NBER WORKING PAPER SERIES

CREDIBLE NUMBERS: A PROCEDURE FOR REPORTING STATISTICAL PRECISION IN PARAMETER ESTIMATES

Nicolas Astier
Frank A. Wolak

Working Paper 32124
http://www.nber.org/papers/w32124

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2024

We are grateful to seminar participants at the Paris School of Economics for their feedback and suggestions on an earlier draft. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Nicolas Astier and Frank A. Wolak. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Credible Numbers: A Procedure for Reporting Statistical Precision in Parameter Estimates
Nicolas Astier and Frank A. Wolak
NBER Working Paper No. 32124
February 2024
JEL No. C01,C10,C12

ABSTRACT

Econometric software packages typically report a fixed number of decimal digits for coefficient estimates and their associated standard errors. This practice misses the opportunity to use rounding rules that convey statistical precision. Using insights from the testing statistical hypotheses of equivalence literature, we propose a methodology that only reports decimal digits in a parameter estimate that reject a hypothesis of statistical equivalence. Applying this methodology to all articles published in the American Economic Review between 2000 and 2022, we find that over 60% of the printed digits in coefficient estimates do not convey statistically meaningful information according to our definition of a significant digit. If one additional digit beyond the last significant digit is reported for each coefficient estimate, then approximately one-third of the printed digits in our sample would not be reported.

Nicolas Astier
48 Bd Jourdan
Paris School of Economics
Ecole des Ponts
PARIS 75014
France
nicolas.astier@psemail.eu

Frank A. Wolak
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
wolak@zia.stanford.edu

1 Introduction

Software packages such as R, SAS/SPSS, and Stata can report parameter estimates up to an arbitrary number of decimal digits. A well-known reason to report few decimal digits, typically three or four, is the numerical reliability of the underlying computations. The same econometric estimator applied to the same data can produce different exact values for coefficient and standard error estimates depending on the software package or the computer used to perform the calculations. Tables 1 and 2 of McCullough and Vinod (1999) provide examples of this phenomenon.

Improvements in computational power and accuracy have significantly reduced the likelihood that the same estimation routine applied to the same data on different computers (or on the same computer using different software packages) produces different estimates up to at least three decimal digits.[1] However, the question of whether all the decimal digits reported, although numerically accurate, are statistically meaningful remains.

For example, a reported demand price elasticity estimate of -1.136 with a standard error estimate of 0.342 implies much less statistical precision and economic content than a demand price elasticity estimate of -0.951 with a standard error estimate of 0.021. In the first case, there is support for the conclusion of a downward sloping demand curve in price, but there is insufficient statistical precision to conclude that the demand for this product is price elastic or inelastic.[2] In the second case, the parameter estimate and associated standard error support both conclusions: (i) that the demand curve is downward sloping in price and (ii) that the demand for the product is price inelastic. However, the current dominant practice in empirical work in economics is to report three decimal digits for both coefficient estimates.

We believe that reporting a fixed number of decimal digits for econometric estimates represents a missed opportunity to increase the statistical and economic content of tables of parameter estimates. For instance, in the above example, we believe that it makes more sense to report fewer decimal digits for the coefficient estimate and its standard error in the

[1] The requirement that authors produce replication packages containing the raw data and software programs used to produce all empirical results reported in published papers in economics journals has helped achieve this outcome.

[2] Greater or less than one in absolute value, respectively.

first case than in the second one, both to convey the appropriate level precision in each price elasticity estimate and to avoid cluttering tables of parameter estimates with statistically meaningless digits.

How can a researcher decide how many digits to report in coefficient estimates and standard error estimates? Would such a reporting rule be easy to implement? Does current practice significantly depart from this reporting rule? Building on insights from the *testing statistical hypotheses of equivalence* literature,[3] we propose a procedure to determine which decimal digits to report in empirical work. We first derive exact closed-form expressions for the number of digits to report for a coefficient estimate for the normal linear conditional mean model.[4] This approach is asymptotically valid for a range of more general econometric models. Our procedure only involves straightforward calculations from variables that are already computed by standard software packages. It could therefore easily become the default setting of post-processing routines that researchers typically use to generate tables of estimation results.

To investigate whether our procedure would make a substantial difference in the way empirical results are reported in economics, we retrieve all coefficient estimates and associated standard errors published in the *American Economic Review* between 2000 and 2022 to obtain a dataset of over 90,000 coefficient estimates each with its associated standard error estimate. We find that 81% of coefficient estimates are reported with either two (23%) or three (58%) decimal digits. We then show that over 60% of the decimal digits of the coefficient estimates would not be printed according to this criteria. If the procedure is relaxed to also report the first non-statistically-meaningful digit, then over 33% of the published digits would not be printed according to this criteria. We conclude that adopting either the restricted or relaxed version of our procedure would imply non-trivial changes in how estimation results are reported. Tables of coefficient estimates would generally become more reader-friendly by conveying precision in the number of reported decimal digits. The risk of subsequent misuses of parameter estimates in applications that are sensitive to their reported precision

[3]Wellek (2010) provides an accessible book-length treatment of the theory of testing statistical hypotheses of equivalence.

[4]We also derive tight lower bound for the number of digits to report for the standard error estimate associated with a coefficient estimate from the normal linear conditional mean model.

would likely decrease.

Our proposal is inspired by current practice in other scientific fields. In particular, researchers in natural sciences pay particular attention to "significant digits" (also called "significant figures") when reporting empirical results.[5] Specifically, scientists typically report decimal digits only up to the point where measurement errors prevent further confidence in the precision of the number recorded. The rationale for this practice is that errors are inevitable in measuring a magnitude of interest. For example, the metering device may be impacted by small changes in background conditions. Therefore, measurements in natural sciences only report the digits whose order of magnitude is higher than the likely magnitude of possible idiosyncrasies in the measurement process.[6]

We argue that empirical research in economics could use an analogous concept of "significant digits," based on insights from the *testing statistical hypotheses of equivalence* literature. If the econometric model and associated dataset is thought as the measurement device employed by a researcher to obtain an estimate of an economic magnitude of interest, such as the above demand price elasticity, then our procedure can be seen as a way to assess the precision of this measurement device. The empirical estimate of this economic magnitude would then only be reported at the number of "significant digits" justified by our assessment of this precision.

Different measurement devices—in our case, datasets and econometric models—could imply reporting different numbers of significant digits. For example, applying our procedure to the first case above could suggest reporting -1 as the estimated demand price elasticity. In the second case, our procedure could instead imply reporting -0.95 as the estimated demand price elasticity.

Our approach has several advantages, particularly for reporting estimates of economic magnitudes. First, it would direct attention to the precision of a parameter estimate, which can often have important economic implications. In the case of a demand price elasticity, a negative value implies that a price increase reduces the quantity demanded, but a value of

[5]See for example Britannica: `https://www.britannica.com/science/significant-figures`(last accessed December 17, 2023).

[6]An example of this phenomenon is the estimated distance between the orbits of planets in the solar system which are typically reported with only a few decimal digits for distances of tens to hundreds of millions of miles.

-0.95 implies that a price increase also increases total revenue, an economic conclusion not supported by a value of -1. Second, our procedure would generally greatly simplify tables of estimation results, and thus make it easier to assess their economic content. Third, it would typically decrease the sensitivity of the values reported for parameter estimates and standard errors estimates to uncertainty in the precision of the underlying data.[7]

The rest of the paper is organized as follows. Section 2 discusses issues related to the current lack of standardized practice to convey precision in empirical research in economics. Section 3 formally introduces our concept of a "significant digit" and shows how to implement it for the normal linear conditional mean model. Section 4 uses a simple example where the true data generating process is known to illustrate how our procedure for reporting "significant digits" could be implemented in practice. Section 5 assembles an original dataset of all coefficient estimates and associated standard errors from the articles published in the *American Economic Review* between 2000 and 2022. We use this dataset to study quantitatively the extent to which current practice differs from the two reporting rules we propose. We also provide evidence from a number of other top economics journals that support the view that we would have obtained similar results had we chosen different publication outlets and/or restricted attention to the main estimates of interest instead of including all reported parameter estimates in our dataset. Section 6 concludes.

2 Conveying Precision in Empirical Studies

Empirical results in economics are derived from identification assumptions and estimated on small samples, making them inherently uncertain, as emphasized by Manski (2011, 2019). For this reason, detecting or failing to detect a non-zero parameter estimate has been the main focus of statistical inference in empirical economic research.

As the cost of data collection declines, larger and larger datasets have become available.

[7]How the same economic magnitude is measured can change over time and across locations because of technical change in how this magnitude is collected—manually versus using electronic monitors, for example—and the extent to which components of this magnitude must be estimated. Although these changes are unlikely to impact the value of the first one or two digits of the measured magnitude, they can impact values of its remaining digits. The procedure we propose would make reported parameter estimates largely invariant to reasonably low levels of uncertainty in the underlying input data.

As a result, the statistical power to detect non-zero effects has increased dramatically. For this reason, empirical research in economics seems likely to shift from *detection* to *measurement*. In other words, the key question in statistical inference may no longer be "*is a coefficient estimate different from zero?*" but instead "*how precise is a (non-zero) coefficient estimate?*".

This paper proposes to report point estimates using a standardized approach that communicates their precision. In doing so, we hope to contribute to improving the credibility of empirical research in economics and increase the emphasis on economic significance relative to statistical significance of a parameter estimate.

2.1 Credibility of Empirical Economic Research

In recent years, economists have dedicated significant attention to the credibility and transparency of their empirical research.[8] A major concern has been that journals may not publish a random sample of the underlying distribution of possible estimation results. A widely held view is that "statistically significant" (Franco et al., 2014; Andrews and Kasy, 2019; Brodeur et al., 2023) and "surprising" (DeLong and Lang, 1992) results are more likely to be published.

Establishing *precision* as another salient attribute of reported results would likely reduce the emphasis put on obtaining "statistically significant" results and highlight that uncertainty in each reported decimal digit is an essential feature of empirical work in economics, that can often lead to different conclusions about an economic magnitude of interest, as shown in our earlier demand price elasticity estimates example.

2.2 Economic versus Statistical Significance

A greater emphasis on the precision of parameter estimates also highlights the distinction between"statistical" and "economic" significance. The latter refers, broadly speaking, to the magnitude of the estimated effects, while the former simply tests whether a point estimate is

[8] See the essays in the symposium "Con out of Economics" of the *Journal of Economic Perspectives* (Angrist and Pischke, 2010; Leamer, 2010; Keane, 2010; Sims, 2010; Nevo and Whinston, 2010; Stock, 2010). Notable earlier contributions to this literature include Hendry (1980) and Leamer (1983).

distinct from zero. Many authors have stressed that putting too much emphasis on statistical significance often diverts researchers away from answering the main research questions of interest (DeLong and Lang, 1992; McCloskey and Ziliak, 1996; Ziliak and McCloskey, 2004).

However, statistical significance is only useful in so far as it acts as a gate-keeper to economic significance. In our example above, both demand price elasticity estimates imply a statistically significant negative relationship between price and demand. Yet, only the second demand price elasticity estimate and standard error imply a significant difference from a unitary price elasticity of demand, an economically significant result. Our methodology reports fewer digits in the first case than in the second case, drawing attention to the economic implications of the greater statistical precision of the second demand price elasticity estimate.

3 Significant Digits in Economics

This section introduces the notation and statistical inference framework necessary to assess whether a given digit in a parameter estimate from an econometric model rejects a non-equivalence hypothesis test relative to its two nearest digits.

3.1 Definitions

Let $\hat{\theta}$ be a scalar parameter estimate from a sample of data. In most of what follows, $\hat{\theta}$ will be a coefficient estimate, but we will also discuss the case where $\hat{\theta}$ is a standard error estimate.[9] The vast majority of empirical estimates reported in results tables in economics articles fall into one of these two categories.

Because it is a real number, $\hat{\theta}$ can be written as an infinite sequence of digits $\{d_i\}_{i=-\infty}^{+\infty}$. For example, the number:

$$\hat{\theta} = 1.234\ldots$$

can be seen as a sequence of digits $\{\ldots, 1, 2, 3, 4, \ldots\}$ where the leading "..." are all zeros and the lagging "..." refer to the infinite sequence of remaining decimal digits.

[9]For simplicity, we sometimes refer to $\hat{\theta}$ as the "estimated parameter" even though the standard error of a coefficient estimate is not strictly speaking a parameter of the estimated model.

We can define this sequence formally for any real number as the sum of the product $d_i \times 10^i$ for $i = -\infty$ to $i = +\infty$ for specific values of $d_{-\infty} \ldots d_{+\infty}$:

$$\hat{\theta} \equiv \sum_{i=-\infty}^{+\infty} d_i 10^i \tag{1}$$

For the above example of $\hat{\theta} = 1.234\ldots$, $d_i = 0$ for $i \geq 1$, $d_0 = 1$, $d_{-1} = 2$, $d_{-2} = 3$, $d_{-3} = 4$, etc.

For obvious practical reasons, researchers only report a limited number of decimal digits. For instance, if the parameter estimate is reported with two decimal digits, the "published" or "printed" value of $\hat{\theta}$, which we denote with $\hat{\theta}_p$, will be:

$$\hat{\theta}_p = 1.23$$

In empirical work in economics, the choice of how many digits to report is typically arbitrary. Researchers usually report estimated parameters with either 2 or 3 decimal digits (see Section 5), which is also the default setting of most software routines available to print results tables. This choice does not typically reflect in any way the researchers' confidence in the precision of reported parameter estimates.

We argue that researchers could instead consider rounding the estimates they report in a way that reflects their precision. Specifically, rather that reporting an exogenously fixed number of decimal digits for all parameter estimates and standard errors, researchers could first assess whether each digit they report meets a pre-specified standard of precision. Indeed, as illustrated in Figure 1, the same estimated parameter "1.234…" may be interpreted differently depending on the available precision.

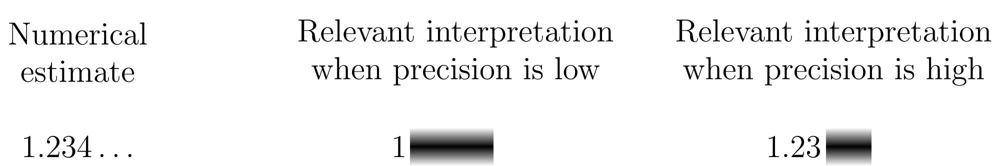

Figure 1: Stylized representation of how to interpret a parameter estimate depending on the available precision.

For each digit of a parameter estimate, the researchers face the following question: Does this digit convey meaningful information or is it simply the result of statistical noise?

This question can be asked iteratively. Specifically, when the estimated parameter is 1.234..., one may first ask: is the first digit of the population parameter value very likely to be "1" or not? If the answer is yes, then one can turn to the second digit ("2") and ask: am I confident that this digit is likely to be a "2" or could it just as well be a "1" or "3"? If the answer is again that the digit is very likely to be correctly estimated, the same question can be asked about the third digit ("3"), and so on and so forth.

To implement this procedure, we build on the *testing statistical hypotheses of equivalence* literature (Wellek, 2010). The general test of statistical equivalence performs the following hypothesis test:

$$H : \theta \leq a \text{ or } \theta \geq b \qquad \text{versus} \qquad K : \theta \in (a, b)$$

where θ is an unknown population parameter value. A typical application of equivalence tests is clinical trials, where two different drugs are deemed to be "bioequivalent" when they trigger similar biological responses. In these applications, θ is the population mean difference in responses and $\hat{\theta}$ is assumed to be normally distributed with unknown mean θ. Similarity between the responses is assessed by specifying a range (a, b) of acceptable population differences (typically centered around zero). If the data rejects H, then the two drugs are considered to be bioequivalent.

Returning to our example of the parameter estimate $\hat{\theta} = 1.234\ldots$, one can assess whether its first digit is significant by performing the following equivalence test:

$$H^{i=0}\colon\ |\theta - 1.234\ldots| > 1 \qquad \text{versus} \qquad K^{i=0}\colon\ \theta \in (0.234\ldots, 2.234\ldots)$$ [10]

We thus introduce a statistical concept of "significant digit" in a parameter estimate as follows.

Definition 1 (Significant digit) *Let* $\hat{\theta} \equiv \sum_{i=-\infty}^{+\infty} d_i 10^i$ *be an estimate of the scalar population parameter* θ*.*

We define digit d_i *to be significant at the level* α *if the following equivalence test* ψ_i *rejects the null hypothesis* H^i *at the level* α*:*

$$H^i\colon\ |\theta - \hat{\theta}| > 10^i \qquad \textit{versus} \qquad K^i\colon\ \theta \in \left(\hat{\theta} - 10^i, \hat{\theta} + 10^i\right).$$

[10] Note that $(0.234\ldots, 2.234\ldots)$ is the open interval that does contain either endpoint.

Continuing our example, suppose that the null hypothesis $H^{i=0}$ associated with the equivalence test ψ_0 is rejected so that the digit $d_0 = 1$ is deemed significant according to Definition 1. We can then turn to the next digit and perform the equivalence test ψ_{-1}:

$$H^{i=-1}\colon\ |\theta - 1.234\ldots| > 0.1 \qquad \text{versus} \qquad K^{i=-1}\colon\ \theta \in (1.134\ldots, 1.334\ldots)$$

If the null hypothesis $H^{i=-1}$ associated with the equivalence test ψ_{-1} is rejected as well, so that the digit $d_{-1} = 2$ is also deemed to be significant, we can again turn to the next digit $d_{-2} = 3$ and perform the equivalence test ψ_{-2}:

$$H^{i=-2}\colon\ |\theta - 1.234\ldots| > 0.01 \qquad \text{versus} \qquad K^{i=-2}\colon\ \theta \in (1.224\ldots, 1.244\ldots)$$

Intuitively, this procedure can be iterated until we fail to reject the null hypothesis for some decimal rank i. In order to ensure that the procedure indeed converges, we make the following assumption.

Assumption 1 (Monotonicity of equivalence tests) *For any rank i, if ψ_i rejects the null hypothesis, then the equivalence test ψ_{i+1} also rejects the null hypothesis.*

Note that, by a recurrence argument, Assumption 1 is equivalent to stating that if ψ_i rejects the null hypothesis, then $\psi_{i'}$ rejects the null hypothesis for all $i' > i$. This assumption is very unlikely to prove restrictive in practice. In particular, it trivially holds for the equivalence tests we use below in the context of the normal linear conditional mean model. We nonetheless state it as an assumption in order to introduce in full generality the following concept of "last significant digit".

Definition 2 (Last significant digit) *Under Assumption 1, we define the rank $i^*(\alpha)$ of the "last significant digit" as the smallest value of i such that digit d_i in Equation (1) is significant at level α according to Definition 1:*

$$i^*(\alpha) \equiv \min_i \{ i \,|\, d_i \textit{ is a significant digit at level } \alpha \}$$

Under Assumption 1, it is straightforward to see that, if d_i is a significant digit, then all $d_{i'}$ with $i' > i$ are also significant digits. Because finite sample sizes necessarily put an upper bound on precision, there exists a threshold rank $i^*(\alpha)$ such that a digit d_i is significant if, and only if, $i \geq i^*(\alpha)$. We call the corresponding digit d_{i^*} the "last significant digit." The rank of the last significant digit summarizes the "precision" of the estimated parameter. According to our definition of "significant," the value of this parameter can only be precisely inferred at the order of magnitude 10^{i^*}.

It is important to note that $i^*(\alpha)$, the rank of the last significant digit, may be strictly greater than $D(\hat{\theta})$, the rank of the first non-zero digit of the estimated parameter, which can be formally defined as:

$$D(\hat{\theta}) \equiv \max_i i\mathbf{1}\left[d_i \neq 0\right] \tag{2}$$

where $\mathbf{1}\left[d_i \neq 0\right]$ is a dummy variable that is equal to 1 when $d_i \neq 0$ and zero otherwise. In such cases, the parameter estimate does not have any non-zero significant digit. Even then, the rank of the last significant digit nonetheless provides useful information about the "precision" of the estimate. For example, if the estimated parameter is 0.436 and $i^*(\alpha) = 0$, our procedure implies that we can rule out a first digit strictly greater than 1 for this value of α, but the same inference cannot be made if $i^*(\alpha) = 1$. The "precision" of the estimated parameter is 10^0 in the former case while it is only 10^1 in the latter.

Finally, when the absolute value of a parameter estimate is strictly lower than one, as in the above example of 0.436, the zero(s) preceding the first non-zero digit provide little information.[11] Under Definition 1, initial zeros are considered to be significant digits as long as their rank is weakly greater than $i^*(\alpha)$. We therefore introduce the concept of the "number of sizable significant digits" to refer to the number of precisely estimated digits when one ignores any zero(s) preceding the first non-zero digit of a parameter estimate.

Definition 3 (Number of sizable significant digits) *Let $\hat{\theta} \equiv \sum_{i=-\infty}^{\infty} d_i 10^i$ be an estimate of the scalar population parameter θ. Let $D(\hat{\theta})$ equal the largest rank of its non-zero digits:*

$$D(\hat{\theta}) \equiv \max_i i\mathbf{1}\left[d_i \neq 0\right] \tag{3}$$

Let $i^*(\alpha)$ equal the rank of the last significant digit of $\hat{\theta}$. We define the "number of sizable significant digits," $\nu(\alpha)$ as:

[11] One reason for the use of scientific notation is to avoid reporting such initial zeros.

$$\nu(\alpha) \equiv \max(D(\hat{\theta}) - i^*(\alpha) + 1, 0)$$

The number of sizable significant digits is the number digits that: (i) contribute to measuring a non-zero coefficient estimate and (ii) can be precisely estimated from the available data. For example, if the parameter estimate is $1.234\ldots$, only the first digit "1" is precisely estimated when $i^*(\alpha) = 0$, while all four non-zero digits are precisely estimated when $i^*(\alpha) = -3$. Because $D(\hat{\theta}) = 0$ for this example ($i = 0$ and $d_0 = 1$), the formula in Definition 3 yields $\nu(\alpha) = 1$ in the former case, and $\nu(\alpha) = 4$ in the latter.

3.2 Equivalence Testing

Definition 1 does not specify which equivalence test should be run in practice. We propose to adopt the widely used "two one-sided test" (TOST). This test rejects the null hypothesis H^i at the level α if, and only if, the following two one-sided tests reject their null hypothesis at level α:

$$H_a^i\colon\ \theta = \hat{\theta} - 10^i \qquad \text{versus} \qquad K_a^i\colon\ \theta > \hat{\theta} - 10^i \qquad\qquad (\psi_{ia})$$

and

$$H_b^i\colon\ \theta = \hat{\theta} + 10^i \qquad \text{versus} \qquad K_b^i\colon\ \theta < \hat{\theta} + 10^i \qquad\qquad (\psi_{ib})$$

Because both tests ψ_{ia} and ψ_{ib} are of size α, it may seem counter-intuitive that the resulting test ψ_i is also of size α. Indeed, in standard settings, induced tests for multiple hypothesis testing have to use a lower significance level for each individual test in order for the overall test to remain of size α.[12] The TOST is a special case of the so-called "intersection-union tests" (Berger, 1982). The underlying intuition for the lack of need to adjust the size of individual tests to account for multiple testing is that H^i is rejected only if *both* H_a^i and H_b^i are rejected.

Several tests besides the TOST have been proposed in the literature on equivalence testing (e.g. Brown et al. (1997)). However, because our procedure requires "testing" many digits for numerous parameter estimates in order to assess their statistical precision, we believe that the simplicity of implementing the TOST makes it ideally suited to this task.

[12] For example, the Bonferroni bound implies a size of $\frac{\alpha}{M}$ for each of the M individual tests for an overall size of α for the joint null hypothesis (see for example Savin (1984)).

3.3 Results for the Normal Linear Conditional Mean Model

This section presents finite sample results for our procedure for individual elements of the coefficient vector from the model:

$$y = X\beta + \epsilon \tag{4}$$

where y is a $T \times 1$ vector of outcome variables, β is a $K \times 1$ vector of coefficients, X is a $T \times K$ fixed matrix of rank K where $K < T$, and ϵ is a $T \times 1$ random vector. The conditional distribution of ϵ given the matrix X is a multivariate normal $T \times 1$ vector with mean zero and covariance matrix $\sigma^2 I_T$, where I_T is a $T \times T$ identity matrix and σ^2 is an unknown positive scalar.

Under the above assumptions, y is $\mathcal{N}(X\beta, \sigma^2 I_T)$, where $\mathcal{N}(\mu, \Sigma)$ denotes a multivariate normal distribution with mean vector μ and covariance matrix Σ. In addition, the ordinary least squares estimate of β, $b \equiv (X'X)^{-1}X'y$, is a $K \times 1$ random vector distributed as $\mathcal{N}(\beta, \sigma^2(X'X)^{-1})$. Define $s^2 \equiv \frac{e'e}{T-K}$, where $e \equiv y - Xb$. Under the above assumptions, $\frac{s^2(T-K)}{\sigma^2}$ is distributed as a χ^2 random variable with $T - K$ degrees of freedom.

We now derive expression for $i^*(\alpha)$, the (size α) rank of the last significant digit, as well as a lower bound for this rank for the case in which we do not know $T - K$. This lower bound is useful for our empirical exercise discussed in Section 5 where we do not observe the number of degrees of freedom. It relies on an large T approximation.

Define a_k as a $K \times 1$ vector with zeros in all elements except the k^{th} element which is equal to one. The inner product of a_k with b, yields:

$$b_k = a_k' b \tag{5}$$

the k^{th} element of the vector b. The standard error estimate for b_k is equal to:

$$SE(b_k) \equiv s\sqrt{a_k'(X'X)^{-1}a_k} \tag{6}$$

As shown in Proposition 1 in Appendix B, a closed-form expression for the level α rank of the last significant digit for b_k is:

$$i^*(\alpha) = \lceil \frac{\ln(SE(b_k)) + \ln(t_{1-\alpha}(T-K))}{\ln(10)} \rceil \tag{7}$$

where $t_{1-\alpha}(T-K)$ is the $1-\alpha$ percentile of a Student's t-distribution with $T-K$ degrees of freedom and $\lceil x \rceil$ denotes the ceiling of x (closest integer value from above).[13]

For a given level of statistical significance, α, computing the rank of the last significant digit of an estimated coefficient only requires knowledge of the number of degrees of freedom $T-K$ and the standard error estimate $SE(b_k)$. Both quantities are readily available as outputs of the software routines used to estimate the parameters of linear conditional mean models. Therefore, any post-processing routine used to generate estimation results tables can easily compute the rank of the last significant digit of a coefficient estimate.

Because published articles do not always report the number of degrees of freedom associated with a coefficient estimate and standard error, it is useful to note that $t_{1-\alpha}(T-K)$ is a decreasing function of $T-K$ that converges to a finite value $t_{1-\alpha}(\infty) > 0$ as T goes to infinity. This observation is just a re-statement of the well-known result that $t_{1-\alpha}(\infty) = Z_{1-\alpha}$, the $1-\alpha$ percentile of a $\mathcal{N}(0,1)$ random variable. Therefore, an asymptotic lower bound for the size α rank of the last significant digit is (see Corollary 1 in Appendix B):

$$i^*_\infty(\alpha) = \lceil \frac{\ln(SE(b_k)) + \ln(Z_{1-\alpha})}{\ln(10)} \rceil. \tag{8}$$

Our definition of significant digits can be applied to standard error estimates, such as $SE(b_k)$, and more generally to any random scalar for which one can devise a TOST equivalence test. For example, Proposition 2 in Appendix B derives a closed-form lower bound for the rank of the last significant digit of the standard error estimate, $SE(b_k)$, of a coefficient estimate from normal linear conditional mean model. We make use of this expression in the examples discussed in Section 4.

[13]For example, the ceiling of 1.234 is 2 and the ceiling of -1.234 is -1.

3.4 Other Statistical Models

As shown in the proof of Proposition 1 in Appendix B, the closed-form expression for the rank of the last significant digit of a coefficient estimate only relies on the fact that the t-statistic has a Student's t-distribution with $T - K$ degrees of freedom. Although this property does not hold exactly for models other than the normal linear conditional mean model discussed above, it is often a reasonable asymptotic approximation. Because conveying precision is most relevant in applications with large sample sizes, our previous results are likely to be a useful approach to determining which digits of an estimated coefficient are statistically significant.

Deriving a similar asymptotic result for standard error estimates (see Proposition 2 in Appendix B) from other econometric models is not possible in general. However, the main purpose of reporting standard error estimates is to give to readers a sense of the precision of coefficient estimates. Reporting standard errors is therefore to some extent redundant with the use of significant digits.[14].

4 Use of Significant Digits Procedure

This section employs a simple example to illustrate how our procedure for determining "significant digits" may be used to report empirical results in economics. We draw data from a known true data generation process to provide greater insight into the properties of our procedure.

4.1 True Data Generating Process

The outcome variable y_t depends on three independent variables X_{1t}, X_{2t} and X_{3t}, and an error term ϵ_t. We assume the true data generating process is:

$$y_t = \beta_0 + \sum_{k=1}^{3} \beta_k X_{kt} + \epsilon_t \tag{9}$$

[14]If the rank of the last significant digit of a given coefficient estimate is $i^*(\alpha)$, the use of significant digits can already convey the information that the coefficient estimate is precise up to a measurement error of at most $10^{i^*(\alpha)}$

where $(\beta_0, \beta_1, \beta_2, \beta_3)$ is the vector of true parameter values. To keep things simple while still being able to illustrate our main points, we set $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$.

We assume the researcher observes T draws of the following vector indexed by t:

$$\{(y_t, X_{1t}, X_{2t}, X_{3t})\}_{t=1...T}$$

To produce these vectors, we take T independent identically distributed draws of X_{1t}, X_{2t}, X_{3t} and ϵ_t, respectively from the following mutually independent normal distributions:

$$X_{1t} \sim \mathcal{N}(0,1),\ X_{2t} \sim \mathcal{N}(0,5),\ X_{3t} \sim \mathcal{N}(0,10), \text{ and } \epsilon_t \sim \mathcal{N}(0,5).$$

For each draw, we use the true data generation process (Equation (9)) and the assumed values of the model parameters ($\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$) to compute a value of y_t.

Models Estimated

We take the perspective of a researcher who only observes a sample of size T of the vector $(y_t, X_{1t}, X_{2t}, X_{3t})$ and seeks to recover an estimate of β_1. Assume the researcher estimates and reports three models:

$$\begin{aligned}
&\textbf{Model 1:} \quad y_t = \beta_0 + \beta_1 X_{1t} + \eta_{1t} \\
&\textbf{Model 2:} \quad y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \eta_{2t} \\
&\textbf{Model 3:} \quad y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \eta_{3t}
\end{aligned}$$

Note that we include different error terms—η_{1t}, η_{2t}, and η_{3t}—in each model to account for the omitted regressors.

Model 1 corresponds to the simplest empirical strategy where the researcher tries to recover β_1 by regressing y_t on X_{1t}. Model 2 adds X_{2t} as a control variable but not X_{3t}. Finally, model 3 corresponds to a situation where the researcher is actually estimating the true data generating process. Note that because the X_{1t}, X_{2t}, X_{3t} and ϵ_t are independent of each other and across observations, all three estimates of β_1 are unbiased. However, because X_{1t} has by far the smallest variance among X_{1t}, X_{2t}, X_{3t} and ϵ_t, we expect it to be difficult to estimate β_1 precisely without including X_{2t} and X_{3t} in the regression.

4.2 Using Significant Digits to Report Empirical Results

We now discuss a procedure to report parameter estimates that conveys their precision in a transparent and efficient manner. We first consider a sample of size $T = 60$.

Panel (a) of Figure 2 shows the current mainstream approach to report empirical results in economics. The number of digits displayed does not convey any information about their precision. We simply keep the default settings of a popular software routine that generates tables of estimation results. As a consequence, many printed digits actually do not provide any relevant information. For example, the coefficient estimates for X_{1t} in models 1 and 2 are very imprecisely estimated, as expected. Therefore, reporting that the second and third decimal digits of these statistical zeros are respectively "89" and "84" may be considered as statistical noise.

Panel (a)

	Model 1	Model 2	Model 3
X_1	0.689	0.784	0.713
	(0.678)	(0.623)	(0.366)
X_2		0.899	1.005
		(0.260)	(0.153)
X_3			1.010
			(0.097)
Constant	0.945	0.710	0.840
	(0.579)	(0.536)	(0.315)
Observations	60	60	60
R^2	0.017	0.187	0.724

Panel (b)

	Model 1	Model 2	Model 3
X_1	0	0	+0
	(0.6)	(0.6)	(0.3)
X_2		+0	1
		(0.2)	(0.1)
X_3			1
			(0.0)
Constant	0	0	+0
	(0.5)	(0.5)	(0.3)
Observations	60	60	60
R^2	0.017	0.187	0.724

Panel (c)

	Model 1	Model 2	Model 3
X_1	0.7	0.8	0.7
	(0.68)	(0.62)	(0.37)
X_2		0.9	1.0
		(0.26)	(0.15)
X_3			1.0
			(0.10)
Constant	0.9	0.7	0.8
	(0.58)	(0.54)	(0.32)
Observations	60	60	60
R^2	0.017	0.187	0.724

Figure 2: Different possible formats for the table of estimation results. Panel (a): current mainstream approach in economics (the table was generated using the package "stargazer" in R (Hlavac, 2022)); Panel (b): reporting only significant digits for $\alpha = 0.1$; Panel (c): reporting significant digits for $\alpha = 0.1$ as well as the first (rounded) non-significant digit.

We propose instead to build on Propositions 1 and 2 in Appendix B to increase the "signal-to-noise" ratio of the reported results table. Panel (b) of Figure 2 displays the same results as in Panel (a), but only reports significant digits at the level $\alpha = 0.1$, both for coefficients and standard errors. Choosing $\alpha = 0.1$ instead of say $\alpha = 0.01$ is a conservative approach in the sense that the less stringent the equivalence tests are, the more digits they will consider to be significant. A caveat of this approach is that an estimated coefficient may be statistically different from zero but have no non-zero significant digit. This situation occurs for example for the estimate of β_1 for Model 3 and the estimate of β_2 for Model 2. One approach to make a distinction between a statistical zero and a positive coefficient with no non-zero significant digit could be to report "0" in the former case (the estimate of β_1 for Model 1), and "+0" in the latter (the estimate of β_2 for Model 2), as shown in Panel (b) of Figure 2. Another caveat is that such a situation where an estimated number has no non-zero significant digit can also occur for standard errors, as illustrated by the estimated standard error for β_3 in Model 3.

Therefore, one may prefer to report the first non-significant digit, rounded up or down depending on the next non-significant digits. This approach is shown in Panel (c) of Figure 2. Adopting this practice to report empirical results in economics could represent a reasonable trade-off between Panel (a), where most printed digits are not "significant" according to Definition 1, and Panel (b), where some coefficients are statistically different from zero but do not have any non-zero significant digit.

Panel (a)

	Model 1	Model 2	Model 3
X_1	0.7	0.8	0.7
	(0.68)	(0.62)	(0.37)
X_2		0.9	1.0
		(0.26)	(0.15)
X_3			1.0
			(0.10)
Constant	0.9	0.7	0.8
	(0.58)	(0.54)	(0.32)
Observations	60	60	60
R^2	0.017	0.187	0.724

Panel (b)

	Model 1	Model 2	Model 3
X_1	1.02	1.01	1.00
	(0.01817)	(0.01573)	(0.00906)
X_2		0.999	0.988
		(0.00706)	(0.00406)
X_3			0.997
			(0.00287)
Constant	1.00	0.99	0.99
	(0.01824)	(0.01579)	(0.00909)
Observations	60,000	60,000	60,000
R^2	0.050	0.288	0.764

Panel (c)

	Model 1	Model 2	Model 3
X_1	1.02	1.01	1.00
X_2		0.999	0.988
X_3			0.997
Constant	1.00	0.99	0.99
Observations	60,000	60,000	60,000
R^2	0.050	0.288	0.764

Figure 3: Results tables when reporting digits up to the first non-significant digit for $T = 60$ (left panel) and $T = 60,000$ (middle and right panels).

We next increase sample size to $T = 60,000$ to illustrate what happens when precision increases.[15] Figure 3 compares the table of estimation results for $T = 60$ and $T = 60,000$ when reporting digits up to the first non-significant digit. As expected, coefficients are more precisely estimated and therefore reported with additional digits. Similarly, standard errors are more precisely estimated and therefore reported with more digits. This latter observation may be seen as a drawback because it somewhat defeats the purpose of reporting simpler results tables. Yet, the current practice of truncating standard errors to a fixed number of decimal digits may not be satisfactory either. For example, truncating standard errors to two

[15]We chose such a large sample size to provide an example where (i) the rank of the last significant digit differs across coefficient estimates; and (ii) the standard error estimate of at least one coefficient estimate is equal to zero when the researchers rule is to always report two decimal digits.

decimal digits would induce some of the reported standard errors to be 0.00.[16] However, the main purpose of reporting standard errors is to convey the precision of coefficient estimates. Therefore, reporting coefficient estimates in a way that accounts for precision through our concept of "significant digits" may also represent a substitute for reporting standard errors for cases when standard errors are extremely small relative to the coefficient estimates. For large sample sizes, Panel (c) in Figure 3 could thus be a sensible approach to report estimates in a reader-friendly way while still accounting for precision.

4.3 Discussion

We now discuss a number of practical considerations when choosing the units of dependent and independent variables. First, it is important to note that the rank $i = 0$ plays no particular role for our concept of significant digits. Equivalently, whether a given digit is a decimal digit or not is irrelevant. However, the choice of the units of the dependent and independent variables in a regression may matter in several ways. To discuss such situations, consider an estimate $\hat{\theta}$ of a parameter:

$$\hat{\theta} = \sum_{i=-\infty}^{D} d_i 10^i \tag{10}$$

where D is the largest rank for non-zero digits (omitting its dependence on $\hat{\theta}$ to simplify notation). We denote with $i^{\dagger}(\alpha)$ the rank of the last printed digit under the chosen equivalence test size α and reporting rule (either the rank of the last significant digit or the first non-significant one). Printed digits correspond to all digits whose rank lies between $i^{\dagger}(\alpha)$ and $\max(D, 0)$. A few special cases are worth discussing.

First, the largest rank for non-zero digits may be negative (i.e., $D < 0$). For example, a coefficient may be estimated to be 0.023 (i.e., $D = -2$). Although the above reporting rules can of course still be implemented, our empirical exercise of Section 5 is sensitive to the choice of units in such situations. Indeed, even though the two zeros of the estimated

[16] Although this remark may seem anecdotal, in the dataset we collect in Section 5 we observe this outcome for almost 1% of the standard error estimates we recover from articles published in the *American Economic Review*.

coefficient may be seen as relatively uninformative, they are considered in our empirical application as significant digits whenever $i^*(\alpha) \leq -1$. For example, if $i^*(\alpha) = -2$, three out of the four printed digits of 0.023 are considered to be statistically significant. If the unit of the explanatory variable were to be divided by 100, the estimated coefficient would become 2.3 and only one out of the two printed digits would be statistically significant. From this perspective, our assessment in Section 5 of the share of printed digits that are not statistically significant may be seen as conservative.

Another practical concern in situations where $D < 0$ and $i^\dagger(\alpha) < 0$ is that, when researchers print an exogenously fixed number of decimal digits, they may report estimates with too few non-zero digits. For example, they may report 0.001 when the rank of the last significant digit is actually $i^*(\alpha) = -5$. Such rounding practices have been shown to possibly generate biases in meta-analysis studies (Kranz and Pütz, 2022; Brodeur et al., 2022). In particular, Elliott et al. (2022) show that p-hacking tests can be sensitive to whether p-values are rounded or not.

Finally, the rank of the last significant digit may be a positive power of ten ($i^*(\alpha) > 0$). For example, a parameter estimate may be 451.23... while the rank of the last significant digit is $i^*(\alpha) = 2$. Under the approach consisting in reporting significant digits up to the first non-significant one, only "4" and "5" would be printed. However, reporting 450 does not implement this rule because of the terminating "0." This is because when an estimate does not have decimal digits and ends with a zero, the reader cannot tell whether this zero is a significant digit or not. The use of scientific notation would then be required to implement the reporting rules discussed above.[17] Our analysis of Section 5 suggests that the rank of the last significant digit is a positive number in a small minority of cases in empirical research in economics. Therefore, changing measurement units is likely to be a more pragmatic approach to handle these situations than widespread use of scientific notation.

[17]Appendix A illustrates the use of scientific notation for our example of Section 4.

5 Significant Digits in Economic Research

This section quantitatively assesses the extent to which the coefficient estimates printed in published articles in economics consist of significant digits. Our assessment relies on the asymptotic approximation presented in equation (8) that only requires knowledge of the standard error estimate of the coefficient estimate.[18]

5.1 Data Compiled

Our empirical exercise relies on coefficient estimates and associated standard error estimates retrieved from the universe of articles published in the *American Economic Review* between 2000 and 2022.[19] The choice of this specific outlet is motivated by its prominence in the economics profession and the state-of-art nature of the articles it publishes. We restrict attention to the period post-2000 for two reasons. First, downloading full issues is not possible before 2011, making it very time consuming to retrieve all published articles in a given year. Second, the practice of systematically reporting coefficient estimates and standard errors in tables of empirical results is relatively recent.[20]

5.2 Descriptive Analysis

This section presents a number of trends and key features of our dataset of over 90,000 coefficients with their standard error estimates that may be helpful to interpret our main results.

Trend in the Number of Coefficient Estimates

First, we ask whether the number of coefficient estimates reported in the *American Economic Review* has changed over the two decades we consider. Because, the number of volumes of the journal has increased from 4 in 2000 to 12 in 2022, simply looking at the

[18] Facing the same data limitation, Brodeur et al. (2020) take a similar approach, focusing on Z-statistics.

[19] We exclude the *AEA Papers and Proceedings* issues over the entire period.

[20] In their study of empirical articles published in top economics journals during the years 2015 and 2018, Brodeur et al. (2020) note (p. 3639): "*Most [test statistics in our sample] (91 percent) are reported as coefficients and standard errors, others as t-statistics (4 percent), or p-values (5 percent).*" Previous practices were less harmonized, with papers reporting indifferently standard errors, t-statistics or z-statistics. This absence of standardization in how results are reported would make our routine to extract coefficients and standard errors more prone to errors and more likely to miss observations (see Appendix C).

total number of reported coefficients per year would be misleading. Instead, we normalize the total number of reported coefficient estimates per year by the total number of printed pages for all articles (both theoretical and empirical) published that year.

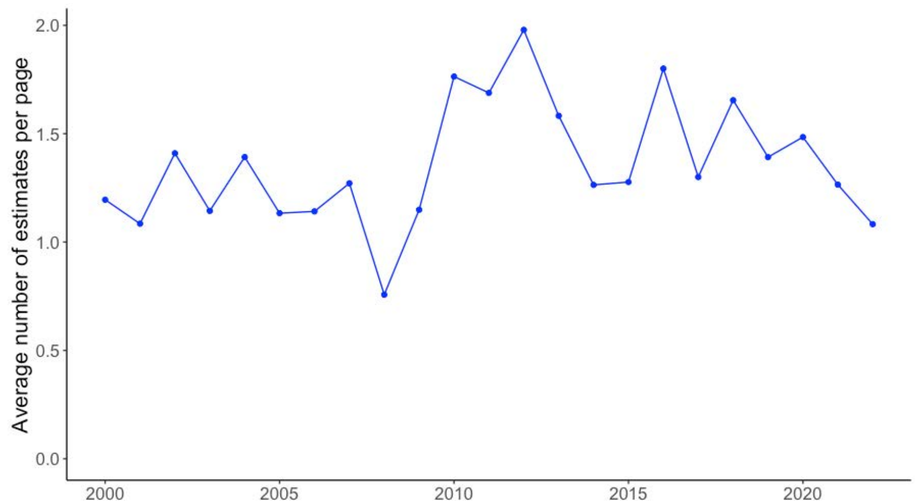

Figure 4: Evolution of the ratio of the number of reported empirical estimates in each year to the number of printed pages for that year.

Figure 4 shows how this ratio has evolved over our period of interest. We find that on average there is about 1.4 coefficient estimates per printed page. This ratio increased in 2010 through 2013, but fell back down to slightly higher than its pre-2010 average in 2014.

Reported Number of Decimal Digits

We now discuss how many decimal digits researchers typically display during our sample period. As shown in Figure 5, the majority of estimated coefficients are reported with three decimal digits. If anything, this practice seems to have become more and more prevalent over time, mostly at the expense of reporting only two decimal digits. Although this observation is consistent with a widespread use of the default setting of econometric or post-processing software routines, it does not necessarily indicate that current practice is significantly at odds with the significant digit approach we derived in Section 4. It could be the case that reporting three decimal digits is justified by the greater precision of the resulting parameter estimates. This interpretation is unlikely because the majority of tables of estimation results report the same number of decimal digits for all coefficient estimates.

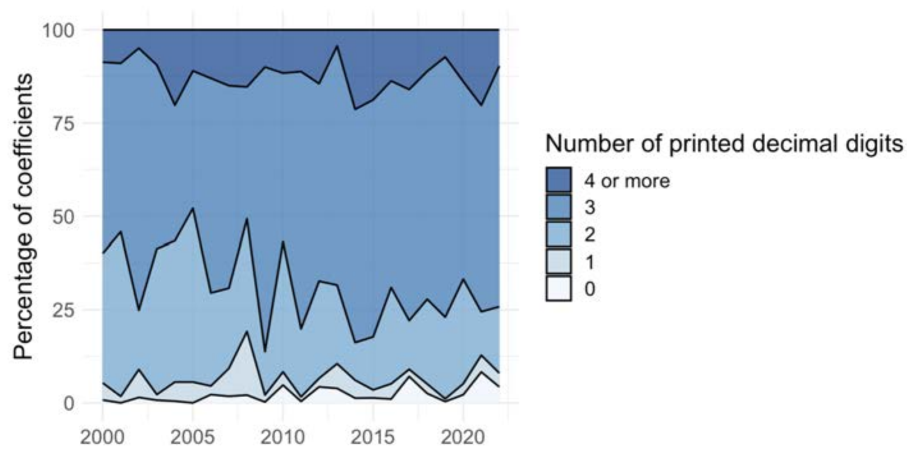

Figure 5: Evolution over time of the number of decimal digits printed for coefficient estimates.

Order of Magnitude of Coefficient and Standard Error Estimates

Figures 6 and 7 show the distributions of the absolute value of respectively coefficients and standard errors, broken down by year. Two main observations are relevant.

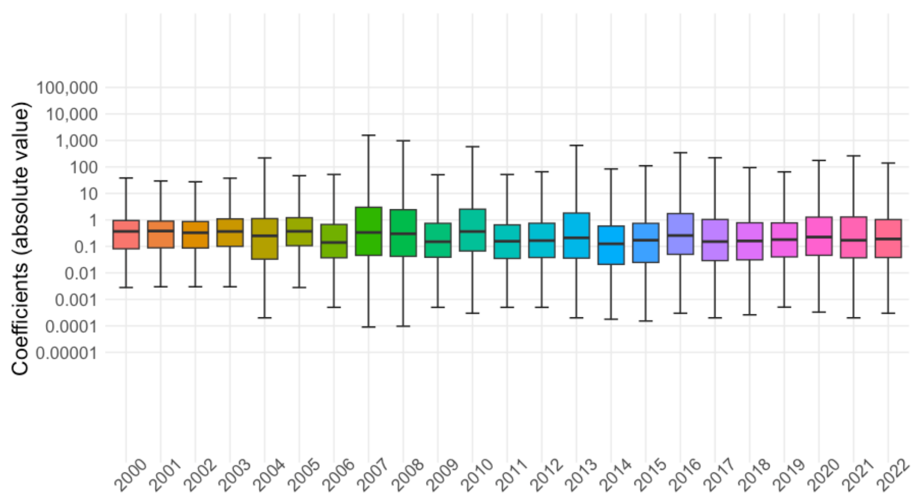

Figure 6: Box plots of the magnitude of estimated coefficients (absolute value in $\log_{10}$ scale). Boxes locate the first, second and third quartiles of the distributions. Top whiskers (resp. bottom whiskers) are drawn at a distance of 1.5 times the interquartile range above the third quartile (resp. below the first quartile).

First, the vast majority of coefficient and standard error estimates have a magnitude in absolute value lower than 10. As a result, it is possible in most cases to convey information about significant digits without having to resort to scientific notation. The few situations with large estimated numbers and low precision could instead be handled with a change in

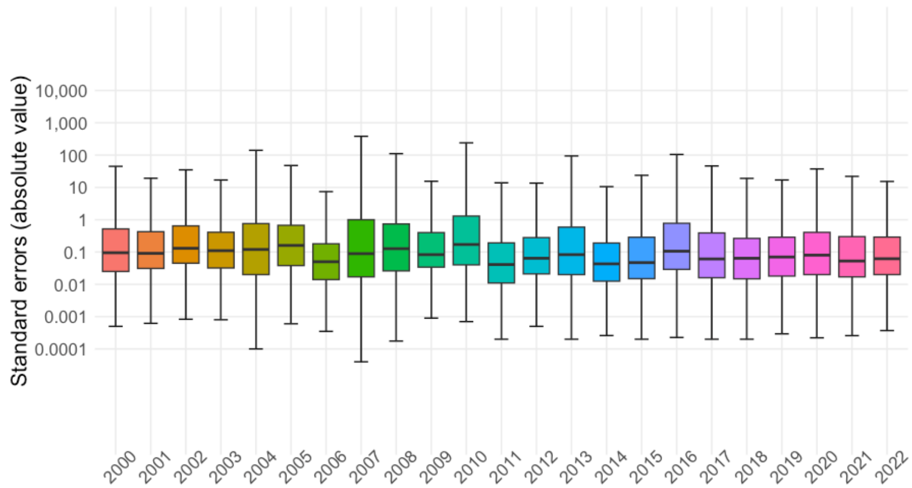

Figure 7: Box plots of the magnitude of estimated standard errors (absolute value in $\log_{10}$ scale). Boxes locate the first, second and third quartiles of the distributions. Top whiskers (resp. bottom whiskers) are drawn at a distance of 1.5 times the interquartile range above the third quartile (resp. below the first quartile).

units.

Second, we observe no obvious temporal trend in the absolute value of the magnitude of reported coefficients and standard errors. In particular, there is no clear evidence that standard errors have become either significantly smaller (e.g. due to improving empirical strategies or increasing sample sizes) or larger (e.g. due to a more prevalent use of robust standard error estimates) over time.

5.3 Main Results

We use our dataset to assess quantitatively the extent to which current practice reports parameter estimates to reflect their precision. Specifically, we derive, for each coefficient estimate, an upper bound of the number of printed digits that are statistically significant. Going back to the example of Section 4, we compare the number of digits that were printed in the published article (panel (a) of Figure 2) to the number of digits that would be printed under alternative reporting rules (panels (b) or (c) of this same figure).

Figure 8 shows, for each year, the fraction of printed digits that are not significant for different values of α. Even for the most conservative level of $\alpha = 10\%$, we find that less than 40% of printed digits for coefficient estimates are significant digits according to

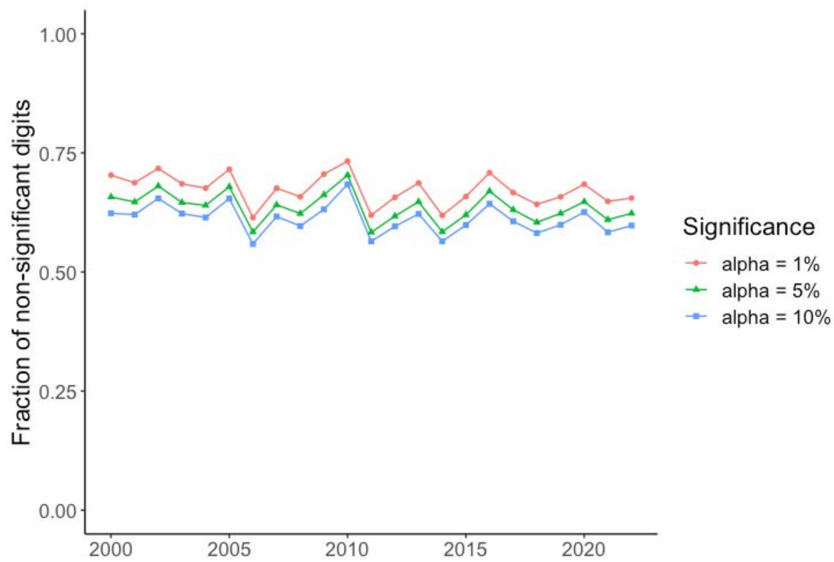

Figure 8: Fraction of the printed digits of estimated coefficients that are not significant digits at different values of $\alpha = 0.1,\ 0.05,\ \ 0.01$.

our approach. In other words, over 60% of printed digits for coefficient estimates are not statistically significant by our definition. In addition, there is no clear evidence that this share has been significantly decreasing over the past two decades.

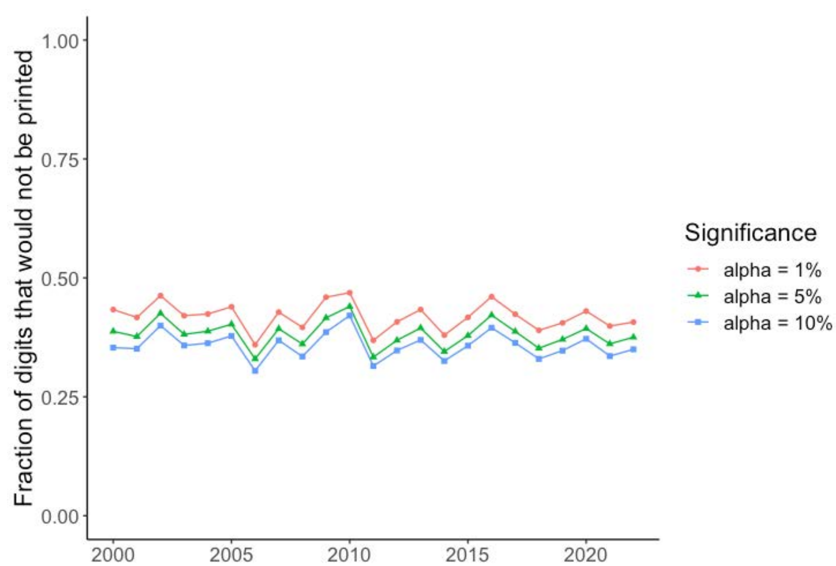

Figure 9: Fraction of the printed digits that would not have been printed if results tables had only reported digits up to the first non-significant digit.

As discussed in Section 4, it may however make sense for empirical papers to also report the first non-significant digit, the approach in panel (c) instead instead of approach in panel (b) in Figure 2. Even under the more generous reporting rule that would print digits up to the first non-significant one, Figure 9 shows that over a third of the digits printed in published articles would not have been reported under this procedure. In other words, more than a third of the digits that have been published as part of a coefficient estimate reported in the *American Economic Review* between 2000 and 2022 does not meet this criteria.

Our framework also allows us to identify coefficient estimates for which our reporting rule would have implied printing *more* digits. For the significance level of $\alpha = 0.1$, we find that only about 1% of the coefficients fall into this category.

Finally, we can use our dataset to compute, for each coefficient, an upper bound of its number of sizable significant digits (see Definition 3). As discussed above, this value represents the number of digits that (i) contribute to a non-zero effect and (ii) can be precisely estimated from the available data. For this exercise, we drop observations from our sample where the printed standard error estimate is zero (1% of coefficients) because we cannot derive a finite value for the rank of the last significant digit for these observations.

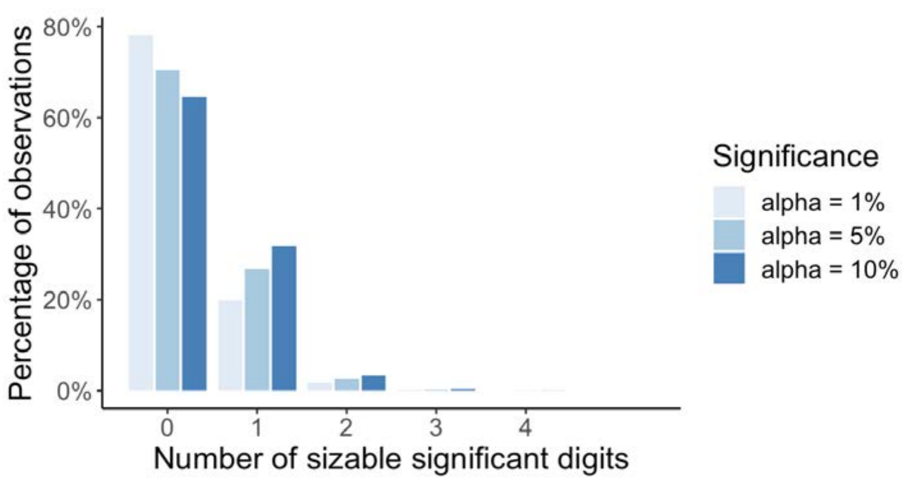

Figure 10: Distribution of the number of sizable significant digits.

Figure 10 shows histograms for the number of sizable significant digits for different levels of statistical significance. First, we observe that the number of sizable significant digits is equal to zero for the vast majority of coefficient estimates (60% to 80% depending on the chosen level of statistical significance, $\alpha = 0.1,\ 0.05,$ and 0.01). For these coefficient estimates, no non-zero digit is precisely estimated. Our dataset contains all coefficient estimates reported in tables of results, which means that it includes coefficient estimates for both the "main effects", that are the focus of the different studies, and coefficient estimates for control variables. It is therefore possible that the "main effects" are actually precisely estimated, but the coefficient estimates for the control variables are not. We however find little support for this hypothesis in what follows. Second, comparing Figures 8 and 9 to Figure 10 suggests that the number of sizable significant digits is more sensitive to the chosen level of statistical significance than our assessment of the number of printed digits that are significant.[21]

5.4 Discussion and Robustness Checks

The dataset we collected has two main limitations. First, we restricted attention to a single outlet, namely the *American Economic Review*. Second, we retrieved all coefficient estimates in tables of results and are therefore not able to infer whether or not a given coefficient corresponds to the main effect of interest in the considered study. As a result, we cannot rule out that our finding that most coefficients have a number of sizable significant digits equal to zero (Figure 10) might be driven by a lack of statistical power to precisely estimate the coefficients associated with control variables. In other words, our results presented thus far do not necessarily imply that the main coefficients of interest are imprecisely estimated.

To address these limitations, we make use of the dataset collected by Askarov et al. (2023).[22] In this study, the authors retrieved the estimated main effect and corresponding standard error from over 300 meta-analysis studies. As in Askarov et al. (2023), we restrict the sample to general interest and field journals in economics. Specifically, we restrict attention to 15 outlets: the "top five" journals,[23] four general interest journals,[24] and the six

[21] This observation is consistent with the fact that the majority of printed coefficients are strictly lower than one. Indeed, the median absolute value of the coefficients in our dataset is 0.2 (see Figure 6). Consider one such coefficient, say for example 0.436, whose rank for the last significant digit is $i^*(\alpha) = -1$. In this example, the number of sizable significant digits is one (corresponding to the digit "4"), that is 25% of printed digits. However, 50% of the printed digits are deemed as significant in the analysis underlying Figures 8 and 9 (both the "0" and the "4" are significant digits by Definition 1). Because the "0" is not the last significant digit, the fact that it is significant is much less sensitive to the chosen level of statistical significance.

[22] Their dataset is publicly available at `https://github.com/anthonydouc/Datasharing/tree/master/Python/data/group1` (last accessed on 15 August 2023).

[23] *American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies* and *Econometrica*.

[24] *Review of Economics and Statistics*, *European Economic Review*, *Economic Journal*, and *Journal of the*

top-field journals included in Askarov et al. (2023) for which we have the largest number of observations.[25] The resulting dataset consists of 18,000+ pairs of coefficient and their associated standard error estimates. In contrast to the dataset used for our main analysis, the coefficients included in this dataset correspond to the "main effect of interest" of the different empirical studies.

We use this complementary dataset to explore (i) the validity of our main results to these journals, and (ii) whether results for the "main effect" coefficient estimates differ from the population of printed coefficients. This external dataset however comes with its own limitations. First, in contrast to the dataset we collected, the sample of coefficients for each outlet is not exhaustive: it is restricted to the main "treatment" effects that where subsequently tabulated in one of the meta-analyses compiled by Askarov et al. (2023). Second, no particular effort was made to keep track of how many digits were actually printed in the original published article. Indeed, in contrast to our dataset where we scraped directly the pdf files of journal issues and carefully retrieved the exact coefficients and standard errors that have been printed, the dataset of Askarov et al. (2023) compiles figures from numerous meta-analyses with different research objectives. As an illustration, Figure A.12 in Appendix A shows that many coefficients (about 25%) are recorded with 15 decimal digits or more, and thus do not correspond to the rounded value published in the original article.[26]

Figure 11 shows, for each journal, the histograms of the number of sizable significant digits. First, we again find that, for the *American Economic Review*, the majority of reported coefficients (70% to 90% depending on the chosen statistical level) do not have any non-zero significant digit. In other words, we do not find evidence that the main effects of interest are on average more precisely estimated than coefficients on control variables. Second, this result holds for virtually all outlets, which suggests a high degree of consistency with our results for the *American Economic Review*. Only two journals stand out for reporting, on average, more precise coefficients, namely *Econometrica* and *Journal of the European Economic Association*. The sample size for these two outlets is significantly smaller than it

European Economic Association.

[25] *Journal of Development Economics*, *Journal of Money Credit and Banking*, *Journal of Public Economics*, *Journal of Finance*, *Journal of Financial Economics* and *Journal of Monetary Economics*.

[26] A possible explanation for this observation is that sequential data manipulations using different software may numerically introduce additional decimal digits.

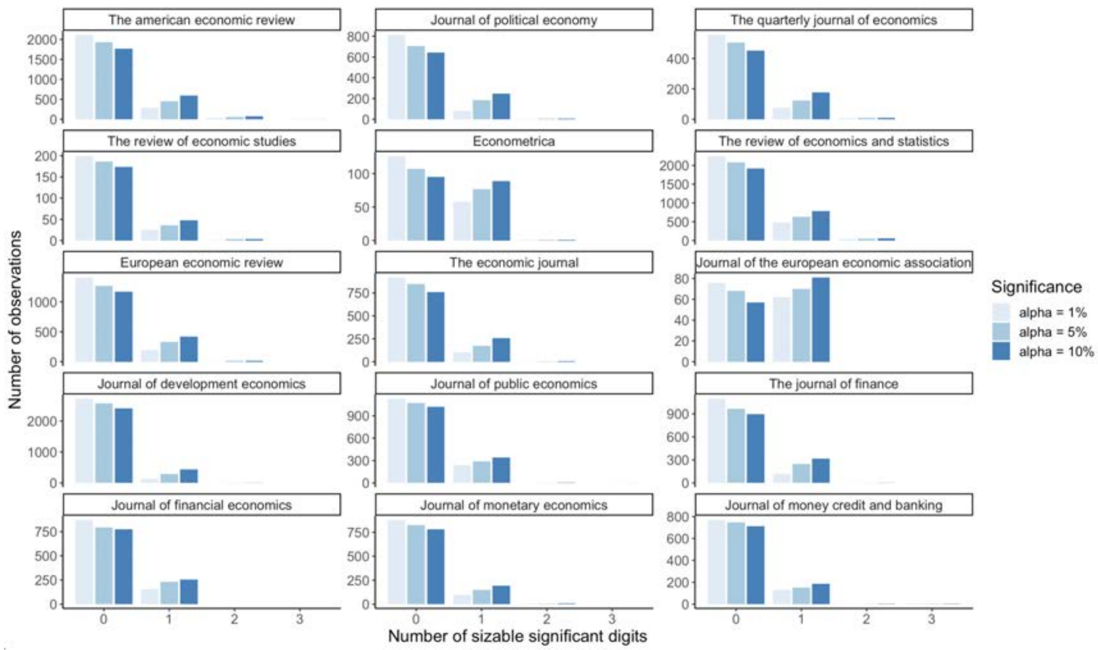

Figure 11: Histograms of the number of sizable significant digits for the different journals.

is for most other journals, which calls for great caution in interpreting this result.

To summarize, we do not find evidence that the "coefficient of interest" significant digit results we report in this section differ from those reported for our *American Economic Review* sample of all coefficient estimates.

5.5 Proposed Procedure for Reporting Statistical Precision

The example of Section 4 has shown that reporting only significant digits can lead to situations where an estimate is printed as "0" while simultaneously rejecting the null hypothesis that it is equal to zero. In addition, Section 5 found that most empirical estimates in economics may have no non-zero significant digits. Only printing significant digits would keep less than half of the digits currently printed, a fraction that includes the initial zeros in parameter estimates. This approach may therefore be a too drastic a change given the level of precision that is typically achievable by empirical work in economics.

We instead propose to adopt a rounding rule that reports estimates up to the rank of the first non-significant digit. Indeed, even such a "relaxed" rule would generally simplify results tables: in our empirical application to coefficient estimates recovered from the *American*

Economic Review, over a third of the published digits would not have been printed under this reporting rule. Besides improving reader-friendliness, we believe that adopting such a rule would increase the emphasis put on precision and economic significance relative to statistical significance.

6 Conclusion

This paper formalizes a statistical concept of "significant digit" that may be used in empirical work in economics to convey precision through the choice of which digits to print in tables of coefficient estimates. We derive a closed-form expression for the rank of the last significant digit for a coefficient estimate in the case of the normal linear conditional mean model. This expression relies on readily available information from econometric estimation packages and can be added seamlessly to any existing post-processing routine used in practice to export empirical results.

We first illustrate our approach on a simple example for which the data generating process is known. We then assemble a dataset composed of the universe of estimated coefficients and their associated standard error estimates published in the *American Economic Review* between 2000 and 2022 (over 90,000 observations). Applying our framework to this dataset, we show that over 60% of printed digits are not statistically significant at standard values of α.

Based on these exercises, we suggest that a compromise solution to current approach to reporting parameter estimates is to report digits up to the first non-significant one.

Using our concept of significant digits as a basis for reporting parameter estimates in economics would thus represent a dramatic change relative to current practice. Reporting digits up to the first non-significant one would be a substantial improvement relative to current practice for a number of reasons. First, it would increase the information content of tables of parameter estimates by not reporting digits that lack statistical significance, and in relatively rare instances, failing to report precisely estimated digits. A greater focus on the economic significance of the coefficient estimates is likely to ensue. Second, it would avoid conveying unwarranted precision in parameter estimates which could be an additional safeguard against subsequent misuse of empirical results in drawing economic conclusions from these coefficient estimates.

References

Andrews, Isaiah and Maximilian Kasy, "Identification of and correction for publication bias," *American Economic Review*, 2019, *109* (8), 2766–94.

Angrist, Joshua D and Jörn-Steffen Pischke, "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of economic perspectives*, 2010, *24* (2), 3–30.

Askarov, Zohid, Anthony Doucouliagos, Hristos Doucouliagos, and TD Stanley, "The significance of data-sharing policy," *Journal of the European Economic Association*, 2023, *21* (3), 1191–1226.

Berger, Roger L, "Multiparameter hypothesis testing and acceptance sampling," *Technometrics*, 1982, *24* (4), 295–300.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes, "Methods matter: P-hacking and publication bias in causal analysis in economics," *American Economic Review*, 2020, *110* (11), 3634–60.

_ , _ , and _ , "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics: Reply," *American Economic Review*, 2022, *112* (9), 3137–39.

_ , Scott Carrell, David Figlio, and Lester Lusher, "Unpacking P-hacking and Publication Bias," *American Economic Review*, November 2023, *113* (11), 2974–3002.

Brown, Lawrence D, JT Gene Hwang, and Axel Munk, "An unbiased test for the bioequivalence problem," *The Annals of Statistics*, 1997, *25* (6), 2345–2367.

DeLong, J Bradford and Kevin Lang, "Are all economic hypotheses false?," *Journal of Political Economy*, 1992, *100* (6), 1257–1272.

Elliott, Graham, Nikolay Kudrin, and Kaspar Wüthrich, "Detecting p-hacking," *Econometrica*, 2022, *90* (2), 887–906.

Franco, Annie, Neil Malhotra, and Gabor Simonovits, "Publication bias in the social sciences: Unlocking the file drawer," *Science*, 2014, *345* (6203), 1502–1505.

Hendry, David F., "Econometrics-Alchemy or Science?," *Economica*, 1980, *47* (188), 387–406.

Hlavac, Marek, *Stargazer: Well-Formatted Regression and Summary Statistics Tables* Social Policy Institute 2022. R package version 5.2.3.

Keane, Michael P, "A structural perspective on the experimentalist school," *Journal of Economic Perspectives*, 2010, *24* (2), 47–58.

Kranz, Sebastian and Peter Pütz, "Methods matter: P-hacking and publication bias in causal analysis in economics: Comment," *American Economic Review*, 2022, *112* (9), 3124–3136.

Leamer, Edward E, "Let's take the con out of econometrics," *American Economic Review*, 1983, *73* (1), 31–43.

_ , "Tantalus on the Road to Asymptopia," *Journal of Economic Perspectives*, 2010, *24* (2), 31–46.

Manski, Charles F, "Policy analysis with incredible certitude," *The Economic Journal*, 2011, *121* (554), F261–F289.

_ , "Communicating uncertainty in policy analysis," *Proceedings of the National Academy of Sciences*, 2019, *116* (16), 7634–7641.

McCloskey, Deirdre N and Stephen T Ziliak, "The standard error of regressions," *Journal of economic literature*, 1996, *34* (1), 97–114.

McCullough, Bruce D and Hrishikesh D Vinod, "The numerical reliability of econometric software," *Journal of Economic Literature*, 1999, *37* (2), 633–665.

Nevo, Aviv and Michael D Whinston, "Taking the dogma out of econometrics: Structural modeling and credible inference," *Journal of Economic Perspectives*, 2010, *24* (2), 69–82.

Savin, Nathan E, "Multiple hypothesis testing," *Handbook of econometrics*, 1984, *2*, 827–879.

Sims, Christopher A, "But economics is not an experimental science," *Journal of Economic Perspectives*, 2010, *24* (2), 59–68.

Stock, James H, "The other transformation in econometric practice: Robust tools for inference," *Journal of Economic Perspectives*, 2010, *24* (2), 83–94.

Wellek, Stefan, *Testing statistical hypotheses of equivalence and noninferiority*, CRC press, 2010.

Ziliak, Stephen T and Deirdre N McCloskey, "Size matters: the standard error of regressions in the American Economic Review," *The Journal of Socio-Economics*, 2004, *33* (5), 527–546.

A Supplementary Tables and Figures

Table A.1 is identical to Panel (c) in Figure 2 except for the fact that the units of X_1, X_2 and X_3 have been divided by 1,000. As a result, coefficient estimates are multiplied by 1,000. The use of scientific notation is then required to implement the reporting rule we propose.

Table A.1: Table of results under the approach of reporting digits up to the first non-significant digit ($\alpha = 0.1$) when the units for X_1, X_2 and X_3 have been divided by 1,000 (e.g. moving from metric tons to kg).

	Model 1	Model 2	Model 3
X_1	7e+02	8e+02	7e+02
	(6.8e+02)	(6.2e+02)	(3.7e+02)
X_2		9e+02	1.0e+03
		(2.6e+02)	(1.5e+02)
X_3			1.0e+03
			(1e+02)
Constant	0.9	0.7	0.8
	(0.58)	(0.54)	(0.32)
Observations	60	60	60
R^2	0.017	0.187	0.724

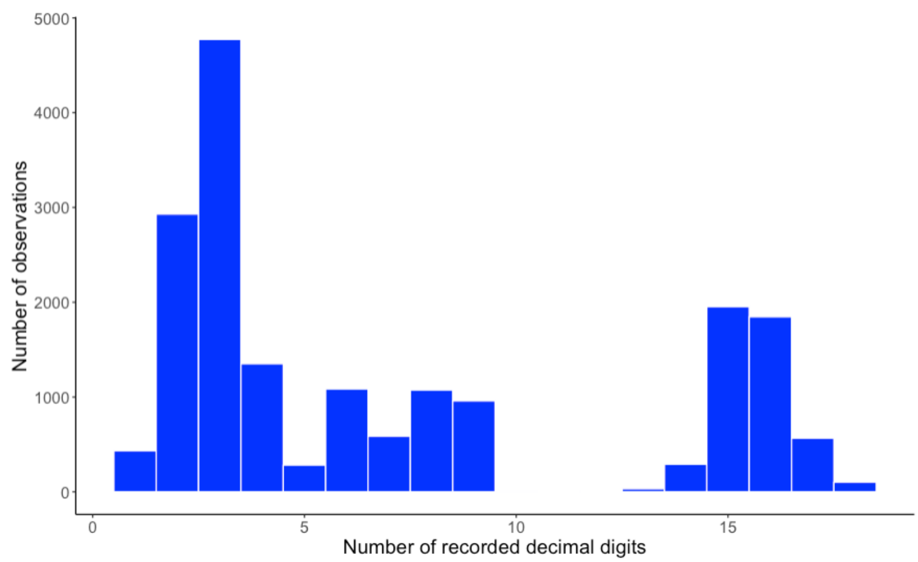

Figure A.12: Histogram of the number of recorded decimal digits for the “effect size” in the subset of the dataset of Askarov et al. (2023) used in Section 5.

B Propositions and Proofs

B.1 Rank of the Last Significant Digit for Coefficient Estimates

Proposition 1 (Coefficients) *Let $t_{1-\alpha}(T-K)$ denote the critical value of the Student's t-distribution with $T-K$ degrees of freedom for the level of significance α. For the normal linear conditional mean model, the rank of the last significant digit at level α for the estimated coefficient b_k is:*

$$i^*(\alpha) = \lceil \frac{\ln SE(b_k) + \ln t_{1-\alpha}(T-K)}{\ln 10} \rceil$$

where $\lceil x \rceil$ denotes the ceiling of x (closest integer value from above).

Proof. Let $b_k \equiv \sum_{i=-\infty}^{D} d_i 10^i$ be a sample estimate of the population parameter β_k, the k-th element of the vector of independent variables in a normal linear regression model.

Let i be a given integer. In order to determine if d_i is a significant digit, we apply Definition 1, using the TOST as our equivalence test.

First, we consider the test ψ_{ia}:

$$H_a^i\colon\ \beta_k = b_k - 10^i \qquad \text{versus} \qquad K_a^i\colon\ \beta_k > b_k - 10^i$$

For the ordinary least squares estimate of the normal linear conditional mean model, the t-statistic has a Student's t-distribution with $T-K$ degrees of freedom, where T is equal to the sample size and K is equal to the number of regressors.

Therefore, the null hypothesis of our one-sided test is rejected at level α if, and only if:

$$\frac{b_k - (b_k - 10^i)}{SE(b_k)} \geq t_{1-\alpha}(T-K)$$

which may be rewritten as:

$$10^i - t_{1-\alpha}(T-K)SE(b_k) \geq 0$$

Second, we consider the test ψ_{ib}:

$$H_b^i\colon\ \beta_k = b_k + 10^i \qquad \text{versus} \qquad K_b^i\colon\ \beta_k < b_k + 10^i$$

The null hypothesis of our one-tailed test is then rejected at a level α if, and only if:

$$\frac{b_k - (b_k + 10^i)}{SE(b_k)} \leq t_\alpha(T-K)$$

By symmetry of the Student's t-distribution, we have $t_\alpha(T-K) = -t_{1-\alpha}(T-K)$ so that the null hypothesis is rejected at the level α if, and only if:

$$10^i - t_{1-\alpha}(T-K)SE(b_k) \geq 0$$

Therefore, both tests ψ_{ia} and ψ_{ib} reject their null hypothesis if, and only if:

$$10^i - t_{1-\alpha}(T-K)SE(b_k) \geq 0$$

This condition may be rewritten:

$$i \geq \frac{\ln SE(b_k) + \ln t_{1-\alpha}(T-K)}{\ln 10}$$

By Definition 2, the decimal rank $i^*(\alpha)$ such that:

$$i^*(\alpha) - 1 < \frac{\ln SE(b_k) + \ln t_{1-\alpha}(T-K)}{\ln 10} \leq i^*(\alpha)$$

is thus the decimal rank of the last significant digit. This expression is precisely the definition of the ceiling of $\frac{\ln SE(b_k)+\ln t_{1-\alpha}(T-K)}{\ln 10}$. ■

B.2 Asymptotic Bound on Rank of the Last Significant Digit for Coefficient Estimates

Corollary 1 (Asymptotic bound) *If we denote:*

$$i^*_\infty(\alpha) \equiv \lceil \frac{\ln SE(b_k) + \ln Z_{1-\alpha}}{\ln 10} \rceil$$

then:

$$i < i^*_\infty(\alpha) \Rightarrow d_i \textit{ is not a significant digit at level } \alpha$$

Proof. We have:

$$i^*_\infty(\alpha) \equiv \lceil \frac{\ln SE(b_k) + \ln t_{1-\alpha}(\infty)}{\ln 10} \rceil \leq \lceil \frac{\ln SE(b_k) + \ln t_{1-\alpha}(T-K)}{\ln 10} \rceil \equiv i^*(\alpha)$$

where $t_{1-\alpha}(\infty) \equiv Z_{1-\alpha}$, the $1-\alpha$ percentile of a standard normal distribution. Therefore, if a digit d_i is such that $i < i^*_\infty(\alpha)$, then we also have $i < i^*(\alpha)$. Therefore, by Proposition 1, d_i is not a significant digit at level α. ■

B.3 Rank of the Last Significant Digit for Standard Error Estimate

Proposition 2 (Standard error estimates) *Let $\chi^2_\alpha(T-K)$ denote the critical value of the χ^2 distribution with $T-K>1$ degrees of freedom for the level of significance α. A lower bound for the rank of the last significant digit at level α for the estimated standard error $SE(b_k)$ is given by:*

$$i^* = \left\lceil \frac{\ln SE(b_k) + \ln\left(\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}} - 1\right)}{\ln 10} \right\rceil$$

where $\lceil x \rceil$ denotes the ceiling of x.

Proof. Let $SE(b_k) \equiv \sum_{i=-\infty}^{D} d_i 10^i$ be a standard error estimate whose (unknown) population value is se_k, and $i \leq D$ be a given integer.

First, we consider the test ψ_{ia}:

$$H_a^i\colon\ se_k = SE(b_k) - 10^i \qquad \text{versus} \qquad K_a^i\colon\ se_k > SE(b_k) - 10^i$$

The null hypothesis of our one-sided test is thus rejected at the level α if, and only if:

$$(T-K)\frac{SE(b_k)^2}{(SE(b_k)-10^i)^2} \geq \chi^2_{1-\alpha}(T-K)$$

that is:

$$\frac{T-K}{\chi^2_{1-\alpha}(T-K)} \geq \left(1 - \frac{10^i}{SE(b_k)}\right)^2$$

and thus (taking the positive root[27] for the right-hand side):

$$\sqrt{\frac{T-K}{\chi^2_{1-\alpha}(T-K)}} - 1 + \frac{10^i}{SE(b_k)} \geq 0$$

Second, we consider the test ψ_{ib}:

$$H_b^i\colon\ se_k = SE(b_k) + 10^i \qquad \text{versus} \qquad K_b^i\colon\ se_k < SE(b_k) + 10^i$$

The null hypothesis of this one-sided test is rejected at the level α if, and only if:

[27] Using the fact that $i \leq D$ and thus that $SE(b_k) \geq 10^i$. When $SE(b_k) - 10^i < 0$, the test ψ_{ia} becomes irrelevant since the true standard error is known to be non-negative.

$$(T-K)\frac{SE(b_k)^2}{(SE(b_k)+10^i)^2} \leq \chi^2_\alpha(T-K)$$

that is:

$$\frac{T-K}{\chi^2_\alpha(T-K)} \leq \left(1+\frac{10^i}{SE(b_k)}\right)^2$$

and thus:

$$1-\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}}+\frac{10^i}{SE(b_k)} \geq 0$$

Therefore, both tests ψ_{ia} and ψ_{ib} reject their null hypothesis if:

$$\sqrt{\frac{T-K}{\chi^2_{1-\alpha}(T-K)}}-1+\frac{10^i}{SE(b_k)} \geq 0 \text{ and } 1-\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}}+\frac{10^i}{SE(b_k)} \geq 0$$

Hence, a sufficient condition for both tests to reject their null hypothesis is:

$$\min\left(\sqrt{\frac{T-K}{\chi^2_{1-\alpha}(T-K)}}-1;1-\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}}\right)+\frac{10^i}{SE(b_k)} \geq 0$$

One can however prove the following Lemma (see proof below):

Lemma 1 *For all $T-K>1$ and α we have:*

$$1-\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}} \leq \sqrt{\frac{T-K}{\chi^2_{1-\alpha}(T-K)}}-1$$

A sufficient condition for digit i being to be non-significant is therefore:

$$1-\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}}+\frac{10^i}{SE(b_k)} \geq 0$$

which may be rewritten:

$$i \geq \frac{\ln SE(b_k)+\ln\left(\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}}-1\right)}{\ln 10}$$

Finally, by Definition 2, the decimal rank $i^*(\alpha)$ such that:

$$i^*(\alpha)-1 < \frac{\ln SE(b_k)+\ln\left(\sqrt{\frac{T-K}{\chi^2_\alpha(T-K)}}-1\right)}{\ln 10} \leq i^*(\alpha)$$

is thus a lower bound of the decimal rank of the last significant digit. ■

B.4 Proof of Lemma 1

We want to show that, for all $k > 1$, α:

$$1 - \sqrt{\frac{k}{\chi^2_\alpha(k)}} \leq \sqrt{\frac{k}{\chi^2_{1-\alpha}(k)}} - 1 \tag{11}$$

that is:

$$\frac{2}{\frac{1}{\sqrt{\frac{\chi^2_\alpha(k)}{k}}} + \frac{1}{\sqrt{\frac{\chi^2_{1-\alpha}(k)}{k}}}} \leq 1$$

Because the harmonic mean is weakly lower that the geometric mean, we have:

$$\frac{2}{\frac{1}{\sqrt{\frac{\chi^2_\alpha(k)}{k}}} + \frac{1}{\sqrt{\frac{\chi^2_{1-\alpha}(k)}{k}}}} \leq \left(\frac{\chi^2_\alpha(k)\chi^2_{1-\alpha}(k)}{k^2}\right)^{\frac{1}{4}}$$

Therefore, in order to prove Lemma 1, it is sufficient to show that:

$$\text{For all } k,\ \alpha,\ \frac{\chi^2_\alpha(k)\chi^2_{1-\alpha}(k)}{k^2} \leq 1$$

Let f denote the probability distribution function of the chi-squared distribution,[28] that is:

$$f(x) \equiv \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{\frac{x}{2}}$$

By definition of critical levels, we have:

$$\int_{\chi^2_{\frac{1}{2}}(k)}^{\chi^2_{1-\alpha}(k)} f(x)dx = \int_{\chi^2_\alpha(k)}^{\chi^2_{\frac{1}{2}}(k)} f(x)dx = \frac{1}{2} - \alpha$$

Using the change of variables $t = x/\chi^2_{\frac{1}{2}}(k)$ (for the left term) and $t = \chi^2_{\frac{1}{2}}(k)/x$ (for the middle term), we get:

$$\int_1^{\frac{\chi^2_{1-\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)}} f(\chi^2_{\frac{1}{2}}(k)t)dt = \int_1^{\frac{\chi^2_{\frac{1}{2}}(k)}{\chi^2_\alpha(k)}} \frac{1}{t^2} f(\frac{\chi^2_{\frac{1}{2}}(k)}{t})dt = \frac{1}{\chi^2_{\frac{1}{2}}(k)}\left(\frac{1}{2} - \alpha\right) \tag{12}$$

We then define the function $\Delta(x)$ for $x \geq 0$ as follows:

[28] We do not explicitly mention the number of degrees of freedom as a parameter to simplify notations.

$$\Delta(x) \equiv \int_1^x \left[f(\chi^2_{\frac{1}{2}}(k)t) - \frac{1}{t^2} f(\frac{\chi^2_{\frac{1}{2}}(k)}{t}) \right] dt$$

We have:

$$\Delta(1) = 0 \text{ and } \lim_{x \to +\infty} \Delta(x) = 0$$

where the second equality is obtained by letting α go to zero in Equation 12. In addition:

$$\Delta'(x) \equiv f(\chi^2_{\frac{1}{2}}(k)x) - \frac{1}{x^2} f(\frac{\chi^2_{\frac{1}{2}}(k)}{x})$$

so that we have:

$$\Delta'(x) \geq 0 \text{ if, and only if, } r(x) \equiv \frac{x^2 f(\chi^2_{\frac{1}{2}}(k)x)}{f(\frac{\chi^2_{\frac{1}{2}}(k)}{x})} \geq 1$$

Given the expression of the pdf of the chi-squared distribution, we have:

$$r(x) = x^k \exp\left[\frac{\chi^2_{\frac{1}{2}}(k)}{2} \left(\frac{1}{x} - x \right) \right]$$

Taking the derivative of $r(.)$, and using the facts that (i) the median $\chi^2_{\frac{1}{2}}(k)$ of the chi-squared distribution is lower than its mean k, and that (ii) $\frac{k}{\chi^2_{\frac{1}{2}}(k)} \leq 2$ for $k > 1$, it is straightforward to show that:

$$r'(x) = \begin{cases} > 0 & \text{for } x \in \left[1, \frac{k}{\chi^2_{\frac{1}{2}}(k)} + \sqrt{\left(\frac{k}{\chi^2_{\frac{1}{2}}(k)} \right)^2 - 1} \right[\\ = 0 & \text{for } x = \frac{k}{\chi^2_{\frac{1}{2}}(k)} + \sqrt{\left(\frac{k}{\chi^2_{\frac{1}{2}}(k)} \right)^2 - 1} \\ < 0 & \text{for } x \in \left(\frac{k}{\chi^2_{\frac{1}{2}}(k)} + \sqrt{\left(\frac{k}{\chi^2_{\frac{1}{2}}(k)} \right)^2 - 1}, +\infty \right) \end{cases}$$

Since $r(1) = 1$, the fact that $r(.)$ increases and then decreases implies that there exists a threshold value $\hat{x}$ such that:

$$r(x) \geq 1 \text{ if, and only if, } x \leq \hat{x}$$

Therefore, we have:

$$\Delta'(x) = \begin{cases} \geq 0 & \text{for } x \in [1, \hat{x}] \\ < 0 & \text{for } x \in (\hat{x}, +\infty) \end{cases}$$

Since $\Delta(1) = 0$ and $\lim_{x \to +\infty} \Delta(x) = 0$, the fact that Δ first increases and then decreases imply that:

$$\text{For all } x,\ \Delta(x) \geq 0$$

To complete the proof, we note that, from Equation 12, we have:

$$\Delta\left(\frac{\chi^2_{1-\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)}\right) = \int_{\frac{\chi^2_{1-\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)}}^{\frac{\chi^2_{\frac{1}{2}}(k)}{\chi^2_{\alpha}(k)}} \frac{1}{t^2} f\left(\frac{\chi^2_{\frac{1}{2}}(k)}{t}\right) dt$$

Since $\Delta \geq 0$ and the integrand of the right-hand-side term is non-negative, we have:

$$\frac{\chi^2_{1-\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)} \leq \frac{\chi^2_{\frac{1}{2}}(k)}{\chi^2_{\alpha}(k)}$$

and thus:

$$\frac{\chi^2_{1-\alpha}(k)\chi^2_{\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)^2} \leq 1$$

Finally, since $\chi^2_{\frac{1}{2}}(k) \leq k$, we have:

$$\frac{\chi^2_{1-\alpha}(k)\chi^2_{\alpha}(k)}{k^2} \leq \frac{\chi^2_{1-\alpha}(k)\chi^2_{\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)^2} \leq 1$$

which completes the proof.

C Details on Data Extraction

Our input data are issues of the *American Economic Review* in pdf format. We use the R package "pdftools" to convert these pdf files into a list of pages, where each pdf page is broken down into a list of individual strings, defined as any sequence of characters preceded and followed by a space/tab (a word, a number, etc.). These strings are spatially located on a 2D Cartesian grid.

We then apply the following three-step approach to each page in order to extract coefficient estimates and their associated standard error estimates:

- **Step 1:** Determine whether the page is likely to contain a table reporting estimation results;
- **Step 2:** If it is, then locate and extract candidate coefficient estimates and their associated standard errors estimates;
- **Step 3:** Clean the obtained data.

Steps 1 & 2 build on the observation that a coefficient estimate b_k and its associated standard error $SE(b_k)$ are reported in a results table as $b_k\,(SE(b_k))$ where the standard error $(SE(b_k))$ of coefficient b_k is located directly below or directly to the right of b_k. The odds of having a number directly followed (either to the right or just below) by a number into brackets are very low in other context than results tables. If a page contains several occurrences of such a pattern, it is thus very likely to display a results table. We detail below how we implement this intuition in practice.

Step 1: Is the page likely to contain a results table?

A given page is deemed likely to contain a results table if the following conditions are met:

1. the page contains the word "Table";
2. the page contains at least four numbers (i.e. a numeric preceded and followed by either a tab or a space) and four numbers into brackets;
3. the page contains one of the strings ("Error", "error","(SE)") OR contains the words "treatment"/"Treatment" and "effect"/"Effect".

A page that does not meet these conditions is screened out. Otherwise, the routine continues to the next step.

Step 2: Extract coefficients and standard errors

When a given page is deemed in Step 1 to be likely to contain a table of estimation results, we run a loop on all the strings that compose it. For each string S:

1. we test whether it is a number, that is if "as.numeric(S)" does not return NA (removing, when relevant, the "," for thousands and the terminating "*", "†", "‡","a", "b" or "c" indicating statistical significance).

2. if it is, we retrieve:

   - the closest neighbor string to the right (if any within a distance of 20 units);
   - the closest neighbor string below (if any within a distance of 15 units).

3. for each of the (up to) two neighbors of string S, we check whether it has the format of a standard error, meaning:

   - it includes the regular expression "\\([0-9]*.*[0-9]*\\)"
   - it can be converted to a number after removing "(" and ")", along with any "," or"*", "†" or "‡".

4. if one of the two neighbor strings meets this format, we store the string S and its neighbor string as a pair (coefficient estimate, standard error).

Step 3: Data cleaning

Having applied steps 1 and 2 to all pdf pages, we perform several data cleaning steps:

1. we drop the pairs (coefficient estimate, standard error) which are the only observation found in a given page. A manual check of the corresponding observations indeed confirms that are false positive.

2. several results tables report coefficients and their standard errors along with statistical tests and their test values. We remove manually the latter observations we were able to detect by looking at the obtained dataset (about 300 observations removed).

3. when columns in results tables are labelled as (1), (2) etc. and the above text label is a number (e.g. a date), the resulting combination may be mistaken for a coefficient and its standard error. Similarly, a number of false positives occur when the text around a result table include reference to model specifications (e.g. "specification (1) shows that etc.") or literature references (e.g. "according to John et al. (1900) etc."). We remove such outliers by checking manually observations for which the standard error (in string format) is an integer between 1 and 10 (for specifications) and between 1700 and 2050 (for dates). We remove about 1,100 false positives through this procedure.

4. a number of results tables report p-values or t-values rather than standard errors. We remove manually the occurrences of such cases we were able to detect by looking at the obtained dataset (about 500 observations removed).

5. when (manually) identified, tables presenting summary statistics without hypothesis testing, as well as idiosyncatric formatting (e.g. percentage points given into brackets), are removed (about 600 observations removed).

6. finally, in all issues between the 4th issue of 2007 and the 4th issue of 2009, the "–" sign is read as a "2" by our pdf scraper.[29] We manually correct the corresponding coefficients (about 1,100 observations corrected).

Overall, our manual data cleaning corrects the value over $1,000$ coefficients and removes almost $3,000$ false positives out of $98,043$ initial observations in the raw dataset coming out of the scraping procedure. While we do not claim that our final dataset is completely error free,[30] our data extraction procedure is entirely automated and transparent. It can therefore be easily replicated and improved.

[29] The issue carries over when copy-pasting the numbers from the original, which suggests it may come from the encoding of the pdf files rather than from our scraper.

[30] For example, our manual inspections revealed the existence of some false negatives. Such missed observations however correspond to situations that would be very complicated to scrape using simple automated routines. For example, square brackets are most often used to report p-values but are sometimes used for standard errors as well. Such estimates are not recovered by our scraping routine. We however have no particular reason to believe that false negatives represent a large and selected population.