

Nederlandstalige samenvatting CPB Discussion Paper 444

‘Predicting Firm Exits with Machine Learning: Implications for Selection into COVID-19 Support and Productivity Growth’

‘Effecten van coronasteun op bedrijvendynamiek en productiviteitsgroei: een nieuwe analyse met machine learning’

Evaluaties van steunmaatregelen voor bedrijven vereisen vaak een goede inschatting van de levensvatbaarheid van bedrijven of van de kans dat een bedrijf de markt verlaat. Dit bleek bijvoorbeeld tijdens de coronapandemie. Op 17 maart 2020 werd een *lockdown* en bijbehorende contactbeperkende maatregelen aangekondigd, waardoor sommige sectoren op slot gingen. Om bedrijven en zelfstandigen te compenseren voor het verlies aan inkomsten, werd een omvangrijk pakket steunmaatregelen ontworpen. Deze steunmaatregelen kenden amper restricties, omdat de steun snel uitgekeerd moest worden. Dit roept de vraag op of ongezonde bedrijven disproportioneel veel gebruikgemaakt hebben van steun en in hoeverre deze steunmaatregelen niet-levensvatbare bedrijven overeind gehouden hebben. Dit is een belangrijke economische vraag, omdat het steunen van niet-levensvatbare bedrijven negatieve gevolgen heeft voor productiviteitsgroei. Dit gebeurt bijvoorbeeld wanneer steun de herallocatie van kapitaal en arbeid van niet-levensvatbare bedrijven naar levensvatbare bedrijven belemmert.

De coronasteunmaatregelen waren effectief, maar hebben ook onbedoelde bijeffecten gehad. In het algemeen is het aannemelijk dat de coronasteunmaatregelen succesvol waren en ervoor hebben gezorgd dat minder bedrijven de markt verlaten hebben. CPB-onderzoekers concludeerden eerder dat bedrijven zonder coronasteun vaker de markt verlieten dan bedrijven met coronasteun ([link](#)). Deze auteurs laten ook zien dat dit effect groter is voor laagproductieve bedrijven dan voor hoogproductieve bedrijven. Dit impliceert dat coronasteunmaatregelen de herallocatie van laagproductieve bedrijven naar hoogproductieve bedrijven kan hebben belemmerd ([link](#)).

Eerdere analyses van de coronasteunmaatregelen in Nederland corrigeerden niet voor selectie-effecten of baten van steun. Stel dat niet-levensvatbare bedrijven vaker steun aanvragen dan levensvatbare. De resultaten beschreven in de eerder genoemde publicaties *onderschatten* dan het effect van de coronasteunmaatregelen op de kans om de markt te verlaten en op de doorwerking op de productiviteitsgroei. Als juist levensvatbare bedrijven vaker steun aanvragen, *overschatten* de auteurs de effecten van coronasteunmaatregelen op de bedrijvendynamiek en op de groei van de productiviteit.

We bepalen de levensvatbaarheid van bedrijven met *machine learning* en *big data* over hun financiële positie. Dit doen we door de kans te schatten dat een bedrijf in het volgende jaar de markt verlaat. Deze schattingen gebruiken we om een *counterfactual* index te maken van de levensvatbaarheid van bedrijven, waarbij de *counterfactual* de levensvatbaarheid van een bedrijf aangeeft in een wereld zonder corona.

We laten zien dat het *machine learning*-model beter in staat is om de kans te voorspellen dat een bedrijf de markt verlaat dan conventionele methoden. Meer dan de helft van de bedrijven binnen de top-1procent bedrijven die volgens de *machine learning*-voorspellingen het hoogste risico hebben om de markt te verlaten, doet dit ook daadwerkelijk. Bij de logistische regressie is dit minder dan een derde.

Daarnaast laten we zien dat het gebruik van de (betere) *machine learning*-voorspellingen leidt tot andere conclusies van economische analyses dan het gebruik van de conventionele voorspellingen. In het bijzonder laten we zien dat de aanwijzingen dat gezonde bedrijven vaker en meer gebruik maken van coronasteunmaatregelen robuuster zijn als dit bestudeerd wordt op basis van de *machine learning*-schattingen. Ook gebruiken we de voorspellingen om de markt te verlaten om te corrigeren voor deze vorm van selectie bij het vergelijken van het gemiddeld percentage bedrijven dat verdwijnt naar productiviteitsniveau en gebruik van steun. Hier vinden we dat correctie volgens de conventionele methode een conclusie oplevert die meer lijkt op die van de analyse van de ruwe data (zonder correctie), maar dat het beeld sterk verandert als *machine learning*-methoden gebruikt worden. Dan zien we nog steeds dat laagproductieve bedrijven meer profijt gehad hebben van de coronasteunmaatregelen dan hoogproductieve bedrijven (waar profijt gemeten is in de vorm van voorkomen *exits*), maar dat de mate waarin dit gebeurt ongeveer de helft kleiner is.

Het onderzoek is gebaseerd op gegevens van alle niet-financiële ondernemingen (bv's en nv's) tussen 2013 en 2020. Voor deze bedrijven observeren we karakteristieken zoals leeftijd, sector en omvang, maar ook posten uit de balans en de winst- en verliesrekening en financiële kengetallen (zoals de solvabiliteit, etc.) die hierop gebaseerd zijn. Deze karakteristieken zijn gemeten op jaarniveau voor de ondernemingsgroep. Bij het *machine learning*-model nemen we ook vertragingen van de financiële variabelen op. We bepalen of een bedrijf het jaar erop de markt verlaat (een 'exit'), waar we bedrijven die overgenomen worden of samengaan met een ander bedrijf of die geherstructureerd worden niet meenemen.

We vergelijken *out-of-sample*-voorspellingen van *machine learning* en conventionele modellen. We selecteren tien procent van de ondernemingen als trainingsdata, waarbij we observaties in 2019 en 2020 verwijderen. Deze data gebruiken we om een verband te schatten tussen de kenmerken van bedrijven en de kans om het jaar erop de markt te verlaten. Omdat observaties in 2019 en 2020 niet meegenomen zijn, vormen we zo *counterfactual* voorspellingen van de verwachte kans om de markt te verlaten in een wereld zonder corona. In het bijzonder vergelijken we een logistische regressie conform de Z-score specificatie van Altman met een *gradient boosting regression trees* model die de hierboven beschreven kenmerken van bedrijven meeneemt (het zijn er meer dan 200). Het model van Altman gebruikt vier financiële ratio's en indicatoren voor leeftijd, omvang en sector om de kans te schatten dat een bedrijf de markt verlaat en is één van de populairste methoden om *financial distress* bij bedrijven te meten binnen de *business*- en *financeliteratuur*. Om goed te vergelijken hoe verschillen in voorspelkwaliteit samenhangen met methode (logistische regressie of *gradient boosting regression tree*) en gebruikte data (Z-score of alle data) schatten we ook een logistische regressie met alle data en een *gradient boosting regression trees* met variabelen van de Z-score-specificatie. We analyseren de vier combinaties van voorspelmodel en gebruikte data om *out-of-sample* voorspellingen te maken voor de ongeveer 90 procent van de observaties die niet in de *train data* ingedeeld zijn.

We vinden dat een *machine learning*-model met big data het best in staat is om bedrijven die de markt verlaten te onderscheiden van bedrijven die dat niet doen. De statistische maatstaf AUC (de *Area Under the ROC-Curve*) is 84 procent, terwijl de AUC van een logistische regressie op basis van de Z-score 70 procent is. Deze hogere AUC komt zowel door het gebruik van *machine learning* als het gebruik van meer data: als ons *machine learning*-model alleen de Z-score variabelen mag gebruiken om te voorspellen of bedrijven verdwijnen, is de AUC 76 procent, vergelijkbaar met de AUC van een logistische regressie die gebruikmaakt van de big data (AUC = 75 procent).

Ook laten we zien dat *machine learning*-voorspellingen een structuur vinden die economisch logisch is. In het paper delen we de voorspellingen van *machine learning* (op basis van big data) en de logistische regressie (op basis van de Z-score specificatie) op in honderd even grote groepen naar grootte van de geschatte kans om de markt te verlaten. Vervolgens berekenen we per groep (percentiel) de gemiddelde waarde van financieel-economische kenmerken van bedrijven, zoals de schuldquote, winstgevendheid, omvang van de balans, etc. Hieruit blijkt dat het *machine learning*-model een vergelijkbare economische structuur vindt als de conventionele logistische regressie.

Gebruik van coronasteunmaatregelen neemt af met de kans om de markt te verlaten, en dit effect is robuuster wanneer de *machine learning* -voorspellingen gebruikt worden. In het paper vergelijken we het gemiddeld gebruik van minimaal een coronasteunmaatregel per percentiel van de voorspellingen voor het *machine learning*-model of de logistische regressie. Dit doen we zowel zonder als met correctie voor karakteristieken van het bedrijf. Wat opvalt is dat de relatie tussen gebruik van steun en kans om de markt te verlaten negatief blijft voor de *machine learning*-schattingen, terwijl de relatie grotendeels plat wordt bij de logistische regressie. De resultaten op basis van *machine learning* zijn dus robuuster dan die van de logistische regressie.

Het resultaat uit de economische analyse van coronasteunpakketten ([link](#)) dat bedrijven met steun de markt minder vaak verlieten dan bedrijven zonder steun wordt gedreven door ‘zwakke bedrijven’, bedrijven met een hoge kans om de markt te verlaten. Ons paper is een verfijning van het eerdere CPB-onderzoek over de coronasteun met meer data en *machine learning*. In het paper berekenen we het verschil in het percentage bedrijven dat in 2020 de markt verlaat ten opzichte van 2019 (de *exit rate differential*). Dit doen we apart voor bedrijven die minimaal 1 coronasteunmaatregel gebruikt hebben en bedrijven die geen enkele maatregel gebruikt hebben. Wanneer we kijken naar bedrijven die géén gebruikmaakten van de coronasteunmaatregelen, blijkt dat het percentage bedrijven dat de markt verliet in 2020 enigszins hoger lag dan in 2019. Echter, wanneer we kijken naar bedrijven die wel gebruikmaakten van deze maatregelen, dan zien we dat in 2020 fors minder bedrijven de markt verlieten én dat dit effect zeer sterk gedreven is door bedrijven met een hoge kans om de markt te verlaten.

Deze resultaten werken door op de invloed van de coronasteunmaatregelen op de groei van de productiviteit. Wanneer we onze *machine learning*-schattingen gebruiken om te corrigeren voor selectiviteit vinden we dat de coronasteunmaatregelen het aantal bedrijven dat uittreedt heeft verlaagd met 16 procent. Dit percentage halveert ongeveer als we het aantal voorkomen uittrekkingen wegen naar productiviteit, wat aangeeft dat laagproductieve bedrijven meer geprofiteerd hebben van de coronasteunmaatregelen dan hoogproductieve bedrijven. Dit geeft aanleiding om te veronderstellen dat de coronasteunmaatregelen de herallocatie van arbeid en kapitaal tussen deze typen bedrijven verstoord heeft.