Authorship, collaboration, topics, and research gaps in Environmental and Resource Economics 1991-2015

Maksym Polyakov ^a, Morteza Chalak ^a, Md Sayed Iftekhar ^a, Ram Pandit ^a,
Sorada Tapsuwan ^{a, c}, Fan Zhang ^a, and Chunbo Ma ^{b *}

^a Centre for Environmental Economics and Policy, UWA School of Agriculture and Environment, University of Western Australia, Perth, WA 6009, Australia

^b UWA School of Agriculture and Environment, University of Western Australia, Perth, WA 6009, Australia

^c CSIRO Land and Water Flagship, Commonwealth Scientific and Industrial Research Organisation, Floreat, WA 6014, Australia

* Corresponding author. E-mail address: chunbo.ma@uwa.edu.au; Tel.: +61 8 6488 2534; fax: +61 8 64881098

Abstract

Environmental and Resource Economics is one of the premier journals in the field of environmental economics. It was established with an aspiration to focus more on applied and policy relevant research compared to other established journals, and to establish better channels of communication and collaboration between researchers from Europe and other parts of the world. We present a text based exploratory analysis of 1,630 articles published in the Journal from 1991 to 2015 that suggests the Journal has been somewhat successful in meeting both these aims. Perhaps more importantly, it shows the Journal continues to progress toward these goals. The European authors are the largest contributors to the Journal, which is in contrast to other prominent journals (such as *Journal of Environmental Economics and Management* and *Ecological Economics*). And while most of the collaboration has occurred within this geographic region (e.g., European authors collaborated with other European authors more frequently), this trend appears to be changing as the proportion of articles written by international collaborators is gradually increasing. Topic analysis reveals that almost all of the articles could be grouped under applied and / or policy relevant topics, and almost two-thirds of the articles are empirical in nature, which suggest that the journal has been able to fulfil both of its commitments. We also investigate trends in research foci over the last twenty-five years and what kind of research gaps can be discerned.

Keywords: Topic analysis; Latent Dirichlet Allocation; Co-authorship; Environmental and Resource Economics.

JEL Classification A11, A14, Q0.

1 Introduction

In the first issue of *Environmental and Resource Economics* (ERE) Opschoor and Jansen (1991, p. v) suggested that in addition to the *Journal of Environmental Economics and Management* (JEEM) and *Ecological Economics* (EE), there is a need for another journal with more applied and policy relevant content. They also suggested that ERE can be a bridge between European and international environmental economists. We investigate how, in the past 25 years, the Journal has progressed towards the goals of focusing more on applied and policy relevant research and acting as a bridge between European and other researchers.

In addition, this paper sheds light on how the field and the research enterprise have evolved. Given that ERE is one of the well-respected outlets in the field of environmental and resource economics, results of our paper will provide an interesting comparison with other works, particularly on the evolution of JEEM (Shogren and Durden 1991; Brookshire and Scrogin 2000; Schymura and Löschel 2014). A systematic review of the structure and dynamics of topics published in ERE will also provide an alternative account of how the field has evolved over a quarter of a century. Examining the dynamics of publication and citation of a broad range of topics allows identification of the areas where ERE has been influential to suggest future research direction. Combined with the analysis of topics co-occurrence in multidisciplinary research, it allows uncovering research gaps, where potential may exist to make progress at the intersection of the fields. In addition, trends in national, institutional and individual contributions, co-authorship, gender and geographical coverage of authorship in ERE reveal the evolution of the research enterprise itself over the years. For example, are authors collaborating more? Is more research getting funded? What is the extent of informal collaboration? Is there any geographical segregation or pattern in terms of topic published or nature of collaboration?

Similar analyses have been conducted for general interest economics journals (Hamermesh 2013) as well as field journals in the area of environmental and resource economics. There have been plenty of reviews done for the JEEM (Shogren and Durden 1991; Brookshire and Scrogin 2000; Schymura and Löschel 2014), the American Journal of Agricultural Economics (Debertin and Pagoulatos 1992), the Australian Journal of Agricultural and Resource Economics (Polyakov et al. 2016), Ecological Economics (Costanza et al. 2004; Ma and Stern 2006; Costanza et al. 2016). But, to our knowledge, no reviews of this nature have been done for ERE.

In this article, 1,630 research articles published in ERE between 1991 and 2015 have been analysed to identify and provide a quantitative assessment of the changing interests and influences in research foci, as well as patterns of authorship and collaboration among environmental and resource economists as reflected in the articles. The remainder of the paper proceeds as follows. Section 2 presents the data and method used for analysis. Section 3 presents the results and discussion which are organized in terms of contributors to the journal and content published in the journal. Finally, Section 4 concludes the paper.

2 Data and Methods

This section first describes the protocol for data collection, followed by the introduction of models that define the research topics to be used for subsequent analyses. We then present our approaches to the analysis of topic popularity, similarity, specificity, and research gaps.

2.1 Data

Our analysis covers all research articles published in ERE from 1991 to 2015, excluding editorials, comments and replies, and book reviews (1,630 articles in total). The data for every article, including title, authors and their affiliations, year of publication, volume, issue,

keywords and abstracts, were downloaded from Springer (www.springer.com) in November 2015. Data on the number of times each article was cited were obtained from the Scopus database in April 2016.

Information on article categories, number of tables, figures, and equations, and the presence of acknowledgement to colleagues and funders were recorded by manually inspecting each article. To ensure robustness, each article was inspected independently by two authors. If inconsistencies between the two occurred, points of contention were resolved in group meetings with all authors. Following classifications by Backhouse (1998) and Schymura and Löschel (2014), the articles were grouped into five categories: theoretical, empirical-qualitative, empirical-quantitative, theoretical-empirical, and review. A “*theoretical*” article develops a theoretical (e.g., algebraic) model based on abstract assumptions and a set of conclusions are reached based on that model. An “*empirical-qualitative*” article uses empirical data. The article could have a theoretical or methodological component, but results are based on data analysis. However, articles in this category are mostly descriptive in nature and any statistical analysis does not go beyond providing descriptive statistics. An “*empirical-quantitative*” article uses empirical data. Articles could have a theoretical or methodological component, but results are based on the analysis of data using a statistical, simulation, or optimisation model. A “*theoretical-empirical*” article contains both theoretical and empirical components. There is a theoretical (e.g. algebraic) model based on abstract assumptions and a set of conclusions is reached. In addition, the model is applied to a specific case using real data based on a statistical, simulation, or optimisation technique. Finally, we distinguish “*review*” articles.

The gender of authors was determined using names as the identifier or by an internet search for additional information, including photographs. This allowed us to assign a gender to 99%

of the 3,468 co-authors. Institutions, localities and countries of the authors were extracted from the authors' affiliations using regular expression also known as search patterns (*prxchange* function in SAS v 9.3). The coordinates of each author's location were calculated using Google Maps Application Programming Interface (API). The spatial links between the co-authors in each article were created in ArcGIS.

2.2 Topic modelling

With emergence of large collections of digital text documents researchers need new analysis tools capable of handling large datasets and capable utilising readily available text data. Information scientists have developed topic modelling tools based on text analysis and mining techniques, which can be used to discover of themes and topics of large unstructured collections of documents (Karl et al. 2015). Topic modelling tools have been widely applied in many disciplines including political science (Grimmer and Stewart 2013), journalism (Rusch et al. 2013), health (Paul and Dredze 2014), conservation biology (Westgate et al. 2015) and general science (Griffiths and Steyvers 2004). They can be useful to investigate trends and evolution of specific terms and identify potential research gaps. Applications of these methods can also provide information that is complementary to that given by systematic reviews and meta-analyses. However, applications of these tools in the field of economics have been very limited so far.

There are several topic modelling techniques, the most popular being Latent Semantic Analysis (LSA) (Furnas et al. 1988; Deerwester et al. 1990) and Latent Dirichlet Allocation (LDA) (Blei et al. 2003). In this study we model topics of the articles using LDA. The LDA model assumes that each document in a corpus of documents can be described as a mixture of latent (unobservable) topics with certain probabilities. Each latent topic can be described by a collection of terms (words) with certain joint probability distribution. We observe occurrence

of terms in the documents, and the goal of LDA is to automatically discover latent topics using a statistical model which assigns the probabilities that the topics are contained in a specific document and the probabilities of the terms occurrences in a specific topic.

Formally, the probability of term w_i occurring in a given document d_k can be expressed as

$$P(w_i | d_k) = \sum_{z=1}^Z P(w_i | z) P(z | d_k), \quad (1)$$

where z is the latent topic, $P(w_i | z)$ is the probability of word i within the topic z , and $P(z | d_k)$ is the probability of choosing a word from topics z in the document k . The number of latent topics Z has to be pre-defined and allows adjusting the degree of specialisation of the latent topics. LDA estimates the topic-term distribution $P(w | z)$ and the document-topic distribution $P(z | d)$ in the corpus of documents using an Expectation Maximisation algorithm. In practice, the true number of topics is unknown. However, one could fit LDA model with different number of topics and choose the “best” model based on a selection criterion; in this study we use the Akaike Information criterion (AIC).

We used words from title, keywords, and abstracts for topic modelling. Text pre-processing was done using R package *tm* (Feinerer 2015). It included converting to American spelling, removing common words (such as “the”), and stemming (words like “fisheries” and “fishery” were converted to stem “fishery” in order to analyse them as one term). LDA was performed using package *topicmodel* (Grün and Hornik 2011) in R statistical software (R Core Team 2015). The optimal number of topics was selected by estimating a set of LDA models with the number of pre-defined topics ranging from 5 to 30 and selecting the model using either most meaningful grouping or AIC as optimality criteria. For this analysis, we selected the

model with 15 topics because it had the lowest AIC and produced the most meaningful grouping of words and articles. The names of the topics were assigned by inspecting the words with the highest probabilities of occurrence in each topic $P(w|z)$. The names were verified by examining the titles of the articles that had the highest probability for each topic $P(z|d)$.

2.3 Topics similarity, popularity, specificity, and research gaps

To explore the nature and meaning of topics, as well as topic similarities and differences, we used cluster analysis (Sievert and Shirley 2014). We used Jensen-Shannon Divergence (JSD) (Endres and Schindelin 2003; Hall et al. 2008) between two topic-term probability distributions as a measure of distance between a pair of topics. JSD is a metric used to measure the difference and similarity between two probability distributions, which is based on the Kullback-Leibler divergence (KLD) (Kullback and Leibler 1951). While KLD is the directed divergence between two distributions that could be interpreted as information gain achieved when we use one distribution instead of another, JSD is a symmetric metric, that is $JSD(z_1, z_2) = JSD(z_2, z_1)$. Given a matrix of divergences, the clustering of topics can be analysed using graphical methods such as dendrograms (Westgate et al. 2015). To build the dendrogram, we used Ward's Agglomerative Hierarchical Clustering Method that is based on a sum-of-squares criterion and produces groups that minimise within-group dispersion at each junction (Murtagh and Legendre 2014).

The analysis reduced a corpus of articles to a set of topics, thereby enabling scrutiny on the level and the dynamics of the popularity of these topics. This further analysis gained insight into the prevalence of particular research and into the dynamics of the discipline. It has previously been recognised that topics in a scientific discipline not only generate different

amounts of interest, but the level of interest can rise or fall (Griffiths and Steyvers 2004).

Identifying “hot topics”, which are garnering lots of interests as measured through the number of papers and / or number of citations, may be useful both from the historical perspective and for anticipating future development of a discipline.

Popularity (level and dynamics) of topics can be analysed by modelling average annual probabilities of topics in published articles (Griffiths and Steyvers 2004) or number of articles assigned to a set of topics (Westgate et al. 2015). In this study, we calculated the average probability of topics by year as $\bar{p}_{z,t} = \sum_{k=1}^{K_t} P(z | d_{k,t}) / \sum_{z=1}^Z \sum_{k=1}^K P(z | d_{k,t})$ and estimated a linear mixed model $\bar{p}_{z,t} = \alpha + (\beta + v_z)t + u_z$ where α and β are the fixed intercept and slope, v_z is the random slope, u_z is the random intercept, and t is time (year relative to median year). Topics with positive random intercepts can be interpreted as having higher than average contribution to the body of literature, while positive random slopes indicate those topics growing at above average rates (“hot” topics).

A similar analysis can be conducted from the “demand” side by looking at the dynamics of citations attracted by the topics. For this, we calculated the number of citations attracted by each article for each year, attributed these citations to topics proportionally to the weight of topic in each article, and calculated the proportion of citation attributed to each topic z in each citation year τ : $\bar{c}_{z,\tau} = \sum_{t=1}^T \sum_{k=1}^K P(z | d_{k,t}) C_{k,\tau,t} / \sum_{z=1}^Z \sum_{t=1}^T \sum_{k=1}^K P(z | d_{k,t}) C_{k,\tau,t}$. The proportions of citations attracted by the topics in each year is modelled using a linear mixed model as well: $\bar{c}_{z,\tau} = \gamma + (\delta + y_z)\tau + w_z$, where γ and δ are the fixed intercept and slope, y_z is the random slope, w_z is the random intercept. The high values of random intercepts can be interpreted as indicating higher than average interest attracted to the topic, while positive

random slopes indicate the topics that are attracting more citations with time, or “hot” topics based on citations.

The topics identified using LDA could represent a broad range of themes, some of the themes being specific and very narrowly defined, while others, more general, broad and overlapping with other topics (Westgate et al. 2015). The articles were assumed to consist of mixtures of topics, and the weight of each topic in each article was captured in the document-topic probability matrix $P(z|d)$. Each document can be assigned the dominant topic, which is a topic that has the highest weight. For example document A consists of the mixtures of topics 1, 2 and 3 with weights, respectively, 85%, 10%, and 5%, while document B consists of topics 4, 5 and 6 with weights 50%, 30% and 20%. Topic 1 is dominant in document A and topic 4 is dominant in document B. The average weight of a topic for articles where the topic is selected as dominant is a measure of specificity. A high measure of topic’s specificity indicates that this topic heavily dominates the articles where it is selected and there is no place for other topics. In our example, topic 1 is dominant in document A and topic 4 is dominant in document B. Using averages for one document with topic 1 dominant and one document with topic 4 dominant, topic 1 is more specific than topic 4. The average weight of a topic in articles where it is not selected as dominant is a measure of generality. A high value of generality indicated that this topic has a higher co-occurrence with other topics. Here, topics 2, 3, 5 and 6 are not selected, with topic 3 least general and topic 5 most general.

It has previously been suggested that scientific breakthroughs and discoveries often happen at the intersections of disciplines or research fields within a discipline (Kates 1987). Analysis of the structure and dynamics of a discipline may help in identifying research gaps, and where new efforts could potentially make breakthroughs or discoveries. Specifically this can be done by looking at the growing (“hot”) research topics that did not have much interaction

with other topics (co-occurrence) in the body of articles. To calculate the co-occurrence $O_{i,j}$ of topics i and j we use cross-products of topic vectors from topic-document matrix:

$O_{i,j} = \sum_{i=1}^Z \sum_{j=1}^Z \sum_{k=1}^K P(z_i | d_k) P(z_j | d_k)$. The co-occurrences of topic pairs can be plotted and analysed in conjunction with the data about popularity and specificity to identify potential research gaps (Westgate et al. 2015).

3 Results and discussion

We start by examining the authorship of ERE: where are they located, who they work for and collaborate with. We then explore what research topics tend to be published in ERE and how they have evolved over time. Finally, we consider what research gaps there may be.

3.1 Journal contributors: spatial patterns and trends in authorship and collaboration

Figure 1 shows the proportions of author affiliations by regions and countries to the total number of articles published in the Journal. The affiliations were weighted by inverse number of authors in each article, so the sum of weighted affiliations adds to the number of articles. The European authors in ERE constituted 56% of the contributors, while most other leading environmental and resource economics journals were dominated by North Americans (for example, North American authors constituted more than 60% of authors in JEEM). Among the European countries, contributors from the UK (contributing 9% of the total number of articles), Germany (8%), Norway (6%) and the Netherlands (6%) dominated ERE. Scandinavian countries (Norway, Sweden, Denmark and Finland) together contributed 15% of the total number of papers. Nevertheless, the US was the largest contributing country (29%). Asia and Oceania contributed 6.1% and 3.5% of authorships, respectively, with Japan and Australia being the largest contributing country in Asia and Oceania respectively. South America and Africa contributed a mere 0.7% and 0.6%, respectively.

Figure 1. Authorship by countries and regions. The articles are assigned to countries and regions proportionally to number of co-authors

Figure 2 presents the spatial patterns of national and international co-authorship. The sizes of the dots indicate the weighted number of articles published by authors in each location and the lines represent co-authorship links between different locations. Consistent with Figure 1, most of the co-authors are located in Europe (greatest concentration in northern part of Germany, the Netherlands, and the UK) and in the US (greatest concentration in the eastern states). Similarly, the greatest concentration of collaboration links was between European countries followed by collaboration within North America, and between Europe and North America. Other noticeable collaboration patterns were observed between Australia and Europe, Australia and North America, Asia and North America, and Africa and North America. Remarkably, there was more collaborative links between Australia and Europe or North America than within Australia or between Australia and New Zealand. Similar patterns of inter- but not intra-continental collaboration are observed for Asia, Africa, and South

America. One of the aspirations of ERE was to bridge the distances between international environmental economists (Opschoor and Jansen 1991). These collaboration patterns seem to suggest that ERE has served the purpose and played an important role in linking European researchers with scholars in other parts of the world.

The dynamics of co-authorship and collaboration using 5-year moving averages are presented in Figure 3. Figure 3a shows two contrasting trends between the proportions of single-authored articles and three or more-authored articles. The proportion of single-authored articles steadily declined from over 50% in 1991 to approximately 20% in 2015. This decline has been compensated by the increasing trends of the proportions of articles by 3 or more authors, with a higher rate of increase for the proportion of four-plus authored articles since 2009. The proportion of two-authored articles shows no clear trend over the entire period, but it decreased initially from 1991 to 1996 then remained more or less stable. These trends have also been observed in JEEM (Schymura and Löschel 2014) as well as in *American Economic Review*, *Journal of Political Economy*, and *Quarterly Journal of Economics* (Hamermesh 2013).

Figure 2. Spatial patterns of authorship and collaboration. Size of the dots represents number of articles in each location. Thickness of the lines represents number of collaboration connections.

Trends in types of collaboration are presented in Figure 3b. International collaboration, where the co-authors are from different countries, increased steadily from close to 10% in 1991 to close to 40% in 2015, which is a four-fold increase over 25 years. National collaboration, where the co-authors are affiliated with different institutions in one country, showed an increasing trend between 1995 (about 20%) and 2005 (just over 30%), but at a slower rate compared to international collaboration, then remained stable until 2015 at around 28%. However, the institutional collaboration, where the co-authors are from the same institution, remained relatively stable over the entire period. Both increases in the number of authors and the proportion of international collaboration could be attributed to advances in communication technologies (Adams 2013).

There is a clear downward trend in the proportion of all-male-authored publications, and a clear upward trend in the proportion of male and female co-authorships (Figure 3c). An increasing contribution by female authors has also been observed in JEEM (Schymura and Löschel 2014) and Australian Journal of Agricultural and Resource Economics (Polyakov et al. 2016). The increase of proportion of female authors (from 4% to 22%) along with the increase of proportion of co-authored articles (Figure 3a) resulted in the increase of articles with male and female co-authors. This is backed up by the fact that the proportion of articles with female-only authors has remained relatively stable for 25 years, whilst male-only has fallen. The relatively low contribution of female authors could be further corroborated by the estimates presented in the Table A1 in the Appendix which shows that out of the top 50 most cited authors only two were female.

Figure 3. Trends of the proportions of articles by (a) number of authors; (b) type of collaboration; and (c) collaboration by gender.

3.2 Journal content: research topics and article categories

The topics that have been identified using LDA analysis are presented in Figure 4 using word clouds. Each cloud shows the 50 terms most frequently associated with respective topic. The sizes of the words are proportional to their probabilities in respective topics. It should be noted that some topics are dominated by one word that has a much higher probability than other words defining a topic. For example, for the topic ‘Environmental tax’ it was the word “tax”, for ‘Nonmarket valuation’ it was “WTP”, for ‘Recreation and amenity’ it was “transfer”, for ‘Fishery’ it was “fishery”, for ‘Energy’ it was “energy”, for ‘Forestry’ it was

“forest”, and for ‘Water’ it was, not surprisingly, “water”. Other topics were defined by several words with similarly high probabilities. For example, for ‘Conservation’ these words were “conserve”, “agriculture” and “biodiversity”, for ‘Voluntary environmentalism’ these were “green”, “adopt” and “innovation”, and for ‘Waste’ these words were “waste”, “recycle” and “household”. This means topics such as ‘Conservation’, ‘Voluntary environmentalism’ and ‘Waste’ were more heterogeneous and had higher potential for overlaps with other topics.

To further test the selection of topics we conducted a topics similarity analysis by calculating the Jensen-Shannon Divergence (2) and presented the results using a dendrogram (Figure 5a), which shows a clear intuitive grouping of topics. For example, ‘Fishery’, ‘Conservation’ and ‘Forestry’ are close to each other and far from the rest of the topics. They are close because both ‘Forestry’ and ‘Fishery’ study renewable resources and are often analysed with respect to conservation and protection of the environment. Similarly, ‘Global climate policy’ is often discussed in connection with ‘Environment and health’. ‘Nonmarket valuation’ is frequently used to estimate recreation and amenity benefits.

Figure 4. Word clouds of 50 most frequent terms by topic. Font size is proportional to the probability of a word in the topic.

Figure 5. Topics similarity (a), total number of articles by topic (b) and average citations per article by 2015 by topic (c)

Figure 5 (panels b and c) also presents distribution of articles among individual topics and average number of citations received by articles in each topic. The bars of both charts are aligned with the topics in the dendrogram. Each article was assigned probabilities of belonging to each topic (or weights of different topics in the article), therefore the numbers of articles and citations shown in Figure 5 are calculated as sums of these probabilities. The sums of probabilities were used elsewhere in this paper when we discussed distributions of articles by topics or citations attracted by the topics. Based on the average number of citations, the topics can be categorized into three groups: highly cited, moderately cited and least cited. The highly cited group ('Nonmarket valuation', 'Recreation and amenity', and 'Conservation') all had on average more than 20 citations per article, reflecting substantial interest to the developments of nonmarket valuation techniques and the associated applications in these subject areas. The top 30 most cited ERE articles have also been presented in the Appendix (Figure A2) which shows that more than half of the top 30 articles were from this highly cited group of topics. An article by Carson et al. (2001) with more than 400 citations and an article by Hanley et al. (1998) with more than 300 citations were classified as 'Nonmarket valuation' and 'Recreation and amenity', respectively. The moderately cited group had 16 to 20 citations per article, including 'Sustainable growth', 'Voluntary environmentalism', 'Global climate policy', 'Energy', 'Water', 'Forestry', and 'Waste'. The first three topics in this group also contributed a fair number of articles to the top 30 list (Figure A2). The least cited group had 15 or less citations per article.

Nonmarket valuation not only attracts a substantial number of citations, but it also produces a large number of articles (Figure 5b), indicating great interest in the developments and use of nonmarket valuation techniques in a range of disciplines and topics (as well as the ability of the Journal to attract high quality papers on this topic). However, the number of published articles did not necessarily follow the trend of citations – that is, topics producing a higher

number of articles were not always the ones attracting a higher number of citations. This is not surprising as the former may be influenced by many factors including the need for more research in those areas or choices made by editors, reviewers and researchers, publication of specialized journals (like *Nature Climate Change* and *Water Economics and Policy*) and the latter reflects on specific areas of the field the journal is having the most influence. For example, ‘Environmental tax’ had the highest number of published articles (160), although, it had one of the lowest average numbers of citations per article. This may be explained by the need for this research to be undertaken in order to support policy making, producing a high number of articles, while at the same time, receiving a relatively low interest outside of the field, reflected in lower citation rates. Conversely, the topics ‘Water’, ‘Forestry’ and ‘Waste’ all received substantial citations per article, while having the smallest number of articles published (77-81). This maybe a reflection of a general trend that applied and interdisciplinary topics, such as ‘Water’, ‘Forestry’ and ‘Waste’, as well as ‘Nonmarket valuation’ and ‘Recreation and amenity’ are being cited both in and outside of the field, thereby receiving more citation than theoretical topics such as ‘Environmental tax’, ‘Sustainable growth’, or ‘Global climate’.

The Journal’s content and specific topic’s impact is influenced by the Journal’s deliberate effort to address real-world environmental problems and policies by publishing special or themed issues. In a number of widely received special/themed issues (e.g., April 2016 Issue 4; October 2013 Issue 2; May 2007 Issue 1; July 2006 Issue 3; March 2006 Issue 3; June 2005 Issue 2), ERE organised focused discussions on how various environmental and resource economics theories and practices can help improve public decision making, design efficiency regulatory instruments and provide useful input to policy assessment. Specific policy areas where ERE has made dedicated contributions include economic valuation and management of various ecosystem services (May 2016, Issue 1), sustainable use of natural

resources (May 2015, Issue 1) and regional as well as global climate policies (October 2015, Issue 2). These special issues were often organized before a global policy forum (e.g., October 2015, Issue 2 before the Paris Forum) with the intention of influencing the discussion or to summarize a set of works to influence national environmental policies (e.g., May 2016, Issue 1 on Valuing Ecosystem Services for Improved National Accounting). Publications of these special issues indicate the journal's willingness to contribute to the policy debate but also attracting papers on specific topics. In addition, articles published in special/themed issues (SI) are much better received than those published in regular issues (RI). ERE started to publish special or themed issues since 1998. Figure 6 presents the average number of citations per article in SI and in the RI published in the same year. On average, SI attracted 90% more citations than RI.

Figure 6. Average number of citations per article for special issues and regular issues by year

Of the total 1,630 articles in our sample, about 42%, 36%, 16%, 3% and 3% articles fall into empirical-quantitative, theoretical, theoretical-empirical, empirical-qualitative and review categories, respectively. The 5-year moving average of the proportion of articles in theoretical and empirical-quantitative categories showed decreasing and increasing trends, respectively, while the articles in the other three categories generally remained stable over the years (Figure 7a). The proportion of theoretical papers started from approximately 50% and later fluctuates between 35% and 40%. The proportion of empirical-quantitative articles increased at a faster rate from about 25% in 1991 to about 50% in 2004, which then decreased to about 40% in 2009. The proportion of theoretical-empirical articles were similar in ERE and JEEM, but the proportion of empirical articles seemed greater in ERE than in JEEM (Schymura and Löschel 2014).

Analysis of the distribution of article categories by topics (Figure 7b) showed that most of the articles on ‘Nonmarket valuation’, ‘Recreation and amenity’, ‘Conservation’, ‘Energy’, ‘Water’, ‘Health’ and ‘Waste’ were empirical quantitative. These topics also had the smallest proportion of theoretical papers. Approximately equal proportion of theoretical, theoretical-empirical, and empirical-quantitative categories were published on ‘Fishery’, ‘Tradable permits’, ‘Global public policy’, ‘Trade and investment’, and ‘Forestry’ topics. These topics have a larger number of theoretical-empirical articles than other topics. Majority of the articles on the topics ‘Environmental tax’, ‘Voluntary environmentalism’, and ‘Sustainable growth’ were theoretical, and consequently, had the fewest articles under empirical quantitative and theoretical-empirical categories.

Figure 7. Article categories: dynamics (a) and distribution by topics (b)

In addition to the analysis of topics and article categories, it is important to look at the extent of informal collaboration and funding received by research resulting in different categories of articles published in the journal. Figure 8 presents the proportion of articles that acknowledge external funding and colleagues by the number of co-authors and article category. From panel (a) it was evident that the proportion of articles with acknowledgement of funding and the proportion of articles with acknowledgement of colleagues showed opposite trends with

respect to number of authors. The proportion of articles acknowledging people decreased as the number of authors increased. It is well known that acknowledgement of colleagues is a substitute to number of co-authors (Laband and Tollison 2000). On the contrary, the proportion of articles acknowledging external funding increased with the increases in number of co-authors. It signifies the potential that with a bigger project team there is a greater chance of getting more funding, hence more co-authors in publications.

Theoretical-empirical and empirical quantitative papers are more applied and therefore relevant to policy and practice making the outputs more useful to the funding agencies. Theoretical-empirical and empirical quantitative with original data collection (as opposed to those that use existing data sources) also require more data hence require more interaction with end users and funding bodies. However, we did not make distinction between two types of data used in empirical studies. Figure 8b shows that papers in these categories acknowledged research funding the most. Review and empirical qualitative articles acknowledged the funding bodies the least suggesting that they have the lowest funding requirements. Theoretical papers acknowledged funding more often than review and empirical qualitative papers possibly because theoretical research can take more time and effort (given challenges of contributing to the theory) than a review or an empirical quantitative study. This requires more contribution to the author's salary by the funding bodies, which results in higher acknowledgement of funding. In terms of proportion of articles acknowledging colleagues, there was not much difference between the different article categories.

Figure 8. Acknowledgment of external funding and colleagues by: (a) number of authors and (b) article category.

3.3 Topics popularity, specificity, and research gaps

Figure 9 presents a measure of level and change of a topics' popularity based on the distribution of topics in published articles and on the citations attracted to topics in published articles. The popularity based on published articles (Panel a) provides information on the allocation of total research effort to topics, while popularity based on citations (Panel b) provides information about academic demand for research on a particular topic. The changes in popularity can be used to assess which topics show growth ("hot" topics) or decline ("cold" topics) (Griffiths and Steyvers 2004). Figure A3 in Appendix presents the same information in greater detail by displaying the dynamics of 3-year moving averages of the proportions of topics in articles and of the proportion of citations attracted by articles published on each topic in all previous years.

Figure 9. Topics popularity, hot and cold topics based on articles (a) and citations (b)

Three topics, ‘Nonmarket valuation’, ‘Recreation and amenity’, and ‘Conservation’, were identified as popular and growing (“hot”) when measured by both number of articles and citations (Figure 9) as well as judged from trends shown in Figure A3. ‘Forestry’ and ‘Waste’ were less popular and exhibited downward trends by both measures indicating a declined interest in these topics. Declining number of articles on ‘Forestry’ may be explained by the emergence of *Journal of Forest Economics* (1995) and *Forest Policy and Economics* (2000). Several topics demonstrated relatively high popularity as measured by proportion of articles and citations, but declining trends. These topics were ‘Sustainable growth’, ‘Tradable permits’, ‘Energy’, and ‘Environmental tax’ (later being the topic with most published articles). ‘Sustainable growth’ had a large share of articles published in the early 1990s and accounted for nearly a quarter of all citations to articles in this Journal around 1995, when sustainability became a popular topic after the publication of Brundtland et al. (1987) report and the 1992 Earth Summit. ‘Global climate policy’ was popular by both publications and citation measures. However, while its share of publications increased, especially in the last five years, the share of citations it attracted declined. Whilst contributing a large proportion of articles to ERE, these topics attracted fewer citations than other topics published by the journal. In an opposite quadrangle of the graphs are two other topics, ‘Voluntary environmentalism’ and ‘Fishery’, which were relatively less popular as measured by both proportions of articles and citations but showed upward trends suggesting that these topics had a potential for future growth.

Finally, some topics showed different dynamics by publication and citation popularity metrics. Specifically, ‘Trade and investments’ was relatively less popular, however it displayed the fastest growth as measured by the number of articles and declining proportion of citations. Whilst the number of articles about ‘Water’ has declined, its citations remains

stable, suggesting that research on water economics may be increasingly being published in specialised journals.

Figure 10 presents topics' specificity and generality. Mean weight when a topic was selected (Y axis) showed the strength of topic domination in articles where it was a dominant topic. Mean weight of a topic when it was not selected (X axis) indicated prevalence of a topic in articles where it was not the dominant topic, suggesting that a topic was more general. More specific topics concentrated around the top left corner, while more general topics concentrated around the bottom right corner of Figure 10. Based on Figure 10, it could be suggested that 'Environmental tax', 'Global climate policy' and 'Trade and investment' are more general topic as they co-occur in the articles with other dominant topics. For instance, Climate change policy could have co-occurred with environmental and resource economics articles since the first publication of assessment reports by the Intergovernmental Panel on Climate Change in 1990. At the same time, 'Fishery', 'Waste' and 'Nonmarket valuation' were more specific topics as they strongly dominated in articles where there were dominant topics. Both 'Fishery' and 'Nonmarket valuation' were "hot" topics in terms of popularity and citations (Figure 9). These results suggest that there might be a positive relationship between popularity and specificity of a topic, as a topic becomes more popular (such as non-market valuation) when more papers focusing on that particular topic get published. On the other hand, a less popular topic may still get researched and published but it does not dominate the entire focus of the article.

Figure 10. Topics specificity and generality

Results of the analysis of co-occurrence of specific topics pairs in the body of articles published in ERE are presented as cross-tabulation in Figure 11. The numbers in grids show extent of co-occurrence ($O_{i,j}$). The shading of the grids vary from dark grey for pairs of topics with high co-occurrence to light grey for pairs of topics with low co-occurrence. The figure shows high scores for pairs of topic where co-occurrence is expected, for example, ‘Environmental tax’ and ‘Global climate policy’, ‘Recreation and amenity’ and ‘Nonmarket valuation’, and ‘Tradable permits’ and ‘Environmental tax’. On the other hand, low co-occurrence may indicate research gaps between topics suggesting scope for future research. This maybe especially true for the topics that are gaining popularity that we labelled as “hot”. For example, ‘Voluntary environmentalism’ was one of the hot topics which had not been extensively studied in relation to ‘Fishery’ and, to a lesser extent, ‘Water’. Another reason for low co-occurrence maybe that studies of particular combinations of topics are published by specialised outlets. For example, very few articles published in ERE study ‘Forestry’ in

conjunction with 'Energy'. There is, however, a journal dedicated to these two topics, *Biomass and Bioenergy*. Whilst Figure 11 also indicated a gap between 'Fishery' and 'Tradable permits'. However, 'Fishery' does include studies of tradable fishing quotas and the topic 'Tradable permits' covered trading of pollution permits. The low co-occurrence between 'Tradable permits' and 'Recreation and amenity' indicated another possible research gap. As expected, 'Nonmarket valuation', another hot topic, had co-occurred with most of the other topics, except 'Tradable permits' and 'Trade and investment'. However, the later topics implied that there was a market for permits or investments, therefore no need to use nonmarket valuation methods to study them. The research gaps identified using analysis of topics co-occurrence in the body of articles published by ERE offer novel avenues for research on unique combination of ideas to understand the linkages between these topics in further details.

Figure 11. Cooccurrence of topics in articles

4 Conclusion

Analysis of ERE over the past 25 years showed that it has attracted articles on a wide range of topics in environmental economics, covering theoretical to applied and review articles written by a diverse group of authors. Our analysis revealed some interesting patterns in terms of contributors and content. As expected, European authors constituted more than half of the contributors, which contrasted some other top environmental and resource economics journals. Most of the collaborations were between European authors, which is consistent with the objective of the European Association of Environmental and Resource Economists (EAERE) that published the journal. The next largest concentration of collaboration occurred among North American authors. However, there was an increasing interest in collaborative works as evidenced by the gradual increase in the proportion of multi-authored papers, in international collaboration (increased from 10% to 40% in the last 25 years); and collaboration between male and female authors. In our opinion, ERE is currently providing a platform for bridging the distance between European and international environmental economics studies.

To analyse the content of the papers, we have used topic modelling technique, which has been widely applied in many scientific disciplines but are relatively new to the field of economics. Despite various pitfalls of automated text analysis tools, careful application can still provide useful information that is complementary to that given by systematic reviews and meta-analyses, the two most common analytical tools for research synthesis in the field of economics.

Topic analysis revealed that all the articles published in the journal could be classified as applied and policy orientated research. ‘Nonmarket valuation’, ‘Recreation and amenity’, and ‘Conservation’ are three highly published and cited topics. Conservation became an

important lexicon of research articles when the Convention on Biological Diversity, an outcome of Earth Summit, was enacted in 1993. Similarly, it is likely that nonmarket valuation became more popular since the release of NOAA Blue Ribbon Panel report in 1993 on Exxon Valdez oil spill in 1989. The popularity of nonmarket valuation papers, as demonstrated by the growth in the number of publications, as well as the growth in the number of citations, showed that the ERE is filling in the gap where more applied and policy relevant papers have a place to be published. Publication of special issues on applied and policy-oriented topics is also an indication of the journal's role in contributing to the wider national and international natural resource management and environmental policies.

Findings from this study provided insights into areas of growing research interest, funding opportunities, and potential collaborators from other institutions. Our analysis revealed that there are potential research gaps, with possible opportunities for novel contribution by combining topics such as: 'Voluntary environmentalism' with 'Fishery' and 'Water', and 'Forestry' with 'Environment and health'. In conclusion, it was evident that ERE is an avenue for researchers in environmental economics, both theoretical and applied, to share their knowledge and advance the research in this field.

5 Acknowledgement

This research was conducted with the support of funding from the Australian Research Council Centre of Excellence for Environmental Decisions. We acknowledge helpful comments and suggestions of Ms Tammie Harold and Mr Tas Thamo of UWA.

6 References

- Adams J (2013) Collaborations: The fourth age of research. *Nature* 497(7451): 557-560
- Backhouse RE (1998) The transformation of US economics, 1920-1960, viewed through a survey of journal articles. *History of Political Economy*: 85
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3(4-5): 993-1022
- Brookshire DS, Scroggin DO (2000) Reflections upon 25 years of the *Journal of Environmental Economics and Management*. *Journal of Environmental Economics and Management* 39(3): 249-263
- Brundtland G, Khalid M, Agnelli S, Al-Athel S, Chidzero B, Fadika L, Hauff V, Lang I, Shijun M, de Botero MM (1987) Our common future ('brundtland report'). In Carson RT, Flores NE, Meade NF (2001) Contingent valuation: Controversies and evidence. *Environmental and Resource Economics* 19(2): 173-210
- Costanza R, Howarth RB, Kubiszewski I, Liu S, Ma C, Plumecocq G, Stern DI (2016) Influential publications in ecological economics revisited. *Ecological Economics* 123: 68-76
- Costanza R, Stern D, Fisher B, He L, Ma C (2004) Influential publications in ecological economics: A citation analysis. *Ecological Economics* 50(3-4): 261-292
- Debertin DL, Pagoulatos A (1992) Research in agricultural economics 1919–1990: Seventy-two years of change. *Rev. Agric. Econ.* 14(1): 1-22
- Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. *Journal of the American society for information science* 41(6): 391
- Endres DM, Schindelin JE (2003) A new metric for probability distributions. *IEEE Transactions on Information Theory* 49(7): 1858-1860
- Feinerer I (2015) Introduction to the tm Package Text Mining in R. In Furnas GW, Deerwester S, Dumais ST, Landauer TK, Harshman RA, Streeter LA, Lochbaum KE (1988) Information retrieval using a singular value decomposition model of latent semantic structure. In *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval*: ACM, 465-480
- Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proceedings of the National academy of Sciences* 101(suppl 1): 5228-5235
- Grimmer J, Stewart BM (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3): 267-297
- Grün B, Hornik K (2011) Topicmodels: An r package for fitting topic models. *Journal of Statistical Software* 40(13): 1-30
- Hall D, Jurafsky D, Manning CD (2008) Studying the history of ideas using topic models. In *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL*, 363-371
- Hamermesh DS (2013) Six decades of top economics publishing: Who and how? *Journal of Economic Literature* 51(1): 162-172
- Hanley N, Wright RE, Adamowicz V (1998) Using choice experiments to value the environment: Design issues, current experience and future prospects. *Environmental and Resource Economics* 11(3-4): 413-428
- Karl A, Wisnowski J, Rushing WH (2015) A practical guide to text mining with topic extraction. *Wiley Interdisciplinary Reviews: Computational Statistics* 7(5): 326-340
- Kates RW (1987) The Human Environment: The Road Not Taken, The Road Still Beckoning. *Annals of the Association of American Geographers* 77(4): 525-534

- Kullback S, Leibler RA (1951) On information and sufficiency. *The annals of mathematical statistics* 22(1): 79-86
- Laband DN, Tollison RD (2000) Intellectual collaboration. *Journal of Political Economy* 108(3): 632-662
- Ma C, Stern DI (2006) Environmental and ecological economics: A citation analysis. *Ecological Economics* 58(3): 491-506
- Murtagh F, Legendre P (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification* 31(3): 274-295
- Opschoor JB, Jansen HMA (1991) The raison d'être of ERE. *Environmental and Resource Economics* 1(1): v-viii
- Paul MJ, Dredze M (2014) Discovering health topics in social media using topic models. *PLoS ONE* 9(8)
- Polyakov M, Gibson FL, Pannell DJ (2016) Six decades of agricultural and resource economics in Australia: an analysis of trends in topics, authorship and collaboration. *Australian Journal of Agricultural and Resource Economics*: n/a-n/a
- R Core Team (2015) R: A language and environment for statistical computing. In, Vienna, Austria: R Foundation for Statistical Computing
- Rusch T, Hofmarcher P, Hatzinger R, Hornik K (2013) Model trees with topic model preprocessing: An approach for data journalism illustrated with the Wikileaks Afghanistan war logs. *Annals of Applied Statistics* 7(2): 613-639
- Schymura M, Löschel A (2014) Incidence and extent of co-authorship in environmental and resource economics: Evidence from the *Journal of Environmental Economics and Management*. *Scientometrics* 99(3): 631-661
- Shogren JF, Durden GC (1991) The first 15 years: Contributors and contributions to the *Journal of Environmental Economics and Management*, 1974-1988. *Journal of Environmental Economics and Management* 20(3): 205-209
- Sievert C, Shirley KE (2014) LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, Baltimore, Maryland, USA, 63-70
- Westgate MJ, Barton PS, Pierson JC, Lindenmayer DB (2015) Text analysis tools for identification of emerging topics and research gaps in conservation science. *Conservation Biology* 29(6): 1606-1614

Appendix: additional figures

Figure A1. Top 50 most cited ERE authors

Figure A2. Top 30 most cited ERE articles

Figure A3. Trends in ERE topics: 1991-2015