Biometrical Letters
Vol. 45 (2008), No. 1, 45-54

Analysis of residuals and adjustment in JRA

A. Oliveira[1], T. Azinheira Oliveira[1,2], João T. Mexia[3]

[1]UAb, Department of Exact and Technologic Sciences, R. Fernão Lopes, 9, 2º D,
1000-132 Lisbon, Portugal, aoliveira@univ-ab.pt, toliveir@univ-ab.pt
[2]CEAUL, Center of Statistics and Applications, University of Lisbon
[3]Department of Mathematics, FCT- New University of Lisbon, Quinta da Torre,
2825 Monte da Caparica, Portugal, jtm@fct.unl.pt

SUMMARY

Joint Regression Analysis (JRA) is based in linear regression applied to yields, adjusting one linear regression per cultivar. The environmental indexes in JRA correspond to a non observable regressor which measures the productivity of the blocks in the field trials. Usually zig-zag algorithm is used in the adjustment. In this algorithm, minimizations for the regression coefficients alternate with those for the environmental indexes. The algorithm has performed very nicely but a general proof of convergence to the absolute minimum of the sum of squares of residues is still lucking. We now present a model for the residues that may be used to validate the adjustments carried out by the zig-zag algorithm.

Key words: Joint Regression Analysis, ziz-zag algorithm, residuals.

1. Introduction

JRA is a well known technique for the comparison of cultivars. This technique is based in the adjustment of linear regressions, one per cultivar; see Mexia et al. (2001). The regression has the yield as dependent variable. The zig-zag algorithm is always applicable, but does not ensure convergence. We presented a linear model for residuals to validate the adjustments. The usual zig-zag algorithm, usually performs well, see Mexia and Pinto (2004), Pereira (2003), Pereira and Mexia (2003), Pereira and Mexia (2003a) and Pinto (2005), but does not guarantee the convergence to the absolute minimum of the square of the residues. We now present a linear model for the residues to validate the adjustment obtained and discuss the significance of the results therein obtained.

2. Zig-zag Algorithm

Joint Regression Analysis rests on the adjustment of linear regression $\alpha_j + \beta_j x,\ j = 1,...,J$, of the yields of the different cultivars on a synthetic measure of productivity: the environmental index.

Let $y_{i,j}$ be the yield of cultivar *j* on block *i* if that cultivar is present in that block, otherwise $y_{i,j}$ may take any value.

With x_i the environmental index for block *i*, we are led to minimize

$$S = \sum_{j=1}^{J} \sum_{i=1}^{b} q_{i,j} \left(y_{i,j} - \alpha_j - \beta_j x_i \right)^2 \tag{2.1}$$

where $q_{i,j} = 1[0]$ when cultivar *j* is present[not present] in block *i*.

We point out that both the regression coefficients (α_j, β_j), $j = 1,...,J$, and the environmental indexes are unknown, so that we have to minimize both for the regression coefficients and the environmental indexes. We thus obtain the corresponding least square estimators.

One technique that has been extensively used to minimize *S* and thus adjust both the regression coefficients and the environmental indexes is the zig-zag algorithm. To apply such algorithm we start by choosing a reasonable set of initial values for the environmental indexes. Usually the block average yields may be used. Alternatively, if we are using an α-design, the super-block average yields can be used as initial values for the environmental indexes of the constituting blocks. Let x' and x'' be the smallest and largest initial values for the environmental indexes.

ZIG-ZAG ALGORITHM

Choose x_1^n

↓

Minimize $S\left(\alpha_J, \beta_J \middle| x_1^n\right) \; : \begin{cases} \tilde{\alpha}(x_1^n) \\ \tilde{\beta}(x_1^n) \end{cases}$

↓

To minimize $S\left(x^n \middle| \tilde{\alpha}_J(x_1^n), \tilde{\beta}_J(x_1^n)\right)$

Minimize the functions $h_i(x), \quad i = 1, \ldots, n$, to get x'^n_1

↓

Standardize taking: $\begin{cases} a_1 = min\{x_{11}, \ldots, x_{n1}\} \\ b_1 = max\{x_{11}, \ldots, x_{n1}\} \\ a'_1 = min\{x'_{11}, \ldots, x'_{n1}\} \\ b'_1 = min\{x'_{11}, \ldots, x'_{n1}\} \end{cases}$

$$x_{i2} = a_1 + \frac{b_1 - a_1}{b'_1 - a'_1}\left(x'_{i1} - a'_1\right)$$

↓

x_2^n

↓

Repeat this procedure until successive sums of sums of squares of weighted residuals differ by less than a fixed amount

Now the zig-zag algorithm is an iterative technique. The name of the technique is following from, in each iteration, minimization being carried first for the regression coefficients and then for the environmental indexes. At each iteration end, the values obtained for environmental indexes are rescaled to keep the range $[x';x'']$ unchanged. The algorithm is above presented.

3. Model for residues

As stated above the zig-zag algorithm usually performs quite well but we still do not have a proof that it converge to the absolute minimum of the goal function.

One may think that if the adjustment was defective so that, with (α_j, β_j) and x_i $[(\tilde{\alpha}_j, \tilde{\beta}_j)$ and $\tilde{x}_i]$ the exact [adjusted] regression coefficients and environmental indexes, we would have

$$\begin{cases} \alpha_j = \tilde{\alpha}_j + \gamma_j \\ \beta_j = \tilde{\beta}_j + \eta_j \;; \quad j = 1,\dots,J \;, \quad i = 1,\dots,b \\ x_i = \tilde{x}_i + u_i \end{cases} \tag{3.1}$$

with not small deviations (γ_j, η_j) and u_i. We are thus led to test the significance of these deviations.

To carry out these tests we consider a model for the regression residuals

$$\overset{\circ}{y}_{i,j} = y_{i,j} - \left(\tilde{\alpha}_j + \tilde{\beta}_j \tilde{x}_i\right) \tag{3.2}$$

Replacing the $(\tilde{\alpha}_j, \tilde{\beta}_j)$ and $\tilde{x}_i$ by there expressions we get

$$\overset{\circ}{y}_{i,j} = \gamma_j + \tilde{x}_i \eta_j + \tilde{\beta}_j u_i + \eta_j u_i \tag{3.3}$$

Assuming that the terms $\eta_j u_i$, may be discarded we may write the model (3.3) in matrix form as

$$\underline{\overset{\circ}{y}} = \overset{\circ\circ}{X}\,\underline{\theta}' + e \tag{3.4}$$

where

$$\begin{cases} \underline{\overset{\circ}{y}} = \left[\underline{\overset{\circ}{y}}'_1 \cdots \underline{\overset{\circ}{y}}'_J \right] \\ \overset{\circ\circ}{X} = \left[I_J \otimes \overset{\circ}{X} \quad \underline{\tilde{\beta}} \otimes I_b \right] \\ \underline{\theta}' = \left(\gamma_1, \eta_1, \ldots, \gamma_j, \eta_j, u_1, \ldots, u_b \right) \end{cases} \tag{3.5}$$

and e represents de vector of error connected with $y_{i,j}$, with the usual assumptions.

We have

$$\overset{\circ}{X} = \begin{bmatrix} 1 & \tilde{x}_1 \\ \vdots & \vdots \\ 1 & \tilde{x}_b \end{bmatrix} \tag{3.6}$$

The least square estimator for $\underline{\theta}$ will be

$$\underline{\tilde{\theta}} = \left(\overset{\circ\circ}{X}' \overset{\circ\circ}{X} \right)^{-1} \overset{\circ\circ}{X}' \underline{\overset{\circ}{y}} \tag{3.7}$$

and the sum of squares of the residues for the adjusted model will be

$$SS_{model} = \overset{\circ}{y}' \overset{\circ}{y} - \overset{\circ}{y}' \overset{\circ\circ}{X} \underline{\tilde{\theta}} \tag{3.8}$$

To evaluate the zig-zag adjustment quality we can use

$$R^2 = 1 - \frac{SS_{model}}{SS_{zig-zag}} \tag{3.9}$$

where $SS_{zig-zag}$ is the sum of squares of the original residues. Contrarily of what is usual, we are interested in low values of R^2. These will indicate that no significant amount of information was extracted by the model adjusted for the residues. This lack of information carried by the residues validates the initially adjusted model, since it, then, would carry the relevant information.

4. Application

This case consists of yield data of oats obtained in experiments carried out, from 1993 to 1997, by the Estação Nacional de Melhoramento de Plantas (the Portuguese plant breeding board), who kindly allowed us to use them. These experiments were designed as randomized blocks, with four blocks and eleven cultivars per experiment. The locations and cultivars were not the same for the different years, being presented in table 1.

Table 1. Locations and years where took place the trials

Location / Experimental station	Year
Elvas (ENMP)	1993,1994,1995,1997
Elvas (Herdade da Comenda)	1994,1996,1997
Benavila	1994
Évora	1993,1994,1995,1996,1997
Beja	1993,1994,1995,1996,1997
Pegões	1993,1995,1996

Applying the zig-zag algorithm for each year and each block (row) we obtained, using (3.2), the adjusted regression coefficients as well as the adjusted environmental indexes. These results are shown in Tables 2 and 3.

Table 2. Adjusted regression coefficients

	1993		1994		1995		1996		1997	
Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$
AE8901	-108.46	1.2702	-	-	-	-	-	-	-	-
AE9004	-137.07	1.2046	208.49	0.9293	-	-	-	-	-	-
AE8902	-37.35	1.0257	-	-	-	-	-	-	-	-
AE9005	-7.76	1.0091	-332.07	1.2472	-	-	-	-	-	-
S.VICENTE	-8.26	0.9877	-	-	-	-	-	-	-	-
AE9002	12.96	0.9822	-228.88	1.1113	67.51	0.9724	-595.26	1.2553	346.49	0.9014
ST ALEIXO	49.60	0.9499	209.49	0.8843	32.11	1.0037	679.48	0.9060	96.87	1.1276
AE9001	59.11	0.9087	-55.15	0.9024	-	-	-	-	-	-
AE9003	67.56	0.8967	22.45	1.0042	-79.20	1.2348	886.51	0.9145	-372.23	1.5035
AE8801	91.26	0.8904	-	-	-	-	-	-	-	-
AVON	60.58	0.8150	96.36	0.9282	73.94	0.9027	128.66	0.9327	252.54	0.8008
AE9303	-	-	-119.61	1.0839	-0.26	0.9928	67.49	0.9702	-91.49	1.0037
AE9101	-	-	-82.66	0.9710	-249.91	1.2102	63.97	0.9470	-84.54	0.9704
AE9302	-	-	175.34	0.9490	-24.32	1.0100	-197.36	1.1164	-10.61	1.0981
AE9301	-	-	171.85	0.9373	130.79	0.9104	835.84	0.7276	361.36	0.6333
AE9401	-	-	-	-	-72.59	1.0525	-1079.9	1.2625	-346.72	1.1831
AE9402	-	-	-	-	68.19	0.8716	-156.39	0.9913	15.03	0.8923
AE9403	-	-	-	-	73.46	0.7130	-268.83	0.8236	31.50	0.8587

Table 3. Adjusted environmental indexes

Year	1993	1994	1995	1996	1997
Number of blocks	16	20	16	16	16
	818.59	3073.93	1502.61	4353.88	1163.74
	681.89	3715.50	1357.96	4787.51	1512.78
	692.94	3793.97	1084.18	4835.70	1544.79
	684.85	3891.24	1347.66	4897.19	1306.06
	512.38	3376.61	1182.01	1976.45	1928.69
	583.62	3290.09	1452.43	3693.10	2061.29
	564.36	3934.17	1513.96	3638.71	2183.35
	606.74	3838.16	1465.87	3573.67	2384.54
	1362.70	4112.04	1394.12	2517.13	813.82
	1298.35	5032.46	1848.48	3087.93	995.58
	1325.97	5004.31	1881.31	3287.93	1090.89
	1539.50	4888.40	1680.99	3516.68	935.58
	134.34	642.85	614.40	1807.90	1450.49
	147.60	949.77	774.51	1988.30	2103.36
	161.45	847.68	389.83	1807.13	2535.30
	169.13	913.36	380.41	1634.20	2836.36
		3289.87			
		2976.09			
		2673.59			
		3492.01			

Of followed we used the environmental indexes to adjust models. In table 4 we present for each block (row) the adjusted corrections for regression coefficients and environmental indexes.

Lastly in table 5 we present the sum of sums of squares for the zig-zag method and for the model for the residues as well as the values of R^2.

Table 4. Adjusted corrections for regression coefficients end environmental indexes

1993			1994			1995			1996			1997		
γ_j	η_j	u_i	γ_j	η_j	u_i	γ_j	η_j	u_i	γ_j	η_j	u_i	γ_j	η_j	u_i
68.63	0.015	17.24	-10.81	0.018	-47.99	-29.66	0.017	-39.74	-45.64	-0.003	17.50	-17.59	-0.010	-70.92
-35.14	0.046	-36.58	-48.11	0.010	-84.80	-13.41	0.001	-90.32	37.54	-0.003	-27.53	49.05	-0.002	17.57
-79.00	0.068	21.82	4.35	-0.002	-178.8	-0.55	-0.002	69.59	35.89	0.002	-12.25	-0.34	0.006	7.89
42.54	0.022	26.34	-40.83	0.004	-123.4	-76.42	0.004	14.65	54.00	-0.002	-13.76	-56.13	0.009	91.54
45.65	-0.024	-37.53	18.98	-0.018	-138.4	48.03	-0.018	-10.13	66.69	0.001	17.73	-74.74	-0.012	24.24
34.25	0.007	-62.97	-3.69	0.008	-20.49	-17.41	0.008	-44.19	10.33	0.688	-86.86	-25.30	0.011	-36.71
47.97	-0.017	17.83	-9.53	-0.008	-41.40	87.05	-0.008	-12.92	-14.12	0.003	-60.85	-12.06	-0.025	-17.82
31.91	-0.013	11.07	-12.36	-0.010	142.48	28.56	-0.010	20.01	13.90	0.011	40.38	78.95	0.002	-10.97
58.91	0.019	22.83	14.27	0.038	49.06	32.97	0.038	-89.17	26.36	0.002	14.00	47.02	-0.028	-0.029
-13.32	-0.046	54.94	-3.49	0.032	185.44	-56.28	0.032	-138.2	69.24	0.005	-22.54	-51.76	0.006	30.42
37.48	-0.083	-29.07	-27.85	-0.008	106.67	26.81	-0.008	-54.77	35.91	-0.001	-30.18	-0.61	0.001	57.00
		31.64			-175.1			-26.81			-17.04			-0.91
		-61.71			31.15			48.53			19.13			-59.80
		85.29			60.38			14.17			-13.26			-29.70
		28.36			30.68			-18.64			61.60			30.12
		-55.25			17.24			-37.70			-0.07			-25.99
					-0.61									
					18.80									
					73.65									
					-88.52									

Table 5. Comparison of the sum of squares of residues

	1993	1994	1995	1996	1997
Sum of squares (zig-zag)	$1.9283{\cdot}10^{6}$	$5.4067{\cdot}10^{7}$	$1.0998{\cdot}10^{7}$	$3.1285{\cdot}10^{7}$	$2.0115{\cdot}10^{7}$
Sum of squares (model)	$1.9283{\cdot}10^{6}$	$5.4067{\cdot}10^{7}$	$1.0998{\cdot}10^{7}$	$3.1285{\cdot}10^{7}$	$2.0104{\cdot}10^{7}$
R^2	$1.4{\cdot}10^{-5}$	$1.8{\cdot}10^{-5}$	$4.9{\cdot}10^{-7}$	$2.6{\cdot}10^{-5}$	$5.5{\cdot}10^{-4}$

5. Conclusion

The low values of R^2 validate the use of the zig-zag algorithm in this case.

REFERENCES

Mexia J.T., Pereira, D. G. and Baeta, J. (2001): Weighted linear joint regression analysis. Biometrical Letters 38: 33–40.

Mexia J.T., Pinto I. (2004): Comparison of wheat cultivars in Portugal (1986-2000). Colloquium Biometryczne 34: 137–145.

Morais, V., Vieira C. (2006): MATLAB 7&6 Curso Completo. FCA.

Pereira D.G.(2003): Análise Conjunta Pesada de Regressões em Redes de Ensaios. Tese de Doutoramento. Universidade de Évora.

Pereira D.G., Mexia J.T. (2003): Reproducibility of Joint Regression Analysis. Colloquium Biometryczne 33: 279–291.

Pereira D.G., Mexia J.T. (2003a): The use of Joint Regression Analysis in selecting recommended cultivars. Biuletyn Oceny Odmian (Cultivar Testing Bulletin) 31: 19–25.

Pinto I. (2005): Análise Conjunta de Regressões e Planos de Melhoramento. Tese de Doutoramento. FCT- Universidade Nova de Lisboa.