Incidental findings on multimodel comparison, hypothesis testing error, and its solution:
The case of information, knowledge, and career pursuit

Weng Marc Lim
Swinburne University of Technology
Adjunct Professor, Swinburne Business School, Swinburne University of Technology, Australia
Professor and Head of School, School of Business, Swinburne University of Technology, Malaysia
lim@wengmarc.com
marclim@swin.edu.au,
wlim@swinburne.edu.my

Symeon Mandrinos
Lecturer, School of Business,
Swinburne University of Technology
Malaysia

Abstract

This article sheds light on an incidental discovery of hypothesis testing error and its solution in a study that compared the multimodel effects of information and knowledge on career decisions. Using a sample of 500 second and third year undergraduates who were simple randomly recruited from 10 simple randomly selected universities, the study demonstrates how hypothesis testing error could occur when multimodel comparison of the effects of the same set of independent constructs (information and knowledge) on a different but related set of dependent constructs (career pursuit and rewarding career pursuit) is performed and how that error can be avoided through detailed scrutiny of the psychic distance between the dependent constructs. It is hoped that the insights herein will be useful for scholars to avoid and overcome the pitfall of hypothesis testing error arising from multimodel comparison involving a different but related set of dependent constructs.

Keywords: Multimodel comparison; hypothesis testing error; psychic distance; information; knowledge; career pursuit; rewarding career pursuit

1   Introduction

Incidental findings, which encapsulate peculiar insights discovered beyond the goals of the study, have become increasingly common due to the proliferation and omnipresence of advanced technologies and methodologies that empower researchers to perform highly sophisticated and rigorous investigations (Booth et al., 2012; Wolf et al., 2008). However, these unexpected, but potentially relevant, findings are often perceived as trivial due to the initial impression that the term "incidental findings", unfortunately, radiates (Booth et al., 2012; *Nature*, 2005).

In this paper, we contend that incidental findings can be ground-breaking. To support our contention, we report our incidental discovery of hypothesis testing error and its solution when we conducted an investigation that compared the multimodel effects of information and

knowledge on career decisions. In the sections that follow, we explain how and why the hypothesis testing error and its solution in our study, which we uncovered incidentally, differ from that which has been traditionally discussed and reported in the existing literature. In doing so, we demonstrate how our incidental findings contribute to theory, specifically by delineating an unconventional way on how hypothesis testing error could occur and how such an error could be resolved, and to practice, specifically by guiding the development of accurate managerial recommendations through rigorous scrutiny of empirical findings.

2   Theoretical background

2.1   Multimodel comparison

Multimodel comparison is a method that is often used but rarely defined, explicitly, by past scholars, as evidenced by a Google Scholar search for "multimodel (and multi-model) comparison" that yield no result that could offer a concrete, explicit definition of the method. Nonetheless, we did consider, in consultation with experts, how multimodel comparisons have been operationalized in the extant literature, which led to our conclusion that it is arguably safe to say, explicitly, that multimodel comparison generally refers to "the appraisal of two or more models." This appraisal, which may be conceptual (e.g., Lim, 2018; Wilson, 1999) or empirical (e.g., Ba et al., 2014; Gillingham et al., 2018), is typically undertaken to observe the similarities and differences in the models under study. In doing so, we find that multimodel comparison allows appraisers to delineate the distinct and overlapping boundaries of different models as well as the magnitude and replicability of effects across studied models (e.g., Ba et al., 2014; Gillingham et al., 2018; Lim, 2018; Wilson, 1999).

2.2   Hypothesis testing

A hypothesis is a testable assumption made about a target population. To test this assumption, a formal procedure in the form of hypothesis testing is undertaken, whereby the goal of hypothesis testing is not the verification but the falsification of the null hypothesis (H_0) (i.e., non-existence of association) in favour of the alternative hypothesis (H_a) (i.e. existence of association) using statistical tests on usable data from a sample of the target population (Banerjee et al., 2009; Wasserstein & Lazar, 2016). The probability value (i.e., *p*-value), which indicates the probability to obtain an effect equal to or more extreme than the one observed when the null hypothesis is presumed to be true (and not the probability of the null hypothesis being true), is a statistical measure of the strength of evidence that can be used to guide the rejection or non-rejection of the null hypothesis in hypothesis testing (Biau et al., 2010; Wasserstein & Lazar, 2016).

Four outcomes are possible from hypothesis testing: two outcomes, whereby one is a true positive that rejects the null hypothesis and the other is a true negative that fails to reject the null hypothesis, suggest that the sample and the reality in the population are concordant, whereas the remaining two outcomes suggest that an incorrect inference would have been made as a result of making a type one error or a type two error (Biau et al., 2010; Rothman, 2010). More specifically, a type one error is a false positive, whereby a null hypothesis that is true is rejected, whereas a type two error is a false negative, whereby a null hypothesis that is false is not rejected (Rothman, 2010). To mitigate hypothesis testing error, the existing literature suggests that using a smaller level of statistical significance for rejecting or not rejecting the null hypothesis can help to minimize type one error, that using a larger sample size can help to reduce the probability of type two error, and that reporting effect size (i.e., the

arithmetic or geometric mean of the two confidence interval limits) and precision (i.e., the narrowness of the confidence interval) is encouraged to avoid misinterpretation and to enrich interpretation of study results (Banerjee et al., 2009; Biau et al., 2010; Rothman, 2010; Wasserstein & Lazar, 2016).

3   Method

3.1   The study (and its hypothesis development)

In our study, we examined the relationship between information, knowledge, and career decisions. In particular, we wanted to investigate the extent to which information and knowledge had an influence on career decisions. To do so, we made two contentions. First, we contended that information is sustainable on the basis that it can be stored permanently (e.g., information centre, such as library) whereas knowledge is transient on the basis that it is an elusive entity that can either diminish or evolve over time (e.g., our understanding of what we have learnt at the university) (Zins, 2007). Second, we contended that investments in career decisions (e.g., to pursue not only a career [e.g., full-time job], but also a rewarding career [e.g., high income full-time job]) is determined by both information and knowledge, though information is posited to have a greater effect than knowledge on the basis that the sustainability of information provides greater assurance than the transience of knowledge in decision making (Gupta et al., 2018). Thus, we hypothesized that:

H1.   Information encourages more than just a career pursuit.

H2.   Knowledge encourages more than just a career pursuit.

H3.   Information is more profound than knowledge for encouraging more than just a career pursuit.

3.2   The instrument

A survey approach by means of questionnaire distribution was used to collect data for the study. More specifically, four single-item measures on a five-point Likert scale (i.e., '1' indicating 'strongly disagree' and '5' indicating 'strongly agree') were used to measure the four constructs under study. More specifically, the items measuring the constructs pertaining to information, knowledge, career pursuit, and rewarding career pursuit are presented accordingly, as follows: "The career centre provides me with career information", "I am able to see clearly the potential career rewards", "I plan to be employed full-time upon graduation", and "I am willing to work hard to achieve a rewarding career". The rationale for taking a unidimensional approach to construct measurement was to allow for a macro-understanding of the evaluation of information, knowledge, and career decisions, which is consistent with existing studies that suggest that taking a unidimensional approach to form a macro-understanding of the researched phenomenon is acceptable as no significant differences generally exist between multi- and single-item measures (Bergkvist & Rossiter, 2007; Walsh & Mitchell, 2010). More important, the questionnaire containing these items were pre-tested and pilot-studied prior to administration in the main study.

3.3   The sample

The sample consisted of 500 second and third year undergraduates who were simple randomly recruited from 10 simple randomly selected universities over two years from 2016 to 2018 (see Lim and Ting [2012] for more information on simple random sampling). The

sample size of 500 is deemed to be large and thus allowed us to mitigate type two error. The choice of target population for our sample, which we used as a sampling criterion, was predicated on the fact that second and third year undergraduates have normally decided on their academic major that will influence the choice of the professional career in the future, which is in line with the focus of the dependent constructs under study (i.e., career pursuit and rewarding career pursuit). More important, the students in our sample participated in our study voluntarily and were assured of their anonymity as no personal identifiable data was collected. Most of them were locals (99.0%), females (65.6%), and second and third year undergraduate students enrolled across eight public universities (87.2%). No missing data is reported as participants in the study had to complete all questions before returning the questionnaire to the surveyor.

4 Findings

Multiple regression analyses were performed to examine whether information (H1) and knowledge (H2) encourages more than just a career pursuit using a sample of second and third year undergraduates. Importantly, we used a small *p*-value of .01 to mitigate type one error and we furbished details on confidence interval, effect size, and precision to enrich our reporting of results.

Figure 1 presents the results of the first regression, which suggests that information (β = .213, *p* = .000 < .01, CI = .125, .304, ES = .215, P = .179) and knowledge (β = .214, *p* = .000 < .01, CI = .124, .302, ES = .213, P = .178) encourage career pursuit, whereas Figure 2 presents the results of the second regression, which suggests that information (β = .129, *p* = .000 < .01, CI = .050, .208, ES = .129, P = .158) and knowledge (β = .297, *p* = .000 < .01, CI = .219, .376, ES = .298, P = .157) encourage a rewarding career pursuit.[1] Multimodel comparison of the first and second regression suggests that a strong case can be built to support H1 and H2.
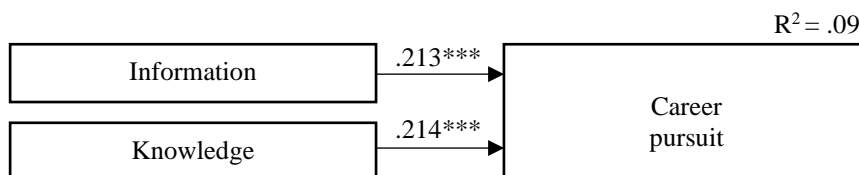

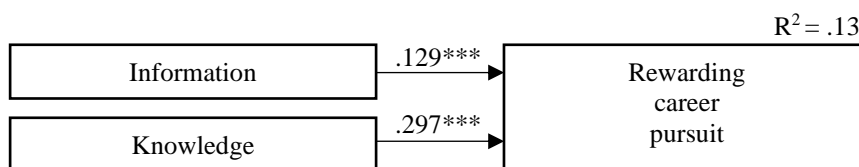*Figure 1. Information, knowledge, and career pursuit*

Figure 2. Information, knowledge, and rewarding career pursuit

More specifically, the first regression offers evidence to first establish that information and knowledge encourage career pursuit, which in turn, allows the second regression, which offers evidence to then establish that information and knowledge encourage rewarding career

---

[1] β = regression coefficient, *p* = significance value, CI = confidence interval, ES = effect size, and P = precision.

pursuit, to build on those findings to collectively provide support for H1 and H2 with medium effect size and reasonably narrow precision (Rothman, 2010; Durlak, 2009).

However, a multimodel comparison of the first and second regression suggests that knowledge has a more profound impact than information for encouraging more than just a career pursuit, which is in contrast to our third hypothesis. This finding, in turn, does not provide support for H3.

Upon further scrutiny, we incidentally discovered that we had conducted a type two error as we mistakenly rejected an alternative hypothesis in the form of H3, and in doing so, not rejecting a null hypothesis that is false. To elaborate, our further scrutiny was undertaken through the computation of psychic distance, which is the perceived distance between two different but related dependent constructs (Sousa & Bradley, 2006). When contextualized in our study, psychic distance offers an alternative measure to test our third hypothesis, specifically by examining whether information and knowledge can encourage more than just a career pursuit when the psychic distance between career pursuit and rewarding career pursuit is reduced, an advantage that we were able to leverage as a result of collecting data for the purpose of multimodel comparison. More specifically, psychic distance in our study was computed by measuring the gap between career pursuit and rewarding career pursuit, where the gap of '0', '1', '2', and '3' that emerged were coded as 'no', 'small', 'medium', and 'large' psychic distance. Following the computation of psychic distance, we ran an analysis of variance to scrutinize the effect of information and knowledge in encouraging more than just a career when psychic distance is reduced. Findings from this analysis, which are reported in Table 1, suggest that information had a significant effect in encouraging more than just a career pursuit when psychic distance is reduced from large to no psychic distance ($F = 20.33$, $p = .00 < .01$), whereas knowledge did not produce any significant improvements in encouraging more than just a career pursuit when psychic distance is reduced from large to no psychic distance ($F = .73$, $p = .39 > .01$). These findings, in turn, nullified type two error to provide support for H3.

Factor	Psychic distance	N	Mean	Standard deviation	Standard error	Weighted linearity		Equality of means			
						F	Sig.	Welch	Sig.	Brown - Forsythe	Sig.
Information	No	203	3.45	.86	.06	20.33	.00	6.62	.00	5.96	.00
	Small	199	3.17	.95	.07						
	Medium	68	2.94	1.18	.14						
	Large	30	2.90	1.09	.20						
Knowledge	No	203	3.79	.83	.06	.73	.39	.72	.54	.68	.57
	Small	199	3.87	.91	.07						
	Medium	68	3.68	1.30	.16						
	Large	30	3.63	1.40	.26						

Table 1. Information, knowledge, and psychic distance of career pursuit

5 Conclusion

In summary, this article has shed new light on an incidental discovery of hypothesis testing error and its solution in a study that compared the multimodel effects of information and

knowledge on career decisions in the form of career pursuit and rewarding career pursuit. The incidental findings reported herein should arguably be substantively significant as the discovery was in light of the implementation of the recommendations by past scholars to mitigate hypothesis testing error. While it is hoped that the insights herein will be useful for scholars to avoid and overcome the pitfall of hypothesis testing error arising from multimodel comparison involving a different but related set of dependent constructs, we concede that further exploration remains necessary to uncover the potential of other situations where hypothesis testing error may occur in multimodel comparison and the potential solutions that can correct for that error.

Declaration of conflicting interests

Funding

References

Ba, J., Keenlyside, N. S., Latif, M., Park, W., Ding, H., Lohmann, K., Mignot, J., Menary, M., Otterå, O. H., Wouters, B., & Melia, D. S. (2014). A multi-model comparison of Atlantic multidecadal variability. *Climate Dynamics*, *43*(9–10), 2333–2348. https://doi.org/10.1007/s00382-014-2056-1

Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal*, *18*(2), 127–131. https://doi.org/10.4103/0972-6748.62274

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, *44*(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175

Biau, D. J., Jolles, B. M., & Porcher, R. (2010). P value and the theory of hypothesis testing: An explanation for new researchers. *Clinical Orthopaedics and Related Research*, *468*(3), 885–892. https://doi.org/10.1007/s11999-009-1164-4

Booth, T. C., Waldman, A. D., Wardlaw, J. M., Taylor, S. A., & Jackson, A. (2012). Management of incidental findings during imaging research in "healthy" volunteers: Current UK practice. *British Journal of Radiology*, *85*(1009), 11–21. https://doi.org/10.1259/bjr/73283917

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928. https://doi.org/10.1093/jpepsy/jsp004

Gillingham, K., Nordhaus, W., Anthoff, D., Blanford, G., Bosetti, V., Christensen, P., McJeon, H., & Reilly, J. (2018). Modeling uncertainty in integrated assessment of climate change: A multi-model comparison. *Journal of the Association of Environmental and Resource Economists*, *5*(4), 791–826. https://doi.org/10.1086/698910

Gupta, G., Tan, K. T., Ee, Y. S., & Phang, C. S. (2018). Resource-based view of information systems: Sustainable and transient competitive advantage perspectives. *Australasian Journal of Information Systems*, *22*. https://doi.org/10.3127/ajis.v22i0.1657

Lim, W. M. (2018). Dialectic antidotes to critics of the technology acceptance model: conceptual, methodological, and replication treatments for behavioural modelling in technology-mediated environments. *Australasian Journal of Information Systems*, 22. https://doi.org/10.3127/ajis.v22i0.1651

Lim, W. M., & Ting, D. H. (2012). *A toolkit of sampling and data analysis techniques for quantitative research*. Munich, Germany: GRIN Verlag.

Nature. (2005). How volunteering for an MRI scan changed my life. *Nature, 434*(7029), 17. https://doi.org/10.1038/434017a

Rothman, K. J. (2010). Curbing type I and type II errors. *European Journal of Epidemiology, 25*(4), 223–224. https://doi.org/10.1007/s10654-010-9437-5

Sousa, C. M., & Bradley, F. (2006). Cultural distance and psychic distance: Two peas in a pod? *Journal of International Marketing, 14*(1), 49–70. https://doi.org/10.1509/jimk.14.1.49

Walsh, G., & Mitchell, V. W. (2010). The effect of consumer confusion proneness on word of mouth, trust, and customer satisfaction. *European Journal of Marketing, 44*(6), 838–859. https://doi.org/10.1108/03090561011032739

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *American Statistician, 70*(2), 129-33. https://doi.org/10.1080/00031305.2016.1154108

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249–70. https://doi.org/10.1108/EUM0000000007145

Wolf, S. M., Lawrenz, F. P., Nelson, C. A., Kahn, J. P., Cho, M. K., Clayton, E. W., Fletcher, J. G., Georgieff, M. K., Hammerschmidt, D., Hudson, K., & Illes, J. (2008). Managing incidental findings in human subjects research: Analysis and recommendations. *Journal of Law, Medicine & Ethics, 36*, 219–248. https://doi.org/10.1111/j.1748-720X.2008.00266.x

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology, 58*(4), 479–493. https://doi.org/10.1002/asi.20508