MANAGEMENT SCIENCE
Vol. 56, No. 4, April 2010, pp. 699–711
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5604 | 0699

DOI 10.1287/mnsc.1090.1123

The Analytic Hierarchy Process and the Theory of Measurement

Michele Bernasconi
Dipartimento di Scienze Economiche and SSE, Università "Ca' Foscari," I-30121 Venezia, Italy, bernasconi@unive.it

Christine Choirat
Department of Quantitative Methods, School of Economics and Business Management, Universidad de Navarra, E-31080 Pamplona, Spain, cchoirat@unav.es

Raffaello Seri
Dipartimento di Economia, Università dell'Insubria, I-21100 Varese, Italy, raffaello.seri@uninsubria.it

The analytic hierarchy process (AHP) is a decision-making procedure widely used in management for establishing priorities in multicriteria decision problems. Underlying the AHP is the theory of ratio-scale measures developed in psychophysics since the middle of the last century. It is, however, well known that classical ratio-scaling approaches have several problems. We reconsider the AHP in the light of the modern theory of measurement based on the so-called separable representations recently axiomatized in mathematical psychology. We provide various theoretical and empirical results on the extent to which the AHP can be considered a reliable decision-making procedure in terms of the modern theory of subjective measurement.

Key words: ratio scales; subjective weighting; decision making
History: Received January 8, 2009; accepted November 5, 2009, by Peter Wakker, decision analysis. Published online in *Articles in Advance* February 12, 2010.

1. Introduction

The analytic hierarchy process (AHP) is a decision-making procedure originally developed by Saaty (1977, 1980, 1986). Its primary use is to offer solutions to decision and estimation problems in multivariate environments. The AHP establishes priority weights for alternatives by organizing objectives, criteria, and subcriteria in a hierarchic structure.

Central to the AHP is the process of measurement, in particular measurement on a ratio scale. Given a set $A_1, \ldots, A_n$ of items of a decision problem (for example, n alternatives) to be compared for a given attribute (for example, one criterion), in the AHP, a response matrix $A = [a_{ij}]$ is constructed with the decision maker's assessments a_{ij}, taken to measure on a subjective ratio scale the relative dominance of item i over item j. For all pairs of items i, j, it is assumed that

$$a_{ij} = \frac{w_i}{w_j} \cdot e_{ij}, \qquad (1)$$

where w_i and w_j are underlying subjective priority weights belonging to a vector $w = (w_1, \ldots, w_n)'$, with $w_1 > 0, \ldots, w_n > 0$ and, by convention, $\sum w_j = 1$, and where e_{ij} is a multiplicative term introduced to account for random errors and inconsistencies often observed in practice. It is assumed that e_{ij} is close to 1, is reciprocally symmetric $e_{ij} = e_{ji}^{-1}$, with $e_{ii} = 1$ for all i.

The AHP has spawned a large literature. Critics have referred to both technical and philosophical aspects of the AHP. Concerns have especially been expressed about the distance of the AHP from the axioms of classical utility theory (Dyer 1990, Smith and von Winterfeldt 2004). Defenders of the AHP have always rejected the criticism, arguing that the normative foundations of the AHP are not in utility theory, but in the theory of measurement (Harker and Vargas 1990, Saaty 1990, Forman and Gass 2001). Appeal has often been made to the psychophysicist Stevens (1946, 1951) and his famous ratio-scaling method.

In this paper, we reconsider the AHP in the light of the newer theory of psychological measurement. Indeed, in spite of the reference of the AHP defenders to the work of Stevens (1946, 1951), it is well known in mathematical psychology that Stevens' theory had several problems (Michell 1999). For mathematical psychologists, a major drawback of the theory has always been seen in the lack of proper mathematical and philosophical foundations justifying the proposition that, when assessing a ratio judgment, a "subject is, in a scientific sense, 'computing ratios'" (Narens 1996, p. 109).

In recent years, however, there has been an important stream of research clarifying the conditions and giving various sets of axioms that can justify ratio estimations. A relevant achievement of the recent literature has been the axiomatization of various theories of subjective ratio judgments belonging to a class of so-called *separable representations* (Narens 1996, 2002; Luce 2002, 2004).

Separable representations are important because they make it clear that individual ratio judgments can be subject to various cognitive distortions. In particular, they imply that Equation (1) of the classical AHP approach can be recast as

$$a_{ij} = W^{-1}\left(\frac{w_i}{w_j}\right) \cdot e_{ij}, \qquad (2)$$

where $W^{-1}(\cdot)$ is the inverse of a subjective weighting function $W(\cdot)$ relating elicited subjective proportions to numerical ratios. Clearly, Equations (1) and (2) are equivalent only when W is the identity.

In the following, we will analyze Equation (2) under various perspectives. We will start with a review of the ratio-scaling method of the classical AHP. We will move on to consider the relationships between the AHP and the modern theory of measurement, and will develop a statistical method to estimate the priority vector $w = (w_1, \ldots, w_n)'$ from Equation (2), which takes into account possible nonlinearities in the subjective weighting function W. Afterward, we will apply the method to some experimental data we have obtained from subjective ratio estimation experiments. We will then compare the results obtained by the classical AHP with those obtained by our method and will discuss the implications of the findings for the status and the practice of the AHP.

2. Scaling and Prioritization in the Classical AHP

In the 1977 paper by Saaty and subsequent book (Saaty 1980), the AHP was developed as a set of operational procedures without axiomatic foundations. Axioms were later added by Saaty (1986). Central in Saaty's (1986) system of axioms is the primitive notion of "fundamental scales." Let $\mathcal{A}$ be a set of alternatives A_i with $i = 1, \ldots, n$ and n finite, and let C be one among a set of criteria to compare the alternatives. A fundamental scale for criterion C is a mapping P_C, which assigns to every pair $(A_i, A_j) \in \mathcal{A} \times \mathcal{A}$ a positive real number $P_C(A_i, A_j) \equiv a_{ij}$ such that (1) $a_{ij} > 1$ if and only if A_i dominates (or "is strictly preferred to") A_j according to criterion C, and (2) $a_{ij} = 1$ if and only if A_i is equivalent (or "indifferent") to A_j according to criterion C.

Operationally, the a_{ij}'s are obtained as the decision maker's assessments of the subjective response matrix $A = [a_{ij}]$ in the Introduction. It is also important to remark that in the AHP, the notion of fundamental scale does not apply only to preference criteria, but is also relevant for carrying out estimation.[1]

Four axioms are given by Saaty (1986) to characterize various operations that can be performed with fundamental scales. The first axiom establishes $A = [a_{ij}]$ as a positive reciprocally symmetric matrix, that is, $a_{ij} \cdot a_{ji} = 1$ and $a_{ii} = 1$ for all $A_i, A_j \in \mathcal{A}$. The second, the so-called homogeneity axiom, suggests aggregating or decomposing the items of a decision into clusters or hierarchy levels so that the stimuli do not differ too much in the property being compared. Otherwise, large errors in judgment may occur. Based on empirical research, the AHP has elaborated various scale models to elicit judgments, including Saaty's (1977, 1980) famous verbal scale with integers from 1 to 9 for intensities of relative importance. The third and fourth axioms give conditions for hierarchic composition. They are not relevant for the present discussion.[2]

A fundamental scale P_C does not give directly a scale of priorities. A scale of priorities is, in Saaty's (1986) language, an n-dimensional vector $w(A) = (w_1, \ldots, w_n)'$ obtained from matrix $A = [a_{ij}]$, with $1 \geq w_i \geq 0$, such that the ith component of $w(A)$ accurately represents the relative dominance of alternative A_i among the n alternatives in $\mathcal{A}$. One question that has triggered off a lot of debate over the AHP concerns the best prioritization method, that is, the best method to obtain the vector $(w_1, \ldots, w_n)'$ from $A = [a_{ij}]$.

The issue is complicated by the fact that the AHP acknowledges that people can be subject to random errors and inconsistencies. This feature of the AHP is transparent in Equation (1) of the Introduction for the relationship between the a_{ij}'s and the priority weights with the multiplicative error term e_{ij}.

The error terms e_{ij} have various implications. First of all, they imply that the ratio judgments obtained by the AHP may violate an important property known as cardinal consistency, $a_{ij} = a_{ik} \cdot a_{kj}$, or even a weaker requirement of ordinal consistency, namely, that when $a_{ij} > 1$ and $a_{jk} > 1$, then also $a_{ik} > 1$.

Second, the possibility of consistency violations poses the problem of estimating the priority weights $w = (w_1, \ldots, w_n)'$ in an appropriate way. The technique proposed by Saaty in his original paper (Saaty 1977) and defended ever since is the maximum eigenvalue (ME) method. It uses the response

[1] For example, in choosing which car to buy, a decision maker may be interested in estimating the relative fuel consumption of various cars, which is then weighted with other preference criteria, before making the final decision.

[2] They are also more controversial due to the problem of rank reversal, which may plague the AHP analyses (Dyer 1990). Extensions of the AHP techniques can avoid rank reversal (Pérez 1995). In this paper, we will not deal with this issue.

matrix $A = [a_{ij}]$ to solve for the column vector $w = (w_1, \ldots, w_n)'$ the linear system of equations

$$Aw = \lambda_{\max} w, \qquad \sum_{i=1}^{n} w_i = 1, \tag{3}$$

where $\lambda_{\max} > 0$ is the largest eigenvalue in modulus (that is, the Perron eigenvalue) of A. It is known from the Perron-Frobenius theorem that system (3) has a unique solution, the Perron eigenvector, henceforth denoted by $\overline{w} = (\overline{w}_1, \ldots, \overline{w}_n)'$. Moreover, if $e_{ij} = 1$ for all pairs (i, j) so that A is cardinally consistent, the ME method delivers the correct priority weights $\overline{w}_i = w_i$ for all i, and the maximum eigenvalue is at its minimum value $\lambda_{\max} = n$. When A is not cardinally consistent, $\overline{w} = (\overline{w}_1, \ldots, \overline{w}_n)'$ usually differs from the correct priority vector, and $\lambda_{\max} > n$. The normalized difference $(\lambda_{\max} - n)/(n-1)$ is proposed in the AHP as a consistency index. In particular, if the index for a certain response matrix is larger than a given cutoff, the AHP suggests correcting the matrix restating the subjective judgments of the individuals until near consistency is reached (various methods are offered in the AHP for conducting such a revision; Saaty 2003).

The ME method has, however, been criticized for not paying attention to the stochastic structure of the data (de Jong 1984). Several alternative techniques for estimating w have therefore been proposed. The logarithmic least squares method (LLSM), which remains the most popular alternative for statistical standards (de Jong 1984, Crawford and Williams 1985, Genest and Rivest 1994), is mentioned below.

3. Separable Representations and the AHP

As is well known, the original idea of developing analytical procedures and experimental techniques for constructing subjective ratio measurement scales was due to Stevens (1946, 1951, 1975). A long history of controversies has followed the approach and ratio-scaling methods more generally (Anderson 1970, Shepard 1978, Michell 1999). Much of the concern has focused on the philosophical and theoretical justifications for the correspondence, also assumed by the AHP, between the number names in the instruction of the subjective scaling procedures and scientific numbers.

In the last 10 years or so, however, a great effort has been made by mathematical psychologists, notably Narens (1996, 2002) and Luce (2002, 2004), to comprehend in a deeper perspective the structural assumptions inherent in the ratio-scaling approach. To illustrate, consider a ratio estimation task in which an individual is provided with two stimuli, z and x, and then is asked to state the value p that corresponds to the subjective ratio of z to x.

Formally, a separable representation is said to hold in a ratio estimation if there exist a *psychophysical function* ψ and a *subjective weighting function* W such that the ratio p is in the following relation with z and x:

$$W(p) = \frac{\psi(z)}{\psi(x)}. \tag{4}$$

Equation (4) corresponds to Narens' (1996) original model and incorporates the notion that independent distortions may occur both in the assessment of subjective intensities and in the determination of subjective ratios (see also Luce 2002). Stevens's (1951, 1975) classical model applies when W can be represented as the identity function and ψ is a power function.

Narens (1996) developed the model in the tradition of representational measurement theory. Central in the axiomatization is the distinction between numerals, which are the response items p provided by the subject to the experimenter, and scientific numbers. Narens (1996), in particular, argued that the case $W(p) = p$ is anything more than a coincidence. He showed that a specific behavioral property, which he called "multiplicativity," must hold. He instead suggested that a weaker property, which underlies representation (4), may hold. The condition is called "commutativity." It implies that a subjective proportion, say, of 2 multiplied by a subjective proportion of 3 is equivalent to a subjective proportion of 3 multiplied by a subjective proportion of 2, though neither products of subjective proportions is equivalent to a subjective proportion of 6, for which the full force of the multiplicative property is necessary.

Form (4) has been also axiomatized by Luce (2002, 2004, 2008) as a special case of a global theory of psychophysics, developed from empirically testable assumptions relating sensorial stimuli intensities (like auditory or visual). Tests rejecting multiplicativity but not commutativity have been conducted by Ellermeier and Faulhammer (2000) and Zimmer (2005) in loudness magnitude production experiments.[3] Other recent evidence in favor of separable representations has been obtained by Steingrimsson and Luce (2005a, b) and Bernasconi et al. (2008).

The axiomatic approach underlying separable representations also has direct application in the context of utility theory for gambles where ψ is called utility, with domain represented by valued goods, and W is a subjective weighting function of probabilities or events that generalizes the classical expected utility model. Examples include cumulative prospect theory

[3] Ratio production is a dual scaling procedure widely used in psychophysics in which an observer is required to produce a stimulus x that appears p times more intense than a reference stimulus.

(Tversky and Kahneman 1992) and the class of rank-dependent utility models.[4]

3.1. AHP in Separable Form

Separable representations also have a natural interpretation in the AHP. In the AHP, the primitives are the alternatives $(A_1, \ldots, A_n)$ and the ratio estimation task that leads to the subjective assessments a_{ij}. The AHP also assumes the existence of a fundamental scale of alternatives, but it does not deal with the question of the qualitative properties that must be satisfied by the ratio estimation task for the priority vector $(w_1, \ldots, w_n)'$ to represent a ratio scale.

On the contrary, by directly applying the representational theory of measurement to the alternatives $A_1, \ldots, A_n$, the relationships between the alternatives and the subjective responses a_{ij} should be written as $W(a_{ij}) = \psi(A_i)/\psi(A_j)$, which, using the normalization $w_1 = \psi(A_1)/(\sum \psi(A_k)), \ldots, w_n = \psi(A_n)/(\sum \psi(A_k))$, yields

$$W(a_{ij}) = \frac{w_i}{w_j}. \qquad (5)$$

Representation (5) and Equation (1) of the classical AHP differ in two respects: the ratio form of Equation (1) restricts the weighting function $W(\cdot)$ to be the identity, whereas representation (5) ignores considerations of errors e_{ij}. As noted in §2, errors e_{ij} represent an important source of possible inconsistencies in the AHP. In the following, we will recomprehend the effect of errors in a more general stochastic version of model (5). Here it is important to emphasize the implications of separable representations for the classical AHP, even in the idealized situation in which the model is thought to hold exactly.

In particular, it should be transparent that in terms of form (5), the property of cardinal consistency in the AHP, namely, $a_{ij} = a_{ik} \cdot a_{kj}$, is equivalent to Narens' (1996) multiplicative property, requiring that the subjective weighting $W(\cdot)$ is the identity, or that it takes the slightly more general power specification $W(p) = p^k$ with $k > 0$ and $W(1) = 1$.[5]

[4] Consider a gamble (x, p, y) giving x with probability p and y otherwise. A rank-dependent utility form is given by $U(x, p, y) = U(x)W(p) + U(y)[1 - W(p)]$. Many authors have worked with this form (see Köbberling and Wakker 2003 for the most general results on this form). When W is linear, the form collapses to the classical expected utility model. When $y = 0$, the expression $U(x, p, y) = U(x)W(p)$ is a separable form.

[5] In particular, it is trivial that the power function $W(p) = p^k$ always satisfies multiplicativity $(W(p) \cdot W(q) = W(p \cdot q))$ when $W(1) = 1$. It is interesting to note that a similar power model for the AHP was considered by Saaty (1980, p. 189) himself and referred to as the eigenvalue power law. A subtlety investigated by Steingrimsson and Luce (2007) is that when multiplicativity fails, $W(p)$ may still be a power function with $W(1) \neq 1$.

The implication is that whenever $W(\cdot)$ is not the identity or the power model, violations of cardinal consistency are inherent in any subjective ratio assessment. At the same time, separable representations show that when multiplicativity fails but form (5) holds, ratio estimations may still result into a ratio scale, though it is necessary to pass through the function W to interpret the subjects' subjective measures of ratios as numerical ratios.

In this respect, an important property of function $W(\cdot)$ is monotonicity, which follows from the mathematical derivation of separable representations (Luce 2002, p. 522). In the AHP, monotonicity implies ordinal consistency (that if $a_{ij} > 1$ and $a_{jk} > 1$, then also $a_{ik} > 1$). Moreover, if $W(\cdot)$ is monotonic, W^{-1} is invertible so that the actual entries of the AHP response matrix $A = [a_{ij}]$ are given by

$$a_{ij} = W^{-1}\left(\frac{w_i}{w_j}\right). \qquad (6)$$

This is quite important for the AHP, because it implies that if one knows how to estimate the function $W^{-1}(\cdot)$ and if $W^{-1}(\cdot)$ is invertible,[6] then one can pinpoint between the elicited numerals a_{ij} and the numerical ratios w_i/w_j to obtain the priority weights $w = (w_1, \ldots, w_n)'$.

Another important condition on the subjective weighting function is $W(1) = 1$, which is necessary for $a_{ii} = 1$ in the AHP. Reciprocal symmetry, that is, $a_{ij} = 1/a_{ji}$, requires instead that $W(\cdot)$ is reciprocally symmetric, namely, $W(1/\cdot) = 1/W(\cdot)$. Both conditions are implied by several derivations of separable representations, including a specification developed by Luce (2001, 2002), similar to the one that Prelec (1998) proposed in the context of utility theory for risky gambles. The more recent specification proposed by Luce (2004), however, does not impose either. Both $W(1) = 1$ and symmetry seem fairly natural in the context of the AHP. Therefore, we maintain both conditions in what follows.

4. A Separable Statistical Model for the AHP

The mathematical theories of separable forms "are about idealized situations and do not involve considerations of error" (Narens 1996, p. 109). This is acknowledged as a limitation. People are not like

[6] An earlier example of form (6) was studied by Birnbaum and Veit (1974) with a function J_R conceptually identical to the inverse function $W^{-1}(\cdot)$ introduced directly as a monotonic judgmental transformation of a ratio model relating overt magnitude estimation of ratios to subjective impressions of ratios, and put in connection to a judgmental transformation J_D of a model relating overt rated differences to subjective differences.

robots. Various elements, including lapses of reason or concentration, states of mind, trembling, rounding effects, and computational mistakes imply the obvious notion that no model of human behavior can be thought to hold deterministically.

This was also remarked by Saaty (1977, 1986) in the AHP with the notion of the multiplicative errors e_{ij}. We now recomprehend the errors in the separable specification of the AHP. In particular, introducing the multiplicative terms e_{ij} to model (6), we obtain the separable statistical form (2) of the Introduction: $a_{ij} = W^{-1}(w_i/w_j) \cdot e_{ij}$.

Form (2) raises several issues for the AHP. Here we focus on a procedure for conducting a rigorous statistical analysis of Equation (2), which we then apply to the data of a ratio estimation experiment in the AHP. We proceed first showing how to obtain a regression model from Equation (2), and then present an inference method to obtain the unknown parameters of the model. We remark that the procedure applies on an individual basis.

4.1. Regression Model

As a first step to transform Equation (2) into a regression model amenable to statistical analysis, we take the log transformation

$$\ln a_{ij} = \ln W^{-1}[\exp(\ln w_i - \ln w_j)] + \varepsilon_{ij}, \qquad (7)$$

where $\varepsilon_{ij} = \ln e_{ij}$. We now assume that the deterministic function $\ln W^{-1}[\exp(\cdot)]$ can be approximated through a polynomial in its arguments. This is generally possible: according to the Weierstrass approximation theorem, any continuous function on a compact domain can be approximated to any desired degree of accuracy by a polynomial in its arguments. Here we stop at a third-order approximation that yields the expression

$$\ln W^{-1}[\exp(z)] \simeq \beta_0 + \beta_1 z + \beta_2 z^2 + \beta_3 z^3, \qquad (8)$$

with $z = \ln w_i - \ln w_j$. We emphasize that an approximation to the third order is sufficient to characterize all the various restrictions discussed in the previous sections for the AHP.[7] In particular, notice that

- the restriction $W(1) = 1$ (from $a_{ii} = 1$) implies $\beta_0 = 0$,
- the fact that W is reciprocally symmetric (from $a_{ji} = 1/a_{ij}$) implies $\beta_2 = 0$,
- the classical AHP where W is the identity (or the power model $W(p) = p^k$ with $k > 0$ and $W(1) = 1$) restricts $\beta_1 = 1$ and $\beta_3 = 0$,
- finally, the case in which $\beta_1 = 1$ and β_3 is left free to vary corresponds to a (third-order) logarithmic approximation of the inverse separable model (6) of §3.1, namely,

$$\ln W^{-1}[\exp(z)] \simeq z + \beta_3 z^3. \qquad (9)$$

Substituting in Equation (7), we finally obtain the statistical inference model:

$$\ln a_{ij} \simeq (\ln w_i - \ln w_j) + \beta_3(\ln w_i - \ln w_j)^3 + \varepsilon_{ij}. \qquad (10)$$

When $\beta_3 = 0$, model (10) collapses to a well-known version of Equation (1) of the classical AHP analyzed with the LLSM by many previous authors (de Jong 1984, Crawford and Williams 1985, Genest and Rivest 1994).

4.2. Statistical Analysis

In fact, we now propose a method for conducting the statistical inference in model (10) that can be viewed as a generalization of the LLSM method (in particular, of the analysis of Genest and Rivest 1994). We derive the estimators $\widehat{w}_i$ and $\hat{\beta}_3$ minimizing the sum of squares:[8]

$$(\widehat{w}_1, \ldots, \widehat{w}_n, \hat{\beta}_3)' = \underset{(w_1, \ldots, w_n, \beta_3)}{\arg\min} \sum_{i,j=1}^{n} \varepsilon_{ij}^2,$$

where

$$\varepsilon_{ij} = \ln a_{ij} - (\ln w_i - \ln w_j) - \beta_3(\ln w_i - \ln w_j)^3,$$

under the constraint that $\sum_{i=1}^{n} \widehat{w}_i = 1$. As in Genest and Rivest (1994), we assume that the errors ε_{ij} for $1 \le i \le n$, $1 \le j < i$ are independent with common variance σ^2. Also recall that $\varepsilon_{ij} = -\varepsilon_{ji}$ and $\varepsilon_{ii} = 0$. Under suitable hypotheses it is possible to show that the estimator $(\widehat{w}_1, \ldots, \widehat{w}_n, \hat{\beta}_3)'$ is consistent and asymptotically normal when $\sigma \downarrow 0$, but the formulas of its asymptotic variance depend on the unknown parameter σ. It is slightly more complicated to find an estimator of σ^2, but if we define the residuals

$$\hat{\varepsilon}_{ij} = \ln a_{ij} - (\ln \widehat{w}_i - \ln \widehat{w}_j) - \hat{\beta}_3(\ln \widehat{w}_i - \ln \widehat{w}_j)^3,$$

the estimator

$$\widehat{\sigma}^2 = \frac{1}{n^2 - 3n} \cdot \sum_{i,j=1}^{n} \hat{\varepsilon}_{ij}^2$$

[7] More generally, in the statistical approach presented below, the order of the approximation can be extended to any desired degree (provided of course that one has enough data to estimate the model), and then the order of the polynomial can be selected using various techniques, for example, the statistical theory of model selection, as in Bernasconi et al. (2008).

[8] The precise statement of the results of this section and the full derivation of the asymptotic theory of estimator $(\widehat{w}_1, \ldots, \widehat{w}_n, \hat{\beta}_3)'$ are in the online appendix in the e-companion to this paper (which is available as part of the online version that can be found at http://mansci.journal.informs.org/).

is asymptotically unbiased (in the sense that $\lim_{\sigma\downarrow 0}(\mathbb{E}\widehat{\sigma}^2/\sigma^2)=1$). The problem is that this estimator is not consistent. Also neither is the LLSM estimator of Genest and Rivest (1994). This is not dramatic: it implies only that when $\sigma\downarrow 0$ and σ^2 is replaced with its estimator $\widehat{\sigma}^2$, the classical t-tests on coefficients are $t((n^2-3n)/2)$-distributed (whereas in classical regression theory they are Gaussian). Tests and confidence intervals can however be built even if asymptotic theory is nonstandard.

5. The Experiment

In the following, we present a test of the AHP based on a pure estimation experiment in which participants were asked to give their estimates of (i) ratios of distances of pairs of Italian cities from a reference city, (ii) ratios of probabilities resulting from games of chance, and (iii) ratios of rainfall in pairs of European cities.

For the three experiments we presented participants five items (see Table 1) compared in 10 pairs and asked them, first, to state for each comparison which of the two items they thought was dominant in the relevant experimental dimension (distances, chances, rainfall), and then to quantify with a number chosen in the interval of integers from 1 to 9 the relative dominance of the two items: for example, in the comparisons of the distances experiment, how many times the city that they considered more distant from Milan was, according to them, actually more distant from Milan than the city they considered less distant.

The three experiments are based on 69 individuals who performed all the experiments in a random order. The interval of integers 1–9 was used in accordance with Saaty's (1977, 1980) "scale of relative importance." Also notice that the integers 1–9 cover the proportions between the physical stimuli, so that the design satisfies Saaty's (1986) homogeneity axiom.

Participants were undergraduate students in economics from the University of Insubria in Italy. It was decided to use a monetary reward as an incentive for subjects to perform the experiment as well as possible; namely, the payment for each subject was proportional to a measure of the accuracy of his or her assessments computed in each experiment according to the formula

$$\Delta=\sum_{i=1}^{n}(v_i-\overline{w}_i)^2,$$

where $v=(v_1,\ldots,v_n)'$ is the vector of the normalized true values in the experiment, and $\overline{w}=(\overline{w}_1,\ldots,\overline{w}_n)'$ is the ME eigenvector computed for the subject.[9] In particular, because when ψ and W are linear (and there are no random errors e_{ij}) the ME eigenvector delivers unbiased estimates, the method implies that when there is no bias, the more accurate the subjects are in their assessments, the higher their payoffs in the experiments. Obviously, this also means that the design gives an incentive for subjects to be consistent with the classical AHP.

5.1. Empirical Estimates

We have estimated the priority vector $(w_1,\ldots,w_n)'$ for every individual participating in the three experiments using both the ME eigenvector method and our theory of polynomial approximation based on the joint estimation of vector $(\widehat{w}_1,\ldots,\widehat{w}_n)'$ and parameter $\hat{\beta}_3$ to account for the possible nonlinearity of the subjective weighting function.[10]

Figure 1 reports the estimates for $w_1,\ldots,w_n$: the first row displays the graphs of the distances experiment; the second row, the graphs of the chances experiment; the third row, the graphs of the rainfall experiment. Each graph displays on the x-axis the weight $\widehat{w}_j$ as obtained by our procedure and on the y-axis the weight $\overline{w}_j$ as obtained by the ME method. The vertical dashed lines in each graph correspond to the true values. Overall, the graphs show that there are differences in the precision of the assessments in the three experiments: perhaps as expected, the

Table 1 Items Compared in the Three Experiments

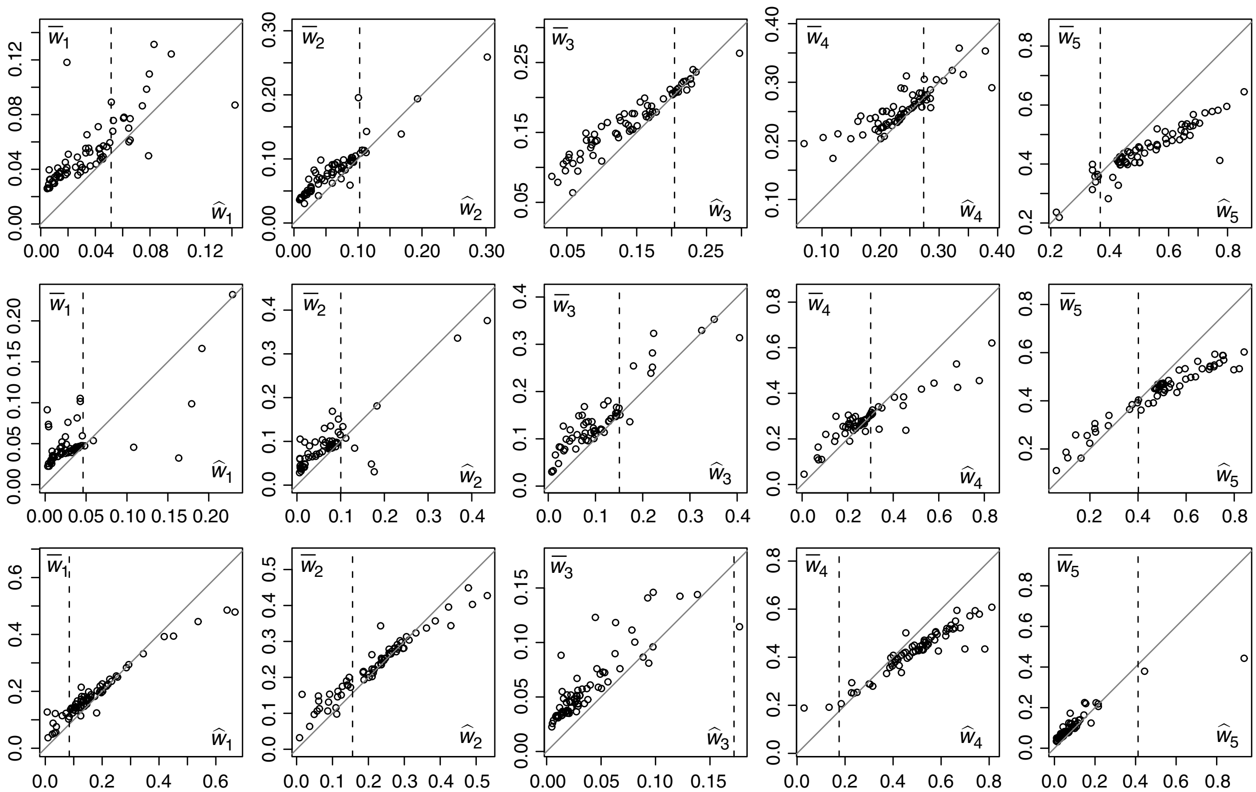
Distances from Milan (km)	Games of chance (probability)	Rainfall in November 2001 (mm)
1. Naples (658 km)	Take 1 heart out of a pack of 52 cards (1/4)	Prague (26 mm)
2. Venice (247 km)	Take 1 ace out of a pack of 52 cards (1/13)	Athens (53 mm)
3. Rome (491 km)	Get at least 3 in a roll of a 6-sided dice (4/6)	Copenhagen (48 mm)
4. Turin (124 km)	Get 6 in a roll of a 6-sided dice (1/6)	London (54 mm)
5. Palermo (885 km)	Get heads in a toss of a coin (1/2)	Rome (127 mm)

Sources. http://www.chemical-ecology.net/java/lat-long.htm for distances, Fremy and Fremy (2001) for rainfall.

[9] The actual payment for each subject, in euros, was given by $\Pi=\max\{3, 25-100\cdot(\Delta_1+\Delta_3+\Delta_3)\}$, where Δ_1, Δ_2, Δ_3 are the values of Δ computed for each of the three experiments (distances, chances, rainfall). Thus, each subject was sure to obtain at least three euros from the experiment. The average payment was about 13 euros per subject. In the actual instructions, the formula for computing payoffs was given and explained in an appendix. The general instructions explained the experimental task to subjects and said that the payment in euros was computed according to a formula that gave them incentives to be as accurate as possible in their assessments.

[10] We have also produced estimates using the LLSM proposed by de Jong (1984), Crawford and Williams (1985), and Genest and Rivest (1994). The LLSM estimates, not reported for brevity, are very similar to those obtained by the ME method, confirming previous results and theoretical expectations (Zahedi 1986, Budescu et al. 1986, Genest and Rivest 1994).

Figure 1 **Individual Estimates of Priority Weights Obtained by Various Methods**



Note. First row, distances experiment; second row, chances experiment; third row, rainfall experiment.

greater precision is in the chances experiment, followed by the distances experiment, and finally by the rainfall experiment (in which the majority of subjects actually ordered cities in the wrong way, guessing that rainfall in November 2001 was heavier in Copenhagen, Prague, and London than in Athens and Rome, whereas the converse was true). The main interest in the graphs in Figure 1 is that they document clear differences in the estimates obtained by the two methods, because in all the three experiments the weights obtained by the two methods do not lie around the 45° line (gray in the graphs) and display significant departures from linearity. Moreover, the differences are systematic in that the diagrams document a tendency of the ME method in comparison to ours to overestimate for most subjects the weights when these are perceived in some way to be low, and underestimate the weights when they are somehow perceived to be greater (the cutoff weight for separating overestimation from underestimation is around 0.4, see Figure 1). We also see that this tendency is fairly robust across all the three experiments.

The graphs in Figure 2 document the reason for this systematic distortion. The figure shows the functions $W^{-1}(x) \simeq x \cdot \exp\{\beta_3 \ln^3 x\}$ estimated for each subject in the three experiments. In each individual graph, the gray solid lines represent the identity function, the black solid lines represent the estimated functions, and the black dashed lines show the confidence intervals at 95%. The confidence intervals allow us to conduct graphically for each subject the test of the null hypothesis that W is the identity. Out of 69 subjects, the hypothesis is rejected in 47 cases in the distances experiment, in 37 cases in the chances experiment, and in 57 cases in the rainfall experiment.

The individual diagrams also indicate that the great majority of subjects show a similar pattern of distortions, in the sense that they substantially underestimate the ratios (fitted W^{-1} below the 45° line). Moreover, a concave shape function $W^{-1}(x)$ is estimated for the majority of subjects, indicating that the tendency to underestimate ratios increases as the ratios get increasingly larger above one. This tendency directly explains why the ME eigenvector method, which does not account for the distortions in the subjective weighting functions, overestimates low weights w_j and underestimates larger weights w_j as seen previously. In this respect, we also remark that the parameter $\hat{\beta}_3$ accounting for the departure from nonlinearity of the subjective weighting function has similar estimates in the three experiments: in particular, the median value of $\hat{\beta}_3$ is -0.0321 in the distances experiment, -0.0353 in

Figure 2 **Individual Estimates of the Inverse Subjective Weighting Function W^{-1}**

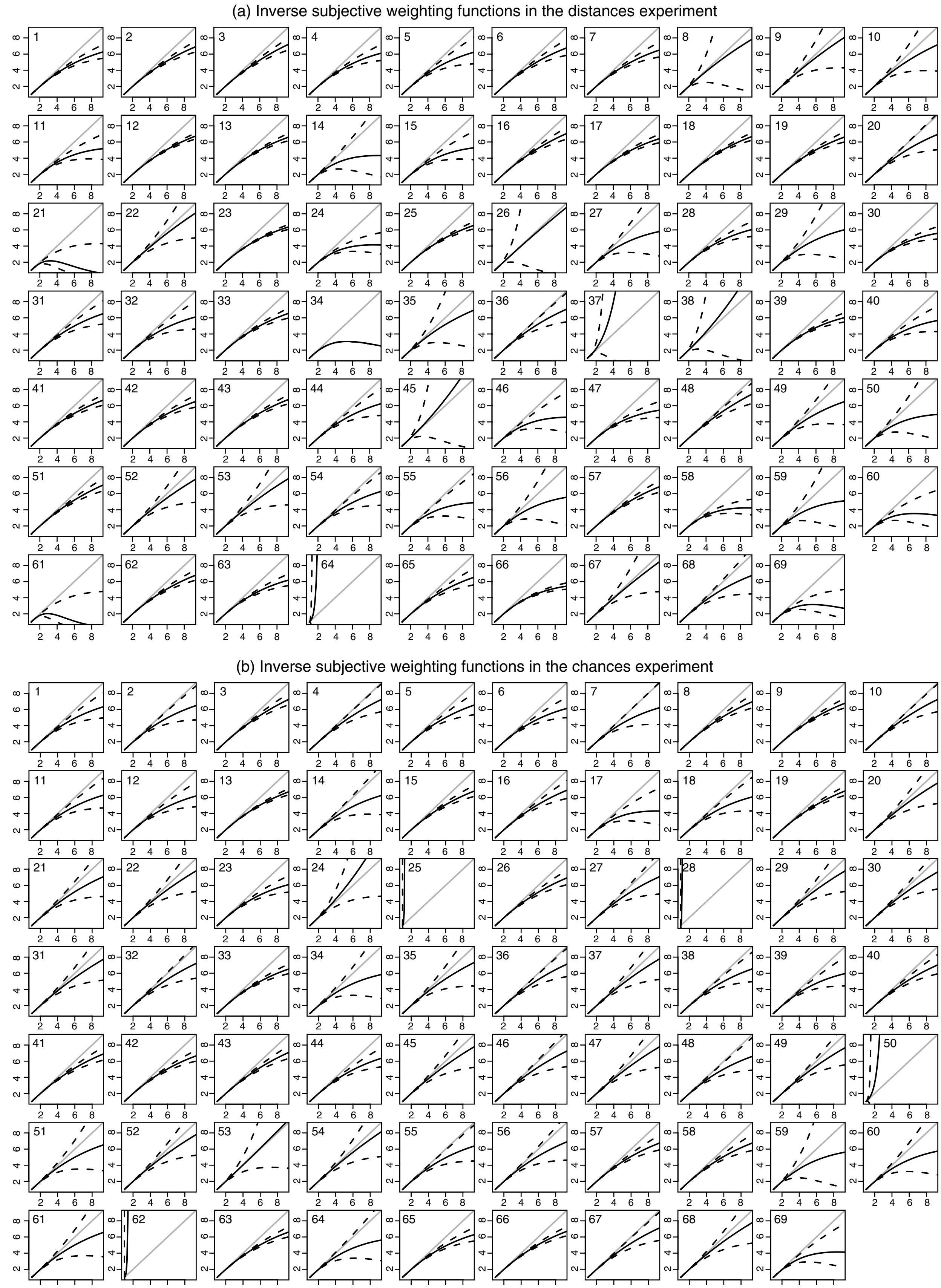
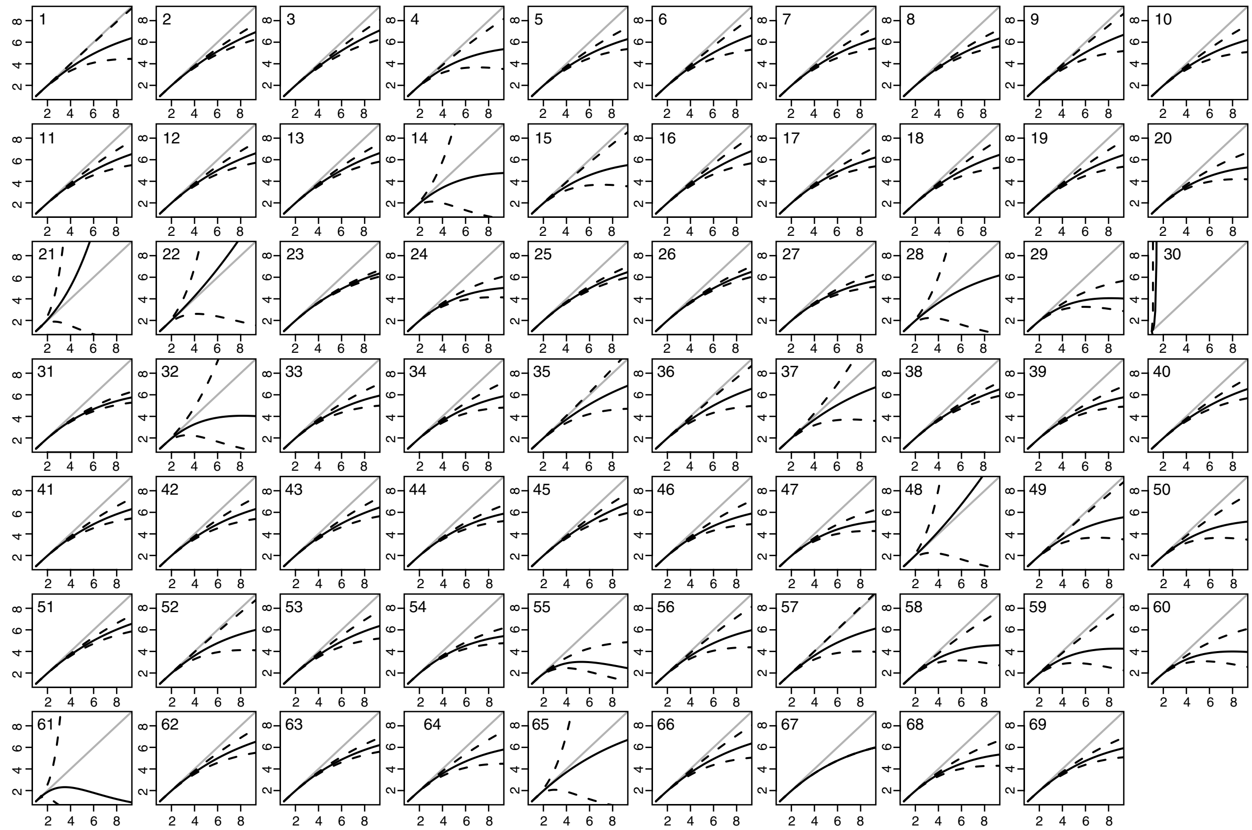


Figure 2 (Continued)

(c) Inverse subjective weighting functions in the rainfall experiment



the rainfall experiment, and −0.0271 in the chances experiment.

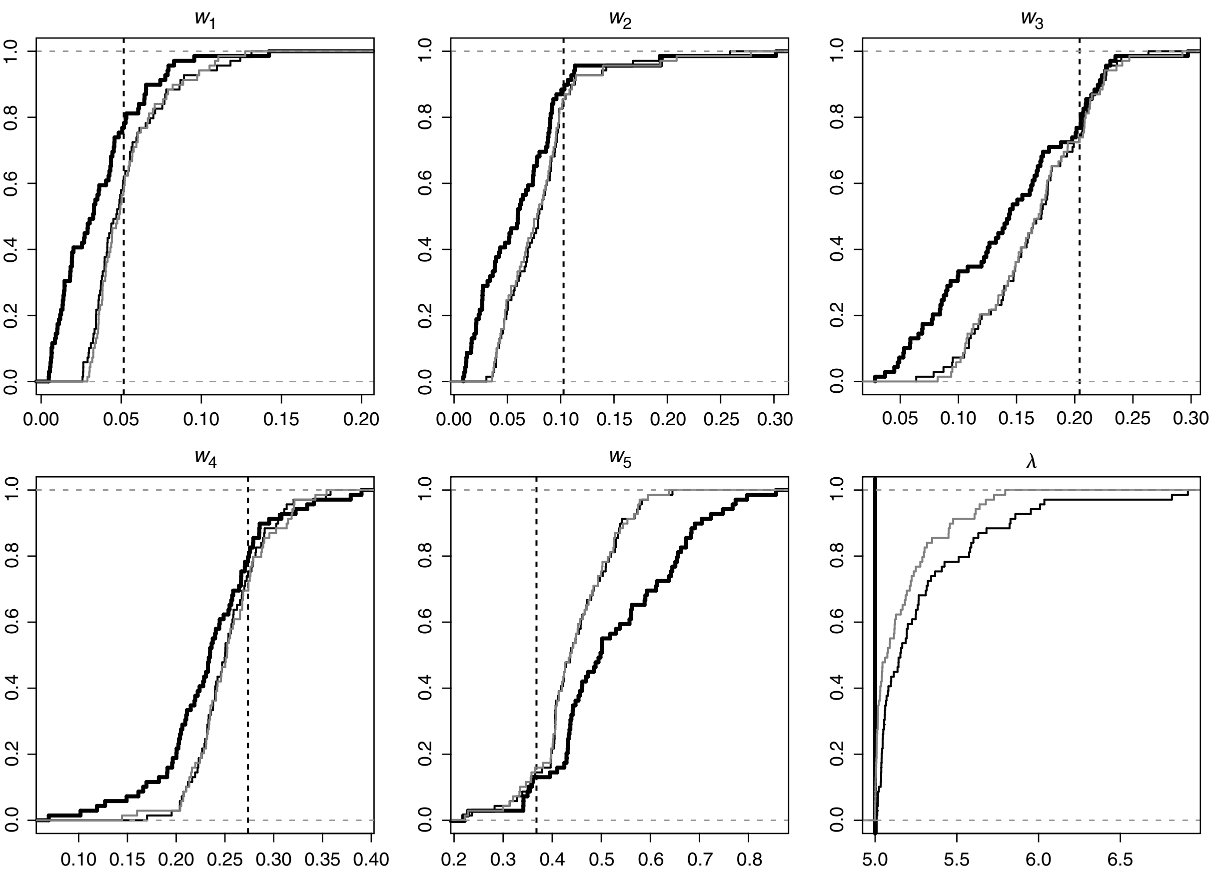
The diagrams in Figure 2 also allow us to conduct some tests of ordinal consistency. In particular, recall that if W^{-1} is not an invertible function, then the subjects' ratio assessments violate ordinal consistency. In our estimation procedure, we have decided not to impose invertibility. Nevertheless, we find that only very few subjects violate ordinal consistency: 5 subjects out of 69 in the distances experiment (subjects 21, 34, 60, 61, 69) and 2 subjects in the rainfall experiment (subjects 55 and 61), whereas all subjects in the chances experiment satisfy ordinal consistency.

5.2. The Effects of W and Errors ε_{ij}

As emphasized, the nonlinearity of the subjective weighting function implies that the AHP response matrix necessarily contains consistency violations. In this perspective, perhaps the most empirically relevant question for the AHP is whether the inconsistencies due to W (and thus β_3) are larger or smaller than the ones due to the noise e_{ij} (or ε_{ij}). To address such a question, we compare eigenvector $\overline{w} = (\overline{w}_1, \ldots, \overline{w}_n)'$ and eigenvalue $\bar{\lambda}$ computed from the matrix of the elicited responses $A = [a_{ij}]$, with the eigenvector-eigenvalue obtained by first removing from the matrix the effect of the noise, thus computing w and λ on the basis of $\widehat{W}^{-1}(\widehat{w}_i/\widehat{w}_j)$ (with $\widehat{W}^{-1}$ and $\widehat{w} = (\widehat{w}_1, \ldots, \widehat{w}_n)'$ estimated according to our method), and then removing also the effect of the distortions due to W, thus computing w and λ directly on the basis of matrix $[\widehat{w}_i/\widehat{w}_j]$.

The results of this decomposition for the distances experiment are shown in Figure 3. For the other experiments the results are similar. The six subgraphs in the figure display the empirical cumulative distribution functions (cdfs) of the 69 values for the five components of the maximum eigenvectors obtained from the different matrices indicated above, followed by the empirical cdfs of the 69 values of the maximum eigenvalues computed for the various matrices. In particular, the thin black lines refer to the cdfs of the five components of the ME eigenvector $\overline{w} = (\overline{w}_1, \ldots, \overline{w}_n)'$ and the ME eigenvalue $\bar{\lambda}$; the gray curves represent the empirical cdfs of the 69 values of the corresponding quantities (eigenvector and eigenvalue) computed for each individual on the basis of the matrix with generic element $(\widehat{w}_i/\widehat{w}_j) \cdot \exp\{\hat{\beta}_3 \cdot \ln^3(\widehat{w}_i/\widehat{w}_j)\} = a_{ij}/\hat{e}_{ij}$; the thick black curves represent the empirical

Figure 3 **Empirical Cumulative Distributions of Eigenvector and Eigenvalue Estimates in the Distances Experiment**



cdfs of the 69 values of the same quantities computed for each individual on the basis of the matrix with generic element $\widehat{w}_i/\widehat{w}_j$. The dashed vertical lines represent the ratios based on the true distances.

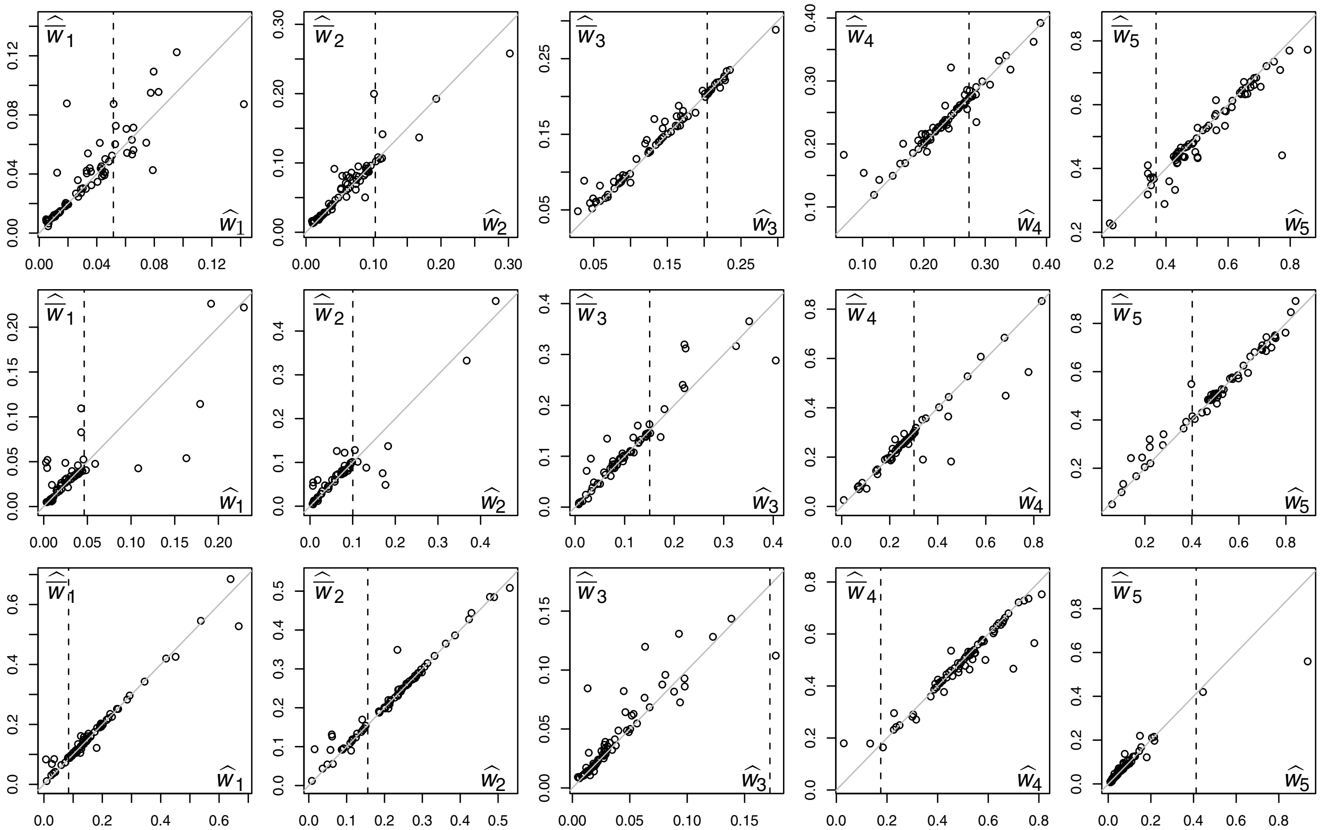
In the graphs, the differences between the thick-black and gray curves are apparent, whereas the gray and the thin black curves are almost indistinguishable. This indicates that the distortions due to the errors ε_{ij} have a definitively smaller effect on the elicited responses than the distortions due to W. The implication is that the errors ε_{ij} have only a limited impact on consistency violations. This is demonstrated by the last graph for the eigenvalue, which shows that the gray line computed with the data cleaned by the noise is only a bit lower than the thin black line for the ME eigenvalue (recall that perfect consistency holds when the Perron eigenvalue is $\lambda_{\max} = 5$), whereas inconsistencies are completely removed only when the effect of W is also taken away from the elicited responses (see the thick black line).

5.3. Discussion

The final objective of the AHP is to obtain the vector of priority weights $(w_1, \ldots, w_n)'$ from the response data $A = [a_{ij}]$. The previous experiment confirms that the response matrix may be affected by fundamental inconsistencies due to the subjective weighting function W. The results are in line with various recent experimental tests of separable representations, including a distance experiment similar to the one described above, but with quantifications on the continuous space of real numbers (Bernasconi et al. 2008). Moreover, in that previous experiment, subjects had an actual map to locate the cities, so as to exert some control over inconsistencies due to subjects not knowing the true distances between cities and inconsistencies introduced by the assessment procedure.

The AHP can be used both for preference measurement and estimation. The above investigation refers to estimation experiments. Estimation tasks are logically simpler than experiments on preference measurement. The distance experiment, in particular, reproduces a standard example used by the AHP to illustrate the force of the measurement method (Saaty 1977, p. 273). Thus, we see the above estimation experiments as a sort of "gold standard" for a test of the classical AHP, in the sense that it seems to us that the evidence that the AHP assessment procedure distorts perception in the context of something that is easy to measure, like distances, makes it hard

Figure 4 Priority Weights Corrected with a "Representative" Weighting Function



Note. First row, distances experiment; second row, chances experiment; third row, rainfall experiment.

to believe that the classical AHP could work better in the context of something that is difficult to measure, like preferences.

From a practical perspective, the above analysis can yield two possible revisions of the classical AHP. First of all, subjective weighting functions can be used to provide suggestions to the decision makers about how to improve the accuracy of their elicited responses. For example, it is clear that in a repetition of the present experiments, most participants should be advised of their tendency to increasingly underweight the ratios as the ratios get larger above one (see again the plots in Figures 2(a)–2(c)).

Alternatively, the estimates $\widehat{w}_1, \ldots, \widehat{w}_n$ obtained by our statistical procedure could be used directly as the priority weights of interest. The procedure is, however, based on subject-specific estimates. The concern is therefore that the procedure is prone to individual variability in the sense that one may be concerned with the stability of subject-specific estimates.[11] To address this issue, we compare in Figure 4 the subject-specific weights with the weights obtained by a "representative" agent model applied to correct the individual assessments prior to estimating the individual weights. That is, we derive individual weights $(\widehat{\widehat{w}}_1, \ldots, \widehat{\widehat{w}}_n)$ following the method underlying the thick black lines in Figure 3, but with a model in which the parameter β_3 has been constrained to take the median value $\hat{\hat{\beta}}_3$ of the individuals' estimates.[12]

The results based on the overall medians of all three experiments are presented in Figure 4.[13] The first row compares, for the distances experiment, on the x-axis the weight $\widehat{w}_j$ as obtained through the subject-specific procedure, and on the y-axis the weight $\widehat{\widehat{w}}_j$ as obtained through the "representative" agent procedure; the second row conducts the same comparison for the chances experiment, and the third row conducts the comparison for the rainfall experiment. The results show that in all of the three experiments, the

[11] When the quantifications are on the continuous space of real numbers, there seem indeed to be more differences in the estimated parameters both across subjects and contexts (Bernasconi et al. 2008).

[12] The values obtained in this way should be representative of an "average" weighting function, without being too sensitive to outliers. Other alternative methods to the median for estimating a "representative" weighting function (like trimmed means, etc.) have also been tried, and the results are largely insensitive to these choices.

[13] The analysis using the medians of each individual experiment obtained very close results.

weights obtained by the "representative" procedure are very similar to the ones obtained by the subject-specific weighting functions because the points in the graphs lie very near to the 45° line except for a few exceptions (see also the difference between these graphs and the graphs in Figure 1 comparing $\widehat{w}_j$'s with the ME estimates $\overline{w}_j$'s). These results give some confidence on the stability of subject-specific estimates. They also suggest that a "representative" weighting function constructed from the median of the individual estimates could represent a possible basis from which to obtain more consistent subjective assessments.[14]

6. Conclusion

The AHP is a problem-solving technique for establishing priorities in multivariate environments. It is based on a method of direct subjective measurement similar to the classical psychophysical ratio-scaling approach. Critiques have concerned both the normative and descriptive status of the AHP (Smith and von Winterfeldt 2004). In this paper, we have reexamined the descriptive and normative foundations of the AHP in the light of the modern theory of psychological measurement. We have emphasized that, as far as psychophysical scaling is concerned, several problems may arise with the AHP assuming a direct correspondence between the number names used in the instruction of the subjective scaling procedures and scientific numbers.

In recent years, however, mathematical psychology has provided various axiomatizations based on different psychological primitives that have made explicit the structural assumptions inherent in representing direct measurement data. It has been shown how the key to a rigorous analysis is the subjective weighting function $W(\cdot)$, which allows one to pinpoint, on the basis of normatively justified arguments and descriptively supported hypotheses, between the subjects' perception of proportions and their underlying scientific ratio-scale representation.

We have proposed a method with which to estimate the priority weights in the AHP that takes into account the distortions caused by the subjective weighting function and have conducted an experimental investigation to illustrate the use of the method. We have found that the distortions due to the subjective weighting function are general and fairly robust across estimation experiments, and have shown that our method can be applied to obtain greater consistency in the subjects' ratio assessments. A caveat is that we have applied and tested the method in a pure estimation context. A question remains whether the method can be equally applied in the case of preference measurement.

[14] The idea of correcting judgments elicited from individuals affected by psychological biases to make them more consistent with prescriptive principles has been advocated also in other areas of decision sciences, as, for example, in the measurement of utilities in the context of risk and uncertainty (Bleichrodt et al. 2001).

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The authors are particularly grateful to four anonymous referees and, especially, to the editor in charge, Peter Wakker, for providing several helpful comments on an earlier version of this paper. They also thank David Budescu, Michael Birnbaum, Duncan Luce, and Ragnar Steingrimsson for discussions. They are indebted to participants of the 29th Annual Conference of the Society for Judgment and Decision Making, the 42nd Annual Meeting of the Society for Mathematical Psychology, the 39th European Mathematical Psychology Group Meetings, the XXIX International Congress of Psychology, and the 2009 Conference of Italian Experimental Economists, and seminar participants at the University of Verona, the University of Brescia, and Polytechnic of Milan. They also thank Steve Hare and Will Hui for providing linguistic and stylistic corrections. This work was supported by a grant from MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca).

References

Anderson, N. H. 1970. Functional measurement and psychophysical judgment. *Psych. Rev.* **77**(3) 153–170.

Bernasconi, M., C. Choirat, R. Seri. 2008. Measurement by subjective estimation: Testing for separable representations. *J. Math. Psych.* **52**(3) 184–201.

Birnbaum, M. H., C. T. Veit. 1974. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception Psychophysics* **15**(1) 7–15.

Bleichrodt, H., J. L. Pinto, P. P. Wakker. 2001. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Sci.* **47**(11) 1498–1514.

Budescu, D. V., E. Zwick, A. Rapoport. 1986. A comparison of the eigenvalue method and the geometric mean procedure for ratio scaling. *Appl. Psych. Measurement* **10**(1) 69–78.

Crawford, G. B., C. Williams. 1985. A note on the analysis of subjective judgment matrices. The geometric mean procedure for estimating the scale of a judgment matrix. *J. Math. Psych.* **29**(4) 387–405.

de Jong, P. 1984. A statistical approach to Saaty's scaling method for priorities. *J. Math. Psych.* **28**(4) 467–478.

Dyer, J. S. 1990. Remarks on the analytic hierarchy process. *Management Sci.* **36**(3) 249–258.

Ellermeier, W., G. Faulhammer. 2000. Empirical evaluation of axioms fundamental to Stevens' ratio-scaling approach. I. Loudness production. *Perception Psychophysics* **62**(8) 1505–1511.

Forman, E. H., S. I. Gass. 2001. The analytic hierarchy process—An exposition. *Oper. Res.* **49**(4) 469–486.

Fremy, D., M. Fremy. 2001. *Quid 2001.* Robert Laffont, Paris.

Genest, C., L.-P. Rivest. 1994. A statistical look at Saaty's method of estimating pairwise preferences expressed on a ratio scale. *J. Math. Psych.* **38**(4) 477–496.

Harker, P. T., L. G. Vargas. 1990. Reply to "Remarks on the analytic hierarchy process" by J. S. Dyer. *Management Sci.* **36**(3) 269–273.

Köbberling, V., P. P. Wakker. 2003. Preference foundations for nonexpected utility: A generalized and simplified technique. *Math. Oper. Res.* **28**(3) 395–423.

Luce, R. D. 2001. Reduction invariance and Prelec's weighting functions. *J. Math. Psych.* **45**(1) 167–179.

Luce, R. D. 2002. A psychophysical theory of intensity proportions, joint presentations, and matches. *Psych. Rev.* **109**(3) 520–532.

Luce, R. D. 2004. Symmetric and asymmetric matching of joint presentations. *Psych. Rev.* **111**(2) 446–454.

Luce, R. D. 2008. "Symmetric and asymmetric matching of joint presentations": Correction to Luce (2004). *Psych. Rev.* **115**(3) 601.

Michell, J. 1999. *Measurement in Psychology: A Critical History of a Methodological Concept*. Cambridge University Press, Cambridge, UK.

Narens, L. 1996. A theory of ratio magnitude estimation. *J. Math. Psych.* **40**(2) 109–129.

Narens, L. 2002. The irony of measurement by subjective estimations. *J. Math. Psych.* **46**(6) 769–788.

Pérez, J. 1995. Some comments on Saaty's AHP. *Management Sci.* **41**(6) 1091–1095.

Prelec, D. 1998. The probability weighting function. *Econometrica* **66**(3) 497–527.

Saaty, T. L. 1977. A scaling method for priorities in a hierarchichal structure. *J. Math. Psych.* **15**(3) 234–281.

Saaty, T. L. 1980. *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Saaty, T. L. 1986. Axiomatic foundation of the analytic hierarchy process. *Management Sci.* **32**(7) 841–855.

Saaty, T. L. 1990. An exposition of the AHP in reply to the paper "Remarks on the analytic hierarchy process." *Management Sci.* **36**(3) 259–268.

Saaty, T. L. 2003. Decision-making with the AHP: Why is the principal eigenvector necessary? *Eur. J. Oper. Res.* **145**(1) 85–91.

Shepard, R. N. 1978. On the status of "direct" psychological measurement. C. W. Savage, ed. *Minnesota Studies in the Philosophy of Science*, Vol. 9. University of Minnesota Press, Minneapolis, 441–490.

Smith, J. E., D. von Winterfeldt. 2004. Decision analysis in *Management Science*. *Management Sci.* **50**(5) 561–574.

Steingrimsson, R., R. D. Luce. 2005a. Evaluating a model of global psychophysical judgments—I: Behavioral properties of summations and productions. *J. Math. Psych.* **49**(4) 290–307.

Steingrimsson, R., R. D. Luce. 2005b. Evaluating a model of global psychophysical judgments—II: Behavioral properties linking summations and productions. *J. Math. Psych.* **49**(4) 308–319.

Steingrimsson, R., R. D. Luce. 2007. Empirical evaluation of a model of global psychophysical judgments: IV. Forms for the weighting function. *J. Math. Psych.* **51**(1) 29–44.

Stevens, S. S. 1946. On the theory of scales of measurement. *Science* **103**(2684) 677–680.

Stevens, S. S. 1951. *Handbook of Experimental Pyschology*. Wiley, New York.

Stevens, S. S. 1975. *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects*. Wiley, New York.

Tversky, A., D. Kahneman. 1992. Advances in Prospect Theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* **5**(4) 297–323.

Zahedi, F. 1986. A simulation study of estimation methods in the analytic hierarchy process. *Socio-Econom. Planning Sci.* **20**(6) 347–354.

Zimmer, K. 2005. Examining the validity of numerical ratios in loudness fractionation. *Perception Psychophysics* **67**(4) 569–579.

MANAGEMENT SCIENCE
DOI 10.1287/mnsc.1090.1123ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"The Analytic Hierarchy Process and the Theory of Measurement" by Michele Bernasconi, Christine Choirat, and Raffaello Seri, *Management Science*, DOI 10.1287/mnsc.1090.1123.

Submitted to *Management Science*
manuscript

Electronic Companion - Statistical theory for "The Analytic Hierarchy Process and the Theory of Measurement"

Michele Bernasconi
Dipartimento di Scienze Economiche and SSE, Università "Ca' Foscari", Cannaregio 873, I-30121 Venezia, Italy, bernasconi@unive.it

Christine Choirat
Department of Quantitative Methods, School of Economics and Business Management, Universidad de Navarra, E-31080 Pamplona, Spain, cchoirat@unav.es

Raffaello Seri
Dipartimento di Economia, Università dell'Insubria, Via Monte Generoso 71 , I-21100 Varese, Italy, raffaello.seri@uninsubria.it

1. Introduction

In the following Electronic appendix we provide the precise derivation of the asymptotic theory of estimator $(\widehat{w}_1,\dots,\widehat{w}_n,\widehat{\beta}_3)$ presented in Section 4 of the paper "The Analytic Hierarchy Process and the Theory of Measurement".

2. Model and assumptions

The model is:

$$\begin{aligned} aij &= \frac{w_i}{w_j}\cdot\exp\left[\beta_3\cdot\left(\ln\frac{w_i}{w_j}\right)^3\right]\cdot\exp\varepsilon_{ij} \\ \ln a_{ij} &= \ln w_i-\ln w_j+\beta_3\cdot(\ln w_i-\ln w_j)^3+\varepsilon_{ij}. \end{aligned}$$

We want to obtain estimates of $\mathbf{w}=(w_i,\dots,w_n)'$ and β_3. In order to do so, we minimize the following objective function:

$$Q(\mathbf{w},\beta_3)\triangleq\sum_{i\neq j=1}^{n}\varepsilon_{ij}^2=\sum_{i\neq j=1}^{n}\left[\ln a_{ij}-\ln w_i+\ln w_j-\beta_3\cdot(\ln w_i-\ln w_j)^3\right]^2. \tag{1}$$

We will indicate the objective function as $Q^{(\sigma)}$ in order to stress the dependence on the asymptotic parameter σ; $Q^{(0)}$ is the objective function when $\sigma=0$, while $Q_0^{(\sigma)}$ and $Q_0^{(0)}$ are the previous quantities when evaluated at the true parameters $\mathbf{w}_0$ and $\beta_{3,0}$; it is clear that $Q_0^{(0)}\equiv 0$. When needed, we will write $\boldsymbol{\theta}=(\mathbf{w},\beta_3)$ and we will indicate respectively the gradient and the Hessian with respect to $\boldsymbol{\theta}$ with one and two dots.

We make the following assumptions.

Ass. 1 The estimator $\widehat{\boldsymbol{\theta}}$ is obtained minimizing the function (1) under the constraint $\sum_{i=1}^n w_i=1$.

Ass. 2 Let $\mathbf{E}_0$ be the skew-symmetric matrix obtained when $\mathbf{w}=\mathbf{w}_0$ and $\beta_3=\beta_{3,0}$, with generic (i,j) −element $\varepsilon_{0,ij}$. Let $\boldsymbol{\varepsilon}_0=\widetilde{v}(\mathbf{E}_0)$ be the $\left(\left(\frac{n^2-n}{2}\right)\times 1\right)$ vector obtained stacking the subdiagonal elements of $\mathbf{E}_0$; $\boldsymbol{\varepsilon}_0$ is such that $\sigma^{-1}\boldsymbol{\varepsilon}_0\xrightarrow{\mathcal{D}}\mathcal{N}\left(\mathbf{0},\mathbf{I}_{\left(\frac{n^2-n}{2}\right)}\right)$. Asymptotic results are stated for $\sigma\downarrow 0$.

Ass. 3 The parameter w_i takes its value in the interval $[\varepsilon_i,1-\varepsilon_i]$ for some $\varepsilon_i>0$; the parameter β belongs to a compact interval $[\beta_L,\beta_U]$; the weights respect the equality $\sum_{i=1}^n w_i=1$. The parameter space is $\Theta = \{\prod_{i=1}^{n}[\varepsilon_i, 1-\varepsilon_i] \cap \{\sum_{i=1}^{n} w_i = 1\}\} \times [\beta_L, \beta_U]$; moreover $\boldsymbol{\theta}_0 \in \text{relint}\Theta = \{\prod_{i=1}^{n}(\varepsilon_i, 1-\varepsilon_i) \cap \{\sum_{i=1}^{n} w_i = 1\}\} \times (\beta_L, \beta_U)$.

We remark that the asymptotic parameter is σ and not n. This implies that even if these results may seem standard, they are not. In particular, normalizations through functions of n are very important since different normalizations (e.g. by n and by $n-1$) do not lead to the same asymptotic behavior. We also remark that the requirement that $w_i \in [\varepsilon_i, 1-\varepsilon_i]$ is in line with Axiom 2 of Saaty (1986).

3. Propositions

Consider the indexes $1 \le i, j \le n$ with $i > j$, and let $k = (j-1)\cdot n + i - {}^{j(j+1)}/_2$. Then we will need the $\left((n+1) \times \binom{n^2-n}{2}\right)$ matrix $\mathbf{Q}_0$ given by:

$$
\begin{aligned}
[\mathbf{Q}_0]_{(h,k)} = \quad & 4 \cdot \mathbf{1}_{i=h} \cdot \left(-\frac{1}{w_{0,i}} - \frac{3\beta_0}{w_{0,i}} \cdot (\ln w_{0,i} - \ln w_{0,j})^2\right) \\
& +4 \cdot \mathbf{1}_{j=h} \cdot \left(\frac{1}{w_{0,j}} + \frac{3\beta_0}{w_{0,j}} \cdot (\ln w_{0,i} - \ln w_{0,j})^2\right),
\end{aligned}
$$

for $1 \le h \le n, \quad 1 \le k \le \binom{n^2-n}{2}$,

$$
[\mathbf{Q}_0]_{(n+1,k)} = 4 \cdot (\ln w_{0,j} - \ln w_{0,i})^3
$$

for $1 \le k \le \binom{n^2-n}{2}$.

Matrix $\mathbf{Q}_0$ is such that $\dot{Q}_0^{(\sigma)} = \mathbf{Q}_0 \cdot \boldsymbol{\varepsilon}_0$, where $\boldsymbol{\varepsilon}_0$ is defined in Ass. 2 (it can be checked that the $k-$th element of $\boldsymbol{\varepsilon}_0$ is $\varepsilon_{0,ij}$, where $k = (j-1)\cdot n + i - {}^{j(j+1)}/_2$ and $i > j$).

The following Proposition shows that consistency and asymptotic normality (when σ^2 is known) hold for this estimator.

Proposition 3.1 *Under Ass. 1 and 3, the estimator $\widehat{\boldsymbol{\theta}}$ is weakly consistent for $\boldsymbol{\theta}_0$ as $\sigma \downarrow 0$. Under Ass. 1-3 it has the following asymptotic distribution:*

$$
\sigma^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, 16 \cdot \mathbf{D}_n \left(\mathbf{D}_n' \mathbf{Q}_0 \mathbf{Q}_0' \mathbf{D}_n\right)^{-1} \mathbf{D}_n'\right)
$$

where $\mathbf{D}_n$ is the $((n+1)\times n)$ matrix $\mathbf{D}_n = \begin{bmatrix} -\left(\mathbf{e}_{n-1}', 0\right) \\ \mathbf{I}_n \end{bmatrix}$.

Now we consider the situation in which σ is replaced with an estimator.

Proposition 3.2 *Under Ass. 1-3, the estimator:*

$$\widehat{\sigma}^2 = \frac{1}{n^2 - 3n} \cdot \sum_{i \neq j = 1}^{n} \widehat{\varepsilon}_{ij}^2$$

is asymptotically unbiased (in the sense that $\mathbb{E}\frac{\widehat{\sigma}^2}{\sigma^2} \to 1$*) and has asymptotic distribution:*

$$\left(\frac{n^2 - 3n}{2}\right) \cdot \frac{\widehat{\sigma}^2}{\sigma^2} \to_{\mathcal{D}} \chi^2\left(\frac{n^2 - 3n}{2}\right).$$

Consider a full row rank $(m \times (n+1))$ *matrix* $\mathbf{\Gamma}$*. Then:*

$$\frac{\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \mathbf{\Gamma}' \left\{\mathbf{\Gamma}\mathbf{D}_n \left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1} \mathbf{D}_n'\mathbf{\Gamma}'\right\}^{-1} \mathbf{\Gamma}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)}{16\widehat{\sigma}^2 m} \to_{\mathcal{D}} F\left(m, \frac{n^2 - 3n}{2}\right).$$

When $m = 1$*, in particular:*

$$\frac{\mathbf{\Gamma}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)}{4\widehat{\sigma}\left\{\mathbf{\Gamma}\mathbf{D}_n \left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1} \mathbf{D}_n'\mathbf{\Gamma}'\right\}^{1/2}} \to_{\mathcal{D}} t\left(\frac{n^2 - 3n}{2}\right).$$

Consider the residual $\widehat{\varepsilon}_{ij}$ *with* $i > j$*. Let* $\mathbf{e}_k$ *be the* $((n+1) \times 1)$ *vector with* 1 *in the* $k-th$ *position and all other components equal to 0. Then the following result holds with* $k = (j-1) \cdot n + i - j(j+1)/2$*:*

$$\frac{\widehat{\varepsilon}_{ij}}{\widehat{\sigma} \cdot \sqrt{1 - \mathbf{e}_k'\mathbf{Q}_0'\mathbf{D}_n \left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1} \mathbf{D}_n'\mathbf{Q}_0\mathbf{e}_k}} \to_{\mathcal{D}} t\left(\frac{n^2 - 3n}{2}\right).$$

Proof of Proposition 3.1. We start with a result of consistency. When $\sigma = 0$, $\ln a_{ij} = \ln w_{0,i} - \ln w_{0,j} - \beta_0 \cdot (\ln w_{0,i} - \ln w_{0,j})^3$; thus when we have

$$\begin{aligned} Q^{(\sigma)}(\mathbf{w}, \beta) - Q^{(0)}(\mathbf{w}, \beta) = {} & 2 \sum_{i \neq j = 1}^{n} \varepsilon_{0,ij} \cdot \left[\ln w_{0,i} - \ln w_{0,j} + \beta_0 \cdot (\ln w_{0,i} - \ln w_{0,j})^3\right] \\ & -2 \sum_{i \neq j = 1}^{n} \varepsilon_{0,ij} \cdot \left[\ln w_i - \ln w_j + \beta \cdot (\ln w_i - \ln w_j)^3\right] \\ & + \sum_{i \neq j = 1}^{n} \varepsilon_{0,ij}^2. \end{aligned}$$

This converges uniformly in probability to 0 under Ass. 3 when $\sigma \downarrow 0$. $Q^{(0)}$ is continuous on a compact space and is uniquely minimized at $w_i = w_{0,i}$ and $\beta = \beta_0$. Therefore Theorem 2.1 in Newey and McFadden (1994) applies.

For the asymptotic distribution, we reason in terms of $\tilde{\boldsymbol{\theta}} = \left[\mathbf{0}_{n\times 1}\ \mathbf{I}_n\right]\cdot\boldsymbol{\theta}$, that is $\boldsymbol{\theta}$ without the first component; remark that $\boldsymbol{\theta}$ can be recovered as $\boldsymbol{\theta} = \begin{bmatrix} 1 \\ \mathbf{0}_{n\times 1}\end{bmatrix} + \begin{bmatrix} -\left(\mathbf{e}'_{n-1}, 0\right) \\ \mathbf{I}_n \end{bmatrix}\tilde{\boldsymbol{\theta}}$. Therefore, the gradient and the Hessian are $\frac{\partial Q^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}} = \mathbf{D}'_n\cdot\frac{\partial Q^{(\sigma)}}{\partial\boldsymbol{\theta}}$ and $\frac{\partial^2 Q^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}\partial\tilde{\boldsymbol{\theta}}'} = \mathbf{D}'_n\cdot\frac{\partial^2 Q^{(\sigma)}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\cdot\mathbf{D}_n$. An expansion of the first order conditions $\frac{\partial Q^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}}\left(\widehat{\tilde{\boldsymbol{\theta}}}\right) = \mathbf{0}$ yields:

$$\begin{aligned}\frac{\partial Q_0^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}} + \frac{\partial^2 Q^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}\partial\tilde{\boldsymbol{\theta}}'}\left(\boldsymbol{\theta}^\star\right)\cdot\left(\widehat{\tilde{\boldsymbol{\theta}}} - \tilde{\boldsymbol{\theta}}_0\right) &= 0 \\ \sigma^{-1}\left(\widehat{\tilde{\boldsymbol{\theta}}} - \tilde{\boldsymbol{\theta}}_0\right) &= -\left(\frac{\partial^2 Q^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}\partial\tilde{\boldsymbol{\theta}}'}\left(\boldsymbol{\theta}^\star\right)\right)^{-1}\cdot\sigma^{-1}\frac{\partial Q_0^{(\sigma)}}{\partial\tilde{\boldsymbol{\theta}}}.\end{aligned}$$

Therefore the distribution of $\sigma^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ can be obtained as:

$$\sigma^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = -\mathbf{D}_n\left(\mathbf{D}'_n\cdot\frac{\partial^2 Q^{(\sigma)}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left(\boldsymbol{\theta}^\star\right)\cdot\mathbf{D}_n\right)^{-1}\mathbf{D}'_n\cdot\sigma^{-1}\frac{\partial Q_0^{(\sigma)}}{\partial\boldsymbol{\theta}}.$$

We start from the behavior of the gradient.

The gradient is:

$$\begin{aligned}\frac{\partial Q^{(\sigma)}\left(\mathbf{w},\beta\right)}{\partial\beta} &= -2\cdot\sum_{i\neq j=1}^{n}\left(\ln a_{ij} - \ln w_i + \ln w_j - \beta\cdot\left(\ln w_i - \ln w_j\right)^3\right)\cdot\left(\ln w_i - \ln w_j\right)^3 \\ \frac{\partial Q^{(\sigma)}\left(\mathbf{w},\beta\right)}{\partial w_i} &= -4\cdot\sum_{j\in\{1,\dots,n\}\setminus i}\frac{\left(\ln a_{ij} - \ln w_i + \ln w_j - \beta\cdot\left(\ln w_i - \ln w_j\right)^3\right)}{w_i} \\ &\quad\cdot\left[1 + 3\beta\cdot\left(\ln w_i - \ln w_j\right)^2\right].\end{aligned}$$

Now we consider $\dot{Q}_0^{(\sigma)}$:

$$\begin{aligned}\frac{\partial Q^{(\sigma)}\left(\mathbf{w}_0,\beta_0\right)}{\partial\beta} &= -2\cdot\sum_{i\neq j=1}^{n}\varepsilon_{ij,0}\cdot\left(\ln w_{i,0} - \ln w_{j,0}\right)^3 \\ \frac{\partial Q^{(\sigma)}\left(\mathbf{w}_0,\beta_0\right)}{\partial w_i} &= -4\cdot\sum_{j\in\{1,\dots,n\}\setminus i}\frac{\varepsilon_{ij,0}}{w_{i,0}}\cdot\left[1 + 3\beta_0\cdot\left(\ln w_{i,0} - \ln w_{j,0}\right)^2\right].\end{aligned}$$

The fact that $\dot{Q}_0^{(\sigma)} = \mathbf{Q}_0\cdot\boldsymbol{\varepsilon}_0$ implies that $\mathbb{V}\left(\sigma^{-1}\dot{Q}_0^{(\sigma)}\right) = \mathbf{Q}_0\mathbf{Q}'_0$ and $\sigma^{-1}\dot{Q}_0^{(\sigma)} = \mathbf{Q}_0\cdot\sigma^{-1}\boldsymbol{\varepsilon}_0 \to_{\mathcal{D}} \mathcal{N}\left(\mathbf{0},\mathbf{Q}_0\mathbf{Q}'_0\right)$

The asymptotic Hessian for $\sigma\downarrow 0$ is:

$$\frac{\partial^2 Q^{(\sigma)}\left(\mathbf{w},\beta,\lambda\right)}{\partial\beta^2} \to \frac{\partial^2 Q^{(0)}\left(\mathbf{w},\beta,\lambda\right)}{\partial\beta^2} = 2\cdot\sum_{i\neq j=1}^{n}\left(\ln\frac{w_i}{w_j}\right)^6$$

$$
\begin{aligned}
\frac{\partial^2 Q^{(\sigma)}(\mathbf{w},\beta,\lambda)}{\partial w_i^2} &\to \frac{\partial^2 Q^{(0)}(\mathbf{w},\beta,\lambda)}{\partial w_i^2} = 4\cdot\sum_{j\in\{1,\ldots,n\}\setminus i}\frac{1}{w_i^2}\cdot\left(1+3\beta\cdot\left(\ln\frac{w_i}{w_j}\right)^2\right)^2\\
\frac{\partial^2 Q^{(\sigma)}(\mathbf{w},\beta,\lambda)}{\partial w_i\partial w_j} &\to \frac{\partial^2 Q^{(0)}(\mathbf{w},\beta,\lambda)}{\partial w_i\partial w_j} = -4\cdot\frac{1}{w_iw_j}\cdot\left[1+3\beta\cdot\left(\ln\frac{w_i}{w_j}\right)^2\right]^2\\
\frac{\partial^2 Q^{(\sigma)}(\mathbf{w},\beta,\lambda)}{\partial w_i\partial \beta} &\to \frac{\partial^2 Q^{(0)}(\mathbf{w},\beta,\lambda)}{\partial w_i\partial \beta} = 4\cdot\sum_{j\in\{1,\ldots,n\}\setminus i}\frac{\left(\ln\frac{w_i}{w_j}\right)^3}{w_i}\cdot\left[1+3\beta\cdot\left(\ln\frac{w_i}{w_j}\right)^2\right].
\end{aligned}
$$

Under Ass. 2 and 3 convergence is uniform. Therefore, we have $\mathbb{V}\left(\dot{Q}_0^{(\sigma)}\right) = \sigma^2\cdot\mathbf{Q}_0\mathbf{Q}_0' = 4\sigma^2\cdot\ddot{Q}_0^{(0)}$, that is the variance of the gradient is $4\sigma^2$ times the limiting Hessian. Therefore:

$$
\begin{aligned}
\sigma^{-1}\left(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0\right) &= -\mathbf{D}_n\left(\mathbf{D}_n'\cdot\frac{\partial^2 Q^{(\sigma)}}{\partial\boldsymbol\theta\partial\boldsymbol\theta'}(\boldsymbol\theta^\star)\cdot\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\cdot\sigma^{-1}\frac{\partial Q_0^{(\sigma)}}{\partial\boldsymbol\theta}\\
&\sim -4\cdot\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\cdot\frac{\boldsymbol\varepsilon_0}{\sigma}\\
\sigma^{-1}\left(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0\right) &\to_{\mathcal{D}} \quad \mathcal{N}\left(\mathbf{0},16\cdot\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\right).
\end{aligned}
$$

Remark that $\boldsymbol\theta_0$ can be replaced with any consistent estimator such as $\widehat{\boldsymbol\theta}$.

Proof of Proposition 3.2. Let $\mathbf{E}$ be the skew-symmetric matrix with generic (i,j) −element $\varepsilon_{ij} = \ln a_{ij} - \ln w_i + \ln w_j - \beta\cdot(\ln w_i - \ln w_j)^3$; consider the $\left(\left(\frac{n^2-n}{2}\right)\times 1\right)$ vector $\boldsymbol\varepsilon = \widetilde{v}(\mathbf{E})$. An expansion around $\boldsymbol\varepsilon_0$ can then be obtained as $\boldsymbol\varepsilon\sim\boldsymbol\varepsilon_0+\frac{\partial\boldsymbol\varepsilon}{\partial\boldsymbol\theta'}\cdot(\boldsymbol\theta-\boldsymbol\theta_0)$. Through direct computation, it is simple to show the equality $\frac{\partial\boldsymbol\varepsilon}{\partial\boldsymbol\theta'}=\frac{1}{4}\cdot\mathbf{Q}_0$, so that through Proposition 3.1 $\widehat{\boldsymbol\varepsilon}\sim\boldsymbol\varepsilon_0+\frac{\partial\boldsymbol\varepsilon}{\partial\boldsymbol\theta'}\cdot\left(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0\right)$ becomes:

$$
\begin{aligned}
\widehat{\boldsymbol\varepsilon} &\sim \boldsymbol\varepsilon_0-\mathbf{Q}_0\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\cdot\boldsymbol\varepsilon_0\\
&=\left\{\mathbf{I}_{\frac{n(n-1)}{2}}-\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\right\}\cdot\boldsymbol\varepsilon_0.
\end{aligned}
$$

This implies:

$$
\mathbb{V}(\widehat{\boldsymbol\varepsilon})\sim\sigma^2\cdot\left\{\mathbf{I}_{\frac{n(n-1)}{2}}-\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\right\}.
$$

On the other hand:

$$
\sum_{i\neq j=1}^{n}\frac{\widehat{\varepsilon}_{ij}^2}{\sigma^2}=\frac{2\widehat{\boldsymbol\varepsilon}'\widehat{\boldsymbol\varepsilon}}{\sigma^2}\sim 2\cdot\frac{\boldsymbol\varepsilon_0'}{\sigma}\cdot\left\{\mathbf{I}_{\frac{n(n-1)}{2}}-\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\right\}\cdot\frac{\boldsymbol\varepsilon_0}{\sigma}.
$$

Therefore:

$$\begin{aligned}\mathbb{E}\left(\sum_{i\neq j=1}^{n}\frac{\widehat{\varepsilon}_{ij}^{2}}{\sigma^{2}}\right) &\sim 2\cdot\mathbb{E}\mathrm{tr}\left\{\mathbf{I}_{\frac{n(n-1)}{2}}-\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\right\}\\ &= n\left(n-1\right)-2\mathrm{tr}\left(\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\right)\\ &= n^2-3n.\end{aligned}$$

We consider the estimator:

$$\widehat{\sigma}^2=\frac{1}{n^2-3n}\cdot\sum_{i\neq j=1}^{n}\widehat{\varepsilon}_{ij}^{2}.$$

Clearly this estimator is asymptotically unbiased.

The matrix $\mathbf{A}_1=\mathbf{I}_{\frac{n(n-1)}{2}}-\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0$ is idempotent and has rank equal to its trace, that is $\frac{n^2-3n}{2}$. The asymptotic distribution is:

$$\begin{aligned}&\left(\frac{n^2-3n}{2}\right)\cdot\frac{\widehat{\sigma}^2}{\sigma^2}=\\ &\quad\frac{1}{2}\sum_{i\neq j=1}^{n}\frac{\widehat{\varepsilon}_{ij}^{2}}{\sigma^{2}}\sim\frac{\boldsymbol{\varepsilon}_0'}{\sigma}\cdot\left\{\mathbf{I}_{\frac{n(n-1)}{2}}-\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\right\}\cdot\frac{\boldsymbol{\varepsilon}_0}{\sigma}\rightarrow_{\mathcal{D}}\chi^2\left(\frac{n^2-3n}{2}\right).\end{aligned}$$

Now we work out the distribution of a vector of linear combinations of the regression parameters. Consider the k combinations $\sigma^{-1}\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)$ of the regression parameters $\widehat{\boldsymbol{\theta}}$ represented by the $(m\times(n+1))$ matrix $\boldsymbol{\Gamma}$, of full row rank. Remark that the quadratic form $\left(\frac{n^2-3n}{2}\right)\cdot\frac{\widehat{\sigma}^2}{\sigma^2}$ is asymptotically independent of the linear form $\sigma^{-1}\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)$.

From:

$$\begin{aligned}&\sigma^{-2}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\boldsymbol{\Gamma}'\left\{\mathbb{V}\left[\sigma^{-1}\cdot\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\right]\right\}^{-1}\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\ &\sim\frac{1}{16\sigma^2}\cdot\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\boldsymbol{\Gamma}'\left\{\boldsymbol{\Gamma}\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\boldsymbol{\Gamma}'\right\}^{-1}\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\rightarrow_{\mathcal{D}}\chi^2\left(k\right)\end{aligned}$$

we get:

$$\begin{aligned}&\frac{\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\boldsymbol{\Gamma}'\left\{\boldsymbol{\Gamma}\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\boldsymbol{\Gamma}'\right\}^{-1}\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)}{16\widehat{\sigma}^2m}\\ &\rightarrow_{\mathcal{D}}\frac{\chi^2(m)/m}{\chi^2\left(\frac{n^2-3n}{2}\right)/\left(\frac{n^2-3n}{2}\right)}=F\left(m,\frac{n^2-3n}{2}\right).\end{aligned}$$

When $m=1$, in particular:

$$\frac{\left\{\mathbf{\Gamma}\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{\Gamma}'\right\}^{-1/2}\mathbf{\Gamma}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)}{4\widehat{\sigma}}\to_{\mathcal{D}} t\left(\frac{n^2-3n}{2}\right).$$

Finally, we develop the distribution of the residuals. Consider $\widehat{\varepsilon}_{ij}=\mathbf{e}_k'\widehat{\boldsymbol{\varepsilon}}$ where $\mathbf{e}_k$ is a vector of zeros with 1 at position $k=(j-1)\cdot n+i-{}^{j(j+1)}/2$:

$$\sigma^{-1}\cdot\widehat{\varepsilon}_{ij}\sim\mathcal{N}\left(0,1-\mathbf{e}_k'\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\mathbf{e}_k\right)$$

$$\frac{\widehat{\varepsilon}_{ij}}{\sigma\cdot\sqrt{1-\mathbf{e}_k'\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\mathbf{e}_k}}\to_{\mathcal{D}}\mathcal{N}(0,1).$$

If we replace σ with $\widehat{\sigma}$, we get:

$$\frac{\widehat{\varepsilon}_{ij}}{\widehat{\sigma}\cdot\sqrt{1-\mathbf{e}_k'\mathbf{Q}_0'\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{Q}_0\mathbf{Q}_0'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'\mathbf{Q}_0\mathbf{e}_k}}\to_{\mathcal{D}} t\left(\frac{n^2-3n}{2}\right).$$

References

Newey, W.K., and McFadden, D. (1994). Large sample estimation and hypothesis testing. In D. McFadden and R. Engle (eds.), *Handbook of Econometrics*,(Vol. 4). Amsterdam: Elsevier, North-Holland.