



# Economische effecten schatten voor individuen bij difference-in-differences

We passen het causal forest-algoritme aan, zodat het gebruikt kan worden bij difference-in-differences. We noemen onze aanpassing causal forest with fixed effects, omdat het mogelijk is om te corrigeren voor effecten die constant zijn binnen observaties van een individu of tijd. Onze aanpassing zorgt er ook voor dat er robuustere, individuele effecten van interventies geschat kunnen worden.

Ons algoritme laat zien dat payrolling het uurloon van een specifieke groep werknemers verlaagt met gemiddeld 3,7 procent. We vinden grotere verschillen tussen mensen dan uit een handmatige subgroepanalyse bleek.

# 1 Inleiding

**Het schatten van het effect van economisch beleid vormt een belangrijk onderdeel van economisch onderzoek.** Zo leren we wat werkt en wat niet werkt, zodat er kan worden bepaald welk beleid effectief is. Deze schattingen leveren doorgaans een gemiddeld effect op over het effect van bijvoorbeeld het eigen risico op zorggebruik. Maar dat is vaak nog onvoldoende informatie om beleid volledig te informeren. Als beleidsmaker wil je bijvoorbeeld ook weten of het effect op zorggebruik verandert door iemands gezondheid, inkomen of kenmerken zoals migratieachtergrond.

**Met een nieuwe techniek, *causal forests*, is het mogelijk om heel nauwkeurig effecten voor subgroepen te schatten.** Deze *causal forest* (Athey et al., 2019; Wager & Athey, 2018) kunnen effecten van interventies schatten tot op het niveau van het individu. Dit gebeurt door te kijken of effecten van een interventie wijzigen met de kenmerken van individuen.

**Een tekortkoming van deze nieuwe techniek is dat die ongeschikt is voor één van de meest gebruikte onderzoeksmethodes: *difference-in-differences*.** Dit is een methode om causale effecten vast te stellen, met andere woorden: om oorzakelijke verbanden te vinden. Deze methode is erg populair (Currie et al., 2020) en wordt ook vaak door het CPB toegepast op Nederlandse data (Goos et al., 2022; Remmerswaal et al., 2019; Roos et al., 2021). Hierbij wordt het verschil in uitkomsten van een zogeheten ‘behandelgroep’ voor en na een interventie (een beleidswijziging of andere gebeurtenis) vergeleken met die van een controlegroep die niet door de interventie geraakt werd.

**We passen in ons onderzoek het *causal forest*-algoritme daarom aan, zodat het wel gebruikt kan worden bij *difference-in-differences*.** We noemen onze aanpassing *causal forest with fixed effects* (CFFE), omdat het mogelijk is om te corrigeren voor effecten die constant zijn binnen observaties van een individu (*individual fixed effects*) of tijd (*time fixed effects*). Onze aanpassing zorgt er ook voor dat er robuustere, individuele effecten van interventies geschat kunnen worden. Dit kan ook als de onderzoeker geen gebruik maakt van *difference-in-differences*, omdat het -bij paneldata- mogelijk wordt om te corrigeren voor niet waargenomen individuele kenmerken die constant zijn.

**We vergelijken onze methode met twee andere methoden om *causal forests* te gebruiken bij *difference-in-differences* en laten zien dat onze aanpassing efficiënter is.** We vergelijken een CFFE met een zogenoemde naïeve methode, een *manually recentered causal forest* (MRCF), waarin we de uitkomst en de interventie eerst handmatig corrigeren voor individuele en tijdseffecten voordat een *causal forest* deze uitkomst en variabele gebruikt. Daarnaast vergelijken we onze methode met een *dynamic causal forest* (DCF), die zeer recent ontwikkeld is (Gavrilova et al., 2023). Maar waar onze methode -bij iedere uitsplitsing van een *causal tree*- gemiddelden van individuen en over tijd gebruikt om te corrigeren voor *fixed effects*, doet hun methode dit door het transformeren van de inputdata. De simulaties laten zien dat onze methode efficiënter is, mits aan bepaalde, vaak gemaakte veronderstellingen voldaan is.

**Tot slot gebruiken we onze methode in een praktijkvoorbeeld waarin we nagaan of het effect van *payrolling* op het uurloon verschilt als gekeken wordt naar kenmerken van werknemers.**

Een handmatige analyse laat zien dat *payrolling* het uurloon gemiddeld verlaagt met ongeveer 17 cent. Voor veel subgroepen schommelt deze daling tussen de 20 en 25 cent, maar de daling loopt op tot wel 32 cent per uur voor hoogopgeleiden en mensen met een tijdelijk contract. Wanneer we een CFFE gebruiken, blijkt dat voor alleen de twee decielen met de grootste daling in uurloon het uurloon daalt met gemiddeld ongeveer 40 cent per uur. Voor de overige decielen wordt geen effect gevonden. Dit verschil tussen de handmatige subgroepanalyse en onze analyse met een CFFE is dat bij de handmatige analyse groepen gemaakt worden op basis van een enkel kenmerk (een werknemer is een vrouw of een werknemer heeft een hoog opleidingsniveau), terwijl onze analyse kenmerken combineert (de werknemer is een hoogopgeleide vrouw, etc.). Dit zorgt ervoor dat gevonden effecten voor sommige subgroepen groter zijn en voor andere juist kleiner, wat een nauwkeuriger beeld geeft van de mate waarin individuele effecten van de interventie kunnen verschillen.

## 2 Gebruik van *causal forests* bij *difference-in-differences*

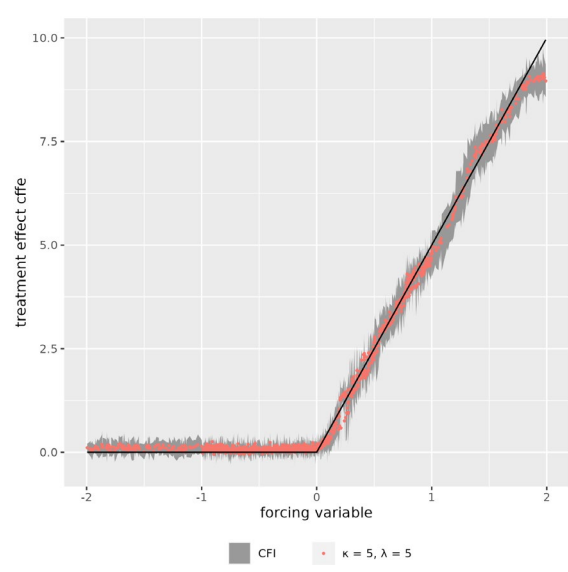
**Het idee achter een *causal forest* is betrekkelijk eenvoudig: om het effect van een interventie voor een bepaalde groep te meten, kijk je alléén naar deze groep.** Doorgaans maken onderzoekers zelf zulke groepen, bijvoorbeeld door de analyse apart te doen voor oude en jonge mensen, of voor mannen en vrouwen. Maar het maken van subgroepen kan ook op basis van de data. Dit heeft als voordeel dat ook kleine groepen gevonden worden waarvoor het effect van de interventie groot is, iets wat niet snel zou gebeuren als onderzoekers handmatig groepen maken omdat de groep zo klein is.

**Maar hoe maakt het algoritme dan groepen?** Een *causal forests* combineert meerdere *causal trees* (Athey & Imbens, 2016) om te bepalen welke mensen een groep vormen. *Causal trees* zijn ontwikkeld door Susan Athey en Guido Imbens, de Nederlandse winnaar van de Nobelprijs voor de economie in 2021. Een *causal tree* is een datagedreven beslisboom om causale effecten te schatten. Het algoritme verdeelt de data steeds verder in twee groepen, zodat het verschil in de effecten van de interventie tussen beide groepen zo groot mogelijk is. Op een gegeven moment stopt het algoritme, zodat observaties met anderen in één groep zitten op basis van hun verwachte effect van de interventie. Een *causal forest* combineert heel veel verschillende *causal trees*, standaard wel tweeduizend, die geschat worden op wisselende selecties binnen de data. Door veel *causal trees* te combineren, worden de geschatte effecten robuust en precies.

### 3 Simulatie

We onderzoeken de werking van de drie *causal forest*-modellen met behulp van simulaties. We richten ons hier op het standaardmodel voor *difference-in-differences* waarin personen in de behandelgroep allemaal op hetzelfde moment te maken krijgen met de interventie. Een moeilijkere opzet, waarin dit per individu verschilt, behandelen we in de paper. Hierin staan ook de details hoe we de data gesimuleerd hebben.

De simulatie laat zien dat ons model de effecten van de interventie goed voorspelt, ook als ze verschillen met individuele kenmerken van personen. De zwarte lijn toont de waarde van het echte effect van de interventie, de rode punten zijn de geschatte waarden, terwijl het 95-procentbetrouwbaarheidsinterval met grijs aangegeven is. De figuur laat duidelijk zien dat ons algoritme een waarde dicht bij nul voorspelt als de waarde nul is. Ook neemt het geschatte effect toe vanaf het moment dat  $x_1$  positief wordt en blijven de voorspellingen geconcentreerd rondom het echte effect. Voor zeer grote waarden van  $x_1$  zien we dat de voorspellingen lager zijn dan de echte waarde. Dit heeft een technische reden en is altijd het geval als groepen herhaaldelijk in tweeën gesplitst worden (Athey & Imbens, 2019).



Figuur 1: Schattingen van een causal forest with fixed effects

De figuur toont de gesimuleerde effecten van de interventie (zwarte lijn) en de geschatte effecten (rode stippen) door een causal forest with fixed effects op de y-as. Op de x-as staat de forcing variable, oftewel het product van  $x_1$  en het aantal perioden nadat de interventie begon. Het grijze gebied vormt het 95-procentbetrouwbaarheidsinterval van de geschatte effecten.

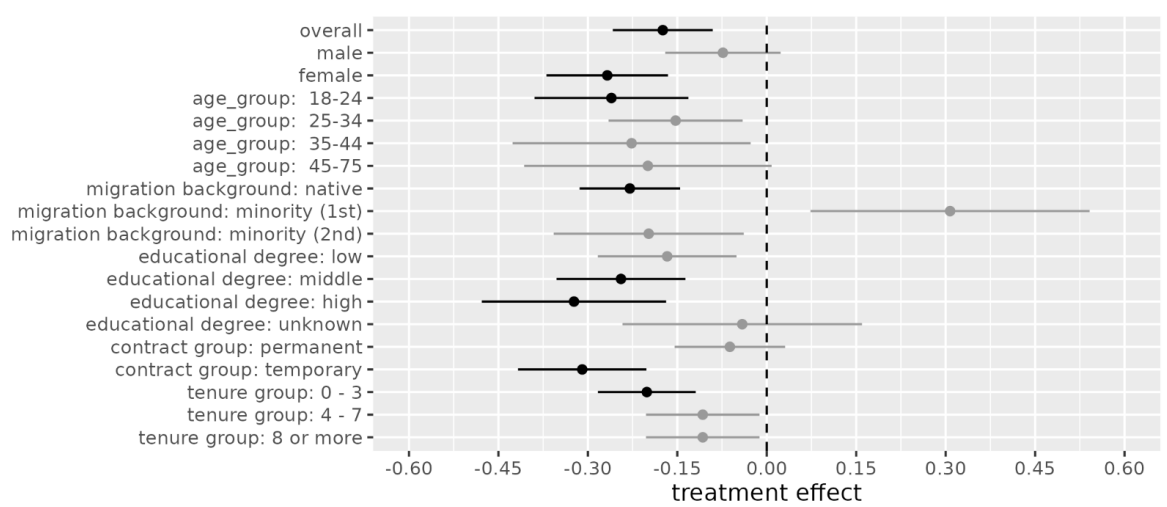
In de paper *Causal forest with fixed effects for heterogeneous treatment effects in difference-in-differences* laten we zien dat de voorspellingen voor het DCF een vergelijkbaar patroon laten zien, al zijn zowel de spreiding rondom het echte effect van de interventie als de afwijking voor grote waarden van  $x_1$  groter. Voor het MRCF zien we daarin dat voorspellingen niet in de buurt

liggen van het echte effect van de interventie en dat de afwijking toeneemt met de omvang (in absolute termen) van  $x_1$ . In de publicatie leggen we uit hoe dit komt. Ook laten we in ons paper zien dat het CFFE een kleiner gemiddeld verschil heeft tussen de geschatte en de door ons gesimuleerde waarde van het effect dan de MRCF of het DCF.

## 4 Praktijkvoorbeeld: Het effect van *payrolling* op het uurloon

**We passen de methode in ons onderzoek toe op Nederlandse data door het effect *payrolling* op het uurloon te schatten.** Hierbij volgen we de onderzoeksmethode beschreven in een recent CPB Discussion Paper over dit onderwerp (Goos et al., 2022). In het vervolg verwijzen we naar deze studie als GMSSB. Bij *payrolling* wordt het salaris en de werkgeverslasten betaald door de ‘payroller’, maar blijft een werknemer in dienst bij zijn huidige werkgever en verandert zijn takenpakket daar niet. We vergelijken werknemers die in een bepaald kwartaal gepayrolld worden met werknemers waarbij dit later gebeurt. We vergelijken deze groepen vier kwartalen voor en vier kwartalen na het moment dat *payrolling* begint.

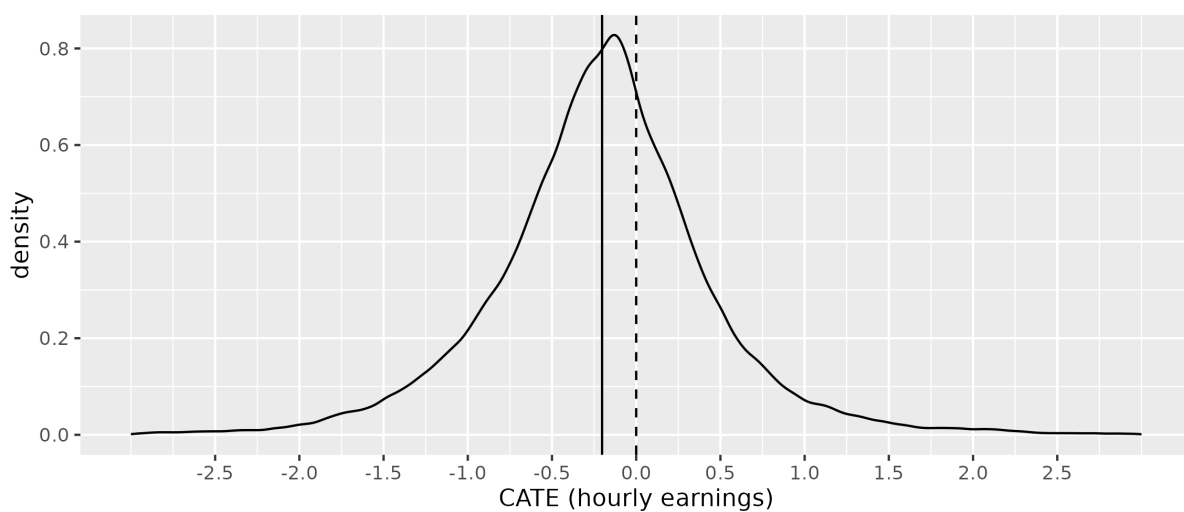
**Wanneer we de handmatige subgroepanalyse van GMSSB uitvoeren, zien we dat het gemiddelde effect van *payrolling* op het uurloon een daling van 17 cent per uur is.** We zien echter dat veel subgroepen een sterker effect kennen. Zo is de daling voor hoogopgeleiden en werknemers met een tijdelijk contract ongeveer 32 cent per uur. Voor de meeste andere groepen schommelt de gevonden daling rond de 20 tot 25 cent per uur, zie figuur 2. Gevonden effecten in het zwart zijn robuuster dan die in het grijs, omdat ze ook significant zijn na de Holm-Bonferronicorrectie (Holm, 1979). Deze correctie voorkomt dat misleidende conclusies getrokken worden wanneer er meerdere hypothesen getest worden, een risico dat bij analyse van een groot aantal subgroepen reëel is.



**Figuur 2: Effecten van *payrolling* (handmatige subgroepen)**

De figuur toont het gemiddelde effect van payrolling (overall) en het effect voor op de y-as gespecificeerde subgroepen. De stip laat het geschatte effect zien, de lijnen vormen het 95-procentbetrouwbaarheidsinterval. Als de waarde 0 op de x-as in het betrouwbaarheidsinterval ligt, wijkt het geschatte effect niet significant af van 0 en wordt de nulhypothese dat er geen effect is niet verworpen. In het zwart getoonde effecten en betrouwbaarheidsintervallen doorstaan ook de Holm-Bonferronicorrectie en zijn daarmee robuuster.

**Wanneer we een CFFE gebruiken, blijkt dat geschatte effecten van payrolling sterk variëren met kenmerken van de werknemer.** Deze conclusie volgt uit figuur 3, die de verdeling van geschatte effecten weergeeft via een *density plot*. Het CFFE heeft gebruik gemaakt van dezelfde variabelen die gebruikt zijn in de handmatige analyse door GMSSB, om een eerlijke vergelijking mogelijk te maken. Het algoritme zoekt dus naar individuele verschillen in het effect van payrolling op basis van het geslacht van werknemers, hun leeftijd, hun migratieachtergrond, hun opleidingsniveau, hun type contract (vóór payrolling begon) en hun aantal jaren dienstverband. Figuur 3 toont de verdeling van het geschatte effect voor payrolling op het uurloon. Het gemiddelde effect is een daling van ongeveer 20 cent, wat in grote lijnen overeenkomt met het gemiddelde effect dat we volgens de conventionele methode vonden. De spreiding is echter groot wanneer gekeken wordt naar de karakteristieken van de werknemer, het kwartaal na payrolling of het kwartaal waarin overgegaan werd op payrolling.

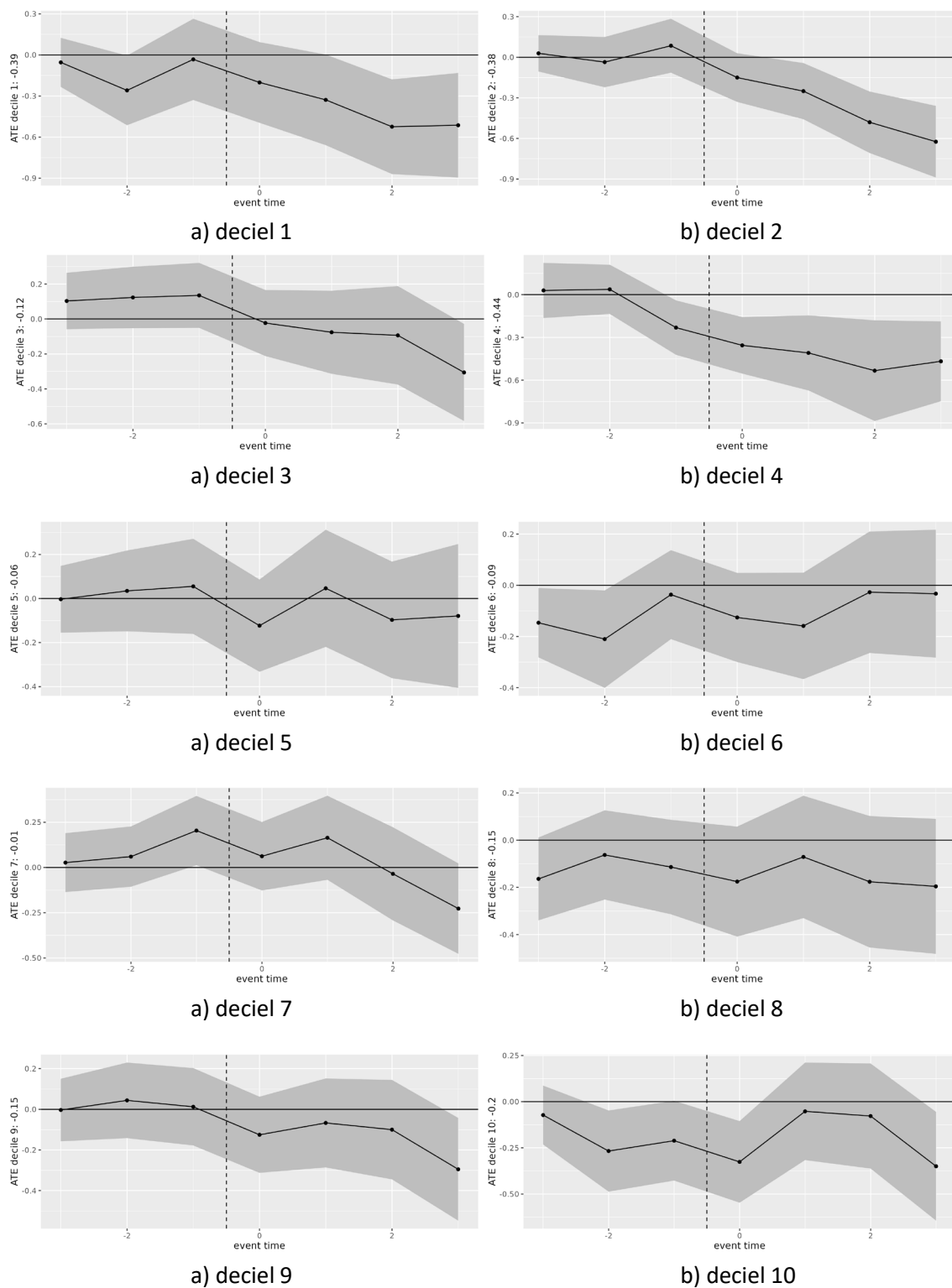


**Figuur 3: Histogram van de effecten van payrolling (CFFE)**

De figuur toont de geschatte effecten van een causal forest with fixed effects op het uurloon van werknemers. Het algoritme gebruikt de volgende variabelen om verschillen in effecten te vinden: geslacht, leeftijd, migratieachtergrond, opleidingsniveau, type contract en aantal jaren dienstverband. Doordat effecten per kwartaal geschat zijn, is impliciet ook rekening gehouden met de conjunctuur (kwartaaleffecten).

**Met een CFFE vinden we sterkere subgroep-specifieke effecten van payrolling dan bij een handmatige analyse voor een kleine groep werknemers** (Wager & Athey, 2018). Dit volgt uit een analyse waarbij we werknemers indelen in decielen op basis van hun met een CFFE geschatte effect van payrolling. Vervolgens schatten we het dynamische effect van payrolling. Hieruit volgen

twee conclusies. Ten eerste is het zo dat we alleen voor een specifieke groep werknemers een negatief effect vinden van *payrolling* op het uurloon. Dit bewijs vinden we alleen voor de twintig procent werknemers die het sterkst reageren op *payrolling*, nadat werknemers ingedeeld zijn in twee groepen op basis van hun geschatte effect van *payrolling*. Voor de ongeveer **tachtig** procent van de werknemers vinden we geen aanwijzingen dat *payrolling* het uurloon beïnvloedt, nadat ze ingedeeld zijn in acht groepen naar hun geschatte effect. Ten tweede is het gemiddelde effect voor werknemers in deze twee groepen fors groter (in absolute waarde) dan het gemiddelde effect voor de in figuur 4 getoonde subgroepen. Voor de twee decielen is het effect een daling van ongeveer 40 cent. De daling van het uurloon door *payrolling* neemt toe in het aantal kwartalen na de interventie.



**Figuur 4: dynamische effecten van payrolling op het uurloon voor decielen**

*Werknemers zijn ingedeeld in tien decielen op basis van hun gemiddelde, met een causal forest with fixed effects geschatte effect van payrolling op het uurloon. Vervolgens is het dynamische treatmenteffect geschat per deciel.*



## 5 Conclusie

**We hebben het bestaande *causal forest*-algoritme aangepast, waardoor deze methode efficiënter gebruikt kan worden bij *difference-in-differences*.** In het bijzonder hebben we het bestaande *causal forest*-algoritme uitgebreid, zodat het rekening kan houden met *fixed effects*. Hierdoor wordt het *causal forest*-algoritme nauwkeuriger en kan het gebruikt worden bij *difference-in-differences*. Onze simulaties laten zien dat een CFFE in staat is om individuele verschillen in causale effecten te vinden. Onze methode is efficiënter dan andere methoden die *causal forests* gebruiken bij *difference-in-differences* en kent bovendien een lagere gemiddelde afwijking.

**We vinden fors grotere verschillen tussen mensen wanneer het gaat om het effect van *payrolling* dan in eerder onderzoek.** GMSSB vonden dat het uurloon daalde met ongeveer 20 tot maximaal ongeveer 30 cent per uur voor een groot aantal subgroepen. Onze analyse laat zien dat het effect ongeveer 40 cent per uur is, maar dat dit alleen opgaat voor de twintig procent werknemers die volgens een CFFE het meest geraakt worden door *payrolling*. Voor bepaalde werknemers is het effect dus groter dan eerder aangetoond, maar voor anderen is er geen duidelijk effect waarneembaar. Dit laat zien dat de negatieve effecten van *payrolling* zich meer concentreren bij één specifieke groep mensen dan uit een handmatige analyse van subgroepen bleek.

**Onze aanpassing is een methodologische bijdrage die ons helpt om meer te leren over de werking van de economie in de praktijk.** Met onze aanpassing is het mogelijk om óók bij *difference-in-differences* op basis van de data te bepalen welke mensen het sterkst reageren op een beleidswijzigingen of interventie. Dit levert belangrijke inzichten op of het beleid verschillend uitpakt voor mensen en in welke mate dit het geval is. Dit helpt onderzoekers om meer te leren over de werking van de economie in de praktijk.

## 6 Literatuur

- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353-7360.
- Athey, S., & Imbens, G. (2019). Machine learning methods economists should know about. *Annual Review of Economics*, 11, 685-725.
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148-1178.
- Currie, J., Kleven, H., & Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. *AEA Papers and Proceedings*, 110, 42-48.
- De Chaisemartin, C., & d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, 26(3), C1-C30.
- Gavrilova, E., Langørgen, A., & Zoutman, F. (2023). *Dynamic Causal Forests, with an Application to Payroll Tax Incidence in Norway*.
- Goos, M., Manning, A., Salomons, A., Scheer, B., & van den Berge, W. (2022). *Alternative Work Arrangements and Worker Outcomes: Evidence from Payrolling*. CPB Netherlands Bureau for Economic Policy Analysis.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2), 65-70.
- Remmerswaal, M., Boone, J., Bijlsma, M., & Douven, R. (2019). Cost-sharing design matters: A comparison of the rebate and deductible in healthcare. *Journal of Public Economics*, 170, 83--97.
- Roos, A.-F., Diepstraten, M., & Douven, R. (2021). *When Financials Get Tough, Life Gets Rough?: Problematic Debts and Ill Health* (CPB Discussion Paper).
- Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.
- Somainsi, P., & Wolak, F. A. (2016). An algorithm to estimate the two-way fixed effects model. *Journal of Econometric Methods*, 5(1), 143-152.
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.