The idea that linear models suffice, in comparison to nonlinear ones, for modeling time series associated with business cycles, is welcomed, strictly from the well-known finding in time series analysis that simple models tend to outperform their more elaborate counterparts in genuine forecasting settings (as opposed to forecast backtesting exercises, or, worse, model selection based on use of Fisherian p-value assessment of parameters; see below). In particular, as forcefully and elegantly argued in Zellner (2001) in a general econometric modeling context, it is worthwhile to have an ordering of possible models in terms of complexity, with higher probabilities assigned to simpler models. See also Keuzenkamp and McAleer (1997, p. 554). This agrees with the general findings of Makridakis and Hibon (2000, p. 458), who state that "statistically sophisticated or complex models do not necessarily produce more accurate forecasts than simpler ones".

Notice that the above support of simple models is tied to forecasting. If the authors are instead interested in answering the question "Are there non-linearities in time series associated with the business cycle?", then they are wasting their time: The answer is, of course, yes. Undoubtedly, the underlying data generating process (DGP) is extraordinary complicated and surely nonlinear, especially as the number of observations increases, either by increasing the frequency of the data, or by increasing the available calendar time. This is not in conflict with the use of a linear model: The purpose of statistical modeling is to develop a reasonably parsimonious model that is able to capture as much of the signal as possible, while avoiding the noise. This implies that the amount of available data will influence the choice of model. Appealing again to some quotes, this is well-stated by Burnham and Anderson (2003, p. 143): "The purpose of the analysis of empirical data is not to find the 'true model'—not at all. Instead, we wish to find a best approximating model, based on the data, and then develop statistical inferences from this model. Data analysis involves the question, 'What level of model complexity will the data support?' and both under- and over-fitting are to be avoided."

As such, the question of interest is not whether there are nonlinearities in a macroeconomic time series, but rather, given the amount of data available, what model (or mixture of models, as produced, for example, by a Bayesian model averaging paradigm or a frequentist forecast averaging exercise) results in the best forecasts. This can be determined via cross-validation, and/or out-of-sample point or density forecasting. Indeed, Yang (2005, p. 937) reminds us that "A traditional approach to statistical inference is to identify the true or best model first with little or no consideration of the specific goal of inference in the model identification stage." If point forecasts are of interest, then one should keep in mind the non-trivial role played by the measure used; see, e.g., Gneiting (2011).

As a final issue worth discussing, and as alluded to above, the use of significance testing (Fisherian p-values) and/or (Neyman-Pearson) hypothesis testing should play virtually no role in model selection. The authors rely on the use of statistical tests, but this is highly questionable. As discussed in detail in Paolella (2017, Sec. 2.8), the original use of p-values has little in common with modern usage in model selection, and are not appropriate. This is no longer a fringe viewpoint: Several journals in, notably, psychology have literally *banned* the reporting of p-values, in recognition that they offer nothing in the way of scientific progress. Further appealing to quotations, Briggs (2016, p. xiii) believes that "Hypothesis testing should immediately and forever be tossed onto the scrap heap of intellectual history and certainly never taught to the vulnerable." While this might seem like a recent conjecture, the idea goes back quite a while: As the (recently deceased) Lindley (1968, p. 321) stated, "The frequency theory of probability and statistics is the most misleading and irrelevant idea that has ever clouded our subject and ought to be forgotten." Much later, Lindley (1999, p. 75) stated "My personal view is that p-values should be relegated to the scrap heap and not considered by those who wish to think and act coherently." Arguably less eloquently, albeit more comically, Briggs (2016, p. 178) makes the emphatic statement "Die, p-value, Die Die Die."

Concrete examples in econometrics abound: For example, Nakamura and Nakamura (1978) investigated the pretest estimator of β_2 in the regression model with time trend, when the choice of model is determined by the outcome of the Durbin-Watson test, for a given significance level α. Nakamura and Nakamura (1978, p. 207) conclude: "Our results so far suggest that tests of significance for autocorrelation might best be dispensed with in estimating [regression relationships] in favor of a practice of always transforming."[1] Their results were corroborated by Fomby and Guilkey (1978), who showed that, if a pretest estimator for $\boldsymbol{\beta}$ is used based on the Durbin-Watson statistic, then the optimal significance level α *is far greater than 0.05*, and more like 0.50, when measuring the performance of $\widehat{\boldsymbol{\beta}}$ based on MSE. Similar findings regarding the inappropriateness of the traditional significance levels have been recently shown to be the case in the unit root testing framework; see Kim and Choi (2017).

References

Briggs, W. (2016). *Uncertainty: The Soul of Modeling, Probability & Statistics*. Springer Verlag, Switzerland.

Burnham, K. P. and Anderson, D. (2003). *Model Selection and Multi-Model Inference*. Springer Verlag, New York, second edition.

Fomby, T. B. and Guilkey, D. K. (1978). On Choosing the Optimal Level of Significance for the Durbin–Watson Test and the Bayesian Alternative. *Journal of Econometrics*, 8:203–213.

[1]Here, "transform" refers to estimating the regression model with AR(1) disturbance term, which, at the time, was not so trivial. They used the Cochrane-Orcutt method of estimation, which is not the same as, and inferior to, the MLE.

Gneiting, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Keuzenkamp, H. A. and McAleer, M. (1997). The Complexity of Simplicity. In *11th Biennial Conference on Modelling and Simulation*, pages 553–561.

Kim, J. H. and Choi, I. (2017). Unit Roots in Economic and Financial Time Series: Re–Evaluation at the Optimal Level of Significance. *Forthcoming in: Econometrics*, 1(1):1–1. Article 1.

Lindley, D. V. (1968). Discussion of Nelder (1968). *Journal of the Royal Statistical Society, Series A*, 131(3):320–321.

Lindley, D. V. (1999). Comment on Bayarri and Berger. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics*, volume 6, page 75. Oxford University Press, Oxford.

Makridakis, S. and Hibon, M. (2000). The M3 Competition: Results, Conclusions and Implications. *International Journal of Forecasting*, 17:567–570.

Nakamura, A. and Nakamura, M. (1978). On the Impact of the Tests for Serial Correlation Upon the Test of Significance for the Regression Coefficient. *Journal of Econometrics*, 7:199–210.

Paolella, M. S. (2017). *Fundamental Statistical Inference: A Computational Approach*. John Wiley & Sons, Chichester. Forthcoming.

Yang, Y. (2005). Can the Strengths of AIC and BIC be Shared? A Conflict Between Model Indentification and Regression Estimation. *Biometrika*, 92(4):937–950.

Zellner, A. (2001). Keep it Sophisticatedly Simple. In Zellner, A., Keuzenkamp, H. A., and McAleer, M., editors, *Simplicity, Inference and Modelling*, pages 242–262. Cambridge University Press, Cambridge.