Accounting for Multiplicity in Inference on Economics Journal Rankings

William C. Horrace* Christopher F. Parmeter†

July 15, 2016

Abstract

Nearly all journal rankings are constructed as deterministic, yet they are clearly stochastic. To remedy this, Stern (2013) calculates standard errors and performs inference on the ranks of economics journal based on impact factors. However, this inference is essentially a series of univariate tests that do not control for the overall error rate of the inferential exercise. Using multiple comparison and ranking procedures, we reevaluate the inference while controlling for the multiplicity (the implied multivariate inference) in the rank statistic. The results are compared and differences highlighted.

Keywords: Ranking and Selection, Subset Selection, Multiple Comparisons, Journal Impact Factors

JEL Codes: A14, C12

*Department of Economics, Syracuse University, Syracuse, NY, 13244-1020. whorrace@maxwell.syr.edu.
†Corresponding author, Department of Economics, University of Miami, Miami, FL, 33146. cparmeter@bus.miami.edu.
1 Introduction

Ranking economic journals is important for a variety of reasons. Journal rankings inform university administrators and faculty during both promotion and tenure processes, draw attention to important economic research, and help administrators assess the quality of their faculty. Journal rankings also aid researchers in matching their work to appropriate outlets for submission. Unfortunately, most journal rankings are presented as if they are deterministic. However, journal rankings based on citations (or impact) need to account for the inherent randomness of the data generating process and the citation analysis. A journal with a large estimated impact factor may not have as much impact as a journal with a smaller estimated impact factor, as statistical uncertainty may confound the comparison.

Aside from the usual sampling variability, there are other factors that may confound our ability to judge relative journal quality. For example, Card and DellaVigna (2013) note that the number of yearly submissions to the “top 5” economics journals nearly doubled from 1990 to 2012 while the number of total articles published declined from approximately 400 per year in the late 1970s to around 300 per year in 2012. Such a structural shift is likely to induce (or have been induced by) behavioral changes in journal policies and in the submission choices of authors, making the publication process and citation analysis less certain. The ranking of economics journals is further complicated by the fact that the number of economics journals, as listed in EconLit, is steadily growing, and it may be difficult to assess the quality of newer journals. Another complication is that journal rankings are aggregate indices that do not take into account the quality of individual articles. Using journal rankings to assess the quality of a research portfolio is clearly less informative than reading the individual papers, or determining the citation stream for each paper in the portfolio. Nonetheless, journal rankings are considered by many to provide a useful measure of researcher impact.

Another source of uncertainty is the method by which journals are ranked. For example, Kalaitzidakis et al. (2003) use iterative impact factors to rank journals, Palacio-Huerta and Volij (2004) use a recursive
method based on impact factors, Bergstrom (2007) develops the Eigenfactor scoring system, while Combes and Linnemer (2010) use a combination of JCR citation indices and Google Scholar h-indices. Even ignoring the uncertainty in any specific rank of journals, there also is uncertainty as to the appropriate way in which to rank journals. Further still, there is not an accepted uniform ranking system for journals as each favors different aspects of the publication/citation process.

A step in the right direction is the recent work of Chang et al. (2016), who propose a robust method of 15 different metrics using the harmonic mean. The implication of their analysis is that a ranking based on any single metric can be seen as extreme relative to the harmonic mean, which is constructed using weights across four classes of journal ranking metrics (# of citations, journal policy, # of high quality papers and journal/article influence), based on the type of ranking metric being used. The benefit of Chang et al.’s (2016) robust ranking is that nearly all of the essential features of publishing are captured, producing a measure which cuts across various aspects of publication quality. However, even with this robust metric, the rankings produced are treated as deterministic.

Recognizing the importance of accounting for uncertainty in journal ranks, Stern (2013) estimates impact factors for each of 230 economics journal over a five-year period, as well as a standard error for each estimate. He then constructs confidence intervals for each journal and conducts a barrage of t-tests for differences between the impact factors of a subset of “top 30” journals and all other journals. Unfortunately, his individual pairwise tests do not control for the overall error rate (the simultaneity) of the testing exercise, so probabilistic statements concerning the relative journal rankings are likely wrong. Additionally, individual tests of the hypotheses like “journal A is better than journal B,” or “journal A is better than journal C,” or “journal A is better than journal D” are useful, but are not as informative as the multivariate test of the hypothesis that “journal A is simultaneously better than journals B, C, and D,” particularly in the context of a ranking exercise. When Stern conducts individual pairwise tests, he is ignoring the inherent multiplicity

\footnote{Stern uses the top 30 journals of Kalaitzidakis et al. (2003).}
(or simultaneity) in the ranking. That is, while one might be familiar with the construction of a $t$-statistic and comparing it to the ubiquitous 1.96 cutoff, when multiple comparisons are possible, this cutoff needs to be increased, sometimes dramatically so. In this sense one can think of multiple pairwise comparisons as conservative, tending to favor outcomes with higher means. The requisite amount to increase the cutoff depends on the number of journals to be compared. Fortunately, there is a significant literature devoted to this problem called Multiple Comparison Procedures, and applying these techniques to the problem of journal ranking is the purpose of this note. The goal here is to adapt the inference of Stern (2013) to account for multiplicity, and see how/if his results change.

Using statistical techniques that account for multiplicity, we analyze the five-year impact factors calculated by Stern (2013). As expected some differences in the ranks of journals arise. Whereas Stern (2013, pg. 184) determines two journals (Journal of Economic Literature and Quarterly Journal of Economics) that are “...clearly separated from the rest of the field,” we find that there are two additional journals that cannot be ruled out as belonging to the best group (Journal of Finance and Journal of Economic Perspectives) at the 95% level. Additionally, Stern’s (2013) analysis notes that there is a large overlap in individual confidence intervals, which precludes distinguishing specific groups across a wide range of journals. Accounting for multiplicity allows easier categorization of journals into specific ranks of groups, such as a first or second tier.

This paper is organized as follows. We detail the statistical underpinnings of multiple comparison procedures in Section 2. Section 3 discusses the data and the results of the journal rankings when these techniques are used. Finally, Section 4 concludes and offers avenues for further research.

2 Multiplicity and Ranking Uncertainty

Multiple comparison procedures are concerned with drawing simultaneous inferences on differences in population means across $n$ populations, while controlling for multiplicity and (hence) the overall error rate of
the exercise. This is accomplished by constructing simultaneous confidence intervals for all \((2^n)\) population differences (or some subset of them), using critical values drawn from multivariate distributions (typically a multivariate normal or multivariate \(t\) distribution), as opposed to univariate distributions.\(^2\) Multivariate critical values are necessarily larger than univariate critical values (e.g., \(z_{\text{crit}} = 1.96\) for a univariate standard normal), so multiple comparison confidence intervals are typically wider and more conservative. It is in this sense that they control for multiplicity. A related body of literature on ranking and selection is due to Bechhofer (1954); additional ranking procedures followed: Gupta (1956, 1965), Fabian (1962) and Desu (1970). Ranking and selection procedures are concerned with identifying a subset of the \(n\) populations which contains the population with the largest mean (“the best” population, say) at a pre-specified confidence level. Like multiple comparison procedures, ranking and selection is also based on critical values drawn from multivariate distributions.\(^3\) It is Gupta’s ranking and selection procedures that form the basis of our analysis.

A related literature is called “multiple comparisons with a control” or MCC, which is primarily due to Dunnett (1955, 1964). Given a preselected control population, MCC constructs \(n - 1\) simultaneous confidence intervals on mean differences between the control and the rest of the populations, again relying on the same multivariate critical values to control for multiplicity and (hence) the overall error rate of the exercise.\(^4\) In fact, if the preselected control population has all positive MCC upper bounds on the mean of the control minus the rest, then the control population is selected for inclusion in Gupta’s (1965) “subset of the best.” Therefore, determination of Gupta’s subset follows by performing MCC with each population as the control and selecting all control populations with all positive MCC upper bounds for inclusion in the subset. A good textbook treatment of MCC can be found in Hsu (1996, Chapter 3). Finally, multiple comparisons with the best, or MCB, evolved in the early 1980s with the work of Hsu, and are concerned with construction of

\(^2\)Stern (2013) uses univariate critical values to construct confidence intervals and to conduct \(t\)-tests. He does not control for multiplicity.

\(^3\)Stern’s group of top 30 journals is an educated guess at a subset of the best. Ranking and selection uses inference to identify this set using statistical rigor.

\(^4\)Stern’s group of top 30 journals can be thought of as 30 individual controls, but with multiplicity ignored in his analysis.
Multiple comparisons procedures have started to gain traction in various applied economic settings (Horrace and Schmidt, 2000; Horrace, 2005a; Flores-Lagunes et al., 2007; Horrace et al., 2008; Horrace and Schnier, 2014; Horrace et al., 2015), and this paper adds to this growing economics literature.

Our focus here is on the ranking and selection of economics journals that have the largest (unknown) impact factors, which follows from MCC with each journal as the control as previously described. In what follows, we introduce notation for the problem at hand.\(^5\) Let \(\theta_i\) be the impact factor of the \(i^{th}\) of \(n\) journals, and we have a random sample of repeated observations of the impact factors for each journal over time: 

\[y_{it}, \ i = 1, \ldots, n, \ t = 1, \ldots, T.\]

Following Stern (2013), we assume that our impact factor data are randomly drawn from independent normal populations 

\[y_{it} \sim N(\theta_i, \sigma_i^2), \ i = 1, \ldots, n,\]

but the techniques that follow are easily adaptable to general dependence structures. All the difficulty lies in the calculation of critical values, and independence simplifies this calculation.\(^6\) A consistent estimator of \(\theta_i\) is the sample average \(\hat{\theta}_i = \bar{y}_i\) with variance 

\[V(\hat{\theta}_i) = s_i^2/T\]

with the usual population variance estimator 

\[s_i^2 = (T - 1)^{-1} \sum_t (y_{it} - \bar{y}_i)^2.\]

These are exactly the assumptions and calculations of Stern (2013), leading to construction of confidence intervals (based on univariate \(t\) distributions) and a barrage of univariate \(t\)-tests of a hypothesis of the form 

\[H_0 : \theta_i = \theta_j, \]

each at an error rate of \(\alpha = 0.05\). This approach ignores the inherent multiplicity of the problem, and the overall error rate will be something larger than 0.05, with the error increasing in the number of tests performed. In what follows we use multivariate ranking and selection procedures to ensure

\(^5\) More general problems are considered in Horrace and Schmidt (20002).

\(^6\) The \(N\)-dimensional probability integral in the dependent case reduces to a single integral in the independent case, which facilities numerical calculation for the critical values. However, in economic applications, simulation of critical values is likely to be acceptable, in which case critical values can be easily generated based on estimates of the dependence structure. See Horrace and Schmidt (2002) for a discussion.
that the error rate for the entire exercise is at least $\alpha = 0.05$.

Let the unknown population ranks of the $n$ journals be: $\theta_{[n]} > \theta_{[n-1]} > \cdots > \theta_{[1]}$ with sample ranks $\hat{\theta}_{(n)} > \hat{\theta}_{(n-1)} > \cdots > \hat{\theta}_{(1)}$. Notice that $(i)$ does not equal $[i]$ in general, and determining the extent to which they are equal is the primary purpose of multiple comparison techniques. Notice that as $T \to \infty$, $(i) \to [i]$, we can ignore the uncertainty in the sample ranks, but in finite samples we cannot. Let $N = \{1,2,\ldots,n\}$ be the set of journal indices, then ranking and selection procedures identify a subset of the best journals, $S_\alpha \subseteq N$, such that $\Pr([n] \in S_\alpha) \geq 1 - \alpha$. That is, the subset contains the index of the best journal with at least probability $1 - \alpha$. The cardinality of $S_\alpha$ is increasing in the uncertainty of the problem (as $T$ decreases) and in the multiplicity of the problem (as $n$ increases). In other words, it is easiest to identify the best journal $[n]$, when we have a lot of repeated observations (large $T$) and when there are fewer journals to compare (small $n$). Therefore, Stern’s problem with $T = 5$ and $n = 230$, is likely to produce a subset of the best with large cardinality, implying a great deal of uncertainty in our ability to identify a single best journal.

Following Gupta (1956) and Horrace et al. (2008), we have the subset:

$$S_\alpha = \{j : \bar{y}_j - \overline{y}_i + t_{j,v,\alpha}(s_j^2/T + s_i^2/T)^{1/2} \geq 0; \forall i \neq j\},$$

for all $j \in N$, where $t_{j,v,\alpha}$ is the two-sided critical value for an $(n - 1)$-dimensional $t$ distribution with $v$ degrees of freedom and diagonal variance matrix with typical element $(s_j^2/T + s_i^2/T)$ such that $\Pr(\max_j |t_j| \leq t_{j,v,\alpha}) = 1 - \alpha$. The subset $S_\alpha$ is guaranteed to be non-empty and have minimal cardinality for a given confidence level, $1 - \alpha$. If it is a singleton, the analysis has identified the best journal with confidence level at least $1 - \alpha$, otherwise there will be more than one journal in contention for the best. Conversely, if it is a particularly noisy exercise (i.e., $s_j^2 \gg 0$ for all $j \in N$) or if the population means are approximately equal,

\footnote{See Horrace and Schmidt (2002, eq. 6) for a precise statement of the probability integral based on independence of the $n$ populations.}

7
then it is possible that $S_\alpha = N$ with cardinality $n$. In any case, it will be true that the probability that $[n] \in S_\alpha$ is at least $(1 - \alpha)$.

The subset selection criteria is related to and can be adapted from Dunnett’s (1955, 1964) multiple comparisons with a control. Let $j \in N$ be a preselected control population, then simultaneous MCC $(1 - \alpha) \times 100\%$ confidence intervals for $\theta_j - \theta_i$ are the set of bounds $[L_i^j, U_i^j]$ such that

$$L_i^j = \bar{y}_j - \bar{y}_i - t_{j,v,\alpha}(s_j^2/T + s_i^2/T)^{1/2}$$

$$U_i^j = \bar{y}_j - \bar{y}_i + t_{j,v,\alpha}(s_j^2/T + s_i^2/T)^{1/2},$$

for all $i \neq j$. Therefore, the selection criteria for $S_\alpha$ can be written as: $j \in S_\alpha$ if $U_i^j > 0$, $i \neq j$. That is, if $\theta_j$ is potentially larger than $\theta_i$ for all $i \neq j$, then population $j$ is “in contention for the best” and is selected to be in $S_\alpha$. In what follows, we simulate $t_{j,v,\alpha}$ as per Horrace and Schmidt (2000), “…there is no compelling reason to prefer numerical approximations over simulated (stochastic) ones in an economic setting.” Following Horrace et al. (2008), if we remove the indices in $S_\alpha$ (the journals in contention for the best) from $N$, and redo the analysis based on the remaining journal indices, then we can identify a subset of second best journals: $S_{\alpha^*} \subseteq N - S_\alpha$, based on the remaining indices in $N - S_\alpha$. Then, we can use the Bonferroni inequality to conclude that the overall error rate of the exercise is at most $2\alpha$.

Edwards and Hsu (1983) provided a general technique for adapting these MCC intervals to MCB intervals. While our focus is to account for multiplicity in inference on journal rankings using Gupta’s subset $S_\alpha$, for completeness simultaneous MCB $(1 - \alpha) \times 100\%$ confidence intervals for $\max_j \theta_j - \theta_i$ is the set of bounds $[L_i, U_i]$:

$$L_i = \max \left[ 0, \min_{j \in S_\alpha} \left\{ \bar{y}_j - t_{j,v,\alpha}(s_j^2/T + s_i^2/T)^{1/2} \right\} - \bar{y}_i \right]$$

$$U_i = \max \left[ 0, \max_{j \neq i} \left\{ \bar{y}_j + t_{j,v,\alpha}(s_j^2/T + s_i^2/T)^{1/2} \right\} - \bar{y}_i \right]$$

8
Then the joint probability of \( \max_j \theta_j - \theta_i \in [L_i, U_i] \) and \([n] \in S_{\alpha} \) is at least \((1 - \alpha)\). That is, once we have determined membership in \( S_{\alpha} \), there is no additional error associated with the construction of confidence intervals on all differences from the unknown best population. See Horrace and Schmidt (2000) for a more complete discussion of results from MCC and MCB.

3 Data and Results

We use the data from Stern (2013), who collects all citations in the year 2011 for articles published between 2006 and 2010 for the 232 journals in the JCR economics category with a five-year impact factor. Overall this produces 54,416 articles with a total of 88,676 citations. Impact factors and associated standard errors were calculated based on the average number of citations received by the articles over the five-year period. Since Stern implicitly assumed that impact factors are normal distributed (i.e., he used 1.96 as the confidence interval critical value), we will make the same assumption, so that our critical values are simulated from a multivariate normal distribution and not from a multivariate \( t \)-distribution.

Results for the top 16 journals in Stern’s ranking are in Table 1. The first column contains the journal name, the second column contains the ranked impact factor estimates from Stern’s paper, and the third column contains the standard errors of the estimated impact factors. The results of the ranking analysis for the 230 journals of Stern \((n = 231)\) is in columns 4 and 5, while the results for the 30 journals of KMS \((n = 30)\) are in columns 6 and 7. Column 4 contains the simulated two-sided critical values \( (t_{j,\infty,0.95}) \) used in the ranking and selection analysis of Stern’s 230 journals. Because we are making 230 simultaneous comparisons, the critical values are all approximately equal to 3.0. The multiplicity of the inferential statement causes the critical values to be much larger that the univariate 1.96 that Stern exploits in his analysis. While Stern concludes that Journal of Economic Literature and Quarterly Journal of Economics are the best journals based on impact factor (and controlling for standard errors), the ranking analysis in the fifth column tells a

\(^8\)We excluded Hacienda Publica Espanola, because in Stern’s Table 1 it has standard error equal to zero (no variability in the data). Therefore we have \( n = 231 \).
slightly different story.

First, with 230 journals in the rank statistics, we cannot reject the hypothesis that *Journal of Finance* and *Journal of Economic Perspectives* are best at the 95% level. While the impact factors of these two journals (6.173 and 6.027, respectively) are lower than those of *J ECON LIT* and *Q J ECON* (9.281 and 8.261, respectively), the scores are not significantly different when we account for the multiplicity in the rank statistic. There are just too many journals being compared for us to confidently say that J FINANC and J ECON PERSPECT are not among the best. Second, the *American Economic Review* is in the second best subset in column 3, despite its impact factor being larger than that of the *Journal of Finance* (compare 6.059 to 6.173, respectively). This may be due to multiplicity, but it is likely due to the fact that the impact factor for AM ECON REV is estimated with relatively high precision (standard error equal to 0.215). That is, its precise impact factor gives us confidence that AM ECON REV is not first best, while the relative imprecise impact factor of J FINANC (and J ECON PERSPECT) cannot reject the hypothesis that it is (they are) not first best. This fact is not captured in Stern’s Figure 2, despite his attempts to control for multiplicity by calculating the maximal whisker of 229 univariate confidence intervals for each journal in his ranking (not a precise statement). Finally, *Journal of Economic Growth* (*J ECON GROWTH*) is contained in the second best subset despite its relatively low impact factor (compare 4.117 for *J ECON GROWTH* to 6.059 for AM ECON REV). The relatively large standard error for *J ECON GROWTH* (0.670) precludes us from rejecting the hypothesis that it may be in the second best subset. The journals that receive a “-” in the fifth column are part of the ranking analysis, but did not make it into the subsets of 1st or 2nd best.

The cardinality of our subset of 1st best ($S_{\alpha}$) is four at a error rate of at most 5% ($\alpha = 0.05$), while Stern’s implied “subset of 1st best” has cardinality two at an assumed error rate of exactly 5%. However, the actual error rate of Stern’s result will be larger once we account for multiplicity in the ranking. This begs the question, how much larger is this error rate for the set of 230 journals once we account for multiplicity? More specifically, how large do we have to make $\alpha$ for the cardinality of $S_{\alpha}$ to be two (the implied Stern
result). We conducted experiments and found that as we increased $\alpha$ from 0.05, the cardinality $S_\alpha$ became two at an error rate of about 9% ($\alpha = 0.092$ to be more precise). The experiment qualifies the extent to which Stern’s inference may be inaccurate.

The sixth and seventh columns of Table 1 contain the ranking analysis based on the 30 journals of KMS. We only show the results for nine of these 30 journals with the largest impact factors. Journals with an NR (not Ranked) in these columns are excluded from the KMS 30 economics journals, and are not part of the analysis. The sixth column contains the multivariate critical values for the top nine journals in that list, although the critical values themselves are based on all 30 journals in the larger list. Because there are fewer comparisons to be made in this case (30 vs. 230), the critical values are necessarily smaller (on the order of 2.8) when compared to those in column four. Obviously, with only 30 journals to compare, the extent of the multiplicity in the rank statistic is less. Now our inference at the top of the rank statistic in column seven is quite similar to Stern’s ranks. the *Journal of Economic Literature* and *Quarterly Journal of Economics* are contained in the subset, while *American Economic Review*, *Journal of Economic Perspectives*, *Journal of Financial Economics*, and *Journal of Political Economy* are contained in the subset of the second best, each with 95% confidence. Using the Bonferonni inequality, we can conclude that both statements are simultaneously true with 90% confidence. Journals that receive a “-” in the seventh column are part of the ranking analysis of 30 journals, but did not make it into the subsets of 1st or 2nd best. One can conclude that journals like *Econometrica* and *Journal of Economic Geography* (J ECON GEOGR) are not among the 1st or 2nd best journals. One cannot conclude this from Stern’s Figure 2.

### 3.1 Discussion of Comparisons

Conventional wisdom might suggest that our result that *Econometrica* and *American Economic Review* are not in the subset of the best, while *Journal of Economic Perspectives* is, is simply wrong and that this is indicative of a fundamental shortcoming of classical statistical inference. The inference penalizes
Econometrica and American Economic Review for their relatively precise impact factor estimates, while rewarding Journal of Economic Perspectives for its imprecisely estimated impact. This may be driven by the number of articles published in each journal (a sample size argument), but it may also be due to the fact that Journal of Economic Perspectives may occasionally publish some highly influential papers, despite its lower average impact. If this is the case, then perhaps Journal of Economic Perspectives deserves to be in the subset of the best.

That said, the issue may be related to differences in the editorial strategies at different journals. Journal of Economic Perspectives publishes many survey articles written by experts in various subfield of economics. This retrospective approach may lead to citation counts and impacts that are different (by design) than those at Econometrica and American Economic Review, whose editorial staff may choose a more forward-looking approach to publishing. More generally, the heterogeneity of precision (the perceived heteroskedasticity) may simply be due to differences in editorial philosophy, and perhaps a “better” ranking would only include journals with similar philosophies (for example, Baltagi’s (2007) ranking of Econometrics journals). Clearly, Journal of Economic Perspectives and Journal of Finance are markedly different journals than Econometrica and American Economic Review not only in terms of their readership, but in terms of their publication goals. Perhaps comparing them is simply not justified. Nonetheless, our multiple comparison inference approach sheds light on these issues that individual t-tests do not.

With multiple comparison procedures one must be careful not to let preconceived notions of journal quality interfere with the results of the statistical exercise. Moreover, these rankings here are based exclusively on variation in the five year impact factor. If an alternative metric was used, a different ranking outcome may arise. The robust metric of Chang et al. (2016) could supplement here.

9This is seen in Table 1 with non-monotonocity of the rankings and the impact factor point estimates.
4 Conclusions

This study has extended the work of Stern (2013) to account for multiplicity, as well as uncertainty in the ranking of economics journals. This is an important consideration given the weight that many academic departments and universities place on journal rankings in promotion and tenure decisions as well as how academics themselves rank journals and departments. Moreover, hiring decisions of new PhDs are likely to be influenced by the perceived quality of the journals which the candidates have either already published in, or are expected to publish in over the course of their tenure. We see that accounting for multiplicity produces several distinct groups of journals which are statistically equivalent in their rankings. This procedure allows for the fact that citations of articles, the foundation for many journal rankings, are random and can vary substantially from year to year and issue to issue. Further, Stern’s (2013) inferential exercise, while carried out at the 5% level, actually possesses an error rate of about 9% given the avoidance of multiplicity; this casts more uncertainty on the statistical equivalence of the journal rankings.

An interesting extension would be to map the uncertainty in journal rankings into uncertainty into department rankings. At present it is not clear exactly how the uncertainty in citations that leads to stochasticity in journal rankings is transferred to department rankings. Given the large number of PhD programs globally, multiple comparison procedures would certainly be useful when making confidence statements regarding department rankings. A further extension would be to account for uncertainty in various forms of journal rankings which have been produced; some methods of ranking have less variation, which will necessarily produce more accurate rankings.
References


Table 1: Top 16 Journals from Stern’s Ranking with Subset Rankings at 95% Level. KMS is the rankings from Kalaitzidakis et al. (2003).

<table>
<thead>
<tr>
<th>Title</th>
<th>Impact $(\overline{y}_j)$</th>
<th>Std. Err. $(s_j/\sqrt{T})$</th>
<th>Stern 230 $(t_{j,∞,0.95})$</th>
<th>Stern 230 Subset</th>
<th>KMS 30 $(t_{j,∞,0.95})$</th>
<th>KMS 30 Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>J ECON LIT</td>
<td>9.281</td>
<td>1.380</td>
<td>2.552</td>
<td>1st best</td>
<td>2.369</td>
<td>1st best</td>
</tr>
<tr>
<td>Q J ECON</td>
<td>8.261</td>
<td>0.573</td>
<td>2.964</td>
<td>1st best</td>
<td>2.713</td>
<td>1st best</td>
</tr>
<tr>
<td>J FINANC</td>
<td>6.173</td>
<td>0.327</td>
<td>3.211</td>
<td>1st best</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>AM ECON REV</td>
<td>6.059</td>
<td>0.215</td>
<td>3.358</td>
<td>2nd best</td>
<td>2.998</td>
<td>2nd best</td>
</tr>
<tr>
<td>J ECON PERSPECT</td>
<td>6.027</td>
<td>0.596</td>
<td>2.942</td>
<td>1st best</td>
<td>2.693</td>
<td>2nd best</td>
</tr>
<tr>
<td>J FINANC ECON</td>
<td>5.730</td>
<td>0.344</td>
<td>3.183</td>
<td>2nd best</td>
<td>2.873</td>
<td>2nd best</td>
</tr>
<tr>
<td>J POLIT ECON</td>
<td>5.050</td>
<td>0.439</td>
<td>3.084</td>
<td>2nd best</td>
<td>2.805</td>
<td>2nd best</td>
</tr>
<tr>
<td>REV FINANC STUD</td>
<td>4.827</td>
<td>0.475</td>
<td>3.048</td>
<td>2nd best</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>ECONOMETRICA</td>
<td>4.567</td>
<td>0.381</td>
<td>3.147</td>
<td>-</td>
<td>2.840</td>
<td>-</td>
</tr>
<tr>
<td>J ECON GEOGR</td>
<td>4.293</td>
<td>0.424</td>
<td>3.099</td>
<td>-</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>J ACCOUNT ECON</td>
<td>4.171</td>
<td>0.382</td>
<td>3.136</td>
<td>-</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>AM ECON REV P&amp;P</td>
<td>4.135</td>
<td>0.182</td>
<td>3.423</td>
<td>-</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>J ECON GROWTH</td>
<td>4.117</td>
<td>0.670</td>
<td>2.886</td>
<td>2nd best</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>REV ECON STUD</td>
<td>4.097</td>
<td>0.317</td>
<td>3.215</td>
<td>-</td>
<td>2.901</td>
<td>-</td>
</tr>
<tr>
<td>AM ECON J - MACRO</td>
<td>4.073</td>
<td>0.584</td>
<td>2.950</td>
<td>-</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>REV ECON STAT</td>
<td>3.764</td>
<td>0.307</td>
<td>3.222</td>
<td>-</td>
<td>2.902</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1.
NR = Not Ranked: journal excluded from this particular ranking.
1st best = Journal is contained in the subset of best ($S_{0.95}$) at the 95% level.
2nd best = Journal is contained in the subset of second best ($S^*_{0.95}$) at the 95% level.
“-“ = Journal included in analysis, but is not in the subset of best ($S_{0.95}$) or 2nd best ($S^*_{0.95}$) at the 95% level.