AI in energy: Advancing new research paradigms
A bibliometric narrative on AI / ML / DL in the economics of energy

David C. Broadstock

Presented at: IAEE AI Webinar, 2020 – Online.
Slides last updated: August 17, 2020
I. AI in economics is a rapidly emerging topic - with legitimate potential for high durability
I don’t want to sound repetitive, or to oversimplify, but...

\[ AI \neq ML \neq DL \]

They diverge in focus:

- **Artificial intelligence:** The term ‘artificial intelligence’ was coined by the authors of a proposal for a seminar to study the subject, to be held at Dartmouth College, USA, in 1956. Though Alan Turing published in 1950 on ‘Computing Machinery and Intelligence’

- **Machine learning:** With the Samuel-Checkers playing Program (1952), Arthur Samuel developed the very first self-learning computer program while at IBM. Tom Mitchell, in 1998 offered a more formal definition: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”

- **Deep learning:** Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits or letters or faces.

But they are intricately connected:

\[ \{ DL \in ML \in AI \} \in Deep\ Analytics \]
Outline for today’s talk

In today’s talk I hope to establish the foundations of a story that will help you position my thinking, and possibly provide a new trajectory to your own.

I. Bibliometrics/scientometrics (and meta-research/analysis)
Bibliometric reflection on growth in the literature, and are there gaps to address?

II. Bibliometrics is Used widely
In the life sciences, (medicine), library related research and archiving.

III. These tools are far less familiar to economists
But offer powerful complementary insights on underlying narratives, & help isolate literature gaps/strengths

IV. How are they useful?
This varies according to the ‘reader’ - junior and senior scholars will focus on different insights.
The ‘descriptive statistics’.

<table>
<thead>
<tr>
<th>Measure</th>
<th>All</th>
<th>Artificial Intelligence</th>
<th>Machine Learning</th>
<th>Deep Learning</th>
<th>Energy Economics</th>
<th>Environmental Economics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>1,276</td>
<td>664</td>
<td>625</td>
<td>91</td>
<td>56</td>
<td>117</td>
</tr>
<tr>
<td>Sources (Journals, Books, etc)</td>
<td>419</td>
<td>267</td>
<td>279</td>
<td>49</td>
<td>27</td>
<td>69</td>
</tr>
<tr>
<td>Average citations per document</td>
<td>9.187</td>
<td>11.61</td>
<td>6.496</td>
<td>4.264</td>
<td>10.82</td>
<td>9.009</td>
</tr>
<tr>
<td>Authors</td>
<td>2,999</td>
<td>1,563</td>
<td>1,513</td>
<td>250</td>
<td>182</td>
<td>344</td>
</tr>
<tr>
<td>Author Appearances</td>
<td>3,372</td>
<td>1,690</td>
<td>1,684</td>
<td>265</td>
<td>186</td>
<td>349</td>
</tr>
<tr>
<td>Authors of single-authored</td>
<td>261</td>
<td>169</td>
<td>104</td>
<td>17</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Authors of multi-authored</td>
<td>2,738</td>
<td>1,394</td>
<td>1,409</td>
<td>233</td>
<td>176</td>
<td>321</td>
</tr>
<tr>
<td>Documents per Author</td>
<td>0.425</td>
<td>0.425</td>
<td>0.413</td>
<td>0.364</td>
<td>0.308</td>
<td>0.34</td>
</tr>
</tbody>
</table>

The truth is relevant academic literature in energy economics remains quite thin.

- Academic (bibliographic) meta-data are obtained from Scopus, and processed with R
- I opted to expand my horizon to AI / ML / DL in economics as a broad discipline (1,276 papers)
- To put this in context, EJ, EE, EP, REE, ERE and JEEM between them have circa 25k articles
- One should heed caution in the coverage of the data* (e.g. EJ) & read “Machine learning in energy economics and finance: A review” by Ghoddusi, Creamer and Rafizadeh (2019).
How is the scale & rate of change of scientific production? (I/III)

Were I to glance at the literature up until 2017...

Less than 50 papers per year, but both econometrics and ‘eye’-conometrics points towards the emergence of the growth phase of the ‘topic diffusion curve’.

However, it is less than clear this is enough to be excited about either as a scholar or an editor...

Literature covers all papers until the end of 2017
How is the scale & rate of change of scientific production? (II/III)

Moving forward a year

Momentum is sustained and the rate of growth in the literature exceeds 2017 predictions.

But still only between 100-300 papers per year up until 2025: And remember, this is for economics as a whole!

Literature covers all papers until the end of 2018
How is the scale & rate of change of scientific production? (III/III)

By the end of 2019

▶ The landscape has fundamentally shifted with AI / ML / DL visibly entering the mainstream
▶ Forecasts are quite literally off the scale

And 2020 is on course for another ‘strong’ year: With early indications suggesting the trend will continue.
I want to leave you with three thoughts... (II/III)

I. AI in economics is a rapidly emerging topic - with legitimate potential for high durability

II. I do not believe applications are absent from practitioners work - but academic coverage in energy economics is disproportionately thin.
Energy related studies in economics covering AI / ML / DL

The Venn diagram to the left shows the prevalence of energy research and also the relative share of AI / ML / DL

- AI and ML take a roughly equal share in the literature.
- This however masks the relative trajectories of AI and ML
- Indications are that within economics ML will overtake AI, within just a few years.

This is highly intuitive: ML can embrace concepts of Bayesian/evolutionary econometrics in a direct and intuitive fashion.
A slightly ‘richer’ Venn diagram

This graph points towards the uniqueness of topic coverage between AI / ML

▶ Little more than 10% of literature (keywords) bridge AI with ML. Of course there is a connection, but there may be a dichotomy in use of keywords

▶ The energy focused literature contains around half the volume of research compared with environmental economics

▶ There are many zeros, some of which highlight useful gaps e.g. $DL \cap Energy = \emptyset$

**DL literature is sparse**: My assumption is that practical relevance is high, but ‘academic translation’ is complex against the backdrop of conventional publishing ‘norms’.
I want to leave you with three thoughts... (III/III)

I. AI in economics is a rapidly emerging topic - with legitimate potential for high durability

II. I do not believe applications are absent from practitioners work - but academic coverage in energy economics is disproportionately thin.

III. AI / ML / DL applications are yet to adequately target the core topics in energy economics
## (LDA) Topic classification of ‘analytical energy economics’

<table>
<thead>
<tr>
<th>Topic number:</th>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic name:</td>
<td>Oil shocks</td>
<td>Technology/innovation</td>
<td>Renewables</td>
<td>Electricity</td>
<td>Energy</td>
<td>Climate</td>
<td>Agricultural</td>
<td>Fuel mix</td>
<td>Oil shocks</td>
<td>Productivity</td>
</tr>
<tr>
<td>Keyword 1:</td>
<td>oil</td>
<td>environmental</td>
<td>wind</td>
<td>electricity</td>
<td>forecasting</td>
<td>fossil</td>
<td>emissions</td>
<td>climate</td>
<td>water</td>
<td>oil</td>
</tr>
<tr>
<td>Keyword 2:</td>
<td>shocks</td>
<td>firms</td>
<td>renewable</td>
<td>forecasting</td>
<td>gas</td>
<td>scenario</td>
<td>respondents</td>
<td>percent</td>
<td>returns</td>
<td>output</td>
</tr>
<tr>
<td>Keyword 3:</td>
<td>stock</td>
<td>technological</td>
<td>electricity</td>
<td>regime</td>
<td>carbon</td>
<td>carbon</td>
<td>respondents</td>
<td>percent</td>
<td>returns</td>
<td>output</td>
</tr>
<tr>
<td>Keyword 4:</td>
<td>crude</td>
<td>technologies</td>
<td>solar</td>
<td>forecast</td>
<td>gas</td>
<td>scenario</td>
<td>land</td>
<td>tax</td>
<td>crude</td>
<td></td>
</tr>
<tr>
<td>Keyword 5:</td>
<td>shock</td>
<td>abatement</td>
<td>generation</td>
<td>forecasts</td>
<td>oil</td>
<td>scenarios</td>
<td>attributes</td>
<td>coal</td>
<td>futures</td>
<td>outputs</td>
</tr>
<tr>
<td>Keyword 6:</td>
<td>returns</td>
<td>innovation</td>
<td>capacity</td>
<td>fig</td>
<td>intensity</td>
<td>mitigation</td>
<td>uranium</td>
<td>utility</td>
<td>stock</td>
<td>productivity</td>
</tr>
<tr>
<td>Keyword 7:</td>
<td>test</td>
<td>climate</td>
<td>investment</td>
<td>spot</td>
<td>electricity</td>
<td>emission</td>
<td>biomass</td>
<td>fuel</td>
<td>conditional</td>
<td>inputs</td>
</tr>
<tr>
<td>Keyword 8:</td>
<td>kilian</td>
<td>emissions</td>
<td>firms</td>
<td>day</td>
<td>manufacturing</td>
<td>technologies</td>
<td>environmental</td>
<td>electricity</td>
<td>commodity</td>
<td>inefficiency</td>
</tr>
<tr>
<td>Keyword 9:</td>
<td>countries</td>
<td>learning</td>
<td>renewables</td>
<td>spikes</td>
<td>countries</td>
<td>gas</td>
<td>crop</td>
<td>capital</td>
<td>risk</td>
<td>sueyoshi</td>
</tr>
<tr>
<td>Keyword 10:</td>
<td>fig</td>
<td>investment</td>
<td>mwh</td>
<td>daily</td>
<td>rfs</td>
<td>baseline</td>
<td>renewable</td>
<td>figure</td>
<td>garch</td>
<td>plants</td>
</tr>
</tbody>
</table>

### Frequency distribution of topics:

<table>
<thead>
<tr>
<th></th>
<th>En. Econ.</th>
<th>AI</th>
<th>ML</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>En. Econ.</td>
<td>6.39</td>
<td>13.41</td>
<td>1.84</td>
<td>2.3</td>
</tr>
<tr>
<td>AI</td>
<td>6.23</td>
<td>0.51</td>
<td>0.84</td>
<td>3.45</td>
</tr>
<tr>
<td>ML</td>
<td>6.15</td>
<td>0.17</td>
<td>3.85</td>
<td>24.14</td>
</tr>
<tr>
<td>DL</td>
<td>5.21</td>
<td>1.53</td>
<td>2.84</td>
<td>2.3</td>
</tr>
</tbody>
</table>

### Cumulative Frequency distribution of topics:

<table>
<thead>
<tr>
<th></th>
<th>En. Econ.</th>
<th>AI</th>
<th>ML</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>En. Econ.</td>
<td>6.39</td>
<td>13.41</td>
<td>14.43</td>
<td>26.37</td>
</tr>
<tr>
<td>AI</td>
<td>12.62</td>
<td>13.92</td>
<td>14.6</td>
<td>16.64</td>
</tr>
<tr>
<td>ML</td>
<td>18.85</td>
<td>14.43</td>
<td>15.11</td>
<td>15.23</td>
</tr>
<tr>
<td>DL</td>
<td>25.00</td>
<td>15.11</td>
<td>15.4</td>
<td>15.77</td>
</tr>
</tbody>
</table>

---

Slide 13 of 18.
So what are the top ‘connected’ topics in Artificial Intelligence

AI embraces many direct connections:
These are across a range of topics, which are seemingly somewhere between management/operations research and mainstream economics.
There are many nebulous concepts/constructs shown. Arguably, this presents a literature still in a stage of theorization, and not distilled.
So what are the top ‘connected’ topics in Machine Learning

Machine learning is visibly more technical in nature
Only four other keywords are connected (*frequently). While there is also an indirect connection to DL.

The indirect connections to various different topics imply the potential to ‘bridge a connection’.
For deep learning, the connections are sparse

This of course follows from the sparse DL literature, but also indicates a lot of room for potential growth.

Given how scant the literature is here, it is not worth over-analyzing the patterns.
Closing remarks

“In the field of AI expectations seem to always outpace the reality.”

Ting Huang, Brian McGuire, Chris Smith*, Gary Yang, (December, 2006) “The History of Artificial Intelligence”

Some practical realities preserve the salience of this statement:

- **Access to data is a constraint:** My peers may touch more on this, but it is not just that we do not know where to find the data, it is also the cost of access and the ability to store and process suitably complex data.

- **Access to computing power is a hurdle:** The econometric ‘game’ is one of combinatorial statistics in the face of too many permutations to handle. Paraphrasing this challenge to my 10 year old son, it is akin to having certainty in what not to analyze, never having looked at it in the first place.

- **We need more benchmarking against traditional/conventional methods:** Arguably too many applications focus on what AI / ML / DL can achieve, without being visibly cognizant that the same could already be achieved at ‘low cost’ with conventional analytics. Academics are often asking to be shown what AI / ML / DL can do that standard econometrics tools cannot.
Thanks for listening!

Any questions/comments are warmly welcomed.

david.broadstock@polyu.edu.hk

Let me also give a quick plug to the econometric society world congress
AI and ML in Empirical Research - Live Policy Session (tomorrow!)

Victor Chernozhukov, Esther Duflo, Guido Imbens, Vasilis Syrgkanis