

## Nederlandstalige samenvatting CPB Discussion Paper 440

### **'Temporal Patterns in Economics Research'**

#### **'Temporele patronen in economisch onderzoek'**

Steeds meer tekstuele data zijn digitaal beschikbaar en daarmee vormen ze een rijke bron voor data-analisten; zowel op macro- als op microniveau. Denk bij data op macroniveau bijvoorbeeld aan financieel nieuws, en op microniveau aan vacatureteksten voor specifieke banen. Uit recente economische literatuur blijkt dat tekstdata informatie kunnen leveren die de nauwkeurigheid van voorspelmodellen verbetert. Het gebruik van tekst als data heeft dus veel potentieel voor economisch onderzoek.

Het verwerken en gebruik van tekstdata vergen echter nieuwe expertise. Het CPB heeft tot dusver weinig onderzoek op dit gebied uitgevoerd. Om deze expertise op te bouwen heeft het Data Science Team een pilot-onderwerp gekozen dat voor economen bekend terrein is: de tekst van korte samenvattingen (zogenoemde abstracts) van economische academische publicaties. Dit project is uitgevoerd om ervaring op te doen met methoden voor tekstverwerking, die we later ook op andere data kunnen toepassen, zoals vacatureteksten.

Een van de meest voor de hand liggende vragen voor tekstanalyse is: Lijken deze twee stukken tekst op elkaar?. Deze vraag staat bijvoorbeeld centraal in zoekmachines: die vinden alle webpagina's met teksten die relevant zijn voor je zoekopdracht. Er zijn ook economische vraagstukken die je hiermee kunt onderzoeken. Denk daarbij aan vacatureteksten uit twee verschillende sectoren, vragen die misschien om vergelijkbare vaardigheden? Als dat zo is, kunnen werknemers makkelijk tussen die sectoren van baan verwisselen.

Ook in ons onderzoek staat deze vraag centraal. We maken verbindingen tussen teksten van economische onderzoeken uit verschillende onderzoeksvelden en over diverse jaren. Hiervoor gebruiken we de abstracts van economische papers uit de RePEc-dataset. Soortgelijke teksten clusteren we met elkaar. Zo kunnen we vaststellen over welke tijdsspanne economische onderwerpen worden beschreven.

Voor de meerderheid van de papers vinden we dat vergelijkbare papers in hetzelfde jaar gepubliceerd worden. Dit suggereert dat er sterke gemeenschappelijke trends zijn in de economische literatuur: veel papers schrijven over hetzelfde onderwerp op hetzelfde moment. Er zijn echter ook vergelijkbare papers die wel 14 jaar van elkaar zijn gepubliceerd. Er bestaan dus ook economische onderwerpen die langdurig worden beschreven.

Ten slotte zien we dat het gemiddelde verschil in publicatiejaar is afgenomen van ongeveer 4 jaar tussen 1990 en 2005 tot ongeveer 1 jaar begin 2010. Dit wekt de suggestie dat economisch onderzoek op hetzelfde onderwerp zich meer concentreert in een korte periode.