

## USING EXTERNAL INFORMATION FOR MORE PRECISE INFERENCES IN GENERAL REGRESSION MODELS

MARTIN JANN 

UNIVERSITY OF HAMBURG

MARTIN SPIESS 

UNIVERSITY OF HAMBURG

Empirical research usually takes place in a space of available external information, like results from single studies, meta-analyses, official statistics or subjective (expert) knowledge. The available information ranges from simple means and proportions to known relations between a multitude of variables or estimated distributions. In psychological research, external information derived from the named sources may be used to build a theory and derive hypotheses. In addition, techniques do exist that use external information in the estimation process, for example prior distributions in Bayesian statistics. In this paper, we discuss the benefits of adopting generalized method of moments with external moments, as another example for such a technique. Analytical formulas for estimators and their variances in the multiple linear regression case are derived. An R function that implements these formulas is provided in the supplementary material for general applied use. The effects of various practically relevant moments are analyzed and tested in a simulation study. A new approach to robustify the estimators against misspecification of the external moments based on the concept of imprecise probabilities is introduced. Finally, the resulting externally informed model is applied to a dataset to investigate the predictability of the premorbid intelligence quotient based on lexical tasks, leading to a reduction of variances and thus to narrower confidence intervals.

Key words: external information, generalized method of moments, imprecise probabilities.

### 1. Introduction

When planning new empirical studies, researchers are confronted with a variety of information from previous studies, including statistical quantities such as means, variances or confidence intervals. However, this external information is mostly used qualitatively, i.e., to develop new theories, and rarely in a quantitative way, i.e., to estimate parameters. One advantage of using external information to estimate a parameter is that some parameter values can be excluded or considered less likely than without the external information, potentially leading to more efficient estimators. The usage of informed prior distributions, where the external information can be used to specify (certain aspects of) the prior distribution, is well known in Bayesian statistics (Bernardo & Smith, 1994). The underlying goal for its use must be clear. On the one hand, external information can facilitate the fitting or tuning of a model. On the other hand, it can make estimators more robust or efficient. This paper aims to achieve the latter of the two goals. Bayesian statistics refers to this as statistical elicitation (Kadane & Wolfson, 1998). The objective is to translate expert knowledge into a prior distribution. Therefore, many psychological biases, such as judgment by representativeness, availability, anchoring, adaptation, or hindsight bias and the intentional misleading by experts, must be considered. It should be noted that the aim is not

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11336-024-09953-w>.

Correspondence should be made to Martin Jann, Department of Psychology, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany. Email: [martin.jann@uni-hamburg.de](mailto:martin.jann@uni-hamburg.de)

to achieve objectivity but to ensure a proper statistical representation of subjective knowledge (Garthwaite et al., 2005; Lele & Das, 2000). However, we believe that in applied psychological research, the researcher is usually the one who selects the external information, but is susceptible to the same psychological biases, e.g., in deciding which studies to include. Moreover, the difficulties in eliciting a (multivariate) prior distribution are well documented (Garthwaite et al., 2005, pp. 686–688). The method proposed in this paper allows a simplification of the elicitation compared to Bayesian statistics, since only moments need to be elicited. The elicitation of moments has been well studied for correlations, means, medians, or variances (Garthwaite et al., 2005). In Bayesian elicitation, there are several possible prior distributions for these externally given moments, e.g., with the same expected value or the same correlation, leading to different posterior distributions and thus potentially different results. This problem of prior sensitivity was addressed by Berger (1990) and led to work on robust Bayesian analysis (for an overview, see Insua & Ruggeri, 2000). However, it is somewhat arbitrary to choose the class of distributions for which one wants to make the analysis robust (Garthwaite et al., 2005, p. 695). In our framework, no restriction to a particular class of distributions is required, since it relies solely on moment information and a central limit theorem.

Another important point is that external information may not in general be precise and correct. As nearly all of the external quantities are estimates themselves, they are at least prone to sampling variation. If the external information is not correct (e.g., due to poor sampling or measurement protocols), its use can lead to biased conclusions that may even be worse than without external information. To address this problem, we suggest using an interval for the external information instead of point values, enabling researchers to incorporate any uncertainty about the external moments into the analysis. Inserting external intervals into estimators results in the imprecise probabilistic concept of *feasible probability* (*F-probability*) discussed in Sect. 4 (Augustin et al., 2014; Weichselberger, 2001). This approach provides an alternative way to enhance the robustness of elicitation compared to the classical Bayesian paradigm: Using intervals can reflect uncertainty about moments, and the resulting inference is still coherent if the interval contains the true value. However, researchers must be cautious of and avoid overconfidence bias when eliciting intervals; that is, the tendency to select intervals that are too narrow to represent current uncertainty (Winman et al., 2004). A test of the latter assumption is available, more specifically a test of the compatibility of the external interval and the data, which could serve as a pretest before applying the methods proposed here (Jann, 2023).

The insertion of intervals into estimators resembles creating fuzzy numbers (Kwakernaak, 1978; Zadeh, 1965), for which generalizations of traditional statistical methods already exist. This is particularly true for the special case of triangular numbers (Buckley, 2004). The possibility distributions induced by triangular numbers constitute special cases of imprecise probabilities and are constructed based on only one distribution (Augustin et al., 2014, pp. 84–87). This is the key difference between triangular numbers and *F-probabilities*, since the latter are constructed from a set of possible probability distributions, which can enhance the robustness of the outcomes compared to constructions based on only one distribution. Another difference lies in the fact that triangular numbers are constructed by varying the confidence probability of a confidence interval based on the estimator, while the external interval we use in this paper is fixed. Moreover, there is no probabilistic statement about the values within that interval.

In the present study, we analyze the frequentist properties of estimators if external information is used, that can be expressed as moment conditions and thus does not use complete distributions as prior information. To our knowledge, there is no general framework for robustly incorporating such quantitative external information into frequentist analysis. Since this would offer the advantage of improving upon classical inference procedures widely used in psychology, our goal is to present such a framework. The use of these external moment conditions in addition to the moment conditions used to estimate the model parameters leads to an overidentified system of moment

conditions. The main idea to find well performing estimators for such “externally” overidentified systems is the framework of the Generalized Method of Moments (GMM) (Hansen, 1982). This idea has already been used in the econometric literature, for example, by Imbens and Lancaster (1994) who combine micro- and macro-economic data and by Hellerstein and Imbens (1999) by constructing weights for regression models based on auxiliary data. A different yet related way to incorporate external moment-information is the empirical likelihood approach (Owen, 1988). This technique is quite frequently used in the literature, for example, in finite population estimation (Zhong & Rao, 2000) and for externally informed generalized linear models (Chaudhuri et al., 2008). Both approaches have in common that the use of external information may increase the efficiency of an estimator and/or reduce its bias.

Actually, in Sect. 3, we show that there will always be a variance reduction, if the external moment conditions and the ones for the model are correlated and if the covariance matrix of all moment conditions is positive definite. As the GMM allows the estimation of a large class of models, and many statistical measures like proportions, means, variances and covariances are statistical moments, the range of possible applications is large but far from being implemented in psychological research. For a multiple linear model, we derive the estimators analytically in Sect. 3. The use of imprecise probabilities will increase the overall variation of the estimator, and moreover, the effect of the variance reduction will decrease. As we will demonstrate, however, variance reduction will still be possible while increasing the robustness of the estimation. The proposed method and techniques allow more precise and robust inferences, which is particularly relevant in small samples. To illustrate the small sample performance of the externally informed models in multiple linear models, a simulation study is presented in Sect. 5. An application to a real data set analyzing the relation of premorbid (general) intelligence and performance in lexical tasks (Pluck & Ruales-Chieruzzi, 2021) is presented in Sect. 6.

## 2. Externally Informed Models

In a first step, we assume that precise external information is available, an assumption that will be relaxed in Sect. 4. Throughout, we assume that all variables will be considered as random variables if not given otherwise. For notational clarity, we will always write single-valued random variables in italic small letters. Vectors as well as vector-valued functions will be written in small bold letters and matrices in bold capital letters.

Although the basic concepts are presented in the following section, for the class of general regression models, we will consider the family of linear models for their illustration in a concrete class of models due to their frequent use. Note that, for example, ANOVA models are special cases of this model, however, with fixed factors instead of random covariates. Nevertheless, the results derived in this paper carry over to these models.

Let  $\mathbf{z} = (z_1, \dots, z_p)^T$  be a real-valued random vector and  $\mathbf{z}_i, i = 1, \dots, n$ , be i.i.d. random vectors distributed like  $\mathbf{z}$ , representing the data. Suppose we want to fit a regression model to this data set with fixed parameter  $\boldsymbol{\theta} \in \mathbb{R}^p$ , where the adopted model reflects the interesting aspects of the true data-generating process and  $\boldsymbol{\theta}_0$  is the true parameter value. In linear regression models, the parameter of scientific interest is usually the parameter of the mean structure denoted as  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  with true value  $\boldsymbol{\beta}_0$ . The notation  $\boldsymbol{\beta}$  will only be used for linear regression models, while we will use  $\boldsymbol{\theta}$  to denote the regression coefficients in general regression models. The random vector  $\mathbf{z}$  is given by  $\mathbf{z} = (\mathbf{x}^T, y)^T$  with random explanatory variables  $\mathbf{x} = (x_1, \dots, x_p)^T$  and dependent variable  $y$ . Accordingly, the unit specific i.i.d. random vectors  $\mathbf{z}$  are written as  $\mathbf{z}_i = (\mathbf{x}_i^T, y_i)^T$  for  $i = 1, \dots, n$ . Hence, the random  $(n \times p)$ -design matrix is  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ , and we write  $\mathbf{y} = (y_1, \dots, y_n)^T$ .

The multiple linear model can now be written as  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}$  with random error terms  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ . As an illustration, suppose we want to investigate the effect of the explanatory variables *fluid intelligence* and *depression* on the dependent variable *mathematics skills*. We could design a study, in which fluid intelligence and math skills are measured via Cattell's fluid intelligence test, in short CFT 20-R, ( $x_2$ ) and the number sequence test ZF-R ( $y$ ), respectively (Weiss, 2006). Depression could be measured as a binary variable indicating if a person has a depression-related diagnosis ( $x_3$ ). The model could be a linear multiple regression of the ZF-R score on the depression indicator and the CFT 20-R score for fluid intelligence. To include the intercept,  $x_1$  is a degenerate variable with value 1.

In addition to the observed data and the assumptions justifying the model, we often have available external information like means, correlations or proportions, e.g., through official statistics, meta-analyses or already existing individual studies. In our applied example, there are various German norm groups for the CFT 20-R and the ZF-R, even for different ages (Weiss, 2006). Hence, we could always transform the results into scores with known expected value and variance, i.e. the CFT 20-R score can be transformed into an IQ-score based on a recent calibration sample from 2019, reported in the test manual (Weiss, 2019). Regarding the relation of fluid intelligence and math skills, a recent meta-analysis based on more than 370,000 participants in 680 studies from multiple countries suggests a correlation of  $r = 0.41$  between the two variables (Peng et al., 2019). In addition, based on a study covering 87% of the German population aged at least 15 years, Steffen et al. (2020) report a prevalence of depression, defined as a F32, F33 or F34.1 diagnosis following the ICD-10-GM manual, of 15.7% in 2017.

Let us assume that these values can be interpreted as true population values, an assumption that will be relaxed later. Note that they have the form of statistical moments. For example, the observable depression prevalence is assumed to equal the expected value of the binary depression indicator (first moment), the mean (now considered as expected value) and variance of the test scores are set equal to the first moment and the second central moment, respectively, of the random variables CFT 20-R-score and ZF-R-score. Finally, the correlation is assumed to equal the mixed moment of the standardized CFT 20-R-score and ZF-R-score. Taking  $q$  to be the number of known external moments, we state

**Definition 1.** Let  $M$  be a statistical model. Further let  $\mathbf{u}$  be a  $(q \times 1)$ -vector of statistical moment expressions and  $\boldsymbol{\mu}_{\text{ex}}$  the corresponding  $(q \times 1)$ -vector of externally determined values for the statistical moments in  $\mathbf{u}$ . Then the model combining  $M$  and the conditions  $\mathbf{u} = \boldsymbol{\mu}_{\text{ex}}$  is called *externally informed model*.

To illustrate the definition, we will use the applied example from above in which case the model  $M$  is a multiple linear regression model. Interpreting the norms for the dependent variable ZF-R from the calibration sample as population values, external knowledge about the corresponding moments, for example the means of ZF-R, is available. Let us assume that ZF-R is transformed into the IQ-scale. Then, if  $\mathbf{u} = E(y)$  and  $\boldsymbol{\mu}_{\text{ex}} = 100$ , we get  $E(\mathbf{y}) = 100 \times \mathbf{1}_n = E(\mathbf{X})\boldsymbol{\beta}_0$ , where  $\mathbf{1}_n$  is a  $(n \times 1)$ -vector of ones. Thus,  $\mathbf{u} = \boldsymbol{\mu}_{\text{ex}}$  imposes conditions on  $\boldsymbol{\beta}$ .

### 3. Estimation and Properties of Externally Informed Models

#### 3.1. Generalized Method of Moments with External Moments

The GMM approach (Hansen, 1982) allows to estimate (general) regression models and to incorporate external moments into the estimation (Imbens & Lancaster, 1994). To estimate the parameter of a general regression model, a "model moment function"  $\mathbf{m}(\mathbf{z}, \boldsymbol{\theta})$  must be given, which satisfies the conditions  $E[\mathbf{m}(\mathbf{z}, \boldsymbol{\theta})] = \mathbf{0}$  only for the true parameter value  $\boldsymbol{\theta}_0$ . The corresponding

“sample moment function” for  $\mathbf{z}_i$  will be denoted as  $\mathbf{m}(\mathbf{z}_i, \boldsymbol{\theta})$ . In case of the linear regression model from Sect. 2, the model moment function corresponding to the method of Ordinary Least Squares (OLS) is  $\mathbf{m}(\mathbf{z}, \boldsymbol{\beta}) = \mathbf{x}(y - \mathbf{x}^T \boldsymbol{\beta})$  (Cameron & Trivedi, 2005, p. 172). Given the model is correctly specified, for true parameter value  $\boldsymbol{\beta}_0$ ,  $E[\mathbf{m}(\mathbf{z}, \boldsymbol{\beta}_0)] = E[\mathbf{x}(y - \mathbf{x}^T \boldsymbol{\beta}_0)] = \mathbf{0}$  holds. Replacing these population model moment conditions by corresponding sample model moment conditions,

$$\mathbf{0} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}(\mathbf{z}_i, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) = \frac{1}{n} \mathbf{X}^T (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}),$$

and solving these estimating equations for  $\boldsymbol{\beta}$ , leads to an estimator  $\hat{\boldsymbol{\beta}}$  for  $\boldsymbol{\beta}_0$ . The above conditions are identical to the estimating equations resulting from the least-squares or, if normality of the errors is assumed, the maximum likelihood method. Furthermore, the general classes of  $M$ - and  $Z$ -estimators can be written using estimating equations that have this moment form. This leads to broad applicability, since these classes, for example, include the median and quantiles (Vaart, 1998).

The possibly vector-valued “external moment function” will be denoted as  $\mathbf{h}(\mathbf{z}) = \mathbf{u}(\mathbf{z}) - \boldsymbol{\mu}_{\text{ex}}$ , where the functional form of  $\mathbf{u}(\mathbf{z})$  depends on the external information included into the model. We assume that  $\boldsymbol{\mu}_{\text{ex}} = E[\mathbf{u}(\mathbf{z})]$ , so that  $E[\mathbf{h}(\mathbf{z})] = \mathbf{0}$ . If, for example, the expected value of  $y$  is known to be  $E(y) = 100$ , then  $u(z) = y$ ,  $\boldsymbol{\mu}_{\text{ex}} = 100$  and  $h(z) = y - 100$ . The corresponding sample moment condition is  $\mathbf{0} = \frac{1}{n} \sum_{i=1}^n (y_i - 100)$  (Imbens & Lancaster, 1994).

To simplify the presentation, we define the combined moment function vector in general regression models as  $\mathbf{g}(\mathbf{z}, \boldsymbol{\theta}) = [\mathbf{m}(\mathbf{z}, \boldsymbol{\theta})^T, \mathbf{h}(\mathbf{z})^T]^T$  in what follows and assume that  $E[\frac{1}{n} \sum_{i=1}^n \mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta}_0)] = \mathbf{0}$  holds. Note that the number of moment conditions exceeds the number of parameters to be estimated, i.e. the externally informed model is overidentified. This means that there will in general be no estimator  $\hat{\boldsymbol{\theta}}$  that solves the corresponding sample moment conditions  $\frac{1}{n} \sum_{i=1}^n \mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta}) = \mathbf{0}$ . To deal with the overidentification problem, we will use the GMM approach (Hansen, 1982), that finds an estimator as “close” as possible to a solution of the sample moment conditions. This is done by maximizing a quadratic form defined by a chosen symmetric, positive definite weighting matrix  $\mathbf{W}$  in the moment functions of the sample. The efficiency of the estimator is affected by  $\mathbf{W}$ , and this can be chosen to maximize the asymptotic efficiency of the estimator in the class of all GMM-estimators based on the same sample moment conditions (Hansen, 1982). This optimal weighting matrix is  $\mathbf{W} = \boldsymbol{\Omega}^{-1}$ , where  $\boldsymbol{\Omega} = E[\mathbf{g}(\mathbf{z}, \boldsymbol{\theta}_0) \mathbf{g}(\mathbf{z}, \boldsymbol{\theta}_0)^T]$ . However, this optimal  $\mathbf{W}$  is unknown in practice and must be estimated by a consistent estimator  $\hat{\mathbf{W}}$ .

**Definition 2.** (Newey & McFadden, 1994, p. 2116) Let  $\mathbf{g}(\mathbf{z}, \boldsymbol{\theta})$  be a vector-valued function with values in  $\mathbb{R}^K$ , that meets the moment conditions  $E[\mathbf{g}(\mathbf{z}, \boldsymbol{\theta}_0)] = \mathbf{0}$ . Further let  $\hat{\mathbf{W}} \in \mathbb{R}^{K \times K}$  be a positive-semidefinite, possibly random matrix, such that  $(\mathbf{r}^T \hat{\mathbf{W}} \mathbf{r})^{1/2}$  is a measure of distance from  $\mathbf{r}$  to  $\mathbf{0}$  for all  $\mathbf{r} \in \mathbb{R}^K$ . Then, the **GMM-estimator**  $\hat{\boldsymbol{\theta}}_{\text{ex}}$  is defined as the  $\boldsymbol{\theta}$ , which maximizes the following function:

$$\hat{Q}_n(\boldsymbol{\theta}) = - \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta}) \right]^T \hat{\mathbf{W}} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{g}(\mathbf{z}_i, \boldsymbol{\theta}) \right].$$

The GMM approach provides consistent and normally distributed estimators under mild regularity conditions (Newey & McFadden, 1994, p. 2148) for a wide range of models, like linear or nonlinear, cross-sectional or longitudinal regression models. Note that we have not assumed that

$\hat{\mathbf{W}}$  is invertible because we will mainly derive asymptotic expressions based on  $\mathbf{W}$  for which the invertibility of  $\hat{\mathbf{W}}$  is not necessary. However, when deriving estimators, additional assumptions about invertibility must be made, which we explain in Sect. 3.2. Let  $\mathbf{G} = E[\nabla_{\theta} \mathbf{g}(\mathbf{z}, \theta_0)]$  be a fixed matrix and  $\mathbf{W}$  the optimal weighting matrix, then  $\text{Var}(\hat{\theta}_{\text{ex}}) = \frac{1}{n}(\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1}$ . This variance expression is not informative with respect to a possible efficiency gain of the GMM-estimator if external information is used. Hence, the following corollary explicitly shows the effect of the external information on the variance of  $\hat{\theta}_{\text{ex}}$ .

**Corollary 1.** *Assume  $\hat{\theta}_M$  is the GMM-estimator based on the model estimating equations alone (ignoring the external moments), and that  $\mathbf{m}(\mathbf{z}, \theta)$  and  $\theta$  have the same dimension. Using the prerequisite  $\mathbf{g}(\mathbf{z}, \theta) = [\mathbf{m}(\mathbf{z}, \theta)^T, \mathbf{h}(\mathbf{z})^T]^T$  it follows, that  $\mathbf{\Omega}$  has the block form*

$$\mathbf{\Omega} = \begin{pmatrix} E[\mathbf{m}(\mathbf{z}, \theta)\mathbf{m}(\mathbf{z}, \theta)^T] & E[\mathbf{m}(\mathbf{z}, \theta)\mathbf{h}(\mathbf{z})^T] \\ E[\mathbf{h}(\mathbf{z})\mathbf{m}(\mathbf{z}, \theta)^T] & E[\mathbf{h}(\mathbf{z})\mathbf{h}(\mathbf{z})^T] \end{pmatrix} = \begin{pmatrix} \mathbf{\Omega}_M & \mathbf{\Omega}_R^T \\ \mathbf{\Omega}_R & \mathbf{\Omega}_h \end{pmatrix}$$

and that

$$\left\{ E[\nabla_{\theta} \mathbf{m}(\mathbf{z}, \theta_0)]^T \right\}^{-1} \mathbf{\Omega}_R^T \mathbf{\Omega}_h^{-1} \mathbf{\Omega}_R \{ E[\nabla_{\theta} \mathbf{m}(\mathbf{z}, \theta_0)] \}^{-1} \quad (1)$$

A proof of Corollary 1 can be found in the supplementary materials online. Note that (1) shows that  $\text{Var}(\hat{\theta}_{\text{ex}})$  is equal to the conditional variance of  $\hat{\theta}_M$  under the external moment conditions, since the asymptotic distribution is normal. This equality shows why there is a reduction in the variance. Let the second term on the right-hand side of (1) be denoted by  $\mathbf{D}$ , then  $\text{Var}(\hat{\theta}_{\text{ex}})$  can be written as  $\text{Var}(\hat{\theta}_{\text{ex}}) = \text{Var}(\hat{\theta}_M) - \mathbf{D}$ . If  $\mathbf{D}$  is nonnegative definite and not equal to  $\mathbf{0}$ , then including external moments leads to an expected efficiency gain in  $\hat{\theta}_{\text{ex}}$  as compared to  $\hat{\theta}_M$ . That  $\mathbf{D} \neq \mathbf{0}$  is nonnegative definite if  $\mathbf{\Omega}_R \neq \mathbf{0}$  is easily seen by noting that  $\mathbf{\Omega}_h^{-1}$  is positive definite and therefore can be written as  $\mathbf{\Omega}_h^{-1} = \mathbf{\Omega}_h^{-1/2} \mathbf{\Omega}_h^{-1/2}$ , where  $\mathbf{\Omega}_h^{-1/2}$  is the positive definite square root of  $\mathbf{\Omega}_h^{-1}$ . Since  $n\mathbf{D}$  can be written as the product of  $\{E[\nabla_{\theta} \mathbf{m}(\mathbf{z}, \theta_0)]^T\}^{-1} \mathbf{\Omega}_R^T \mathbf{\Omega}_h^{-1/2}$  with its transpose,  $\mathbf{D}$  is nonnegative definite. In summary,  $\mathbf{\Omega}_R \neq \mathbf{0}$  is a necessary and sufficient condition for the presence of variance reduction based on Corollary 1. Finally, it should be noted that  $\text{Var}(\hat{\theta}_{\text{ex}})$  can consistently be estimated via the plug-in approach (e.g. Newey & McFadden, 1994, pp. 2171–2173) by replacing all unknown expected values by sample means.

### 3.2. The Externally Informed Multiple Linear Model

In linear models,  $\hat{\theta}_{\text{ex}}$  is denoted as  $\hat{\beta}_{\text{ex}}$ . For analytical simplicity, in this section, we assume the Gauss–Markov assumptions hold, specifically  $E(\epsilon_i) = 0$ ,  $\text{Var}(\epsilon_i) = \sigma^2$ ,  $\text{Cov}(\epsilon_i, \epsilon_j) = 0$  for all  $i \neq j$  with  $i, j = 1, \dots, n$ , and independence of the explanatory variables and the error terms  $\epsilon$ . Furthermore, we assume the errors to be normally distributed in small samples. Analytical solutions to the estimating equations exist under these assumptions:

**Theorem 1.** *Let  $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1, y_1), \dots, \mathbf{h}(\mathbf{x}_n, y_n)]^T$  be the  $(n \times q)$  random matrix containing the externally informed sample moment functions and  $\mathbf{I}_n$  a  $(n \times 1)$ -vector of ones. Further let  $\hat{\mathbf{\Omega}}_h$  and  $\hat{\mathbf{\Omega}}_R$  be consistent estimators of the corresponding matrices in Corollary 1. Then, the (consistent) externally informed OLS estimator is:*

$$\hat{\beta}_{\text{ex}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - (\mathbf{X}^T \mathbf{X})^{-1} \hat{\mathbf{\Omega}}_R^T \hat{\mathbf{\Omega}}_h^{-1} \mathbf{H}^T \mathbf{I}_n$$

and its variance is

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{\beta}}_{\text{ex}}) &= \text{Var}(\hat{\boldsymbol{\beta}}) - \mathbf{D} \\ &= \frac{1}{n}\sigma^2 \left[ E(\mathbf{xx}^T) \right]^{-1} - \frac{1}{n} \left[ E(\mathbf{xx}^T) \right]^{-1} \boldsymbol{\Omega}_R^T \boldsymbol{\Omega}_h^{-1} \boldsymbol{\Omega}_R \left[ E(\mathbf{xx}^T) \right]^{-1},\end{aligned}$$

where  $\sigma^2$  is the variance of the error in the assumed linear model.

The proof of Theorem 1 is given in the supplementary materials online. Note that only the assumption of invertibility of  $\hat{\boldsymbol{\Omega}}_h$  is made, which is weaker than the assumption that  $\hat{\boldsymbol{\Omega}}$  is invertible. From Theorem 1, it is not immediately obvious which of several possibly available functions may lead to a variance reduction. Therefore, let us consider some external moment functions and their possible effects on the variance of  $\hat{\boldsymbol{\beta}}_{\text{ex}}$ . Note that inclusion of external moment functions into the estimating equations may lead to expected efficiency gains only if  $\boldsymbol{\Omega}_R^T = E[\mathbf{x}(y - \mathbf{x}^T \boldsymbol{\beta}_0) \mathbf{h}(\mathbf{x}, y)^T] = E[\mathbf{x} \epsilon \mathbf{h}(\mathbf{x}, y)^T] \neq \mathbf{0}$  holds.

Let the expressions  $\sigma_{x_j}$  and  $\sigma_y$  denote the population standard deviations of  $x_j$  and  $y$ , respectively, whereas  $\sigma_{x_j, y}$  indicates the covariance of  $x_j$  and  $y$ . To denote the covariance vector  $(\sigma_{x_1, x_j}, \dots, \sigma_{x_p, x_j})^T$  of  $\mathbf{x}$  and  $x_j$  the expression  $\boldsymbol{\sigma}_{x, x_j}$  is used, including  $\sigma_{x_j}^2$  at the  $j$ -th position. Finally,  $\rho_{x_j, y}$  is the population correlation of  $x_j$  and  $y$ .

First, consider some function  $\mathbf{f}(\mathbf{x})$  of  $\mathbf{x}$ , i.e.  $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})]_{\text{ex}}$ , where  $E[\mathbf{f}(\mathbf{x})]_{\text{ex}}$  is the known expected value of  $\mathbf{f}(\mathbf{x})$ . If the assumptions underlying the linear model hold, then  $\boldsymbol{\Omega}_R = E[\mathbf{x} \epsilon \mathbf{h}(\mathbf{x})^T] = \mathbf{0}$  because  $\epsilon$  is independent of  $\mathbf{f}(\mathbf{x})$  and  $E(\epsilon) = 0$ . Thus, according to the results of Sect. 3.1, there will be no variance reduction if the external moment function is a function of the explanatory variables only. In the example described in Sect. 2, there will be no efficiency gain if the 15.7%-prevalence of depression is used as external information to estimate the linear regression model.

On the other hand, if the external moment function is a function of  $\epsilon$ , then generally,  $E[\mathbf{x} \epsilon \mathbf{h}(\mathbf{x}, y)^T] \neq \mathbf{0}$ . In the example, assume that the correlation between fluid intelligence and math skills reported in Peng et al. (2019) is taken as external information, in which case  $\mathbf{h}(\mathbf{x}, y) = h(x_2, y) = [y - E(y)][x_2 - E(x_2)]/(\sigma_{x_2} \sigma_y) - \rho(x_2, y)_{\text{ex}}$ , where  $\rho(x_2, y)_{\text{ex}} = 0.41$ . Then  $E[\mathbf{x} \epsilon h(x_2, y)] = [\sigma^2/(\sigma_{x_2} \sigma_y)] \boldsymbol{\sigma}_{x, x_2}$  will not in general be zero, and hence, there will, in general, be efficiency gains with respect to  $\hat{\boldsymbol{\beta}}_{\text{ex}}$ . For more examples, see Table 1 and for the derivation of the results, see the supplementary materials online. It should be noted that if the distribution of the errors is not symmetrical, then  $E(\mathbf{x})E(\epsilon^3)$  has to be added to the entries in column  $\boldsymbol{\Omega}_R^T$  of Table 1 for the cases  $E(y^2)$  and  $\sigma_y^2$ , see the supplementary materials online for further details.

Table 2 presents, in the second column, the absolute variance reduction for the parameters if the external information given in the first column is used to estimate the regression model. The third column in Table 2 shows which entries of the parameter  $\boldsymbol{\beta}$  can be estimated more precisely if the external information is used. The results of Table 2 are derived in the supplementary materials online. Note that  $\boldsymbol{\Omega}_h$  is written as  $\omega_h$  here, as it is single-valued. It holds that  $\omega_h = E[h(\mathbf{x}, y)^2]$ , where  $h(\mathbf{x}, y)$  is of the form given for various moments in Table 1. However, this expression often includes the terms  $E(\epsilon)$  and  $E(\epsilon^3)$ , which are already set to zero in  $\boldsymbol{\Omega}_R^T$  (see supplementary materials online). In order to avoid invalid estimates,  $E(\epsilon)$  and  $E(\epsilon^3)$  should be set to zero in  $\omega_h$ . For example, if the correlation between fluid intelligence and math skills reported in Peng et al. (2019) would be used in the regression from math skills on fluid intelligence and depression, then the variance of the estimator weighting the variable fluid intelligence would be reduced by:

$$\frac{\sigma^4}{n\omega_h\sigma_y^2\sigma_{x_2}^2} = \frac{\sigma^4}{n\text{Var}\{[x_2 - E(x_2)][y - E(y)]\}}.$$

TABLE 1.  
Forms of  $\Omega_R^T$  for various single moments.

Moments	$h(\mathbf{x}, y)$	$\Omega_R^T$
$E(y)$	$y - E(y)_{\text{ex}}$	$\sigma^2 E(\mathbf{x})$
$E(x_j y)$	$x_j y - E(x_j y)_{\text{ex}}$	$\sigma^2 E(x_j \cdot \mathbf{x})$
$E(y^2)$	$y^2 - E(y^2)_{\text{ex}}$	$2\sigma^2 E(\mathbf{x}\mathbf{x}^T)\boldsymbol{\beta}_0$
$\sigma_y^2$	$[y - E(y)]^2 - (\sigma_y^2)_{\text{ex}}$	$2\sigma^2[E(\mathbf{x}\mathbf{x}^T)\boldsymbol{\beta}_0 - E(y)E(\mathbf{x})]$
$\sigma_{x_j, y}$	$[y - E(y)][x_j - E(x_j)] - (\sigma_{x_j, y})_{\text{ex}}$	$\sigma^2 \sigma_{x_j, x_j}$
$\rho_{x_j, y}$	$\frac{[y - E(y)][x_j - E(x_j)]}{\sigma_x \sigma_y} - (\rho_{x_j, y})_{\text{ex}}$	$\frac{\sigma^2}{\sigma_{x_j} \sigma_y} \sigma_{x_j, x_j}$
$\beta_{x_j, y}$	$\frac{[y - E(y)][x_j - E(x_j)]}{\sigma_{x_j}^2} - (\beta_{x_j, y})_{\text{ex}}$	$\frac{\sigma^2}{\sigma_{x_j}^2} \sigma_{x_j, x_j}$

The subscript ex indicates externally determined values. In the last line,  $\beta_{x_j, y}$  represents the expected value of the estimator of the slope from a simple linear regression model, which is identical to the true value of the slope only if  $x_j$  is independent of the other explanatory variables.

TABLE 2.  
Effects of various single moments in terms of variance reduction.

Moments	<b>D</b>	Effect on
$E(y)$	$\frac{\sigma^4}{n\omega_h} \mathbf{e}_1 \mathbf{e}_1^T$	Only $\beta_1$
$E(x_j y)$	$\frac{\sigma^4}{n\omega_h} \mathbf{e}_j \mathbf{e}_j^T$	Only $\beta_j$
$E(y^2)$	$\frac{4\sigma^4}{n\omega_h} \boldsymbol{\beta}_0 \boldsymbol{\beta}_0^T$	All $\beta_j \neq 0$
$\sigma_y^2$	$\frac{4\sigma^4}{n\omega_h} [\boldsymbol{\beta}_0 - E(y)\mathbf{e}_1][\boldsymbol{\beta}_0 - E(y)\mathbf{e}_1]^T$	All $\beta_j \neq 0$ and $\beta_1$
$\sigma_{x_j, y}$	$\frac{\sigma^4}{n\omega_h} \tilde{\mathbf{e}}_j \tilde{\mathbf{e}}_j^T$	$\beta_j$ and $\beta_1$
$\rho_{x_j, y}$	$\frac{\sigma^4}{n\omega_h \sigma_y^2 \sigma_{x_j}^2} \tilde{\mathbf{e}}_j \tilde{\mathbf{e}}_j^T$	$\beta_j$ and $\beta_1$
$\beta_{x_j, y}$	$\frac{\sigma^4}{n\omega_h \sigma_{x_j}^4} \tilde{\mathbf{e}}_j \tilde{\mathbf{e}}_j^T$	$\beta_j$ and $\beta_1$

The expression  $\mathbf{e}_j$  denotes the  $(p \times 1)$ -vector with 1 at the  $j$ -th position and zeros elsewhere. Further we set  $\tilde{\mathbf{e}}_j := -E(x_j) \cdot \mathbf{e}_1 + \mathbf{e}_j$ . In the last line,  $\beta_{x_j, y}$  represents the expected value of the estimator of the slope from a simple linear regression model, which is identical to the true value of the slope only if  $x_j$  is independent of the other explanatory variables.

This means that there will be a variance reduction in all practically relevant cases, where  $\sigma^2 \neq 0$  and  $\text{Var}\{[x_2 - E(x_2)][y - E(y)]\} < \infty$  hold. For a comparison of the effects of the different external moments, the corresponding relative variance reductions may be of interest. These are obtained by dividing the  $j$ -th diagonal element of the absolute reductions in Table 2 by  $\frac{1}{n}\sigma^2 E(\mathbf{x}\mathbf{x}^T)_{(j,j)}^{-1}$ , where  $E(\mathbf{x}\mathbf{x}^T)_{(j,j)}^{-1}$  denotes the element of the inverse of  $E(\mathbf{x}\mathbf{x}^T)$  in the  $j$ -th row and the  $j$ -th column. For the resulting expressions it is clear that  $n$  factors out, as **D** also includes  $\frac{1}{n}$  as the only factor depending on  $n$ , while the rest are fixed values. Hence, the relative efficiency gains do not vanish with increasing  $n$ , but are constant. In our example, the known correlation  $\rho_{x_2, y} = .41$  exerts an expected relative variance reduction of

$$\frac{\sigma^2}{E(\mathbf{x}\mathbf{x}^T)_{(2,2)}^{-1} \text{Var}\{[x_2 - E(x_2)][y - E(y)]\}},$$



which is independent of  $n$  and does not vanish for large  $\sigma^2$ . Including more than one external moment is straightforward. In that case  $\mathbf{\Omega}_h$  includes not only variances but also covariances of the external moments which may lead to additional variance reduction. To illustrate this effect, consider the example from Sect. 2 using the external moments  $\rho(x_2, y)_{\text{ex}} = 0.41$  and  $E(x_2)_{\text{ex}} = 100$ . For the sake of simplicity and without loss of generality we assume  $x_2$  and  $y$  to be centralized. In this example, the external moments  $\rho_{x_2, y}$  and  $E(x_2)$  are included in the externally informed multiple linear model, leading to  $\mathbf{\Omega}_R^T = \left( \mathbf{0} \quad \frac{\sigma^2}{\sigma_{x_2} \sigma_y} \boldsymbol{\sigma}_{x, x_j} \right)$  according to Table 1, and

$$\mathbf{\Omega}_h = \begin{pmatrix} \text{Var}(x_2) & \frac{\text{Cov}(x_2^2, y)}{\sigma_{x_2} \sigma_y} \\ \frac{\text{Cov}(x_2^2, y)}{\sigma_{x_2} \sigma_y} & \frac{\text{Var}(x_2 y)}{\sigma_{x_2}^2 \sigma_y^2} \end{pmatrix}$$

via definition, wherein  $\text{Var}(x_2 y)$  is the scalar variance of  $x_2$  times  $y$ . Using the notation of Table 2, the explicit inversion formula for  $(2 \times 2)$ -matrices implies

$$\begin{aligned} \mathbf{D} &= \frac{1}{n} \left[ E(\mathbf{xx}^T) \right]^{-1} \mathbf{\Omega}_R^T \mathbf{\Omega}_h^{-1} \mathbf{\Omega}_R \left[ E(\mathbf{xx}^T) \right]^{-1} \\ &= \frac{1}{n} \left[ E(\mathbf{xx}^T) \right]^{-1} \frac{\sigma^2}{\sigma_{x_2} \sigma_y} \boldsymbol{\sigma}_{x, x_j} (\mathbf{\Omega}_h^{-1})_{(2,2)} \boldsymbol{\sigma}_{x, x_j}^T \frac{\sigma^2}{\sigma_{x_2} \sigma_y} \left[ E(\mathbf{xx}^T) \right]^{-1} \\ &= \frac{\sigma^4 (\mathbf{\Omega}_h)_{(1,1)}}{n \det(\mathbf{\Omega}_h) \sigma_{x_2}^2 \sigma_y^2} \tilde{\mathbf{e}}_2 \tilde{\mathbf{e}}_2^T = \frac{\sigma^4}{n \left[ \text{Var}(x_2 y) - \frac{\text{Cov}(x_2^2, y)^2}{\sigma_{x_2}^2} \right]} \tilde{\mathbf{e}}_2 \tilde{\mathbf{e}}_2^T, \end{aligned}$$

where  $\det(\mathbf{A})$  denotes the determinant of matrix  $\mathbf{A}$ . Assuming both variances to be finite and positive and invoking the Cauchy–Schwartz inequality, the fraction  $\text{Cov}(x_2^2, y)^2 / \sigma_{x_2}^2$  will not exceed  $\text{Var}(x_2 y)$  and hence  $\mathbf{D}$  will be nonnegative. Further, if  $x_2^2$  and  $y$  have a covariance different from 0, the variance will decrease even further, compared to the reduction due to  $\rho_{x_2, y}$  alone. Hence,  $\beta_1$  and  $\beta_2$  can in general be estimated even more efficiently, if  $E(x_2)$  is used in addition.

### 3.3. Additional Remarks

Using many moments, however, increases the risk of a near-singular  $\mathbf{\Omega}$  matrix, especially if the moments are strongly mutually (linear) dependent. Calculation of the GMM-estimator with additional external moment functions often includes unknown population moments, like  $E(\mathbf{x})$  or  $\sigma_y^2$  (see Table 1), which may be replaced by the corresponding sample moments. However,  $\mathbf{\Omega}_R$  and  $\mathbf{\Omega}_h$  may in addition be functions of unknown  $\sigma^2$  or  $\beta_0$ , as can be seen in Table 1. Hence, the externally informed GMM-estimator is calculated iterating over the following steps until convergence: First estimate the model using the ordinary least squares approach without external moments to get  $\hat{\sigma}^2$  and  $\hat{\boldsymbol{\beta}}$  and then estimate  $\hat{\boldsymbol{\beta}}_{\text{ex}}$  based on the estimates from the former step.

Statistical inference with a GMM-estimator can be based on the Wald test, which simplifies to a t-test if single regression coefficients are tested and its approximative normality can be used to construct confidence intervals (Cameron & Trivedi, 2005). However, in small samples or when dealing with complex models, it is sometimes better to use a bootstrap method (Cameron & Trivedi, 2005; Spiess et al., 2019, p. 177).

As this approach combines data from different sources, one should take into account the issues arising in meta-analyses in general. The *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins et al., 2019) and the PRISMA statement (Page et al., 2021) should be considered

to select proper sources of external information, which are as up-to-date and as close as possible to the same population, method and design of the study one wishes to use the externally informed model in. This is important because a core regularity condition of the GMM is that the expected values of the moment functions are zero, which can be violated, if the external moment and the data were taken from different populations. As a possible approach to deal with this compatibility issue, the GMM framework incorporates the Sargan–Hansen test to test if the overidentification due to the additional moment conditions causes a  $\hat{Q}_n(\hat{\theta}_{\text{ex}})$  significantly larger than 0 (Hansen, 1982; Sargan, 1958). Another option to test for incompatibility especially in linear regression models is the Durbin–Wu–Hausman test (Hausman, 1978), as it compares two estimators of the same parameter. We will take a different approach here, as we will instead relax the assumption of correct external point-values to intervals containing the true value.

#### 4. Robustness due to Interval Probability

External information is only an estimate itself and thus prone to uncertainty. A classical approach to analyze and prevent the issues of misspecification and thus misleading inferences is to use robust models (Huber, 1981). Hence, it is important to use techniques to robustify the estimation of the externally informed model. In this paper, we will adopt an approach based on the theory of imprecise probabilities due to Weichselberger (2001), that is capable of dealing with probabilistic and non-probabilistic uncertainty, not depending on a fully specified stochastic model. The advantage is that instead of distributional assumptions we only need bounds for the true external values. It would be possible to model the uncertainty in the external information within a probabilistic, e.g., a Bayesian, framework. However, this framework would replace uncertainty in the external information by assuming an additional parametric model of its estimation process in form of precise prior distributions. Moreover, it is not straightforward to represent only certain distributional aspects (moments) within a Bayesian approach, e.g., the external information  $100 = E(y) = E(\mathbf{x})^T \beta_0$  presented in Sect. 2.

##### 4.1. Externally Informed Models Based on Interval Information

Assume that  $I_{\text{ex}}$  is an interval containing the true value of an unknown external moment. Hence every value in the interval could be the true one. To illustrate a possible way to construct an  $I_{\text{ex}}$ , we use our earlier example. In our application example, we have a 95% confidence interval of [0.39, 0.44] for the correlation between fluid intelligence and mathematical skills (Peng et al., 2019). This is, of course, an interval that includes the true value only with a positive probability, but not with certainty. However, combining this confidence interval with the results of other studies on this or a similar correlation, and thus possibly widening the interval, the resulting interval serves as a subjective, rough approximation for  $I_{\text{ex}}$ . We illustrate the use of this technique in Sect. 6.

In this section, we discuss another way of constructing  $I_{\text{ex}}$ . Regarding the estimated depression prevalence of 0.157 in Steffen et al. (2020), we know that 87% of the population has been investigated. Thus, we can construct an interval by the technique proposed, e.g., in Manski (1993, 2003), Manski and Pepper (2013), Cassidy and Manski (2019). The two extreme cases that could occur are that no person of the 13% unobserved individuals has a depression and on the other extreme that all of these individuals have a depression. As 87% of 0.157 is 0.137, we get the interval [0.137, 0.267] for the prevalence. The advantage of such intervals is that they completely compensate for the missing values without any further assumptions. Having available an interval for the external information, one can adopt a technique denoted as cautious data completion proposed by Augustin et al. (2014, p. 182) to determine based on  $I_{\text{ex}}$  the sets of possible values for the estimator itself and its variance estimator. In our setting, this amounts to evaluating the

estimator for the externally informed linear model and its variance estimator from Theorem 1 traversing  $I_{\text{ex}}$ . This leads to a set  $\mathcal{B}_{\text{ex}}$  of possible parameter estimates and a set  $\mathcal{V}_{\text{ex}}$  of possible variance estimates. These sets of estimates are compact and connected in the strict mathematical sense, since both estimators are continuous functions on the external interval.

*4.1.1. F-Probability* Interval-based inferences can be justified by adopting the concept of F-probabilities (Augustin, 2002; Weichselberger, 2000).

**Definition 3.** (Augustin 2002) Let  $\Omega$  be a set and  $\mathcal{A}$  be a  $\sigma$ -algebra on  $\Omega$ . Further, let  $\mathcal{K}(\Omega, \mathcal{A})$  be the set of all probability measures on  $(\Omega, \mathcal{A})$ . Then, a set-valued function  $F(\cdot)$  on  $\mathcal{A}$  is called *F-probability* with structure  $\mathcal{M}$ , if

1. there are functions  $L(\cdot), U(\cdot) : \mathcal{A} \rightarrow [0, 1]$  such that for every event  $A \in \mathcal{A}$  it holds that  $L(A) \leq U(A)$  and  $F(\cdot)$  has the form

$$F(\cdot) : \mathcal{A} \rightarrow \{[a, b] \mid a, b \in [0, 1] \text{ and } a \leq b\}$$

$$A \mapsto F(A) := [L(A), U(A)] \text{ for every event } A \in \mathcal{A},$$

2. the set  $\mathcal{M} := \{P(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid L(A) \leq P(A) \leq U(A), \text{ for all } A \in \mathcal{A}\}$  is not empty,
3. for all events  $A \in \mathcal{A}$  it holds that  $\inf_{P(\cdot) \in \mathcal{M}} P(A) = L(A)$  and  $\sup_{P(\cdot) \in \mathcal{M}} P(A) = U(A)$ .

For most applications, it is sufficient to restrict to the case  $\Omega = \mathbb{R}^d$  and let  $\mathcal{A}$  be the corresponding Borel  $\sigma$ -algebra. F-probabilities are best understood as a representation of a “continuous” set of probability measures. For example, consider all normal distributions with a variance of 1 and a mean between  $-0.5$  and  $0.5$ . If we consider all these distributions as possible true distributions for a random variable  $X$  and evaluate an event in terms of its probability, we obtain a set of possible probability values. Consider the event  $A = \{X \leq 0\}$ , its possible probability ranges from 0.3085 (for mean 0.5) to 0.6915 (for mean  $-0.5$ ) and thus  $P(A) \in F(A) := [0.3085, 0.6915]$ . If this procedure is performed for all  $A \in \mathcal{A}$ , the resulting  $F(\cdot)$  is an F-probability. In general, given any nonempty set  $\mathcal{P}$  of probability measures, one can construct the narrowest F-probability containing  $\mathcal{P}$  by defining  $F(A) := [\inf_{P \in \mathcal{P}} P(A), \sup_{P \in \mathcal{P}} P(A)]$  for each event  $A \in \mathcal{A}$ , cf. Remark 2.3. in Augustin (2002). If the intervals  $F(A)$  consist of one element for all  $A$ , the F-probability simply corresponds to a single probability measure. Thus, it is a natural generalization of the conventional notion of probability, using simultaneously a range of probability measures between a lower bound and an upper bound. An important property of F-probabilities for ensuring robustness is that their structure  $\mathcal{M}$  (all the probability measures covered by  $F(\cdot)$ , in the sense of condition 2 in Definition 3) is generally larger than the set  $\mathcal{P}$  (called pre-structure) of probability measures used to construct them, since the structure is closed under convex combinations (Augustin, 2002). For two probability measures  $P$  and  $Q$ , this follows by the basic inequality that for all  $0 \leq \epsilon \leq 1$  and  $A \in \mathcal{A}$  it holds that

$$\min(P(A), Q(A)) \leq \epsilon P(A) + (1 - \epsilon)Q(A) \leq \max(P(A), Q(A)).$$

For example, convex combinations of normal distributions are not themselves normally distributed and include skewed and bimodal distributions. This illustrates that robustness with respect to distributional assumptions increases compared to using normal distributions alone. Unlike other concepts that reflect uncertainty about probability measures, such as triangular numbers (fuzzy numbers), there is no preference for one distribution over another caused by weighting functions or

possibility distributions. This agnosticism regarding the true distribution also covers deterministic ambiguity to some extent. For instance, in our example, a deterministic alteration of  $\mu$  over time, where  $\mu(t) \in [-0.5, 0.5]$  for all  $t$ , like  $\mu(t) = 0.5 \sin(t)$ , would still be covered by the F-probability at any time  $t$  because the F-probability covers the range of  $\mu(t)$ . In applied research, the exact form of deterministic variation of  $\mu$  is typically unknown, but if its bounds are known to lie within an interval, the F-probability based on this interval would account for it. Of course, these advantages come at the cost of greater conservatism than using a single probability distribution.

In our framework, the assumption of knowing the true moment value can be relaxed to assuming that an interval is known containing the unknown true moment value. As the GMM-estimator is asymptotically normally distributed for the true value of the external moment, we asymptotically get a pre-structure consisting of all normal distributions for estimator  $\hat{\beta}_{\text{ex}}$  with expected value inside  $\mathcal{B}_{\text{ex}}$  and with variance inside  $\mathcal{V}_{\text{ex}}$ . This pre-structure is guaranteed to contain the normal distribution based on GMM asymptotics, since the true external moment value is assumed to be in  $I_{\text{ex}}$ . Therefore, for each event, the probability assigned to an event by this true normal distribution will lie between the lower and upper bounds assigned to that event by  $F(\cdot)$ , possibly leading to more conservative but valid statistical inference. Based on this pre-structure, we get an F-probability. Statistical inference based on F-probabilities is done by treating the probability intervals as a whole, e.g., by interval arithmetic. We demonstrate this principle by constructing an equivalent to confidence intervals in the context of F-probabilities in the next section.

*4.1.2. Confidence Intervals for the Externally Informed Model Under F-Probabilities* The construction of confidence intervals (point-CIs) is in general not possible in the framework of F-probabilities, because instead of a single probability value lower and upper bounds are assigned to an event. One possibility, however, is to use the union of all possible point-CIs traversing  $I_{\text{ex}}$ . The idea to calculate unions of intervals already has been investigated for Bayesian highest density intervals in an imprecise probability setting by Walter and Augustin (2009). Let  $\hat{\theta}_{e,j}$  be the  $j$ -th entry of the externally informed GMM-estimator  $\hat{\theta}_{\text{ex}}$  using external value  $e$ , we define the  $(1 - \alpha) \cdot 100\%$  confidence union for  $\theta_j$  to be

$$\bigcup \text{CI}_{1-\alpha} := \left[ \inf_{e \in I_{\text{ex}}} [\hat{\theta}_{e,j} - t_{1-\frac{\alpha}{2}, n-p} \sqrt{\widehat{\text{Var}}(\hat{\theta}_{e,j})}], \sup_{e \in I_{\text{ex}}} [\hat{\theta}_{e,j} + t_{1-\frac{\alpha}{2}, n-p} \sqrt{\widehat{\text{Var}}(\hat{\theta}_{e,j})}] \right]$$

Because the true external moment value is in  $I_{\text{ex}}$ , the borders of the point-CI constructed via the true moment value lie between the infimum and the supremum of the lower and upper borders, respectively, of all point-CIs on  $I_{\text{ex}}$ . Therefore,  $\bigcup \text{CI}_{1-\alpha}$  covers the point-CI constructed via the true moment value. The asymptotic normal distribution of  $\hat{\beta}_{\text{ex}}$  at the true value of the external moment implied by the asymptotic properties of GMM-estimators described in Sect. 2 ensures that the confidence union covers the true parameter asymptotically with probability at least  $1 - \alpha$ . An approximation of the confidence union can be calculated using grid search traversing  $I_{\text{ex}}$ . If the point-CIs used to construct  $\bigcup \text{CI}_{1-\alpha}$  differ, then the resulting interval is wider than every of these point-CIs. This demonstrates that the positive effect of the variance reduction (a shorter CI) can be reversed by the length of  $I_{\text{ex}}$ . The reason is that a broader  $I_{\text{ex}}$  increases the set over which infimum and supremum are taken, possibly expanding  $\bigcup \text{CI}_{1-\alpha}$ . However, we will show in a simulation study in Sect. 5 that in some cases it is possible to get a  $\bigcup \text{CI}_{1-\alpha}$  shorter than the  $(1 - \alpha)$  confidence interval based on the OLS multiple linear regression. Hence, the variance reduction can compensate the broadening of  $\bigcup \text{CI}_{1-\alpha}$  introduced by  $I_{\text{ex}}$ . Finally, using  $\bigcup \text{CI}_{1-\alpha}$  strengthens the robustness through the F-probability, on which  $\bigcup \text{CI}_{1-\alpha}$  is based, as it also includes, e.g., bimodal and skewed distributions.

## 5. A Simulation Study

### 5.1. Settings

To test the externally informed GMM approach for multiple linear models in small samples, we conducted two simulation studies. The first setting illustrates possible variance reduction if correctly specified external moments are used and shows that the usage of small external moment intervals can lead to confidence unions that may even be shorter than the OLS confidence interval. In the second setting, we focus on misspecified external information and non-normal errors. In this case, it is interesting to see, if inferences are still valid and whether the effects of the variance reduction illustrated in the first setting still occur. The simulation script was written and executed in R version 4.2.1 (R Core Team, 2022), the script can be found in the supplementary materials online. The function `interval_gmm()` implements the calculation of intervals of estimators and of their standard deviation, as well as confidence interval unions. In both settings, we used an intercept ( $x_1 = 1$ ), a normally distributed variable  $x_2 \sim N(2, 4)$  and a binary variable distributed according to Bernoulli distribution  $x_3 \sim \text{Bernoulli}(0.4)$  as explanatory variables. The response variable was generated according to  $y = x_1 + 0.5x_2 + 2x_3 + \epsilon$ , where  $\epsilon \sim N(0, 9)$  in the first setting. In the second setting, the errors were generated by affine transformation of a  $\chi_1^2$ -distributed random sample, so that its mean is 0 and its variance is 9. The settings were selected, so that all required moments can easily be calculated, which is done before the simulations. The ratio of explained variance to total variance was  $1 - 9/\text{Var}(y) = 1 - 9/10.96 = 0.178$  a value which is similar to often reported values in psychological research. This amounts to a relatively high error variance, a factor for possibly large variance reduction for some external moments (see Sect. 3).

Different moments have different scales, so a similar interval width of  $I_{\text{ex}}$  does not imply similar “sharpness” of the external information across scales. To create intervals for the external information, that are comparable across the different scales of the external moments, we have used external intervals where the ratio of half their width to their center is the same for all external moments in each setting. It should be noted that this technique is different from the design techniques discussed in Sect. 4.1. The reason for this difference is that the simulation study aims to compare the different moments in terms of their effectiveness and statistical validity in a context where the  $I_{\text{ex}}$  are comparable in magnitude and contain the true value. To motivate this, one could compare the given ratio to the coefficient of variation. For the standard IQ-scale, the coefficient of variation is  $15/100 = 0.15$ . For the first setting, we arbitrarily chose a ratio of 0.1 to represent somewhat more precise external information than one standard deviation in the IQ-scale around the center. For the second setting, we have chosen a ratio of 0.3 to represent a radius of two standard deviations in the IQ-scale and thus an approximate confidence interval width that takes the IQ-scale as a basis. In the first setting, we created intervals that were symmetrical around the true external value. Hence, if the true external value was  $e$ , then the interval was  $I_{\text{ex}} = [0.9e, 1.1e]$ . In the second setting, we first multiplied all true external moment values by 1.3. Since none of these true external values were equal to zero, this resulted in misspecified point values. These misspecified values were used as external point values during the simulation to test the sensitivity of the externally informed model based on point information. The constant 1.3 was again chosen arbitrarily and leads to a relative bias of 30%. Then, as in the first setting we generated a symmetric interval around the misspecified value. If  $e$  again denotes the true external value,  $0.7 \cdot 1.3e = 0.91e$  was the lower limit and  $1.3 \cdot 1.3e = 1.69e$  the upper limit of  $I_{\text{ex}}$ , i.e.  $I_{\text{ex}} = [0.91e, 1.69e]$ , which contains the true value  $e$ . As for sensitivity, tests with center width ratio and misspecification values similar to 0.1, 0.3 and 1.3 gave similar results.

Sample sizes  $n$  chosen are 15, 30, 50, 100. The moments used are those listed in Table 2 for both,  $x_2$  and  $x_3$ . Given the results in Sect. 3, the expected relative variance reductions were calculated to check if these settings are capable of providing enough variance reduction. For every

moment condition in each setting we run 500 simulations. Only single moment conditions were used.

In a first step, all explanatory variables were generated and  $y$  was calculated as described above. In the second step,  $\hat{\beta}_{\text{ex}}$  and  $\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$  were calculated according to the following two-step GMM algorithm:

1. Calculate  $\hat{\beta}$  and  $\hat{\sigma}^2$  via the classical OLS method
2. Determine  $\hat{\Omega}_R$ ,  $\hat{\omega}_h$  and  $\hat{\beta}_{\text{ex}}$  based on  $\hat{\beta}$  and  $\hat{\sigma}^2$
3. Recalculate  $\hat{\sigma}^2$ ,  $\hat{\Omega}_R$  and  $\hat{\omega}_h$  based on  $\hat{\beta}_{\text{ex}}$
4. Update  $\hat{\beta}_{\text{ex}}$  and calculate  $\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$

Then, 95% confidence intervals were calculated based on  $\hat{\beta}_{\text{ex}}$  and its estimated variance, using a t-distribution with  $n - 3$  degrees of freedom. Let  $\hat{\beta}_{\text{ex}}$  be one element of  $\hat{\beta}_{\text{ex}}$ , then its 95% confidence interval is

$$\text{CI}_{0.95} = \left[ \hat{\beta}_{\text{ex}} - t_{n-3,0.975} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})}, \hat{\beta}_{\text{ex}} + t_{n-3,0.975} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})} \right].$$

To calculate  $\bigcup \text{CI}_{0.95}$  a grid search algorithm was adopted. First we determined 101 equidistant points in the given  $I_{\text{ex}}$  (including the bounds of the interval). The number 101 was chosen after some preliminary tests of the algorithm as a compromise between precision and computing time. Then we traversed these grid points calculating  $\hat{\beta}_{\text{ex}}$  and  $\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$  using the two step procedure from above at each point. Comparing the bounds of the CIs sequentially, the minimal lower and maximal upper CI bounds on the grid points were determined and served as approximation for the bounds of  $\bigcup \text{CI}_{0.95}$ .

## 5.2. Results

As criteria to evaluate the statistical inferences, we calculated the mean  $\bar{\hat{\beta}}_{\text{ex}}$  of the estimates  $\hat{\beta}_{\text{ex}}$  and their variances  $\text{Var}(\hat{\beta}_{\text{ex}})$  over 500 simulations. The latter will be compared to the corresponding means of estimated variances,  $\overline{\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})}$ . To evaluate possible variance reduction for  $\beta_j$ , the mean ratio of variance reduction to OLS-variance,  $\hat{\Delta}_j := \frac{[\widehat{\text{Var}}(\hat{\beta}_{OLS}) - \widehat{\text{Var}}(\hat{\beta}_{\text{ex}})]_{(j,j)}}{[\widehat{\text{Var}}(\hat{\beta}_{OLS})]_{(j,j)}}$ , will be considered. In addition, the actual coverage is calculated over simulations. For given  $\alpha = 0.05$  and 500 simulations, the actual coverage should be between 0.93 and 0.97 for the point-valued moments (Spiess, 1998) and equal to or greater than 0.93 for the external moment intervals, as the confidence union is used to calculate the coverage in this case. Finally,  $|\text{CI}| := \overline{\text{CI}}_{0.95} - \underline{\text{CI}}_{0.95}$  and  $|\bigcup \text{CI}| := \overline{\bigcup \text{CI}}_{0.95} - \underline{\bigcup \text{CI}}_{0.95}$  were computed. They can be compared to the OLS-CI-length to evaluate the possible precision gains or losses.

**5.2.1. Results for the Correctly Specified Setting** The detailed results for sample size  $n = 15$  are presented in Table 3, while the results for the other sample sizes are given in Tables 7 to 9 in the supplementary materials. Consistent with the theory in Sect. 3, the use of the moment  $E(x_2)$  had no effect on the variances, neither for the correctly specified nor for the misspecified setting, and estimation results were equal to OLS estimation results. The corresponding results are presented for comparison. For all moments except  $E(y^2)$  and  $\sigma_y^2$  both the coverages for the point valued moments as well as the coverages for the external intervals exceeded 0.93. The coverages for  $\sigma_y^2$  were in the valid range only for  $n = 100$ , while for  $E(y^2)$  they were in the valid range already for  $n = 50$ . The undercoverage for sample sizes below  $n = 100$  can be explained by the skewness

TABLE 3.  
Results of the simulations with correctly specified external moments for sample size  $n = 15$ .

Moments	$\beta_j$	$\tilde{\beta}_{\text{ex}}$	$\text{Var}(\hat{\beta}_{\text{ex}})$	$\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$	$\hat{\Delta}_j$	Cov	$\text{Cov}_I$	ICI	$\cup$ CI
$E(x_2)$ = OLS	$\beta_1$	1.045	1.929	1.968	0	0.944	0.944	5.866	5.866
	$\beta_2$	0.499	0.205	0.211	0	0.950	0.950	1.909	1.909
	$\beta_3$	1.955	3.422	2.976	0	0.940	0.940	7.292	7.292
$E(y)$	$\beta_1$	1.048	1.453	1.593	0.217	0.942	0.964	5.215	5.639
$E(x_2y)$	$\beta_2$	0.560	0.174	0.177	0.169	0.950	0.956	1.740	1.832
$\sigma_{x_2,y}$	$\beta_1$	0.802	1.478	1.456	0.239	0.954	0.954	5.054	5.222
	$\beta_2$	0.625	0.096	0.080	0.616	0.960	0.978	1.153	1.237
$\rho_{x_2,y}$	$\beta_1$	0.928	1.405	1.418	0.258	0.944	0.948	4.984	5.141
	$\beta_2$	0.562	0.069	0.074	0.639	0.976	0.984	1.110	1.188
$\beta_{x_2,y}$	$\beta_1$	0.967	1.263	1.414	0.255	0.960	0.964	4.982	5.131
	$\beta_2$	0.543	0.054	0.072	0.633	0.986	0.986	1.107	1.180
$E(x_3y)$	$\beta_3$	2.150	2.745	2.380	0.192	0.936	0.946	6.525	6.837
$\sigma_{x_3,y}$	$\beta_1$	0.959	1.552	1.704	0.141	0.964	0.966	5.433	5.561
	$\beta_3$	2.202	0.783	0.949	0.689	0.976	0.976	3.952	4.273
$\rho_{x_3,y}$	$\beta_1$	0.977	1.659	1.694	0.148	0.960	0.960	5.415	5.541
	$\beta_3$	2.136	0.888	0.919	0.699	0.964	0.970	3.891	4.208
$\beta_{x_3,y}$	$\beta_1$	0.942	1.545	1.700	0.143	0.964	0.970	5.426	5.559
	$\beta_3$	2.182	0.692	0.908	0.698	0.976	0.980	3.888	4.213
$E(y^2)$	$\beta_1$	1.072	1.965	1.935	0.018	0.914	0.936	5.807	5.950
	$\beta_2$	0.507	0.207	0.204	0.031	0.942	0.946	1.875	1.929
	$\beta_3$	1.995	3.414	2.871	0.030	0.916	0.930	7.164	7.367
$\sigma_y^2$	$\beta_1$	0.905	1.951	1.708	0.142	0.920	0.928	5.434	5.667
	$\beta_2$	0.540	0.228	0.184	0.148	0.896	0.910	1.766	1.839
	$\beta_3$	2.092	3.646	2.584	0.151	0.858	0.878	6.725	7.005

The expressions  $\tilde{\beta}_{\text{ex}}$ ,  $\text{Var}(\hat{\beta}_{\text{ex}})$ ,  $\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$ ,  $\hat{\Delta}_j$ , ICI and  $\cup$  CI are defined in the beginning of Sect. 5.2. The results for the moment  $E(x_2)$  are equivalent to the OLS results. Cov is the coverage for the external point value and  $\text{Cov}_I$  symbolizes the coverage for the confidence interval union based on the external interval. Only the affected coefficients are reported per moment. The true values are  $\beta_1 = 1$ ,  $\beta_2 = 0.5$  and  $\beta_3 = 2$ .

of the distributions of their sample moment functions in small samples caused by the quadratic terms  $y^2$ , leading to higher sample size required for the asymptotic results to be applicable. Using confidence unions only reduced these required sample sizes to  $n = 50$  and  $n = 30$ , respectively, showing that high skewness is also problematic for  $\cup$  CI-based coverage in small samples.

For  $\beta_{x_j,y}$  with  $j = 2, 3$ , the coverage for  $\beta_j$  was in many cases above 0.97 (up to 0.994) for all  $n$ . This was also the case, though not as pronounced, when the external information about the covariance between  $x_j$  and  $y$  was used. The reason for this is that the variances were mostly overestimated in these cases, as can be seen in Tables 3 and 4 as well as in Tables 7 to 12 in the supplementary materials by the fact that  $\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$  was larger than  $\text{Var}(\hat{\beta}_{\text{ex}})$  for the respective  $\beta_j$ . Although variances are overestimated, the true and estimated variances nevertheless tend to be smaller than the variance of the OLS-estimators. Thus, inferences still tend to be more precise, suggesting a possible relationship with superefficiency (Bahadur, 1964).

As shown in Sect. 3, the relative variance reduction for each estimator of  $\beta_j$ , reported in column  $\hat{\Delta}_j$  of Table 3 as well as Tables 7 to 9 in the supplementary materials, did not change significantly under the various conditions over the different sample sizes realized. The smallest relative variance reduction per  $\beta_j$  was attained by using the external information  $E(y^2)$ , ranging from 0.018 to 0.059, followed by  $\sigma_y^2$  with a maximal relative variance reduction of 0.180. The

largest relative variance reduction was attained by using the covariance, the correlation and  $\beta_{x_j,y}$  regarding  $\beta_j$ , ranging from 0.633 to 0.734 for  $j = 2$  as well as from 0.698 to 0.857 for  $j = 3$ . For all other moments the values varied between 0.169 and 0.294, see Table 3 and Tables 7 to 9 in the supplementary materials.

These variance reductions translated for all moments directly into a reduction of the length of the confidence interval for the external point value. For the external interval, the length of the union of the confidence intervals is always greater than the one derived from a single external point. These differences increase with larger samples, as the variance estimator decreases with increasing sample sizes, while it can be seen from the formulas in Theorem 1 that the interval for  $\hat{\beta}_{\text{ex}}$  is only affected by the difference between the estimators and the true values of  $\Omega_R$  and  $\Omega_h$ , not directly by  $n$ . Finally, with regard to  $|\bigcup \text{CI}|$  compared to  $|\text{CI}|$  the results imply that at the sample sizes 15 and 30, using any moment except  $E(y^2)$  resulted in a shorter confidence interval union than the OLS confidence interval. For  $n = 50$ , this was the case for all moments except  $E(y^2)$ ,  $E(y)$  and  $\sigma_y^2$ . Finally, for  $n = 100$ , only the moments  $\sigma_{x_2,y}$ ,  $\rho_{x_2,y}$ ,  $\beta_{x_2,y}$ ,  $\sigma_{x_3,y}$ ,  $\rho_{x_3,y}$  and  $\beta_{x_3,y}$  resulted in shorter confidence unions than point-CIs. This can be explained by the constancy of  $I_{\text{ex}}$  while  $n$  increases. There is always an interval inside  $\bigcup \text{CI}$  which does not vanish for large  $n$ , while  $|\text{CI}|$  converges to 0.

*5.2.2. Results for the Misspecified Setting* The detailed results for sample size  $n = 50$  are presented in Table 4, while the results for the other sample sizes are given in Tables 10 to 12 in the supplementary materials. The coverage rates using the point-valued moments illustrate the expected sensitivity of the models due to misspecification. Even at  $n = 15$  more than half of the coverage rates are below 0.93, although in most cases they are still above 0.9. The severeness increases with increasing  $n$ : For  $n = 30$  only five coverage rates are in the acceptable range of at least 0.93. As seen in Table 4 for  $n = 50$  the coverage is as low as 0.586 in the worst case for  $\beta_3$  if  $\sigma_{x_3,y}$  is used. Finally, for  $n = 100$  all coverage rates are invalid, see Table 12 in the supplementary materials. Except for the moments  $E(y^2)$  and  $\sigma_y^2$ , this is corrected by the union of confidence intervals based on the external interval, since all coverage rates in these cases are above 0.93, except the one for  $\beta_1$  using  $\sigma_{x_2,y}$  while  $n = 15$ . Like in the correctly specified setting, there are considerably larger coverage rates for the moments  $\beta_{x_j,y}$  and lower coverage rates for  $\sigma_y^2$  or  $E(y^2)$  even in the cases  $n = 30$  and  $n = 15$ . The explanations for these over- and undercoverages are the same as for the correctly specified case in Sect. 5.2.1. However, only the use of covariance, correlation or  $\beta$  for  $x_j$  and  $y$  for  $j = 2, 3$  resulted in narrower confidence unions as compared to OLS confidence intervals, not the use of other moments. Regarding  $\beta_j$  for  $j = 2, 3$  this is the case for every  $n$ , regarding  $\beta_1$  this is only the case for  $n = 15$ . We conclude that the use of external intervals for covariances, correlations or  $\beta$  not only corrects low coverage rates due to misspecified point values for external moments, but can also lead to narrower (unions of) confidence intervals.

## 6. Application

To illustrate the possible benefits of using external information in a linear model, we reanalyze a dataset of Pluck and Ruales-Chieruzzi (2021), who investigated the estimation of premorbid intelligence based on lexical reading tasks in Ecuador. We will focus on their Study 2. Since the purpose of this analysis is to illustrate the proposed use of external information, we will only shortly sketch the theoretical background of the study. For a more detailed description, see Pluck and Ruales-Chieruzzi (2021). The dataset was downloaded from PsychArchives (Pluck, 2020a).

To quantify the cognitive impairment of patients, it is necessary to have an accurate baseline estimate observed in the premorbid state (Pluck & Ruales-Chieruzzi, 2021). As psychometric



TABLE 4.  
Results of the simulations with misspecified external moments for sample size  $n = 50$ .

Moments	$\beta_j$	$\bar{\beta}_{\text{ex}}$	$\text{Var}(\hat{\beta}_{\text{ex}})$	$\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$	Cov	$\text{Cov}_I$	ICII	$ \cup \text{CI}$
$E(x_2)$ = OLS	$\beta_1$	1.012	0.440	0.509	0.936	0.936	2.778	2.778
	$\beta_2$	0.488	0.041	0.049	0.960	0.960	0.865	0.865
	$\beta_3$	2.042	0.751	0.803	0.954	0.954	3.494	3.494
$E(y)$	$\beta_1$	1.636	0.412	0.376	0.782	0.992	2.397	4.032
$E(x_2y)$	$\beta_2$	0.615	0.031	0.040	0.924	0.984	0.779	1.062
$\sigma_{x_2,y}$	$\beta_1$	0.733	0.307	0.380	0.912	0.954	2.409	2.922
	$\beta_2$	0.627	0.016	0.016	0.864	0.984	0.496	0.747
$\rho_{x_2,y}$	$\beta_1$	0.756	0.243	0.380	0.952	0.968	2.401	2.914
	$\beta_2$	0.617	0.019	0.016	0.880	0.996	0.494	0.744
$\beta_{x_2,y}$	$\beta_1$	0.768	0.271	0.376	0.926	0.960	2.397	2.901
	$\beta_2$	0.610	0.008	0.015	0.948	0.998	0.488	0.735
$E(x_3y)$	$\beta_3$	2.505	0.486	0.622	0.934	0.982	3.095	4.165
$\sigma_{x_3,y}$	$\beta_1$	0.827	0.341	0.413	0.930	0.966	2.502	3.026
	$\beta_3$	2.549	0.116	0.139	0.586	1.000	1.445	2.751
$\rho_{x_3,y}$	$\beta_1$	0.830	0.291	0.413	0.956	0.978	2.498	3.025
	$\beta_3$	2.525	0.341	0.138	0.642	0.978	1.434	2.739
$\beta_{x_3,y}$	$\beta_1$	0.822	0.346	0.412	0.922	0.966	2.499	3.030
	$\beta_3$	2.537	0.092	0.135	0.618	1.000	1.428	2.740
$E(y^2)$	$\beta_1$	1.136	0.512	0.513	0.914	0.944	2.802	3.151
	$\beta_2$	0.558	0.053	0.049	0.894	0.948	0.863	1.011
	$\beta_3$	2.318	0.846	0.790	0.894	0.958	3.477	4.093
$\sigma_y^2$	$\beta_1$	0.512	0.894	0.433	0.698	0.828	2.543	3.361
	$\beta_2$	0.626	0.067	0.044	0.754	0.896	0.806	1.029
	$\beta_3$	2.597	1.054	0.697	0.750	0.896	3.234	4.167

The expressions  $\bar{\beta}_{\text{ex}}$ ,  $\text{Var}(\hat{\beta}_{\text{ex}})$ ,  $\widehat{\text{Var}}(\hat{\beta}_{\text{ex}})$ , ICII and  $|\cup \text{CI}$  are defined in the beginning of Sect. 5.2. The results for the moment  $E(x_2)$  are equivalent to the OLS results. Cov is the coverage for the external point value and  $\text{Cov}_I$  symbolizes the coverage for the confidence interval union based on the external interval. Only the affected coefficients are reported per moment. The true values are  $\beta_1 = 1$ ,  $\beta_2 = 0.5$  and  $\beta_3 = 2$ .

intelligence tests can be too long or cumbersome for elderly people with emerging cognitive impairments, it is important to have short, yet reliable tests for general intelligence. It is argued in Pluck and Ruales-Chieruzzi (2021) that vocabulary has a high positive correlation with general intelligence, hence using short lexical tests could be helpful to estimate general intelligence. Following Cattell's classical theory, general intelligence can be divided into fluid and crystallized intelligence (Cattell, 1963). In this context, the variance reduction property of the externally informed linear model could provide an asymptotically unbiased estimate with higher precision than the estimates in Pluck and Ruales-Chieruzzi (2021), because external information about the correlation of general, fluid or crystallized intelligence and lexical tests is available. Although the different factors of intelligence are not identical, combining external information about them leads to a broader and thus more reliable external interval than using information about general intelligence alone, as the correlation between lexical tasks and fluid or crystallized intelligence may be lower or higher than for general intelligence.

In their Study 2 Pluck and Ruales-Chieruzzi (2021) used a Spanish, validated seven-subtest version of the Wechsler Adult Intelligence Scale in the 4th edition (WAIS-IV) (Meyers et al., 2013) to measure general intelligence, as well as three lexical tests, the Word Accentuation Test (WAT) in Spanish (Del Ser et al., 1997), the Stem Completion Implicit Reading Test (SCIRT) (Pluck,

2018) and the Spanish Lexical Decision Task (SpanLex) (Pluck, 2020b). The sample consists of 106 premorbid participants without neurological illness. As one participant has not completed the WAT, this person was excluded from the analysis regarding the WAT score. Simple linear regression models with the WAIS-IV as dependent and the lexical tests as independent variable, respectively, were conducted to determine the percentage of explained variance and to test the predictability of general intelligence through every single test. Therefore, the sample was randomly divided into two halves; hence, the net sample size for the linear regression models was 53 as the other half was used to test the prediction based on the regression models. We compared the widths of the 95% confidence intervals for the parameters of these regression models to the widths of the 95% confidence unions resulting from externally informed versions of the linear models. Because the OLS estimation does not account for heteroscedastic errors, which are common in practice, the standard errors are often too small (White, 1980). To correct for heteroscedasticity, we have computed robust standard errors of type HC3 using the package *sandwich* (Zeileis, 2004; Zeileis et al., 2020). Since the dependent variable is the WAIS-IV, an intelligence test with calibration sample, we calculated  $E(y) = 100$ . In the simulation study using the external information about  $\rho$  was found to lead to high variance reduction. Hence, by reviewing the literature, we identified the upper bound for the correlation between general intelligence and lexical tasks to be .85. This value was reported as correlation between WAT and the vocabulary scale of the Wechsler Adult Intelligence scale in Burin et al. (2000). A lower bound for the correlation between general intelligence and lexical tasks was found using the meta-analysis of Peng et al. (2019) or the study of Pluck (2018). Pluck (2018) argued based on a couple of studies that the correlation of general intelligence and lexical skills is typically higher than .70. In the meta-analysis of Peng et al. (2019), the reported 95% confidence interval for the correlation of fluid intelligence and reading is [0.36, 0.39]. To compare the results, both sources were used separately, leading to the lower bounds 0.4 and 0.7, where 0.4 is very conservative as it is derived from a correlation including a different variable (fluid intelligence). Together this amounts to the intervals [0.4, 0.85] and [0.7, 0.85], which are adopted for each of the three lexical tests. The confidence unions were calculated in the same way as in the simulations using grid search, but with 10001 grid points instead of 101 and  $\hat{\Omega}_h = \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{z})\mathbf{h}(\mathbf{z})^T$ . The details of the analysis can be found in the R script in the online supplements to this article. The results for the interval [0.7, 0.85] are shown in Table 5 and the results for the interval [0.4, 0.85] are in Table 13 in the supplementary materials. First, the results of Pluck and Ruales-Chieruzzi (2021) were recalculated, showing no differences from the results reported in their Study 2. In addition, the corresponding OLS confidence intervals for the parameters were calculated based on the HC3 estimator (see column five of Table 5). Then, estimator and standard error intervals, as well as the unions of confidence intervals, were calculated for the externally informed model. For both [0.4, 0.85] and [0.7, 0.85], the maxima of all standard error intervals were below the respective standard errors calculated for the OLS models of Pluck and Ruales-Chieruzzi (2021). This clearly shows the variance reduction property of the externally informed model and was most pronounced for the SpanLex. For [0.4, 0.85] all estimation intervals included the OLS estimates and all confidence unions were larger than the corresponding OLS confidence intervals, indicating that [0.4, 0.85] is very conservative. For [0.7, 0.85], the estimation interval  $[\hat{\beta}_j, \overline{\hat{\beta}}_j]$  included the OLS estimator only for the slope and intercept of the regression on SCIRT and the one based on the WAT. In this case, however, all confidence unions overlapped with the OLS-based confidence intervals. Using [0.7, 0.85], for every lexical test, the widths of the confidence unions from the externally informed model were smaller than the confidence intervals from the simple linear regression models, for both slopes and intercepts, except for the intercept of WAT. Since the prediction interval is calculated based on the distribution of parameter estimators, this would lead to shorter prediction intervals for a participant's general intelligence based on the externally informed model. In addition, the confidence union approach

TABLE 5.  
Results using  $\rho_{x,y} \in [0.7, 0.85]$  and  $E(y) = 100$ .

$j$	Test	Pluck and Ruales-Chieruzzi			Externally informed estimates		
		$\hat{\beta}_j$	$s(\hat{\beta}_j)$	$CI_{0.95}$	$[\underline{\hat{\beta}}_j, \overline{\hat{\beta}}_j]$	$[s(\underline{\hat{\beta}}_j), s(\overline{\hat{\beta}}_j)]$	$\cup CI_{0.95}$
1	SpanLex	54.61	8.864	[37.06, 72.15]	[37.41, 47.15]	[2.373, 2.663]	[32.06, 51.91]
	WAT	62.81	4.701	[53.51, 72.12]	[60.02, 62.89]	[3.587, 3.612]	[52.77, 70.10]
	SCIRT	60.81	4.395	[52.11, 69.51]	[59.01, 61.28]	[3.910, 3.920]	[51.14, 69.13]
2	SpanLex	1.821	0.332	[1.163, 2.480]	[2.068, 2.430]	[0.124, 0.132]	[1.818, 2.696]
	WAT	2.083	0.240	[1.607, 2.559]	[2.041, 2.186]	[0.190, 0.191]	[1.659, 2.568]
	SCIRT	3.292	0.358	[2.583, 4.001]	[3.213, 3.393]	[0.309, 0.310]	[2.592, 4.015]

The third and fourth columns contain the recomputed results of Pluck and Ruales-Chieruzzi (2021) in terms of the OLS regression coefficients  $\hat{\beta}_j$ , where  $\hat{\beta}_1$  is the intercept and  $\hat{\beta}_2$  is the slope and the robust standard errors  $s(\hat{\beta}_j)$  of the coefficients. The (robust) 95% confidence intervals  $CI_{0.95}$  for the parameters were computed in addition. The estimator interval  $[\underline{\hat{\beta}}_j, \overline{\hat{\beta}}_j]$ , the standard error interval  $[s(\underline{\hat{\beta}}_j), s(\overline{\hat{\beta}}_j)]$  and the 95% confidence interval union  $\cup CI_{0.95}$  are shown as results of the estimation of the externally informed model.

is more robust than OLS confidence intervals with respect to deviations from the assumed normal distribution. Taken together, this amounts to possibly more precise yet robust parameter estimation and prediction, if the external information is correct.

## 7. Discussion

In this paper, we show that incorporating external moments into the GMM framework by using intervals instead of point values can lead to more robust analyses, while a possible variance reduction can prevent the confidence unions from being too wide.

The results of the simulation study for point values show that the variance reduction can be considerable, over 70% using external information about covariances, correlations or  $\beta_{x_j,y}$ . However if the external moments deviate from the true values, the inferences will be biased, getting worse with increasing sample size. Instead, the use of external intervals often leads to correct inferences. However, the F-probability couldn't completely correct the undercoverages caused by using the moments  $\sigma_y^2$  as well as  $E(y^2)$ , though it slightly improved them. The reason for these undercoverages is the skewed distribution induced by  $y^2$ , indicating a limitation of the distributional robustness in the presence of large deviations from the normal distribution. As these two moments also showed low variance reduction when used, one should thoughtfully decide on basis of their relative variance reduction if one wants to use them in small samples. However, bootstrap methods, like the bias-corrected accelerated bootstrap (Efron & Tibshirani, 1993), could be used instead to try to correct the undercoverage.

For small sample sizes, the use of covariances, correlations, and  $\beta_{x_j,y}$ ,  $j = 2, 3$ , leads to variance reduction despite the use of external intervals. However, this was mostly the case for certain entries  $\beta_j$  of  $\beta$  in this setting, not for all elements in  $\beta$ . Interestingly, the use of covariances and  $\beta_{x_j,y}$ ,  $j = 2, 3$ , still resulted in overcoverage caused by overestimation of the variance. This means that inferences based on these moments would be more conservative than necessary, yet they had the highest variance reduction of all the moments tested, providing an interesting link to the concept of superefficiency (Bahadur, 1964). Further research on the variance estimator is needed to potentially correct for its overestimation.

Taken together, the simulation study showed promising results regarding very small sample sizes as  $n = 15$ , and however, one should still be cautious as the estimators are only proved to be consistent, not unbiased. To be sure that the inference will be valid in the sample at hand, a simulation to test the adopted scenario, i.e., model to be estimated and data set, is advised. In Sect. 6, we showed the applicability of the theoretical results to real data, where for the variable SpanLex the width of the confidence unions was significantly smaller than the width of the corresponding point-CI, if an appropriately small external interval is used. This shows the usefulness of adopting an externally informed model for applied problems.

A possible limitation of GMM is the assumption of the covariance matrix of the external moments being positive definite, which excludes distributions for which the required covariance matrix does not exist, e.g., the Cauchy distribution. Nevertheless, in many psychological applications the variables have a constrained range of values, so that at least the existence of the covariance matrix can be assumed. In general, the applicability of the method is not overtly limited by its assumptions. Another limitation is that the true value of the external moment must be within the external interval. However, this identifiability assumption, or an analogous assumption, exists in other approaches, and it is much weaker than point identifiability. Thus, a more robust use of external information is possible, up to using the full range of possible values, which would definitely lead to a valid, more robust, but also very conservative inference. The construction of the external moment interval in Sect. 6 was based on a rough, subjective approximation. The question of how to construct the external intervals requires further research. In particular, further links to existing techniques for eliciting intervals and preventing overconfidence bias would be important.

An application of the theory to generalized linear models or multi-level models is of inherent interest for psychological research, especially as Corollary 1 sets the foundation for research on more complex models. At first glance, the results appear to be in conceptual “conflict” with multi-level-models, since these often assume the random effects to be normally distributed and in this case there is no bounded interval, that includes the true parameter. However, even in these models there are fixed (hyper-)parameters one could know bounds for, and hence, it would be interesting for future research to analyze the behavior of these models in the external GMM framework. With respect to the limitation of robustness found in the simulation study, it would be interesting to investigate how robust the estimators are as a function of the length of the external interval. Finally, research on (the properties of) significance tests based on the use of an external intervals would be of great interest.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Data availability** The dataset of Pluck and Ruales-Chieruzzi (2021) analyzed during the current study is available in the PsychArchive repository, <https://doi.org/10.23668/psycharchives.2897>. All data generated by the simulations in this study are included in this article and its supplementary information files.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory

regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Augustin, T. (2002). Neyman–Pearson testing under interval probability by globally least favorable pairs: Reviewing Huber–Strassen theory and extending it to general interval probability [Imprecise probability models and their applications]. *Journal of Statistical Planning and Inference*, 105(1), 149–173. [https://doi.org/10.1016/S0378-3758\(01\)00208-7](https://doi.org/10.1016/S0378-3758(01)00208-7)
- Augustin, T., Coolen, F. P., De Cooman, G., & Troffaes, M. C. (2014). *Introduction to imprecise probabilities*. Hoboken: Wiley.
- Bahadur, R. R. (1964). On Fisher's bound for asymptotic variances. *The Annals of Mathematical Statistics*, 35(4), 1545–1552.
- Berger, J. O. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25(3), 303–328. [https://doi.org/10.1016/0378-3758\(90\)90079-A](https://doi.org/10.1016/0378-3758(90)90079-A)
- Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. Hoboken: Wiley.
- Buckley, J. J. (2004). *Fuzzy statistics*. Berlin, Heidelberg: Springer.
- Burin, D. I., Jorge, R. E., Arizaga, R. A., & Paulsen, J. S. (2000). Estimation of premorbid intelligence: The word accentuation test—Buenos Aires version. *Journal of Clinical and Experimental Neuropsychology*, 22(5), 677–685. [https://doi.org/10.1076/1380-3395\(200010\)22:5;1-9;FT677](https://doi.org/10.1076/1380-3395(200010)22:5;1-9;FT677)
- Cameron, A., & Trivedi, P. (2005). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511811241>
- Cassidy, R., & Manski, C. F. (2019). Tuberculosis diagnosis and treatment under uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, 116(46), 22990–22997.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1.
- Chaudhuri, S., Handcock, M. S., & Rendall, M. S. (2008). Generalized linear models incorporating population level information: An empirical-likelihood based approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(2), 311–328.
- Del Ser, T., González-Montalvo, J.-I., Martínez-Espinoza, S., Delgado-Villalpalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the word accentuation test and its application to the diagnosis of dementia. *Brain and Cognition*, 33(3), 343–356. <https://doi.org/10.1006/brcg.1997.0877>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap* (Vol. 57). New York, NY: Chapman & Hall.
- Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470), 680–701. <https://doi.org/10.1198/016214505000000105>
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029–1054.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251–1271.
- Hellerstein, J. K., & Imbens, G. W. (1999). Imposing moment restrictions from auxiliary data by weighting. *The Review of Economics and Statistics*, 81(1), 1–14.
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Hoboken: Wiley.
- Huber, P. J. (1981). *Robust statistics*. Hoboken: Wiley. <https://doi.org/10.1002/0471725250.ch7>
- Imbens, G. W., & Lancaster, T. (1994). Combining micro and macro data in microeconomic models. *The Review of Economic Studies*, 61(4), 655–680.
- Insua, D. R., & Ruggeri, F. (Eds.). (2000). *Robust Bayesian analysis*. New York: Springer. <https://doi.org/10.1007/978-1-4612-1306-2>
- Jann, M. (2023). Testing the coherence of data and external intervals via an imprecise Sargan–Hansen test. In *International symposium on imprecise probability: Theories and applications* (pp. 249–258).
- Kadane, J. B., & Wolfson, L. J. (1998). Experiences in elicitation. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(1), 3–19.
- Kwakernaak, H. (1978). Fuzzy random variables—I. Definitions and theorems. *Information Sciences*, 15(1), 1–29. [https://doi.org/10.1016/0020-0255\(78\)90019-1](https://doi.org/10.1016/0020-0255(78)90019-1)
- Lele, S. R., & Das, A. (2000). Elicited data and incorporation of expert opinion for statistical inference in spatial studies. *Mathematical Geology*, 32, 465–487. <https://doi.org/10.1023/A:1007525900030>
- Manski, C. F. (1993). Identification problems in the social sciences. *Sociological Methodology*, 23, 1–56.
- Manski, C. F. (2003). *Partial identification of probability distributions*. Berlin: Springer.
- Manski, C. F., & Pepper, J. V. (2013). Deterrence and the death penalty: Partial identification analysis using repeated cross sections. *Journal of Quantitative Criminology*, 29(1), 123–141.
- Meyers, J. E., Zellinger, M. M., Kockler, T., Wagner, M., & Miller, R. M. (2013). A validated seven-subtest short form for the Weis-IV. *Applied Neuropsychology: Adult*, 20(4), 249–256. <https://doi.org/10.1080/09084282.2012.710180>
- Newey, W. K., & McFadden, D. (1994). *Chapter 36 large sample estimation and hypothesis testing*. Amsterdam: Elsevier. [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4)

- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2), 237–249.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine*, 18(3), 1–15. <https://doi.org/10.1371/journal.pmed.1003583>
- Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, 145(2), 189–236. <https://doi.org/10.1037/bul0000182>
- Pluck, G. (2018). Lexical reading ability predicts academic achievement at university level. *Cognition, Brain, Behavior*, 22(3), 175–196.
- Pluck, G. (2020a). *Datasets for: Estimation of premorbid intelligence and executive cognitive functions with lexical reading tasks*. <https://doi.org/10.23668/psycharchives.2897>
- Pluck, G. (2020b). A lexical decision task to measure crystallized-verbal ability in spanish. *Revista Latinoamericana de Psicología*, 52, 1–10.
- Pluck, G., & Ruales-Chieruzzi, C. B. (2021). Estimation of premorbid intelligence and executive cognitive functions with lexical reading tasks. *Psychology and Neuroscience*, 14, 358.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.Rproject.org/>
- Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3), 393–415.
- Spieß, M. (1998). A mixed approach for the estimation of probit models with correlated responses: Some finite sample results. *Journal of Statistical Computation and Simulation*, 61(1–2), 39–59. <https://doi.org/10.1080/00949659808811901>
- Spieß, M., Jordan, P., & Wendt, M. (2019). Simplified estimation and testing in unbalanced repeated measures designs. *Psychometrika*, 84(1), 212–235. <https://doi.org/10.1007/s11336-018-9620-2>
- Steffen, A., Thom, J., Jacobi, F., Holstiege, J., & Bätzing, J. (2020). Trends in prevalence of depression in Germany between 2009 and 2017 based on nationwide ambulatory claims data. *Journal of Affective Disorders*, 271, 239–247. <https://doi.org/10.1016/j.jad.2020.03.082>
- Vaart, A. W. (1998). M- and z-estimators. In *Asymptotic statistics* (pp. 41– 84). Cambridge University Press. 10.1017/CBO9780511802256.006
- Walter, G., & Augustin, T. (2009). Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice*, 3(1), 255–271. <https://doi.org/10.1080/15598608.2009.10411924>
- Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2), 149–170. [https://doi.org/10.1016/S0888-613X\(00\)00032-3](https://doi.org/10.1016/S0888-613X(00)00032-3)
- Weichselberger, K. (2001). *Elementare grundbegriffe einer allgemeineren wahr-scheinlichkeitsrechnung I: Intervall-wahrscheinlichkeit als umfassendes konzept* (Vol. 1). Berlin Heidelberg: Springer.
- Weiss, R. H. (2006). *Cft 20-r: Grundintelligenztest skala 2-revision*. Gottingen: Hogrefe.
- Weiss, R. H. (2019). *Cft 20-r mit ws: Grundintelligenztest skala 2-revision (cft 20-r) mit wortschatztest und zahlenfolgentest-revision (ws/zf-r)* (2nd ed.). Gottingen: Hogrefe.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1167.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.
- Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10), 1–17. <https://doi.org/10.18637/jss.v011.i10>
- Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object oriented implementation of clustered covariances in R. *Journal of Statistical Software*, 95(1), 1–36. <https://doi.org/10.18637/jss.v095.i01>
- Zhong, B., & Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, 87(4), 929–938.