An Equivalence Test Approach to Rounding Parameter Estimates*

Nicolas Astier† Frank A. Wolak‡

March 26, 2026

Abstract

We address a question faced by empirical economists almost on a daily basis: How to report the results from statistical analyses? We first document current practice by compiling the universe of coefficient estimates and their associated standard errors from all articles published in the *American Economic Review* (AER) between 2000 and 2022. We find that the overwhelming majority of articles rely on a simple heuristic: always report a fixed number of decimal digits (most often three) for all parameter estimates and standard errors. We construct an alternative approach to determining the number of digits to report that builds on tests equivalence hypotheses. Specifically, we introduce the concept of a "distinct" digit, which refers to the situation where an equivalence test rejects the null hypothesis that the parameter estimate and the true value, both rounded at a given decimal rank, cannot be equal. We recommend reporting digits in parameter estimates up to the first non-distinct digit. Applying this rounding rule to our AER dataset, we find that about one-third the digits reported would not have been printed. Our rounding heuristic is straightforward to implement and already embedded into a widely used R package.

Keywords: rounding, hypothesis tests of equivalence, distinct digits, statistical precision
JEL: A12, B41, C12

*A prior version of this work was circulated in February 2024 under the title "*Credible Numbers: A Procedure for Reporting Statistical Precision in Parameter Estimates.*" This project is supported by the French Government's Investment for the Future Program (PIA), now part of France 2030, operated by ADEME. The views expressed in this paper are those of the authors alone and might not represent the views of ADEME or any other institution. We are grateful to seminar participants at the Paris School of Economics, the Urban Institute, the 2024 European Association of Young Economists conference, the 2024 EEA-ESEM conference and the 2025 California Econometrics Conference for their feedback and suggestions.

†Paris School of Economics and Ecole des Ponts, 48 Bd Jourdan 75014 Paris, France, nicolas.astier@psemail.eu.

‡Program on Energy and Sustainable Development and Department of Economics, Stanford University, Stanford, CA 94305-6072, wolak@zia.stanford.edu.

1 Introduction

Empirical economists estimate coefficients from statistical models almost on a daily basis. Because "best unbiased" estimators are usually characterized by analytical formulas, the most "precise" coefficient estimate researchers could compute is an infinite series of digits, which is practically impossible to report.[1] Therefore, researchers must round parameter estimates at some decimal rank.

Historically, heuristics such as reporting only two or three decimal digits could be rationalized by high computational costs and limited numerical reliability. Indeed, the same estimator applied to the same data could produce different exact values for parameter estimates depending on the software package or the computer used to perform the calculations.[2] Nowadays, however, the same estimation routine applied to the same data on different computers (or on the same computer using different software packages) is expected to yield identical estimates up to a large number of decimal digits. In this context, should empirical researchers report their coefficient estimates with as many decimal digits as possible? If not, is there a principled approach to rounding parameter estimates?

Choosing how many decimal digits to report for parameter estimates does not currently appear to be a critical issue that empirical researchers face. When retrieving the universe of coefficient and standard error estimates published in the *American Economic Review* (AER) between 2000–2022 (see below), we find that the vast majority of researchers always report two or three decimal digits in their papers, suggesting that they delegate the decision of how to round parameter estimates to the default setting of the software routine they are using.

While it may be expedient for them not to invest too much time on this issue, it may lead empirical economists to report either "too many" digits, perhaps fueling over-confidence in available precision and/or hampering the credibility of reported estimates, or "too few" digits, when available precision is actually high. As a striking example of this second situation, we observe that, under current rounding heuristics, about 1% of the coefficients printed in the AER between 2000–2022 are reported with a standard error of exactly zero.

[1] A parameter estimate, $\hat{\theta}$, can be expressed as $\sum_{i=-\infty}^{D} \hat{d}_i 10^i$, where each $\hat{d}_i$ takes on values in {0,1,2,...,9}. We call the integers i (positive or negative) "decimal ranks."

[2] Tables 1 and 2 of McCullough and Vinod (1999) provide examples of this phenomenon.

Could such shortcomings be avoided by setting a different but equally simple default rounding heuristic in software routines? If so, even if these issues might be perceived as minor, what would be the rationale for not fixing them?

This article proposes a general approach to rounding parameter estimates that relies on equivalence testing (Wellek, 2010). Specifically, equivalence tests posit as a null hypothesis the statement that the true parameter value lies outside of a given interval.[3] Rejection of this null hypothesis provides evidence that this parameter value lies in this interval. Our proposal builds on the observation that the decision of whether or not to report a given digit in a parameter estimate should be informed by the likelihood that this digit is estimated with sufficient precision. Equivalence tests precisely address this question.

We define a given digit as "precisely estimated", or "distinct," if one can reject the null hypothesis that the parameter estimate and the true value, both rounded at the decimal rank of the considered digit, cannot be equal. To make this definition more concrete, consider two numbers a and b and a decimal rank i. If $|a - b| < 10^i$, it is then possible, depending on the exact values of a and b, that the rounded values of a and b at rank i are equal. For example, for $a = 1.250001$, $b = 1.349999$ and $i = -1$, the absolute difference between a and b is very close to $10^{-1} = 0.1$, and both a and b rounded at the decimal rank $i = -1$ are equal to 1.3.[4] Conversely, if $|a - b| \geq 10^i$, the rounded values of a and b at decimal rank i are necessarily different. Therefore, the rounded values of two randomly chosen numbers a and b at decimal rank i cannot be equal if, and only if, $|a - b| \geq 10^i$. Applying this property to the estimate $\hat{\theta}$ of a quantity θ, the null hypothesis that the rounded values of both the estimate and the true value of a parameter rounded at decimal rank i are equal corresponds to $|\hat{\theta} - \theta| \geq 10^i$. We define the digit of decimal rank i as "distinct" if an equivalence test rejects this hypothesis.

Loosely speaking, for each decimal digit in a parameter estimate, our procedure asks whether the null hypothesis that this digit is one value higher or lower is rejected. For example, for a parameter estimate equal to 1.234, can the equivalence test with the null

[3]Equivalence testing is often used in biostatistics determine whether two treatments are not significantly different (Wellek, 2010).

[4]Note that the difference between a and b can be made arbitrarily close to 0.1 with slightly different choices of a and b.

hypothesis that the true value of the coefficient is 0.234 or more or 2.234 or less be rejected? Based on the above discussion, although this test is centered on the parameter estimate, it actually only hinges on the pre-specified length ± 1 of the equivalence interval. More precisely, one could think of this test as asking the question: given a pre-specified value of the parameter, can we reject the null hypothesis that the digit of a given decimal rank is one value higher or lower? The pre-specified value that is least likely to reject the null is the empirical estimate of the parameter, so that applying an equivalence test centered on the parameter estimate makes sure that, if the null hypothesis of this test is rejected, then the null hypothesis of any alternative test with the same interval length but centered at a different pre-specified value would have been rejected too.

If the null hypothesis of the equivalence test of length ± 1 is rejected, we define the digit "1" in 1.234 as "distinct," that is, as likely (although not certain) to be equal to the estimated value, in this case 1. As one continues scanning digits to the right, that is, tests equivalence intervals of length ± 0.1, ± 0.01, etc., there will exist some decimal rank after which estimated digits are no longer distinct. Suppose for example that the digit "2" in 1.234 is found not to be distinct. Our proposal is then to round parameter estimates at the rank of their first non-distinct digit, that is, to report 1.2 as the parameter estimate.

How much "precision" is lost by rounding parameter estimates at the decimal rank of their first non-distinct digit? Figure 1 illustrates sampling uncertainty by drawing many datasets from a known data generating process where the true value of a parameter of interest is equal to 1.234 (see Section 4 for more detail). Specifically, it shows the histograms, over one thousand data draws for two different sample sizes (T=100 and T=1,000,000), of the digits of the parameter estimate for the decimal ranks 0 to -3. The empirical distribution of the first non-digit distinct is observed to have more mass around its population value, the subsequent digits are essentially drawn from $\{0, 1, 2, ..., 9\}$ according to a uniform distribution. Therefore, not reporting them, that is, reporting 1.2 when the sample size is equal to 100 observations, boils down to truncating a tail of decimal digits that, based on sampling uncertainty alone, essentially consists of pure statistical noise. In particular, a parameter estimate that is statistically different from zero will by construction be reported with at least one non-zero digit. Conversely, if a parameter estimate is found to have at least

one non-zero distinct digit, it is then statistically different from zero.

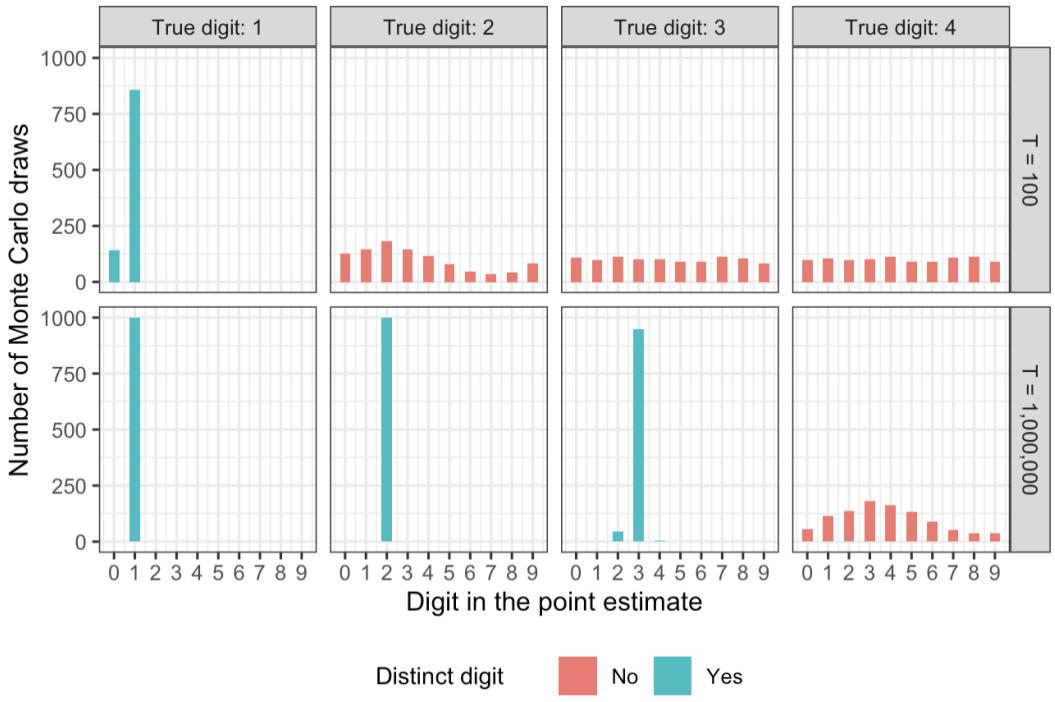


Figure 1: Results of Monte Carlo simulations (using Model 1 of Section 4). Each row corresponds to a different sample size ($T \in \{100; 1,000,000\}$). Each column corresponds to a given digit of the point estimate $\hat{\beta}_1$ of the price elasticity. Denoting $\hat{\beta}_1 = \hat{d}_0 + 0.1\hat{d}_{-1} + 0.01\hat{d}_{-2} + 0.001\hat{d}_{-3} + \ldots$, the columns report respectively (from left to right) the distributions of $\hat{d}_0$, $\hat{d}_{-1}$, $\hat{d}_{-2}$ and $\hat{d}_{-3}$. The corresponding values of these digits in the true parameter are respectively 1, 2, 3 and 4. See Section 4 for more detail.

Can equivalence tests be applied in a tractable way in practice?[5] We derive an exact closed-form expression to characterize the decimal rank of the first non-distinct digit for the normal linear conditional mean model, from which we derive a large sample approximation for the decimal rank that is valid for a wide range of parameter estimates. This formula only involves straightforward calculations from variables that are already computed by standard software packages. It can therefore be easily added to post-processing routines that researchers typically use to generate tables of estimation results. It has for example been included in the R package "modelsummary" (Arel-Bundock, 2022) and only requires a few lines of code to be added to any existing software routine.

[5]One brute force approach to generate Figure 1 and assess which digits to report could be to bootstrap the estimation dataset. However, such an approach would prove very computationally intensive in many applications.

While the practical implementation of our rounding heuristic relies on the asymptotic approximation that parameter estimates follow a normal distribution, equivalence tests provide a statistical framework that remains valid for arbitrary distributions of parameter estimates. For example, we derive in the Appendix the rank of the first non-distinct digit of standard error estimates for the normal linear conditional mean model.

Contrary to most parameter estimates, however, the use value of standard errors lies in their magnitude, not their precision. Therefore, the reporting heuristic we implemented in the R package "modelsummary" rounds standard errors two decimal ranks after their first non-zero digit, irrespective of their precision.

Obviously, the same reasoning could apply to coefficient estimates: a precisely estimated digit may not always be "economically significant." In such cases, researchers may prefer to report coarser parameter estimates. The practical relevance of such situations is, however, an empirical question. In order to explore it further, we retrieve all coefficient estimates and associated standard errors printed in the AER between 2000 and 2022. We obtain a dataset of over 90,000 coefficient estimates each with its associated standard error estimate. We find that coefficient estimates tend to be imprecisely estimated: 60% of the digits reported for coefficient estimates are not distinct. If empirical researchers were to adopt the rounding rule we propose, over 33% of the published digits in our sample would not be printed. We also provide evidence from a number of other economics journals supporting 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 the parameter estimates reported in results tables. We therefore conclude that the rounding rule we propose would very likely represent a significant improvement relative to the current status quo.

The rest of the paper is organized as follows. Section 2 argues that the way precision is conveyed when reporting parameter estimates may matter, documents current practice in economics and discusses how such practices may differ in other empirical sciences. Section 3 introduces our concept of a "distinct digit" and derives closed-form expressions to characterize the rank of the first non-distinct digit. Section 4 illustrates with a numerical example why researchers may consider rounding parameter estimates to the first non-distinct digit. Finally, Section 5 studies quantitatively the extent to which current practice differs from the

rounding rule we propose. Section 6 concludes.

2 Background

2.1 Conveying Precision in Empirical Studies

Empirical results in economics are derived from identification assumptions and estimated on small samples, making them inherently uncertain (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 this context, many authors have stressed that putting too much emphasis on statistical significance may divert researchers away from discussing economic significance (DeLong and Lang, 1992; McCloskey and Ziliak, 1996; Ziliak and McCloskey, 2004). In addition, such an over-emphasis of "statistically significant" results have raised an important debate about the credibility and transparency of empirical research.[6]

As the cost of data collection has been declining, larger and larger datasets have become available, and the statistical power to detect non-zero economic magnitudes using conventional tests of statistical significance 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 different from zero?*" but instead "*how precise is a (non-zero) coefficient estimate?*". This change of perspective offers an opportunity to establish *precision* as a salient attribute of reported results. Such a practice would likely reduce the emphasis put on simply obtaining statistically significant results and highlight that uncertainty in each reported decimal digit is an essential feature of empirical work in economics.

Although a large theoretical literature on scientific communication has developed models to think about the relationship between the supply of scientific evidence and its influence

[6]See for example 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), as well as Hendry (1980) and Leamer (1983). A major concern in this literature is 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.

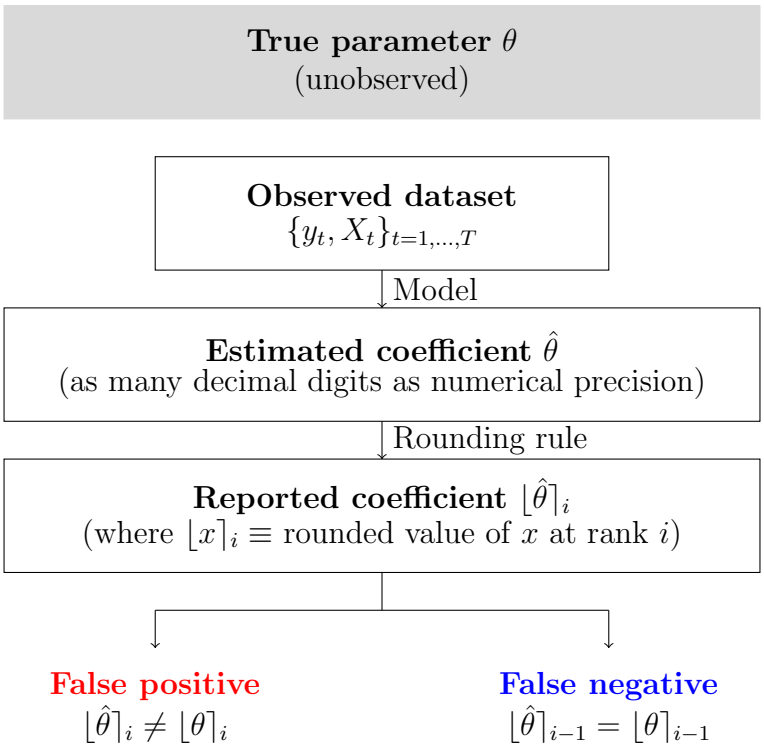


Figure 2: Possible definitions of "false positives" and "false negatives" when applying a rounding rule (i.e., choosing the rank i) to a parameter estimate.

on policy-making,[7] such considerations are beyond the scope of this article. Instead, we simply note that a rounding rule must be used in practice, and that this rule will make a given number (the reported rounded point estimate) particularly salient.[8] In this context, two questions are of particular interest (see Figure 2). First, is this rounded number indeed equal to the underlying population parameter (rounded at the same decimal rank)? Second, would the next (non-reported) digit actually prove to be accurate? Loosely speaking, the first (resp. second) question may be thought as assessing whether the reported estimate is a "false positive" (resp. "false negative"). Of course, no rounding rule will be unambiguously better along both dimensions under all circumstances. In order to be able to compare different rounding heuristics, we start by documenting current practices in economics.

[7]This literature studies for example strategic reporting by researchers (Henry and Ottaviani, 2019) or optimal reporting rules (Andrews and Shapiro, 2021).

[8]Recent work has shown that the interpretation of scientific evidence can be subject to behavioral biases. For example, Korting et al. (2023) show that the chosen graphical representation influences how readers interpret regression discontinuity graphs.

2.2 Current Practice

We retrieve all coefficient estimates and associated standard error estimates from the universe of articles published in the *American Economic Review* between 2000 and 2022.[9] 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.[10] We obtain a dataset of over 90,000 coefficients with their standard error estimates.

We start by assessing 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 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 3 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 in 2014 to slightly higher than its pre-2010 average.

We next discuss how many decimal digits researchers typically reported during our sample period. As shown in Figure 4, 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 could be the case that reporting three decimal digits is always

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

[10] 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).

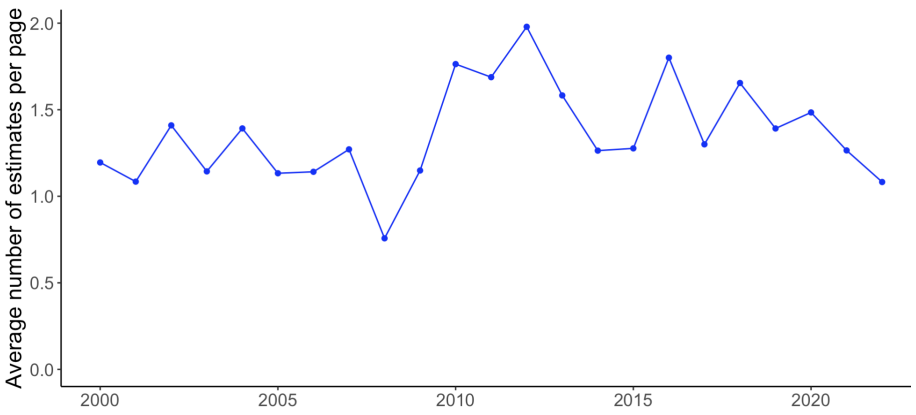


Figure 3: Evolution of the ratio of the number of reported empirical estimates in each year to the number of printed pages for that year.

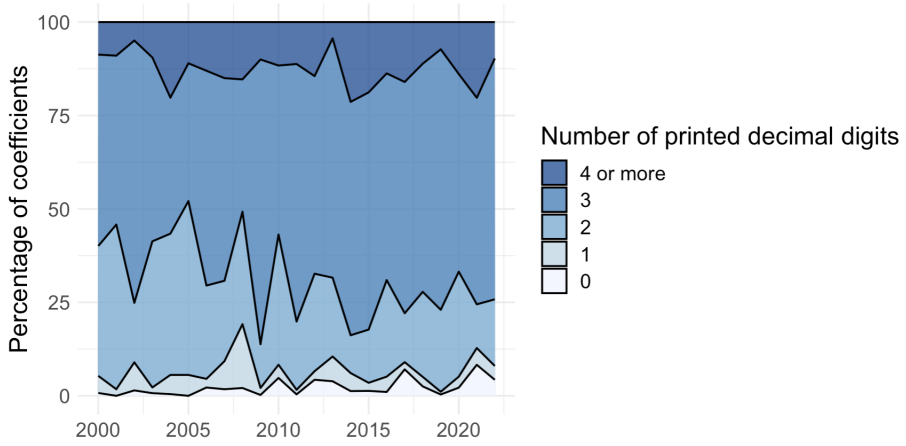


Figure 4: Evolution over time of the number of decimal digits printed for coefficient estimates.

warranted given available precision. This interpretation is, however, hard to reconcile with the observation that the majority of tables of estimation results are found to report the same number of decimal digits for all coefficient estimates.

When looking at the value of coefficients (see Figure A.11 in the Appendix), we observe that units of measurements are typically chosen such that estimated coefficients have an absolute magnitude lower than 10. In addition, we find 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 (see Figure A.12 in the Appendix).

Interestingly, about 1% of printed standard error estimates are equal to zero. In other

words, by relying on the default heuristic of always printing say three decimal digits, about 1% of coefficients are reported with a standard error of (0.000). This finding shows that the current rounding heuristic sometimes leads to reporting "too few" digits. Indeed, standard errors should, at a minimum, be printed with their first two or three non-zero digits, and a coefficient estimate whose standard error is 10^{-4} or smaller should be reported with more than three decimal digits.

Overall, we conclude that current rounding heuristics barely attempt to account for the available precision of parameter estimates, if at all.

2.3 Practices in Other Empirical Siences

In natural sciences, researchers pay particular attention to "significant digits" (also called "significant figures") when reporting empirical results. Specifically, scientists typically report decimal digits only up to the point where random errors in the measurement device prevent further confidence in the precision of the digit recorded. The rationale for this practice is that errors are inevitable in measuring a magnitude of interest. 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.

By analogy, an econometric model and its associated dataset may be thought as the measurement device employed by empirical researchers to retrieve estimates of magnitudes of interest. We thus propose a procedure to characterize what we call "distinct digits", which can be seen as a way to assess the precision of the measurement device of the econometrician and determine the number of digits to report in a parameter estimate. We favor the terminology "distinct digits" over "significant digits" in order to avoid confusing our approach with traditional significance testing in economics, which simply tests whether one call confidently rule out that a given parameter is equal to zero.

This change in focus offers an opportunity to establish the *number of distinct digits* and their values as the salient attribute of reported results. Such a practice might contribute to reduce the emphasis put on simply obtaining statistically significant results.

3 Distinct Digits in Economics

This section introduces the framework we use to assess whether a given digit is "distinct" and derives an analytical formula for the decimal rank of the first non-distinct digit in a coefficient estimate with an asymptotically normal distribution.

3.1 Equivalence Testing versus Significance Testing

Equivalence testing asks whether an effect is *small enough to be practically negligible*, rather than whether it is exactly zero, as is the case with traditional significance testing.

Let θ denote the population quantity of interest (e.g. a regression coefficient). A *standard two-sided test of statistical significance* takes the form:

$$H: \ \theta = 0 \qquad \text{(the effect is zero)}$$

$$K: \theta \neq 0 \qquad \text{(the effect is not equal to zero)}$$

Rejection of $H: \ \theta = 0$ provides evidence that θ is non-zero.

In contrast, equivalence hypotheses reverse the roles of "null" and "alternative." Specifically, let $\Delta > 0$ be a *pre-specified negligible margin* representing the largest magnitude considered practically unimportant. An equivalence test takes the form:

$$H: \ |\theta| \geq \Delta \qquad \text{(not equivalent; effect is meaningfully large)}$$

$$K: \ |\theta| < \Delta \qquad \text{(equivalent; effect is sufficiently small)}$$

Equivalence is concluded only if the data provide evidence that θ lies in the interval $(-\Delta, \Delta)$. Rejection of the non-equivalence null hypothesis $H: \ |\theta| \geq \Delta$ is that evidence.

Equivalence tests are routinely used in the context of drug effectiveness (Romano, 2005). More recently, they have been advocated for in political science (Rainey, 2014) and economics (Fitzgerald, 2024) as the correct approach to argue for negligible effects, for which the failure to reject a significance test provides insufficient evidence.

A common implementation of tests of equivalence, which we adopt in what follows, is the

two one-sided tests (TOST) approach at level α.[11] These tests reject the null hypothesis H at the level α if, and only if, the following two one-sided tests reject their null hypothesis at the level α:

$$H_L\text{: } \theta \leq -\Delta \qquad \text{versus} \qquad K_L\text{: } \theta > -\Delta \qquad (\psi_L)$$

and

$$H_U\text{: } \theta \geq \Delta \qquad \text{versus} \qquad K_U : \theta < \Delta \qquad (\psi_U)$$

Because both tests ψ_L and ψ_U are of size α, it may seem counter-intuitive that the resulting test 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, however, a special case of the so-called "intersection-union tests" (Berger, 1982). The intuition for the lack of need to adjust the size of individual tests to account for multiple testing is that H is rejected only if *both* H_L and H_U are rejected.

In practice, a major challenge hampering a widespread use of equivalence tests is the choice of, and justification for, the *equivalence margin* Δ. In the case of the "distinct digit" framework that we next introduce, a natural equivalence margin for the digit of decimal rank i is $\Delta = 10^i$. This choice is indeed anchored in the use of a decimal system to report numbers.

3.2 Equivalence Tests and Distinct Digits

Consider a scalar parameter estimate,[13] $\hat{\theta}$, decomposed into an infinite series of digits $\{\hat{d}_i\}_{i=-\infty}^{D}$, where $\hat{d}_i$ multiplied by 10^i for each i as:[14]

$$\hat{\theta} \equiv \sum_{i=-\infty}^{D} \hat{d}_i 10^i$$

[11] Although several tests besides the TOST have been proposed in the literature on equivalence testing (e.g. Brown et al. (1997)), we believe that the simplicity of implementing the TOST makes it ideally suited to our purpose.

[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)).

[13] Our framework is actually more general and can be applied to any estimated magnitude, such as for example standard error estimates. For simplicity, however, we 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 model.

[14] When $\hat{\theta} < 0$, this decomposition is preceded by a minus sign.

Econometric software packages report a limited number of decimal digits, so that the "printed" value of $\hat{\theta}$, which we denote with $\hat{\theta}_p$, is:

$$\hat{\theta} \equiv \sum_{i=i_0}^{D} \hat{d}_i 10^i$$

For $\hat{\theta}_p = 1.234$, $i_0 = -3$, $D = 0$, and $\hat{d}_0 = 1$, $\hat{d}_{-1} = 2$, $\hat{d}_{-2} = 3$ and $\hat{d}_{-3} = 4$.

As discussed above, current practice among empirical researchers is to pick a default value for i_0 for all reported parameter estimates, most commonly $i_0 = -3$.

Instead, the choice of i_0 could be coefficient-specific and chosen according to how precisely estimated this parameter is. Based on the above discussion (e.g., Figure 2), one statistically-grounded way to make this assessment is to ask the question: can we reject with a high level of confidence the null hypothesis that the estimate and the true value rounded at decimal rank i cannot be equal? Indeed, if this assumption can be rejected, then there is statistical evidence that the corresponding digit is informative about the true parameter value.

For two randomly chosen real numbers a and b, we have the following lemma:

Lemma 1 *For a given real number x, let us denote with $\lfloor x \rceil_i$ the rounded value of x at the decimal rank i. We then have:*

$$|a - b| \geq 10^i \quad \Leftrightarrow \quad \lfloor a \rceil_i \neq \lfloor b \rceil_i \text{ for all possible values of } (a, b)$$

By lemma 1, the null hypothesis "*the rounded values at decimal rank i of the estimate and the true value cannot be equal*" may be written:

$$H^i\text{: } |\theta - \hat{\theta}| \geq 10^i$$

An alternative approach to rationalize the use of this null hypothesis goes as follows. Let us set a pre-specified standard of precision, say 10^0. Ultimately, we are interested in determining whether a parameter estimate can be confidently assessed to lie within an interval $\pm 10^0$. For a given guessed value θ_0 of the parameter, this characterization corresponds to the following equivalence test:

$$H\text{: } |\theta - \theta_0| \geq 10^0 \qquad \text{versus} \qquad K\text{: } |\theta - \theta_0| < 10^0$$

Because $|\hat{\theta} - \theta_0|$ is minimized for $\theta_0 = \hat{\theta}$, the value of θ_0 for which H is least likely to be rejected is $\theta_0 = \hat{\theta}$. In other words, if H is rejected for $\theta_0 = \hat{\theta}$, then it is also rejected for any other value of θ_0.

Therefore, we consider the digit of decimal rank i to be "distinct" if an equivalence test rejects the null hypothesis H^i.

Definition 1 (Distinct digit) *Let $\hat{\theta} \equiv \sum_{i=-\infty}^{D} \hat{d}_i 10^i$ be an estimate of the scalar population parameter θ. We define digit $\hat{d}_i$ to be distinct at the level α if the following equivalence test ψ_i rejects the null hypothesis H^i at the level α:*

$$H^i\text{: } |\theta - \hat{\theta}| \geq 10^i \qquad \textit{versus} \qquad K^i\text{: } \theta \in \left(\hat{\theta} - 10^i, \hat{\theta} + 10^i\right)$$

Equivalently, the digit of decimal rank i is defined as "distinct" if:

- *the null hypothesis that the estimate and the true value rounded at decimal rank i cannot be equal is rejected at the level α.*
- *for any pre-specified value θ_0, the null hypothesis H^i: $|\theta - \theta_0| \geq 10^i$ is rejected at the level α.*

Concretely, to assess whether the digit of rank $i = 0$ is distinct, we simply perform the equivalence test:

$$H^{i=0}\text{: } |\theta - 1.234| \geq 1 \text{ vs. } K^{i=0}\text{: } \theta \in \ (0.234, 2.234)$$

Suppose that digit $\hat{d}_0 = 1$ is assessed to be "distinct" according to Definition 1. We can next turn to the next digit and perform the equivalence test ψ_{-1}:

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

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

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

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 consider in what follows. We nonetheless state it as an assumption in order to introduce in full generality the following concept of "first non-distinct digit".

Definition 2 (First Non-Distinct Digit) *Under Assumption 1, we define the rank $i^*(\alpha)$ of the "first non-distinct digit" as the largest value of i such that digit $\hat{d}_i$ is not distinct at level α according to Definition 1:*

$$i^*(\alpha) \equiv \max_i \{i \,|\, \hat{d}_i \text{ is not distinct at level } \alpha\}$$

Under Assumption 1, it is straightforward to see that, if $\hat{d}_i$ is a distinct digit, then all $\hat{d}_{i'}$ with $i' > i$ are also distinct digits. Because finite sample sizes necessarily put an upper bound on precision, there exists a threshold rank $i^*(\alpha)$ such that a digit $\hat{d}_i$ is distinct if, and only if, $i > i^*(\alpha)$. We call the corresponding digit $\hat{d}_{i^*(\alpha)}$ the "first non-distinct digit." One may interpret $i^*(\alpha)$ as a measure of "precision": loosely speaking, digits of decimal ranks $i^*(\alpha)$ and below cannot be confidently identified.

It is important to note, however, that the rank of the first non-distinct digit is sensitive to the choice of units.[15] For example, in a univariate regression, converting the unit of the explanatory variable from "meters" to "kilometers" increases the magnitude of the coefficient estimate by 10^3, therefore shifting $i^*(\alpha)$ by three. In order to obtain a notion of "precision" that does not suffer from this limitation, we introduce the concept of the "number of sizable distinct digits."

[15] We discuss the choice of units in greater length in Appendix D.

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

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

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

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

In words, the number of sizable distinct digits is the number of digits that: (i) contribute to measuring a non-zero parameter 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 digits are precisely estimated when $i^*(\alpha) = -3$. Because $D(\hat{\theta}) = 0$ for this example, the formula in Definition 3 indeed yields $\nu(\alpha) = 1$ in the former case, and $\nu(\alpha) = 4$ in the latter.

3.3 Closed-form Expressions

We now derive, for the normal linear conditional mean model, an exact closed-form expression for the rank of the first non-distinct digit of a coefficient estimate. Specifically, we consider the textbook model:

$$y = X\beta + \epsilon \tag{2}$$

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 these 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 $\mathcal{N}(\beta, \sigma^2(X'X)^{-1})$ and, if we define $s^2 \equiv \frac{e'e}{T-K}$, where $e \equiv y - Xb$, then $\frac{s^2(T-K)}{\sigma^2}$ is distributed as a χ^2 random variable with $T - K$ degrees of freedom.

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:

$$b_k \equiv a_k' b \tag{3}$$

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

$$SE(b_k) \equiv s\sqrt{a_k'(X'X)^{-1}a_k} \tag{4}$$

Using these notations, we can derive the exact closed-form expression for the level α rank of the last distinct digit for b_k.

Proposition 1 (Rank of the first non-distinct digit) *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 first non-distinct digit at level α for the estimated coefficient b_k is:*

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

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

Proof. See Appendix B. ■

For a given level of statistical significance α, computing the rank of the first non-distinct 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 first non-distinct digit of a coefficient estimate.

Moreover, the proof of Proposition 1 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, it is often a reasonable asymptotic approximation. A convenient asymptotic lower bound for the size α rank of the first non-distinct digit of a coefficient estimate can then be derived as follows.

Corollary 1 (Asymptotic formula) *if the t-statistic is asymptotically distributed as a Student's t-distribution, then a lower bound formula for the rank of the first non-distinct digit of a coefficient estimate is:*

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

where $Z_{1-\alpha}$ is the $1-\alpha$ percentile of a $\mathcal{N}(0,1)$ random variable.

Proof. See Appendix B ■

It is worth noting that our definition of distinct digits can also be applied to the standard error estimate $SE(b_k)$.[16] Proposition 2 in Appendix B derives, for the normal linear conditional mean model, a closed-form expression for a tight lower bound of the level α rank of the first non-distinct digit for $SE(b_k)$.

Unfortunately, the asymptotic distribution of a test statistic for standard error estimates is not generally available. However, in contrast to most coefficient estimates, the instrumental value of standard error estimates lies in their magnitude rather than their precision. We therefore proposes in what follows that a pragmatic rounding rule for standard error estimates is to report a fixed number of digits after the first non-zero one, which allows to make sure that reported standard errors are never equal to zero.

3.4 Graphical illustration

Figure 5 illustrates graphically our definition of "distinct" digits, using the asymptotic formula of Corollary 1. The two vertical blue bars materialize the confidence interval at the level 2α. Indeed, under an asymptotic approximation, the confidence interval of size 2α is the interval of length $2Z_{1-\alpha}SE(\hat{\theta})$ centered on the coefficient estimate, that is, the segment with lower bound:

$$\underline{CI}_\alpha(\hat{\theta}) \equiv \hat{\theta} - Z_{1-\alpha}SE(\hat{\theta})$$

and upper bound:

$$\overline{CI}_\alpha(\hat{\theta}) \equiv \hat{\theta} + Z_{1-\alpha}SE(\hat{\theta})$$

[16] And, more generally, to any random scalar for which one can devise a TOST equivalence test.

From Corollary 1, the digit of rank i is distinct if, and only if:

$$Z_{1-\alpha} SE(\hat{\theta}) \leq 10^i$$

In other words, digit i is distinct at the level α if, and only if, the confidence interval of size 2α is included in the interval $\left[\hat{\theta} - 10^i, \hat{\theta} + 10^i\right]$.

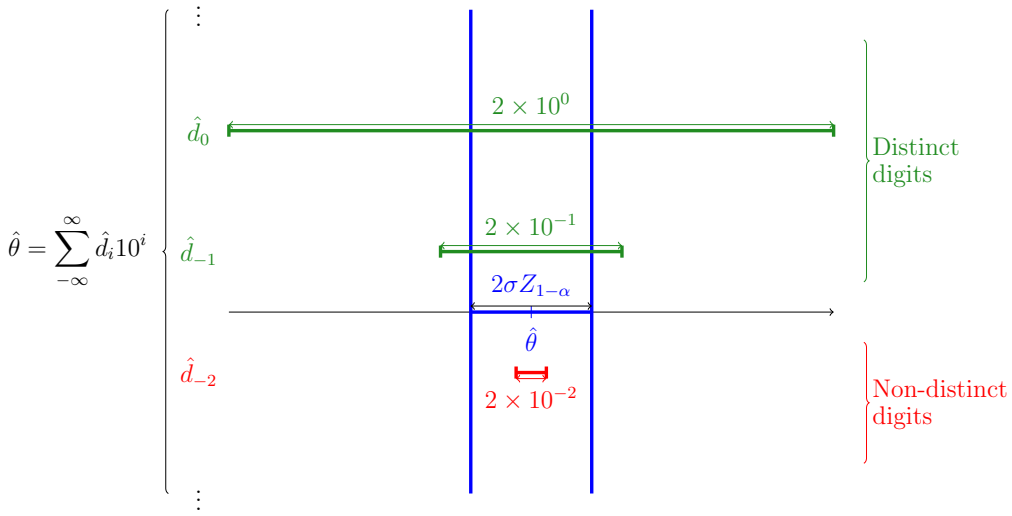


Figure 5: Illustration of distinct digits in a coefficient estimate (asymptotic formula).

The intuition behind Figure 5 may be understood as follows. Consider a pre-specified standard of precision, say 10^i. The confidence interval materializes a bandwidth that contains the parameter estimate with a very high probability $(1 - 2\alpha)$. If this bandwidth is included in an interval of size 2×10^i, that is, the distance between one digit below and one digit above at the decimal rank i, we then have statistically evidence to be able to tell two digits apart with a very high probability at this level of precision. In other words, the corresponding digit may be considered as "distinct." As the value of i decreases, the statistical evidence needed to confidently tell two digits apart increases. Consistently, the ability to declare a given digit as "distinct" requires tighter and tighter confidence intervals.

Figure 5 further clarifies why our definition of distinct digits hinges on the pre-specified width of the equivalence region, not on the fact that the test that implements our definition is in practice, similarly to confidence intervals, centered on the estimated value $\hat{\theta}$. While confidence intervals are routinely displayed on graphs, they are rarely displayed in results tables.

A rounding heuristics accounting for precision should therefore be seen as a complementary approach to the use of confidence intervals.

4 Using Distinct Digits for Rounding Parameters

This section first uses a numerical example to illustrate the properties of "distinct" digits. We next propose a rounding heuristic to report parameter estimates.

4.1 Numerical Example

True Data Generating Process

Consider an outcome variable y_t that 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:

$$y_t = \beta_0 + \sum_{k=1}^{3} \beta_k X_{kt} + \epsilon_t \tag{5}$$

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 $\beta_1 = 1.234$ and $\beta_0 = \beta_2 = \beta_3 = 1$.

We assume the researcher observes T draws of the vector $(y_t, X_{1t}, X_{2t}, X_{3t})$, which is indexed by t. To generate 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 5) and the above assumed true values of the model parameters to compute a value of y_t.

Estimated Models

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:

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

Note that we include different error terms—η_{1t}, η_{2t}, and η_{3t}—in each model to account for the omitted regressor(s). Model 1 corresponds to the simplest empirical strategy where the researcher tries to recover β_1 by regressing y_t on constant and 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 mutually independent for the same value of t and across observations, all three estimates of β_1 are unbiased. However, because X_{1t} has 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.

Monte Carlo simulations

We first embed our linear regression example into the Monte Carlo exercise sketched in the introduction. Specifically, we fix a sample size, T, and draw from our data generating process $1,000$ datasets of this sample size. We then estimate Model 1 on each of these datasets. Each of the $1,000$ Monte Carlo runs gives us a point estimate $\hat{\beta}_1$ of the coefficient of X_{1t}. This point estimate may be written as:

$$\hat{\beta}_1 = \hat{d}_0 + 0.1\hat{d}_{-1} + 0.01\hat{d}_{-2} + 0.001\hat{d}_{-3} + \ldots \tag{6}$$

where $\hat{d}_0$, $\hat{d}_{-1}$, $\hat{d}_{-2}$ and $\hat{d}_{-3}$ are integers. We recover the empirical distributions, over the Monte Carlo draws, of these digits. We perform our numerical exercise for two different sample sizes T, 100 and $1,000,000$. Figure 1shows the histograms of $\hat{d}_0$, $\hat{d}_{-1}$, $\hat{d}_{-2}$ and $\hat{d}_{-3}$ for each value of T, where each row corresponds to a sample size (from lowest to highest) and each column corresponds to a digit (from $\hat{d}_0$ to $\hat{d}_{-3}$).

Conditional on a given sample size T, we find that the rank $i^*(\alpha)$ of the last distinct digit ($\alpha = 0.1$) is assessed to be the same in all Monte Carlo runs. Figure 1 therefore uses two different colors to distinguish estimated digits that are distinct under Definition 1 from

digits that are not. We find that the distribution of $\hat{d}_i$ is tightly centered around its true value when the digit is distinct ($i > i^*(\alpha)$). The distribution of the first non-distinct digit is much more spread out but can have more probability mass around its true value. The first non-distinct digit therefore conveys useful information, despite not being a very accurate estimate of the population value of the digit. The empirical distributions of decimal digits of lower ranks assign sizable positive probabilities on all possible digit values —(0,1,2,...,9)— and resemble pure statistical noise.

4.2 Proposed Rounding Rule

We propose to adopt the following rounding heuristic for coefficient estimates: "*round parameter estimates at the rank of the first non-distinct digit, that is,* $i^*(\alpha)$." In other words, if a parameter estimate is $\hat{\theta} = 1.234\ldots$, then a researcher would report this estimate as $\hat{\theta}_p = 1$ if $i^*(\alpha) = 0$ and $\hat{\theta}_p = 1.23$ if $i^*(\alpha) = -2$. Situations where $i^*(\alpha) > 0$ can be handled by changing measurement units or using of scientific notation.[17]

The rationale for this proposal follows the conceptual framework outlined in Section 2. It essentially builds on the question: when does the information contained in an additional digit start vanishing? Our concept of distinct digits allows us to answer this question in a statistically-grounded way: digits after the last distinct digit are no longer precisely estimated, so that reporting more than one additional digit after it has negligible informative value.

To illustrate how this rounding rule differs from current practice, let us first consider a sample of size $T = 100$. Panel (a) of Figure 6 shows the current mainstream approach to report empirical results in economics. The number of digits displayed does not convey any information about their precision. Panel (b) of Figure 6 displays the same results as in Panel (a), but applies the rounding rule we propose for $\alpha = 0.1$, both for coefficients and standard errors (using the closed-form expressions we derived for the normal linear

[17]For example, a parameter estimate may be 451.23... while $i^*(\alpha) = 1$. If a researcher reports digits up to the first non-distinct 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 distinct digit or not. Table A.1 in Appendix A illustrates the use of scientific notation for our example of Section 4.

conditional mean model). Choosing $\alpha = 0.1$ instead of say $\alpha = 0.01$ is a conservative approach since the less stringent the equivalence tests, the higher the number of distinct digits. We observe that, when statistical precision is low, our procedure tends to simplify results tables without inducing any sizable loss of information.

Panel (a)

	Model 1	Model 2	Model 3
X_1	1.174	1.176	1.104
	(0.510)	(0.468)	(0.250)
X_2		0.864	0.945
		(0.196)	(0.105)
X_3			1.074
			(0.069)
Constant	1.131	1.204	1.118
	(0.459)	(0.422)	(0.225)
Observations	100	100	100
R^2	0.051	0.209	0.777

Panel (b)

	Model 1	Model 2	Model 3
X_1	1.2	1.2	1.1
	(0.51)	(0.47)	(0.25)
X_2		0.9	0.9
		(0.20)	(0.10)
X_3			1.07
			(0.069)
Constant	1.1	1.2	1.1
	(0.46)	(0.42)	(0.23)
Observations	100	100	100
R^2	0.051	0.209	0.777

Figure 6: 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 distinct digits ($\alpha = 0.1$) as well as the first (rounded) non-distinct digit.

We next increase sample size to $T = 1,000,000$ to illustrate what happens when precision increases (Figure 7). As expected, coefficients are more precisely estimated and, therefore, most coefficient estimates are reported with three decimal digits, as under the current prevailing approach to report parameter estimates. This latter approach would, however, fail to report informative digits, such as the fourth decimal digit of $\hat{\beta}_3$ in Model 3.

Figure 7 shows two alternative approaches to report standard errors. Panel (a) rounds standard errors at the rank of their first non-distinct digit as well (as characterized in Appendix B). We observe that standard errors are actually estimated with a degree of precision that exceeds by far the second or third decimal digit. For example, rounding standard errors to two decimal digits would imply that all reported standard errors would be 0.00! However, reporting all precisely estimated digits in standard errors may make results tables somewhat less reader-friendly. Because the instrumental value of standard errors lies in their magnitude rather than their precision, a pragmatic approach could be to report a fixed number of digits after the first non-zero one, as shown in Panel (b) with two digits.

In summary, our proposed reporting heuristic consists in (i) rounding coefficient estimates

Panel (a)

	Model 1	Model 2	Model 3
X_1	1.237	1.235	1.234
	(0.004474)	(0.003875)	(0.002237)
X_2		0.999	1.000
		(0.001732)	(0.0010001)
X_3			1.0000
			(0.0007072)
Constant	0.999	0.999	1.000
	(0.004474)	(0.003876)	(0.002238)
Observations	1,000,000	1,000,000	1,000,000
R^2	0.071	0.303	0.768

Panel (b)

	Model 1	Model 2	Model 3
X_1	1.237	1.235	1.234
	(0.0045)	(0.0039)	(0.0022)
X_2		0.999	1.000
		(0.0017)	(0.0010)
X_3			1.0000
			(0.00071)
Constant	0.999	0.999	1.000
	(0.0045)	(0.0039)	(0.0022)
Observations	1,000,000	1,000,000	1,000,000
R^2	0.071	0.303	0.768

Figure 7: Results tables with a sample size of $T = 1,000,000$. Left panel: standard errors rounding at the first non-distinct digit. Right panel: standard errors rounded at the decimal rank following their first non-zero digit.

at their first non-distinct digit, using the asymptotic formula of Corollary 1; and (ii) rounding standard error estimates one or two decimal ranks after their first non-zero digit. We suggest to use conservative (i.e., large) values of α (e.g. $\alpha = 0.1$).[18] Our procedure is straightforward to implement and already embedded into the widely used R package "modelsummary" (Arel-Bundock, 2022). This feature only requires a few lines of code and relies on readily available magnitudes. It could therefore seamlessly be set as the default option to any routine used to report parameter estimates. Note that our implementation accounts for the fact that either a change of units or the use of scientific notation is warranted for large but imprecise estimates that are different from zero. Because such situations rarely arise in practice, we discuss this point in Appendix D.

Finally, one may use the conceptual framework sketched in Figure 2 to assess whether, relative to current rounding heuristics, the proposed rounding rule tends to report "too many" (i.e. high rate of "false positives") or "too few" (i.e. high rate of "false negatives") decimal digits. Table A.2 in the Appendix report the obtained results from Monte Carlo simulations. Except for very large sample sizes, always reporting two or three decimal digits is found to overstate available precision (high rate of "false positives"). In contrast, the proposed heuristic rounds coefficient induces negligible loss of information (low rate of "false negatives") and, at the same time, makes sure that a reasonably low rate of false "positives"

[18]This implies reporting more digits but the sensitivity of $i^*(\alpha)$ to α is small given the logarithm term in the formula of Corollary 1.

is obtained when rounding instead at one decimal rank to the left. In other words, the proposed rule makes it possible for readers to directly interpret the point estimate in the manner that best fit their intended use of the estimated value.

5 Empirical Relevance

This section quantitatively assesses the extent to which the coefficient estimates printed in published articles in economics consist of distinct digits.

5.1 Main Results

We use our dataset of printed coefficients and standard errors in the AER to compare our proposed rounding rule to current practice. Specifically, we derive, for each coefficient estimate, the decimal rank of the first non-distinct digit using Corollary 1.



Figure 8: Fraction of the printed digits that would not have been printed if results tables had only rounded coefficient estimates at their first non-distinct digit.

Figure 8 shows, for each year, the fraction of printed digits that would not have been reported with our rounding heuristic for different values of α. We find that over a third of the digits printed in published articles would not have been reported. Our framework also allows us to identify coefficient estimates for which our reporting rule would have implied printing *more* digits. We find that only about 1% of the coefficients fall into this category

(for $\alpha = 0.1$), mainly consisting of the parameter estimates with a reported standard error of zero.

Next, we use our dataset to compute, for each coefficient, an upper bound of its number of sizable distinct 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.[19]

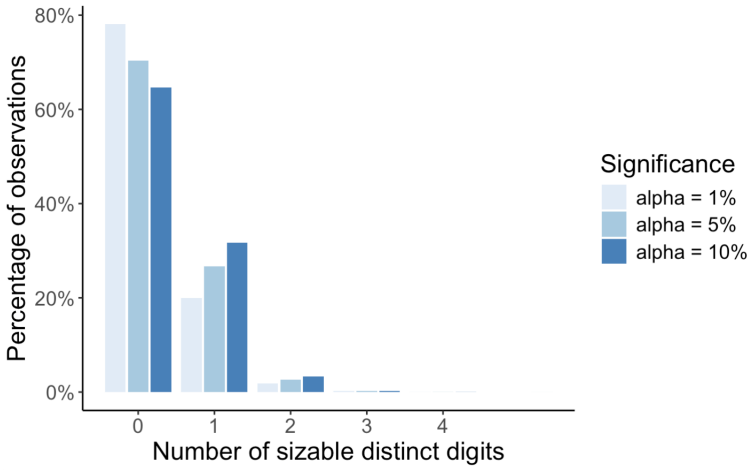


Figure 9: Distribution of the number of sizable distinct digits.

Figure 9 shows histograms for the number of sizable distinct digits for different values of α. We find that the number of sizable distinct 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 distinct, meaning that most coefficient estimates in economics would be reported with a single non-zero digit.[20]

5.2 Robustness Checks

The dataset we collected has two main limitations. First, we restricted attention to a single outlet. Second, we retrieved all coefficient estimates and are therefore not able to infer

[19]For this exercise, we drop the observations from our sample for which the printed standard error estimate is zero (1% of coefficients) because we cannot derive a finite value for the rank of the last distinct digit for such observations.

[20]Nonetheless, our reporting rule would preserve statistical significance in the sense that, if the coefficient estimate reported with many decimal digit is significantly different from zero at $\alpha = 5\%$ or less, then its reported rounded value also rejects the null hypothesis of the significance test.

whether or not a given coefficient corresponds to the main effect of interest in the considered studies. As a result, we cannot rule out that our finding that most coefficients have a number of sizable distinct digits equal to zero (Figure 9) might be driven by a lack of statistical power to precisely estimate the coefficients associated with control variables.

To address these limitations, we make use of the dataset collected by Askarov et al. (2023).[21] 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,[22] four general interest journals,[23] and six top-field journals for which we have a large number of observations.[24] 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 for 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, only a subset of articles is considered for each outlet (the studies that where subsequently tabulated in one of the included meta-analyses). 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.13 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.[25]

Figure 10 shows, for each journal, the histograms of the number of sizable distinct dig-

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

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

[23] *Review of Economics and Statistics*, *European Economic Review*, *Economic Journal*, and *Journal of the European Economic Association*.

[24] *Journal of Development Economics*, *Journal of Money Credit and Banking*, *Journal of Public Economics*,

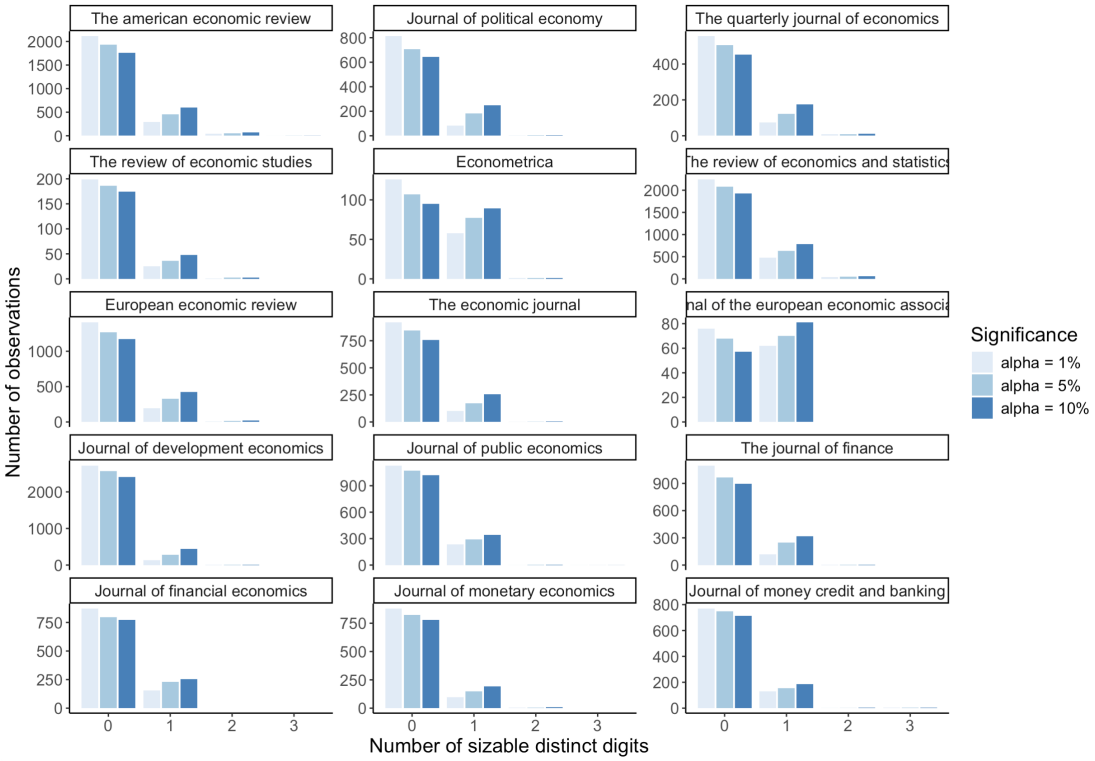


Figure 10: Histograms of the number of sizable distinct digits for the different journals.

its. 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 distinct 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 similar results would have been obtained if we had studied a different journal. Only two journals stand out for reporting, on average, more precise coefficients, namely *Econometrica* and *Journal of the European Economic Association*. The sample size is, however, significantly smaller for these two outlets than for most other journals, which calls for great caution when interpreting this result. Overall, we do not find evidence that our empirical results would have been different if we had focused attention on the main coefficients of interest and/or on a different journal.

Journal of Finance, *Journal of Financial Economics* and *Journal of Monetary Economics*.

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

6 Conclusion

This article proposes a principled rule to round parameter estimates, building on the statistical concept of "distinct digits," which are characterized through equivalence tests. While the currently prevailing practice is to always report either two or three decimal digits, we suggest instead to round parameter estimates at the first non-distinct digit. We derive an asymptotically-valid closed-form expression for the rank of the last distinct digit of a coefficient estimate for a broad range of econometric models, as well as exact results for the case of the normal linear conditional mean model. These formulas rely on readily available information from econometric estimation packages. Therefore, our rounding approach can be added seamlessly to any existing post-processing routine used in practice to export empirical results. For example, the R package "modelsummary" (Arel-Bundock, 2022) includes it as a formatting option for reporting results.

We make two main points to support our proposal. First, our statistical framework provides guarantees that the tail of digits beyond the first non-distinct digit essentially conveys statistical noise. As a result, our rounding rule only induces a negligible loss in precision. We illustrate this property through Monte Carlo simulations using a simple example for which the data generating process is known. Second, we assemble an original 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). Using this dataset, we find that current practice departs significantly from the reporting rule we propose. In particular, we show that over a third of printed digits in coefficient estimates would not be reported should our reporting rule be used, while only a minority of estimates would be printed with more decimal digits. Interestingly, we observe that most coefficient estimates have no non-zero distinct digit, which suggests that current rounding heuristics often convey a misguided sense of precision.

We believe that rounding parameter estimates at the rank of their first non-distinct digit would be a valuable improvement relative to current practice. In particular, it would increase the information content of tables of parameter estimates by both avoiding to report digits that are essentially uniformly drawn between 0 and 9, and failing to report precisely

estimated digits. A better use of economic results in practical applications is likely to ensue. In addition, the incentives to generate more precise estimates of important economic parameters may increase.

References

Andrews, Isaiah and Jesse M Shapiro, "A model of scientific communication," *Econometrica*, 2021, *89* (5), 2117–2142.

_ **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.

Arel-Bundock, Vincent, "modelsummary: Data and Model Summaries in R," *Journal of Statistical Software*, 2022, *103*, 1–23.

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.

Fitzgerald, Jack, "The need for equivalence testing in economics," Technical Report, I4R Discussion Paper Series 2024.

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.

Henry, Emeric and Marco Ottaviani, "Research and the approval process: The organization of persuasion," *American Economic Review*, 2019, *109* (3), 911–955.

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.

Korting, Christina, Carl Lieberman, Jordan Matsudaira, Zhuan Pei, and Yi Shen, "Visual inference and graphical representation in regression discontinuity designs," *The Quarterly Journal of Economics*, 2023, *138* (3), 1977–2019.

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.

Rainey, Carlisle, "Arguing for a negligible effect," *American Journal of Political Science*, 2014, *58* (4), 1083–1091.

Romano, Joseph P., "Optimal Testing of Equivalence Hypotheses," *The Annals of Statistics*, 2005, *33* (3), 1036–1047.

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.

Appendices

A Supplementary Figures and Tables

A.1 Figures

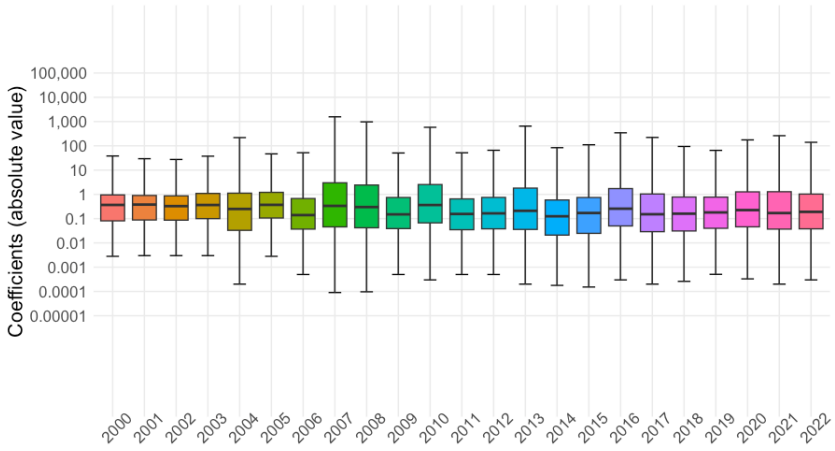


Figure A.11: Box plots of the magnitude of estimated coefficients (absolute value in $\log_{10}$ scale). Boxes locate the first, second and third quartiles of the distributions. Top whiskers (resp. bottom whiskers) are drawn at a distance of 1.5 times the interquartile range above the third quartile (resp. below the first quartile).

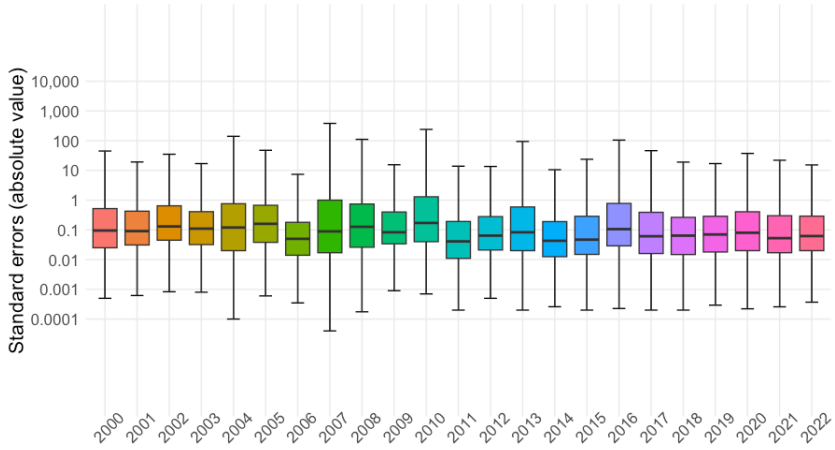


Figure A.12: Box plots of the magnitude of estimated standard errors (absolute value in $\log_{10}$ scale). Boxes locate the first, second and third quartiles of the distributions. Top whiskers (resp. bottom whiskers) are drawn at a distance of 1.5 times the interquartile range above the third quartile (resp. below the first quartile).

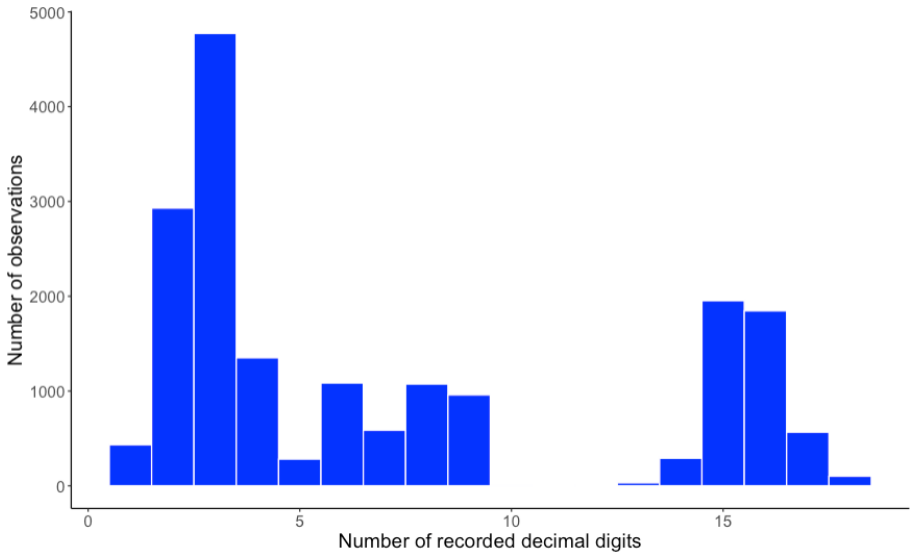


Figure A.13: Histogram of the number of recorded decimal digits for the "effect size" in the subset of the dataset of Askarov et al. (2023) used in Section 5.

A.2 Tables

Table A.1 is identical to panel (c) in Figure 6 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-distinct 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	1.2e+03	1.2e+03	1.1e+03
	(5.1e+02)	(4.7e+02)	(2.5e+02)
X_2		9e+02	9e+02
		(2.0e+02)	(1.0e+02)
X_3			1.07e+03
			(69)
Constant	1.1	1.2	1.1
	(0.46)	(0.42)	(0.23)
Observations	100	100	100
R^2	0.051	0.209	0.777

Table A.2 compares, based on Monte Carlo simulations, the rate of "false positives" and "false negatives" (as defined in Figure 2) for the proposed rounding rule versus current

heuristics.

Sample size T:	100	1,000	10,000	100,000	1,000,000
Rounding at 2 decimal digits					
False positive rate	97.9%	93.1%	81.3%	55.9%	30.6%
False negative rate	0.3%	0.7%	2.1%	5.0%	16.7%
Rounding at 3 decimal digits					
False positive rate	99.7%	99.3%	97.9%	95.0%	83.3%
False negative rate	0%	0%	0.3%	0.3%	1.1%
Rounding at first non-distinct digit - $\alpha = 0.1$					
Average # of decimal digits reported	1	2	2	3	3
False positive rate	82.1%	93.1%	81.3%	95.0%	83.3%
False negative rate	2.1%	0.7%	2.1%	0.3%	1.1%

Table A.2: Rates of false positives and false negatives (as defined in Figure 2) for different rounding rules for coefficient $\hat{\beta}_1$ of Model 3.

B Proofs and Additional Results

B.1 Proof of Proposition 1

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 distinct 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}:

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

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 last decimal rank $i^\dagger(\alpha)$ such that the digit of decimal rank i is distinct is such that:

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

This expression is precisely the definition of the ceiling of $\frac{\ln SE(b_k) + \ln t_{1-\alpha}(T-K)}{\ln 10}$. The rank of the first non-distinct digit it then:

$$i^*(\alpha) \equiv i^\dagger(\alpha) - 1$$

B.2 Proof of Corollary 1

We have:

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

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 $i \leq i_\infty^*(\alpha)$, then we also have $i \leq i^*(\alpha)$. Therefore, by Proposition 1, d_i is not a distinct digit at level α.

B.3 Rank of the Last Distinct Digit for Standard Error Estimate in the Normal Linear Regression Model

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 distinct digit at level α for the estimated standard error $SE(b_k)$ is given by:*

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

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}:

$$H^i_a\text{: } se_k = SE(b_k) - 10^i \qquad \text{versus} \qquad K^i_a\text{: } se_k > SE(b_k) - 10^i$$

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[26] 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}:

$$H^i_b\text{: } se_k = SE(b_k) + 10^i \qquad \text{versus} \qquad K^i_b\text{: } se_k < SE(b_k) + 10^i$$

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

[26] 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 because 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 2 *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-distinct 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 $i^\dagger(\alpha)$ such that:

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

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

B.4 Proof of Lemma 2

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

$$1 - \sqrt{\frac{k}{\chi^2_\alpha(k)}} \leq \sqrt{\frac{k}{\chi^2_{1-\alpha}(k)}} - 1 \tag{7}$$

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 2, 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,[27] 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:

$$\int_1^{\frac{\chi^2_{1-\alpha}(k)}{\chi^2_{\frac{1}{2}}(k)}} f(\chi^2_{\frac{1}{2}}(k)t)dt = \int_1^{\frac{\chi^2_{\frac{1}{2}}(k)}{\chi^2_\alpha(k)}} \frac{1}{t^2} f(\frac{\chi^2_{\frac{1}{2}}(k)}{t})dt = \frac{1}{\chi^2_{\frac{1}{2}}(k)}\left(\frac{1}{2} - \alpha\right) \tag{8}$$

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

[27] 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 8. 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 8, we have:

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

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 in brackets are very low in other contexts 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 they are false positive.

2. several results tables report coefficients and their standard errors along with statistical tests and their test values. We manually remove manually the latter observations we were able to detect by inspecting the 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 estimate and its standard error. Similarly, a number of false positives occur when the text around a results table includes 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 manually checking 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 idiosyncratic 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.[28] We manually correct the corresponding coefficients (about 1,100 observations corrected).

Overall, our manual data cleaning corrects the value of 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,[29] our data extraction procedure is entirely automated and transparent. It can therefore be easily replicated and improved.

D On the Choice of Units

This Appendix discusses a number of practical considerations when choosing the units of dependent and independent variables. A common feature of these considerations is that

[28]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.

[29]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.

they emerge from the lack of use of scientific notation. However, because the units used in economics tend to be chosen such that parameter estimates are neither very high nor very low (see Figure A.11), reporting plain numbers may represent a simpler and more pragmatic approach than scientific notation. In this section, we therefore discuss situations where the choice of units can have an impact on the empirical exercise of Section 5 and/or on the applicability of the reporting rule we propose.

As a starting point, it is important to note that the rank $i = 0$ plays no particular role for our concept of distinct 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:

$$\hat{\theta} = \sum_{i=-\infty}^{D(\hat{\theta})} \hat{d}_i 10^i \tag{9}$$

where, as before, $D(\hat{\theta})$ is the largest rank for non-zero digits.

Under the reporting rule we propose, for a given parameter estimate, results tables would print all digits whose rank lies between $i^*(\alpha)$ (rank of the first non-distinct digit) and $\max(D(\hat{\theta}), 0)$. A few special cases are worth discussing.

First, the largest rank for non-zero digits may be negative (i.e., $D(\hat{\theta}) < 0$). For example, a coefficient may be estimated to be 0.023 (i.e., $D(\hat{\theta}) = -2$). Although our reporting rule 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 coefficient may be seen as relatively uninformative, they are considered in our empirical application as distinct digits whenever $i^*(\alpha) \leq -2$. For example, if $i^*(\alpha) = -3$, three out of the four printed digits of 0.023 are considered to be deemed distinct. 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 deemed distinct. From this perspective, our assessment in Section 5 of the share of printed digits that are not distinct may be seen as conservative.

Another practical concern in situations where $D(\hat{\theta}) < 0$ and $i^*(\alpha) < 0$ 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

first non-distinct digit is actually $i^*(\alpha) = -6$. Such rounding practices have been shown to possibly generate biases in meta-analysis studies (Kranz and Pütz, 2022; Brodeur et al., 2022).

Finally, the rank of the first non-distinct 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 first non-distinct digit is $i^*(\alpha) = 1$. Under our approach consisting in reporting distinct digits up to the first non-distinct 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 distinct digit or not. The use of scientific notation would then be required to implement the reporting rules discussed above.[30] Our analysis of Section 5 suggests that the rank of the first non-distinct 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 rather than imposing a widespread use of scientific notation.

[30]Table A.1 in Appendix A illustrates the use of scientific notation for our example of Section 4.