JOURNAL OF ECONOMIC METHODOLOGY
2025, VOL. 32, NO. 1, 14–32
https://doi.org/10.1080/1350178X.2025.2468462

Mostly harmless econometrics? Statistical paradigms in the 'top five' from 2000 to 2018

John-Oliver Engler [a,b], Julius J. Beeck [b] and Henrik von Wehrden [b]

[a]Faculty of Natural and Social Sciences, University of Vechta, Vechta, Germany; [b]Faculty of Sustainability, Leuphana University of Lüneburg, Lüneburg, Germany

ABSTRACT
We explore the connection between four major inferential paradigms in statistical science and inferential practice in current econometrics. We develop the argument that econometrics is still largely characterized by John Stuart Mill's conception of statistical inference from data, who saw a distinction between 'theorists' and 'practical men'. We follow up with a review of all empirical papers published in the Top 5 economics journals in the period 2000–2018 ($N = 2,258$). In spite of Rodrik's [(2015). *Economics rules: The rights and wrongs of the dismal science*. W. W. Norton & Company] much-debated notion of economics that sees issues of model selection at the core of the discipline, the 'theory first' / 'pre-eminence of theory' approach vastly dominates in the sample (94.0%). When model selection and model uncertainty is accounted for, this largely happens under the frequentist statistical paradigm. This finding may be explained by frequentism's special role as an 'orientational paradigm' (Hoyningen-Huene and Kincaid, [2023]. What makes economics special: Orientational paradigms. *Journal of Economic Methodology*, *30*(2), 188–202) in economics.

ARTICLE HISTORY
Received 8 November 2023
Accepted 13 February 2025

KEYWORDS
Bayesianism; econometrics; frequentism; likelihoodism; philosophy of science; paradigm

JEL CODES
A12; B16; B23 B41; C18

1. Introduction

One of the cornerstones of any scientific endeavor, including economics, is the confrontation of theories and models with empirical data (Allais, 1991; Feynman, 1967; Spanos, 2010b). Econometric analyses inform real-world policy and decision making and therefore create ripples that affect the world well beyond the realm of academic research. Moreover, unlike natural scientists, economists, to some extent with the exception of randomized controlled field trials (see Duflo, 2020), typically have no way of experimenting with the economy. While quasi-experiments and laboratory settings have enabled substantial knowledge advancements in the recent past, lab experiments study *individual* behavior, decision making or economic reasoning and, by definition, quasi-natural settings offer no way to *control the intervention*, i.e. to actively experiment, with the economy as a whole.[1] Hence, in the words of Koopmans (1947, p. 166),

> Economists are not in a position to perform experiments with an economic system as a whole for the sole purpose of establishing scientific truth (although deliberate changes in parts of the system have been undertaken at various occasions for other than scientific purposes, and have incidentally added to our information).

CONTACT John-Oliver Engler john-oliver.engler@uni-vechta.de Faculty of Natural and Social Sciences, University of Vechta, Driverstraße 22, D-49377 Vechta, Germany; Faculty of Sustainability, Leuphana University of Lüneburg, Universitätsallee 1, D-21335 Lüneburg, Germany

The go-to method to confront economic theory with real-world data (henceforth: theory-data confrontation problem) is regression analysis. Typically, the aim of regression analysis in economics is to make statistical inferences from models, which represent economic theory, using empirical data, which is a snapshot of some aspect of an economic system of interest. In this sense, models test hypotheses about the system under study, which follow from or build on the theory available for the particular problem at hand. For example, Haavelmo (1944, p. iii) writes

> The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier.

There are two pivotal questions of statistical inference (cf. Burnham & Anderson, 2001). First, one would like to know what the magnitude of model parameters is and with what precision they can be quantified. In essence, the researcher would like to know whether some model parameter is economically meaningful. This first problem is one of *model estimation*. Second, one would like to know whether parameters should be included in a model for further inference at all. This is a problem of *model selection*. The purpose of model selection is sometimes described as finding the single-best fitting model out of a set of candidate models. This statement, while not generally false, misses a critical point: model selection can account for the fact that we do not fully know the underlying economic process or mechanism that generated the data, which means that it can quantify model uncertainty.[2]

Thus, because an economist can generally not expect to know exactly what process generated the data, it seems straightforward to assume that model selection from a set of candidate models should play a prominent role in econometric analyses. Indeed, Rodrik (2015) has argued that economics is about the selection of the right model(s) for the purpose at hand. However, Spanos (2009, 2010b) laments a historically grown predominance of econometric approaches that merely use data as a means to 'instantiate' theories rather than testing them against the data (Spanos, 2010b, p. 202f), with many of his other works critiquing specific aspects of this approach to empirical evidence in economics over the years, the common theme being invalid inferences due to statistical model misspecification (e.g. Do & Spanos, 2024; McGuirk et al., 1993; Spanos, 1989, 1995, 2011; Spanos & Taylor, 1984).

Against this background, this paper makes three main contributions. First, we connect the model-selection problem in econometrics with different perspectives on inference from data according to the four major schools of statistical thinking, Bayesian inference, frequentism, likelihoodism and information theory ('Akaikeanism', see Bandyopadhyay & Forster, 2011).[3] Second, we provide empirical evidence for the claim (see Reiss, 2013; Spanos, 2010b) that econometrics is dominated by a theory-first narrative that does not address the model-selection issue. In particular, we provide a bibliometric analysis of how the issues of model selection and model uncertainty have been represented in relevant publications in the Top Five (or 'T5'; Heckman & Moktan, 2020) academic journals in economics according to the rankings available (Bornmann et al., 2018; Heckman & Moktan, 2020; Kalaitzidakis et al., 2011) from 2000 through 2018. We quantify what fraction of T5 publications accounted for model uncertainty in any given way and checked in addition which statistical paradigm was used. Third, we discuss our findings in light of some recent contributions to the philosophy of economics, most notably Hoyningen-Huene and Kincaid's (2023) notion of operational paradigms and Rodrik's (2015) account of economics as a science that selects between different models.

The paper proceeds as follows. Section 2 clarifies setting and notation (Section 2.1) and elaborates on the different schools of thinking in econometrics (Section 2.2) and statistics (Section 2.3) on inference from data. Section 3 describes and discusses the case study, i.e. the survey of model uncertainty and statistical paradigms in the T5 from 2000 through 2018. In Section 4, we discuss our findings in a broader context, primarily with respect to Rodrik's (2015) account of economics as a science of model selection and Hoyningen-Huene and Kincaid's (2023) concept of 'orientational paradigms'. Section 5 concludes.

2. Perspectives on evidence and inference: econometrics vs. statistics

2.1. Setting and notation

The typical theory-data confrontation problem in economics involves observational data $X = \{x_1, \ldots, x_N\}$ where $x_i \in \mathbb{R}^m$, at least one linear model $\mu(X) = E(Y|X) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in}$ and the normality assumption $y_i \sim N(\mu(X), \sigma^2)$. Importantly, the normality assumption can be relaxed to obtain generalized linear models, for example logit or probit models, to name the most common use cases in econometric practice. With regard to terminology, Rodrik (2015, p. 113ff) has recently pointed out that while economists frequently speak of 'economic theories', as in 'theory-data confrontation' here, they actually imply 'economic models'. Rodrik is right to argue that theories are universal and that universality is impossible in the social sciences. It would therefore be better to refer to all econometric analyses as model-data confrontation or model-selection problem from the outset (see also Grüne-Yanoff & Marchionni, 2018). However, as will be clear from the next section, there is a historically grown propensity towards the notion of 'theory', so that it will not be possible to entirely avoid it.

2.2. Econometrics, evidence and model selection

Spanos (2010b) describes three broad traditions with respect to the theory-data confrontation problem in econometrics. These traditions follow distinct worldviews. The first and oldest tradition sees the data as a means to the ultimate end of theory quantification. Reiss (2013) traces this understanding back to John Stuart Mill's (1844, p. 142) distinction between 'practical men' and 'theorists'. Importantly, Mill's theorists use data to 'instantiate theories' (Spanos 2010b, p. 202f) rather than testing them, equivalent to confirmatory modeling. Hence, this longstanding econometric tradition, which has been referred to as 'pre-eminence of theory' (Spanos, 2009) or simply 'theory first' (Reiss, 2013), concerns itself with parameter estimation in econometric models deduced from economic theory. In this theorist understanding of econometrics, model uncertainty and model selection are not an econometric issue, because the economic model underlying the econometric specification is logically deduced from general higher-level principles or axioms before any data analysis. Therefore, the theorist procedure is perfectly analogous to most data-analysis problems in physics, for example when determining the decay constant of some radioactive substance from empirical data, in that the law of exponential decay follows deductively from theoretical nuclear physics. Thus, what remains to be done is parameter estimation. The remaining two traditions that Spanos (2010b) describes are then variants of Mill's practical men at the other end of the spectrum.

Generally, Mill's practical men seek to apply inductive reasoning to generalize from specific observations. According to Reiss (2013), they share Mill's skepticism towards induction in economics in general, but contend that quasi-experimental setups can be found and used to apply inductive reasoning. Specifically, Spanos' second tradition revolves around data-driven explorative modeling with the aim of uncovering economic regularities and 'laws' by maximizing goodness-of-fit. Econometric modeling in this tradition is about prediction. Initially advocated by Burns and Mitchell (1946), this approach prompted scorn and mockery from proponents of the first tradition (e.g. 'measurement without theory', Koopmans, 1947). Reiss (2013) points out that the ensuing 'measurement without theory' controversy was essentially in many aspects identical to the so-called *Methodenstreit* between the inductivist camp led by Gustav Schmoller and the deductivist side led by Carl Menger in the beginning of the twentieth century. Importantly, the data-driven explorative approach has been referred to as 'algorithmic modeling culture' by statistician Breiman (2001), and recent voices have been more appreciative in calling for more diversity in econometric tools (Athey & Imbens, 2019). Lastly, the third tradition or 'third way', according to Spanos (2010b), takes the middle route and aims at finding a compromise between theory-driven and data-driven econometric modeling (Granger et al., 1995; e.g. Hendry, 2009). That is, the third tradition starts with a model (or a set of models) informed by economic theory and tries to select, or rather de-select (see Engler & Baumgärtner, 2015), models to narrow down the set of plausible models given the data.[4] That is, the 'third way' strongly emphasizes accounting for the statistical regularities in the data while still being informed by the relevant economic theory for the phenomena under study. Motivated by the principle that 'empirical models that do not account for the statistical regularities in the data are likely to give rise to untrustworthy empirical evidence and very poor predictive performance' (Spanos, 2010b, p. 236), this approach to econometrics embraces the importance of *both* model estimation *and* model selection. It resembles the currently dominant approaches to data analysis in environmental science and ecology that emphasize model uncertainty and multi-model inference (Aho et al., 2014; Grace, 2006; Grueber et al., 2011; Johnson & Omland, 2004). On the other hand, it entails difficult methodological questions such as how to meaningfully compare different models in light of data, what constitutes good criteria for testing models and hypotheses, and how to deal with our epistemological uncertainty. Hence, the following section aims to disentangle the role of statistical paradigms in answering these questions.

2.3. Statistical paradigms, evidence and model selection

In statistics, four major paradigms exist to date, in particular with respect to statistical inference (Bandyopadhyay & Forster, 2011; but see also, e.g. Cox, 2006; Lillestol, 2014; Young & Smith, 2005). These include classical or frequentist inference, Bayesian inference, likelihood inference and information-theoretic (Akaikean) inference. All paradigms model the conditional distribution $f(X|\Theta)$ of observed data X given some parameter Θ, generally from $\mathbb{R}^n$, but differ in how they exploit the model (see Lillestol, 2014 and examples therein). In particular, the paradigms differ in what they consider evidence in favor (or against) some hypothesis or model. Following Royall (1997) and Bandyopadhyay and Forster (2011), statistical inference relates to the four following questions:

(Q1) Belief: What should we believe and to what degree, given the data?

(Q2) Evidence: What do the data say regarding evidence for H against its alternative?

(Q3) Decision: What should one do, given the data?

(Q4) Prediction: What do the data tell us about the predictive accuracy of the hypothesis?

Infamously, Royall (1997) argued that Bayesians addressed (Q1), frequentists (Q3) and only likelihood inference addressed (Q2). Of course, this view is contentious to say the least, and there have been many contributions that argue to the contrary, e.g. Mayo (1996) and Mayo and Spanos (2010) in defense of frequentist inference, Morey et al. (2016) in defense of a Bayesian account of evidence, and Burnham and Anderson (2002) have advocated the information-theoretic paradigm to assess the relative evidence of competing models. From a foundational point of view, all approaches make use of the likelihood function as a central concept, but differ with respect to their conformity to the Likelihood Principle (see, e.g. Gandenberger, 2016). Informally speaking, the Likelihood Principle is a proposition that states that, given a statistical model, all the evidence in a sample relevant to the model parameters is contained in the likelihood function. Specifically, the Likelihood Principle states that 'the evidential import of a body of data for a set of hypotheses depends *only* on how probable that body of data would be if each hypothesis in the set were true' (Gandenberger, 2016, p. 4, emphasis added). Many arguments and lines of reasoning support this principle (Birnbaum, 1962; Gandenberger, 2015), but it is far from a universally accepted notion in statistics and philosophy of science (see, e.g. Akaike, 1982; Berger & Wolpert, 1988; Bjørnstad, 1996; Evans et al., 1986; Kalbfleisch, 1975; Mayo, 2014).[5]

Crucially, Bayesian inference, likelihoodism and Akaikeanism conform to the Likelihood Principle, but frequentism does not (Gandenberger, 2016; Royall, 1997). As a consequence, significance testing, one of the prime and most abundantly used tools of frequentist inference, does also not conform to the Likelihood Principle. What does this mean? Consider the following example, a similar version of which was first presented by Lindley and Phillips (1976). Suppose a coin is tossed 12 times and lands heads 3 times. A researcher wants to assess the hypothesis that this coin is fair, $H_0{:}p_H = .5$, as opposed to $H_A{:}p_H < .5$. We then have for the p-value

$$P_{12}(k \leq 3 \mid H_0) = \left(\frac{1}{2}\right)^{12}\left[\binom{12}{0} + \binom{12}{1} + \binom{12}{2}\right] = \frac{299}{4096} \approx 7.3\% \tag{1}$$

Hence, based on this calculation, we fail to reject H_0 at the typical threshold level of significance of $\alpha = 5\%$. However, this calculation is only correct if our experimental design was to toss the coin 12 times and count the number of heads. Suppose that the experimental design was to toss the coin *until* 3 heads were observed instead, and that the coin had to be tossed 12 times in order to achieve this. Under this experimental design, the p-value is

$$P_{11}(k \leq 2 \mid H_0) \cdot\ P_1(k = 1 \mid H_0) = \left(\frac{1}{2}\right)^{11}\left[\binom{11}{0} + \binom{11}{1} + \binom{11}{2}\right] \cdot \frac{1}{2} = \frac{67}{4096} \approx 1.6\% \tag{2}$$

We thus reject H_0 at $\alpha = 5\%$, and see that in null hypothesis significance testing, the experimental design plays a crucial role in determining whether a result is statistically significant, which violates the Likelihood Principle. Specifically, the Likelihood Principle implies that the evidential import of the data in favor of H_0 only depends on the likelihood of the parameter being equal to $\frac{1}{2}$, given the data. Much of the implicit or explicit quarreling between proponents of the four main statistical paradigms, most prominently the skirmish between frequentists and Bayesians, boils down to a disagreement about whether violating the Likelihood Principle, as in the simple example above, is normatively desirable or not. Frequentists maintain that it is desirable to violate the Likelihood Principle, while advocates of the other three paradigms maintain the this is not the case. It goes without saying that, as with almost any normative issue, none of the camps is obviously right or wrong as strong arguments for both positions exist.[6]

Nonetheless, the stance of the different statistical paradigms towards the Likelihood Principle has far-reaching repercussions as to the question of evidence. Based on the Likelihood Principle, Hacking (1965) formulated the Law of Likelihood, which states that the degree to which data X favors some hypothesis H_1 over another hypothesis H_2 is expressed by the likelihood ratio $\Lambda = \frac{L(H_1|X)}{L(H_2|X)} = \frac{P(X|H_1)}{P(X|H_2)}$. Λ is sometimes also referred to as evidence ratio. If $\Lambda > 1$, then H_1 is favored over H_2 and if $\Lambda < 1$ then H_2 is favored over H_1 by the data X. Royall's (1997) book on the likelihoodist paradigm can be seen as a formalization and exploration of Hacking's notion of evidence. In Bayesian statistics, Λ is known as Bayes factor (see Kass & Raftery, 1995). Rules of thumb for the interpretation of Λ exist, e.g. according to Royall (1997), a value of 8 indicates 'fairly strong' and a value of 32 'strong' favoring of H_1 over H_2. For the Bayes factor, Kass and Raftery (1995) or Stefan et al. (2019) have compiled various suggestions for threshold values from the literature, referring in particular to the works of Harold Jeffreys (e.g. Jeffreys, 1935). As Gandenberger (2016, p. 5) points out, all of these threshold values are somewhat arbitrary. Information-theoretic inference (Akaikeanism) is based on Akaike's (1973) groundbreaking finding of an estimator $\hat{I}(f, g)$ for the information loss when describing an unknown data-generating process f by some model g (with K parameters) *without* knowing f. Akaike found that, for large samples and models f and g, $\hat{I}(f, g) = \log L(\hat{\theta}|X) - K$. Specifically, it is possible to go beyond likelihood or evidence ratios, and quantify the relative strength of evidence, or 'weight of evidence' (see Burnham & Anderson, 2002) for each model i in a candidate set $\mathcal{M}$ given some data X and $\Delta_i = \text{AIC}(g_i) - \text{AIC}(g_0)$ where

$\text{AIC} = -2\log L(\hat{\theta}|X) + 2K$ and g_0 being the AIC-best model in $\mathcal{M}$ in the sense of minimizing AIC:

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta_r\right)} \tag{3}$$

Frequentism and its related apparatus of null hypothesis significance testing adopts a critical rationalist stance. That is, hypotheses are tested against data in the sense of an evaluation of the probability to obtain X or something more extreme if H_0 were assumed to hold (p-value). Thus, one evaluates $p = P(X|H_0)$ rather than $P(H_0|X)$ and if $p < \alpha$, with $\alpha = .05$ usually, one may say that H_0 has not passed (i.e. it has failed) the test at significance level α, which could be taken as evidence *against* H_0. However, by construction, even a large p-value does *not* constitute evidence *in favor* of H_0, it only constitutes evidence that it would be unreasonable to reject H_0 with data X (see, e.g. Wasserstein & Lazar, 2016). Thus, there could be a large number of other hypotheses that also 'survive' a test in the above sense (underdetermination problem). Mayo (1996) and Mayo and Spanos (2011) claim to have found a solution for the underdetermination problem by developing the concept of 'severe testing', but this issue ultimately remains a subject of dispute among statisticians, economists and philosophers of science (see contributions in Mayo & Spanos, 2010). Recently, Spanos (2024) has provided an encompassing overview on the concept of 'post-data severity' evaluation as an alternative to Royall's (1997) approach to evidence via the likelihood ratio Λ, including examples from ecology, psychology and biology. Frequentism's main tool of model selection, the likelihood ratio test, does *not* assess the likelihood ratio in Royall's sense of an evidence ratio. Instead, it uses the fact that the transformation $-2\log\left\{\frac{L(H_1|X)}{L(H_2|X)}\right\}$ is asymptotically χ^2 distributed which enables calculation of a p-value for any two nested competing models[7] (Burnham & Anderson, 2014). Lastly, there is also a following of procedures of specification searching based on parameter significances, where one typically starts with the maximal (general unrestricted) model, removing statistically insignificant model parameters until only significant ones remain upon fitting. This approach is known as 'LSE approach' in econometrics (Gilbert, 1989; Hendry, 2009, 2011), or simply backward selection/elimination more broadly in statistics (see Kadane & Lazar, 2004). Table 1 gives an overview of each paradigm's main tools for evidence appraisal and model selection.

Finally, while the statistical paradigms disagree on the desirability of the Likelihood Principle, the likelihood function is still a central tool in all four paradigms discussed here, for example through its role in maximum likelihood estimation of parameters in regression, as the sampling distribution in Bayesian inference or as a constituent of all common information criteria (see discussion around equation 3 above). Therefore, no matter what statistical paradigm is followed, because $L(\theta|X) = \prod f(X|\theta)$ all inferences always critically hinge on the adequacy of the probabilistic assumptions (see Spanos, 2019).

2.4. Preliminary conclusion

Econometrics has to deal with major uncertainties from two sources. First, there are three schools of thought within econometrics with respect to the theory-data confrontation problem that we have

Table 1. Main tools for evidence appraisal and model selection according to statistical paradigm.

Frequentism	Bayesianism	Likelihoodism	Akaikeanism
Significance testing (e.g. likelihood ratio or Vuong test, forward/ backward specification search, mis-specification testing)	Bayes factor, Bayesian model averaging	Likelihood function, likelihood (evidence) ratio	Information criteria (e.g. AIC, BIC, DIC), weight of evidence / model probabilities based on differences in information-criteria values, model averaging

argued to be varieties of John Stuart Mill's 'theorists' vs. 'practical men'. Mill's 'theorists' deduce theory from high-level axioms and use empirical data to quantify, or 'instantiate' their theories (cf. Spanos, 2010b, p. 202f). Mill's 'practical men' generally reason inductively, i.e. they seek to generalize or 'learn' from data starting from specific observations, but differ in their perspective on the role of economic theory in this process (Section 2.2). Second, there are four statistical paradigms and their respective notions of evidence and model appraisal (Section 2.3). The variety in statistical paradigms is not much of a problem for the pre-eminence of theory (PET) approach, because, by construction, PET does not concern itself with questions of model selection and model uncertainty. The question that proponents of the PET school of econometrics have to answer is much rather what kind of scientific progress can be generated by emulating physical approaches to data analysis in a setting that is virtually always different from physics, both in terms of stage of development of available theories and in terms of available data.

Non-PET approaches to econometrics, that is current variants of Mill's 'practical men', need to confront the problem of model selection. The different statistical paradigms offer multiple tools for this task (see Table 1), albeit frequentism's tools are not capable of comparing more than two models at once, which may lead to inconclusive results. The likelihood function under the likelihoodist paradigm and Akaikeanism's information criteria are not limited in this particular way, in that all conclusions are ultimately based on the likelihood function, but strongly depend on the quality of candidate models (see, e.g. Burnham & Anderson, 2002). On the other hand, whether a candidate model (or model hypothesis) is statistically plausible at all, i.e. whether it should be ruled out or not, is something that can directly be assessed with statistical significance testing and its focus on $P(X|H_0)$ (Clauset et al., 2009; Engler & Baumgärtner, 2015). As Breiman (2001, p. 202) has put it

> when a model is fit to data to draw quantitative conclusions, the conclusions are about the model's mechanism, not about nature's mechanism. It follows that: If the model is a poor emulation of nature, the conclusions may be wrong.

We now turn to quantifying how theory-data confrontation has been approached in practice in the Top-5 journals in economics in the period 2000–2018.

3. Case study: statistical paradigms and model selection in the Top 5 from 2000–18

3.1. Methods

3.1.1. Data acquisition

Figure 1 provides an overview over the paper sampling, classification and review process. The initial database contained all final research articles as listed in Scopus (www.scopus.com) in the following five peer-reviewed economics journals in the period 2000–2018: *American Economic Review, Econometrica, Quarterly Journal of Economics, Review of Economic Studies* and the *Journal of Political Economy*.[8] The reason for surveying these five journals is that they are commonly considered the top five scientific publishing outlets in economic (Bornmann et al., 2018; Heckman & Moktan, 2020). We excluded conference papers (1549), book reviews (656), notes and comments (161), editorials (52), errata and corrigenda (47), letters and short surveys (3) and those publications listed as 'undefined' in the Scopus database (9). Due to the intent of this research to survey all research in the top five, we did not impose any restrictions regarding minimum number of citations per year. When querying Scopus on February 28, 2020, the resulting database consisted of $N = 5161$ publications.[9]

3.1.2. Paper screening

We screened every publication for the following three criteria:

(1) the study presented must be empirical;
(2) it must use regression analysis;

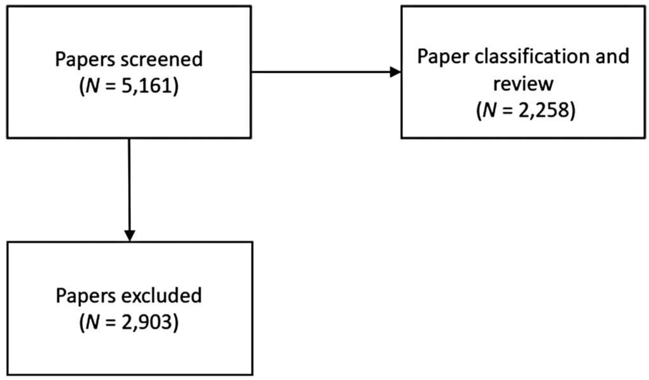

Figure 1. Overview of the paper sampling, classification and review process.

(3) it must *not* use regression for merely illustrative purposes, for example to illustrate a novel estimation technique or procedure.

We thus required that the paper must engage with estimation, assessment and possibly comparison of econometric models. The screening process often included an assessment of the full-text document rather than just abstract, title and keywords and was performed by two researchers (JJB and JOE). $N = 2{,}903$ publications did not fulfill both inclusion criteria and were therefore excluded from the database. The most frequent reason for exclusion was that papers were about development or refinement of economic theory, and thus conceptual but not empirical. Hence, $N = 2{,}258$, i.e. 43.8% of all papers screened, met all three criteria and proceeded to the full-text document review and classification stage (Figure 1).

3.1.3. Paper classification and review process

We classified each of the $N = 2{,}258$ papers according to how it approached the theory-data confrontation problem. Specifically, we assessed whether it followed a single or multi-model approach to the theory-data confrontation problem. The single model approach corresponds to what Spanos (2009) and Reiss (2013) have referred to as 'pre-eminence of theory' or 'theory first', respectively, and which we have introduced in the previous section, in particular Section 2.3. To summarize the methodological perspective of the approach, we quote Kydland and Prescott (1991, p. 171) who pointedly write

> The issue of how confident we are in the econometric answer is a subtle one which cannot be resolved by computing some measure of how well the model economy mimics historical data. The degree of confidence in the answer depends on the confidence that is placed in the economic theory being used.

That is, we classified a paper as following the PET approach, if it followed a confirmatory analysis in the sense of estimating one general econometric model based on arguments and derivations from economic theory. Any kind of model comparison, including the comparison of nested models, led to the paper being classified as following a multi-model approach.

If multiple models were considered in a paper, we decided which statistical paradigm was followed when dealing with model selection and model uncertainty. If tools from different statistical schools of thought were used in the same paper, for example when combining an information criterion and likelihood ratio tests, we tried to determine which statistical paradigm was dominant. In the example given, if an information criterion was only employed as an addendum following a frequentist analysis, we would classify the paper as frequentist. On the other hand, if the frequentist analysis occurred as an addendum to an Akaikean procedure, we would classify the paper as Akaikean. If we were unable to reach a conclusion, we classified the paper in all categories applicable, e.g. both as Akaikean and classical/frequentist.

3.2. Results

3.2.1. Descriptive statistics

Table 1 summarizes the results of our survey with regard to single model (PET, theory first) and multi-model approaches. Of all 5,161 publications in the period under investigation, 2,258 were relevant in the sense of our review procedure (Section 3.2, Figure 1). A clear majority of 94.0% (2,122 papers) in the T5 followed the PET approach, while 6.0% (136 papers) followed a multi-model approach (MMA). The *Quarterly Journal of Economics* featured the largest share of relevant papers relative to all papers published (68.7%), followed by the *Journal of Political Economy* (52.0%) and the *American Economic Review* (51.3%). The journals *Review of Economic Studies* (31.0%) and *Econometrica* (19.5%) featured the smallest share of relevant papers. In terms of prevalence of MMA, the only journal with a percentage share larger than 10% was the *Review of Economic Studies* with 11.7% (31 papers), followed by the Journal of Political Economy (8.1%, 31 papers) and *Econometrica* (8.1%, 17 papers). The *American Economic Review* (4.6%, 43 papers) and the *Quarterly Journal of Economics* (3.0%, 14 papers) remained below the 5% mark in the period investigated.

With regard to the four statistical paradigms (see Section 2.3), frequentism clearly dominated when multi-model approaches were actually employed (Table 3), as 64.7% (88 papers) of all MMA publications that used frequentist tools exclusively or predominantly. The second-most popular statistical paradigm was Akaikeanism (25.0%, 34 papers), followed by Likelihoodism (5.9%, 8 papers) and Bayesianism (4.4%, 6 papers). The *American Economic Review* was most pluralistic with regard to statistical schools of thinking, in that all paradigms were featured in at least three publications. The *Review of Economic Studies* was the only other journal to feature all four statistical paradigms at least once. On the other end of the spectrum, the *Quarterly Journal of Economics* did not feature any MMA publications that used Bayesian or likelihoodist reasoning. The *Review of Economic Studies* featured the largest frequentist share (67.7%, 21 papers) of publications that followed MMA, but this share was consistently in the range between 62.8% (*American Economic Review*) and 67.7% among the journals surveyed. The relative share of Akaikeanism was largest in *Econometrica* (35.7%, 5 papers). Generally, papers operating under a Bayesian or likelihoodist paradigm were rare, with the share of Bayesianism largest in the *Review of Economic Studies* (9.7%, 3 papers) and likelihoodism strongest in *Econometrica* (11.8%, 2 papers).

3.2.2. Temporal dynamics

Generally, we observe a slight increase in the number of relevant empirical publications in the T5 (Figure 2). Figure 3 illustrates the prevalence of multi-model approaches (MMA) to econometric modeling in these empirical publications in the sense explained in Section 2.2 from 2000 through 2018. There is a clear and steep rise in the share of articles that use MMA from 0.7% to 6.3% of all relevant publications in the period 2000–2003, before plateauing in the years 2006 and 2007 (7.7% each). In terms of absolute number of MMA papers published, a maximum of 14 publications (equivalent to an annual share of 9.8%) was attained in 2009. Ten or more MMA papers were published in five years, four of which were in the first decade of the twenty-first century (2001, 2006–07, 2009, 2012). Annual fluctuations in relevant published papers (Figure 2) combined with a stagnation of the number of papers that used MMA (Figure 3), the ten papers that made 2012 the 'record year' of the second decade resulted in a share of 7.0%. In terms of relative share, there is a clear downward trend of MMA in the literature surveyed since reaching its peak in 2009. Generally, the percentage has been hovering around 4–5% since 2010, the only exception being the aforementioned year 2012 (Figure 3).

3.3. Discussion

It has been argued on several occasions that there is a historically grown propensity in economics and econometrics towards a 'theory first' (Reiss, 2013) or 'pre-eminence of theory' (Spanos, 2009,

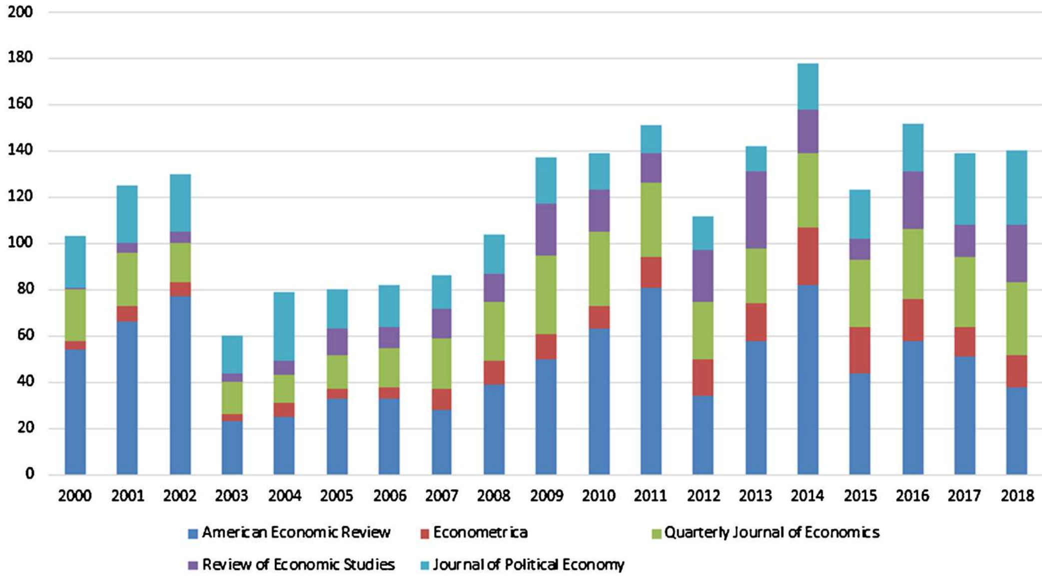

Figure 2. Temporal dynamics of relevant papers published in the T5 from 2000 through 2018.

2010b) approach to empirical data. However, these history-based arguments, though forceful, have not been supported, or tested, by empirical surveys on econometric practices in published papers that featured analyses of empirical data. Based on our survey of all empirical papers published in the T5 in the period 2000–2018, 94% of publications indeed employed a PET approach to empirical data, while the remaining 6% used a multi-model approach (MMA) that entails dealing with the issue of model selection. Clearly, as far as the T5 go in the period surveyed,

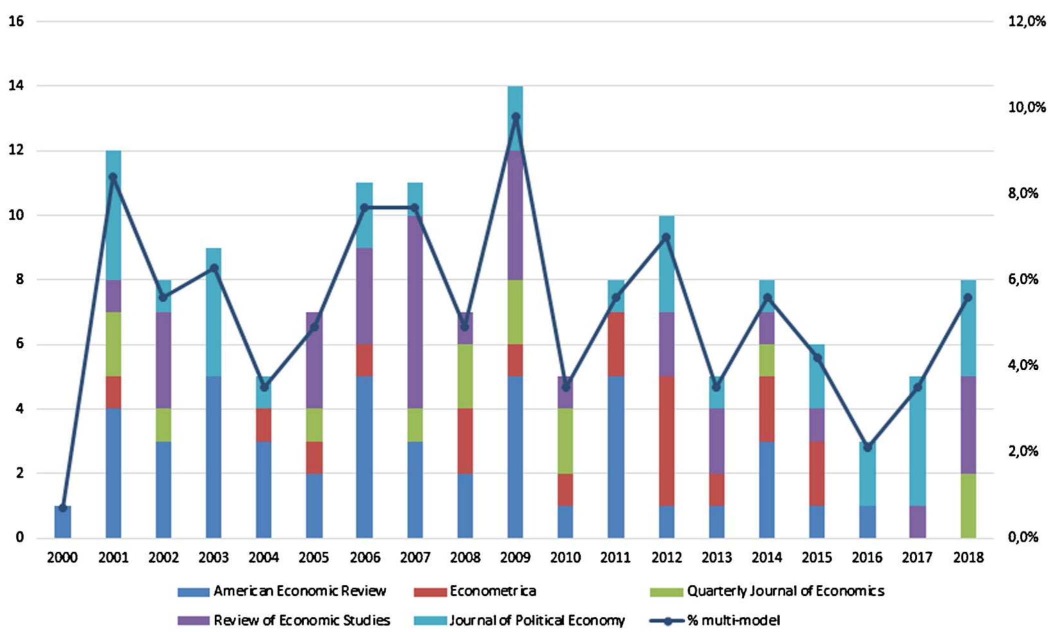

Figure 3. Temporal dynamics of multi-model approaches in the T5 from 2000 through 2018.

the claim that PET approaches to empirical data analysis vastly dominate in econometrics is empirically well supported.

Those papers that actually embraced model selection and thus considered several competitors as candidates to model the data-generating process predominantly followed a classical frequentist line of reasoning (64.7%). We found Akaikeanism (25.0%) to be the second-most frequently used statistical paradigm, mostly due to the popularity of information-theoretic methods in time-series analysis and forecasting. Bayesian (4.4%) or likelihoodist (5.9%) methods were only marginally employed. On the one hand, this finding is not particularly surprising, because frequentism still is the main statistical school of thinking in many fields of empirical research, including economics (Krämer, 2011; Rommel & Weltin, 2021; van Witteloostuijn & van Hugten, 2022). On the other hand, frequentism and its main inferential tool, significance testing, have been subject to harsh criticism over the years within economics (Abadie, 2020; Granger et al., 1995; Leamer, 1983; McCloskey & Ziliak, 1996; Ziliak & McCloskey, 2004), and has come under severe attack at the beginning of the twenty-first century in research in general, with professional statisticians leading much of the way (Ioannidis, 2005; Leek et al., 2017; McShane et al., 2019; McShane & Gal, 2017; Nuzzo, 2014; Wasserstein & Lazar, 2016). Some of the traction of MMA that we observe in particular at the beginning of the twenty-first century might be explainable by these general criticisms of the frequentist approach, in that model comparison is generally more straightforward in non-frequentist settings (see Section 2.3). However, in spite of all this, there is little indication that frequentism will be abandoned anytime soon (see also van Witteloostuijn & van Hugten, 2022) as has been vehemently hoped and argued for by some (McShane et al., 2019).

4. General discussion

This discussion will synthesize and present two general arguments about economic methodology. First, if economics is indeed a 'collection of models, along with a system of navigating among those models' (Rodrik, 2015, p. 208), then model selection should take precedence over mere model estimation in econometrics. Economics would thus benefit from embracing a multi-paradigmatic statistical view on the theory-data confrontation problem (Section 4.1). Second, the different statistical paradigms can be understood as 'orientational paradigms' (Hoyningen-Huene & Kincaid, 2023) in statistics, but not (yet) in economics. Orientational paradigms share some characteristics with Kuhnian paradigms, but add three characteristics that set them apart from paradigms in a Kuhnian sense. In economics, only the frequentist paradigm may be regarded as an orientational paradigm. Thus, apart from the limitations implied by the predominance of the PET approach to the theory-data confrontation problem, econometric analyses often remain limited to frequentist tools (Section 4.2).

4.1. Model selection in economics, and econometrics

Rodrik (2015) has recently argued that economics is a 'collection of models, along with a system of navigating among those models' (Rodrik, 2015, p. 208). Specifically, scientific progress in economics means 'expanding its collection of useful cases', i.e. models (Rodrik, 2015, p. 72). Grüne-Yanoff and Marchionni (2018) have put Rodrik's ideas into a formalized set-theoretic framework. If one accepted Rodrik's (2015) reflections on what economics is and how it works, an immediate practical consequence would be that model selection should take precedence over mere model estimation in econometrics. That is, the main econometric problem is one of model selection, not (only) one of model estimation. Remarkably, Grüne-Yanoff and Marchionni (2018), like Rodrik (2015), refrain from discussing the role of econometrics in theory-data confrontation (see Grüne-Yanoff & Marchionni, 2018, p. 274). According to Grüne-Yanoff and Marchionni (2018, p. 266), Rodrik's model-selection procedure consists of three steps, '(1) selecting candidate models, (2) identifying the critical assumptions, and (3) carrying out verification procedures'. In the context of size-distribution data, Engler and Baumgärtner (2015) have proposed a three-step procedure that is conceptually similar to Rodrik's (2015), which combines concepts from multiple statistical paradigms. From a statistical point of view, Rodrik's procedure closely corresponds to the information-theoretic approach under the Akaikean paradigm (see Burnham & Anderson, 2002), which was the second-most popular MMA in our survey (25.0%, Table 3). This suggests that a pragmatic multi-paradigmatic methodological approach in econometrics towards the theory-data confrontation problem as a problem of model selection would be beneficial to economics, but this is obviously at odds with the PET approach to econometrics, which still dominates the way the subject is taught to students. Moreover, similar to Gigerenzer (2004) and Krämer (2011), Spanos (e.g. Spanos, 2007, 2011) has repeatedly raised concern about the recipe-like employment of frequentist tools and procedures in what he calls 'textbook econometrics' (Spanos, 2010b, p. 232). What is more, Spanos, drawing on ideas from Mayo (1996), has also argued that the frequentist approach to empirical modeling, when supplemented with post-data severity evaluations, was well suited to deal with both model selection and model uncertainty (Mayo & Spanos, 2011; Spanos, 2019, 2024). However, whether these still recent tools will establish as the new standard in econometric modeling is currently an open question.

'Textbook econometrics' (Spanos, 2010b, p. 232) adopts PET under the frequentist paradigm, albeit usually without acknowledging or discussing these stances explicitly. Even in the rare case of textbooks on alternative non-frequentist approaches to econometrics, the exclusive focus is on model estimation rather than selection (e.g. Judge & Mittelhammer, 2012). As a proxy, we surveyed some recent econometrics textbooks in a non-representative sample, focusing in particular at their treatment of the Akaikean paradigm, the second most popular statistical paradigm in our sample. Three books provide worked examples of model selection under the Akaikean paradigm (Baltagi, 2008; Hill et al., 2011; Stock & Watson, 2019). Greene (2003) and Dougherty (2011) at least mention AIC and BIC, but without providing more detail as to how these work in an example or what the differences to measures such as R^2 were. In fact, some of the most popular textbooks on econometrics do not mention Akaikean tools at all (e.g. Angrist & Pischke, 2009), and those that do tend to gloss over the subject rather briefly.[10] None of the books talked about differences between and implications of the different statistical paradigms. With the exception of Spanos (2019), we also did not find a textbook that discussed the issue of drawing inferences about real-world economic systems from observational data and statistical models. Spanos' textbook, while not explicitly mentioning the paradigms of Likelihoodism and Akaikeanism, is also the only one we are aware of that discusses at length the differences between the Neyman-Pearson and Fisher's approach to statistical inference. Lastly, while all textbooks discussed misspecification testing to some extent, Spanos (2019) pointed out most explicitly the importance of statistical adequacy for valid inference, i.e. 'the adequacy of the probabilistic assumptions comprising the postulated statistical model, in the light of the observed data (Spanos, 2019, p. 729).

4.2. Statistical paradigms as orientational paradigms?

In a recent article, Hoyningen-Huene and Kincaid (2023) introduced the notion of 'orientational paradigms'. Its main claim is that economics is set apart from other social sciences by its particularly high number of orientational paradigms, which enable scientific discourse across school boundaries by serving as a common reference point for the community. In what follows, we will argue that while statistical paradigms serve as orientational paradigms *in statistics*, they do *not all* serve as orientational paradigms *in economics*. We think that this is largely because of the way econometrics tends to be taught to economics students (see our previous discussion in Section 4.1).

What are orientational paradigms? Orientational paradigms (OPs) share some characteristics with Kuhnian paradigms (Kuhn, 1962). Importantly, OPs are shared across scientific communities, i.e. they are *known* to all members of the community. For economics, two such concepts would be Nash equilibrium and expected utility theory. The other two characteristics of OPs that are similar to Kuhnian paradigms are that they are historical entities ('they come, flourish, and may go', Hoyningen-Huene & Kincaid, 2023, p. 191) and that their range of applications may vary considerably, i.e. from general theories (like expected utility theory) to models applicable only to rather specific situations (for example, Akerlof's model of asymmetric information, Akerlof, 1970). There is, according to Hoyningen-Huene and Kincaid (2023, p. 191), a set of three additional defining characteristics of OPs that set them apart from paradigms in the Kuhnian sense.

(1) OPs are common points of orientation rather than fixed points of belief, i.e. OPs may provide guidelines for further research to *some* members of the community, but not to all, and some members may 'explicitly distance themselves from them' (Hoyningen-Huene & Kincaid, 2023, p. 191).
(2) OPs are openly and possibly severely criticized by parts of the pertinent scientific community, especially by those who develop alternatives to them, i.e. OPs are subject to publishable criticism as well as defense.
(3) OPs are not further articulated, developed, and applied to new problems by the whole community, i.e. *some* members of the community may indeed use OPs in these ways and create progress doing so, but this progress will be restricted to the respective schools that accept certain OPs.

Given this list of characteristics of OPs, there is no doubt that the statistical paradigms as discussed in Section 2.3 are indeed OPs in statistics. First, all professional statisticians[11] must know the four paradigms, if only for historical reasons. Bayesianism in particular is a good example of how a paradigm can go through the cycle of coming, flourishing, going and possibly even coming back (see Efron, 2005). Because every statistician knows the difference between the frequentist, the Bayesian and, for example, the Akaikean paradigm, these paradigms provide a common point of orientation. Clearly, every paradigm has been, and probably will be for the foreseeable future, subject to publishable criticism and defense across all camps (e.g. Akaike, 1982; Gandenberger, 2016; Howson & Urbach, 1989; Ioannidis, 2005; Spanos, 2010a; Wasserstein & Lazar, 2016), and it seems to have some relevance to professional statisticians whether they identify as, for example, a 'Bayesian' or a 'frequentist' (see Efron, 1986 and comments therein), if only to clarify the perspective taken on a problem in any given piece of statistical research. Thus, virtually all statistical research has a statement that signifies to the reader the paradigm followed. Examples include statements such as 'using an information-theoretic argument' (Spiegelhalter et al., 2002, p. 583), 'we consider the general problem of testing in the Neyman-Pearson framework' (Romano et al., 2010, p. 76) or 'we propose a misspecification-resistant information criterion' (Hsu et al., 2019, p. 1063). However, we maintain that, in contrast to statistics, the four statistical paradigms are *not all* OPs in economics. In particular, this means that *all four* statistical paradigms are generally *not* known to all economists. Therefore, they cannot serve as 'common points of orientation' (Hoyningen-Huene & Kincaid, 2023, p. 191) for the economics community. Only frequentism seems to take the role of an OP in economics.[12] Hoyningen-Huene and Kincaid (2023) argue that OPs are beneficial to a research field or discipline, because OPs increase a field's systematicity, a notion coming from systematicity theory, Hoyningen-Huene's own proposal for a general philosophy of science (Hoyningen-Huene, 2008, 2013). On this account, increased systematicity means scientific progress. Concretely, the argument is that in providing common reference points to all members of the pertinent scientific community, OPs increase 'epistemic connectedness', which constitutes one dimension of systematicity (Hoyningen-Huene & Kincaid, 2023, p. 197f):

> Scientists referring to OPs when explicating alternatives to them or working under their reign by developing new applications of them, generate or utilize epistemic connections among various parts of science. (Hoyningen-Huene & Kincaid, 2023, p. 198)

The claim that only frequentism may be seen as an OP in economics opens up the path to another interpretation of the results of our survey of empirical T5 publications (Section 3). If only the

Table 2. Descriptive statistics of our systematic literature review.

Journal	Papers published 2000–18	Relevant papers	Single- model/pre-eminence of theory	Multi-model
American Economic Review	1,826	936 (51.3%)	893 (95.4%)	43 (4.6%)
Econometrica	1,068	208 (19.5%)	191 (92.4%)	17 (8.1%)
Journal of Political Economy	736	383 (52.0%)	352 (91.9%)	31 (8.1%)
Quarterly Journal of Economics	680	467 (68.7%)	453 (97.0%)	14 (3.0%)
Review of Economic Studies	851	264 (31.0%)	233 (89.8%)	31 (11.7%)
Total	**5,161**	**2,258 (43.8%)**	**2,122 (94.0%)**	**136 (6.0%)**

Table 3. Statistical paradigms in multi-model econometric papers.

Journal	Multi-model papers	Frequentism	Bayesianism	Likelihoodism	Akaikeanism
American Economic Review	43	27 (62.8%)	3 (7.0%)	3 (7.0%)	10 (23.3%)
Econometrica	17	11 (64.7%)	0 (0.0%)	2 (11.8%)	4 (23.5%)
Journal of Political Economy	31	20 (64.5%)	0 (0.0%)	2 (6.5%)	9 (29.0%)
Quarterly Journal of Economics	14	9 (64.3%)	0 (0.0%)	0 (0.0%)	5 (35.7%)
Review of Economic Studies	31	21 (67.7%)	3 (9.7%)	1 (3.2%)	6 (19.4%)
Total	**136**	**88 (64.7%)**	**6 (4.4%)**	**8 (5.9%)**	**34 (25.0%)**

frequentist paradigm, along with its perspective on evidence and model uncertainty, is an OP in economics, adopting this paradigm carries the least methodological risk, but it also limits one's methodological options for model selection to variants of significance testing (see Table 1). There is a sizeable literature about frequentist deficiencies in model selection and its potential alternatives, both in statistics and econometrics (e.g. Burnham & Anderson, 2002; Freedman, 1983; Granger et al., 1995; Kass & Raftery, 1995; Lukacs et al., 2010). Yet, as demonstrated in Section 3, even in the rare cases that practice MMA, the vast majority (64.7%) follows the frequentist paradigm (see Tables 2 and 3), precisely because frequentism is an OP in economics, while the other major statistical paradigms are not, or not yet.[13]

5. Conclusions

Our paper explored the connection between four major inferential paradigms in statistical science and inferential practice in modern-day econometrics. Econometrics is still largely characterized by John Stuart Mill's conception of statistical inference, who saw a distinction between 'theorists' and 'practical men'. In spite of Rodrik's (2015) much-debated notion of economics that sees issues of model selection at the forefront, the 'theory first' / 'pre-eminence of theory' approach vastly dominates empirical research published in economics' T5 journals. When model selection and model uncertainty is accounted for, this largely happens under the frequentist paradigm, in spite of frequentism's limited model-selection capabilities that could be cured with tools from the other statistical paradigms. This finding may be explained by frequentism's special role as an orientational paradigm in economics.

Notes

1. For a helpful categorization and discussion of scientific experiments and evidence, see Caniglia et al. (2017).
2. Importantly, model uncertainty is different from mere parameter uncertainty, which is the uncertainty about the true value of a parameter in a model. Model uncertainty refers to the uncertainty inherent to model selection, which is either the uncertainty about which model from a set of candidates S actually generated the data or the uncertainty about which model from S best approximates the data-generating process when the latter is unknown or not included in S. Hsu et al. (2019) refer to these two situations as model-selection problems of category I and II, respectively.
3. Different authors have expressed various views as to what constitutes the major schools of thinking in statistics in general and statistical inference in particular. For example, Howson and Urbach (1989) distinguish between Bayesian and the two 'classical' theories of statistical inference by Fisher and by Neyman and Pearson, whereas Young and Smith (2005) and Cox (2006) refer to Bayesian, Fisherian and frequentist schools. Lillestol (2014)

draws lines between Early- and Neo-Bayesianism, Fisherianism, the Neyman-Pearson paradigm and Likelihoodism while Gandenberger (2016) talks about Bayesianism, Frequentism and Likelihoodism. Here, we follow the categorization by Bandyopadhyay and Forster (2011). We do so not because we maintain that it is necessarily the best, but the only one to our knowledge that explicitly refers to all four schools of thinking discussed here as statistical paradigms.

4. Note that this general understanding includes a wide variety of settings, including, for example, forward/backward specification searching, imposing restrictions to a general unrestricted model based on economic theory or testing competing theories such as, for example, expected utility theory and rank-dependent utility against choice data. The distinctive feature is that there is a set of models deemed plausible which is compared in terms of model fit based on statistical and economic criteria.
5. Birnbaum (1962) provided a proof of the Likelihood Principle, which Kalbfleisch (1975) and later Mayo (2014) attempted to refute. However, all refutations have been a source of contempt. Berger and Wolpert (1988) argued that Kalbfleisch's refutation only worked for restricted settings and not against Birnbaum's theorem in general. Mayo's attempt was subject to immediate criticism (e.g. Bjørnstad [2014]; Dawid [2014]), before Gandenberger (2015) provided an alternative proof of the Likelihood Principle and Peña and Berger (2017) refuted Mayo (2014)'s argument directly.
6. In 2013, mathematician Christian Hennig, referring to the arguments made much earlier in Evans et al. (1986) has concluded 'frequentists/error statisticians therefore don't need to worry about the likelihood principle because they shouldn't accept the sufficiency principle in the generality that is required for Birnbaum's proof' (see https://errorstatistics.com/2013/02/10/u-phil-gandenberger-hennig-birnbaums-proof/).
7. The nestedness requirement can be relaxed [see Vuong (1989)].
8. The complete search string was as follows: (SRCTITLE ('Econometrica') OR SRCTITLE ('American Economic Review') OR SRCTITLE ('Journal of Political Economy') OR SRCTITLE ('Review of economic studies') OR SRCTITLE ('quarterly journal of economics')) AND (LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010) OR LIMIT-TO (PUBYEAR, 2009) OR LIMIT-TO (PUBYEAR, 2008) OR LIMIT-TO (PUBYEAR, 2007) OR LIMIT-TO (PUBYEAR, 2006) OR LIMIT-TO (PUBYEAR, 2005) OR LIMIT-TO (PUBYEAR, 2004) OR LIMIT-TO (PUBYEAR, 2003) OR LIMIT-TO (PUBYEAR, 2002) OR LIMIT-TO (PUBYEAR, 2001) OR LIMIT-TO (PUBYEAR, 2000)) AND (LIMIT-TO (EXACTSRCTITLE, 'American Economic Review') OR LIMIT-TO (EXACTSRCTITLE, 'Econometrica') OR LIMIT-TO (EXACTSRCTITLE, 'Review Of Economic Studies') OR LIMIT-TO (EXACTSRCTITLE, 'Journal Of Political Economy') OR LIMIT-TO (EXACTSRCTITLE, 'Quarterly Journal Of Economics')) AND (LIMIT-TO (DOCTYPE, 'ar')).
9. When double-checking the search string again on August 25, 2021, we found $N = 5167$ hits, with additional papers now listed as published in *American Economic Review* (+5) and *Econometrica* (+1).
10. For example, Greene's (2003) treatment of model selection criteria covers one page in a book of 828 pages and the exposition in Hill et al. (2011) takes about two out of 790 pages including an exercise that deals with AIC. Stock and Watson (2019) mention the topic on five out of 836 pages of their Introduction to Econometrics. Dougherty (2011, p. 500) devotes half a page out of 537 to mention AIC and BIC in the context of choosing optimal lags in non-stationary time series analysis.
11. Here, we refer to statisticians working in research, i.e. we assume at least a Master's degree in a statistical major.
12. For those readers wondering about Bayesian statistics in economics, this is summed up nicely by Baltagi (2008, p. vii): 'I did not attempt the coverage of Bayesian econometrics simply because it is not my comparative advantage. The reader should consult Koop (2003) for a more recent treatment of the subject'. Thus, *some* economists will know about Bayesian econometrics, but it certainly would take it too far to assume that *all* economists are familiar with it.
13. As one reviewer pointed out, the fact that Bayesianism in particular is not a more popular paradigm in economics and applied econometrics may also be explained by the lack of guidance from economic theory with respect to priors, which would be especially problematic in a discipline dominated by the PET approach. While this is a fair point, it does not explain the relative unpopularity of likelihoodism and Akaikeanism.

Acknowledgments

The authors would like to thank Chan Le and Linda von Heydebreck for their invaluable support in retrieving and organizing thousands of PDFs throughout this project. Michael Heitmann's assistance and doublechecks helped a great deal with Figure 3 and Table 3.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Financial support from Volkswagen Foundation [grant number VWZN3188] is gratefully acknowledged.

Data availability statement

Data is available from the authors upon reasonable request.

Notes on contributors

John-Oliver Engler Studies in physics (Diplom, Konstanz University, Germany) and economics (PhD, Leuphana University, Germany). Since 2022, he is a professor of bioeconomy and resource efficiency at University of Vechta (Germany). His research interests are ecological and transformational economics, in particular the role of risk and uncertainty for sustainability.

Julius J. Beeck Studies in Data Science and Management (MSc) and Business Administration (BA, Leuphana University). Julius's currently works as an investment advisor.

Henrik von Wehrden Studies in geography (Diplom, Marburg University, Germany) and biology (PhD, Halle University, Germany). Since 2016, Henrik's worked as full professor of normativity of methods at Leuphana University. His research interests are methods in sustainability science and on the intersection of ethics and statistics.

ORCID

John-Oliver Engler http://orcid.org/0000-0002-8590-3748

References

Abadie, A. (2020). Statistical nonsignificance in empirical economics. *American Economic Review: Insights*, *2*(2), 193–208. https://doi.org/10.1257/aeri.20190252

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*(3), 631–636. https://doi.org/10.1890/13-1452.1

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Akaike, H. (1982). On the fallacy of the likelihood principle. *Statistics & Probability Letters*, *1*(2), 75–78. https://doi.org/10.1016/0167-7152(82)90019-0

Akerlof, G. A. (1970). The market for "Lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Allais, M. (1991). An outline of my main contributions to economic science. *Theory and Decision*, *30*(1), 1–26. https://doi.org/10.1007/BF00134634

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Baltagi, B. (2008). *Econometrics* (4th ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76516-5

Bandyopadhyay, P. S., & Forster, M. (Eds.). (2011). *Philosophy of statistics: Volume 7 in handbook of the philosophy of science*. Elesvier.

Berger, J. O., & Wolpert, R. L. (1988). *The likelihood principle* (2nd ed.). Lecture Notes – Monograph Series: Vol. 6. Institute of Mathematical Statistics.

Birnbaum, A. (1962). On the foundations of statistical inference. *Journal of the American Statistical Association*, *57*(298), 269–306. https://doi.org/10.1080/01621459.1962.10480660

Bjørnstad, J. F. (1996). On the generalization of the likelihood function and the likelihood principle. *Journal of the American Statistical Association*, *91*(434), 791–806.

Bjørnstad, J. F. (2014). Discussion of "on the Birnbaum argument for the strong likelihood principle". *Statistical Science*, *29*(2), 2. https://doi.org/10.1214/14-STS475

Bornmann, L., Butz, A., & Wohlrabe, K. (2018). What are the top five journals in economics? A new meta-ranking. *Applied Economics*, *50*(6), 659–675. https://doi.org/10.1080/00036846.2017.1332753

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Burnham, K. P., & Anderson, D. R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, *28*(2), 111. https://doi.org/10.1071/WR99107
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
Burnham, K. P., & Anderson, D. R. (2014). *P* values are only an index to evidence: 20th- vs. 21st-century statistical science. *Ecology*, *95*(3), 627–630. https://doi.org/10.1890/13-1066.1
Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. National Bureau of Economic Research.
Caniglia, G., Schäpke, N., Lang, D. J., Abson, D. J., Luederitz, C., Wiek, A., Laubichler, M. D., Gralla, F., & von Wehrden, H. (2017). Experiments and evidence in sustainability science: A typology. *Journal of Cleaner Production*, *169*, 39–47. https://doi.org/10.1016/j.jclepro.2017.05.164
Clauset, A., Shalizi, C. R., & Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. https://doi.org/10.1137/070710111
Cox, D. R. (2006). *Principles of statistical inference*. Cambridge University Press.
Dawid, A. P. (2014). Discussion of "On the Birnbaum argument for the strong likelihood principle". *Statistical Science*, *29*(2), 2. https://doi.org/10.1214/14-STS470
Do, H.-P., & Spanos, A. (2024). Revisiting the Phillips curve: The empirical relationship yet to be validated. *Oxford Bulletin of Economics and Statistics*, *86*(4), 761–793. https://doi.org/10.1111/obes.12605
Dougherty, C. (2011). *Introduction to econometrics* (4th ed.). Oxford University Press.
Duflo, E. (2020). Field experiments and the practice of policy. *American Economic Review*, *110*(7), 1952–1973. https://doi.org/10.1257/aer.110.7.1952
Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician*, *40*(1), 1–5. https://doi.org/10.1080/00031305.1986.10475342
Efron, B. (2005). Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, *100*(2), 1–5. https://doi.org/10.1198/016214505000000033
Engler, J.-O., & Baumgärtner, S. (2015). Model choice and size distribution: A Bayequentist approach. *American Journal of Agricultural Economics*, *97*(3), 978–997. https://doi.org/10.1093/ajae/aau034
Evans, M., Fraser, D. A., & Monette, G. (1986). On principles and arguments to likelihood. *Canadian Journal of Statistics*, *14*(3), 181–194. https://doi.org/10.2307/3314794
Feynman, R. P. (1967). *The character of physical law* (12th ed.). The M.I.T. Press.
Freedman, D. A. (1983). A note on screening regression equations. *The American Statistician*, *37*(2), 152–155. https://doi.org/10.1080/00031305.1983.10482729
Gandenberger, G. (2015). A new proof of the likelihood principle. *The British Journal for the Philosophy of Science*, *66*(3), 475–503. https://doi.org/10.1093/bjps/axt039
Gandenberger, G. (2016). Why I am not a likelihoodist. *Philosophers' Imprint*, *16*(7), 1–22.
Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033
Gilbert, C. L. (1989). LSE and the British approach to time series econometrics. *Oxford Economic Papers*, *41*(1), 108–128. https://doi.org/10.1093/oxfordjournals.oep.a041887
Grace, J. (2006). *Structural equation modeling and natural systems*. Cambridge University Press.
Granger, C. W., King, M. L., & White, H. (1995). Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics*, *67*(1), 173–187. https://doi.org/10.1016/0304-4076(94)01632-A
Greene, W. H. (2003). *Econometric analysis* (5th ed.). Prentice Hall.
Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*, *24*(4), 699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x
Grüne-Yanoff, T., & Marchionni, C. (2018). Modeling model selection in model pluralism. *Journal of Economic Methodology*, *25*(3), 265–275. https://doi.org/10.1080/1350178X.2018.1488572
Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*, iii-vi+1–115. https://doi.org/10.2307/1906935
Hacking, I. (1965). *Logic of statistical inference*. Cambridge University Press.
Heckman, J. J., & Moktan, S. (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, *58*(2), 419–470. https://doi.org/10.1257/jel.20191574
Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T. C. Mills, & K. D. Patterson (Eds.), *Palgrave handbook of econometrics* (pp. 3–67). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244405_1
Hendry, D. F. (2011). Empirical economic model discovery and theory evaluation. *RMM Journal*, *2*, 115–145.
Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). *Principles of econometrics* (4th ed.). Wiley.
Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. Open Court.
Hoyningen-Huene, P. (2008). Systematicity: The nature of science. *Philosophia*, *36*(2), 167–180. https://doi.org/10.1007/s11406-007-9100-x
Hoyningen-Huene, P. (2013). *Systematicity: The nature of science*. Oxford University Press.

Hoyningen-Huene, P., & Kincaid, H. (2023). What makes economics special: Orientational paradigms. *Journal of Economic Methodology*, *30*(2), 188–202. https://doi.org/10.1080/1350178X.2023.2192231
Hsu, H.-L., Ing, C.-K., & Tong, H. (2019). On model selection from a finite family of possibly misspecified time series models. *The Annals of Statistics*, *47*, 2. https://doi.org/10.1214/18-AOS1706
Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124
Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, *31*(2), 203–222. https://doi.org/10.1017/S030500410001330X
Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*(2), 101–108. https://doi.org/10.1016/j.tree.2003.10.013
Judge, G. C., & Mittelhammer, R. C. (2012). *An information theoretic approach to econometrics*. Cambridge University Press.
Kadane, J. B., & Lazar, N. A. (2004). Methods and criteria for model selection. *Journal of the American Statistical Association*, *99*(465), 279–290. https://doi.org/10.1198/016214504000000269
Kalaitzidakis, P., Mamuneas, T. P., & Stengos, T. (2011). An updated ranking of academic journals in economics. *Canadian Journal of Economics/Revue canadienne d'économique*, *44*(4), 1525–1538. https://doi.org/10.1111/j.1540-5982.2011.01683.x
Kalbfleisch, J. D. (1975). Sufficiency and conditionality. *Biometrika*, *62*(2), 251–259. https://doi.org/10.1093/biomet/62.2.251
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572
Koop, G. (2003). *Bayesian econometrics*. Wiley.
Koopmans, T. C. (1947). Measurement without theory. *The Review of Economics and Statistics*, *29*(3), 161–172. https://doi.org/10.2307/1928627
Krämer, W. (2011). The cult of statistical significance – What economists should and should not do to make their data talk. *Journal of Contextual Economics – Schmollers Jahrbuch*, *131*(3), 455–468. https://doi.org/10.3790/schm.131.3.455
Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.
Kydland, F. E., & Prescott, E. C. (1991). The econometrics of the general equilibrium approach to business cycles. *The Scandinavian Journal of Economics*, *93*(2), 161–178. https://doi.org/10.2307/3440324
Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review: Insights*, *73*(1), 31–43.
Leek, J., McShane, B. B., Gelman, A., Colquhoun, D., Nuijten, M. B., & Goodman, S. N. (2017). Five ways to fix statistics. *Nature*, *551*(7682), 557–559. https://doi.org/10.1038/d41586-017-07522-z
Lillestol, J. (2014). *Statistical inference: Paradigms and controversies in historic perspective*. NHH. https://www.nhh.no/globalassets/departments/business-and-management-science/research/lillestol/statistical_inference.pdf.
Lindley, D. V., & Phillips, L. D. (1976). Inference for a Bernoulli process (A Bayesian view). *The American Statistician*, *30*(3), 112–119. https://doi.org/10.1080/00031305.1976.10479154
Lukacs, P. M., Burnham, K. P., & Anderson, D. R. (2010). Model selection bias and Freedman's paradox. *Annals of the Institute of Statistical Mathematics*, *62*(1), 117–125. https://doi.org/10.1007/s10463-009-0234-4
Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. University of Chicago Press.
Mayo, D. G. (2014). On the birnbaum argument for the strong likelihood principle. *Statistical Science*, *29*, 2. https://doi.org/10.1214/13-STS457
Mayo, D. G., & Spanos, A. (2010). *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science*. Cambridge University Press.
Mayo, D. G., & Spanos, A. (2011). Error statistics. In P. S. Bandyopadhyay, & M. Forster (Eds.), *Philosophy of statistics: Volume 7 in handbook of the philosophy of science* (pp. 149–198). Elesvier.
McCloskey, D. N., & Ziliak, S. T. (1996). The standard errors of regression. *Journal of Economic Literature*, *34*(1), 97–114.
McGuirk, A., Robertson, J., & Spanos, A. (1993). Modeling exchange rate dynamics: Non-linear dependence and thick tails. *Econometric Reviews*, *12*(1), 33–63. https://doi.org/10.1080/07474939308800253
McShane, B. B., & Gal, D. (2017). Statistical significance and the dichotomization of evidence. *Journal of the American Statistical Association*, *112*(519), 885–895. https://doi.org/10.1080/01621459.2017.1289846
McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253
Mill, J. S. (1844). On the definition of political economy and on the method of investigation proper to it. In J. S. Mill (Ed.), *Essays on some unsettled questions of political economy* (pp. 120–162). Parker.
Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6–18. https://doi.org/10.1016/j.jmp.2015.11.001
Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*(506), 150–152. https://doi.org/10.1038/506150a
Peña, V., & Berger, J. O. (2017). *A note on recent criticisms to Birnbaum's theorem*. Cornell University. https://arxiv.org/pdf/1711.08093.
Reiss, J. (2013). *Philosophy of economics*. Routledge.
Rodrik, D. (2015). *Economics rules: The rights and wrongs of the dismal science*. W. W. Norton & Company.

Romano, J. P., Shaikh, A. M., & Wolf, M. (2010). Hypothesis testing in econometrics. *Annual Review of Economics*, *2*(1), 75–104. https://doi.org/10.1146/annurev.economics.102308.124342
Rommel, J., & Weltin, M. (2021). Is there a cult of statistical significance in agricultural economics? *Applied Economic Perspectives and Policy*, *43*(3), 1176–1191. https://doi.org/10.1002/aepp.13050
Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. Chapman & Hall.
Spanos, A. (1989). Early empirical findings on the consumption function, stylized facts or fiction: A retrospective view. *Oxford Economic Papers*, *41*(1), 150–169. https://doi.org/10.1093/oxfordjournals.oep.a041890
Spanos, A. (1995). On theory testing in econometrics: Modeling with nonexperimental data. *Journal of Econometrics*, *67*(1), 189–226. https://doi.org/10.1016/0304-4076(94)01633-B
Spanos, A. (2007). Curve fitting, the reliability of inductive inference, and the error-statistical approach. *Philosophy of Science*, *74*(5), 1046–1066. https://doi.org/10.1086/525643
Spanos, A. (2009). The Pre-eminence of theory versus the European CVAR perspective in macroeconometric modeling. *Economics*, *3*(1), 1–16. https://doi.org/10.5018/economics-ejournal.ja.2009-10
Spanos, A. (2010a). Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification. *Journal of Econometrics*, *158*(2), 204–220. https://doi.org/10.1016/j.jeconom.2010.01.011
Spanos, A. (2010b). Theory testing in economics and the error-statistical perspective. In D. G. Mayo, & A. Spanos (Eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* (pp. 202–246). Cambridge University Press.
Spanos, A. (2011). Foundational issues in statistical modeling: Statistical model specification and validation. *Rationality, Markets and Morals*, *2*, 146–178.
Spanos, A. (2019). *Probability theory and statistical inference: Econometric modeling with observational data*. Cambridge University Press.
Spanos, A. (2024). How the post-data severity converts testing results into evidence for or against pertinent inferential claims. *Entropy*, *26*(1), 95. https://doi.org/10.3390/e26010095
Spanos, A., & Taylor, M. (1984). The monetary approach to the balance of payments: A critical appraisal of some empirical evidence. *Oxford Bulletin of Economics and Statistics*, *46*(4), 329–340. https://doi.org/10.1111/j.1468-0084.1984.mp46004004.x
Spiegelhalter, D. J., Best, N. G., Carlin, B. B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353
Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes factor design analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/10.3758/s13428-018-01189-8
Stock, J. H., & Watson, M. W. (2019). *Introduction to Econometrics* (4th ed.). The Pearson series in economics. Pearson.
van Witteloostuijn, A., & van Hugten, J. (2022). The state of the art of hypothesis testing in the social sciences. *Social Sciences & Humanities Open*, *6*(1), 100314. https://doi.org/10.1016/j.ssaho.2022.100314
Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*(2), 307–333. https://doi.org/10.2307/1912557
Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p -values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108
Young, G. A., & Smith, R. L. (2005). *Essentials of statistical inference*. Cambridge University Press.
Ziliak, S. T., & McCloskey, D. N. (2004). Size matters: The standard error of regressions in the American economic review. *The Journal of Socio-Economics*, *33*(5), 527–546. https://doi.org/10.1016/j.socec.2004.09.024