
\frac{Y^s - \hat{Y}_{FGLS}^s}{\sqrt{MSE(\hat{Y}_{FGLS}^s)}} &= \left(\frac{Y^s - \hat{Y}_{FGLS}^s}{\sqrt{MSE(\hat{Y}_{FGLS}^s)}} - \frac{Y^s - \hat{Y}_{FGLS}^s}{\sqrt{MSE(\hat{Y}^s)}} \right) + \left(\frac{Y^s - \hat{Y}_{FGLS}^s}{\sqrt{MSE(\hat{Y}^s)}} - \frac{Y^s - \hat{Y}^s}{\sqrt{MSE(\hat{Y}^s)}} \right) + \frac{Y^s - \hat{Y}^s}{\sqrt{MSE(\hat{Y}^s)}} \\
&= (Y^s - \hat{Y}_{FGLS}^s) \frac{\sqrt{MSE(\hat{Y}^s)} - \sqrt{MSE(\hat{Y}_{FGLS}^s)}}{\sqrt{MSE(\hat{Y}_{FGLS}^s)} \sqrt{MSE(\hat{Y}^s)}} + \frac{\hat{Y}^s - \hat{Y}_{FGLS}^s}{\sqrt{MSE(\hat{Y}^s)}} + \frac{Y^s - \hat{Y}^s}{\sqrt{MSE(\hat{Y}^s)}}. \tag{XI}
\end{aligned}$$

Once $\hat{\beta}_{FGLS} \xrightarrow{P} \beta$ (because $\hat{\beta}_{FGLS} - \hat{\beta}_{GLS} \xrightarrow{P} 0$ and, from H2, $\hat{\beta}_{GLS} \xrightarrow{P} \beta$), it

follows that $Y^s - \hat{Y}_{FGLS}^s \xrightarrow{P} \sum_{j=1}^k \varepsilon_{ij}$. Also, $\frac{\sqrt{MSE(\hat{Y}^s)} - \sqrt{MSE(\hat{Y}_{FGLS}^s)}}{\sqrt{MSE(\hat{Y}_{FGLS}^s)MSE(\hat{Y}^s)}} \xrightarrow{P} 0$ (cf. items

(a) and (b) and the fact that convergence in probability is preserved under continuous functions). Then, using Slutsky Theorem (cf. Shao, 2003), the first parcel after the second equality of (XI) converges in probability to zero. Combining these facts with

$$\frac{\hat{Y}^s - \hat{Y}_{FGLS}^s}{\sqrt{MSE(\hat{Y}^s)}} \xrightarrow{P} 0 \text{ (recall again: } \hat{\beta}_{FGLS} - \hat{\beta}_{GLS} \xrightarrow{P} 0 \text{)} \text{ and since } \frac{Y^s - \hat{Y}^s}{\sqrt{MSE(\hat{Y}^s)}} \xrightarrow{d} N(0,1)$$

trivially (cf.(VII)), and applying Slutsky Theorem once more, the result follows.

e) The proof is entirely analogous to that of item (d). The only modifications required are to replace $\hat{Y}_{MQGF}^s$ by $\tilde{Y}^s$ and $EQM(\hat{Y}_{MQGF}^s)$ by $EQM(\tilde{Y}^s)$. Also, recall that $\tilde{\beta} \xrightarrow{P} \beta$ in view of H2.

Q.E.D.

REFERENCES

AMEMIYA, T.. **A Note on a Heteroscedastic Model**. Journal of Econometrics, 6: 365-370, 1977.

BILLINGSLEY, P.. **Probability and Measure**. 3rd edition. New York: Wiley, 1995.

DAVIDSON, R.; MACKINNON, J.. **Econometric Theory and Methods**. New York: Oxford, 2004.

GREENE W. H.. **Econometric analysis**. 5th edition, New York: MacMillan, 2003.

HAYASHI, F.. **Econometrics**. New Jersey: Princeton, 2000.

SHAO, J.. **Mathematical Statistics**. 2nd edition. New York: Springer, 2003.

WHITE, H. A **Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity**. *Econometrica*, 48: 817-838, 1980.

WOOLDRIDGE, J. M. **Introductory econometrics: a modern approach**. 2nd edition.
Australia: South-Western College Publishing, 2000.