Far East Journal of Theoretical Statistics

Available online at http://pphmj.com/journals/fjts.htm
Volume …, Number …, 2014, Pages …

SOME PROPOSALS OF THEIL USED TO DISCRIMINATE BETWEEN A LINEAR LATENT GROWTH MODEL AND A LINEAR REGRESSION MODEL

Marco Barnabani

Dipartimento di Statistica, Informatica, Applicazioni

"G.Parenti" - Viale Morgagni, 59 -50134 Florence, Italy

Abstract

Theil [12] proposed three instruments to be applied in regression analysis with prior stochastic information. Two of them are: a quadratic form (known as compatibility test) and a measure of the share of sample information on posterior variance. Duly modified, they are proposed in the paper to investigate if data come from a linear regression model or from a linear latent growth model. We compare a covariance matrix due to sample information and a total variance due to both sample and prior information following two approaches: (i) by difference of two quadratic forms based on the structure of the Theil's compatibility test, (ii) by product of two covariance matrices based on a measure of an estimated share of sample information in total variability. The first approach depends on approximations to chi-square distributions which are discussed on the appendix. The second one is based on a comparison of empirical cumulative distribution functions. A simple algorithm based on the bootstrap is proposed for this comparison.

Received: December 23, 2013; Accepted: February 11, 2014

2010 Mathematics Subject Classification: 62F03, 91B62

Keywords and phrases: Linear Latent growth model, Linear Regression model, Empirical cumulative distribution function, chi-square approximation, quadratic form.

1. Introduction

A data set on Tourism in Tuscany (Italy) consists of the Index number (base year 2002) of accommodations (the response variable) on 260 Municipalities from 2003 to 2009. These data have been firstly processed so that to obtain homogeneous groups of units. The upper-left panel of Figure: 1 shows the graph of the data analyzed in the paper, composed of 95 "homogeneous" municipalities. It can be seen that each unit appears to have its own trajectory approximated by linear functions with specific intercept and slope that determine the trend. We define the model for the *i*-th municipality as $y_i = X\beta_i + u_i$ where X is a $T \times k$ matrix containing a column of ones and a column of constant time values; β_i is a column vector whose components are β_{0i} the intercept component of unit *i* and β_{1i} the slope component for unit *i*; u_i is $T \times 1$ column vector whose u_{ti} component is the measurement error at the time point *t*, for individual *i*. We assume $u_i \sim N(0, \sigma^2 I_T)$. The graph of Figure 1 shows that the trajectories are "high" or "low" suggesting two hypotheses from an economic point of view. One is that the growth of the tourism of each municipality at time *t* might see have mean $\theta_0 + \theta_1 t$ and vary about it according to the overall deviation u_{ti}. That is, a regional political economy determines the growth of the tourism of single municipalities. Statistically this is modeled with a vector of fixed population parameters, $\theta = [\theta_0 \ \theta_1]'$ which captures the regional political economy with the constraint $\beta_i = \theta$ for each trajectory. We refer to this model as (constrained) Linear Regression Model (LRM).

On the other hand, the graph of Figure 1 shows different steepness across municipalities, showing that the unit-specific intercepts β_{0i} and slopes β_{1i} are not fixed but they could vary across units suggesting a growth of the tourism influenced not solely by the regional political economy but also by specific characteristics of each municipality. This induces us to introduce a random component on the vectors β_i distinguishing the various trajectories. Then, we write $\beta_i = \theta + v_i$, $i = 1, ..., n$, $v_i = [v_{0i} \ v_{1i}]'$ where v_{0i} and v_{1i}

are unobserved random variables that configure individual growth. The base hypothesis on random component is $v_i \sim N(0, \Omega)$, Ω is a positive semi definite matrix $(\Omega \succeq 0)$, $u_i \perp v_i$, where the symbol $\perp$ indicates independence of random variables. The normality assumptions are introduced for testing purposes. We refer to this model as Linear Latent Growth Model (LLGM). Then, on the base of our data we ask whether the specificity of the municipalities contributes to the growth of the tourism other than the regional political economy. Statistically we ask whether it is more appropriate modeling data with a linear regression model $(\Omega = 0)$ or a linear latent growth model $(\Omega \succ 0)$. Of course, to discriminate between the two models the sign of the matrix Ω plays a crucial role. If Ω is a positive definite matrix, we can state that data come from a linear latent growth model. If $\Omega = 0$, then data come from a LRM. Unfortunately, Ω is unknown and must be estimated. In this case the implications $\hat{\Omega} \succ 0 \Rightarrow LLGM$ and $\hat{\Omega} \not\succ 0 \Rightarrow LRM$ are not always true. ($\not\succ 0$ is for non positive definiteness and $\hat{\Omega}$ is an estimate of Ω). This induced us to investigate alternative approaches not involving explicitly $\hat{\Omega}$. In the work of Theil [12] where the use of prior stochastic constraints in regression analysis is discussed, we found some “instruments” that properly modified allowed us to tackle the problem. One of these “instruments” is the quadratic form associated to the so-called Theil’s compatibility test. The other is a scalar measure of the relative contribution of sample information in total variability.

Let re-propose the Linear latent growth model as a Theil’s mixed model. By substitution of the latent part into the measurement part, we get

$$y_i = X\theta + \varepsilon_i, \quad \varepsilon_i \sim N(0, X\Omega X' + \sigma^2 I_T) \tag{1}$$

and

$$\bar{b} = \theta + (\bar{b} - \theta) = \theta + \omega, \quad \omega \sim N\left(0, \frac{1}{n}V\right), \tag{2}$$

where $b_i = (X'X)^{-1}X'y_i \sim N(\theta, V)$, $V = \sigma^2(X'X)^{-1} + \Omega$ and $\bar{b} = \frac{1}{n}\sum_{i=1}^{n} b_i$

$= (X'X)^{-1} X'\bar{y}$, with $\bar{y}$ a mean vector whose tth element, $\bar{y}_t = (1/n)\sum_{i=1}^{n} y_{it}$, is the mean of the observations at time t. We can look at (2) as to a stochastic prior information compatible with model (1). This way to define a stochastic prior information which is of sample origin can be traced back, for example, in the works of Lee and Griffiths [8] and Buse [2]. From model (1)-(2), we can define the following quadratic form

$$\frac{n}{n-1}(b_i - \bar{b})' V^{-1} (b_i - \bar{b}) \tag{3}$$

which has a structure similar to the compatibility test proposed by Theil and, if data come from a LLGM it is a χ^2 with k degrees of freedom. The sampling nature of prior information and the presence of Ω which is unknown make this quadratic form useless for inferential purposes and of uncertain interpretation if it is used as a compatibility test in the sense given by Theil.

If $\beta_i = \theta$ for each unit $(\Omega = 0)$, we face a set of linear regression models, $y_i = X\theta + u_i$ and $\bar{b} = \theta + \omega^*$, $\omega^* \sim N\left(0, \frac{\sigma^2}{n} I_T\right)$. In this case, the quadratic form (3) is given by

$$\frac{n}{n-1}(b_i - \bar{b})' \frac{X'X}{\sigma^2} (b_i - \bar{b}) \tag{4}$$

which is a χ^2 with k degrees of freedom if data come from a *LRM*.

As a first approach to investigate whether data come from a *LRM* (the null hypothesis, H_0) or from a *LLGM* (the alternative hypothesis, H_1) we propose a sample discriminating function based on the difference of estimated quadratic forms (3) and (4). To this end observe that the variance covariance matrix V can be seen as composed of two components: $\sigma^2(X'X)^{-1}$ which is the variance of b_i if data come from a *LRM* $(\Omega = 0)$

and Ω which is the variance of b_i if data come from prior information $(\sigma^2 = 0)$. Then, we propose to estimate V of the quadratic form (3) with the sample variance covariance matrix of b_i, S_b. This estimate has been proposed by several authors see for example Swamy [11], Judge et al. [7] and Hsiao and Pesaran [5]. By definition S_b can contain implicitly information on Ω without estimating it. If data come from a *LRM*, S_b is a consistent estimate of $\sigma^2(X'X)^{-1}$, if data come from an *LLGM* S_b is a consistent estimate *V*. As to σ^2 of the quadratic form (4) we adopt the estimator given by Swamy [11], $s^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2$, with $s_i^2 = \frac{(y_i - Xb_i)'(y_i - Xb_i)}{T-k}$.

In Section 2, we discuss the definition and the application of the sample discriminating function based on the difference of the estimated quadratic forms (3) and (4). To compare in homogeneous way the two quadratic forms, they are approximated by chi-square distributions which are discussed in the appendix.

As said previously a second purpose of Theil's paper was to propose a measure for the relative contribution of sample information on total variability. This measure is given by the trace (divided by the rank of *V*) of the product $\sigma^2(X'X)^{-1}V^{-1}$. This quantity which is based on the product of variance-covariance matrices, ranges between zero and one when data come from a *LLGM*, $(\Omega \succ 0)$, it is equal to one when data come from a *LRM* $(\Omega = 0)$. This characteristic allowed us to discriminate between a *LLGM* and a *LRM*. The share of sample information on total variance is proposed in Section 3 where an estimated measure based on S_b and s^2 is discussed. In Section 4 this quantity is used to test the hypothesis H_0 by comparing two empirical cumulative distribution functions (*ecdf*) one defined under the null hypothesis, the other under the alternative. A simple algorithm based on bootstrap is proposed for the comparison.

2. A Comparison between Chi-square Approximations

We refer to the upper-left panel of Figure 1 which shows the graph of the index number (base year 2002) of accommodations on 95 “homogeneous” Municipalities of Tuscany (Italy) from 2003 to 2009. As previously observed there is a considerable variability in the response variable both within each municipality and between municipalities. We ask whether it is more appropriate modeling data with a linear regression model or a linear latent growth model.

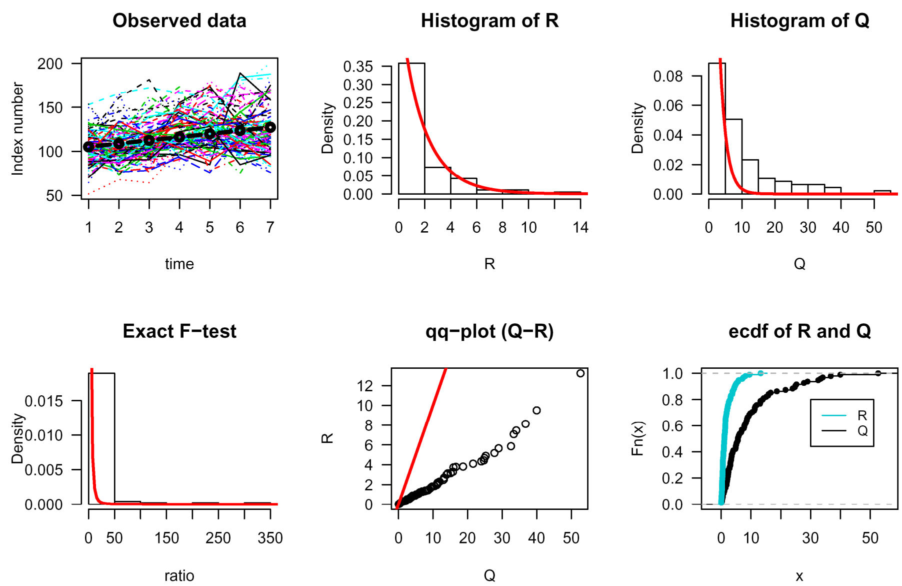

Figure 1. Comparison between chi-square approximations.

As suggested by several authors (for example Swamy [11]), an exact F-test can be derived from the residual sums of squares of the restricted $(H_0 : \beta_i = \theta \ \forall i)$ and unrestricted versions of the model (coefficient vectors are truly different across units). Under the null hypothesis, it is immediate to show that the average-trajectory $\bar{y} \sim N\left(X\theta, \frac{\sigma^2}{n} I_T\right)$, the ith trajectory $y_i \sim N(X\theta, \sigma^2 I_T)$ and $y_i - \bar{y} \sim N\left(0, \frac{n-1}{n}\sigma^2 I_T\right)$. Then,

$$\frac{(y_i - \bar{y})' P(y_i - \bar{y})}{\sigma^2 (n-1)/n} \sim \chi^2(k), \quad P = X(X'X)^{-1}X'.$$

Replacing σ^2 with the usual unbiased estimate we get the standard *F*-test,

$$\frac{n(T-k)}{k(n-1)} \frac{(y_i - \bar{y})' P(y_i - \bar{y})}{y_i'(I-P)y_i} \sim F(k, T-k). \tag{5}$$

The bottom-left panel of Figure 1 shows the histogram of the test statistic (5) and the density of the *F*-Fisher under the null hypothesis applied to data on tourism. The computed *p*-value is approximately zero emphasizing a strong evidence against the null Hypothesis. However, the evidence against the null hypothesis does not allow to state that data come from a linear latent growth model.

In an attempt to get a decision in favor of the alternative hypothesis, introduce the quadratic form $R_i/\lambda = \frac{n}{n-1}(b_i - \bar{b})' \frac{S_b^{-1}}{\lambda}(b_i - \bar{b})$, $i = 1, \ldots, n$, $\lambda > 0$. R_i/λ is the Theil' compatibility test with the matrix $s^2(X'X)^{-1} + \Omega$ replaced by the sample covariance matrix re-scaled by λ. We shall see later that the purpose of re-scaling S_b with λ is to obtain a better approximating distribution of the quadratic form.

S_b is both a consistent estimate of $\sigma^2(X'X)^{-1}$ if H_0 is true, and a consistent estimate of $\sigma^2(X'X)^{-1} + \Omega$ if H_1 is true. Then, S_b reflects the variability of the coefficients due to the sample information under H_0 and the variability due to both sample and prior information under H_1.

The distribution of R_i/λ is unknown. Some authors (for example (Johnson and Wichern [6]) propose to approximate R_i with a $\chi^2(k)$ without re-scaling the quadratic form $(\lambda = 1)$. However, a value of $\lambda \neq 1$ gives a better chi-square approximation. In the appendix the algorithm used to determine λ is described. The algorithm works by maximizing a constrained log-likelihood function, estimating the degrees of freedom (*df*) by the data. Various simulations show that *df* is centered around *k* which is the degrees of freedom suggested by Johnson and Wichern [6] to approximate the chi-square distribution. Then, we assume $\frac{R_i}{\lambda} \,\dot{\sim}\, \chi^2(df),\ \ i = 1, ..., n$ under both hypotheses. The symbol $\dot{\sim}$ is for "approximately distributed as". Also the approximation $\frac{R_i}{\lambda} \,\dot{\sim}\, \chi^2(k)$ could be used.

Because R_i conforms to both hypotheses, the set of observations $R_1/\lambda, ..., R_i/\lambda, ..., R_n/\lambda$ is not able to give enough information to discriminate about the two populations.

We propose to compare the re-scaled R_i with the re-scaled quadratic form

$$\frac{Q_i}{\xi} = \frac{n}{n-1}(b_i - \bar{b})'\left(\frac{X'X}{s^2}\right)(b_i - \bar{b})/\xi, \quad \xi > 0, \quad i = 1, ..., n$$

which is defined under the null hypothesis and can be traced back in the work of Swamy [11] where it was defined with $\xi = 1$. In small samples, the exact or approximate distribution of Q_i is not of practical importance because it depends on σ^2 whose value is usually unknown. However, specifying an appropriate value of ξ could justify the use of an approximate chi-square distribution. In the appendix, we propose $\xi = \frac{n-1}{n}\frac{df}{df - 2}$ with $df = n(T - k)$. Therefore, under the null hypothesis, we assume $\frac{Q_i}{\xi} \,\dot{\sim}\, \chi^2(k)$, $i = 1, ..., n.$ Under the alternative hypothesis, a chi-square approximation to $\frac{Q_i}{\xi}$ is difficult to justify because the unknown eigenvalues of the matrix Ω in the metric $X'X$ are involved other than σ^2. In this case the approximation is affected both by the relative sizes of the eigenvalues and by their variability if an estimate is used. In this case, we leave unknown the distribution.

The following table summarizes the result.

	Approximation of re-scaled	
	R_i	Q_i
H_0 is True	$\chi^2(df)$	$\chi^2(k)$
H_1 is True	$\chi^2(df)$	NOT $\chi^2(k)$

Therefore, we propose to compare two set of observations, $Q = \{Q_1/\xi, ..., Q_i/\xi, ..., Q_n/\xi\}$ and $R = \{R_1/\lambda, ..., R_i/\lambda, ..., R_n/\lambda\}$ through the following sample discriminating function (Anderson [1]),

$$D_i = \frac{n}{n-1}(b_i - \bar{b})'\left(\frac{X'X}{\xi s^2} - \frac{S_b^{-1}}{\lambda}\right)(b_i - \bar{b}), \quad i = 1, ..., n \tag{6}$$

to classify into H_1 if (6) is large and into H_0 if (6) is small.

The top-middle and the top-right panels of Figure 1 show the histograms of R and Q respectively and relative chi-square approximations (continuous lines). A visual comparison of the two graphs suggests an evidence in favor of the alternative hypothesis in confirmation of the comments on the *F*-test.

If H_0 were true, then the difference $Q_i/\xi - R_i/\lambda$ should reflect a random variability. Under H_1, Q_i/ξ is (stochastically) greater than R_i/λ. Even though Q and R are not a set of independent random variables (only $n-1$ are independent) we apply the Kolmogorov-Smirnov test trying to determine if the two “samples” are “the same”. The test applied to the data shows a *p*-value approximately equal to zero emphasizing that they differ significantly. The qq-plot of Figure 1 and a comparison between the two “samples” Q and R through the empirical cumulative distribution functions (*ecdf*) emphasizes this aspect.

3. Estimated Share of Sample Information in Total Variability

Given the identity $[s^2(X'X)^{-1} + \Omega][s^2(X'X)^{-1} + \Omega]^{-1} = I_k$, by applying the trace operator and dividing by k, we get

$$\theta_S = \frac{1}{k} tr s^2 (X'X)^{-1}[s^2(X'X)^{-1} + \Omega]^{-1} \tag{7}$$

and $\theta_P = 1 - \theta_S$ which are taken as a measure of the shares of sample and prior information, respectively, in the total variance. This way of measuring the shares have been firstly proposed by Theil [12] which discussed “reasonable” requirements they satisfy.

In large samples $plim\ s^2 = \sigma^2$ and θ_S ranges from 0 to 1. A value of $\theta_S < 1$ implies $\Omega \succ 0$ and denotes the presence of randomness on the coefficients. The closer θ_S is to 1 the less is the influence of randomness of the coefficients on the data. When $\theta_S = 1$, $\Omega = 0$ and data are generated by a model with no random coefficients. If $\sigma^2 = 0$, $\theta_S = 0$ and data come from prior information.

In small samples because of the presence of the estimate s^2, the above relationships are not exact but valid with probability close to one.

The typical behavior of (7) (and θ_P) obtained by simulation, is shown in the left panel of Figure 2 where the x-axis represents the population standard deviation ranging from 1 to 300 and the y-axis represents the shares. The matrix Ω is assumed known and fixed. Data are generated with a *LLGM*.

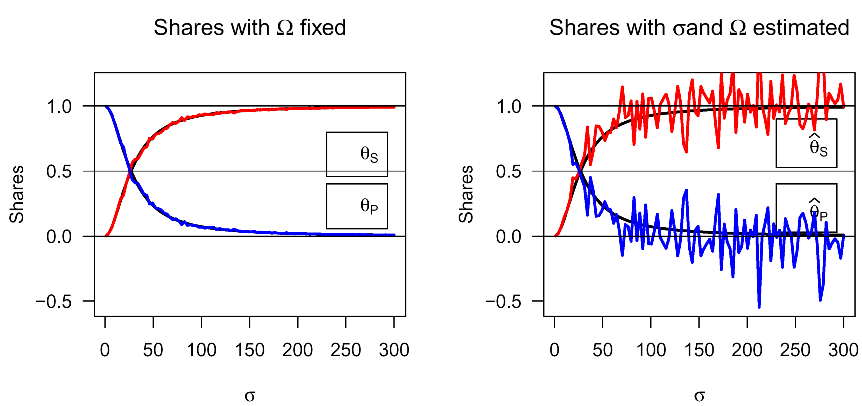

Figure 2. Behavior of the share of sample information on total variance.

In applications, one needs to replace θ_S with an estimate because Ω is unknown. We propose to estimate the shares in total variability as $\hat{\theta}_S = \frac{1}{k} trs^2(X'X)^{-1}S_b^{-1}$ and $\hat{\theta}_P = 1 - \hat{\theta}_S$. The right panel of Figure 2 shows the plot of the estimates $\hat{\theta}_S$ and $\hat{\theta}_P$ from which it emerges the large variability of the shares when Ω is unknown. This variability has consequences on the interpretability of $\hat{\theta}_S$. When σ^2 increases $\hat{\theta}_S$ is no more in the interval $(0, 1)$ but can assume values greater than 1 with consequences on the sign of the matrix $\hat{\Omega}$. More precisely, $\hat{\theta}_S \geq 1 \nRightarrow \hat{\Omega} \not\succeq 0$ and $\hat{\theta}_S < 1$ does not imply that $\hat{\Omega}$ is a positive definite matrix. Therefore, we are not able to associate directly the sign of the matrix $\hat{\Omega}$ to the data generating model. In other terms, the implications $\hat{\theta}_S < 1 \Rightarrow \hat{\Omega} \succ 0 \Rightarrow LLGM$ and $\hat{\theta}_S \geq 1 \Rightarrow \hat{\Omega} \nsucc 0 \Rightarrow LRM$ are not always true. This is especially true when $\hat{\theta}_S$ is close to 1. In this case it can happen that either data come from a *LLGM* but the sample information is large dominating prior information, or data come from a linear regression model without randomness on the coefficients. Therefore, the index $\hat{\theta}_S$ is not able to help to discriminate between a linear latent growth model and a linear regression model. Instead we propose the compare the empirical cumulative distribution function of $\hat{\theta}_S$ under H_0 and H_1. This approach will be discussed in next section.

4. A Comparison between *ecdf*

Let investigate by simulation the distribution of $\hat{\theta}_S$ under the hypothesis that data come from a *LRM*. The top-left and top-right panels of Figure 3 show the box-plots of the simulated $\hat{\theta}_S$ for different values of σ and different values of β. In this investigation it emerges a “stability” of the empirical distributions changing the parameters of the simulation procedure.

A visual inspection induces us to look at the box-plots as describing data coming from a common unknown distribution. To support this statement we tested the null hypothesis of a common unknown distribution through the *K*-sample Anderson-Darling test statistic (Scholz and Stephens [9]), obtaining strong evidences in favor of the claim.

This fact allows us to construct the empirical distribution function of $\hat{\theta}_S$ by re-sampling from a standard normal distribution taking it as an estimate of the common unknown distribution under the null hypothesis. We call prototype $\hat{\theta}_S$ the estimate whose distribution is obtained by re-sampling from a standardized normal distribution without any hypothesis on β and σ^2. The histogram and the empirical cumulative distribution function of the prototype $\hat{\theta}_S$ are shown in the bottom panels of Figure 3.

This availability offers the possibility to compare the empirical cumulative distribution functions of the observed $\hat{\theta}_S$ with the prototype $\hat{\theta}_S$ so that a (probabilistic) decision on the data generating process can be done. More precisely, we propose to compare the empirical cumulative distribution function of the prototype $\hat{\theta}_S$ (which can be considered as “fixed”) with the empirical cumulative distribution function of the observed $\hat{\theta}_S$ which depends on the data. If the two *ecdf* are close to one another then we state that data come from a *LRM*. Otherwise if the *ecdf* of the observed $\hat{\theta}_S$ is stochastically greater than the prototype $\hat{\theta}_S$ the hypothesis H_1 is accepted.

Figure 4 shows the behavior of the two *ecdf* when the weight of the sample information on the total variability increases. This is done by simulation holding Ω fixed and increasing σ^2.

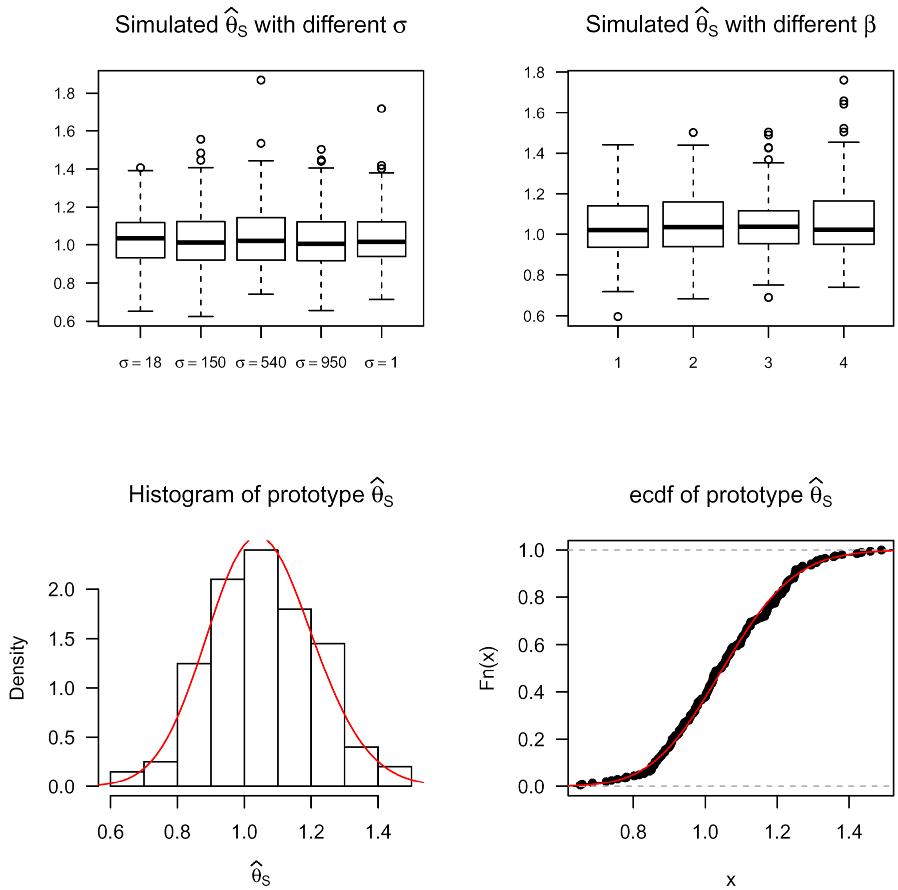

Figure 3. Distribution function of prototype $\hat{\theta}_S$.

From top to bottom, the *ecdf* of the simulated $\hat{\theta}_S$ approaches more and more to the (“fixed”) prototype $\hat{\theta}_S$. Looking at Figure 4, a comment on the bottom panels is necessary. The bottom-left panel is simulated via a linear latent growth model with $\sigma = 180$. In this case, the weight of the sample information on the total variance is close to 1 and the influence of prior information are practically zero. The bottom-right panel is simulated via a *LRM* and prior information are zero. Both show $\hat{\theta}_S \simeq 1$ and the empirical cumulative distribution functions of the observed and prototype $\hat{\theta}_S$ are not distinguishable. In this case, any decision on the data generating model is difficult and a *LRM* could be used in applications assuming non influential prior information.

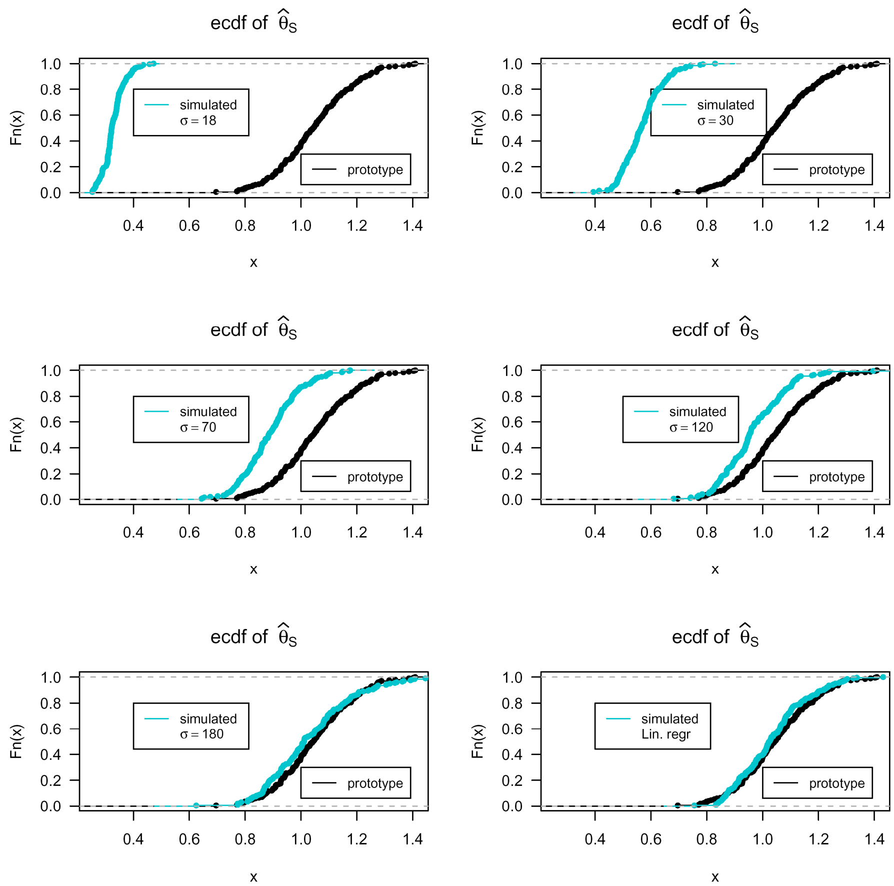

Figure 4. Simulations of observed and prototype $\hat{\theta}_S$.

To quantify the uncertainty due to the comparison of the *ecdf*, we use the Kolmogorov-Smirnov test and the two sample Anderson-Darling test. The test applied to the simulated data of Figure 4 gives a *p*-value approximately equal to zero in the first four panels while in the last two panels the *p*-value is approximately the same, at about 22 confirming the difficulty to choose the model in this case.

Working with observed data, the two empirical cumulative distribution functions can be computed as follows:

Step 1. Re-sample *n* times from a standardized normal distribution.

Step 2. Construct the *ecdf* of prototype $\hat{\theta}_S$.

Step 3. Apply a bootstrap approach to get the *ecdf* of the observed $\hat{\theta}_S$.

Step 4. Compare the two *ecdf*.

This algorithm has been applied to the data on the Tourism. Figure 5 shows the result. The two *ecdf* are strongly different and the observed *ecdf* of $\hat{\theta}_S$ is stochastically greater than the *ecdf* of prototype $\hat{\theta}_S$. We can state that data come from a linear latent growth model. Of course we applied the Kolmogorov-Smirnov test which gives a *p*-value equal to zero.

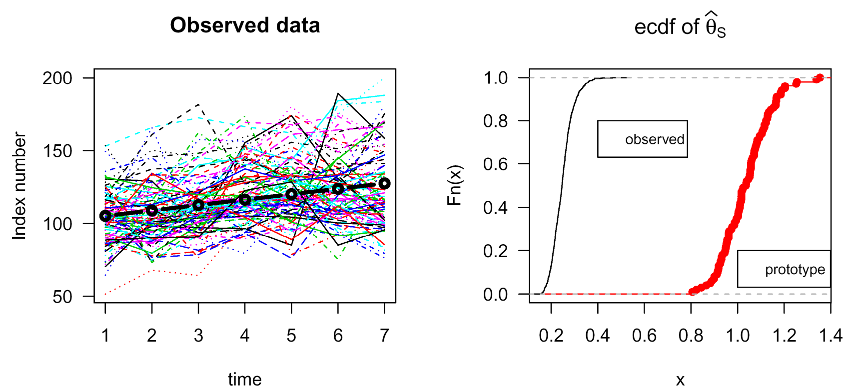

Figure 5. Comparison between *ecdf* of observed and prototype $\hat{\theta}_S$.

5. Conclusions

The paper has discussed two approaches derived from two proposals of Theil [12] to discriminate between a *LRM* (the null hypothesis, H_0) and a *LLGM* (the alternative hypothesis, H_1).

The two approaches are a comparison between quadratic forms approximated by chi-square distributions and a comparison between empirical cumulative distribution functions of a measure of the share of the sample information on total variability defined under H_0 and H_1. In the first approach, a crucial role is played by the approximated chi-square distributions which are discussed and justified in the appendix where an algorithm to obtain the approximations is described. In the second approach, a bootstrap procedure is used to obtain the empirical cumulative distribution of the observed $\hat{\theta}_S$ and it is compared with the *ecdf* of $\hat{\theta}_S$ under H_0. Our study has shown that if data come from a *LRM* the distribution of $\hat{\theta}_S$ does not depend on the parameters of the population and can be considered as “fixed”. Finally a simple algorithm used to compare the two *ecdf* is proposed.

Appendix: Chi-square Approximations to Quadratic Forms

Let $b - M(b) \sim N(0, H_n \otimes A)$ where $H_n = \left(I_n - \frac{1}{n} 1_n 1'_n\right)$ is the centering matrix, $b = [b'_{.1}, ..., b'_i, ..., b'_{.n}]'$ the $nk \times 1$ vector of estimates and $M(b) = [\bar{b}', ..., \bar{b}']'$ is the $nk \times 1$ mean vector, $\bar{b} = \frac{1}{n} \sum_{i=1}^{n} b_i$. Consider the quadratic form $Q_i = (b_i - \bar{b})' \frac{(X'X)}{s^2} (b_i - \bar{b})$ and investigate the distribution with two different definitions of A.

(i) $A = \sigma^2 (X'X)^{-1}$,

For large n, $s^2 \to \sigma^2$ and $H_n \to I_n$. In this case, $Q_i \sim \chi^2(k)$. Let see Q_i as a Gamma distribution with shape parameter $\alpha = k/2$ and scale parameter $\delta = 2$.

In small sample, the distribution of Q_i is unknown and an approximation is needed. To this end let write Q_i as $Q_i = (b_i - \bar{b})' \frac{(P'P)}{s^2} (b_i - \bar{b})$ where P is a non singular matrix such that $P(X'X)^{-1} P' = I$. Then, $Q_i = \frac{y'_i y_i}{s^2}$ with

$y_i \sim N\left(0, \frac{n-1}{n}\sigma^2 I\right)$. This implies that $Q_i \sim \sum_{j=1}^{k} w_j \chi^2(1)$ with $w_j = \frac{n-1}{n}\frac{\sigma^2}{s^2}$. That is, given s^2, Q_i is distributed as a linear combination of independent chi-square variates, each with one degree of freedom.

The exact distribution of a linear combination of independent chi-square variates is difficult to obtain in general and various approximations to its distribution have been proposed (Solomon and Stephens [10]).

Two relatively simple ones are widely used in practice and have been implemented in popular softwares. They work by assuming that Q_i is approximately distributed as a χ^2 just like each of the variables in the linear combination. One approximation known as mean correction, works by rescaling the quadratic form by referring to $\xi^{-1}Q_i$ in such a way that $E(\xi^{-1}Q_i) = E(\chi^2(k)) = k$. In our case because the weights of the linear combination, w_j, do not depend on the index j, we have $Q_i \sim \frac{n-1}{n}\frac{\sigma^2}{s^2}\chi^2(k)$. Then, $\frac{n-1}{n}\frac{\sigma^2}{s^2}$ can be seen as a correction such that $E\left(\frac{n}{n-1}\frac{s^2}{\sigma^2}Q_i\right) = k$.

The other approximation is based on a more sophisticated correction where both the scale and the degrees of freedom are adjusted by matching the first two moments of the quadratic form with those of a chi-square distribution (Yuan and Bentler [13]).

The performance of the two approximations depends on the corrections made which, in turn, depend on unknown quantities. When these corrections can be consistently estimated the approximations are valid asymptotically. In small sample it is not clear how to proceed in estimating the corrections and how to approximate and evaluate the performance of the quadratic form.

In our case, $\frac{n-1}{n}\frac{\sigma^2}{s^2}$ is a random variable involving an unknown parameter, σ^2. To approximate the distribution of Q_i we proceed as follows.

Write $Q_i \sim \text{Gamma}\left(\alpha = k/2,\ \hat{\delta} = 2\frac{n-1}{n}\frac{\sigma^2}{s^2}\right)$. We propose to replace $\hat{\delta}$ with $E(\hat{\delta})$. Then, $Q_i \approx \text{Gamma}(\alpha = k/2,\ \delta = E(\hat{\delta}))$ where the symbol $\approx$ is for "approximately distributed as". Observe that $\frac{\sigma^2}{s^2 df}$ is an inverse-$\chi^2(df)$. This implies that $E(\hat{\delta}) = 2\frac{n-1}{n}\frac{df}{df-2}$. Then,

$$Q_i \approx \frac{n-1}{n}\frac{df}{df-2}\text{Gamma}(\alpha = k/2,\ \delta = 2)$$

letting $\xi = \frac{n-1}{n}\frac{df}{df-2}$, the approximation used is $\frac{Q_i}{\xi} \approx \chi^2(k)$. Of course if *n* is large $\xi \to 1$ and Q_i is an exact χ^2 distribution as seen above.

Observe that in small samples $Q_1/\xi, ..., Q_i/\xi, ..., Q_n/\xi$ is no more a set of independent random variables. Only $n-1$ are independent.

(ii) $A = \sigma^2(X',\ X)^{-1} + \Omega$

Carrying out as in the previous section, there exists a non singular matrix, P such that $P(X'X)^{-1}P' = I$ and $P\Omega P' = D$, then the distribution of the quadratic form $Q_i = (b_i - \bar{b})'\frac{(X'X)}{s^2}(b_i - \bar{b})$ is a linear combination of chi-square variates with weights $w_j = \frac{n-1}{n}\frac{\sigma^2 + d_j}{s^2}$ where d_j are the eigenvalues of Ω in the metric $(X'X)$. In this case, if $A = \sigma^2(X'X)^{-1} + \Omega$, an approximate chi-square distribution of Q_i is difficult to accept given the presence of the unknown eigenvalues of Ω. In this case, we leave the distribution unknown.

Consider now the quadratic form

$$Q_{S_b} = (b - \bar{b})'(I \otimes S_b^{-1})(b - \bar{b}) = \sum_{i=1}^{n} (b_i - \bar{b})' S_b^{-1}(b_i - \bar{b}) = \sum_{i=1}^{n} R_i,$$

S_b is an estimate both of $A = \sigma^2(X'X)^{-1}$ and $A = \sigma^2(X'X)^{-1} + \Omega$ depending on the data generating model. Let justify the approximation of the quadratic form used.

Of course, the exact distribution both of Q_{S_b} and of its components, R_i, are difficult to obtain in general. There exists a nonsingular matrix P such that $A^{-1} = P'P$ and $S_b^{-1} = P'DP$ where D is a diagonal matrix whose diagonal elements, d_j, $j = 1, ..., k$ are the eigenvalues of AS_b^{-1}. Replacing the matrix S_b^{-1}, we get

$$R_i = (b_i - \bar{b})' P'D(b_i - \bar{b}) = y_i' D_{y_i}, \quad y_i \sim N\left(0, \frac{n-1}{n} I\right),$$

then $R_i \sim \sum_{j=1}^{k} \lambda_j \chi^2(1)$, $\lambda_j = \frac{n-1}{n} d_j$, is distributed as a linear combination of independent chi-square variates, and Q_{S_b} is the sum of dependent quadratic forms.

We propose to approximate R_i as a χ^2 just like each of the variables in the linear combination as follows. Let see the distribution of R_i as a linear combination of independent gamma variates each with the same shape but different estimated scale parameter, $R_i \sim \sum_{j=1}^{k} \text{Gamma}(\alpha = 1/2, \delta = 2\lambda_j)$. If we replace λ_j with a scalar $\lambda > 0$, then we have $R_i \sim \lambda\, \text{Gamma}(\alpha = k/2, \delta = 2)$ which allows us to define the following approximation

$$(b_i - \bar{b})' \frac{S_b^{-1}}{\lambda} (b_i - \bar{b}) \sim \chi^2(df), \quad i = 1, ..., n. \tag{8}$$

Of course the problem is how to determine the scalar λ so that the above χ^2 approximation is obtained. To determine λ we propose to proceed as follows. Look for (iteratively) a parameter λ so that the likelihood of $R_1/\lambda, ..., R_i/\lambda, ..., R_n/\lambda$ is a gamma distribution with scale parameter approximately equal to two and shape parameter freely estimated by the data. The scalar λ is linked to the weights λ_j and in turn to the scale parameter of the gamma distribution. More precisely, the scale parameter of the approximated gamma distribution is inversely related to λ. This inverse relationship allows one to jump from λ to the scale parameter and vice-versa so that to reach iteratively an (fractional) estimated shape parameter compatible with a scale parameter approximately equal to two. This approach allows us to use the approximating distribution (8), with λ and *df* estimated through the following simple algorithm:

$\lambda \leftarrow 1, \delta \leftarrow 2.5$
while $\delta > 2.001$ or $\delta < 1.999$ **do**
  for $i = 1 \rightarrow n$ **do**
    compute the quadratic form for each unit
  end for
  estimate scale and shape parameter by ML
  if $\delta > 2.001$ **then**
    increase λ
  end if
  if $\delta < 1.999$ **then**
    reduce λ
  end if
end while

Of course λ and *df* are random variables and their aim is to absorb the randomness contained in the scale parameter which is constrained to be 2 allowing to obtain an approximating chi-square distribution. Moreover, the use of a re-scaled chi square distribution and the algorithm used has also the effects of capture and “approximate” the independence of the random variables R_i (Chuang and Shih [3]).

It is well known that MLE's shape is upward biased then, actually, in our applications, we used the following biased-adjusted estimator proposed by Giles and Feng [4]:

$$\tilde{\alpha} = \alpha - \frac{[\alpha(\Psi_{(1)}(\alpha) - \alpha\Psi_{(2)}(\alpha)) - 2]}{2n[\alpha\Psi_{(1)}(\alpha) - 1]^2},$$

where $\Psi_{(i)}(\alpha) = \partial^i \Psi(\alpha)/(\partial\alpha^i)$, $i = 1, 2$ and $\Psi(\alpha)$ is the usual digamma function. The degrees of freedom are then estimated as $df = 2\tilde{\alpha}$.

References

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd ed., John Wiley, New York, 2003.

[2] A. Buse, On the equivalence of pooled and mixed estimation, Statist. Papers 38 (1997), 231-241.

[3] L. L. Chuang and Y. S. Shih, Approximated distributions of the weighted sum of correlated chi-squared random variables, J. Statist. Plann. Inference 142 (2012), 457-472.

[4] D. E. Giles and H. Feng, Bias of the maximum likelihood estimators of the two parameter gamma distribution revisited, University of Victoria, Dept. of Economics, Econometric Working Paper EWP0908, 2009.

[5] C. Hsiao and M. H. Pesaran, Random coefficient panel data models, CESifo Working Paper n. 1233, 2004.

[6] R. A. Johnson and D. W. Wichern Applied multivariate statistical analysis, 2007.

[7] G. Judge, R. Griffiths, H. Hill, H. Lutkepohl and T. Lee, The Theory and Practice of Econometrics, 2nd ed., Wiley, New York, 1985.

[8] L. F. Lee and W. E. Griffiths, The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models, Center for Economic Research, Dept. of Economics, University of Minnesota, Discussion Paper No. 79, 1979.

[9] F. Scholz and M. A. Stephens, *K*-sample Anderson-Darling tests, J. Amer. Statist. Assoc. 82(399) (1987) 918-924

[10] H. Solomon and M. A. Stephens, Distribution of a sum of weighted chi-square variables, J. Amer. Statist. Assoc. 72 (1977), 881-885.

[11] P. Swamy, Efficient inference in a random coefficient regression model, Econometrica 38 (1970), 311-323.

[12] H. Theil, On the use of incomplete prior information in regression analysis, J. Amer. Statist. Assoc. 58 (1963), 401-414.

[13] K. Yuan and P. M. Bentler, Two simple approximations to the distributions of quadratic forms, British J. Math. Statist. Psych. 63(2) (2010), 273-291.

<table>
<tr>
<td>

Paper No. PPH-1312080-TS

Kindly return the proof after correction to:

The Publication Manager
Pushpa Publishing House
Vijaya Niwas
198, *Mumfordganj*
Allahabad-211002 (*India*)

along with the print charges* by the fastest mail

***Invoice attached**

</td>
<td>

Proof read by: ....

Copyright transferred to the Pushpa Publishing House

Signature: ....

Date: ....

Tel: ....

Fax: ....

e-mail: ....

Number of additional reprints required

....

Cost of a set of 25 copies of additional reprints @ Euro 12.00 per page.

(25 copies of reprints are provided to the corresponding author ex-gratis)

</td>
</tr>
</table>