Regression problems for magnitudes.

S. Castellaro, F. Mulargia, Y. Y. Kagan

March 1, 2006

Abstract

Least squares linear regression is so popular that it is sometimes applied without checking that its basic requirements are satisfied. In particular, in studying earthquake phenomena, the conditions *a*) that the uncertainty on the independent variable is at least one order of magnitude smaller than the one on the dependent variable, *b*) that both data and uncertainties are normally distributed, and *c*) that residuals are constant are at times disregarded. This may easily lead to wrong results. As an alternative to least-squares, when the ratio between errors on the independent and the dependent variable can be estimated, orthogonal regression can be applied. We test the performance of orthogonal regression in its general form against Gaussian and non-Gaussian data and error distributions and compare it with standard least square regression. General orthogonal regression is found to be superior or equal to the standard least squares in all the cases investigated and its use is recommended. We also compare the performance of orthogonal regression versus standard regression when, as often happens in the literature, the ratio between errors on the independent and the dependent variables cannot be estimated and is arbitrarily set to 1. We apply these results to magnitude scale conversion, which is a common problem in seismology, with important implications in seismic hazard evaluation, and analyze it through specific tests. Our analysis concludes that the commonly used standard regression may induce systematic errors in magnitude conversion as high as 0.3-0.4, and, even more importantly, this can introduce apparent catalogue incompleteness, as well as a heavy bias in estimates of the slope of the frequency-magnitude distributions. All this can

be avoided by using the general orthogonal regression in magnitude conversions.

1 Introduction

Linear least squares fitting is the simplest and most commonly applied technique for establishing a linear (or linearized) functional relation between two variables. However, the mathematical and statistical validity of this method is based on stringent constraints, the most important of which is that the independent variable (x) must be known to a much greater accuracy than the dependent variable (y). It follows that this regression can never be inverted, that is, the regression of y against x cannot be inverted to derive the regression of x against y. This suggests that $y = \alpha + \beta\, x$ should more properly be written as $y \leftarrow \alpha + \beta\, x$.

Linear least squares is applied very frequently, to the point where familiarity may induce users to forget these constraints, and this may lead to wrong results.

Orthogonal regression is a more appropriate technique to deal with least squares problems in which dependent and independent variables are both affected by uncertainty. This technique is analyzed here in its general formulation (GOR) and its performance is compared to that of standard least squares (SR) and inverted standard least squares regression (ISR), that is, the standard regression in which the roles of dependent (y) and independent (y) variables are reversed. SR and ISR are expected to give very different results not only when errors on X and Y are very different but also when errors are similar and large. The general orthogonal regression technique is then applied to a couple of important seismological problems: magnitude conversion and the way this affects the magnitude distribution.

Explanation of notation. In this paper, we use the upper case letters (for example, X, Y) to indicate the true value of a variable and the lower case letters (for example, x, y) to indicate the value of the same variable affected by (measurement) errors.

Measuring the size of seismic events. The problem of magnitude conversion. Estimating the hazard potential of an earthquake implies measurement of the severity of the shaking in the frequency band to which buildings are sensitive, that is, typically, 0.1-10 Hz, while the ultimate effect will depend also on event duration and on local resonances. Energy magnitude (M_e, Choy & Boatwright, 1995) and moment magnitude (M_w), which is based on the scalar seismic moment (M_0) are the most relevant in this respect. The seismic moment

(which is a function of the product of the rupture area and the average displacement) does not tell us much about the relative amount of energy released at the frequencies relevant to the seismic hazard estimates (cfr, e.g., NMSOP, Chapter 3).

Ignoring the point that a single indicator of size may be inadequate in seismic hazard estimates, the state of the art is to use M_w on account of its better definition in seismological terms. Unfortunately a moment-tensor solution can be practically worked out only for sizeable events, so, for the bulk of the events in any seismic catalogue, the size estimate is made on the basis of magnitude or, for the historical part of the catalogue, on the basis of intensity. Local magnitude (M_L), surface wave magnitude (M_s), body wave magnitude (m_b) and duration magnitude (M_d) are the magnitudes most commonly reported in seismic catalogues (see Table 1 for magnitude acronyms). This variety of indicators of earthquake size reflects the specific instrument used (e.g. the Wood-Anderson seismometer to produce M_L data) as well as the properties of the seismic sequence. For example, amplitude saturation, inhibiting the evaluation of M_L or m_b for strong earthquakes, led to the introduction of the duration magnitude M_d, which in turn is not immune from problems, not least, that it lacks a unique definition (see e.g. Lee *et al.*, 1972 for the procedure followed in California and Gasperini, 2002, for the procedure adopted in Italy). Additionally, in seismic swarms or aftershock sequences, the beginning of an event may overlap the end of the previous one, making the M_d estimate impossible or highly inaccurate. Finally, a wide variety of events can lead to systematic errors in the reported magnitudes, such as deviations in the instrumental calibration or changes in the seismic equipment (Habermann, 1986), changes of the agency operating the earthquake recording network, introduction of new software for the analysis, removal or addition of seismograph stations as well as changes in the magnitude definition (Zúñiga and Wyss, 1995). Such systematic errors can be very large, as much as 0.5 magnitude units, as found by Pérez (1999). In this jungle of parameterizations of earthquake size, homogeneous unified databases have always been sought and effective magnitude conversion functions are thus necessary. Standard regression is usually applied in the literature to estimate the value of a random variable (y) corresponding to a certain value of the independent (X) variable. If the X-error is small compared to the Y-error then the Y-estimates are correct.

For uniformity we might wish to adopt M_w, but, given the composition of seismic catalogues, we will need to estimate M_w from measurements of M_L, or some other magnitude, which have a much lower accuracy (Kagan, 2002b; 2003; Table 2). Hence, we are interested in an orthogonal regression in which the error on the variable X is larger than or comparable to the error on variable Y. In this paper we analyze this problem with respect to regression. Finally, we demonstrate

the method in action through an application to Italian seismicity.

Magnitude distribution. The frequency-magnitude distribution of earthquakes (Ishimoto and Iida, 1939, Gutenberg and Richter, 1944), often called the Gutenberg-Richter or G-R relation, is a basic ingredient of seismic hazard estimates and is also used to test the completeness of seismic catalogues. The G-R relation states that

$$\log N(m) = a - bm, \tag{1}$$

where $N(m)$ is the number of events with magnitude $\geq m$ and a, b are constants ($b \approx 1.0$). Tinti and Mulargia (1985) studied the effect of magnitude errors on the G-R relation and found that they do not practically affect the b but only the a value. If we consider earthquakes with magnitude m_z and larger, the probability density function for the G-R equation 1 can be written in the form

$$f(d) = B \exp^{-Bd} \qquad d \geq 0, \tag{2}$$

where $d = m - m_z$ and $B = b \log_e 10 \simeq \log_e 10$ (Utsu, 1999; Kagan, 2005). We now show how the b-value of the frequency-magnitude relation is affected when the magnitudes used are converted from other magnitudes through standard or orthogonal regressions. The G-R b-value will be calculated through the Maximum-Likelihood Method, since neither standard nor orthogonal regression would apply (see Section 4).

2 The magnitude conversion problem

However refined it may be, no analysis can give results of better quality than that of the data it employs. Regarding earthquake size, data are often referred to moment magnitude, M_w, which is linked to the seismic scalar moment M_0 through the relation

$$M_w = \frac{2}{3} \log_{10} M_0 - C \quad [M_0 \text{ in } Nm]. \tag{3}$$

Hanks and Kanamori (1979) suggest $C = 6 + \frac{1}{30}$ but for simplicity of expression we use $C = 6.0$, as in Hanks (1992). Moment magnitude, which can be computed when a moment-tensor solution is available, has a number of advantages over other magnitude definitions: 1) it is a physical parameter of the earthquake, which allows the earthquake process to be quantitatively linked to tectonic deformation (cf. Kagan, 2002a; Bird and Kagan, 2004); 2) it does not saturate for large earthquakes; 3) it allows determination of the moment magnitude with an accuracy 2-3 times higher than with other magnitudes (cf. Kagan, 2002b; 2003). However, we must note that these properties apply only when M_w is estimated directly from the

seismic recording and this is generally true for a very small percentage of events (the larger ones). In all other cases M_w is a value converted from other data, so the benefits listed above do not fully apply.

The standard least square linear regression procedure has so far been the method most commonly used to find the relation between different types of magnitude (see, e.g., Gasperini and Ferrari, 2000; Gasperini, 2002 and Bindi *et al.*, 2005 for the Italian catalogues, Braunmiller et al., 2005 for Switzerland). However, one can argue that this method is misapplied: 1) because both the dependent and the non-dependent variable are affected by similar errors and this contradicts the basic assumption of the standard regression method, 2) because magnitude is not a normally distributed variable. Although other methods exist (see Castellaro and Mulargia, 2006), the most general solution is general orthogonal regression. In this paper we test the performance of this method in the case of magnitude conversion (see Section 3).

We start by summarizing the properties of orthogonal regression in its general formulation (GOR) and then proceed to test its performance through simulations on Gaussian and non-Gaussian data sets and error distributions. Next, we apply it to find the relations between M_w and other magnitudes. Last, we explore the bias induced by the use of the standard regression through a specific non-parametric test. We analyze the performance of the different procedures on simulated data sets in order to study the influence of the regression method. Investigating the practical problem would add several difficulties (threshold difference for both magnitudes, non-linearity of regression curves, spatial and temporal inhomogeneities of earthquake catalogs, and other less known aspects of magnitude determination) which are left to future studies.

The dataset used for these analyses on real data is described in Table 2. It consists of 109 events, recorded in Italy between 1981 and 1996, for which both M_s and M_w magnitudes were estimated, of 121 events for which M_L and M_w magnitudes are available and of 204 events with m_b and M_w. For each magnitude type an estimate of the global standard deviation σ (computed for earthquakes with at least 3 station estimates) is also given (Table 2) and it shows clearly that m_b has the largest standard error.

3 General orthogonal regression

Standard least squares regression assumes that the error on the independent variable (X) is zero and that the error on the dependent variable (Y) is normally distributed and approximately constant over the whole regression domain (cf. Draper and Smith, 1998). This means that a bell-shaped Gaussian curve exists on the y

axis, centered at each 'true' value Y_i, and that each measured value y_i is sampled on such a distribution (Figure 1a). The adjustment from the experimental value y_i to the 'true' value Y_i is thus made along vertical lines as shown in Figure 1b (see also York, 1967).

If an error is also present on the X variable, each measurement is sampled from a two-dimensional normal-distribution centred at the 'true' value (X,Y) and with major and minor axes equal to σ_x and σ_y (Figure 2a). In other words the paths from the experimental (x_i, y_i) to the 'true' (X_i, Y_i) follows lines with slopes which depend on the size of the errors affecting x and y. For constant σ_x and σ_y the slope of the lines from the experimental to the 'true' line is the same (Figure 2b and York, 1967) and, in the case of general orthogonal regression, is the weighted orthogonal distance. The use of standard regression in the latter case is like projecting x_i on the abscissa of the sampled point, which again results in a Gaussian distribution. However, in case of magnitudes, non-linearity in the $m_1 \div m_2$ relation, saturation of magnitude scales, additive and multiplicative noise and a host of other phenomena may result in a distortion of the Gaussian distribution on y. Since the least square estimator of a regression coefficient is vulnerable to gross errors and the related confidence interval is sensitive to non-normality of the parent distribution, it is important to test its validity.

General orthogonal regression is designed to account for the effects of measurement error in predictors (Madansky, 1959; Fuller, 1987; Kendall and Stuart, 1979, chap. 28; Carroll and Ruppert, 1996). Its general lines can be outlined as follows: let us assume that two variables Y and X are linearly related and that their measurement errors ϵ and u are independent normal variates with variances σ_ϵ^2 and σ_u^2 respectively, i.e.

$$y = Y + \epsilon, \tag{4}$$

$$x = X + u, \tag{5}$$

and the regression-like model

$$Y = \alpha + \beta X + \hat{\epsilon}, \tag{6}$$

where $\hat{\epsilon} = \epsilon + u$. Now consider the error variance ratio

$$\eta = \frac{\sigma_\epsilon^2}{\sigma_u^2}, \tag{7}$$

where $\sigma_\epsilon^2 = \sigma_y^2$ and $\sigma_u^2 = \sigma_x^2$, provided that σ_ϵ^2 and σ_u^2 are constants. Orthogonal regression (OR) is often defined as the case in which $\eta = 1$. The general orthogonal regression estimator is obtained by minimizing

$$\sum_{i=1}^{n} \left[\frac{(y_i - \alpha - \beta X_i)^2}{\eta} + (x_i - X_i)^2 \right] \tag{8}$$

in the unknowns, that is α, β, X_i. For $\eta = 1$ we have the squared Euclidean distance of the point (x_i, y_i) from the line $(X_i, \alpha + \beta X_i)$. If $\eta \neq 1$ then equation 8 represents a weighted orthogonal distance.

Let us call s_y^2, s_x^2 and s_{xy} the sample variance of the y, x and the sample covariance between y and x. The general orthogonal estimator of slope is then

$$\hat{\beta} = \frac{s_y^2 - \eta s_x^2 + \sqrt{(s_y^2 - \eta s_x^2)^2 + 4\eta s_{xy}^2}}{2s_{xy}}, \tag{9}$$

and the estimator of the intercept is

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}, \tag{10}$$

where $\overline{y}$ and $\overline{x}$ stand for the average values.

Note that while the standard regression predicts Y from X, since it assumes that no error is present in measuraments of X (i.e. $x = X$), orthogonal regression predicts Y from x and y, as shown by equation 8 above.

As demonstrated in Fuller (1987), in practice, the following formulas can be used to estimate the errors on the regression parameters:

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma_x} s_v + \hat{\sigma_u} s_v - \hat{\sigma}_{uv}^2}{(n-1)\hat{\sigma}_x^2}, \tag{11}$$

and

$$\hat{\sigma}_{\hat{\alpha}}^2 = \frac{s_v}{n} + \overline{x}^2 \hat{\sigma}_{\hat{\beta}}^2, \tag{12}$$

where

$$s_v = \frac{\sum_{i=1}^{n} [Y_i - \overline{y} - \hat{\beta}(X_i - \overline{x})]^2}{(n-2)} = \frac{(n-1)(\eta + \hat{\beta}^2)\hat{\sigma_u}}{(n-2)}, \tag{13}$$

$$\hat{\sigma}_{uv} = -\hat{\beta}\hat{\sigma}_u, \tag{14}$$

and

$$\hat{\sigma}_x = \frac{\sqrt{(s_y^2 - \eta s_x^2)^2 + 4\eta s_{xy}^2} - (s_y^2 - \eta s_x^2)}{2\eta}, \tag{15}$$

$$\hat{\sigma}_u = \frac{s_y^2 + \eta s_x^2 - \sqrt{(s_y^2 - \eta s_x^2) + 4\eta s_{xy}^2}}{2\eta}. \tag{16}$$

Under the assumption that the variations of the observations about the line are normal, that is that the errors are all from the same normal distributions $N(0, \sigma^2)$, it can be shown that the $100(1-\alpha)\%$ confidence limits for β and α are

$$\beta \pm t(n-2, 1 - \frac{1}{2}\alpha)\frac{\hat{\sigma_x} s_v + \hat{\sigma_u} s_v - \hat{\sigma}_{uv}^2}{(n-1)\hat{\sigma}_x^2}, \tag{17}$$

$$\alpha \pm t(n-2, 1-\frac{1}{2}\alpha)\left(\frac{s_v}{n} + \overline{x}^2 \hat{\sigma}_{\hat{\beta}}\right), \tag{18}$$

where $t(n-2, 1-\frac{1}{2}\alpha)$ is the $100(1-\frac{1}{2}\alpha)$ percentage point of a t-distribution with $(n-2)$ degrees of freedom (e.g., Kendall & Stuart, 1979).

3.1 Simulations

In principle, the theoretical likelihood function can be obtained for the main variable (magnitude) in x and y distributed according to the exponential (G-R) law and assuming that errors in both are Gaussian. Wetherill (1986, p. 284) states that if the x-distribution is non-Gaussian, then it is possible in principle to obtain estimates of regression parameters and mentions that Reiersol (1950) considered this problem in general. However, one should not expect that such a theoretical derivation would be of great value in the magnitude regression problem since other issues such as magnitude threshold variation (Katsura and Ogata, 2004) and nonlinearity of magnitude relations blur the picture. The most appropriate technique to deal with all these problems seems to be a simulation (see also Stefanski and Cook, 1995).

In this Section we analyze the performance of general orthogonal regression compared to standard least square regression on:

1. normally, lognormally and exponentially distributed variables,
2. normally, lognormally and exponentially distributed errors,
3. different amount of errors.

The performance of the GOR method is first analyzed versus SR and ISR by studying the distribution curves of the β values (that is the slope coefficients) calculated on 10^3 regressions of 50 different (x, y) sets generated according to some fixed criteria: all the (x, y) sets have been produced starting from $x : y = 1 : 1$ and adding noise to x and/or y according to different criteria. The true result of the regression is then expected to be $\beta = 1$ in all cases. This value is chosen as the most representative of the magnitude regression problem. Large deviations from this should suggest the presence of saturation effects or other problems in the magnitude measurement. Noise is added in 4 ways: we explore the case in which $\eta = 25$, where η is defined as in equation 7, that is when an error on Y 5 times larger than the error on X is added. Second, we explore $\eta = 4$, that is the case in which the errors on Y are twice the size of the errors on X; third $\eta = 1$, the case in which errors on Y and X are comparable in size. Last, we use $\eta = 0.25$, that is the case in which the errors on X are twice the errors on Y. Since the requirements of SR apply strictly only when $\eta >> 1$ and when variables and their

errors are normally distributed, we can expect *a priori* SR to perform well only in the first case ($\eta = 25$) and with variables and errors normally distributed. In the fourth case ISR is expected to perform as GOR and as SR in the case $\eta = 25$, so this case will be presented only in Section 3.1.1.

3.1.1 Normal distributions and errors

When data and errors are normally distributed and the error on Y is much larger than the error on X, SR and GOR have a practically identical performance, as shown in Figure 3a which shows the case of errors on Y 5 times larger than the error on X. The β distribution curve of the ISR, as obvious, performs much worse since error is mostly on the dependent variable.

When the error on Y is twice that on X (Figure 3b), SR regression performs slightly worse than GOR while ISR gives estimates in error by 5%.

When errors on Y are comparable to errors on X (Figure 3c), GOR performs better than SR and ISR. The discrepancy between SR and ISR depends on σ_x and σ_y and increases with data scatter (see Section 3.4 and also Fig.2 and 3 in Bormann and Wylegalla, 1975; and Fig. 3 and 4 in Bormann and Khalturin, 1975).

Last, when errors on Y are half the size of errors on X, GOR performs always well while, as expected, SR does not. This case corresponds to Figure 3a where error size on X and Y has been exchanged. ISR is then expected to perform as SR in case *a* above, that is approximately as GOR.

3.1.2 Exponential distributions and Gaussian errors

We now consider sets of data distributed according to exponential functions and with Gaussian errors. This is a case of interest in seismology since frequency-magnitude is an exponential distribution affected by Gaussian errors (see equations 1 and 2).

As might be anticipated, the performance of the various approaches differs. When errors on Y are 5 times larger than errors on X (Figure 4a) and when errors on Y are twice those on X (Figure 4b), the β distributions obtained with GOR and SR are comparable even if the SR distribution appears to be biased towards smaller values by a few percent.

GOR performance remains good while SR decays with decreasing error on Y ratio over error on X (Figure 4c). We note that ISR also performs poorly in the whole range of η considered.

3.1.3 Exponential distributions and errors

We now consider sets of data distributed according to exponential functions and with exponential errors. We take into account this case since we consider it as worse than exponentially distributed data with normal errors (Section 3.1.2) and we want to see whether GOR continues to perform well under such unfavourable circumstances.

As in Section 3.1.2, when errors on Y are 5 times larger than errors on X (Figure 5a), the β distributions obtained with GOR and SR are comparable even if the SR distribution appears to be biased towards smaller values by a few percent.

GOR performance remains good while SR progressively decays with decreasing error on Y ratio over error on X (Figure 5b, c). We note that ISR also performs poorly in the whole range of η considered.

3.1.4 Lognormal distributions and errors

When used on lognormally distributed data and errors, with errors on Y 5 times larger than error on X, GOR performs better than SR, which appears to be biased to values smaller than real by a few percent (Figure 6a). GOR performs much better than SR and ISR in all the other cases investigated (Figure 6b, c).

3.2 What happens if η is unknown?

It often happens (see Section 3.5) that when the ratio between errors on the Y and X variables is unknown, it is assumed to be equal to unity, that is $\eta = 1$, which formally coincides with the OR assumption. In some cases this assumption is justified and is expected to be not too far from reality. We explore the effects of this assumption when $\eta \neq 1$.

First we considered a set of lognormal data and error (as in Figures 6a-c) with a true $\eta = 0.25$, that is, error on X twice that on Y. When $\eta = 1$ is used instead of $\eta = 0.25$, one still obtains a better performance than SR, as it can be seen in Figure 7a, although in a less evident way than Figures 6a-c, where the true η was applied.

In any case, the use of $\eta = 1$ can lead to wrong results if the true η is very large, as shown in Figure 7b. Here the β coefficients of the regressions have been computed for the OR ($\eta = 1$) instead of the correct value $\eta = 25$ on a set of lognormally distributed data and errors. This case corresponds to that studied in Figure 6a with the correct η.

In light of this and of the considerations in the previous Section, the use of GOR is recommended in all cases, provided that at least an order of magnitude of η is available.

3.3 Does magnitude distribution induce a bias in magnitude-magnitude regressions?

Having established the superiority of GOR with respect to SR in general terms, let us now test the GOR on a typical seismological problem. In order to verify whether the exponential distribution of earthquake magnitudes, which has an upper and lower cutoff (cf. Kagan, 2002a), induces a bias in magnitude vs. magnitude regressions according to different scales, we simulate magnitude-frequency relations, add errors, remove events below the lower threshold and then run standard and orthogonal regressions. To this end, we created a synthetic database of 120 (m_1, m_2) pairs of magnitudes distributed according to an exponential-distribution and affected by a Gaussian error. The parameter B of the exponential distribution is set to $\log_e 10 = 2.3$, corresponding to $b \simeq 1$ in the G-R relation (equation 1 and 2, Section 4); σ_{m_2} is fixed to 0.18 and σ_{m_1} is set to 0.22 as in Table 2. We should therefore expect that ISR gives more robust results compared to SR. The number of pairs was fixed to such a small value in order to be similar to the true database available (Table 2). We note that m_1 and m_2 have been generated using the same distribution, hence on average $m_1 = m_2$, i.e., contrary to Figures 10-12 which represent real cases and which will be discussed later on, there is no systematic bias in magnitude relation. The results of the standard and orthogonal regression on this data set are reported in Figure 8, which shows the substantial systematic error (biases) introduced by the use of standard regression. For example, suppose that we measured $m_1 = 6.0$. This value would give $m_2 = 5.70$ according to the SR, $m_2 = 6.24$ according to the ISR but $m_2 = 6.00$ according to the general orthogonal regression. Systematic errors of opposite sign would occur for m_1 approximately lower than 4.5.

Since this set is composed of synthetic data, generated by a simulation which does not include data incompleteness at low magnitude, there is no lower completeness threshold and the cumulative frequency-magnitude distribution is a line with slope equal to -1 without any deviation from linearity. However, if by analogy with what is usually done in seismology we consider only events above a magnitude threshold (say 4.5) higher than the simulation threshold (Figure 9) we still obtain different regression laws and, as expected since $\eta < 1$, the ISR is closer to GOR. For $m_1 = 6$, GOR would estimate $m_2 = 6.18$, SR would estimate $m_2 = 5.79$ and ISR would produce $m_2 = 6.46$. Again, the most correct results are given by GOR.

3.4 Application

We now apply orthogonal regression to the real datasets of Table 2 and find the regression coefficients and error estimates reported in Table 3. The estimates of

standard deviation σ of the different magnitudes, needed to derive the η value, have been computed from the cases in which more than 3 stations recorded and classified the same earthquake. Although they may be rough estimates, the relative values appear reasonable considering how the different magnitudes are derived from earthquake recordings.

As can be seen from Figures 10, 11 and 12, the general orthogonal regression estimator (GOR) has the property of lying between the slope of the standard regression of y on x (SR) and the inverse of the slope of the standard regression of x on y (ISR). It can also be noted, as already found by Bormann and Wylegalla (1975, Fig. 2 and 3) and Bormann and Khalturin (1975, Fig. 3 and 4), that the slope difference between SR and ISR increases with data scatter.

Considering the $M_w \div m_b$ regression (Figure 12) we observe that the general orthogonal regression slope is much closer to the inverted least square regression slope when m_b is taken as the dependent variable, and in fact $\eta < 1$.

The above examples show that normal regression may introduce significant errors, (from +0.1 to -0.3 in the example of Figure 8) during magnitude conversion. Moreover, as Figures 10-12 demonstrate, a substantial systematic bias is present in observational data. Non-linearity in the magnitude relation, which is obvious in the $M_w \div m_b$ plot, adds to conversion errors. In conclusion, this bias, together with improper accounting for magnitudes' uncertainty, leads to significant distortion of seismicity and of any seismic hazard estimate based on them. Therefore, general orthogonal regression should always be used, rather than standard regression, in magnitude conversions, provided that at least an order of magnitude of η is available.

In addition, note that if the differences between magnitudes were due to random errors, we should see α-values of the order of 0.0, and β-values of the order of 1.0 as in Figures 8 and9. Clearly this is not a case, and the reason for this is not only incorrect use of regression. Magnitudes are empirical quantities, and their inter-relationship was established through various regression relations, so one should not expect that they rigorously characterize earthquake size. Since the β-values often are substantially different from 1.0, this should strongly affect the b-value (of the G-R relation) bias (see more in Section 4).

3.5 Orthogonal regression in the seismic literature: a critical review

Orthogonal regression has been used several times to study European seismicity (Bormann and Khalturin, 1975 and Ambraseys, 1990; Panza *et al.*, 1993; Cavallini and Rebez, 1996; Kaverina *et al.*, 1996; Gutdeutsch *et al.*, 2002 and 2005; Grünthal and Wahlström, 2003; Stromeyer *et al.*, 2004) and global magnitude re-

lations (Bormann and Wylegalla, 1975). It has also been applied to find the slope of the magnitude-intensity relation (cf. Cavallini and Rebez, 1996; Gutdeutsch *et al.*, 2002; Stromeyer *et al.*, 2004). All these applications dealt with the special case (OR) of assuming equal errors on the dependent and independent variables, i.e. $\eta = 1$ in equations 7-9, whereas we deal with the general orthogonal regression method (GOR) of $\eta \neq 1$. Note that similar general results would be less comfortably achieved by rescaling the axes as a function of σ_y and σ_x to reach $\eta = 1$.

Gutdeutsch *et al.* (2002) also applied OR also to find the $M_L \div M_s$ relation for the high-quality Kárnik (1996) earthquake database of Central and Southern Europe, which includes also Italy. It is interesting to note that their $M_L \div M_s$ orthogonal regression based on about 250 pairs (M_L, M_s) gives $\beta = 0.893 \pm 0.163$, which is in perfect agreement with our $\beta_{M_L \div M_s} = 0.859$ derived by merging the first two equations in Table 3. This probably implies that the error ratio of M_L and M_s is close enough to 1 (a more detailed discussion of this problem is left to another manuscript currently in preparation).

As an alternative to OR, Grünthal and Wahlström (2003) and Stromeyer *et al.* (2004) proposed the chi-square maximum likelihood regression. They applied it to find the $M_w \div m_x$ relationship for Central Europe earthquakes, including a small part of Northern Italy. An advantage of the chi-square regression is that it is a distribution-free method. It has, however, the same problem as GOR: the ratio of the standard deviations of the data should be known with a reasonable accuracy, for example 30%. Grünthal and Wahlström (2003) and Stromeyer *et al.* (2004), instead, assumed $\eta = 1$ as in OR. In any case, we note that Grünthal and Wahlström's (2003) chi-square regression gave $\beta_{M_w \div M_s} = 0.769$, while our general orthogonal regression gives $\beta_{M_w \div M_s} = 0.765$, i.e. a very similar result in spite of the different method and data set. A comparison between chi-square method and OR can also be found in Gutdeutsch *et al.* (2005).

4 Frequency-Magnitude relation.

Frequency-magnitude relations (G-R relation) are ubiquitously used in seismology, for example, to test the completeness of seismic catalogues, to estimate seismic hazard or in the thermodynamic criticality approach to earthquake modelling (Knopoff, 2000). The G-R relation can itself be treated as a regression but it is not amenable to the case treated in this paper since neither the assumptions of SR, nor those of GOR strictly apply. The preferred technique to estimate the b-value in this case is the Maximum Likelihood Estimator Method (MLEM) given by Aki

(1965) and discussed by Weichert (1980), Utsu (1999) and Kagan (2005):

$$b = \frac{\log_{10} e}{\overline{M} - M_c}, \tag{19}$$

where $\overline{M}$ is the average magnitude and M_c is the lower cutoff magnitude (or the completeness threshold magnitude). The accuracy on the b-value, derived from the statistical estimation theory (Aki, 1965; Utsu, 1999), is given by

$$\sigma_b = \frac{b}{\sqrt{n}}, \tag{20}$$

where n is the total number of earthquakes in the sample.

Since the G-R relations are evaluated on a single magnitude scale, which is in general derivesd from converting other magnitudes, it is nevertheless interesting to see how the b estimations are affected by the use of different regression procedures employed for magnitude conversion. Again, the issue is most effectively treated in simulation.

We start by assuming that we have a set of 10^4 exponentially distributed events with M_2 magnitude values which follow a G-R relation (equation 2) without added error and which represent our 'true' data. We then add random Gaussian errors with variance σ^2_{m2} and obtain a set of 10^4 (m_2) data which represent the 'noisy' data, i.e. the measured data of the same set. These can be seen as the real moment magnitude data (with error). Generally these data are unavailable and are derived by converting estimates measured on other magnitude scales. We consider just one such magnitude scale, which we call m_1 and generate a set of 10^4 (m_1) synthetic data still using an exponential distribution with $B = 2.3$ (equation 2) but with different realizations and Gaussian errors with variance σ^2_{m1}. These can be imagined to correspond, say, to M_L data.

We now derive the conversion law for m_2 data from the m_1 set ($m_2 = \alpha + \beta\, m_1$) by using both SR ($\dot{m}_2$) and GOR ($\ddot{m}_2$) and study how the G-R plot of the derived magnitudes ($\dot{m}_2$) and ($\ddot{m}_2$) is consequently affected. The b-values in both cases are computed through the MLEM (equations 19-20).

Figures 13 to 15 show both the 'true' data relative to M_2, the corresponding noisy ones with directly added noise m_2, and the $\dot{m}_2$ and $\ddot{m}_2$ ones inferred from m_1. We first note that, as predicted by Tinti and Mulargia (1985), a magnitude error affects the a, but not the b-value in the G-R relation. In particular, in the log-linear part of the plot, a is shifted to a' as a function of the variance σ^2_m:

$$a' = a + \frac{b^2 \sigma^2_m}{2 \log_{10} e}, \tag{21}$$

so that the effect of magnitude uncertainties is that the observed number of earthquakes exceeding a given magnitude always appears larger than the true number.

Figure 13 shows that when $\eta = 25$ ($\sigma_{m2} = 0.5$ and $\sigma_{m1} = 0.1$), i.e. when the error on the original variable m_2 is 5 times larger than error on the original variable m_1, SR and GOR give overlapping G-R plots which coincide with the 'true' data. As stated in Section 3, SR (when $\sigma_y >> \sigma_x$) and GOR are capable of retrieving Y from x, because x can be considered $\simeq X$. However, $\sigma_y >> \sigma_x$ hardly occurs in real earthquake catalogues since measurement errors on m_2 are unlikely to be 5 times larger than measurement errors on m_1 (see e.g. Table 2 for realistic error estimates on m_1 and m_2. As mentioned above, in the most common case $m_2 = M_w$ and $m_1 = M_L$). In this example we have used a large σ_{m2} in order to emphasize the effect of the magnitude uncertainties (dashed vs solid line) in Figure 13. Note that, in this case, the true a-value at $m = 5.5$ is 2.51 (corresponding to 326 earthquakes), while the observed a'-value is 2.78 (corresponding to 604 earthquakes) which fits with the theoretical $a' = 2.79$ derived from equation 21 (corresponding to 617 earthquakes).

When η decreases from 25 to 1 ($\sigma_{m2} = \sigma_{m1} = 0.2$) and 0.25 ($\sigma_{m2} = 0.2$, $\sigma_{m1} = 0.4$), (see Figure 14 and 15) the G-R relation derived using $\dot{m}_2$ values (converted from m_1 through SR) strongly differs from the true G-R, while the G-R obtained using $\ddot{m}_2$ values (converted through the GOR) still replicate the true relation.

An example of average b slope coefficients and their error is shown in Table 4 and is obtained through MLEM on events with $m \geq 4.5$. Due to the large sample size on which we run these simulations, we expect that the b-values obtained in this way are less scattered than in the simulations of Figures 3-6.

In summary, Figures 13-15 and Table 4 show that the use of $\dot{m}_2$ data inferred from m_1 through SR produces a strong bias in the slope which is reflected in a biased estimate of the b-value, even when computed through the MLEM. When $\ddot{m}_2$ data inferred from GOR are used, true unbiased M_2 data are always produced. This demonstrates the importance of using GOR in *converting magnitudes* to be used in the G-R relation to obtain correct hazard estimates.

5 Regression problems involving slope: a non-parametric regression test

The performance of a least square estimator of a regression coefficient can be gauged through a non-parametric test.

When there are no ties in the data (i.e., all data points are distinct), Theil's test (p. 416 in Hollander and Wolfe, 1999) is suitable. When ties are present (and this

is easily the case with magnitudes), an estimate of the regression coefficients can be produced based on Kendall's tau (Kendall, 1962; Sen, 1968). This procedure is similar to Theil's but is based on weaker assumptions and does not require $x_1, \ldots, x_n$ to be all distinct. We start from the premise that the model is

$$y_i = \alpha + \beta x_i \qquad i = 1, \ldots, n, \tag{22}$$

where $x_1 \leq x_2 \leq \ldots \leq x_n$ are the known constants and α and β the unknown parameters. Calling N the number of non-zero differences $x_j - x_i$ $(1 \leq i \leq j \leq n)$, Sen (1968) shows that the point estimator is the median of the N slopes $(y_j - y_i)/(x_j - x_i)$ for which $x_j \neq x_i$ and this is shown to be unbiased for β. The confidence interval for β is also obtained in terms of second order statistics of this set of N slopes.

Let us define $c(u)$ to be 1, 0 or -1 if $u > 0$, $u = 0$ or $u > 0$ respectively. Then the number of positive differences $x_j - x_i$ is

$$N = \sum_{1 \leq i < j \leq n} c(x_j - x_i), \tag{23}$$

and $N \leq \binom{n}{2}$. For any real β, $Z_i(\beta) = y_i - \beta x_i$. Relying on the Kendall's tau statistic (Kendall, 1962) between x_i and $Z_i(\beta)$ we thus have

$$U_n(\beta) = \left[N \binom{n}{2} \right]^{-\frac{1}{2}} \sum_{i \leq i < j \leq n} c(x_j - x_i)\, c(Z_j(\beta) - Z_i(\beta)). \tag{24}$$

Since we deal with quite large sample sizes, we can use the asymptotic formula

$$U_n(\beta) = \tau_{1/2\epsilon} \left[V_n / \sqrt{N \binom{n}{2}} \right], \tag{25}$$

where τ_ϵ is the upper $100\epsilon\%$ point of a standard normal distribution and the variance V_n is

$$V_n = \frac{1}{18}[n(n-1)(2n+5) - \sum_{j=1}^{a_n} u_j(u_j - 1)(2u_j + 5)]. \tag{26}$$

Let x_n be composed of $a_n \geq 2$ distinct sets of elements. In each of the i-th $(i = 1, ..., a_n)$ set there are then u_i all equal elements which are the values to be inserted in equation 26. $U_n(\beta)$ is a strictly distribution-free statistic having a distribution symmetric about 0. This implies that one way of estimating β is to make $U_n(\beta)$ as close to zero as possible, by varying β.

The confidence interval for β having the confidence coefficient $1-\epsilon_n$ is thus

$$P\{\beta_L^* < \beta < \beta_U^* | \beta\} = 1 - \epsilon_n, \tag{27}$$

where

$$\beta_U^* = \mathrm{S}up\{\beta : U_n(\beta) \geq -U_n\}, \tag{28}$$

$$\beta_L^* = \mathrm{I}nf\{\beta : U_n(\beta) \leq U_n\}. \tag{29}$$

We recall that these formulas apply exclusively when ties are present in the x variable only. We now consider the set of N distinct pairs (i, j) for which $x_j > x_i$ and define

$$X_{ij} = \frac{y_j - y_i}{x_j - x_i}. \tag{30}$$

Thus the X_{ij} are the slopes of the lines connecting each pair of points (x_i, y_i) and (x_j, y_j) where $x_i \neq x_j$. After arranging the N values of X_{ij} in increasing order, the slope estimator is shown (Sen, 1968) to be the median of the N numbers and the confidence interval is given by the M_1-th and M_2-th values

$$N^* = \sqrt{N\binom{n}{2}}U_n \quad \text{and} \quad M_i = \frac{1}{2}(N + (-1)^i N^*). \tag{31}$$

An extensive study of this test performance in presence of ties, outliers and non-normal data and error distributions can be found in Castellaro and Mulargia (2006).

5.1 Application

The test described above provides a robust (point as well as interval) estimator of β, which is valid also when the parent distributions are not Gaussian (as happens with magnitudes). We apply it to our dataset in order to find the regression coefficients which express one type of magnitude with respect to another, and their significance. We use a 2.5% significance on each tail (i.e. $\tau_{1/2\epsilon} = 1.96$ in equation 25) and find that the slope regression values should be included within the following ranges (Table 5):

$$0.574 < \beta_{M_w - M_s} < 0.827, \tag{32}$$

$$0.770 < \beta_{M_w - M_L} < 0.937, \tag{33}$$

$$0.690 < \beta_{M_w - m_b} < 1.600, \tag{34}$$

when the M_w data are on the y axis. Since the test applies when there are ties only in the x variable, we removed ties on y, when present, by averaging their values and properly accounting for the number of degrees of freedom.

We note that the $M_w \div m_b$ confidence interval on slope is very large and this happens because of the large number of ties on x (m_b, see Figure 12). If one averages the tied-values, the confidence interval is then restricted to $1.01 < \beta_{M_w - m_b} < 1.29$, which does not include the slope value found with the SR, while it still includes the slope value found with GOR (Table 3). In general, the GOR values lie in the middle of the slope ranges derived from Kendall's tau and are more stable than the SR slopes.

6 Discussion and conclusions

We have found that orthogonal regression in its general form (GOR) provides superior performance to standard linear regression (SR) in virtually all cases, provided that at least an order of magnitude estimate of the error ratio on the y and x variables is available. The use of general orthogonal regression is therefore recommended in all practical cases.

As a demonstration of its application we selected the case of magnitude conversions for Italian seismicity. Practically all the regression laws proposed in the literature for converting magnitudes from one scale to another were based on the standard least square regression, which is not a good estimator since both the x and y variables are affected by errors of non-negligible size and since the data are not normally distributed, features which contradict the basic assumptions of standard least square regression. General orthogonal regression, which is specifically formulated to handle the case where both variables are affected by errors of non-negligible size, should instead be used to obtain reliable magnitude conversions and trustworthy seismic hazard estimates. This appears clearly in the final application in which we show that the Gutenberg-Richter frequency-magnitude relation, which is in general calculated on magnitudes converted from various scales, can be heavily biased if magnitudes are converted through standard least squares. By contrast, it is possible to obtain unbiased estimates of a- and b-values (equation 1) by converting magnitudes through generalized orthogonal regression.

The influence of magnitude uncertainties on the a- and b-values provides a strong indication of serious biases in traditional interpretations of earthquake size distribution. Shifts in the a-values due to magnitude errors may have important consequences on seismic data obtained in different time intervals. As a rule, the accuracy of older data is lower, and since the a-value shift is always positive and proportional to the square of magnitude errors (Tinti and Mulargia, 1985 and

equation 21), this means that the seismic activity for older time intervals may spuriously appear to exceed more recent activity by a significant margin. This effect is likely to be at least in part responsible for the often claimed discrepancy between earthquake rates in recent and old catalogs. Such disagreement in earthquake rates might even be taken as evidence for the presence of characteristic earthquakes in a region.

Virtually all traditional (non seismic moment) earthquake catalogs contain several magnitudes, the relationships between which have been established by standard (not orthogonal) regression. Bias and systematic effects in magnitude data are difficult or next to impossible to disentangle. As Figures 13-15 and Table 4 demonstrate, incorrect application of regression methods in the magnitude conversion may induce b-value bias of the order of 20%-40%. Hence, the variations in b-values (claimed by many researchers) may have an additional explanation as an artifact of improper catalog data processing.

Acknowledgments

S. Castellaro and F. Mulargia acknowledge support from MIUR 60% funds and from the INGV-DPC 2004-2006, Project V3. Y.Y. Kagan appreciates partial support from the National Science Foundation through grants EAR 00-01128, EAR 04-09890, DMS-0306526, and from the Southern California Earthquake Center, Publication 960, SCEC. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. Publication 0000, SCEC.

We thank prof. P. Bormann, prof. A. Oncel as well as the three anonymous reviewers for their constructive suggestions.

References

[1] Aki, K., 1965. Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, *Bull. Earthquake Res. Inst. Tokyo Univ.*, **43**, 237-239.

[2] Ambraseys, N.N., 1990. Uniform magnitude reevaluation of European earthquakes associated with strong-motion recordings, *Earthq. Eng. and Struct. Dyn.*, **19**, 1-20.

[3] Bindi, D., Spallarossa, D., Eva, C. & Cattaneo, M., 2005. Local and duration magnitudes in northwestern Italy, and seismic moment versus magnitude relationships, *Bull. Seism. Soc. Am.*, **95**, 592-604.

[4] Bird, P. & Kagan, Y.Y., 2004. Plate-tectonic analysis of shallow seismicity: apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, *Bull. Seism. Soc. Am.*, **94**, 2380-2399.

[5] Bormann, P., & Khalturin,, V.I., 1975. Relations between different kinds of magnitude determinations and their regional variations. *Proceedings of the XIVth General Assembly of the European Seismological Commission*, Trieste, 16 - 22 September 1974, Ed. Nationalkomitee für Geodäsie und Geophysik, AdK of the GDR, Berlin, 27-35.

[6] Bormann, P., & Wylegalla, K., 1975. Untersuchung der Korrelationsbeziehungen zwischen verschiedenen Arten der Magnitudenbestimmung der Station Moxa in Abhängigkeit vom Gerätetyp und vom Herdgebiet, *Publs. Inst. Geophys. Pol. Ac. Sci.*, **93**, 159-175.

[7] Braunmiller J., Deichmann, N., Giardini, D., Wiemer, S. & the SED Magnitude Working Group, 2005. Homogeneous moment-magnitude calibration in Switzerland, *Bull. Seism. Soc. Am.*, **95**(1), 58-74.

[8] Carroll, R.J. & Ruppert, D., 1996. The use and misuse of orthogonal regression in linear errors-in-variables models, *American Statistician*, **50**, 1-6.

[9] Castellaro, S. & Mulargia F., 2006. A procedure for linear regression capable of less biased estimates on all magnitude relations, including Gutenberg-Richter law, submitted.

[10] Cavallini, F. & Rebez, A., 1996. Representing earthquake intensity magnitude relationship with a nonlinear function, *Bull. Seismol. Soc. Amer.*, **86**(1), 73-78.

[11] Choy, G.L. & Boatwright, J.L., 1995. Global patterns of radiated seismic energy and apparent stress, *J. Geophys. Res.*, **100**, 18205-18228.

[12] Cornell, C.A, 1968. Engineering seismic risk analysis, *Bull. Seism. Soc. Am.*, **58**, 1583-1606.

[13] Draper, N.R. & Smith, H., 1998. *Applied Regression Analysis*, Wiley Series in Probability and Statistics, 3rd ed.

[14] Fuller, W.A., 1987. *Measurement Error Models*, John Wiley, New York.

[15] Gasperini, P., 2002. Local magnitude revaluation for recent Italian earthquakes (1981-1996), *J. Seismology*, **6**, 503-524.

[16] Gasperini, P. & Ferrari, G., 2000. Deriving numerical estimates from descriptive information: the computation of earthquakes parameters, *Annali di Geofisica*, **43**, 729-746.

[17] Grünthal, G. & Wahlström, R., 2003. An M_w based earthquake catalogue for central, northern and northwestern Europe using a hierarchy of magnitude conversions, *J. Seismology*, **7**(4), 507-531.

[18] Gutdeutsch, R., Kaiser, D., & Jentzsch, G., 2002. Estimation of earthquake magnitudes from epicentral intensities and other focal parameters in Central and Southern Europe, *Geophys. J. Int.*, **151**(3), 824-834.

[19] Gutdeutsch, R., Kaiser, D., Bormann, P., Riufeng, L. & Ren, X., 2005. Regressionanalyse zur Umrechnung zwischen Magnituden des chinesischen Erdbebendienstes und des NEIC, *Mitteilung der Deutschen Geophysikalischen Gesellschaft*, **3**, 14-22.

[20] Gutenberg, B., & Richter, C.F., 1944, Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.*, **46**, 105-145.

[21] Haberman, R.E., 1986. A test of two techniques for identifying systematic errors in magnitude estimates using data from the Parkfield, California region, *Bull. Seism. Soc. Am.*, **82**, 533-579.

[22] Hanks, T.C., 1992. Small earthquakes, tectonic forces, *Science*, **256**, 1430-1432.

[23] Hanks, T.C. & Kanamori, H., 1979. A moment-magnitude scale, *J. Geophys. Res.*, **84**, 2348-2350.

[24] Hollander, M. & Wolfe, D.A., 1999. *Nonparametric Statistical Methods*, 2-nd ed., New York, Wiley, 787 pp.

[25] Ishimoto, M., & Iida, K., 1939, Observations sur les seisms energistré per le microseismograph construite dernierment (I), *Bull. Earthq. Res. Inst.*, **17**, 443-478, in Japanese.

[26] Kagan, Y.Y., 2002a. Seismic moment distribution revisited: II. Moment conservation principle, *Geophys. J. Int.*, **149**, 731-754.

[27] Kagan, Y.Y., 2002b. Modern California earthquake catalogs and their comparison, *Seism. Res. Lett.*, **73**(6), 921-929.

[28] Kagan, Y.Y., 2003. Accuracy of modern global earthquake catalogs, *Phys. Earth Planet. Inter. (PEPI)*, **135**(2-3), 173-209.

[29] Kagan, Y. Y., 2005. Earthquake slip distribution: A statistical model, *J. Geophys. Res.*, **110**(5), B05S11, doi:10.1029/2004JB003280, 1-15.

[30] Kárnik, V., 1996. *Seismicity of Europe and the Mediterranean*, Klima ed., Academy of Science of the Czech Republic StudiaGeo s.s.r.o. and Geophysical Institute, Praha (KA96).

[31] Katsura, K. & Y. Ogata, 2004. Simultaneous estimation of b-values and detection rates of earthquakes for the application to aftershock probability forecasting, *Eos Trans. AGU*, **85**(47), Fall Meet. Suppl., Abstract S22A-06.

[32] Kaverina, A.N., Lander, A.V. & Prozorov, A.G., 1996. Global creepex distribution and its relation to earthquake-source geometry and tectonic origin, *Geophys. J. Int.*, **125**, 249-265.

[33] Kendall, M.G., 1962. *Rank correlation methods*, (3rd ed.), Griffin, London.

[34] Kendall, M.G. & Stuart, A., 1979. *The Advanced Theory of Statistics*, (Vol. 2, 4th ed.), Griffin, London.

[35] Knopoff, L., 2000. The magnitude distribution of declustered earthquakes in Southern California, *P. Nat. Acad. Sci. USA*, *97*, 11880-11884.

[36] Lee, W.H.K., Bennet, R.E. & Meagher, K.L., 1972. A method of estimating magnitude of local earthquakes from signal duration, *U.S. Geol. Surv. Open-File Rep.*, 28 pp.

[37] Madansky, A., 1959. The fitting of straight lines when both variables are subject to error, *J. Am. Statist. Assoc.*, **54**, 173–205.

[38] NMSOP, New Manual of Seismological Observatory Practice, 2002. GeoForschungsZentrum Potsdam, Bormann P. Editor, 2 Volumes, ISBN 3-9808780-0-7.

[39] Panza, G.F., Prozorov, A.G. &. Pazzi, G., 1993. Extension of global creepex definition ($M_s - m_b$) to local studies ($M_d - M_L$): the case of the Italian region, *Terra Nova*, **5**(2), 150-156.

[40] Pérez, O., 1999. Revised world seismicity catalog (1950-1997) for strong ($M_S \geq 6$) shallow ($h \leq 70$ km) earthquakes, *Bull. Seism. Soc. Am.*, **89**, 335-341.

[41] Reiersol, O., 1950. Identifiability of a Linear Relation between Variables Which Are Subject to Error, *Econometrica*, **18**(4), 375-389.

[42] Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau, *J. Am. Stat. Ass.*, **63**, 1379-1389.

[43] Stefanski, L. A., & J. R. Cook, 1995. Simulation-Extrapolation: The Measurement Error Jackknife (in Theory and Methods), *J. Amer. Statist. Assoc.*, **90**(432), 1247-1256.

[44] Stromeyer, D., Grünthal, G., & Wahlström, R., 2004. Chi-square regression for seismic strength parameter relations, and their uncertainties, with applications to an M_w based earthquake catalogue for central, northern and northwestern Europe, *J. Seismology*, **8**(1), 143-153.

[45] Tinti, S. & Mulargia, F., 1985. Effects of magnitude uncertainties on estimating the parameters in the Gutenberg-Richter frequency-magnitude law, *Bull. Seismol. Soc. Amer.*, **75**, 1681-1697.

[46] Utsu, T., 1999. Representation and analysis of the earthquake size distribution: A historical review and some new approaches, *Pure Appl. Geophys.*, **155**, 509-535.

[47] Weichert, D., 1980. Estimation of earthquake recurrence parameters for unequal observation periods for different magnitudes, *Bull. Seism. Soc. Am.*, **70**, 1337-1356.

[48] Wetherill, G. B., 1986. *Regression Analysis with Applications*, London, Chapman and Hall, 311 pp.

[49] York, D., 1967. The best isochron, *Earth Planet. Sci. Lett.*, **2**, 479-482.

[50] Zúñiga, F.R. & and Wyss, M., 1995. Inadvertent changes in magnitude reported in earthquake catgalogs: their evalutation through *b*-value estimates, *Bull. Seism. Soc. Am.*, **85**, 1858-1866.

Magnitude definition	
M_s	surface wave magnitude
M_L	local magnitude
M_d	duration magnitude
M_w	moment magnitude
M_0	scalar seismic moment
M_e	energy magnitude

Table 1: Symbols used to indicate different magnitudes: M_L indicates a true or synthetic Wood-Anderson magnitude calculated from broad-band seismic stations; M_d is calculated as in Gasperini (2002); M_e as in Choy and Boatwright (1995).

		Data no.
M_s	M_w	109
M_L	M_w	121
m_b	M_w	204

	σ
M_s	0.28
M_L	0.22
m_b	0.37
M_w	0.18

Table 2: Data sets used to perform the tests on the slope. The left table reports the number of events recorded in Italy between 1981 and 1996 for which both the magnitudes shown in the first two columns are available. The right table reports the standard deviation σ of the different magnitudes.

y	**x**	β	α	α''
M_w	M_s	0.765 ± 0.035	1.445 ± 0.175	0.426
M_w	M_L	0.890 ± 0.033	0.770 ± 0.159	0.283
M_w	m_b	1.291 ± 0.046	-1.272 ± 0.271	0.037

Table 3: Results of the general orthogonal regression (GOR) applied to our data set. The model is $y = \alpha + \beta x$, that is $M_w = \alpha + \beta m_x$. α'' applies when magnitudes are reduced so that $m''_w = M_w - 4.5$, $m''_x = m_x - 4.5$ and $m''_w = \alpha'' + \beta m''_x$. See Figures 10-12.

	b_{M2}	b_{m2}	$b_{\dot{m}2}$	$b_{\ddot{m}2}$
$\eta = 25$	1.091 ± 0.009	1.006 ± 0.006	1.149 ± 0.009	1.113 ± 0.008
$\eta = 1$	1.194 ± 0.010	1.044 ± 0.007	1.259 ± 0.010	1.205 ± 0.010
$\eta = 0.25$	1.141 ± 0.009	1.147 ± 0.009	1.399 ± 0.012	1.069 ± 0.008

Table 4: b-values of the G-R distribution computed through the maximum likelihood estimator method (MLEM) regression using $m = 4.5$ as lower cutoff for the M_2 exponential data set without errors (b_{M2}), the m_2 exponential data set with Gaussian errors (b_{m2}), the $\dot{m}_2$ data set retrieved from m_1 through the standard regression ($b_{\dot{m}2}$) and the $\ddot{m}_2$ data set retrieved from m_1 through the general orthogonal regression ($b_{\ddot{m}2}$). η is the variance ratio between m_2 and m_1.

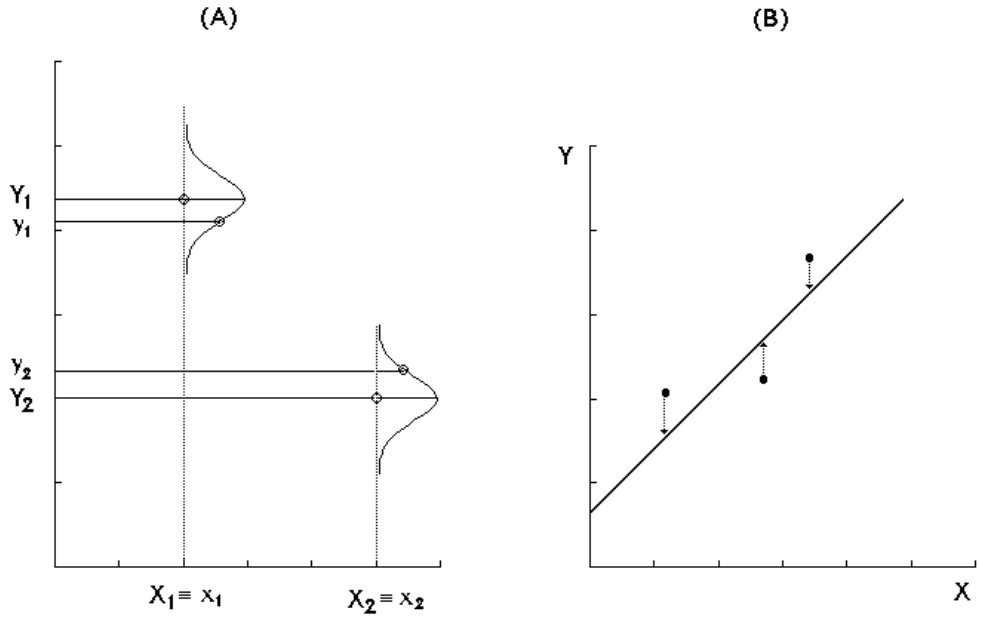

Figure 1: Standard least square regression applies when the dependent variable is affected by a Gaussian error much larger than the error affecting the independent variable (A) and it consists in projecting the independent variable along a vertical line (B).

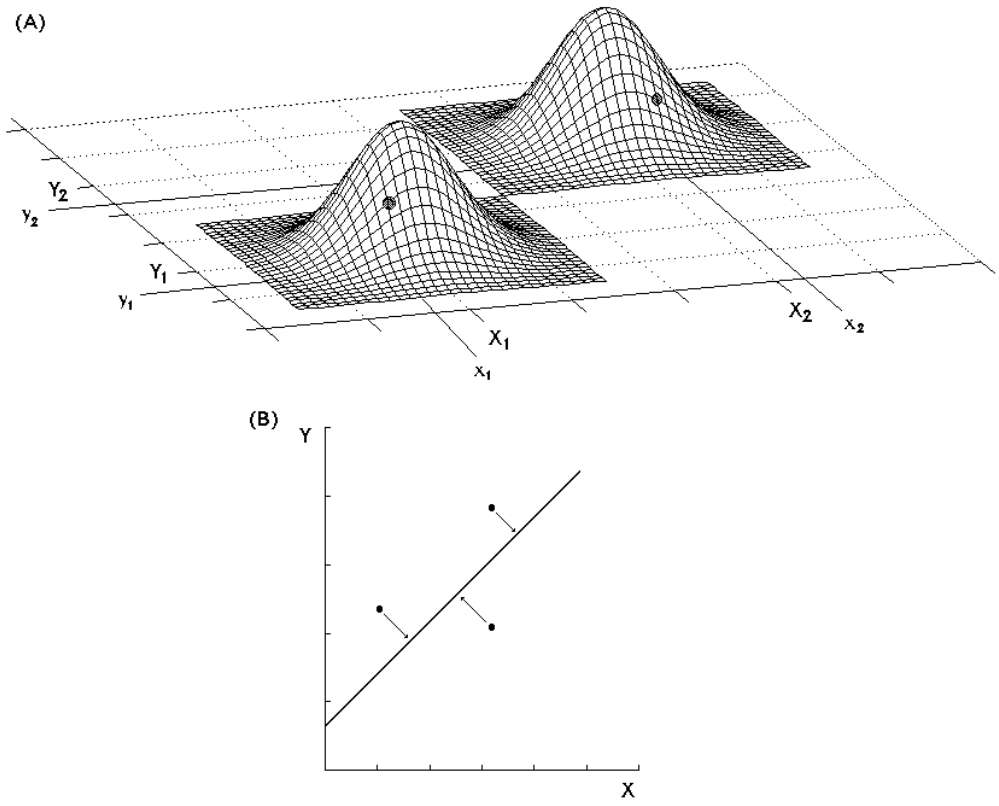

Figure 2: General orthogonal regression (A) applies when both variables are affected by non-negligible Gaussian errors. The path from the experimental (x_i, y_i) points to the 'true' (X_i, Y_i) is the weighted orthogonal distance from the line (B).

		SR	**non parametric test**	
y	**x**	β	β	95% **confidence interval**
M_w	M_s	0.673± 0.0322	0.703	$0.574 < \beta_{M_w - M_s} < 0.827$
M_s	M_w	1.190± 0.0569		
M_w	M_L	0.813± 0.0316	0.856	$0.770 < \beta_{M_w - M_L} < 0.937$
M_L	M_w	1.040± 0.0405		
M_w	m_b	0.971± 0.0429	1.157	$0.690 < \beta_{M_w - m_b} < 1.600$
m_b	M_w	0.737± 0.0325		

Table 5: Comparison between the standard regression and the non-parametric test on β, where the model is intended to be $y = \alpha + \beta x$. *3-rd column*: β coefficient of the standard regression applied to the data set described in Table 2. Confidence bounds are given as one σ. *4-th and 5-th columns*: results of the Kendall's tau test on β in terms of the 95% confidence interval. Parameters derived through the general orthogonal regression on the same dataset are in Table 3.

- **Figure 3** *Left:* examples taken from the 10^3 generations of 50 couples of points (x, y). Data are normally distributed and variable Gaussian errors have been added on the x and y axis. *Right:* frequency-distribution of the β coefficient found with generalized orthogonal regression (GOR), standard least squares (SR) and inverted standard least squares (ISR) method. **(A)** $\eta = 25$, $\sigma_y = 2$, $\sigma_x = \sqrt{\frac{\sigma_y^2}{\eta}} = 0.4$ as from equation 7. **(B)** $\eta = 4$, $\sigma_y = 2$, $\sigma_x = 1$. **(C)** $\eta = 1$, $\sigma_y = 2$, $\sigma_x = 2$. **(D)** $\eta = 0.25$, $\sigma_y = 2$, $\sigma_x = 4$.
- **Figure 4** *Left:* examples taken from the 10^3 generations of 50 couples of points (x, y). Data are exponentially distributed and variable Gaussian errors have been added on the x and y axis. *Right:* frequency-distribution of the β coefficient found with the GOR, SR and ISR method. **(A)** $\eta = 25$, $\sigma_y = 1$, $\sigma_x = 0.2$. **(B)** $\eta = 4$, $\sigma_y = 1$, $\sigma_x = 0.5$. **(C)** $\eta = 1$, $\sigma_y = 1$, $\sigma_x = 1$.
- **Figure 5** *Left:* examples taken from the 10^3 generations of 50 couples of points (x, y). Data are exponentially distributed and variable exponential errors have been added on the x and y axis. *Right:* frequency-distribution of the β coefficient found with the GOR, SR and ISR method. **(A)** $\eta = 25$, $\sigma_y = 1$, $\sigma_x = 0.2$. **(B)** $\eta = 4$, $\sigma_y = 1$, $\sigma_x = 0.5$. **(C)** $\eta = 1$, $\sigma_y = 1$, $\sigma_x = 1$.
- **Figure 6** *Left:* examples taken from the 10^3 generations of 50 couples of points (x, y). Data are log-normally distributed and variable log-normal

errors have been added on the x and y axis. *Right:* frequency-distribution of the β coefficient found with the GOR, SR and ISR method. **(A)** $\eta = 25$, $\sigma_y = 2$, $\sigma_x = 0.4$. **(B)** $\eta = 4$, $\sigma_y = 2$, $\sigma_x = 1$. **(C)** $\eta = 1$, $\sigma_y = 2$, $\sigma_x = 2$.

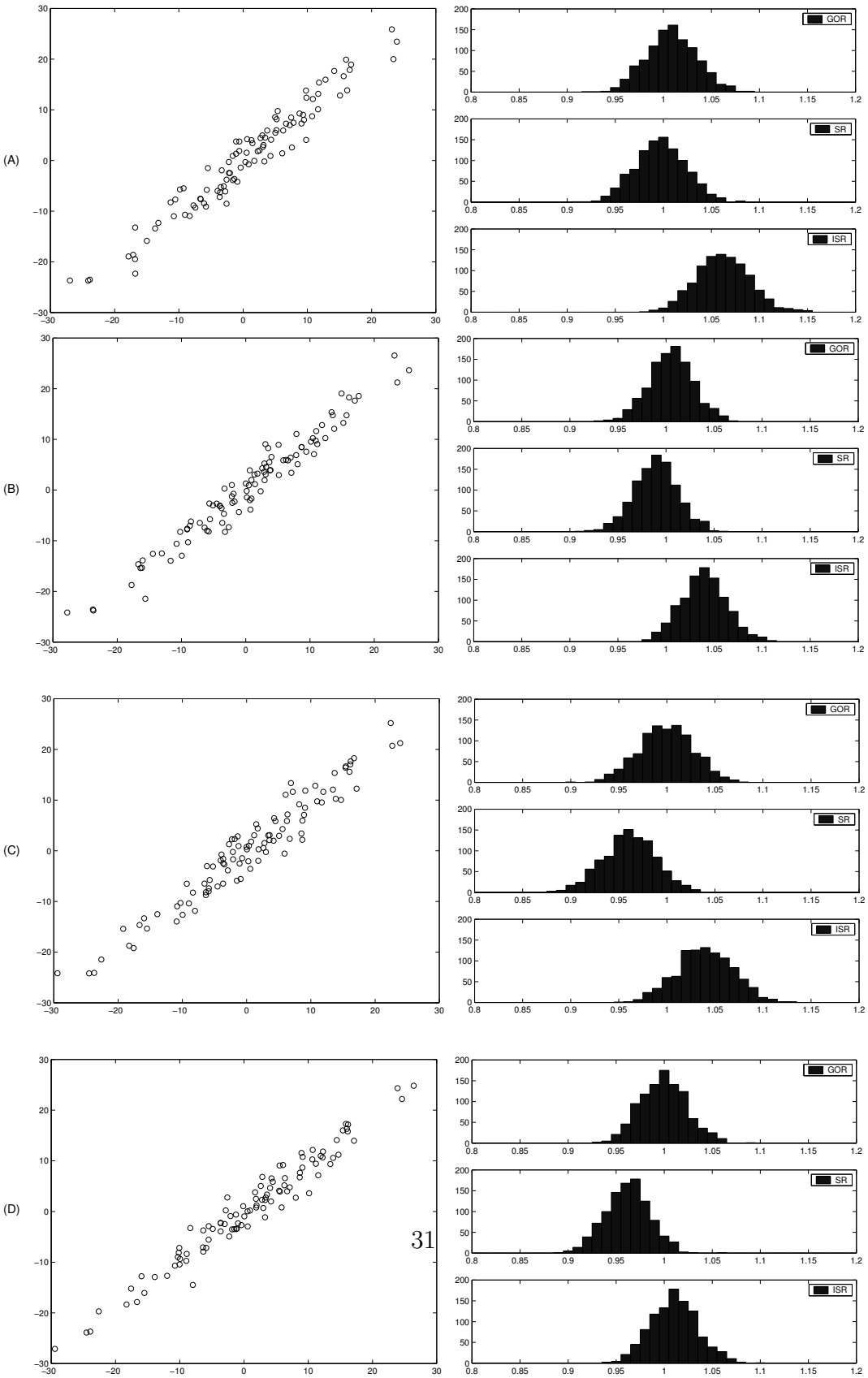

Figure 3:

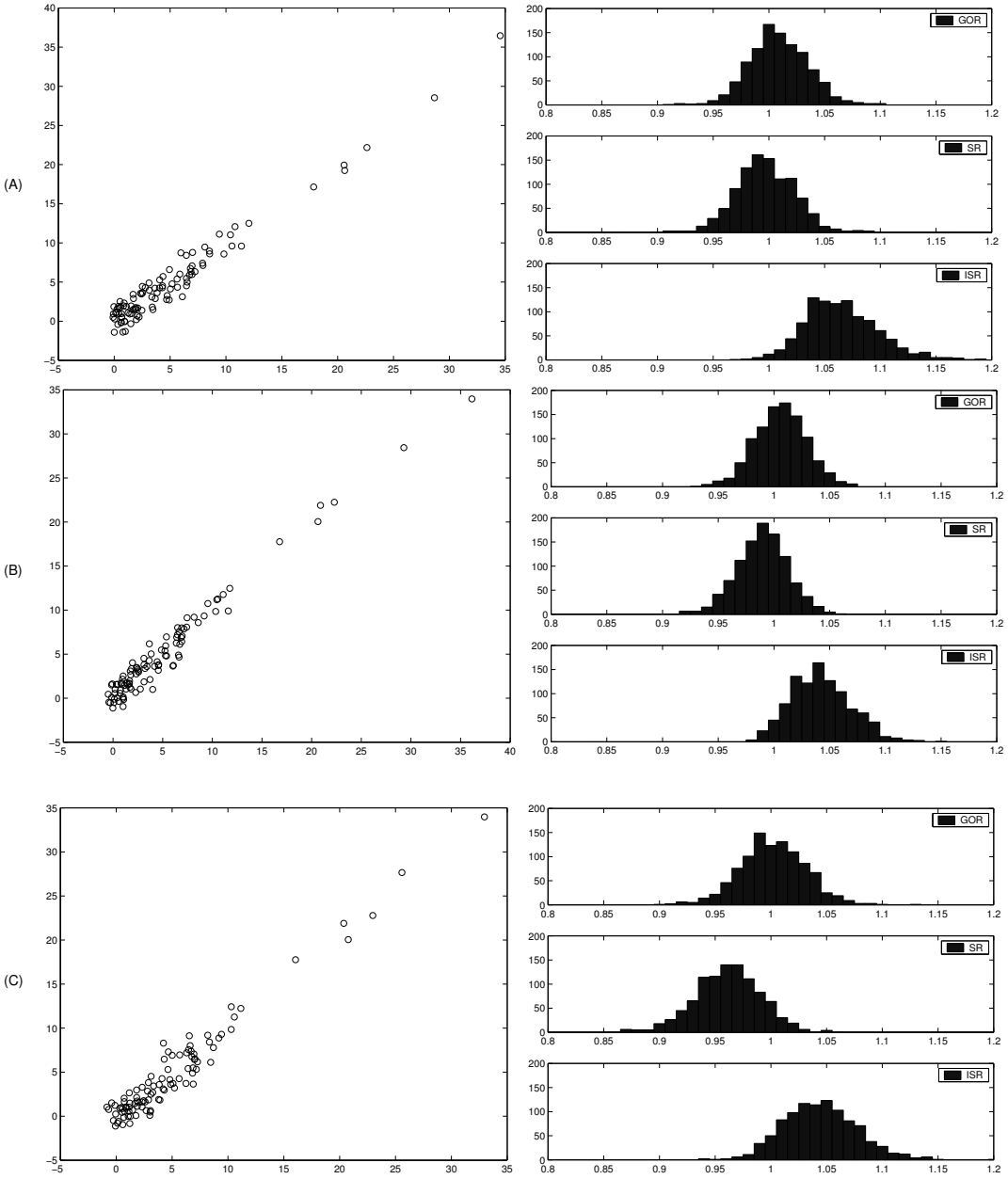

Figure 4:

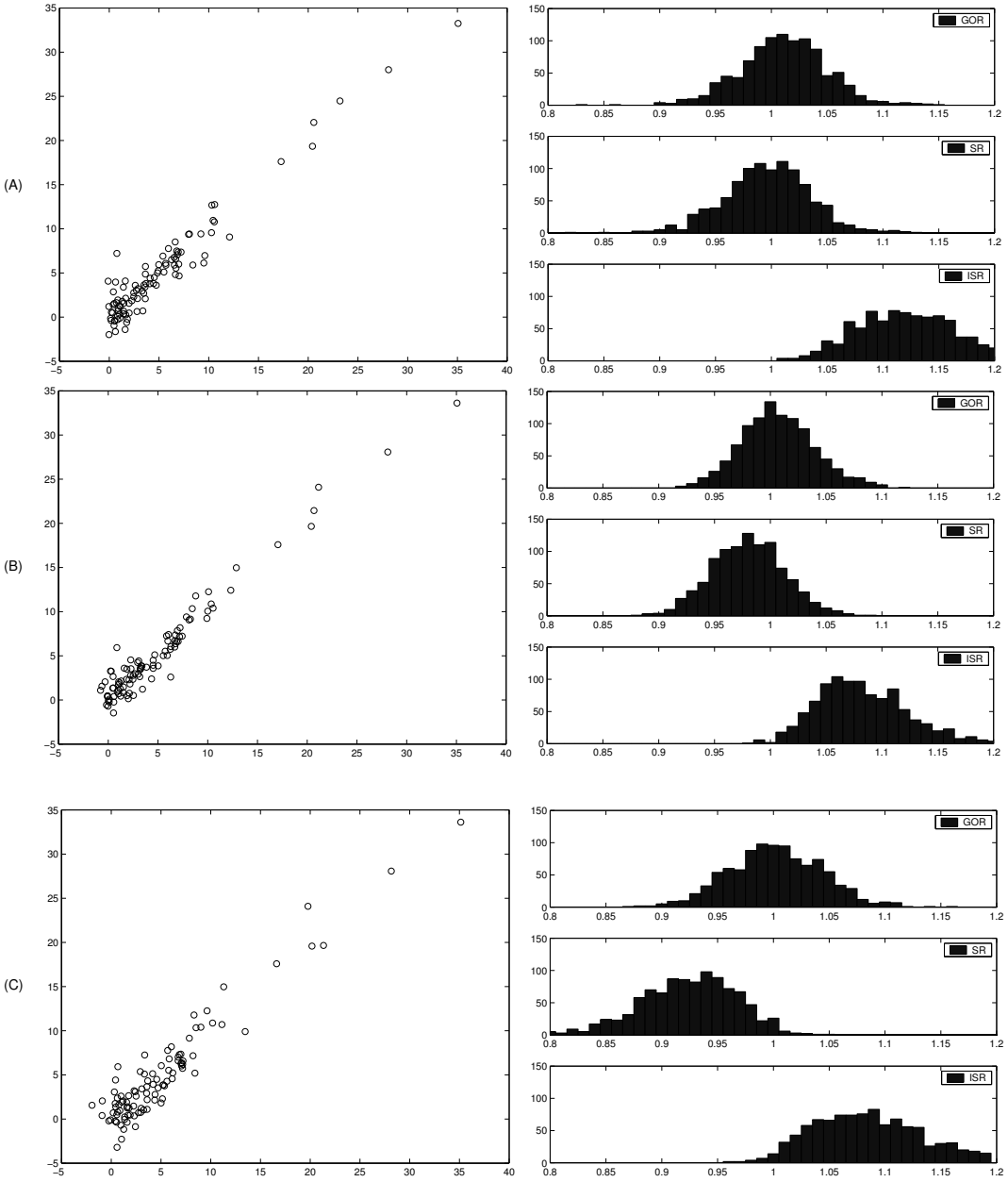

Figure 5:

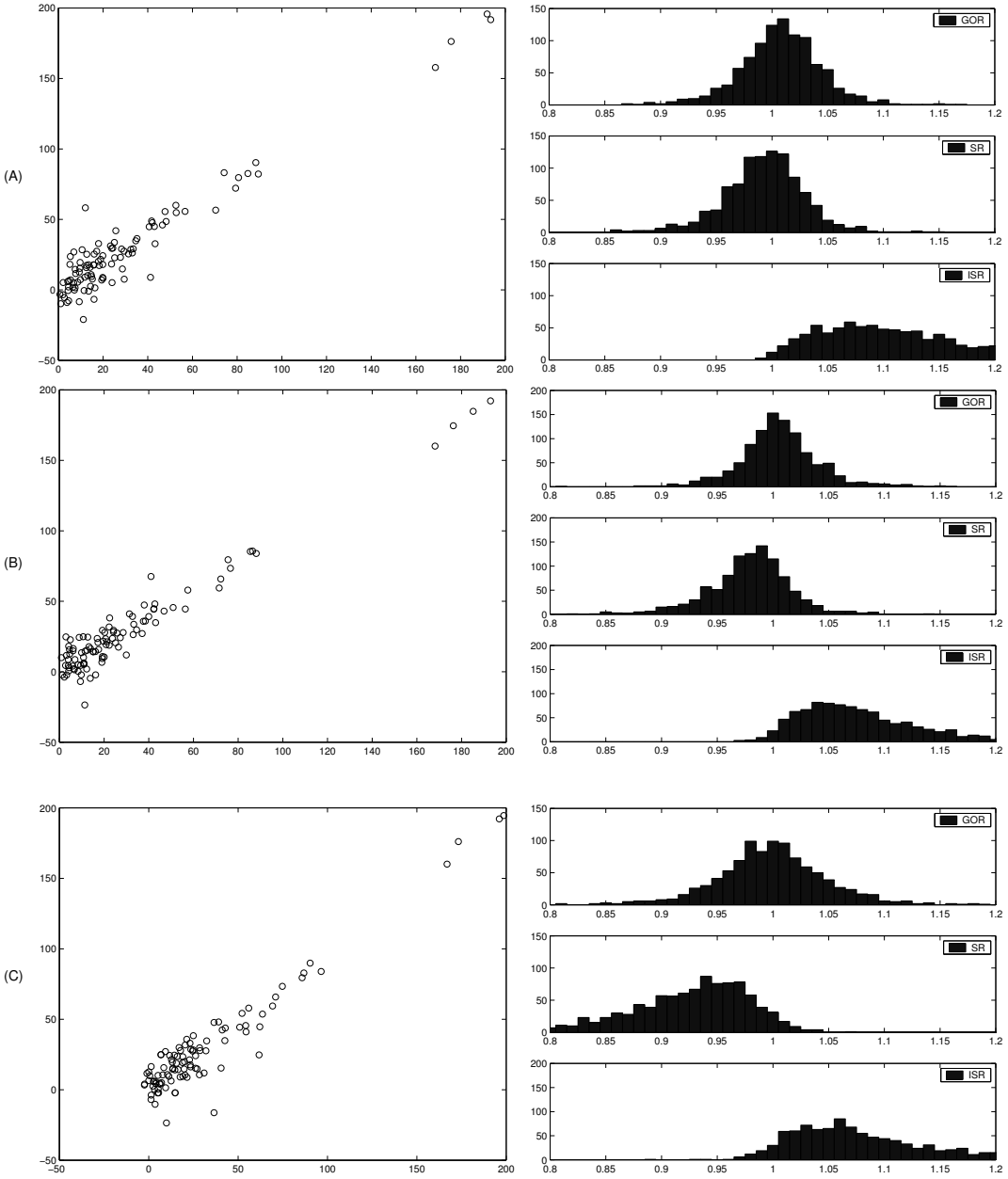

Figure 6:

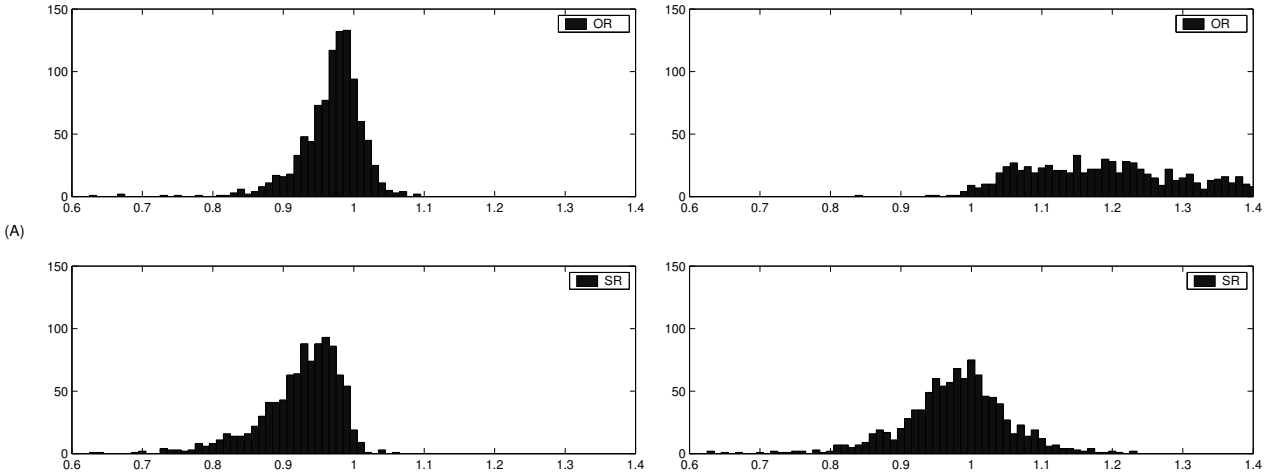

Figure 7: Examples taken from the 10^3 generations of 50 couples of points $(x,\, y)$. Data are log-normally distributed and variable log-normal errors have been added on the x and y axis. Frequency-distribution of the β coefficient found with OR and the SR method. **(A)** in the calculation η has been set to 1 instead of the true 0.25 value ($\sigma_y = 2$, $\sigma_x = 4$). **(B)** in the calculation η has been set to 1 instead of the true 25 value ($\sigma_y = 2$, $\sigma_x = 0.4$).

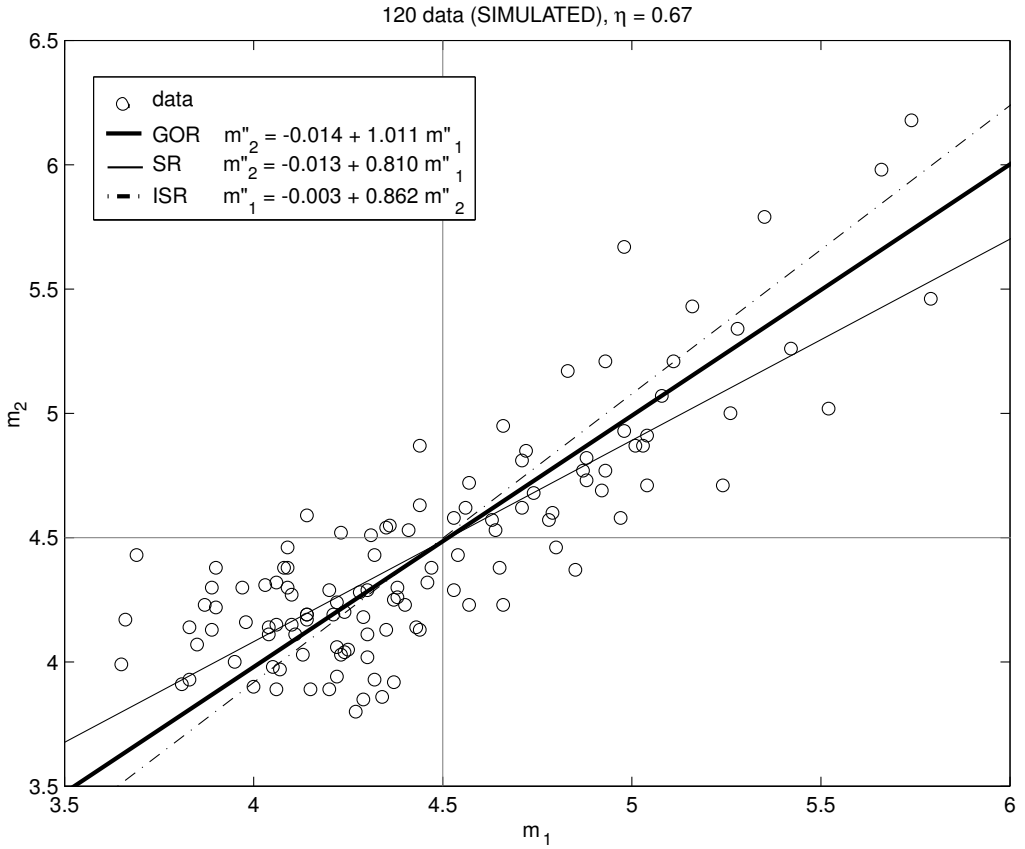

Figure 8: Standard and generalized orthogonal regression values for the 120 couples of simulated events distributed according to an exponential-distribution with parameter $B = 2.30$ and affected by Gaussian error with $\eta = 0.67$ ($\sigma_y = 0.18$, $\sigma_x = 0.22$). Results of the regressions on the whole data set. We define the reduced magnitude $m_1'' = m_1 - 4.5$ and $m_2'' = m_2 - 4.5$. Regression results are shown in inset: GOR $\rightarrow$ general orthogonal regression, SR $\rightarrow$ standard regression, ISR $\rightarrow$ inverted standard regression.

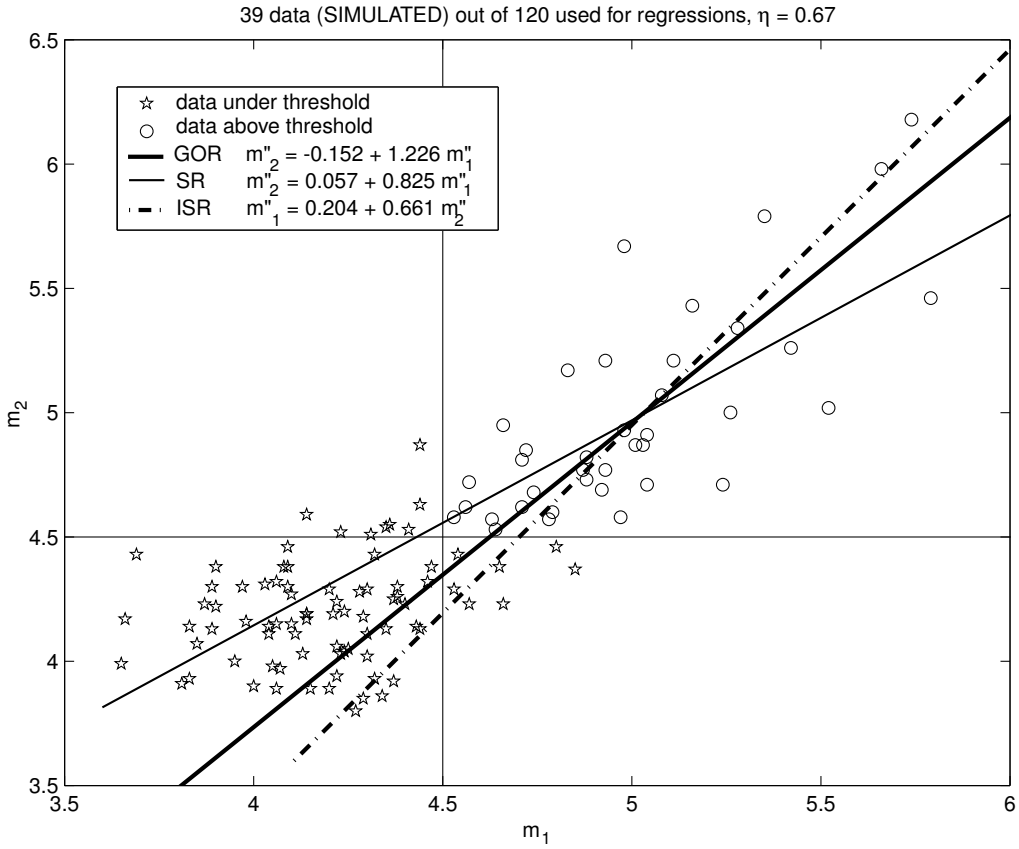

Figure 9: As in Figure 8 but regressions have been computed on data above the $m = 4.5$ threshold only.

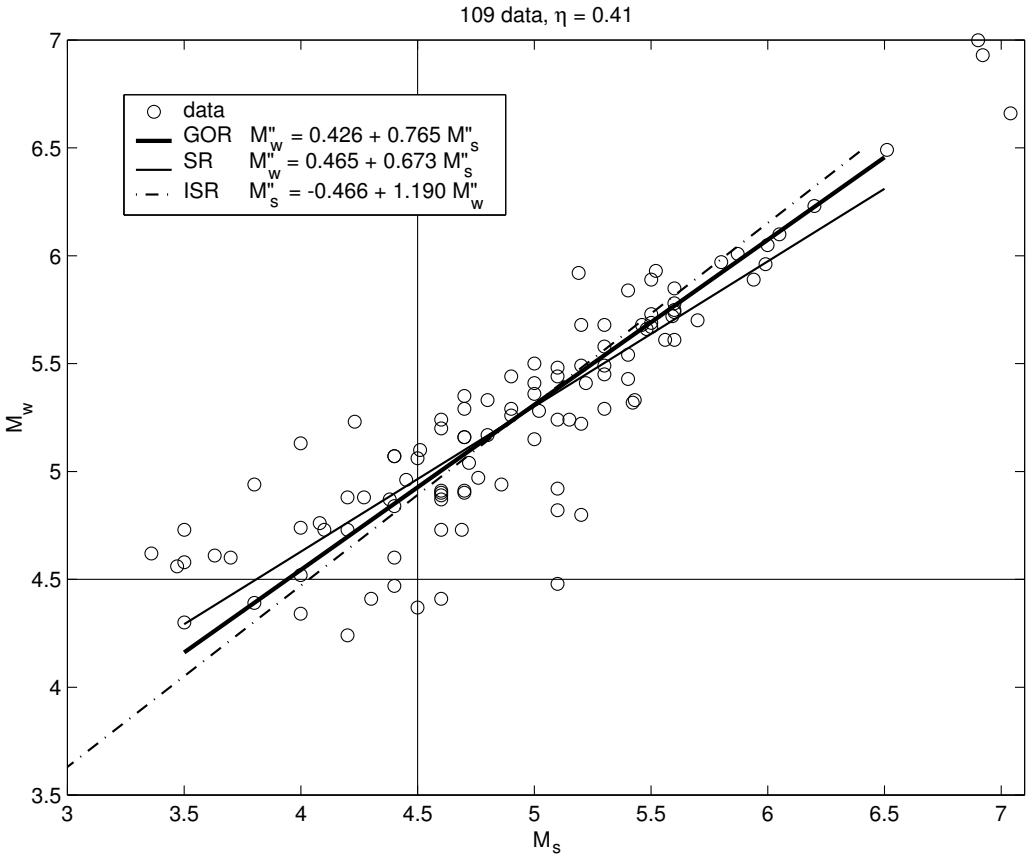

Figure 10: Standard and generalized orthogonal regression values for the 109 couples of events recorded in Italy between 1981 and 1996 gauged with M_w and M_s. We define the reduced magnitudes $M_s'' = M_s - 4.5$ and $M_w'' = M_w - 4.5$. Regression results are shown in inset: GOR $\rightarrow$ general orthogonal regression, SR $\rightarrow$ standard regression, ISR $\rightarrow$ inverted standard regression.

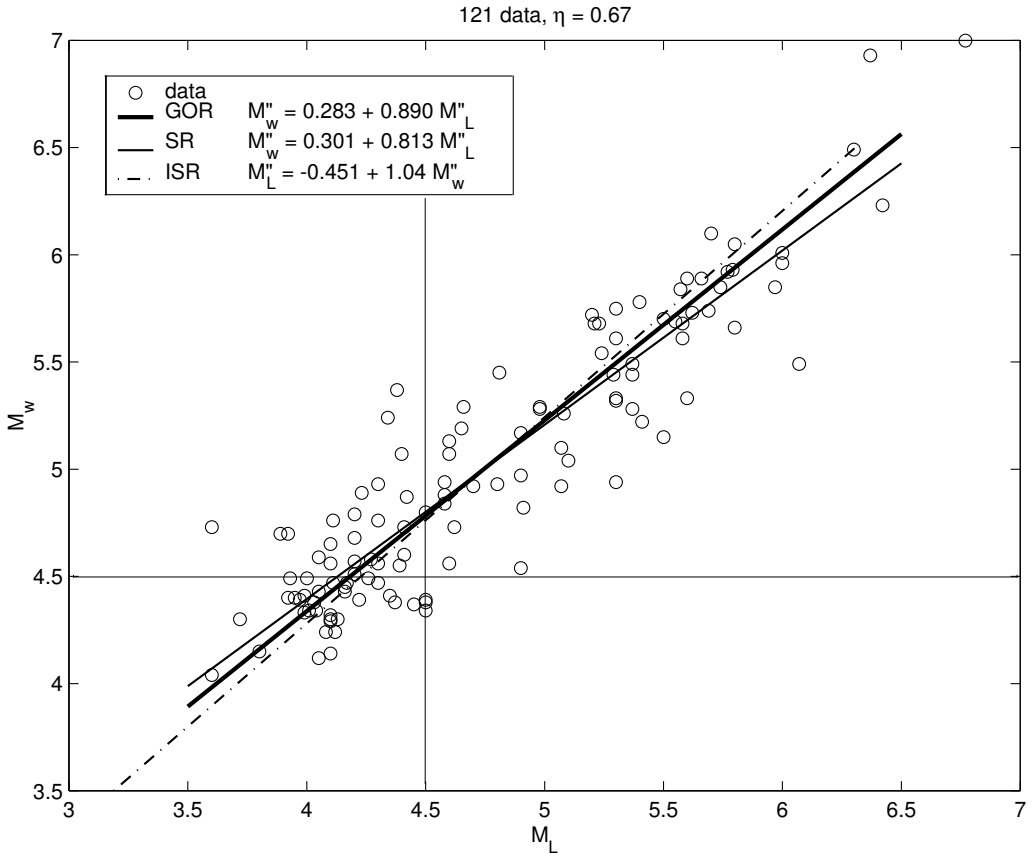

Figure 11: Standard and generalized orthogonal regression values for the 121 couples of events recorded in Italy between 1981 and 1996 gauged with M_w and M_L. We define the reduced magnitudes $M''_L = M_L - 4.5$ and $M''_w = M_w - 4.5$. Regression results are shown in inset: GOR $\rightarrow$ general orthogonal regression, SR $\rightarrow$ standard regression, ISR $\rightarrow$ inverted standard regression.

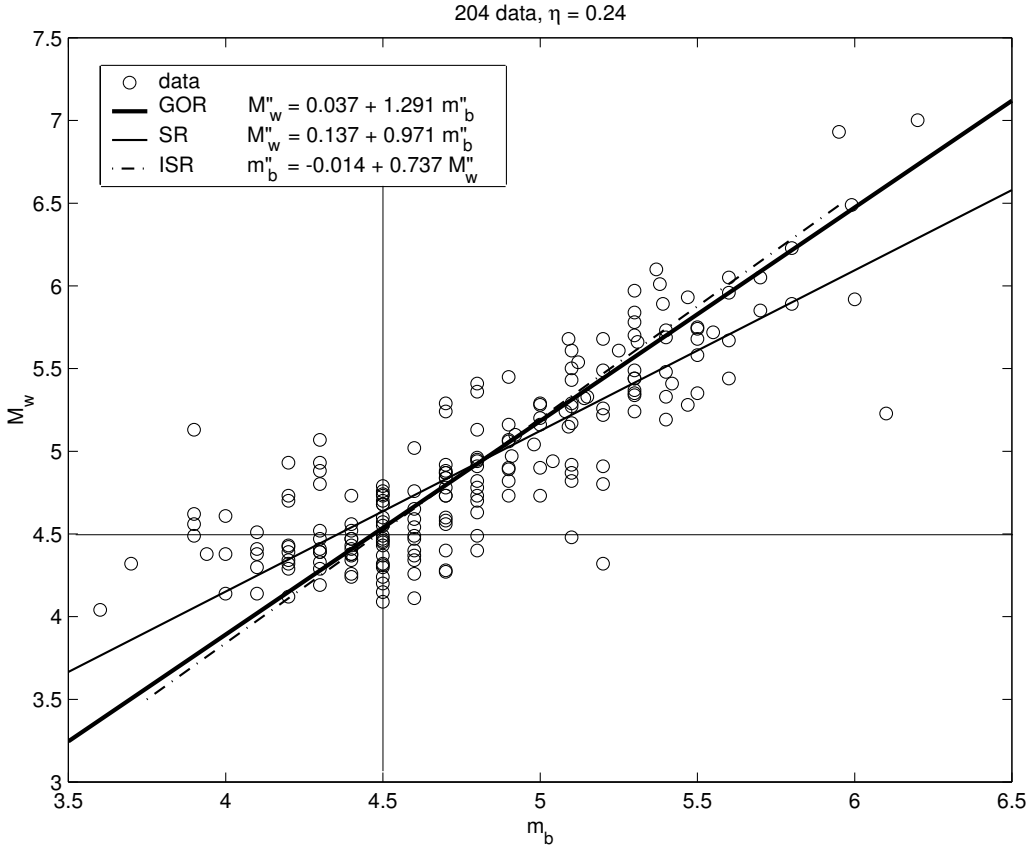

Figure 12: Standard and generalized orthogonal regression values for the 204 couples of events recorded in Italy between 1981 and 1996 gauged with M_w and m_b. Generalized orthogonal regression slope is much closer to the least square regression slope when m_b is on the y axis and treated as the variable affected by the largest error. We define the reduced magnitudes $m_b'' = m_b - 4.5$ and $M_w'' = M_w - 4.5$. Regression results are shown in inset: GOR $\rightarrow$ general orthogonal regression, SR $\rightarrow$ standard regression, ISR $\rightarrow$ inverted standard regression.

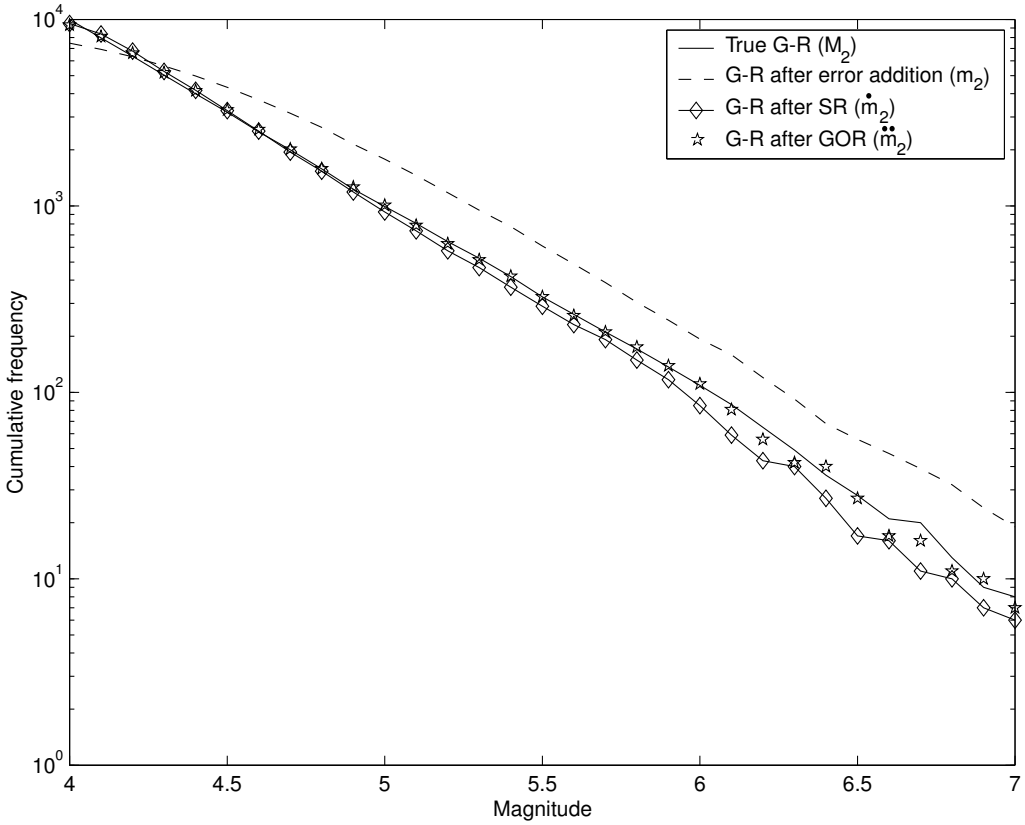

Figure 13: Cumulative frequency-magnitude plots (Gutenberg-Richter, G-R relations) of synthetic magnitude data (M_2) following an exponential distribution with parameter 2.3. Thick solid line: G-R of true data (M_2). Dashed line: G-R of M_2 data after Gaussian error addition (m_2) with $\sigma = 0.2$. Diamonds ($\diamondsuit$): G-R of $\dot{m}_2$ data, estimated from m_1 through the standard regression (SR, $\dot{m}_2 \leftarrow \alpha + \beta m_1$). Stars ($\star$): G-R of $\ddot{m}_2$ data, estimated from m_1 through the general orthogonal regression (GOR, $\ddot{m}_2 = \alpha + \beta m_1$) with true $\eta = 25$ ($\sigma_{m2} = 0.5$, $\sigma_{m1} = 0.1$).

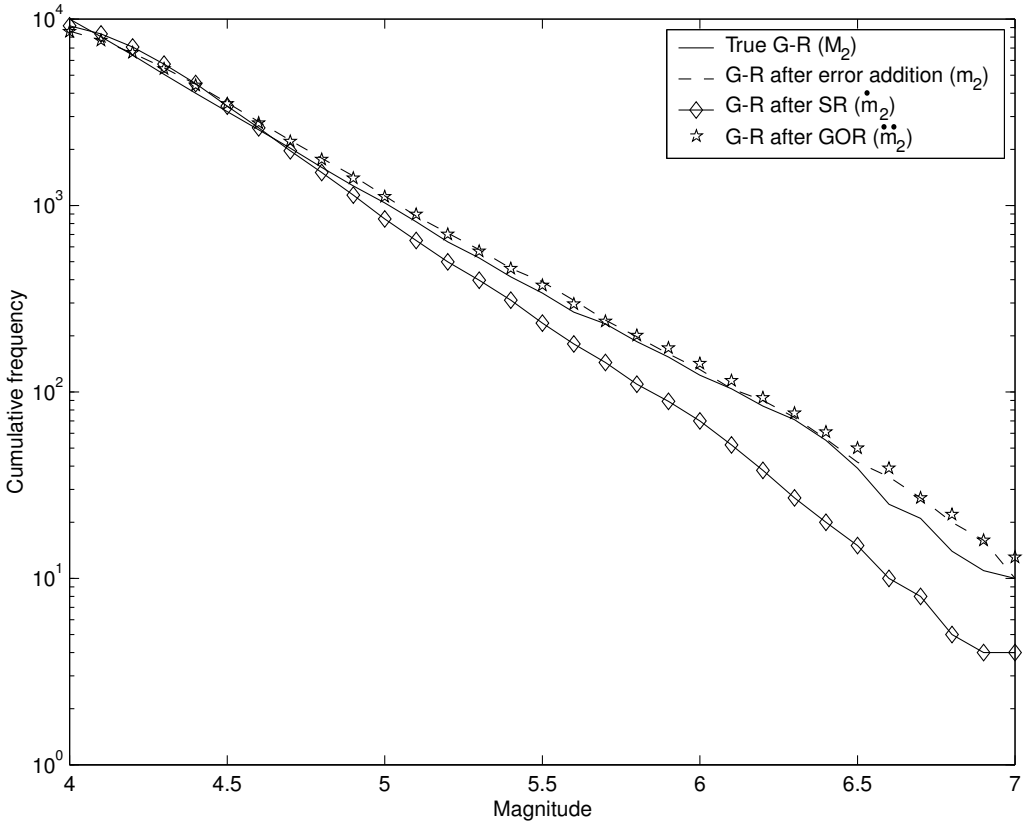

Figure 14: As in Figure 13 but stars ($\star$): G-R of $\ddot{m}_2$ data, estimated from m_1 through general orthogonal regression with $\eta = 1$ ($\sigma_{m2} = 0.2$, $\sigma_{m1} = 0.2$).

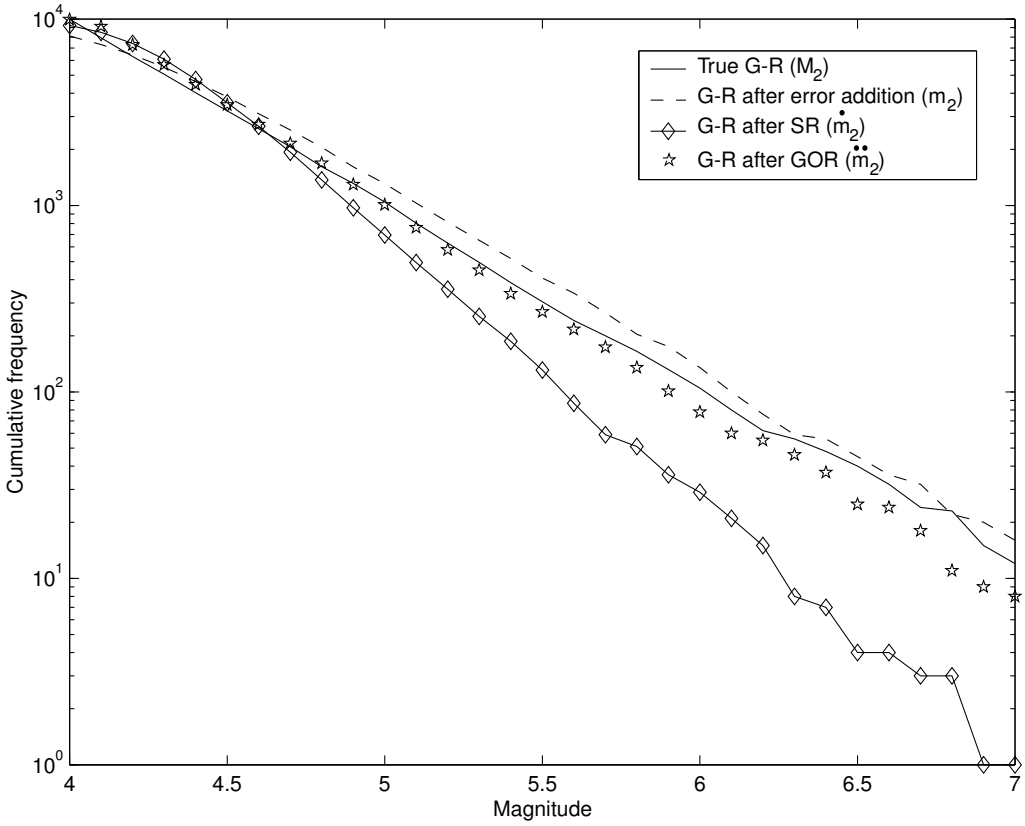

Figure 15: As in Figure 13 but stars ($\star$): G-R of $\ddot{m}_2$ data, estimated from m_1 through general orthogonal regression with $\eta = 0.25$ ($\sigma_{m2} = 0.2$, $\sigma_{m1} = 0.4$).