

Pimeän työn määrä

Mittaamisselvitys osakeyhtiöiden
verotarkastustietojen perusteella

Selvitys 4/2022



Julkaisun nimi:

Pimeän työn määrä – Mittaamisselvitys osakeyhtiöiden verotarkastustietojen perusteella

Tekijät:

Analyttikot Kirsi Ristola ja Lasse Winter; asiantuntijat Riitta Nummila ja Matti Hautala

Diaarinumero: VH/5717/02.98.00.00/2019 / 2013_051

Julkaisija: Verohallinto, Harmaan talouden selvitysyksikkö

Julkaistu: 22.11.2022

Julkaisutapa: Sähköinen (PDF)

Julkisuus: Julkinen

Lisätietoja medially:

Johtaja Janne Marttinen, puh. 029 512 6066

Apulaisjohtaja Marko Niemelä, puh. 029 512 6070

Muut yhteydenotot htsy@vero.fi

Julkaisualusta harmaa-talous-rikollisuus.fi ja harmaatalous.fi (vain luvite-
tuille viranomaisille)

Velvoitteidenhoitoselvityksen voi pyytää ilmiöselvityksessä kuvatus ryhmän perusteella. (Laki Harmaan talouden selvitysyksiköstä (1207/2010) 5:3 §)

Harmaan talouden selvitysyksikön selvityksissä esitetyt näkemykset ja tulkintakannanotot ovat yksikön omia, eivätkä ne sido Verohallintoa tai muita viranomaisia.

Tiivistelmä

Selvityksessä arvioidaan pimeän työn määrää osakeyhtiöissä verotarkastustietojen perusteella. Selvitys tarjoaa ensimmäistä kertaa Suomessa euromääräisiä arvioita yhdestä keskeisestä harmaan talouden ilmiöstä. Tehdyn data-analyysin perusteella pimeän työn laajuus näyttäytyy Suomessa maltillisena, jos sitä verrataan muihin EU-maihin.

Raportissa pimeää työtä ilmiönä taustoitetaan muun muassa Eurobarometri-kyselytutkimuksen avulla. Selvityksessä vertaillaan koko EU-alueella yhdenmukaisesti toteutetun kyselyn tuloksia Ruotsiin, Tanskaan ja Viroon. Osana pimeän työn ilmiön hahmottamista esitellään myös 2000-luvulla Suomessa tehtyjä lainsäädäntöhankkeita ja niitä koskevia arvioita.

Selvityksen tärkein osa käsittelee verotarkastustietojen analyysin perusteella saatuja arvioita pimeän työn määrästä Suomessa. Tarkoituksena on hahmottaa ilmiön euromääräistä laajuutta, ei sen trendiä eli suuntaa.

Data-analyysissä on tarkasteltu neljän eri menetelmän keinoin vuosina 2014–2020 tehtyjen noin 7 500 verotarkastuksen tietoja. Arvio pimeiden palkkojen määrästä on noin 670–850 miljoonaa euroa vuodessa. Summa sisältää sekä osakasyrittäjien että palkansaajien pimeät tulot. Arvioidut ennakonpidätysten ja sosiaalivakuutusmaksujen menetykset ovat arviolta noin 260–300 miljoonaa euroa vuodessa.

Tyypillinen pimeää työtä teettävä yritys on pieni ja osakkaat ovat pimeiden tulojen saajina merkittävässä roolissa. Keskeisiä toimialoja, joilla pimeää työtä esiintyy, ovat rakennusala, vähittäis- ja tukkukauppa sekä ravitsemusala.

Selvityksen perusteella pimeän työn tehokkaan torjunnan toteuttamiseksi on tärkeää, että harmaan talouden toimenpideohjelmat jatkuisivat myös seuraavilla hallituskausilla. Pimeän työn torjuminen edellyttää jatkossakin monenlaisia toimenpiteitä. Näitä ovat esimerkiksi terveen kilpailun ja työmarkkinoiden edistäminen, harmaan talouden ennalta estämiseen liittyvät toimenpiteet, viranomaisten toimintaedellytyksien turvaaminen sekä viranomaisyhteistyön kehittäminen.

Avainsanat: Mittaaminen; Pimeä työ; Verotarkastus; Verovaje

Sisällys

1	Johdanto	1
2	Tutkimusongelma ja -kysymykset	2
2.1	Selvityksen rakenne	3
2.2	Aineisto	4
2.3	Selvityksessä käytetyt laskentamenetelmät.....	4
2.4	Veronmenetysten laskeminen	4
2.5	Selvityksen rajaukset.....	6
3	Pimeä työ tutkimuskirjallisuudessa	8
3.1	Kotimainen tutkimus pimeästä työstä	8
3.2	Tutkimus pimeästä työstä EU-alueella.....	10
4	Toteutetut hankkeet pimeä työn vähentämiseksi	13
4.1	Suomessa toteutetut hankkeet	13
4.1.1	Kotitalousvähennys	13
4.1.2	Tilaajavastuulaki.....	14
4.1.3	Rakennusalan lainsäädäntötoimet.....	14
4.1.4	Palkka pankkiin -lainsäädäntö.....	15
4.1.5	Tulorekisterilainsäädäntö	15
4.1.6	Ulkomaiseen työvoimaan kohdistuneet hankkeet.....	16
4.1.7	Pimeän työn torjunta harmaan talouden torjuntaohjelmissa	16
4.1.8	Yhteenvedoa ja analyysiä kotimaisista hankkeista	17
4.2	Norjan ja Ruotsin ns. henkilöstöpäiväkirjan pitämisvelvollisuus	17
5	Pimeän työn ja palkan käsitteet	18
5.1	Selvityksessä tarkoitettu pimeä työ ja palkka.....	18
5.2	Pimeä työ ja palkka harmaan talouden osa-alueena	19
5.2.1	Pimeä palkka maksajan kannalta sekä vältetyt maksut ja verot.....	20
5.2.2	Pimeä palkka työntekijän ja yhteiskunnan kannalta	21
5.2.3	Yhteenvedo pimeästä palkasta aiheutuvasta ennakonpidätysten ja sosiaalivakuutusmaksujen vajeesta.....	22
5.3	Yleisimmät pimeän työn ja palkan muodot	22
5.3.1	Täysin pimeä työ ja palkka	22
5.3.2	Osittain pimeä työ ja palkka.....	23
5.3.3	Veronalainen palkka verovapaina kustannusten korvauksina.....	23
5.3.4	Veronalaiset luontoisedut jätetään ilmoittamatta.....	23
5.3.5	Osakkaan pimeä palkka ja peitelty osinko	24
6	Selvityksen verotarkastusaineistot ja yritysjoukot	25
6.1	Verotarkastukset	25
6.2	Verotarkastuksissa havaitut virheet.....	26
6.3	Selvityksen perusjoukot.....	29
6.3.1	2014–2020 toimivat yritykset.....	29
6.3.2	2014–2020 toimimattomat yritykset.....	30

7	Pimeiden palkkojen määrä Suomessa eri menetelmin arvioituna	31
7.1	Pimeät palkat ja peitelty osinko 2018–2018: substanssiperusteinen kaltaistamismenetelmä	31
7.1.1	Toimivien osakeyhtiöiden maksamat pimeät palkat ja peitelty osinko	32
7.1.2	Toimimattomien osakeyhtiöiden maksamat pimeät palkat ja peiteltyt osingot	34
7.2	Pimeät palkat ja peiteltyt osingot 2017–2020: algoritmilähtöiset laskentamenetelmät	36
7.2.1	CEM-kaltaistaminen	36
7.2.2	DBSCAN-kaltaistaminen	37
7.2.3	Yrityskohtaisen odottaman laskenta kahdella tavalla	37
8	Pimeän työn euromäärä sekä verojen ja muiden maksujen menetykset	41
8.1	Pimeiden palkkojen määrä 2014–2020 eri menetelmillä laskettuna	42
8.1.1	Pimeät palkat yhteensä	42
8.1.2	Osakkaiden ja palkansaajien pimeät palkat	43
8.1.3	Toimimattomien yritysten pimeät palkat	44
8.2	Pimeistä palkoista aiheutuneet verojen ja muiden maksujen menetykset	45
8.3	Tyypillinen pimeän työn yritys	46
8.3.1	Pimeän työn yritysten koko liikevaihtoluokittain	47
8.3.2	Pimeän työn yritykset toimialoittain	49
9	Yhteenveto ja johtopäätökset	52
10	Lähteet	53
11	Luettelo kuvioista ja taulukoista	56
12	Liitteet	58
	Liite 1. Osakasyrittäjän eläkevakuutusmaksun määräytymisperusteet	1
	Liite 2. Substanssiperusteinen kaltaistamismenetelmä	2
	Liite 3. Algoritmivetoiset kaltaistusmenetelmät	9
	Liite 4. Yrityskohtaisen odottaman laskenta	15

1 Johdanto

Harmaan talouden selvitysyksikössä on laadittu neljä pimeää työtä ja palkkaa käsittelevää selvitystä¹. Tämä selvitys on yksi laajasta selvitetystä kokonaisuudesta. Selvityksessä on arvioitu osakeyhtiömuotoisissa yrityksissä tapahtuvaa veronvälttelyä pimeän palkan maksamisen muodossa. Selvityksessä tutkitaan pimeää työtä, toisin sanoen pimeästi maksettuja palkkoja ja osakkaiden peiteltyjä osinkoja erityisesti verotarkastustietojen avulla. Tarkoituksena on muodostaa mahdollisimman kattava kuva siitä, mitä kaikkea pimeä työ pitää sisällään ja arvioida ilmiön laajuutta data-analyysin avulla erilaisia menetelmiä käyttäen.

Pimeä työ mielletään yleensä palkansaajan asemassa tehdyksi työksi, josta työnantajan ja työntekijän yhteisellä sopimuksella jätetään maksamatta palkan verot ja muut lakisääteiset velvoitteet, kuten työeläkemaksut. Usein pimeästä työstä maksetaan korvaus käteisellä rahalla.

Pimeä työ voidaan kuitenkin perustellusti määritellä myös edellistä huomattavasti laajemmaksi ilmiöksi, joka koskettaa kaikenlaista niin yrittäjän kuin palkansaajankin tekemää työtä, josta maksettu korvaus suoritetaan verot ja muut lakisääteiset maksut laiminlyöden. Laajasti ajatellen pimeää palkkaa ovat esimerkiksi perusteettomasti verovapaana maksetut kilometrikorvaukset, yrityksestä vastikkeetta saadut hyödykkeet ja varat tai vaikkapa liikematkaksi naamioitu lomareissu. Pimeä työ ja pimeästi maksetut palkat ilmenevät usein samanaikaisesti sellaisen työn ja palkanmaksun rinnalla, jossa huolehditaan lakisääteisistä veroista ja maksuista. Näin ollen kaikki pimeä työ ei suinkaan ole sen räikeintä ilmenemismuotoa, jossa kaikki työnantajan lakisääteiset velvoitteet laiminlyödään. Pimeän työn skaala on siten hyvin laaja, eikä sen toteuttamistapoja voi tyhjentävästi määritellä ilmiöiden muuttuessa ajan saatossa.

Suomessa nimenomaisesti pimeään työhön kohdistuvaa tutkimusta on tehty erittäin vähän. Pimeän työn merkitys voi kuitenkin olla varsin suuri, sillä Ruotsin Verohallinto on arvioinut vuonna 2020 julkaistussa selvityksessä pimeän työn muodostavan jopa puolet maan verovajeesta. Verovajeen pitäessä sisällään kaikenlaiset lailliset ja laittomat syyt verojen maksamatta jättämiselle, on pimeän työn merkitys julkiselle taloudelle siten mahdollisesti huomattavan iso. Tämä oli selvityksen laatijoiden hypoteesina selvitystyötä aloitettaessa. Selvityksen tulosten perusteella osakeyhtiöissä tehtävä pimeä työ muodostaa euromääräisesti merkittävän harmaan talouden osa-alueen, mutta toisaalta pimeiden palkkojen ja peiteltyyn osingon osuus palkkasummasta on eri menetelmillä mitattuna maltillinen 1,4–1,5 prosenttia.

¹ Selvitykset on julkaistu 22.11.2022. Pimeän työn ilmiöselvitykset ovat: *Pimeän työn määrä – Mittaamisselvitys osakeyhtiöiden verotarkastustietojen perusteella 4/2022*, *Pimeän työn määrä – mittaamisselvitys elinkeinonharjoittajien verotarkastustietojen perusteella 5/2022*, *Pimeän työn mittaaminen profilointi menetelmällä - Osakeyhtiöiden pimeän työn määrä 06/2022* ja *Pimeän työn mittaaminen LIM-menetelmällä – Yksityisen sektorin palkansaajien pimeän työn määrä 7/2022*. Lisäksi pimeän työn selvityksistä on julkaistu kooste: *Pimeä työ Suomessa – kooste selvityksistä 2022, 8/2022*.

2 Tutkimusongelma ja -kysymykset

Selvityksen tutkimusongelmana on tarjota uutta tietoa pimeästä työstä: luotettavia arvioita siitä, kuinka merkittävästä ilmiöstä on kyse. Selvityksessä termejä *pimeä työ* ja *pimeä palkka* käytetään synonyymeina.

Pimeä työ on ilmiö, joka koskettaa monia eri valvontaviranomaisia. Konkreettisesti pimeisiin työntekijöihin ”törmäävät” lähinnä työsuojeluviranomaiset ja verotarkastajat. Selvitys on rajattu koskemaan verotarkastusten perusteella tehtyjä havaintoja pimeästä työstä. Pimeä työ on siitä hankalasti tutkittava ilmiö, että tosiasiaassa maksettujen pimeiden palkkojen ja pimeiden yrittäjätulojen tarkka selvittäminen on mahdotonta. Pimeiden palkkojen maksutapa on 2020-luvullakin usein käteinen raha, jonka seuranta on haastavaa. Näin ollen ilmiöstä ei ole mahdollista saada täsmällistä tietoa.

Selvityksessä kartoitetaan pimeän työn määrää ja esiintyvyyttä, rajaten tarkastelu kohdistumaan osakeyhtiömuotoisiin yrityksiin. Selvitys ei kata toiminimiyrittäjän pimeää työtä eikä henkilöyhtiöiden vastuualaisten yhtiömiesten pimeää työtä eikä näiden työntekijöilleen maksumia palkkoja. Selvityksessä tarkastellaan osakeyhtiöiden ja ulkomaisten yhteisöjen palkanmaksussa tehtyjä pimeään työhön liittyviä verotarkastushavaintoja. Pimeä työ pitää sisällään niin palkansaajan asemassa tehdyn työn kuin osakkaidenkin vastaavat toimet. Osakeyhtiöt maksavat valtaosan yritysten maksamista palkoista², joten niihin suoritettujen verotarkastusten havaintoja analysoimalla on mahdollista muodostaa kattava kuva pimeään työhön liittyvästä harmaasta taloudesta.

Harmaan talouden kvantitatiiviset tutkimusmenetelmät voidaan yleensä jaotella karkeasti kahteen erilaiseen tutkimusorientaatioon: suoriin ja epäsuoriin menetelmiin (Harmaan talouden selvitysyksikkö, 2014). Tässä selvityksessä käytettävät menetelmät ovat niin sanottuja suorita menetelmiä, jossa tarkastelu perustuu verotarkastuksissa tehtyihin havaintoihin. Tämän selvityksen lisäksi samanaikaisesti julkaistaan kolme muuta pimeää työtä koskevaa selvitystä. Myös selvityksessä *Pimeän työn mittaaminen profilointimenetelmällä – Osakeyhtiöiden pimeän työn määrä* harmaata taloutta analysoidaan suoran lähestymistavan ja verotarkastushavaintojen avulla. Selvityksessä *Pimeän työn mittaaminen LIM-menetelmällä – Yksityisen sektorin palkansaajien pimeän työn määrä* pimeään työhön perehdytään epäsuoran menetelmän avulla. Menetelmän nimi on LIM (labour input method). Selvitys *Pimeän työn määrä – Mittaamisselvitys elinkeinonharjoittajien verotarkastustietojen perusteella* käsittelee elinkeinonharjoittajien pimeää työtä. Myös siinä käytetään suorita harmaan talouden mittausmenetelmiä ja verotarkastushavaintoja.

Selvityksessä pimeillä palkoilla tarkoitetaan palkansaajille tai osakkaille työstä maksettuja erilaisia rahallisia tai rahanarvoisia korvauksia, jotka määrätelmän mukaisesti jätetään

² Yrityssektori maksoi 61 miljardia euroa eli 66 prosenttia kaikista palkoista (93 mrd. €) Suomessa vuonna 2020. Toinen suuri palkanmaksaja oli julkisyhteisöt 27 prosentin osuudella. (Lähde: kansantalouden tilinpito, Tilastokeskus). Yrityssektori käsittää osakeyhtiöiden lisäksi myös kaikki muut taloudellista toimintaa harjoittavat oikeudelliset muodot. Selvityksen perusjoukon yritykset maksoivat 58 miljardia euroa palkkoja vuonna 2020 (Lähde: Verohallinto), eli noin 94 prosenttia kaikista yrityssektorin palkoista. Tiedot on poimittu eri lähteistä, eikä niiden vertailukelpoisuutta ole varmistettu. On kuitenkin selvää, että perusjoukko kattaa suurimman osan yrityssektorin maksamista palkoista.

ilmoittamatta Verohallinnolle³. Selvityksessä arvioidaan myös palkkoihin liittyvien virheiden ja ilmoittamisen laiminlyöntien vuoksi aiheutuvaa verotulojen, työeläkemaksujen ja erilaisen palkkojen sivukulujen⁴ menetystä.

Tutkimusongelmaan haetaan vastauksia seuraavien kysymysten avulla:

- Kuinka paljon Suomessa maksetaan vuosittain pimeitä palkkoja?
- Kuinka paljon ennakonpidätyksiä ja sosiaalivakuutusmaksuja pimeiden palkkojen vuoksi jää kertymättä?
- Kuinka paljon osakeyhtiön osakkaat nostavat yrityksistä varoja ilman, että niistä maksetaan asianmukaisesti verot (peitelty osinko ja ennakoperinnän alaiset väärinkäytökset)?
- Minkä kokoisissa yrityksissä osakkailla on havaittu eniten palkkaan liittyviä väärinkäytöksiä? Onko toimialojen välillä havaittavissa eroja?

Harmaan talouden kokonaismäärää Suomessa on pyritty arvioimaan aikaisemmin, muun muassa vuonna 2010 Eduskunnan tarkastusvaliokunnalle laaditussa tutkimusjulkaisussa⁵. Pimeää työtä ja palkkaa koskevassa tutkimuksessa on selkeä tutkimusaukko: arvioita pimeän työn kokonaismäärästä ei ole aiemmin tehty.

2.1 Selvityksen rakenne

Selvityksen kolmannessa luvussa käydään lyhyesti läpi pimeää työtä koskevia kotimaisia ja ulkomaisia selvityksiä ja tutkimuksia. Neljännessä luvussa esitellään ja analysoidaan muutamia nimenomaisesti pimeää työtä ehkäisemään tehtyjä suomalaisia sekä pohjoismaisia lakihankkeita. Viidennessä luvussa tarkennetaan tässä selvityksessä käytetty pimeän työn määritelmä sekä kuvataan pimeän työn merkitystä muutamien esimerkkitapausten kautta. Kuidennessa luvussa esitellään selvityksen tarkastellut joukot ja seitsemännessä luvussa menetelmät, joilla pimeiden palkkojen määrää on analysoitu. Kahdeksannessa luvussa esitellään ydintulokset eli data-analyysin perusteella tuotetut arviot osakeyhtiömuotoisten yritysten pimeän palkanmaksun suuruudesta ja veronmenetyksen määrästä. Yhdeksäs luku on johtopäätösluku, joka kokoaa selvityksen keskeiset havainnot. Selvityksen lopussa ovat liitteet, joissa käsitellään lyhyesti yrittäjien eläkevakuuttamista sekä data-analyysin menetelmät tarkalla tasolla.

³ On tärkeä huomata, että käsite eroaa Verohallinnon käyttämästä vakiintuneesta terminologiasta. Tavallisesti pimeillä palkoilla tarkoitetaan vain palkansaajan asemassa oleville henkilöille työnteosta maksettuja, verotukseen ilmoittamattomia korvauksia. Yleensä kyse on vakavista harmaan talouden verotarkastuksilla havaituista tapauksista.

⁴ Sivukuluja ovat työntekijän työttömyysvakuutusmaksu (TVR-maksu), työnantajan työttömyysvakuutusmaksu, työnantajan sairausvakuutusmaksu (sava-maksu) ja tapaturma- ja henkivakuutusmaksu. Pimeän palkan maksuun liittyy näiden lisäksi myös muita ulottuvuuksia, kuten ulosoton onnistuminen tai epäonnistuminen sekä yhteys sosiaalietuuksiin. Tässä selvityksessä ei tarkastella näitä ulottuvuuksia.

⁵ Suomen kansainvälistyvä harmaatalous, kesäkuu 2010, Markku Hirvonen, Pekka Lith ja Risto Walden.

2.2 Aineisto

Selvityksen laadullinen osa koostuu relevantin kirjallisuuden ja viranomaisaineistojen analyysistä. Pääasiallisena lähteenä on Eurobarometrin sekä EU:n pimeän työn foorumin tuottama tieto. Kirjallisten lähteiden avulla pyritään muodostamaan kokonaiskuva siitä, millaisena ja kuinka laajana ilmiönä pimeä työ kuvataan erityisesti Suomessa, mutta myös naapurimaisamme sekä laajemmin EU-tasolla. Ennakkokäsitys aiheesta on se, että erityisesti numeraalista tietoa pimeän työn määrästä ei ole juurikaan saatavilla.

Data-analyysin aineisto perustuu Verohallinnon tekemiin riskiperusteisiin ja satunnaisotannalla toteutettuihin verotarkastuksiin. Tarkastelussa käytetään sekä vuosiin 2014–2018 kohdistuneita satunnaisverotarkastuksia että vuosiin 2014–2020 kohdistuneita muita tarkastuksia. Aineisto, josta kohdejoukko on valittu, koostuu niistä yrityksistä, joiden verotarkastuksilla on havaittu palkkaan liittyviä virheitä ja yhtiöille on määrätty veroja maksettavaksi. Peitellyn osingon tapaukset eli yritykset, joissa osakkaiden perusteettomasti yrityksestä nostamat varat on todettu verotarkastuksilla, on otettu mukaan tarkasteluun riippumatta siitä, kuuluuko yritys työnantajarekisteriin vai ei.

2.3 Selvityksessä käytetyt laskentamenetelmät

Pimeiden palkkojen määrää arvioidaan tässä selvityksessä verotarkastuksista saaduilla tiedoilla. Pimeiden palkkojen määrän arviointiin käytetään useampaa menetelmää mahdollisimman luotettavan kokonaiskuvan muodostamiseksi. Käytettävät analyysitekniikat ovat:

- 1.) Substanssiperusteinen kaltaistamismenetelmä
- 2.) Algoritmilähtöinen kaltaistamismenetelmä CEM
- 3.) DBSCAN-klusterointialgoritmi-tekniikka
- 4.) XGBoost-koneoppimismalli, sekä
- 5.) Bayesilainen logistinen regressiomalli

Substanssiperusteisessa kaltaistamismenetelmässä on laskettu arviot pimeiden palkkojen määristä vuosille 2014–2018 kaikkien verotarkastusten perusteella sekä pelkkien satunnaisvalintaan perustuvien verotarkastusten avulla.

Arviot pimeiden palkkojen määrästä vuosille 2017–2020 on laskettu neljällä eri menetelmällä. Tällöin aineistona on käytetty kaikkia verotarkastuksia. Kaksi edellisessä listassa ensin mainittua ovat algoritmilähtöisiä kaltaistusmenetelmiä: CEM (Coarsened Exact Matching) ja DBSCAN-klusterointialgoritmi. Lisäksi asiakaskohtaista odottamaa pimeiden palkkojen riskisyydelle ja esiintyvyydelle on ennustettu XGBoost-koneoppimismallilla ja bayesilaisella logistisella regressiolla. Menetelmien tuloksia on esitelty luvussa 7 ja lisäksi jokaisesta menetelmästä on tarkka kuvaus selvityksen liitteenä omassa kohdassaan.

2.4 Veronmenetysten laskeminen

Selvityksen varsinainen data-analyysi koskee pimeiden palkkojen määrää (luku 7). Myöhemmin pimeiden palkkojen estimaateista johdetaan verojen ja muiden julkisten maksujen

menetyksiä koskevat arviot (luku 8.2). Tässä luvussa kerrotaan, mihin veronmenetysten laskeminen perustuu.

Ennakonpidätykset

Käytämme tässä selvityksessä ennakonpidätysten määrän laskemiseen 25 prosenttia. Taustalla on korkeimman oikeuden (KKO) ratkaisu⁶, jossa KKO on ottanut kantaa vältetyn veron määrän arviointiin törkeän veropetoksen rikosasiassa. Kyse oli pimeästä palkanmaksusta, ja päätöksen perusteluissa on todettu aiempien oikeusasteiden laskentaperusteen (25 prosenttia) olleen soveltuva ennakonpidätyksen määräksi. Selvityksessä tätä prosenttia käytetään sekä palkansaajien että osakkaiden pimeiden palkkojen veronmenetystä arvioitaessa. Osakasyrittäjien pimeän palkan veronmenetys lasketaan koko peiteltyyn osingon arvosta, vaikka vallitseva oikeustila on, että vain 75 prosenttia peittelystä osingosta on veronalaista ansiotuloa.⁷ Pimeiden palkkojen laskennassa arvioidaan, kuinka paljon veroja olisi kertynyt, jos osakkaat olisivat toimineet oikein.⁸ Siksi peiteltyyn osingon 75 prosentin sääntöä ei sovelleta.

Verotarkastuksilla havaittujen pimeiden palkkojen ennakonpidätys (vero) määrätään palkanmaksajalle maksettavaksi enintään 40 prosentin mukaan⁹. Tämä on yleensä huomattavasti enemmän kuin mikä palkansaajan todellinen ennakonpidätysprosentti olisi ollut verokortin mukaisesti. Mikäli tässä selvityksessä käytettäisiin veronmenetyksen laskennassa verotarkastuksilla yhtiölle maksettavaksi määrättyjä ennakonpidätyksiä, olisi määrä liian suuri. Todellista ennakonpidätysprosenttia puolestaan on mahdoton käyttää, koska siihen vaikuttavat kaikki kunkin palkansaajan tulot ja verotuksessa tehtävät vähennykset; verotettavan tulon määrä selviää vasta henkilön lopullisessa verotuksessa.

Palkansaajien sosiaalivakuutusmaksut

Sosiaalivakuutusmaksujen¹⁰ määrästä säädetään erikseen kutakin maksua koskevassa laissa tai niitä täydentävissä asetuksissa¹¹. Sosiaalivakuutusmaksujen veron menetysten laskennassa on käytetty sosiaali- ja terveysministeriön vuodelle 2021 vahvistamia prosentteja (Sosiaali- ja Terveysministeriö). Laskennassa käytetään oletusta, että pimeän palkan saaja on täysi-ikäinen, alle 53-vuotias ja työskennellyt sellaisen yrityksen palveluksessa, jonka vuoden 2020 aikana maksamat palkat ovat olleet alle 2 125 500 euroa¹². Palkansaajien

⁶ KKO 2017:75

⁷ TVL 33 d §:n mukaan peittelystä osingosta on 75 prosenttia veronalaista ansiotuloa ja 25 prosenttia verovapaata tuloa. Peitelty osinko on kuitenkin kokonaan veronalaista ansiotuloa, jos peitelty osinko on TVL 33 c 3 momentissa tarkoitettu ulkomaiselta yhteisöltä.

⁸ Yleisesti verovajeiden laskennassa toteutunutta verokertymää verrataan niin sanottuun potentiaaliseen verokertymään. Potentiaalinen verokertymä on se verojen määrä, joka saataisiin, jos kaikki verovelvolliset maksisivat verot oikein ja lainmukaisesti.

⁹ Laki oma-aloitteisista veroista 51 § 1 mom. Palkanmaksajalle määrätty ennakonpidätykset palautetaan sen jälkeen, kun suorituksen saajat ovat maksaneet tuloista heille määrättyt verot ja lakisääteiset maksut.

¹⁰ Työntekijän työeläkevakuutusmaksu (TyeL-maksu), työntekijän työttömyysvakuutusmaksu (TVR-maksu), työnantajan eläkevakuutusmaksu, työnantajan työttömyysvakuutusmaksu, työnantajan sairausvakuutusmaksu (savamaksu).

¹¹ Työntekijän eläkelain 153 §, Laki työttömyyssetuuksien rahoituksesta 18 §, Laki työnantajan sairausvakuutusmaksusta, Valtioneuvoston asetus sairausvakuutusmaksujen maksuprosenteista vuonna 2021.

¹² Oletukset perustuvat ajatukseen ”keskimääräisestä” pimeän palkan saajasta. Tosiasialliset maksut vaihtelevat eri ikäryhmien kesken, asiaa on esitelty tarkasti luvussa 5.2.1 / sivulla 20.

maksamattomien niin sanottujen palkan sivukulujen eli sosiaalivakuutusmaksujen määrä (noin 28 prosenttia) esitetään taulukossa 1 käyttäen vuoden 2021 arvoja.

Taulukko 1. Palkan sivukulut ja niiden suuruus vuonna 2021

Sosiaalivakuutusmaksut	Sivukulun suuruus (prosenttiosuus bruttopalkasta)
TyEL (työnantajan osuus)	16,95 %
TyEL (työntekijän osuus)	7,15 %
Työnantajan työttömyysvakuutusmaksu	0,50 %
Työntekijän työttömyysvakuutusmaksu	1,40 %
Työnantajan sairausvakuutusmaksu	1,53 %
Työnantajan tapaturmavakuutusmaksu	0,70 %
Yhteensä	28,23 %

Lähde: Sosiaali- ja terveystieteiden tutkimuskeskus 2021.

Selvityksessä ei arvioitu yrittäjien eläkevakuuttamisen (YEL) puutteista tai pimeistä palkoista johtunutta vajetta, koska selvityksessä ei ollut käytettävissä YEL-vakuuttamisen yrittäjäkohtaisia tietoja. Yrittäjien eläkejärjestelmän salliessa alivakuuttamisen suhteessa verotettaviin tuloihin ovat oletettavat maksujen menetykset kohtuullisen pieniä palkansaajiin verrattuna. Osakasyrittäjän YEL-vakuuttamisen perusteita on kuvattu selvityksen liitteessä 1.

2.5 Selvityksen rajaukset

Selvitys on rajattu koskemaan osakeyhtiöitä, jotka ovat olleet tarkasteltavana ajanjaksona palkkoja maksavien työnantajien rekisterissä, ennakkoperintärekisterissä tai liiketoiminnasta arvonlisäverovelvollisten rekisterissä. Osakeyhtiöitä ja ulkomaisia yhteisöjä, jotka eivät ole olleet rekistereissä, on tarkasteltu lyhyesti omana erillisenä joukkonaan¹³.

Pimeää työtä ja sen vuoksi vältettyä verojen maksua tarkastellaan verotarkastuksilla havaittujen virheiden avulla. Selvityksen data-analyysin tuloksiin vaikuttaa se, mille toimialoille ja mihin veroriskeihin Verohallinto on valvontaa kohdistanut. Niillä toimialoilla, joihin verotarkastuksia kohdistuu vähän tai ei lainkaan, estimaatit ovat epävarmempia. Haaste on suurin toimialoilla, joissa yritysten liikevaihdon tunnusluvut ja muut harmaata taloutta ennustavat indikaattorit poikkeavat paljon muista yrityksistä. Tällöin verotarkastuksilla tehtyjä havaintoja ei voida onnistuneesti yleistää näille toimialoille. Kuitenkin verotarkastusten vähäinen määrä tai kokonaan puuttuminen tietyiltä toimialoilta ei ole niin merkittävä ongelma tähän joukkoon tarkastustuloksia yleistettäessä, etteikö tuloksia voisi pitää kohtuullisen luotettavina.

Selvitys kattaa palkkoihin liittyvät virheet, joissa yhtiölle on määrätty verotarkastuksella maksettavaksi lisää ennakonpidätyksiä (eli palkkatulojen veroja) tai lähdeveroa (ulkomaalaiselle maksetun palkan veroa) liian pienenä ilmoitetun tai kokonaan ilmoittamatta jätetyn rahapalkan vuoksi. Selvityksessä ovat mukana myös tilanteet, joissa luontoisedut on ilmoitettu arvotaan vääränä tai jätetty kokonaan ilmoittamatta. Selvitykseen sisältyvät myös tilanteet, joissa rahapalkan virhe on aiheutunut väärin perustein tai puutteellisin perustein maksetuista matkakustannusten korvauksista tai muista virheellisesti verovapaina maksetuista kustannusten korvauksista. Mukana tarkastelussa ovat myös tilanteet, joissa osakeyhtiön on katsottu

¹³ Tähän joukkoon sisältyy yrityksiä, jotka eivät ilmoita mitään tietoja Verohallinnolle, myös toimintansa virallisesti lopettaneet yritykset.

antaneen osakkaalleen rahanarvoisen edun osakasaseman perusteella, toisin sanoen peitel-
lyn osingonjaon tilanteet¹⁴.

Laajasti ajateltuna pimeänä työnä voidaan pitää myös tilannetta, jossa työnantaja teettää
työntekijällä työtä, mutta maksaa siitä alle yleis- tai normaalisitovan työehtosopimuksen mu-
kaista palkkaa. Tätä kutsutaan usein alipalkaksi. Työnantaja hyötyy menettelystä, koska työ
tulee tehtyä pienin kustannuksin, kun osa palkasta jätetään maksamatta. Näiden väärinkäy-
töstilanteiden ja niissä menetettyjen verojen ja lakisääteisten maksujen määrän arvioiminen
on hankalaa. Tapausten selvittäminen edellyttää usein monien viranomaisten yhteistyötä, ta-
pauskohtaista tarkastelua ja työrikosten tutkintaa (työsyryntä, kiskonnantapainen työsyryntä,
jopa ihmiskauppa). Tämä selvitys ei koske alipalkkaustilanteita. Harmaan talouden selvitys-
yksikkö on laatinut ulkomaalaisten alipalkkausta harmaan talouden ilmiönä käsittelevän selvi-
tyksen vuonna 2020, ja selvitys on luettavissa harmaa-talous-rikollisuus.fi-sivustolla¹⁵.

Verolajien näkökulmasta tarkasteltuna selvityksen ulkopuolelle jäävät muun muassa arvonli-
säverotuksen ja yritysten tuloverotuksen virheet. Ennakonpidätysten muita väärinkäytöksiä
kuten esimerkiksi ennakonpidätysten perusteettomia palautuksia ja niistä mahdollisesti ai-
heutuvia petoksia ei käsitellä tässä selvityksessä. Sen sijaan enakkoperintälain näkökul-
masta selvitys kattaa hyvin laajan osan verotuksen virheistä, joiden katsotaan liittyvän pime-
ään työhön. Verolajien näkökulmaa pimeään työhön on avattu tarkemmin selvityksen luvussa
5.2. Selvityksessä ei arvioida verovajeen kokonaismäärää Suomessa eikä sitä, kuinka mer-
kittävä osuus verovajeesta liittyy nimenomaan pimeään työhön.

Selvityksen ulkopuolelle on rajattu pimeän työn teettämisestä työnantajalle ja pimeän työn
tekemisestä työntekijälle itselleen koituvien seurausten käsitteleminen. Yleisesti voidaan to-
deta, että seuraukset vaihtelevat veronkorotuksista aina rikosoikeudellisiin sanktioihin ja va-
hingonkorvauksiin ollen ankarimmillaan ehdotonta vankeutta. Pimeästä työstä aiheutuvat
sanktiot ovat aihealueena laaja ja mahdollisesti erillisen selvityksen kohde.

¹⁴ Laki verotusmenettelystä 18.12.1998/1558, 29 §.

¹⁵ Harmaa talous & talousrikollisuus / Laajuus / Harmaan talouden selvityksiä / *Ulkomaalaisten alipalkkaustapauk-
sista oikeuteen vain jäävuoren huippu* 19.10.2020.

3 Pimeä työ tutkimuskirjallisuudessa

Tässä luvussa käsitellään pimeää työtä koskevia kotimaisia ja kansainvälisiä tutkimuksia ja selvityksiä. Tarkoituksena on muodostaa yleiskuva siitä, millaista tietoa pimeästä työstä on tarjolla erityisesti eurooppalaisessa kontekstissa. Pimeää työtä koskevia vertailuja vaikeuttavat merkittävästi käsitteiden eroavaisuudet ja tutkimusmenetelmien vakiintumattomuus. Systemaattisesti samanlaisena eri EU-maissa näytettäisiin toteutettavan lähinnä Eurobarometri-kyselytutkimus. Puutteistaan huolimatta sen perusteella voidaan varovaisena arviona esittää, että pimeän työn laajuus on Suomessa EU-alueen vähäisimmästä päästä. Lisäksi pimeän työn laajuutta EU-alueella on arvioitu EU:n komission tilaamassa selvityksessä¹⁶ LIM-menetelmän avulla.

3.1 Kotimainen tutkimus pimeästä työstä

Suomessa nimenomaisesti pimeää työtä on tutkittu varsin vähän. Pimeän työn euromääräistä laajuutta tai pimeää työtä tekevien henkilöiden määrää ei ole tiettävästi tutkittu lainkaan, tai ainakaan niistä ei ole saatavilla julkaistuja arvioita. 2010-luvulla pimeää työtä koskevia kyselytutkimuksia on toteutettu kolme.

Eurobarometri on laajin ja merkittävin kyselytutkimus, ja se käsittelee pimeän työn kysyntää, tarjontaa ja pimeään työhön liittyviä asenteita koko EU-alueella. Viimeisin toteutettiin vuonna 2018 ja tulokset julkaistiin 2019¹⁷. Tutkimukseen osallistui 27 565 vastaajaa kaikista 28 EU-maasta¹⁸. Suomalaisia kyselyyn osallistui 1 019. Tutkimus toteutettiin kasvokkain tehdyillä haastatteluilla ja vastaajille painotettiin kyselyn anonymiteettiä.

Kyselytutkimuksen keskeiset Suomea koskevat tulokset voidaan tiivistää seuraavasti, mukaan on myös vertailua naapurimaahamme Ruotsiin sekä Tanskaan ja Viroon:

Pimeän työn kysynnän näkökulma

Vastaajista 14 prosenttia sanoi ostaneensa pimeää työtä viimeisen 12 kuukauden aikana. Perusteena oli se, ettei saanut kuittia tai laskua¹⁹. Ruotsalaisten, tanskalaisten ja virolaisten vastaukset olivat hyvin lähellä suomalaisten vastauksia.

Yleisimmät toimialat, joissa pimeää työtä oli edellä mainituilla perusteilla esiintynyt, olivat kampaamo- tai kauneudenhoitopalvelut ja korjauspalvelut (esim. matkapuhelin, auto). Tältä osin tuloksia ei voi kuitenkaan pitää yleistettävänä vastaajajoukon pienuuden takia (alle 150 vastaajaa).

¹⁶ Euroopan komission tilaama selvitys on julkaistu vuonna 2017 ja on nimeltään *An evaluation of the scale of undeclared work in the European Union and its structural determinants: estimates using the Labour Input Method*. Raportin ovat kirjoittaneet Colin C. Williams, Predrag Bejakovic, Davor Mikulic, Josip Franic, Abbi Kedir ja Ioana A. Horodnic.

¹⁷ Vastaava kysely on toteutettu ensimmäisen kerran vuonna 2007 ja toisen kerran vuonna 2013.

¹⁸ Iso-Britannia oli kyselyn toteuttamishetkellä syyskuussa 2019 vielä EU:n jäsenvaltio.

¹⁹ Tältä osin kyselyn tulosta voidaan pitää melko epäluotettavana. Selvitysyksikön näkemyksen mukaan kuitittomuus ei tarkoita suoraviivaisesti pimeää työtä, mutta viittaa kylläkin harmaaseen talouteen.

Pimeän työn tarjonnan näkökulma

Vastaajista 26 prosenttia kertoi tuntevansa jonkun, joka työskentelee kokonaan tai osin pimeästi. Ruotsissa vastaava luku oli 44 prosenttia, Tanskassa 55 prosenttia ja Virossa 27 prosenttia.

Vastaajista kolme prosenttia kertoi työskennelleensä itse pimeästi joko omaan tai työnantajansa lukuun. Ruotsissa vastaava luku oli seitsemän prosenttia, Tanskassa kahdeksan prosenttia ja Virossa kuusi prosenttia.

Kyselyn vastaajista yksi prosentti kertoi, että työnantaja on maksanut palkkaa pimeästi käteisellä viimeisen vuoden aikana. Ruotsissa vastaava luku oli kaksi prosenttia, Tanskassa kolme prosenttia ja Virossa neljä prosenttia.

Asenteet pimeää työtä kohtaan

Kyselyyn vastanneet suomalaiset suhtautuivat huomattavan kielteisesti erilaisiin pimeän työn hyväksyttävyyttä kartoittaviin väittämiin. Esimerkiksi tapaukseen, jossa yritys palkkaa toisen yrityksen, joka ei ilmoita toimistaan vero- ja sosiaalivakuuttajille, suhtautui kielteisesti 93 prosenttia vastaajista.

Kaikkein hyväksyvimmin suhtauduttiin siihen, että kotitalous palkkaa yksityishenkilön, joka ei ilmoita saamastaan palkkiosta vero- tai sosiaalivakuuttajille, vaikka niin tulisi tehdä. Tapaukseen kielteisesti suhtautui 72 prosenttia vastaajista, sitä jokseenkin hyväksyttävänä piti 14 prosenttia ja hyväksyttävänä 11 prosenttia.

Suomalaiset ja ruotsalaiset suhtautuivat kaikkiin kyselyssä esitettyihin väittämiin kielteisemmin kuin tanskalaiset, mutta erot olivat alle 10 prosenttiyksikköä, pääosin vain muutamia prosenttiyksikköjä. Virolaiset suhtautuivat pimeään työhön paljon myönteisemmin, erojen ollessa 10–30 prosenttiyksikköä Suomeen ja Ruotsiin verrattuna. Suomalalaisten ja ruotsalaisten vastaajien vastausten erot olivat pääosin erittäin pieniä. (Euroopan komissio, 2020)

Edellä referoitu kyselytutkimus antaa kattavan kuvan pimeän työn kysynnästä, tarjonnasta ja siihen liittyvistä asenteista koko EU-alueella. Osin kyselyn tuloksiin on kuitenkin syytä suhtautua kriittisesti, sillä anonyymeista vastauksista huolimatta kaikki eivät välttämättä halua tai uskalla myöntää työskennelleensä pimeästi. Jo ensimmäisen pimeää työtä koskevan Eurobarometrin kohdalla erityisesti Etelä-Euroopan maita koskevia tuloksia ei voitu pitää kovin luotettavina (Harmaan talouden selvitysyksikkö, 2014).

Kolmen pohjoismaan ja Viron vertailussa kyselytutkimus antaa kuvan, jonka mukaan pimeän työn kysyntä on kaikissa neljässä maassa hyvinkin samansuuruisia. Kaikissa maissa noin 13–16 prosenttia vastaajista kertoi ostaneensa tai epäilevänsä ostoon liittyvän pimeää työtä. Suomessa huomattavan suuri osuus (35 prosenttia) vastaajista kertoi, että tajusi ostoon liittyvän pimeää työtä vasta jälkikäteen.

EU-barometrin perusteella pimeän työn tarjontaa koskien erot Suomen, Ruotsin, Tanskan ja Viron välillä ovat kysyntää suurempia. Suomi näyttäätyy kyselyn valossa maana, jossa pimeän työn tarjonta on vertailujoukon vähäisintä.

Asenneilmapiiriltään pimeään työhön kielteisimmin suhtautuvat suomalaiset ja ruotsalaiset, ja vain hivenen hyväksyvämmin asiaan suhtautuvat tanskalaiset. Asenteiltaan selvästi sallivimpia pimeää työtä kohtaan ovat virolaiset. Toisaalta he luottivat sekä vero- että työsuojeluviranomaisiin kaikkein eniten. Matalinta luottamus oli Tanskassa, jossa veroviranomaisiin luotti vain 41 prosenttia vastaajista.

Yhteenvedona voidaan todeta, että EU-barometrin perusteella pimeä työ ei näyttäyty Suomessa merkittävän suurena ilmiönä – verrattiinpa sitä sitten koko EU-tasoon taikka vain Viroon, Ruotsiin ja Tanskaan.

Huomattavasti EU-barometria suppeamman näkymän pimeään työhön tarjoaa tutkimusyhtiö T-Median kevättalvella 2017 toteuttama kyselytutkimus²⁰, jossa yhtiö kartoitti vastaajien näkemyksiä työn tekemisestä ja tarjoamisesta sekä muutamalla kysymyksellä myös pimeästä työstä. Peräti 20 prosenttia vastaajista ilmoitti tehneensä töitä pimeästi. Suurin pimeää työtä tehnyt vastaajajoukko oli 25–34-vuotiaiden ikäryhmä, pienin 55–64-vuotiaat vastaajat. (T-Media, 2017) Tämän selvityksen laatijoiden tiedossa ei ole, mistä T-median toteuttaman kyselytutkimuksen ja Eurobarometrin tulosten merkittävät erot johtuvat.

Toinen, myös EU-barometria selvästi rajatumpaan joukkoon suunnattu kysely toteutettiin vuonna 2016. Tällöin muun muassa niin sanottua kevytyrittäjyyttä tarjoava Ukko.fi toteutti käyttäjäkyselyn²¹, johon sisältyi useita kysymyksiä pimeään työhön liittyen. Vastaajista lähes puolet vastasi myöntävästi kysymykseen *onko sinua pyydetty tekemään pimeää työtä*. (Ukko.fi, 2016) Kysymykseen annettuja vastauksia ei voida yleistää siten, että ne kertoisivat pimeän työn kokonaiskysynnästä Suomessa. Tämä johtuu muun muassa kyselyyn vastanneiden henkilöiden rajautumisesta vain heihin, jotka käyttivät kyselyn teettäjän palveluita.

Yhteenvedona edellä esitellyistä kyselytutkimuksista voidaan todeta, että Eurobarometrin perusteella pimeä työ näyttäätyy Suomessa ilmiönä, joka on koko EU-tasoon vertailtuna kohutuullisen pieni. Toisaalta kahden muun toimijan toteuttamien kyselyiden perusteella pimeän työn voisi otaksua olevan Suomessa laajamittainen ongelma. Tutkimusmetodologia ja kattavuus huomioiden Eurobarometrin tuloksia voidaan pitää näistä kolmesta luotettavimpina. Tämän selvityksen tarkoituksena on selvittää, kuinka merkittävänä ongelmana pimeä työ näyttäätyy verotarkastustietojen perusteella.

3.2 Tutkimus pimeästä työstä EU-alueella

EU-tasolla pimeää työtä on käsitelty erityisesti niin sanotun EU:n pimeän työn vastaisen foorumin julkaisuissa sekä useissa Eurobarometri-kyselytutkimuksissa. Lisäksi pimeän työn laajuutta on arvioitu EU:n komission tilaamassa selvityksessä LIM-menetelmän avulla. Kyseisen raportin tuloksia on esitelty selvityksessä *Pimeän työn mittaaminen LIM-menetelmällä* –

²⁰ N= 1 222, ikäjakauma 18–65 vuotta, kyselyn virhemarginaali 2,8 prosenttiyksikköä.

²¹ N= 1 270.

yksityisen sektorin palkansaajien pimeän työn määrä. Tämän selvityksen innoittajana voidaan pitää ennen kaikkea Ruotsin Verohallinnon vuonna 2020 valmistunutta pimeän työn selvitystä²².

Ruotsin Verohallinnon selvityksessä pimeän työn määrää arvioitiin verotarkastustietojen avulla. Raportin mukaan maan kokonaisverovaje oli 10 prosenttia, josta puolet aiheutui pimeästä työstä. Selvityksen aineisto oli vuosilta 2010–2016, käsittäen 28 173 verotarkastusta, joista 21 810 koski palkkoja ja 6 363 yksinyrittäjiä ja avoimien yhtiöiden yhtiömiehiä. Selvityksessä arvioitiin, että pimeiden tulojen määrä oli kuusi prosenttia kaikista palkoista ja 2,3 prosenttia bruttokansantuotteesta. Selvityksessä hyödynnettiin pääosin riskiperusteisesti tehtyjä verotarkastuksia, joten on mahdollista, että saatu tulos yliarvioi pimeän työn määrää Ruotsissa. Ruotsin raportissa korostettiinkin useaan otteeseen, että tuloksia voi pitää lähinnä suuntaa antavina eikä absoluuttisina. Tuloksia verrattiin vuonna 2006 toteutettuun selvitykseen ja havaittiin, että pimeä työ ei ollut lisääntynyt vertailuajankohtien välillä. (Skatteverket, 2020)

Ruotsin pimeän työn selvityksen tuloksia voidaan verrata maan tuoreimpaan verovajeraporttiin *Skattefelrapport 2021*. Tässä raportissa luonnollisten henkilöiden ansiotuloveroa koskevat tulokset perustuvat satunnaisotannalla tehtyihin määrämuotoisiin verotarkastuksiin. Satunnaisverotarkastuksissa ei oletettavasti havaita kaikkia virheitä, mistä johtuen laskettuja tuloksia on korotettu tiettyjen vakiokertoimien avulla.²³ Raportissa arvioitiin, että luonnollisten henkilöiden ansiotuloveron vuosittainen vaje oli 12 miljardia kruunua eli 1,5 prosenttia maksettavaksi määräytyistä vastaavista veroista ajalla 2014–2018. Tämä kokonaisarvio sisältää yhtenä osana liiketoiminnan tulojen verovajeen, jonka suuruus oli 3,9 miljardia kruunua (24,6 % vastaavista veroista), sekä palkkatulojen verovajeen, jonka suuruus oli 5,3 miljardia kruunua (1,1 % vastaavista veroista).²⁴ (Skatteverket, 2022) Verovaje on käsitteenä laajempi kuin pimeät palkat, mutta verovajeraportin tulosten perusteella voidaan kuitenkin arvioida, että pimeän työn määrä Ruotsissa olisi pienempi kuin edellä esitetty kuusi prosenttia.

Yhtenä vertailutietona pimeän työn määrälle Ruotsissa voidaan pitää myös Eurobarometriä, jonka mukaan seitsemän prosenttia ruotsalaisista kyselyyn vastanneista sanoi työskennelleensä pimeästi viimeisen vuoden aikana. (Euroopan komissio, 2020) Tulosten vertailu on kuitenkin vaikeaa, koska esimerkiksi Eurobarometriin annetuista vastauksista ei voida kuitenkaan päätellä, että kyseisten henkilöiden kaikki tulot olisivat muodostuneet pimeistä työtu-loista.

Yksityiskohtaisimmin pimeän työn laajuutta viranomaisten näkökulmasta ja pimeän työn merkitystä EU-maissa on selvittänyt EU:n pimeän työn foorumi²⁵. Foorumilla on useita eri tehtäviä, kuten esimerkiksi tietojen ja parhaiden käytäntöjen vaihtaminen (Euroopan Unionin neuvosto, 2019). Pimeän työn foorumi on teettänyt maakohtaisen kyselyn, jossa on

²² Svarta arbetsinkomster: beräkningar baserade på Skatteverkets revisioner, Skatteverket 2020.

²³ Esimerkiksi elinkeinonharjoittajien korotuskerroin oli 1,17.

²⁴ Kolmannen osan kokonaisuudesta muodostavat palkkatulojen verosta tehdyt perusteettomat vähennykset. Tämän verolajin vaje rajautuu pimeän työn käsitteen ulkopuolelle.

²⁵ EU:n pimeän työn foorumi on yhteistyöelin, johon osallistuvat kaikki jäsenvaltiot. Foorumi ei ole päätöksenteoelin.

tiedusteltu monia pimeään työhön liittyviä asioita eri EU-jäsenmailta. Kysymykset koskevat esimerkiksi pimeän työn määritelmää, sen laajuutta ja keinoja pimeän työn torjumiseksi. (European Platform Undeclared Work, 2017)

Vain 11 maata kaikista 28 EU-maasta²⁶ ilmoitti pimeän työn foorumin kyselyssä arvion pimeän työn laajuudesta. Osa esitetyistä arvioista oli suhteutettu tehtyihin työn kokonaistuntimääriin, osa maista sen sijaan ilmoitti pimeän työn määrän suhteessa bruttokansantuotteeseen. Myös tietojen keräämistapa vaihteli paljon: osa vastauksista perustui kyselytutkimuksiin, osassa vastauksista tiedon lähde ei käynyt tarkasti ilmi. Pimeän työn esiintyvyydessä oli suuria maakohtaisia eroja. Esimerkiksi Slovenia esitti vastauksessaan, että pimeää työtä ei käytännössä esiinny lainkaan, kun taas Italian vastauksessa pimeän työn määrän arvioitiin olevan noin 15 prosenttia kokoaikaisesta työstä. Muut annetut vastaukset pimeän työn määrästä liikkuvat Tanskan kaikesta ilmoitetusta työstä arvioiman 1,6 prosentin ja Unkarin kokonaistyövoimasta ilmoittaman 10–17 prosentin välillä. (European Platform Undeclared Work, 2017) Kokonaisuutena EU:n pimeän työn foorumin koontiraportista saa kuvan, että pimeän työn laajuudesta on tietoa hyvin vaihtelevasti. Pimeän työn käsite myös sekoitetaan useissa vastauksissa laajempaan harmaan talouden käsitteeseen.

EU-maiden pimeän työn vastaisia strategioita on analysoinut tutkimuksessaan brittiläisen Sheffieldin yliopiston professori Colin C. Williams. Hän peräänkuuluttaa eri viranomaisten yhteisiä strategioita ja koordinoitua lähestymistapaa pimeän työn kitkemiseen. Williamsin mukaan tämä on välttämätöntä, koska useimmissa maissa pimeä työ liittyy kolmeen eri osa-alueeseen, joiden valvonnasta vastaavat eri viranomaiset. Näitä osa-alueita ovat verotus, sosiaaliturva ja työsuojelu. Yhtä lailla Williams jakaa torjuntakeinot kolmeen eri kategoriaan, joita ovat rangaistukset, kannustimet ja epäsuorat menettelyt. Kannustimilla tarkoitetaan keinoja, joilla lain säädöksiä noudattavasta työnteosta tehdään helpompaa. Epäsuorat menettelyt viittaavat keinoihin, joilla lisätään tietoisuutta pimeän työn haitoista ja kustannuksista, ja toisaalta parannetaan tietoisuutta lainmukaisen työnteon hyödyistä.

Ennen Williamsin toteuttamaa tutkimusta ei ollut tehty arviota siitä, missä määrin eri Euroopan Unionin maiden kansalliset hallitukset ovat omaksuneet yhdistetyn ja koordinoitun lähestymistavan pimeään työhön. Williams päätyy tutkimuksessaan johtopäätökseen, jonka mukaan ainoastaan neljäsosalla EU-maista on kokonaisvaltainen, eri hallinnonalojen rajat ylittävä strategia pimeän työn ehkäisemiseksi. Pääpaino torjuntakeinoissa on rangaistuksilla, ja jäsenmaat myös näkevät rangaistukset tehokkaimmaksi keinoksi pimeään työhön puuttamisessa. (Williams, 2019)

Edellä referoitujen lähteiden perusteella pimeän työn laajuudesta eri EU-maissa on saatavilla tietoa varsin vaihtelevasti. Kattavimmin tietoa tarjoaa kyselytutkimuksen keinoin toteutettu Eurobarometri, jota on referoitu edellä luvussa 3.1. Vaikuttaa kuitenkin selvältä, että yhteinäistämispyrkimyksistä huolimatta eri EU-maiden toimintatavat pimeän työn torjuntaan vaihtelevat varsin suuresti, joskin painopisteenä on perinteinen rangaistuksia korostava lähestymistapa.

²⁶ Nykyisin EU:ssa on 27 jäsenmaata, koska Yhdistynyt kuningaskunta erosi Euroopan unionista 31.1.2020.

4 Toteutetut hankkeet pimeä työn vähentämiseksi

Edellä luvussa kolme kerrottiin kotimaisista ja kansainvälisistä pimeää työtä koskevista tutkimuksista. Tässä luvussa listataan hyvin tiiviisti muutamia kotimaisia 2000-luvulla toteutettuja hankkeita, joiden konkreettisenä tavoitteena on ollut pimeän työn vähentäminen²⁷. Lisäksi esitellään lyhyesti Ruotsissa ja Norjassa käytössä olevaa henkilöstöpäiväkirjaa ja niihin liittyvää tutkimustietoa. Hankkeita analysoidaan tässä vain lyhyesti. Esiteltävät lait eivät muodosta tyhjentävää katsausta pimeän työn kitkemiseksi tehtäviin lainsäädäntöhankkeisiin, koska useilla harmaan talouden torjuntaa tehostavilla lakihankkeilla on välillisesti pimeää työtä ja pimeää palkanmaksua ehkäisevä vaikutus.

4.1 Suomessa toteutetut hankkeet

4.1.1 Kotitalousvähennys

Jo 1990-luvun loppupuolella osassa Suomea otettiin kokeiluluonteisesti käyttöön kotitalousvähennys. Kokeilua koskevassa hallituksen esityksessä todetaan muun muassa, että: ”kotitalouksien pimeästi teettämä työ lienee varsin tavallista ainakin tilapäisissä ja pienimuotoisissa työsuorituksissa. Veronsaajilta jäävät tulot näiltä osin saamatta.” Tällöin vielä kotitaloustyön väliaikaisen verotuen nimellä kulkenutta määräaikaisesti voimassa ollutta lakia perusteltiin sen tuottamalla hyödyillä, joita ovat lakiesityksen perusteluiden mukaan työttömyyskorvausten pieneneminen, verotulojen kasvu ja harmaan talouden vähentyminen. (Valtionvarainministeriö, 1997)

Vuoden 2001 alussa tuli voimaan tuloverolain muutos²⁸, jonka nojalla Suomessa otettiin pysyvästi käyttöön aiemmin kokeiluluontoisena voimassa ollut kotitalousvähennys. Hallituksen esityksessä uudenlaista verovähennystä perusteltiin enää vain hyvin niukasti harmaan talouden vähentymisellä:

”Kun kotitalous joutuu rahoittamaan maksamansa määrän verotetuilla varoilla, kotitaloudelle tulee taloudellisesti raskaaksi palkata vierasta työvoimaa kotona tehtävään työhön. Työllisyyspoliittisesti on kuitenkin tarkoituksenmukaista edistää kotityöpalveluiden kulutusta, mikä luo uusia työpaikkoja näitä palveluita tuottavalle pienyrityssektorille. Sosiaalipoliittisista syistä tulisi aktivoida koteja omaehtoisin hoivajärjestelyihin. Yhteiskunnan ja osapuolten kannalta on tärkeää, ettei kotityöhön liittyviä maksuvelvoitteita laiminlyödä.” (HE 140/2000 vp)

²⁷ Todettakoon, että lakihankkeiden lisäksi Suomessa on toteutettu erilaisia kampanjoita pimeän työn vähentämiseksi, esimerkiksi harmaa talous – musta tulevaisuus -projekti. EU-tasolla tietoisuutta pimeän työn vaikutuksista on viestitty #EU4FairWork-kampanjalla. Lisäksi pimeään työhön puututaan mm. iskutyypillisesti eri viranomaisten yhteistyönä toteutettavilla niin sanotuilla valvontaiskuilla, joihin osallistuu muun muassa Poliisin, Verohallinnon ja työsuojeluviranomaisen henkilökuntaa.

²⁸ Lakiin lisättiin 127 a §.

Edellä kerrotun mukaisesti kotitalousvähennyksen tarkoitus on lain perusteluiden mukaan ennen kaikkea luoda uusia työpaikkoja pienyrityssectorille. Sittemmin vähennyksen olemassaoloa on perusteltu sen harmaan talouden torjuntaa edistäväillä vaikutuksilla²⁹.

Valtion taloudellinen tutkimuskeskus on tarkastellut kotitalousvähennystä vuonna 2021 julkaistussa tutkimuksessa. Sen mukaan kotitalousvähennys ei lisää vähennykseen oikeuttavien palveluiden kysyntää merkittävästi, eikä siten vaikuta työllisyyteen kyseisillä aloilla. Tutkimuksessa ei myöskään havaittu kotitalousvähennyksellä olevan vaikutusta harmaaseen talouteen. (Harju, Jysmä, Koivisto & Kosonen, 2021) Voidaan kuitenkin todeta, että suhteessa kotitalousvähennyksen euromääräiseen laajuuteen on tuen harmaata taloutta torjuvista vaikutuksista hyvin vähän tutkimustietoa.

4.1.2 Tilaajavastuulaki

Vuoden 2007 alussa tuli voimaan niin sanottu tilaajavastuulaki, jonka tarkoituksena on muun muassa edistää työehtojen noudattamista ja parantaa lakisääteisten velvoitteiden hoitamista (HE 114/2006 vp). Lakia on sittemmin muutettu useaan otteeseen. Työ- ja elinkeinoministeriön (TEM) kolmikantainen työryhmä on tutkinut lain vaikutuksia vuosien 2012–2013 aikana. Lisäksi lakiin liittyviä opinnäytetöitä ja pro graduja on tehty useampiakin³⁰.

Edellä mainittu TEM:n työryhmän loppuraportti on julkaistu vuonna 2014. Selvityksen mukaan tilaajavastuulaki toimii hyvin rehellisten toimijoiden osalta, mutta ei juurikaan ole vaikuttanut ehkäisevän talousrikollisuutta. Työsuhteen vähimmäisehtojen noudattamisessa ja ulkomaalaisia yrityksiä koskevissa asioissa tilaajavastuulain merkitys on jäänyt vähäiseksi. Vähäisimmäksi lain merkitys arvioitiin palkkaukseen ja työehtoihin liittyvien velvoitteiden hoitamisen kannalta. TEM:n selvityksen mukaan yritykset ja asiantuntijat tunnustavat tilaajavastuulain merkityksen, mutta lähinnä symbolisella tasolla. Loppuraportin mukaan suurimpina haasteina ovat olleet lain valvontaan liittyvät ongelmat, viranomaisten tietojensaanti-oikeudet ja viranomaisyhteistyön ja valvonnan kehittämistarpeet. (Työ- ja elinkeinoministeriö, 2014) Loppuraportin julkistamisen jälkeen tilaajavastuulakia on muutettu ja esimerkiksi viranomaisten tietojensaanti-oikeuksia parannettu. Lain nykytilaa kuvaavaa kokonaisarvioita ei kuitenkaan ole saatavilla.

4.1.3 Rakennusalan lainsäädäntötoimet

Harmaan talouden torjuminen oli yksi pääministeri Jyrki Kataisen hallituksen painopistealueista 2010-luvun alkupuolella. Tällöin rakennustoimialan harmaan talouden ja pimeän työn torjuntaa tehostettiin lukuisilla uusilla laeilla. Pimeää työntekoa pyrittiin vähentämään muun muassa ottamalla käyttöön työmailla veronumero pakollisessa kuvallisessa tunniste-kortissa (HE 58/2011 vp). Lainmuutos toteutettiin vuonna 2011. Vuonna 2014 otettiin

²⁹ Esim. Harmaan talouden selvitysyksikkö 15.10.2019: lausunto eduskunnan valtiovarainvaliokunnalle, Keskukskauppa-kamari 25.4.2020: Palvelualaa elvytettävä koronaviruspandemian jälkeen: Kotitalousvähennyksen korvausastetta korotettava ja omavastuu puolitettava.

³⁰ Esim. Juntunen, Marjo, 2011: Tilaajavastuulaki ja sen valvonta – ongelmat erityisesti työehtosopimuksen noudattamisen valvonnassa, Laurea AMK, Mäkelä, Albert, 2014: Lähetetyt työntekijät ja tilaajavastuu, Helsingin yliopisto, Oikarinen, Satu, 2016: Tilaajavastuulain toteutuminen julkisissa hankinnoissa, Turun AMK.

käyttöön rakennusalalla urakka- ja työntekijätietojen ilmoittamisvelvollisuus. Myös tässä tavoitteena oli vähentää pimeiden palkkojen määrää. (HE 92/2012 vp)

Harmaan talouden selvitysyksikkö on tutkinut edellä kerrottujen, rakennusalaa koskeneiden lakihankkeiden vaikuttavuutta vuonna 2017 julkaistussa selvityksessä. Selvityksen perusteella rakentamisen toimialan lainsäädäntömuutoksista johtuva palkkasumman kasvu oli arviolta noin 300 miljoonaa euroa vuonna 2015. Palkkasumman kasvun arvioitiin selvityksessä kasvattaneen palkanmaksuun liittyviä verotuloja noin 100 miljoonaa euroa. Edelleen selvityksessä todetaan, että urakkatietojen ilmoittamisvelvollisuus on kasvattanut arvonlisäveron tuottoa varovaisen arvion mukaan noin 20 miljoonaa euroa. Lisäksi selvityksen mukaan rakentamisen laajennettu tiedonantovelvollisuus on kasvattanut verotuottoa yhteensä 34–35 miljoonaa euroa vuodessa. (Harmaan talouden selvitysyksikkö, 2017)

4.1.4 Palkka pankkiin -lainsäädäntö

Vuonna 2013 toteutettiin työsopimuslain muutos, jonka jälkeen palkan maksamisen pääsäännöksi tuli maksaminen pankkitilille. Muutoksen keskeisenä perusteluna oli tavoite vähentää käteisen rahan käyttöä ja sitä kautta vaikeuttaa pimeiden palkkojen maksua. (HE 24/2013 vp) Käteisen rahan käytön kieltämistä palkan maksamisessa voitaneen pitää lähinnä symbolisena toimenpiteenä tilanteessa, jossa palkanmaksu muutenkin tapahtui – jo ennen lakimuutosta – valtaosin tilisiirtoina.

4.1.5 Tulorekisterilainsäädäntö

Vuonna 2019 otettiin käyttöön tulorekisteri, joka on kansallinen sähköinen tietokanta, jonne tallennetaan palkka-, etuus- ja eläketietoja yksilötasolla. Tulorekisterin arvioitiin mahdollistavan Verohallinnolle aiempaa reaaliaikaisemman verovalvonnan ja entistä tehokkaamman harmaan talouden torjunnan (HE 134/2017 vp). Hallituksen esityksessä ei mainittu erikseen pimeää työtä, mutta koska tulorekisteriin ilmoitetaan nimenomaisesti henkilöiden tulotiedot, liittyy rekisterin mahdollistama aiempaa tehokkaampi harmaan talouden torjunta eritoten pimeään työhön.

Valtiontalouden tarkastusvirasto (VTV) on arvioinut tulorekisterin käyttöönottoa ja vaikutuksia tuloksellisuustarkastuksessaan vuonna 2021. Tarkastuskertomuksen mukaan tulorekisterin ansiosta harmaaseen talouteen voidaan puuttua nopeammin kuin ennen, mutta rekisterin merkittävänä puutteena on kuitenkin se, ettei tieto yhden käyttäjän havaitsemasta väärästä tiedosta tulorekisterissä välity muille tiedon käyttäjille. Lisäksi tarkastuskertomuksessa todetaan, että tulorekisteri on tuonut uusia mahdollisuuksia väärinkäyttöksiin. Konkreettisesti kyse on esimerkiksi siitä, että jos sekä tulonsaaja että maksaja toimivat tietoisesti yhdessä väärin, virheellinen ilmoitus jää voimaan eikä tietoa virheellisyydestä saada merkityksi tulorekisteriin. Väärinkäyttötilanteissa on havaittu myös ongelmia ilmoittajan tavoittamisessa. (Valtiontalouden tarkastusvirasto, 2021)

4.1.6 Ulkomaiseen työvoimaan kohdistuneet hankkeet

Ulkomaisen työvoiman valvontaan on kohdistettu hyvin runsas määrä erilaisia lainsäädäntömuutoksia, joiden tarkoituksena on ollut ainakin epäsuorasti vähentää pimeän työn määrää³¹. Tällaisia säädöksiä ovat olleet esimerkiksi vuokratyövoiman ilmoitusvelvollisuus³² sekä lähetettyjen työntekijöiden edustajan asettamisvelvollisuus ja työntekijöiden ennakoilmoitukset.

Valvonnan keskeisenä haasteena on se, että vaikka jälkikäteen tapahtuva valvonta saattaa paljastaa laiminlyöntejä, on toimenpiteiden vaikuttavuus lähes olematon silloin, kun yritys on lopettanut toimintansa tai poistunut Suomesta. Vaatimusta lähetettyjen työntekijöiden työnantajan Suomessa olevasta edustajasta ja hänen asiakirjojen säilytys- ja selvitysvelvollisuudesta on noudatettu huonosti. 2010-luvun alkupuolella vain 41 prosenttia tarkastetuista kohteista oli täyttänyt edustajan asettamiseen ja tietojen säilyttämiseen liittyvät velvollisuutensa. (Hirvonen, 2012) Vuoden 2020 tietojen mukaan tilanne on edelleen samanlainen (Aluehallintovirasto, 2021).

4.1.7 Pimeän työn torjunta harmaan talouden torjuntaohjelmissa

Vuosien 2010–2011 ja 2012–2015 harmaan talouden ja talousrikollisuuden torjuntaohjelmissa pimeä työ esiintyy mainintana harmaan talouden yleistilanteen kuvauksissa, mutta ei erillisissä hankkeissa. Aikaisempien vuosien (1996–2009) torjuntaohjelmissa pimeää työtä ei ole mainittu lainkaan. Ensimmäisen kerran pimeä työ ja sen torjunta on esiintynyt omana erillisenä hankkeena vuosien 2016–2020 toimenpideohjelmissa. Tällöin yhtenä kärkihankkeena (2.2) oli pimeän työn torjunta ja kansallisen sekä kansainvälisen yhteistyön edistäminen³³.

Selvitystä laadittaessa käynnissä olleessa vuosien 2021–2023 harmaan talouden torjuntaohjelmassa³⁴ ja sen täydennyksessä³⁵ pimeän työn torjunta on mainittu useassa eri hankkeessa³⁶. Työ- ja elinkeinoministeriö on asettanut pimeän työn vastaista yhteistyötä edistävän kansallisen yhteistyöverkoston 6.4.2020–31.12.2023 väliseksi ajaksi. Lisäksi työ- ja elinkeinoministeriössä on vireillä lukuisia työntekoon ja työllisyyteen liittyviä lakihankkeita. Näissä ei kuitenkaan otsikkotasaisen tiedon mukaan ole kyse suoranaisesti Suomessa tapahtuvan pimeän työn torjumisesta.

³¹ Aihealuetta on selvitetty laajasti valtionvarainministeriön Ulkomaisten työntekijöiden ja yritysten verovalvontaa koskevassa selvityksessä vuonna 2013. Työryhmä esitti yhteensä 14 toimenpidettä, joista vain osa koski ainoastaan ulkomaista työvoimaa. Osa toimenpide-ehdotuksista on toteutettu.

³² Voimassa vuodesta 2007 alkaen.

³³ Toimenpideohjelmissa on kehitetty viranomaisten toimivaltuuksia mm. tietojenvaihtosäännöksiä kehittämällä ja digitaalisia palveluita kuten velvoitteidenhoitoselvityspalvelua tietojen saatavuuden edistämiseksi. Tämän seurauksena eri viranomaisten kyvykyys tunnistaa ja kohdistaa erilaisia toimenpiteitä riskiperusteisesti on kehittynyt. Tämä koskee myös pimeän työn torjuntaa. Myös viranomaisyhteistyömallit ovat kehittyneet.

³⁴ 2021–2023 torjuntaohjelma on nimeltään ”Valtionneuvoston periaatepäätös kansalliseksi harmaan talouden ja talousrikollisuuden torjunnan strategiaksi ja toimenpideohjelmaksi vuosille 2021–2023”. Aikaisemmat ovat valtionneuvoston periaatepäätökset olleet nimeltään vaihtelevasti joko talousrikollisuuden ja harmaan talouden torjuntaohjelmia tai strategioita.

³⁵ Vuoden 2021–2023 torjuntaohjelman 4-luvun hankkeita on täydennetty 10.9.2020 muutamilla lisäyksillä.

³⁶ Pimeän työn torjuntaan liittyviä hankkeita ovat esimerkiksi:

1.5.b Selvitetään palkan ja yritystulon rajanvetotilanteita ja sääntelyn selkeyttämistarpeita

1.4: Selvitetään keinoja ja sanktioita puuttua tahalliseen ja törkeään huolimattomaan alipalkkaukseen

2.1: Veronumerot telakoille

4.1: Tehostetaan viranomaisyhteistyötä ja lisätään valvonnan vaikuttavuutta rikosten ja pimeän työn torjumiseksi.

4.1.8 Yhteenvetoa ja analyysiä kotimaisista hankkeista

Edellä on käsitelty lyhyesti keskeisimpiä 2000-luvulla pimeän työn torjumiseksi toteutettuja lakihankkeita. Huomionarvoista on, että aiempina vuosikymmeninä pimeän työn estämiseksi toteutettujen hankkeiden vaikutusarvioita on tehty vain vähän. Voidaankin ajatella, että ennen uusien pimeää työtä mahdollisesti vähentävien toimien toteuttamista, olisi tärkeää tarkastella nykyisen lainsäädännön toimivuutta ja tehokkuutta sekä mahdollisia muutostarpeita.

Pimeä työ esiintyy mainintana monissa harmaan talouden ja talousrikollisuuden torjuntaan liittyvissä kuvauksissa, lakien valmistelun esitöissä ja hankkeissa. Vaikuttaa kuitenkin siltä, että nimenomaisesti pimeän työn torjuntaan on kohdistettu lähinnä pistemäisiä toimia, joilla yksittäisiä lieveilmiöitä pyritään kitkemään.

4.2 Norjan ja Ruotsin ns. henkilöstöpäiväkirjan pitämismvelvollisuus

Ruotsissa on käytössä tietyillä toimialoilla niin sanottu henkilöstöpäiväkirja. Sillä tarkoitetaan työnantajan velvollisuutta pitää työssä olevasta henkilökunnasta jatkuvaa päiväkohtaista lueteloa, jonka tiedot veroviranomainen voi tarkistaa. (Harmaan talouden selvitysyksikkö, 2011). Suomessa vastaava velvollisuus on osittain olemassa työaikalain perusteella (Työ- ja elinkeinoministeriö, 2012). Ruotsin järjestelmää vastaava henkilöstöpäiväkirjan pitämismvelvollisuus on myös Norjassa. Siellä tehty satunnaistettu kenttäkoe osoitti, että henkilöstöpäiväkirjaan kohdistuva valvonta kasvatti merkittävästi ilmoitettujen palkkojen määrä tarkastuksen kohteeksi joutuneissa yrityksissä verrokkijoukkoon nähden (Bjørneby; Alstadsæter; & Telle, 2018). Suomessa henkilöstöpäiväkirjan pitämismvelvollisuudesta ei ole säädetty erikseen eikä asiaa ole tarkemmin selvitetty.

Ruotsissa valmistui vuonna 2019 laaja, maan eduskunnan teettämä henkilöstöpäiväkirjaa koskeva selvitys. Sen tarkoituksena oli selvittää, kuinka eri sidosryhmät ovat kokeneet lainsäädännön toimivuuden ja hyödyllisyyden. Selvityksen perusteella henkilöstöpäiväkirjan hyödyllisyys ja toimivuus koetaan hyvin eri tavoin eri toimialoilla. Tutkimus päättyi johtopäätökseen, että lainsäädännön toimivuus voidaan kokonaan kyseenalaistaa, tai ainakin lainsäädäntö kaipaa muutoksia. (Riksdagstryckeriet, 2019)

Ruotsin ja Norjan kokemukset henkilöstöpäiväkirjan toimivuudesta pimeän työn estäjänä ovat edellä kerrottujen tietojen perusteella jossain määrin ristiriitaisia.

5 Pimeän työn ja palkan käsitteet

Selvityksessä pimeän työn ja palkan käsitteitä käytetään synonyymeina. Tässä luvussa esitellään selvityksessä käytettyä pimeän työn ja palkan käsitettä sekä hahmotellaan pimeän palkan maksamisen merkitystä osana valtion verotulojen kokonaisuutta. Lisäksi luvussa kuvaillaan pimeän palkanmaksun erilaisia muotoja ja käydään yksityiskohtaisesti läpi palkanmaksuun liittyvät vero-, eläke- ja sosiaalivakuutusmaksuvelvoitteet.

5.1 Selvityksessä tarkoitettu pimeä työ ja palkka

Pimeällä palkalla tarkoitetaan tilannetta, jossa tehdystä työstä maksetaan palkkaa, mutta palkan maksaja jättää ilmoittamatta viranomaisille kokonaan tai osittain työntekijälle maksetun palkan eikä hoida palkkoihin liittyviä työnantajan maksuvelvoitteita (Ansamäki, 2017). Käytännössä pimeän palkan tilanteissa työnantaja välttyy tilittämästä ennakonpidätyksiä ja sairaskorvausmaksuja sekä jättää maksamatta työeläke- ja tapaturmavakuutusmaksut. (Lith, 2010) Työntekijöille maksettujen pimeiden palkkojen lisäksi tässä selvityksessä pimeällä palkalla tarkoitetaan myös osakeyhtiön osakkaan yrityksestä nostamia varoja ja rahanarvoisia etuuksia, jotka olisi oikein meneteltäessä tulleet käsitellä saajan veronalaisena palkkana tai osinkona, mutta näin ei ole toimittu.

Tässä selvityksessä pimeällä palkalla tarkoitetaan kaikkia ennakkoperintälain mukaan palkaksi luettavia suorituksia ja etuuksia, joita ei ole käsitelty palkkana. Pimeisiin palkkoihin katsotaan kuuluvaksi myös esimerkiksi työntekijälle maksetut verovapaat kustannusten korvaukset, jotka eivät ole täyttäneet verovapauden edellytyksiä. Selvitys kattaa myös tilanteet, joissa työntekijälle annetuista rahanarvoisista luontoiseduista on jätetty asianmukaiset ilmoitukset antamatta tai luontoisedun arvo on ilmoitettu liian alhaisena. Pimeää palkkaa voivat olla niin myös sanotuissa kuittikauppatilanteissa esimerkiksi rahannostajien palkkiot ja kuitin kirjoituspalkkiot. Osakkaiden palkkaan liittyviä virheitä käsitellään verotuksessa toisinaan peiteltynä osinkona, toisinaan ennakkoperinnän alaisina palkkoina, mutta molemmat sisältyvät tämän selvityksen määritelmään pimeästä palkasta. Selvityksessä ei pyritä kuvailemaan yksityiskohtaisesti kaikkia erilaisia pimeän palkan maksamisen toteuttamistapoja.

Edellä kerrotuista syistä pimeän työn ja palkan käsite on tässä selvityksessä huomattavan laaja. Laajaan käsitteeseen päädyttiin, jotta tarkasteluun saatiin mukaan myös osakasasemassa tehdyt veronvälttelytoimet, jotka ovat olennainen osa pimeän työn ilmiötä.

Täysin pimeästi eli ilman rekistereihin hakeutumista toimineiden yritysten ja samalla pimeästi palkkoja maksavia tapauksia on mukana rajatusti. Tällaisia tapauksia ovat ne, jotka ovat verotarkastuksilla todettu, ja yritykset on viranomaisaloitteisesti rekisteröity Verohallinnon rekistereihin ja palkanmaksuun liittyvät verot on määrätty maksettavaksi. Selvityksen tarkkuustaso ulkomaisten työntekijöiden pimeästä työstä Suomessa on kotimaista pimeää työtä heikompi, koska ulkomaiseen työvoimaan on kohdistettu huomattavasti vähäisempi määrä verotarkastuksia.

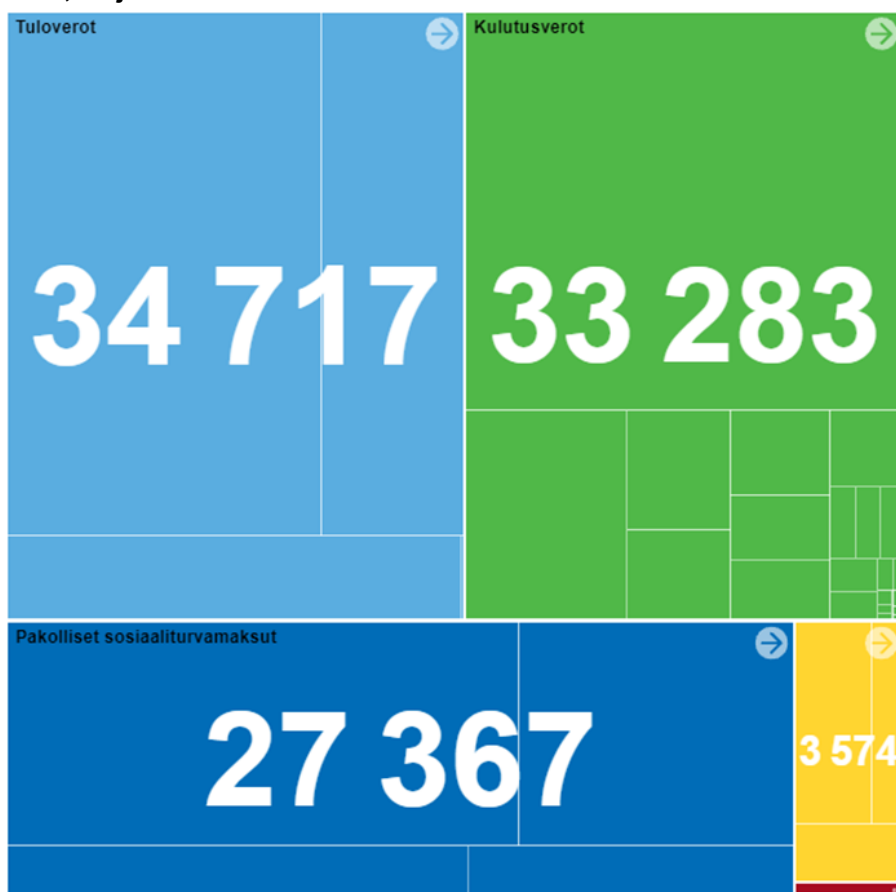
Seuraavaksi kuvataan lyhyesti pimeää työtä harmaan talouden osa-alueena.

5.2 Pimeä työ ja palkka harmaan talouden osa-alueena

Pimeän palkan tilanteissa maksajayrityksessä tehdään usein samanaikaisesti monia erilaisia väärinkäytöksiä, joilla on vaikutusta eri verolajien³⁷ verojen määrään. Käytännössä pimeiden palkkojen maksamiseen irrotetaan varoja usein siten, että osa myyntituloista salataan kokonaan. Näin ollen useiden eri verolajien veroja jää maksamatta. Toisinaan rahavaroja irrotetaan liiketoiminnasta siten, että kirjanpitoon laaditaan tekaistuja kulutositteita, joita vastaan varoja nostetaan pimeiden palkkojen maksamiseen. Tekaistut kulut aiheuttavat usein perusteettomia arvonlisäveron vähennyksiä. Tässä selvityksessä ei huomioida muita kuin ennakkoperinnän virheitä, jotta selvityksen fokus pysyy riittävän tarkasti rajattuna. Siten *pimeä työ* ei ole käsitteenä synonyymi *harmaalle taloudelle*. Tässä selvityksessä tarkasteltu pimeä työ on huomattavasti rajatumpi käsite.

Harmaan talouden potentiaalista laajuutta voidaan luontevimmin tarkastella eri verolajien ja sosiaalivakuutusmaksujen näkökulmista. Vuonna 2020 eri verolajien verokertymät olivat seuraavan kuvion mukaiset³⁸, luvut miljoonaa euroa:

Kuvio 1. Suomen valtion ja kuntien verotulot ja pakolliset sosiaalivakuutusmaksut vuonna 2020, miljoonaa euroa.



Lähde: Tilastokeskus, Veronmaksajain Keskusliitto

³⁷ Esimerkiksi yritysten ja henkilöiden tulovero ja arvonlisäverotus.

³⁸ Kuvion lähde: Veronmaksajain keskusliitto 2021 (luettu 19.11.2021).

Kuvion tarkempi analyysi Verohallinnon tilastotietokannan avulla kertoo, että tuloveroista lähes 30 miljardia koostuu kunnallisveroista ja valtion ansio- ja pääomatuloveroista. Näistä noin 65 prosenttia eli noin 21 miljardia kertyy palkansaajien tuloveroista³⁹. Sosiaalivakuutusmaksut, määrältään yli 27 miljardia euroa, liittyvät kiinteästi työntekoon, koostuen työnantajien ja -tekijöiden eläkemaksuista sekä muun muassa työttömyysvakuutus- ja sairausvakuutusmaksuista. Näin ollen työntekoon olennaisesti liittyvien tuloverojen ja sosiaalivakuutusmaksujen yhteismäärä on noin 48 miljardia euroa, joka vastaa miltei puolta Suomen valtion ja kuntien kokonaisverokertymästä⁴⁰. Siten pimeän työn euromääräinen potentiaali voi olla varsin suuri. Tämän selvityksen ydin on arvioida pimeän työn euromäärää Suomessa.

5.2.1 Pimeä palkka maksajan kannalta sekä vältetyt maksut ja verot

Pimeästi palkkaa maksava työnantaja ei ilmoita tulorekisteriin työntekijälle maksettua palkkaa eikä maksa valtiolle ennakonpidätyksiä ja palkkaan liittyviä sosiaalivakuutusmaksuja. Työnantajalla on velvollisuus pidättää työntekijän palkasta verokortin mukainen ennakonpidätys ja tilittää se Verohallinnolle. Lisäksi palkasta tulee maksaa sosiaalivakuutusmaksut, jotka olivat vuonna 2021 seuraavat:

- työntekijän työeläkevakuutusmaksu, 17–52 ja 63–68-vuotiaalla 7,15 prosenttia palkasta ja 53–62-vuotiaalla 8,65 prosenttia palkasta
- työntekijän työttömyysvakuutusmaksu 1,4 prosenttia palkasta
- työnantajan eläkevakuutusmaksu 16,95 prosenttia palkasta
- työnantajan työttömyysvakuutusmaksu 0,50 prosenttia palkasta ja jos palkkoja on maksettu yli 2,126 miljoonaa euroa, ylimenevältä osalta 1,9 prosenttia
- työnantajan sairausvakuutusmaksu 1,53 prosenttia palkasta
- tapaturma- ja henkivakuutusmaksu (vaihtelee toimiala- ja yrityskohtaisesti).

Palkan sivukulujen yhteismäärä on vuonna 2021 edellisen listauksen mukaisesti 27,53–30,43 prosenttia palkkojen bruttomäärästä.

Tapausesimerkki: pimeä palkka ja työnantajan laiminlyönnit

Oy Koodausfirma Ab maksaa palkan täysin pimeästi 32-vuotiaalle palkansaajalle Erkki Cobolille vuonna 2021. Vuodessa palkkaa maksetaan 40 000 euroa. Erkki Cobolilla ei ole Verohallinnon tietojen mukaan lainkaan verotettavia ansiotuloja, koska hän on saanut kaiken palkan pimeästi.

Oy Koodausfirma Ab:n olisi tullut maksaa palkasta työntekijän työeläkevakuutusmaksua (7,15 prosenttia + 16,95 prosenttia) ja sairausvakuutusmaksua (1,53 prosenttia) sekä tapaturmavakuutusmaksua (0,7 prosenttia). Lisäksi työnantajan tulee maksaa työttömyysvakuutusmaksua (0,50 prosenttia), joka perustuu yrityksen vuoden aikana maksamiin palkkoihin.

³⁹ Verohallinnon tilastotietokannan tilaston *Verot ja maksut tulonsaajaryhmittäin ja tuloluokittain* perusteella, vuoden 2019 tietojen mukaan henkilöasiakkaiden tuloverot jakoutuivat tulonsaajaryhmittäin seuraavasti: palkansaajat 65,1 prosenttia, eläkkeiden saajat 21,8 prosenttia, pääomatulojen saajat 6,5 prosenttia, elinkeinonharjoittajat 2,4 prosenttia ja muut 4,2 prosenttia.

⁴⁰ On kuitenkin huomioitava, että julkisen sektorin ollessa Suomessa kansainvälisesti vertailtuna suuri (Veronmaksajain keskusliitto, 2021), ovat kunnat ja valtio merkittäviä palkkojen maksajia. Tältä osin palkkoihin ei liity harmaan talouden riskiä.

Työnantajan työttömyysvakuutusmaksu on 0,50 prosenttia palkasta, jos yrityksen kokonaispalkkasumma on korkeintaan 2 125 500 euroa. Sen ylittävältä osalta työttömyysvakuutusmaksu on 1,90 prosenttia palkasta. Mikäli Oy Koodausfirma Ab olisi maksanut palkkaan liittyvät maksut, sille olisi tullut maksettavaa seuraavasti:

Taulukko 2. Esimerkki 40 000 euron vuosipalkkaan liittyvistä sosiaalivakuutusmaksuista, jotka työnantajan tulisi maksaa.

Maksu tai vero	Prosentti	Euroa	Selite
TyEL (työnantajan osuus)	16,95 %	6 780 €	Työnantaja maksaa
Sairasvakuutusmaksu	1,53 %	612 €	Työnantaja maksaa
Tapaturmavakuutusmaksu	0,70 %	280 €	Työnantaja maksaa
Työttömyysvakuutusmaksu	0,50 %	200 €	Työnantaja maksaa
Yhteensä	19,68 %	7 272 €	

Esimerkin tapauksessa yritys on pimeällä palkanmaksulla välttynyt maksamasta valtiolle sosiaalivakuutusmaksuja yhteensä lähes 7 300 euroa. Lisäksi yrityksen olisi tullut pitää työntekijälle maksetusta rahapalkasta ennakonpidätys ja työntekijän osuus sosiaalivakuutusmaksuista. Työntekijän osuus on kuvattu seuraavassa luvun 5.2.2 esimerkissä.

5.2.2 Pimeä palkka työntekijän ja yhteiskunnan kannalta

Työntekijälle pimeän työn vaikutukset ovat huomattavan laajat. Seurauksia voi aiheuttaa myös kirjallisen työsopimuksen puuttumisesta eikä pelkästään pimeästä palkasta. Ilman työsopimusta työskentelevän henkilön asema on kiistatilanteissa usein heikko. Pimeän palkan maksaja ei välttämättä maksa ylityökorvauksia ja muita lisiä, koska pimeästä palkasta harvoin pidetään asianmukaista palkkakirjanpitoa, josta työtunnit kävisivät ilmi. Usein pimeää palkkaa saavalla työntekijällä ei ole mahdollisuutta työterveyshuoltoon eikä häntä ole vakuutettu mahdollisten työtapaturmien varalta. Pimeästä palkasta ei myöskään kerry ansiosidonnaisia työttömyysturvaa eikä minkäänlaista eläketurvaa.

Toisinaan työnantaja ja työntekijä sopivat yhdessä, että osa palkasta maksetaan pimeästi ja osan korvauksesta työntekijä hakee avustuksina ja tukina valtiolta. Näissä tilanteissa ei pelkästään jätetä maksamatta veroja ja sosiaalivakuutusmaksuja, vaan käytetään väärin yhteiskunnan varoja, jotka on kerätty muilta veronmaksajilta.

Tapausesimerkki: pimeä palkanmaksu työntekijän näkökulmasta

Oy Koodausfirma Ab maksaa palkan täysin pimeästi 32-vuotiaalle palkansaajalle Erkki Cobolille vuonna 2021. Vuodessa palkkaa maksetaan 40 000 euroa. Erkki Cobolilla ei ole Verohallinnon tietojen mukaan lainkaan verotettavia ansiotuloja, koska hän on saanut kaiken palkan pimeästi. Progressiivisen veroasteikon mukaisesti 40 000 euroa vuodessa ansaitsevan palkansaajan veroprosentti olisi ollut noin 20 prosenttia. Veroprosentti voi tosiasiallisesti kuitenkin vaihdella huomattavasti kunnallisveroprosentin ja mahdollisten verovähennysten suuruuden mukaan.

Mikäli Oy Koodausfirma Ab toimisi lakien ja asetusten mukaisesti, se pidättäisi työntekijän palkasta verokortin mukaisen ennakonpidätyksen 20 prosenttia sekä työntekijän osuuden työeläkevakuutusmaksusta 7,15 prosenttia ja työttömyysvakuutusmaksusta 1,40 prosenttia ja tilittäisi ne valtiolle ja eläkeyhtiölle työntekijän puolesta. Työntekijälle maksettaisiin rahapalkkana se määrä, joka jää jäljelle lakisääteisten pidätysten jälkeen.

Taulukko 3. Esimerkki 40 000 euron vuosipalkkaan liittyvistä sosiaalivakuutusmaksuista, jotka työnantajan tulisi maksaa työntekijän puolesta.

Maksu tai vero	Prosentti	Euroa	Selite
Ennakonpidätys	20,00 %	8 000 €	Maksaja tilittää työntekijän puolesta
TyEI (työntekijän osuus)	7,15 %	2 860 €	Maksaja tilittää työntekijän puolesta
Työttömyysvakuutusmaksu	1,40 %	560 €	Maksaja tilittää työntekijän puolesta
Yhteensä	28,55 %	11 420 €	

Esimerkin tapauksessa pimeän palkan vuoksi valtiolta on jäänyt saamatta viime kädessä työntekijän maksettavaksi kuuluvia maksuja yhteensä noin 11 400 euroa. Palkanmaksajan velvollisuus on pidättää työntekijän maksettavaksi kuuluvat maksut työntekijän palkasta ja tilittää ne työntekijän puolesta. Palkansaajan on tarkistettava esitäytetyssä veroilmoituksessa olevat tiedot. Jos tiedoissa on puutteita tai virheitä, palkansaajan on annettava veroilmoitus ilmoittamalla puuttuvat ja korjatut tiedot Verohallinnolle.

Mikäli esimerkin työntekijä olisi pimeän palkan lisäksi nostanut työttömyyskorvauksia, olisivat valtion tulojen menetykset olleet vielä suuremmat. Tällöin ei pelkästään veroja olisi jäänyt kertymättä valtion kassaan, vaan lisäksi yhteiskunta olisi menettänyt varoja perusteettomien työttömyyskorvausten muodossa. Räikeimmissä tapauksissa pimeän palkan ja perusteettomien työttömyyskorvausten lisäksi maksetaan myös väärin perustein asumistukia ja toimeentulotukia. Niitä maksetaan pienituloisuuden tai työttömyyden vuoksi henkilöille, joiden verotettavat ansiotulot ovat vähäiset tai niitä ei ole.

5.2.3 Yhteenveto pimeästä palkasta aiheutuvasta ennakonpidätysten ja sosiaalivakuutusmaksujen vajeesta

Edellä luvuissa 5.2.1 ja 5.2.2 on havainnollistettu esimerkein pimeästä palkasta aiheutuvaa ennakonpidätysten ja sosiaalivakuutusmaksujen vajeesta. Työnantajan tulee maksaa valtiolle työnantajan osuus palkkaan liittyvistä sosiaalivakuutusmaksuista sekä lisäksi tilittää ennakonpidätykset ja työntekijän osuus sosiaalivakuutusmaksuista. Esimerkkien mukaisesta 40 000 euron suuruisesta bruttopalkasta ennakonpidätyksiä ja sosiaalivakuutusmaksuja tulisi tilittää valtiolle yhteensä lähes 19 000 euroa. Pimeästi maksettujen palkkojen vuoksi valtiolle jää kertymättä huomattavat vero- ja maksukertymät.

5.3 Yleisimmät pimeän työn ja palkan muodot

Verotuksessa sovellettavan palkan käsite määritellään ennakkoperintälaissa⁴¹. Kyseisen lainkohdan mukaan palkalla tarkoitetaan kaikenlaatuista palkkaa, palkkiota, etuutta ja korvausta, joka saadaan työ- tai virkasuhteessa. Pimeäksi palkaksi kutsutaan tässä selvityksessä myös sellaisia työntekijöille maksettuja suorituksia, jotka on käsitelty verovapaina suorituksina, vaikka verovapauden edellytykset eivät ole täyttyneet.

5.3.1 Täysin pimeä työ ja palkka

Palkat, joista ei makseta lainkaan veroja, ovat pimeän palkan muodoista räikein. Tällaista palkanmaksua esiintyy esimerkiksi tilanteissa, jossa työntekijälle maksetaan palkkaa, mutta palkanmaksaja ei ole hakeutunut työnantajarekisteriin eikä ilmoita lainkaan palkanmaksua.

⁴¹ Ennakkoperintälaki 20.12.1996/1118, 13 § 1 mom.

Toisinaan on tilanteita, joissa palkanmaksaja on rekisteröitynyt työnantajaksi, mutta jättää kokonaan ilmoittamatta joidenkin työntekijöiden palkat⁴².

Pimeän palkan tilanteissa maksajayritys välttyy maksamasta palkkoihin liittyvät maksut eikä myöskään palkansaaja maksa verojaan. Toisinaan on tilanteita, joissa palkansaaja ilmoittaa omassa verotuksessaan palkan, jota maksaja ei ole ilmoittanut. Tällaisissa tilanteissa verot määrätään maksettavaksi jälkikäteen palkanmaksajalle. Kyse on silloin lähes poikkeuksetta tilanteista, joissa työntekijä on tietämätön työnantajan laiminlyönneistä.

5.3.2 Osittain pimeä työ ja palkka

Osittain pimeästi maksettavalla palkalla tarkoitetaan tilannetta, jossa palkanmaksaja ilmoittaa osan työntekijälle maksamastaan palkasta asianmukaisesti ja ilmoittaa sekä maksaa myös verot ja muut maksut. Palkanmaksaja maksaa kuitenkin työntekijälle lisäksi myös palkkaa, josta verot jätetään maksamatta ja palkka ilmoittamatta Verohallinnolle. Palkkaa siis ilmoitetaan, mutta liian pienenä.

5.3.3 Veronalainen palkka verovapaina kustannusten korvauksina

Työntekijälle maksetut suoritukset, jotka eivät täytä verottomuuden edellytyksiä ovat veronalaista ansiotuloa. Tässä selvityksessä nämä työntekijälle perusteettomasti verovapaina maksetut korvaukset on katsottu olevan pimeää palkkaa.

Työnantaja voi tietyin edellytyksin maksaa työntekijälle työmatkoista aiheutuneista kuluista korvausta verovapaasti. Matkakustannuksia ovat Suomessa tai ulkomailla tehdystä työmatkasta aiheutuneet kustannukset. Verohallinto julkaisee vuosittain päätöksen⁴³ verovapaista matkakustannustenkorvauksista, joita ovat:

- kilometrikorvaukset
- kotimaan päivärahat
- ulkomaan päivärahat
- ateriakorvaukset
- majoittumiskorvaukset ja yömatkaraha.

Verohallinnon päätöksessä on määritelty verovapaiden korvausten enimmäismäärät ja edellytykset, joiden pitää täytyä, jotta korvausta voidaan pitää verotuksessa verovapaina korvauksena.

5.3.4 Veronalaiset luontoisedut jätetään ilmoittamatta

Tahalliset merkittävät luontoisedujen ilmoittamatta jättämiset ja aliarvostukset ovat yksi pimeän palkan muoto. Luontoisedulla tarkoitetaan työnantajan muuna kuin rahana työstä suorittamaa vastiketta. Luontoisetu on työnantajan järjestämä ja kustantama hyödyke (tavara tai

⁴² Tämä selvitys ei kata sellaisia tilanteita, joissa maksetut palkat ja niihin liittyvät verot ja työnantajamaksut ilmoitetaan, mutta maksut jätetään osittain tai kokonaan maksamatta.

⁴³ Verohallinnon päätös verovapaista matkakustannusten korvauksista vuonna 2021 (Dnro VH/7555/00.01.00/2020)

palvelu), jonka työnantaja luovuttaa palkansaajan käyttöön. Työnantaja voi antaa työntekijän käyttöön esimerkiksi puhelimen, auton, asunnon tms.

Etuudet ovat työntekijälle veronalaista tuloa⁴⁴, josta työnantaja on velvollinen toimittamaan ennakonpidätyksen ja maksamaan palkan sivukulut. Verohallinto julkaisee vuosittain päätöksen luontoisetujen laskentaperiaatteista⁴⁵. Myös tilanteet, joissa luontoisetu ilmoitetaan käypää arvoa alhaisemmasta määrästä aiheuttaa virhettä verojen ja palkan sivukulujen määrään.

5.3.5 Osakkaan pimeä palkka ja peitelty osinko

Osakeyhtiö voi maksaa osakkaalleen palkkaa kuten kenelle tahansa työntekijälleen; tällöin palkanmaksu tulee suorittaa palkkaa koskevien säännösten mukaisesti. Jos osakas nostaa osakeyhtiöstä varoja muutoin kuin veronalaisena palkkana tai osinkona, voi menettely olla verotuksessa niin sanottua peiteltyä osingonjakoa. Tässä selvityksessä peitelty osinko katsotaan pimeäksi palkaksi. Peiteltyllä osingolla tarkoitetaan rahanarvoista etuutta, jonka osakeyhtiö antaa osakkaan tai tämän omaisen käyttöön. Osakas tai omainen saa edun käyttöönsä tavallista halvemmalla hinnalla tai jopa ilman vastiketta⁴⁶.

Osakkuusaseman perusteella saatu etu verotetaan sen laadusta ja määrästä riippumatta lähtökohtaisesti peiteltynä osinkona, jos sitä ei ole käsitelty kirjanpidossa palkkana tai luontoisetuna. Etu verotetaan palkkana, jos se on sellaisena käsitelty kirjanpidossa ja jos sen perusteena on työsuoritus. Näin tehdään esimerkiksi tilanteissa, joissa luontoisetu on palkkakirjanpidossa käsitelty virheellisesti esimerkiksi väärästä arvosta. Tapauskohtaisesti epätavanomaiset ja perusteettomat edut voidaan verottaa peiteltynä osinkona, vaikka ne olisi kirjanpidossa käsitelty palkkana tai luontoisetuna⁴⁷.

⁴⁴ Tuloverolain 64 §:n 1 momentin mukaan luontoisetu on veronalaista ansiotuloa ja se arvioidaan käypään arvoon.

⁴⁵ Verohallinnon päätös vuodelta 2021 toimitettavassa verotuksessa noudatettavista luontoisetujen laskentaperusteista (Dnro VH/7288/00.01.00/2020)

⁴⁶ Laki verotusmenettelystä 29 § 1 momentti.

⁴⁷ Peitelty osinko (Dnro VH/8436/00.01.00/2020)

6 Selvityksen verotarkastusaineistot ja yritysjoukot

Edellä luvussa viisi kerrottiin selvityksessä käytetystä pimeän työn ja palkan käsitteestä. Kuu-dennessa pääluvussa esitellään erilaisia verotarkastustyyppisiä ja verotarkastuksilla tehtyjä havaintoja. Lisäksi tässä luvussa esitellään selvityksessä käytetyt yritysjoukot.

6.1 Verotarkastukset

Verotarkastus on lakisääteistä viranomaistoimintaa⁴⁸, jota koskien Verohallinto on antanut yleisohjeen⁴⁹. Tarkastuskohteiden valinta perustuu veroriskien hallintaan. Verotarkastuksen kohteena voivat olla Suomessa kaikki koti- ja ulkomaiset luonnolliset ja juridiset henkilöt sekä heidän verovelvollisuusasemansa. Myös tietty alue, työmaa tai toimipaikka voidaan tarkastaa, esimerkiksi verovelvollisten selville saamiseksi. Verotarkastus suoritetaan sen laajuusena kuin kulloinkin on tarkoituksenmukaista. Laajuus ratkaistaan etukäteen tehdyn kartoituksen sekä tarkastuksella ilmenneiden seikkojen perusteella. Verotarkastus voi kohdistua yhteen tai useampaan verolajiin, veroriskiin tai sellaisiin liiketapahtumiin, joilla on merkitystä tarkastuskohteen verotuksen kannalta. Tarkastus kohdistetaan myös tarkoituksenmukaiseen ajanjaksoon. (Verohallinto, 2018)

Selvitykseen rajattiin mukaan ne osakeyhtiöiden verotarkastukset, joissa kohteena oli ennakkoperintälain (EPL) ja elinkeinoverolain (EVL) mukaiset tulolajit. Muita tarkastettuja verolajeja, kuten arvonlisäverotusta ja valmisteverotusta, ei huomioitu.

Verotarkastukset jaetaan tässä selvityksessä eri tyyppisiin sen mukaan, miten tarkastuskohteet valitaan (satunnais- tai kohdevalintatarkastus) ja kuinka laajasti tarkastus toteutetaan (kokonais- tai osittaistarkastus).

Satunnaistarkastukset valitaan satunnaisotannalla, ja niissä tutkitaan aina yrityksen toiminta kokonaisuudessaan. Satunnaistarkastuksissa käydään läpi kaikki verolajit ja huomioidaan kaikki löydetty virheet. Satunnaisverotarkastuksilla voidaan selvittää muun muassa verotarkastuksen vaikuttavuutta, kohdevalinnan osuvuutta ja verovajeen määrää sekä muutoksia. Satunnaistarkastukset tuottavat tietoa, jota voidaan käyttää muuhun kohdevalintaan. (Ansamäki, 2017). Satunnaistarkastukset ovat niin sanottuja kokonaistarkastuksia, joiden tarkoituksena on yleensä tuottaa mahdollisimman tarkkaa tietoa erilaisista verotuksen virheistä.

Verohallinnon vuosittain laadittavassa ohjaus- ja valvontasuunnitelmassa määritellään verovalvonnan ja verotarkastuksen painopistealueet. Suunnitelman avulla verotarkastukset pyritään kohdentamaan esimerkiksi toimialoille, jossa veroriskit ovat suurimmat. (Länsisaari, 2020) **Kohdevalinnan perusteella** tehtävät tarkastukset jaetaan kokonais- ja osittaistarkastuksiin. Osittaistarkastukset ovat suppeampia kuin kokonaistarkastukset ja niissä tarkastetaan vain tiettyjä verolajeja tai aihekokonaisuuksia. Kokonaistarkastuksissa sen sijaan käydään läpi yrityksen toiminta kokonaisuudessaan verolajeittain. Kohdevalinnan perusteella

⁴⁸ Laki verotusmenettelystä 14 § ja Laki oma-aloitteisten verojen verotusmenettelystä 24 § sekä asetus verotusmenettelystä 2 § sekä valtioneuvoston asetus oma-aloitteisten verojen verotusmenettelystä 12 §–14 §.

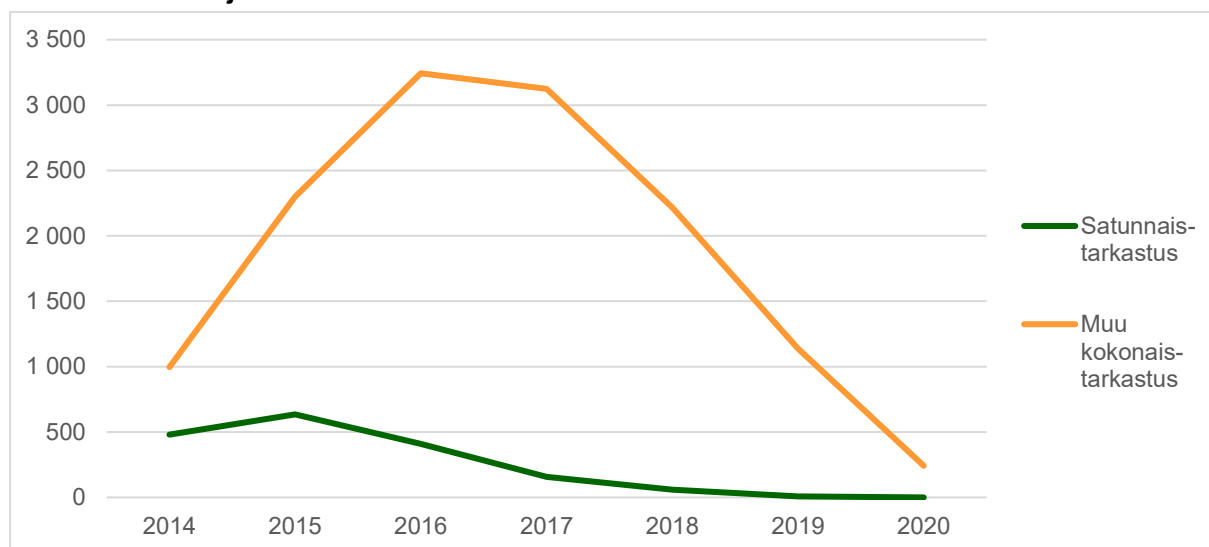
⁴⁹ Hyvä verotarkastustapa 22.1.2018 päivitetty ohje.

tehdyissä kokonaistarkastuksissa ei kuitenkaan aina tutkita yrityksen toimintaa yhtä laajasti kuin satunnaistarkastuksissa, vaan on mahdollista, että tarkastus rajataan johonkin tiettyyn verolajiin.

Tässä selvityksessä käytettiin lähdeaineistona satunnaistarkastuksia ja kohdevalinnan perusteella tehtyjä **kokonaistarkastuksia**. Osittaitarkastukset rajattiin lähdeaineistosta pois, koska niiden perusteella voi muodostua vääristynyt kuva virheiden esiintyvyydestä: esimerkiksi, jos on tarkastettu vain kustannusten korvaukset, ei tällaisen tarkastuksen perusteella voida tehdä johtopäätöksiä pimeiden palkkojen maksamisesta muilla tavoin.

Selvityksessä käytettiin vuosia 2014–2020 koskevia verotarkastusaineistoja⁵⁰. Suurin osa satunnaistarkastuksista tehtiin vuosina 2014–2017⁵¹. Kohdevalinnan perusteella tehtyjä tarkastuksia oli käytettävissä koko tutkitulta ajanjaksolta 2014–2020. Tarkastukset on jaettu eri vuosille tarkastuksen kohteena olleen ajanjakson mukaan. Sama tarkastus voi siten ajoittua usealle eri vuodelle.

Kuvio 2. Selvityksessä mukana olevien verotarkastusten lukumäärä ajalla 2014–2020. Satunnaistarkastukset ja muut kokonaistarkastukset.



Muu kokonaistarkastus sisältää tarkastukset, joissa joko EPL- tai EVL-tarkastus on tehty kokonaistarkastuksena. Lähde: Verohallinto 5–6/2021.

6.2 Verotarkastuksissa havaitut virheet

Verotarkastuksissa havaitut virheet, jotka selvityksessä huomioitiin, olivat:

- Palkansaajan rahapalkka on ilmoitettu liian pienenä tai jäänyt kokonaan ilmoittamatta (yleisesti tai rajoitetusti Suomessa verovelvollinen työntekijä tai osakas).
- Veronalaiset luontoisedut ovat ilmoitettu arvoltaan väärinä tai jätetty kokonaan ilmoittamatta.

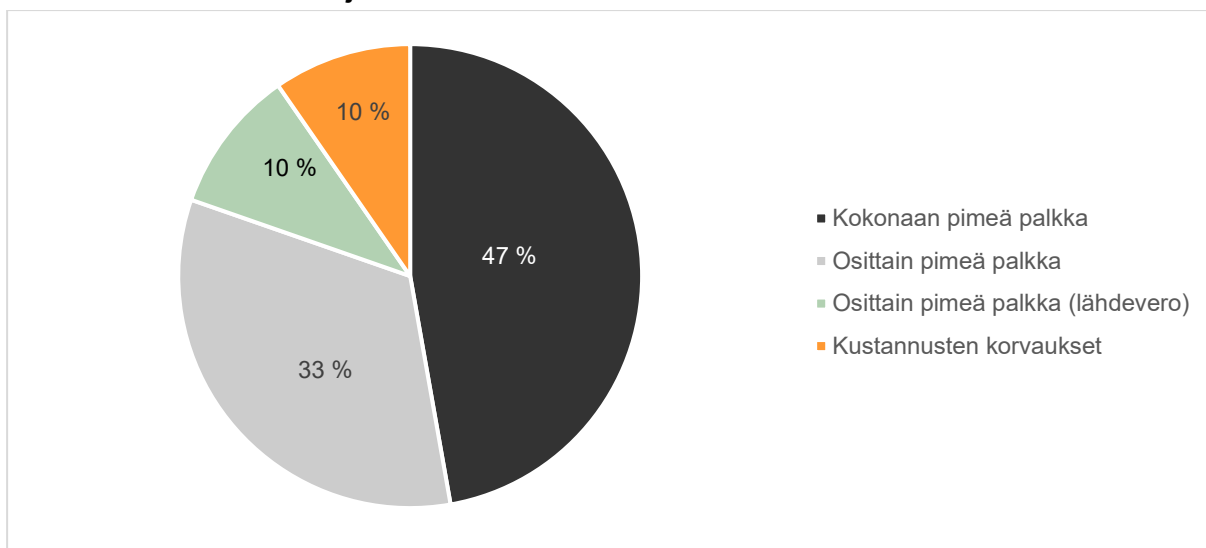
⁵⁰ On huomioitava, että osittaitarkastusten ja muiden verolajien kuin EPL ja EVL puuttumisen vuoksi oheinen graafi ei kerro Verohallinnon kyseisinä vuosina tekemien verotarkastusten kokonaismäärää.

⁵¹ Satunnaistarkastusten otoskehikko muodostui liikevaihdoltaan 300 000–10 000 000 euron yrityksistä, joihin ei ollut tehty verotarkastusta edellisten kolmen vuoden aikana.

- Palkansaajalle on maksettu rahapalkan sijaan väärin tai puutteellisin perustein verovapaita kustannusten korvauksia (esimerkiksi matkakustannusten korvauksia).
- Osakeyhtiö on antanut osakkaalleen rahanarvoisen edun osakasaseman perusteella (peitelty osingonjako).⁵²

Selvityksessä verotarkastuksissa havaitut virheet jaettiin kahteen pääluokkaan. **Työntekijöiden pimeä työ** käsittää yleisesti ja rajoitetusti verovelvollisten työntekijöiden liian pienet rahapalkat, verovapaisiin kustannusten korvauksiin liittyvät virheet ja veronalaisiin luontoisetuihin liittyvät virheet. Tarkastushavaintojen perusteella eniten veronmenetyksiä aiheuttivat kokonaan tai osittain pimeästi maksetut palkat. **Osakkaiden pimeä työ** käsittää vastaavat osakkaiden palkanmaksuun liittyvät virheet sekä peiteltyyn osingonjaon. Tarkastushavaintojen perusteella euromääräisesti suurin virheiden luokka oli peitelty osingonjako.

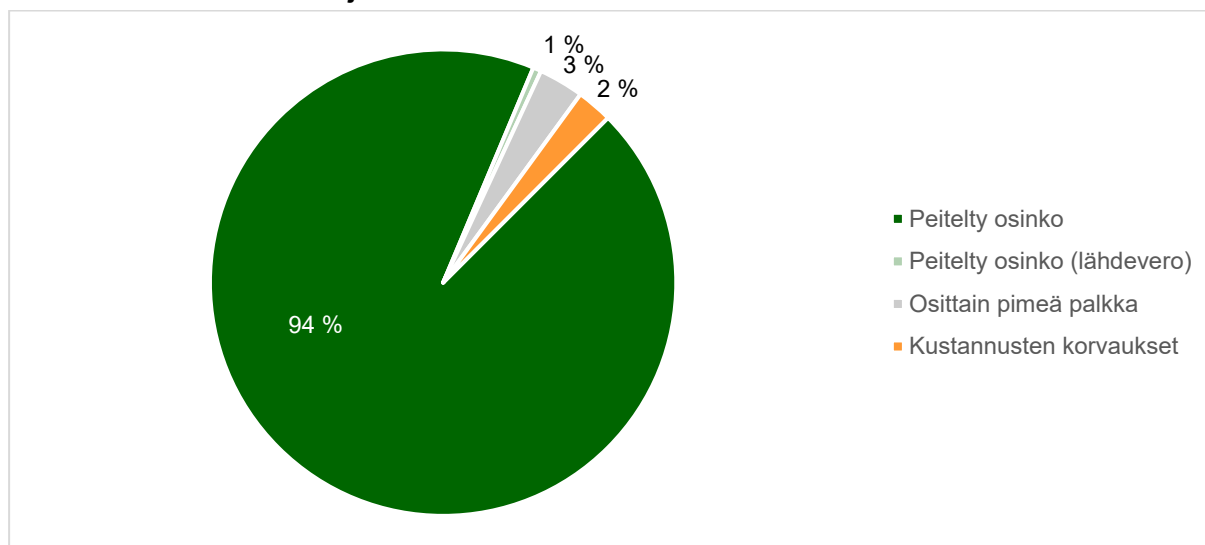
Kuvio 3. Työntekijöiden palkkoja koskevat virrehavainnot verotarkastuksissa. Virheluokkien osuudet tarkastushavaintojen euromääräistä.



Lähde: Verohallinto 5–6/2021.

⁵² Vääränsisältöiset tositteet tulevat osin huomioitua siten, että niiden perusteella maksetut pimeät palkat ja peiteltyt osingot ovat mukana selvityksessä.

Kuvio 4. Osakkaiden palkkoja koskevat virrehavainnot verotarkastuksissa. Virheluokkien osuudet tarkastushavaintojen euromäärästä.



Lähde: Verohallinto 5–6/2021.

Verohallinto keskittää tarkastukset kohdevalinnan avulla yrityksiin, joissa virheiden esiintymisriski on keskimääräistä suurempi. Kohdevalinnalla valituissa kokonaistarkastuksissa löytyikin selvästi enemmän virheitä (28 prosenttia) kuin satunnaisotannalla valituissa tarkastuksissa (12 prosenttia).

Taulukko 4. Verotarkastuksissa havaitut virheiden lukumäärät vuosina 2014–2020.

Tarkastus	Virheet lkm*	Tarkastukset lkm*	Virheet prosenttia
Satunnaistarkastukset	201	1 746	12 %
Muut kokonaistarkastukset	3 720	13 258	28 %
Kaikki yhteensä	3 921	15 004	26 %

* Virheiden ja tarkastusten lukumäärä on laskettu vuosittain. Esimerkki: yrityksen verotarkastus käsitti kolme vuotta ja virheitä esiintyi kaikkina vuosina. Tässä tapauksessa verotarkastuksesta tilastoon kirjautuu sekä tarkastusten että virheiden lukumääräksi kolme.

Lähde: Verohallinto 5–6/2021.

Virheiden suuruutta kuvaavat luvut taulukoissa 5 ja 6 ovat tarkastuksissa havaittuja veronalaisia palkkoja, joista verot ovat jääneet maksamatta. Havaittuihin virheisiin sisältyvät myös sellaiset vähäiset virheet, joista Verohallinto ei tehnyt maksuunpanoa, eli asiakas oli korjannut virheen itse verotarkastuksen aikana.

Satunnaistarkastuksissa havaitut virheet olivat keskimäärin pienempiä kuin muissa kokonaistarkastuksissa havaitut virheet (taulukko 5). Tähän on todennäköisesti vaikuttanut se, että kohdevalinnan tarkoitus on priorisoida tarkastuskohteet ja löytää verotuksen vakavimmat virheet. Kohdevalinnan perusteella tehtyjen verotarkastusten ja satunnaisotannalla toteutettujen verotarkastusten vertailun perusteella tähän tavoitteeseen on myös päästy.

Taulukko 5. Satunnaistarkastuksissa havaittujen virheiden suuruus vuosina 2014–2020.

Virhettä kuvaava tunnusluku	Työntekijät	Osakkaat	Kaikki
Löydetyt virheet yhteensä, €	3 239 100	1 874 700	5 113 800
Keskiarvo, €	36 000	15 100	25 000
Pienin, €	1 200	1 100	1 200
Mediaani, €	12 500	7 500	9 700
Suurin, €	1 247 000	136 600	1 247 000

Lähde: Verohallinto 5–6/2021.

Taulukko 6. Muissa kokonaistarkastuksissa havaittujen virheiden suuruus vuosina 2014–2020.

Virhettä kuvaava tunnusluku	Työntekijät	Osakkaat	Kaikki
Löydetyt virheet yhteensä, €	119 013 400	90 317 700	209 331 000
Keskiarvo, €	67 800	35 000	56 300
Pienin, €	600	300	300
Mediaani, €	21 000	17 700	21 000
Suurin, €	2 872 700	1 487 600	3 383 500

Lähde: Verohallinto 5–6/2021.

Satunnaisverotarkastuksissa ei havaittu lainkaan työntekijöitä, joiden kaikki palkat olisi maksettu täysin pimeästi. Sen sijaan osittain pimeästi maksettuja palkkoja havaittiin myös satunnaistarkastuksissa. Kokonaan pimeästi maksettuja palkkoja on kuitenkin havaittu muissa kuin satunnaisuuteen perustuvissa verotarkastuksissa. Nämä virheet olivat myös suuria. Se, että satunnaistarkastuksissa ei havaittu lainkaan kokonaan pimeitä työntekijöitä ja muissa tarkastuksissa niitä löytyi, selittää suurimman osan palkansaajien keskimääräisten virheiden erosta:

- Satunnaistarkastukset virheiden keskiarvo = 36 000 €
- Muut kokonaistarkastukset virheiden keskiarvo = 67 800 €
- Muut kokonaistarkastukset (pl. kokonaan pimeät palkat) virheiden keskiarvo = 46 900 €.

Osakkaiden kaikkia virhetyyppejä löytyi sekä satunnaistarkastuksista että muista tarkastuksista. Muissa kokonaistarkastuksissa havaitut peiteltyyn osingon virheet olivat suurempia kuin satunnaistarkastuksissa havaitut peiteltyyn osingon virheet, mikä selittää lähdes kokonaan taulukoista 5 ja 6 nähtävän osakkaiden keskimääräisten virheiden eron:

- Satunnaistarkastukset peiteltyjen osinkojen virheiden keskiarvo = 14 900 €
- Muut kokonaistarkastukset peiteltyjen osinkojen virheiden keskiarvo = 35 500 €.

6.3 Selvityksen perusjoukot

Perusjoukolla tarkoitetaan kaikkia laskennassa mukana olleita osakeyhtiöitä ja ulkomaisia yhteisöjä. Perusjoukko sisältää sekä tarkastetut että tarkastamattomat yritykset. Substanssi-perusteisessa menetelmässä aineistona on käytetty vuosien 2014–2018 verotarkastustietoja sekä toimivista että toimimattomista yrityksistä. Algoritmiperusteisissa menetelmissä on käytetty vuosien 2017–2020 verotarkastustietoja toimivista yrityksistä. Laskentojen perusjoukot kuvataan seuraavissa alaluvuissa.

6.3.1 2014–2020 toimivat yritykset

Toimivien yritysten perusjoukkoon kuuluivat ne tarkastelujaksolla Verohallinnon tietokannassa olleet osakeyhtiöt ja ulkomaiset yhteisöt, joilla täytyi vähintään yksi seuraavista toimivuuden kriteereistä:

- Liiketoiminnasta arvonlisäverovelvollinen⁵³
- Työnantajarekisterissä³²
- Ennakkoperintärekisterissä³²

⁵³ Kuulunut rekisteriin vähintään 1 päivän ko. vuoden aikana.

- On ilmoittanut liikevaihtoa lomakkeella 6B (yhteisön tuloveroilmoitus)
- On ilmoittanut palkkoja lomakkeella 7801 (työnantajan vuosi-ilmoitus)

Perusjoukkoon kuului kaikkiaan noin 180 000–215 000 yritystä vuosina 2014–2020, joista suurin osa oli osakeyhtiöitä.

Taulukko 7. Perusjoukon yritysten lukumäärät 2014–2020, toimivat yritykset.

	Yhteensä	Osakeyhtiöt	Ulkomaiset yhteisöt
2014	180 000	175 000	6 000
2015	182 000	177 000	6 000
2016	187 000	181 000	6 000
2017	191 000	185 000	6 000
2018	196 000	190 000	6 000
2019	204 000	198 000	7 000
2020	215 000	208 000	7 000

Lähde: Verohallinto 5–6/2021.

6.3.2 2014–2020 toimimattomat yritykset

Tässä selvityksessä niin sanottujen toimimattomien⁵⁴ yritysten perusjoukkoon kuuluivat vuosina 2014–2018 osakeyhtiöt ja ulkomaiset yhteisöt, jotka:

- olivat merkittyinä kaupparekisteriin,
- eivät kuuluneet alv-, työnantaja tai ennakkoperintärekisteriin,
- eivät olleet ilmoittaneet liikevaihtoa tai palkkoja.

Toimimattomia y-tunnuksia oli vuosittain noin 70 000. Ulkomaisten yhteisöjen osuus oli hieman suurempi kuin toimivien yritysten joukossa.

Taulukko 8. Perusjoukon yritysten lukumäärät 2014–2018, toimimattomat yritykset.

	Yhteensä	Osakeyhtiöt	Ulkomaiset yhteisöt
2014	63 000	57 000	6 000
2015	70 000	63 000	6 000
2016	75 000	68 000	7 000
2017	75 000	68 000	7 000
2018	78 000	69 000	8 000

Lähde: Verohallinto 5–6/2021.

⁵⁴ Myös rekisterimerkintöjen perusteella toimimattomia yrityksiä voidaan käyttää harmaan talouden harjoittamiseen, joten "toimimattomuus" on suhteellinen käsite. Tästä syystä myös toimimattomille yrityksille on laskettu arvio pimeän työn määrästä.

7 Pimeiden palkkojen määrä Suomessa eri menetelmin arvioituna

Tässä luvussa esitellään tarkastusaineistojen pohjalta tehtyjä pimeiden palkkojen kokonaismäärää arvioivia tuloksia sekä käytettyjä data-analyysimenetelmiä. Kaikissa menetelmissä on pohja-aineistona ollut edellä luvussa 6 kuvattu tarkastettujen yritysten joukko ja niitä koskeva data-aineisto. Kaikille analyyseille on myös yhteistä, että niissä on arvioitu pimeästi maksettujen palkkojen määrää. Vasta myöhemmässä vaiheessa on arvioitu pimeästä palkanmaksusta aiheutuneita verojen- ja muiden maksujen menetyksiä (luku 8).

Substanssiperusteisessa menetelmässä aineistona on käytetty vuosien 2014–2018 ja algoritmiperusteisissa vuosien 2017–2020 verotarkastustietoja. Laskennoissa oli eri ajankohdat siksi, että substanssiperusteisessa laskennassa oli mahdollista käyttää aineistona ennen Verohallinnon tietojärjestelmämuutosta tehtyjä verotarkastuksia, joihin kuului myös melko suuri joukko satunnaisverotarkastuksia. Algoritmiperusteinen analyysi edellytti laajemman ja tarkemman tiedon käyttöä, joka on ollut mahdollista vasta Verohallinnon tietojärjestelmämuutoksen jälkeiseltä ajalta. Vuosina 2015–2017 Verohallinnossa toteutettiin laajempi satunnais-tarkastusten projekti, jonka tuloksia myös hyödynnetään ilmiöselvityksen data-analyyseissä.

Todennäköisin arvio pimeiden palkkojen määrästä on noin 670–850 miljoonaa euroa vuodessa. Osa menetelmistä tuotti tätä suurempia tai pienempiä tuloksia, ja ne huomioiden pimeiden palkkojen vaihteluväli oli maksimissaan noin 500–1 300 miljoonaa euroa vuodessa. Pimeiden palkkojen kokonaismäärä jaettiin osakkaiden ja palkansaajien kesken. Kaikilla eri menetelmillä arvioiden osakkaiden osuus pimeiden palkkojen määrästä oli yli puolet.

Pimeän palkan määrä suhteutettiin Verohallinnolle ilmoitettuihin palkkoihin vuosina 2017–2018. Pimeiden palkkojen osuus palkkasummasta oli 1,4–1,5 prosenttia useimmilla menetelmillä mitattuna.⁵⁵ Osakkaiden pimeät palkat olivat noin 9–12 prosenttia osakkaille maksetuista palkoista ja palkansaajien pimeät palkat olivat noin 0,5–0,7 prosenttia palkansaajille maksetuista palkoista.

Vaikka selvityksessä esitetään vuosittaisia arvioita pimeiden palkkojen määrästä, ei tehtyjen analyysien tarkoituksena ole varsinaisesti ollut arvioida pimeiden palkkojen trendiä. Tarkoituksena ei siis ole ollut arvioida, mihin suuntaan pimeä työ Suomessa on kehittynyt. Sen sijaan pyrkimyksenä on ollut saada arvio pimeiden palkkojen suuruusluokasta.

7.1 Pimeät palkat ja peitelty osinko 2018–2018: substanssiperusteinen kaltaistamismenetelmä

Tässä luvussa esitetään niin sanotulla substanssiperusteisella kaltaistamismenetelmällä saatuja tuloksia pimeiden palkkojen ja peiteltyyn osingon määrästä Suomessa vuosina 2014–

⁵⁵ XGBoost-menetelmän arvio pimeiden palkkojen määrästä oli suurempi kuin muiden menetelmien tulokset, yli miljardi euroa vuodessa, eli noin 2,1 prosenttia palkkasummasta (ks. luku 7.2.3 ja taulukko 12). XGBoost-menetelmän tuottama tulos arvioitiin olevan ylöspäin harhainen eli menetelmällä arvioitu pimeiden palkkojen määrä on todennäköisesti liian suuri (ks. luku 8.1.1).

2018. Pääasiallisena aineistona on käytetty sekä satunnais- että riskiperusteisia verotarkastuksia. Tulokset on laskettu sekä toimiville että toimimattomille yrityksille.

Tarkemmat tiedot analyysin toteuttamistavasta ovat liitteessä 2.

7.1.1 Toimivien osakeyhtiöiden maksamat pimeät palkat ja peitelty osinko

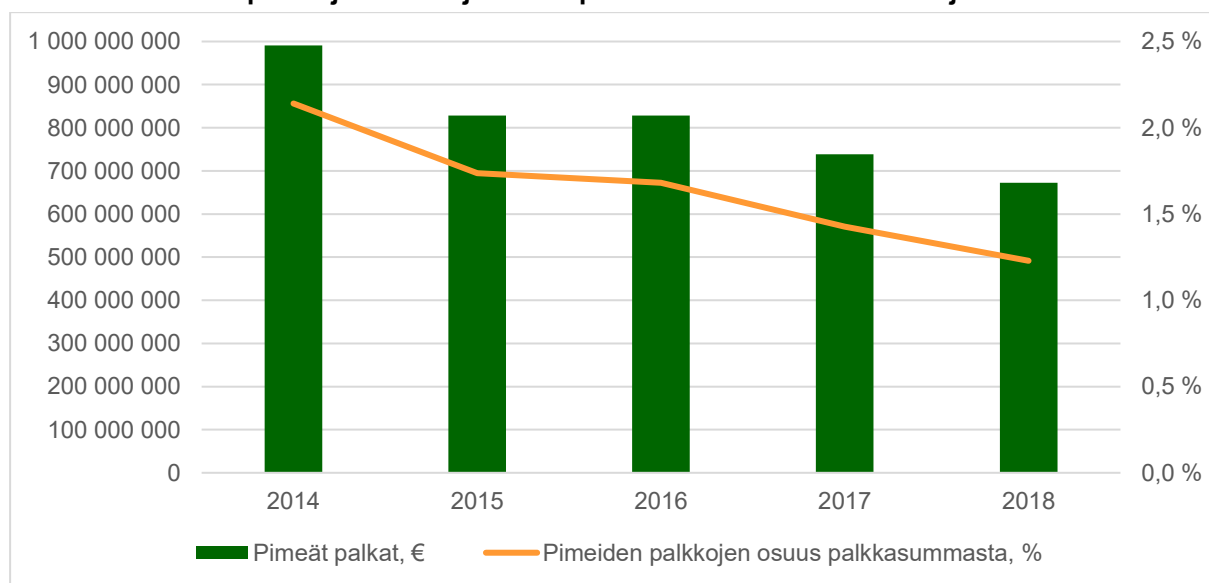
Osakeyhtiöiden ja ulkomaisten yhteisöjen arvioidaan maksaneen pimeitä palkkoja ja jakaaneen peiteltyä osinkoa keskimäärin noin 780 miljoonaa euroa vuodessa ajalla 2014–2018, mikä vastaa noin 1,6 prosenttia palkkasummasta samalla ajanjaksolla.

Maksettujen pimeiden palkkojen arvioitu alaraja oli noin 460 miljoonaa euroa vuodessa ja yläraja noin 1 040 miljoonaa euroa vuodessa. Esiitetty alaraja on kuitenkin oletettavasti harhainen alaspäin ja yläraja puolestaan oletettavasti harhainen ylöspäin. Edellä esitetty tulos, 780 miljoonaa euroa, on saatu korjaamalla aineiston vinoumia menetelmällisesti. Vinoumia ja niiden korjaamista on selitetty tiiviisti myöhemmin tässä luvussa ja yksityiskohtaisemmin liitteessä 2.

Seuraavaksi esitettävät yksityiskohtaisemmat tulokset on laskettu korjatun tuloksen perusteella, olettaen pimeän työn määräksi keskimäärin 780 miljoonaa euroa vuodessa.

Kuviossa 5 on esitetty tuloksen jakautuminen eri vuosille. Pimeiden palkkojen määrä näyttää laskevan tarkastelujaksolla, mutta jako vuosien kesken sisältää epävarmuustekijöitä. Tulokseen on voinut vaikuttaa Verohallinnon oma tietojärjestelmämuudistus, joka otettiin käyttöön vuonna 2017, sekä valtakunnallinen tulorekisterin käyttöönotto vuonna 2019. Tuloksissa olevia ajallisia epäjatkuvuuksia ei ole korjattu aikasarjatilastoinnin näkökulmasta. Trendin epävarmuuden takia tuloksia tulkittaessa on käytetty pääsääntöisesti keskimääräistä arviota pimeän työn määrästä (780 miljoonaa euroa vuodessa).

Kuvio 5. Pimeiden palkkojen määrä ja osuus palkkasummasta vuosittain ajalla 2014–2018.

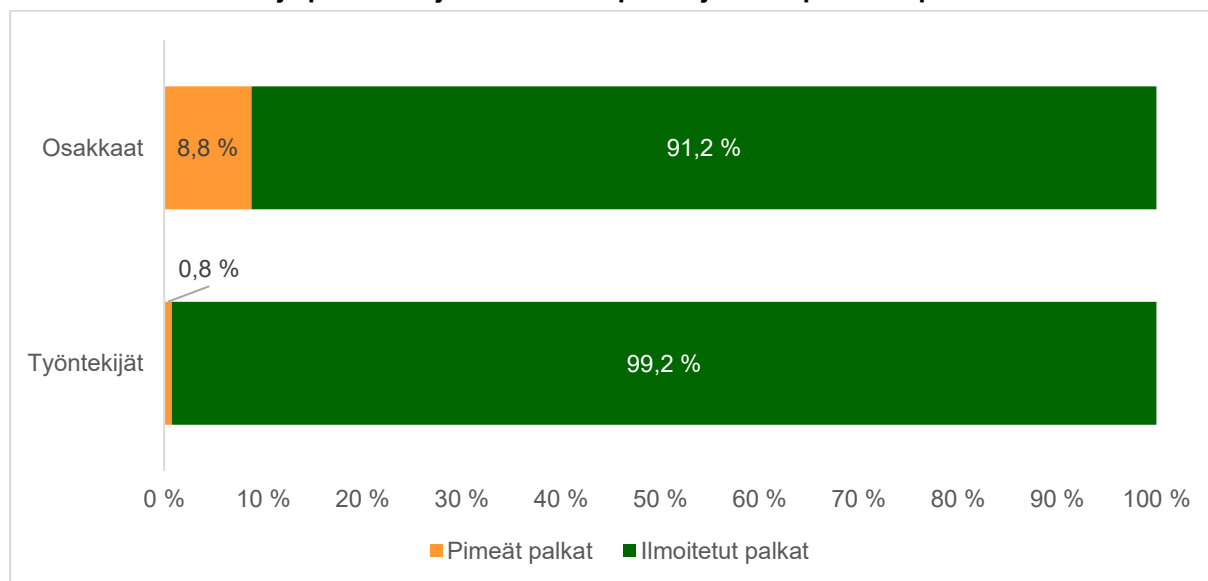


Lähde: Verohallinto 5–6/2021.

Osakkaiden pimeiden palkkojen osuus oli keskimäärin 53 prosenttia ja palkansaajien pimeiden palkkojen osuus keskimäärin 47 prosenttia kaikista pimeistä palkoista.

Osakkaille ja palkansaajille maksetut pimeä palkat voitiin suhteuttaa (kuva 6) näille maksetuihin palkkasummiin ajalla 2017–2018. Arviolta lähes yksi prosentti palkansaajien palkoista ja lähes yhdeksän prosenttia osakkaiden palkoista maksettiin pimeästi vuosina 2017–2018.

Kuvio 6. Osakkaiden ja palkansaajien ilmoitetut palkat ja arvio pimeistä palkoista 2017–2018.



Lähde: Verohallinto 5–6/2021.

Lyhyt kuvaus laskennasta ja sen perusteista

Laskenta toteutettiin niin, että perusjoukon yritykset jaettiin luokittelevilla muuttujilla pienempiin, keskenään samankaltaisten yritysten ryhmiin⁵⁶. Luokittelujen tarkoitus on pienentää riskiperusteisten tarkastusten aiheuttamaa harhaa arvioidussa pimeän työn määrässä. Luokittelujen ansiosta mahdolliset otannan aiheuttamat vääristymät kohdistuvat rajatumpiin joukkoihin. Näin virheet eivät pääse kertautumaan koko perusjoukkoon ja ne jäävät pienemmiksi.

Huomattakoon, että myös puhtaasti satunnaisotantaan perustuvissa laskennoissa käytetään normaalisti luokitteluja ja ne parantavat tulosten tarkkuutta. Mutta, jos otanta ei ole satunnainen, tarvitaan enemmän ja/tai tarkempia luokitteluja. Tämä puolestaan tarkoittaa, että otoshavaintoja – tässä tapauksessa tarkastushavaintoja – tarvitaan enemmän.

Pimeän työn määrä saatiin kertomalla keskenään arvioitu **virheiden lukumäärä** ja verotarkastuksissa havaittu **virheiden keskiarvo**. Tämä tulos laskettiin erikseen jokaiselle ositteelle ja kokonaistulos saatiin ositteiden summana.

Virheiden lukumäärä saatiin kertomalla perusjoukon tarkastamattomien yritysten lukumäärä virheiden **esiintyvyydellä**. Virheiden esiintyvyys puolestaan on tarkastuksissa havaittujen virheiden lukumäärä jaettuna tarkastusten lukumäärällä. Tuloksen alarajaa laskettaessa esiintyvyys arvioitiin yksinomaan satunnaistarkastusten perusteella. Ylärajan ja korjatun tuloksen laskennoissa hyödynnettiin lisäksi tietoja riskiperusteisista tarkastuksista. Kuitenkin koska virheiden esiintyvyys riskiperusteisissa tarkastuksissa suhteutettiin lopulta

⁵⁶ Menetelmän yksityiskohtainen kuvaus on esitetty liitteessä 2.

satunnaistarkastuksiin, voidaan sanoa, että näissäkin laskennoissa esiintyvyys vastasi satunnaistarkastusten esiintyvyyttä keskimäärin.

Erilaiset arviot pimeiden palkkojen määrästä – alaraja, korjattu tulos ja yläraja – johtuivat suurimmaksi osaksi eri tavoin lasketuista keskimääräisistä virheistä. Alaraja saatiin laskemalla virheiden keskiarvo pelkästään satunnaistarkastusten perusteella. Satunnaistarkastusten otoskehikosta puuttuivat hiljattain tarkastetut yritykset, minkä vuoksi virheiden keskiarvo on oletettavasti ollut alaspäin harhainen.⁵⁷ Pimeiden palkkojen yläraja saatiin laskemalla virheiden keskiarvo kaikkia tarkastustietoja käyttäen (sekä satunnaiset että riskiperusteiset). Ylärajan laskennassa käytetty virheiden keskiarvo on oletettavasti ollut ylöspäin harhainen, koska verotarkastuksen kohdevalinta on pyrkinyt poimimaan tarkastettavaksi suurimpia virheitä tehneet yritykset.

Korjattu tulos saatiin laskemalla virheiden keskiarvo samasta aineistosta kuin yläraja, mutta oletettua harhaa pienennettiin niin sanotun trimmatun keskiarvon avulla. Siinä 5 prosenttia tarkastushavainnoista poistettiin jakauman ylä- ja alapäistä eli suurimmat ja pienimmät virheet, ja laskettiin virheiden keskiarvo jäljelle jääneistä havainnoista. Koska virheiden jakaumat olivat voimakkaasti vinoja oikealle, menettely pienensi keskiarvoa ja siten oletettua harhaa.

7.1.2 Toimimattomien osakeyhtiöiden maksamat pimeät palkat ja peitellyt osingot

Tässä luvussa esitetään arvio niin sanottujen toimimattomien yritysten maksamista pimeistä palkoista. Verotarkastuksilla myös näillä yrityksillä on havaittu pimeää palkanmaksua. Verotarkastushavainnot ja selvityksen aihe (pimeä työ) huomioiden ei tätä yritysjoukkoa voitu jättää analyysissä täysin huomiotta.

Toimimattomien yritysten perusjoukko on kuvattu luvussa 6.3. Joukko oli varsin suuri, vuosittain 63 000–78 000 y-tunnusta.

Toimimattomien yritysten arvioidaan maksaneen pimeitä palkkoja keskimäärin noin 70 miljoonaa euroa vuodessa ajalla 2014–2018. Tästä arviolta puolet oli ulkomaisten yhteisöjen maksamia pimeitä palkkoja ja puolet suomalaisten osakeyhtiöiden maksamia pimeitä palkkoja.

Esitettyyn arvioon liittyy paljon epävarmuustekijöitä – esimerkiksi verotarkastuksia oli tehty paljon vähemmän kuin toimivien yritysten joukossa, eikä satunnaistarkastuksia ollut tehty lainkaan. Koska tarkastuksia ja luokittelevia tietoja oli käytettävissä vain vähän, laskentaa ei voitu toteuttaa menetelmällisesti "oikeaoppisesti". Saatua tulos on vain suuntaa antava asiantuntija-arvio, johon liittyy paljon oletuksia ja epävarmuutta. Tarkastelun avulla voidaan kuitenkin hahmottaa toimimattomien yritysten pimeän työn potentiaalia ja riskialueita.

Toimimattomiin yrityksiin oli ajalla 2014–2018 tehty 261 sellaista verotarkastusta, jotka voitiin huomioida pimeän työn arvioinnissa. Havaintoja pimeästi maksetuista palkoista löytyi 82 tapauksessa. Jos tarkastusten määrä suhteutetaan yritysten lukumäärään, havaitaan, että

⁵⁷ Esimerkiksi satunnaisverotarkastuksissa ei havaittu lainkaan työntekijöitä, joiden kaikki palkat oli maksettu pimeästi. Kohdevalintatarkastuksissa tällaisia täysin pimeitä työntekijöitä kuitenkin löydettiin.

suhde on paljon pienempi kuin toimivien yritysten joukossa. Toimimattomista yrityksistä vain 0,04 prosenttia tarkastettiin vuosittain, kun toimivista yrityksistä tarkastettiin vuosittain keskimäärin 1,4 prosenttia.⁵⁸ Toimimattomien yritysten tarkastuksissa havaittiin virheitä noin 4,5 miljoonan euron edestä. Suurin osa (78 prosenttia) oli osakkaiden pimeitä palkkoja ja pienempi osa (22 prosenttia) palkansaajien pimeitä palkkoja.

Esitetty toimimattomien yritysten pimeiden palkkojen suuruus (70 miljoonaa euroa / vuosi) on helpompi hahmottaa, jos se suhteutetaan perusjoukon kokoon eli jaetaan perusjoukon yritystunnusten lukumäärällä. Tällöin osakeyhtiöiden pimeät palkat olivat keskimäärin hieman alle 600 euroa yritystunnusta kohden ja ulkomaisten yhteisöjen pimeät palkat olivat keskimäärin noin 5 400 euroa yritystunnusta kohden. Vertailun vuoksi todettakoon, että toimivien osakeyhtiöiden pimeät palkat olivat keskimäärin noin 4 000 euroa yritystä kohden ja toimivien ulkomaisten yhteisöjen vastaava luku oli noin 12 000 euroa yritystä kohden.

Keskimääräiset luvut paljastavat, että pimeän työn riski arvioitiin paljon suuremmaksi toimimattomien ulkomaisten yhteisöjen keskuudessa kuin toimimattomien osakeyhtiöiden keskuudessa. Toimimattomien yritysten joukko voidaan jakaa neljään päätyyppiin. Näiden lähempi tarkastelu auttaa hahmottamaan mahdollisia pimeän työn riskiryhmiä.⁵⁹

1. **Ulkomaiset yhteisöt, jotka eivät hakeudu tunnuksen perustamisvaiheessa Verohallinnon rekistereihin** (noin 5 200 kpl eli 78 prosenttia toimimattomista ulkomaisista yhteisöistä). Näistä yli puolella ei ollut edes toimialatietoa. Pimeän työn riskin arvioitiin olleen suurin tässä toimimattomien yritysten ryhmässä. Noin puolet kaikista toimimattomiin yrityksiin tehdyistä verotarkastuksista kohdistui tähän ryhmään.
2. **Ulkomaiset yhteisöt, joiden toiminta oli päättynyt**, eli ne olivat aiemmin kuuluneet Verohallinnon rekistereihin (noin 1 500 kpl eli 22 prosenttia toimimattomista ulkomaisista yhteisöistä).
3. **Osakeyhtiöt, jotka eivät hakeudu tunnuksen perustamisvaiheessa Verohallinnon rekistereihin** eli niin sanotut pöytälaatikkoon⁶⁰ perustetut yhtiöt. Näiden osuus kaikista toimimattomista osakeyhtiöistä oli noin 45 prosenttia (noin 29 000 kpl). Toisin kuin ulkomaisilla yhteisöillä, suurimmalla osalla tähän ryhmään kuuluvista osakeyhtiöistä oli kuitenkin toimialatieto (vain 1 prosentilla toimimattomista osakeyhtiöistä toimialatieto puuttui).
4. **Osakeyhtiöt, joiden toiminta oli päättynyt**, eli ne olivat aiemmin kuuluneet Verohallinnon rekistereihin. Niin sanotut pöytälaatikkoon jätetyt yhtiöt. Näiden osuus kaikista toimimattomista osakeyhtiöistä oli noin 55 prosenttia (noin 36 000 kpl).

Suhteellisesti eniten pimeää työtä arvoitiin esiintyvän edellä kuvatussa ryhmässä 1 eli ulkomaisissa yhteisöissä, jotka eivät hakeudu tunnuksen perustamisvaiheessa Verohallinnon rekistereihin. Lähes puolet arvioidusta pimeistä palkoista kohdistui tähän ryhmään (noin

⁵⁸ Suhdeluvut on laskettu selvityksessä huomioitujen tarkastusten perusteella, jotka olivat EPL- ja EVL-verolajien kokonaistarkastuksista. Muita verolajeja sekä EPL- ja EVL-verolajien osittaitarkastuksia ei ole huomioitu.

⁵⁹ Luettelossa ilmoitetut prosenttiluvut on laskettu vuosittaisista havainnoista ajalla 2014–2018 ja luvut ovat vuosittaisia keskiarvoja samalta ajanjaksolta.

⁶⁰ Pöytälaatikkoyrityksellä tarkoitetaan sellaista yritystä, joka on juridisesti olemassa eli sillä on kaupparekisteriin merkitty y-tunnus, mutta se ei ole Verohallinnon rekistereissä (alv-velvollinen, ennakkoperintärekisteri tai työnantajarekisteri).

35 miljoonaa euroa vuodessa). Muissa ryhmissä pimeiden palkkojen määrä arvioitiin vähäiseksi, mikä on luettavissa myös edellä mainitusta yrityskohtaisesta virheestä (alle 600 euroa / yritystunnus).

7.2 Pimeät palkat ja peitelty osingot 2017–2020: algoritmilähtöiset laskentamenetelmät

Tässä luvussa esitellään neljän algoritmilähtöisen laskentamenetelmän tuottamat estimaatit vuosille 2017–2020. Käytetyt menetelmät ovat kaltaistamislähestymistapaa hyödyntävät CEM-kaltaistaminen ja DBSCAN-klusterointi sekä yritystasoisien asiakaskohtaisen pimeän työn odottaman laskenta XGBoost-koneoppimisalgoritmeilla ja bayesilaisella logistisella regressiolla. Kaikilla menetelmillä saadaan tehtyä kokonaisestimaatin jako palkansaajien pimeisiin palkkoihin sekä osakkaiden peiteltyihin osinkoihin, mutta jakamiseen käytetty tekniikka vaihtelee menetelmittäin. Vuodet 2017–2020 ovat valittu rajaukseksi, koska Verohallinnon tietojärjestelmämuutoksen myötä tältä ajalta on ollut käytettävissä laaja, yhtäjaksoinen ja samanmuotoinen data. Sen avulla saadaan muodostettua laadukas yrityksen ominaisuustiedoista koostuva useiden kymmenien eri muuttujien aineisto.

CEM- ja DBSCAN-menetelmät ovat sukulaisia edellä esitetyille substanssiperusteiselle kaltaistamismenetelmälle. Erona on se, että saadut ositteet syntyvät puhtaan datalähtöisesti algoritmien tekemän päättelyn seurauksena. Kuten edellä, niitä käytetään yrityspopulaation jakamiseen algoritmisesti pieniin, keskenään samankaltaisiin joukkoihin, jolloin verotarkastettujen yritysten ominaisuuksia voidaan yleistää pieneen osapopulaatioon. Tällä tavalla saadaan riskiperusteisen valikoinnin aiheuttamaa valikoitumisharhaa poistettua ja siten saadut estimaatit ovat virheettömämpiä.

Kolmantena ja neljäntenä algoritmivetoisena laskentamenetelmänä on käytetty asiakaskohtaisen pimeän työn odottaman laskentaa, jossa kahdella eri tavalla (XGBoost ja bayesilainen regressio) on ennustettu kunkin yrityksen pimeän palkan riskiä ja sen mahdollista kokoluokkaa. Tässä lähestymistavassa kullekin yritykselle päätellään sen ominaisuuksien perusteella oma riskiarvio pimeän työn teettämisestä.

Menetelmiä ja niiden avulla saatuja tuloksia esitellään seuraavissa alaluvuissa. Käytettyjen analyysimenetelmien syvällisemmät kuvaukset ovat liitteissä 3 ja 4.

7.2.1 CEM-kaltaistaminen

Menetelmän avulla laskettuna pimeät palkat ovat olleet vuosien 2017–2020 verotarkastustietojen mukaan noin 780–815 miljoonaa euroa vuodessa. Taulukosta 10 ilmenee, että CEM-kaltaistamisen avulla saadaan arvio, jossa pimeistä palkoista 1/3 koostuu palkansaajan asemassa saaduista palkoista ja noin 2/3 osakasasemassa tehdystä pimeästä työstä.

Taulukko 9. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) CEM-kaltaistamismenetelmällä laskettuna.

Vuosi	Pimeät palkat yhteensä	Palkansaajien pimeät palkat	Osakkaiden pimeät palkat
2017	780	243	537
2018	797	233	564
2019	801	212	589
2020	814	242	572

Lähde: Verohallinto 5–8/2021.

Coarsened Exact Matching eli CEM on yksinkertainen ja tehokas kaltaistamisen menetelmä, jolla voidaan tasoittaa valikointiharhaa. CEM-menetelmässä jokaiselle tarkastetulle yritykselle pyritään löytämään vastinpari tai pareja tarkastamattomien yritysten joukosta, jotka vastaavat valittujen ominaisuuksien osalta täysin verotarkastettua yritystä. Verotarkastettujen yritysten tarkastustulokset kopioidaan heille löydetuille verrokkiyrityksille sellaisenaan ja ilman paria jääneiden tarkastamattomien yritysten pimeän työn määräksi ei jää mitään.

7.2.2 DBSCAN-kaltaistaminen

Menetelmän perusteella pimeät palkat ovat olleet vuosien 2017–2020 verotarkastustietojen mukaan noin 770–850 miljoonaa euroa. Summa jakautuu siten, että noin 40 prosenttia on palkansaajien ja noin 60 prosenttia osakkaiden pimeää palkkaa.

Taulukko 10. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) DBSCAN-algoritmilla laskettuna.

Vuosi	Pimeät palkat yhteensä	Palkansaajien pimeät palkat	Osakkaiden pimeät palkat
2017	769	314	455
2018	798	326	472
2019	826	337	489
2020	852	351	501

Lähde: Verohallinto 5–8/2021.

DBSCAN on klusterointialgoritmi, joka annettujen parametrien valossa päätelee itse optimaalisen määrän klustereita eli yritysryppäitä. Tällä tavalla aineisto jaetaan automaattisesti joukkoon keskenään samankaltaisia ositteita, minkä lisäksi osa yrityksistä jää poikkeaviksi havainnoiksi kaikkien ositteiden ulkopuolelle. Yritysryppäiden sisällä menetelmä päätelee verotarkastettujen yritysten avulla pimeän työn esiintyvyyttä sekä virheen suuruutta, joka yleistetään koko ryppään tarkastamattomiin yrityksiin. Menetelmä päätelee aluksi sopivat joukot verotarkastettujen yritysten ominaisuuksien perusteella ja sitten jakaa kaikki tarkastamattomat yritykset ominaisuuksiltaan vastaaviin joukkoihin.

7.2.3 Yrityskohtaisen odottaman laskenta kahdella tavalla

Kolmantena algoritmivetoisena menetelmänä on käytetty yrityskohtaista pimeiden palkkojen odottamaa. Yrityskohtaisella odottamalla tarkoitetaan kullekin yritykselle algoritmilähtöisesti estimoitua pimeiden palkkojen odotusarvoa. Menetelmä arvioi kunkin yrityksen riskisyyttä verotarkastusten ja muiden verotustietojen avulla.

Tarkastettujen yritysten aineistoon opetetaan tarkastustulosta ennustava algoritmi; tämä tapahtuu yrityksen erilaisia ominaisuuksia, kuten keskipalkkaa hyödyntämällä. Kun tarkastettujen yritysten joukossa on opittu suhde ennustavien muuttujien sekä tarkastustuloksen välille, opitulla mallilla voidaan päätellä pimeiden palkkojen esiintyvyyttä myös tarkastamattomilla yrityksillä, tiettyjen reunaehtojen vallitessa. Jokaiselle yritykselle päätellään kaksi arvoa

kahdella eri mallilla: kuinka todennäköisesti yritys maksaa pimeitä palkkoja sekä pimeiden palkkojen suuruus, jos riski toteutuu. Kertomalla yrityskohtainen pimeiden palkkojen maksamisen todennäköisyys pimeän palkan euromäärällä saadaan jokaiselle yritykselle pääteltyä pimeän työn odottama.

Esimerkiksi jos asiakkaalle X arvioidaan 30 prosentin todennäköisyys pimeiden palkkojen maksamiseen ja mallin arvioima pimeiden palkkojen määrä on 20 000 euroa, tulee asiakkaan odottamaksi tällöin 6 000 euroa. Kun jokaiselle yritykselle on päätelty odottama, saadaan muodostettua arvio pimeiden palkkojen kokonaismäärästä kunakin vuonna.

Todennäköisyydelle, että yritys maksaa pimeitä palkkoja, rakennettiin kaksi erilaista mallia. Toinen on päätöspuihin perustuva XGBoost-malli ja toinen bayesilainen logistinen regressio. Näin saadaan kaksi erillistä arviota, joita esitellään seuraavaksi lyhyesti. Yrityskohtaisen pimeiden palkkojen euromäärää arvioiva regressiomalli tehtiin vain yhdellä menetelmällä ja sitä hyödynnetään molempien luokittelevien mallien yhteydessä.

XGBoost-koneoppimismalli

Extreme Gradient Boosting eli XGBoost on tehokas päätöspuu-pohjainen koneoppimismalli, joka toimii erityisen hyvin luokitteluongelmissa, joissa hyödynnetään niin sanottua rakenteellista dataa. Sen huonona puolena on, että tehtyjä ennusteita ei voida selittää käyttämättä erilisiä selkoälyn (explainable AI eli XAI) menetelmiä, joilla algoritmin musta laatikko saadaan avattua. Sen avulla voidaan kelvollisella tarkkuudella ennustaa, kuinka todennäköisesti yritys maksaa pimeitä palkkoja, kun tunnetaan yrityksen ominaispiirteet.

Tällä tavalla laskettuna pimeiden palkkojen määräksi arvioidaan vuosittain noin 1,1–1,3 miljardia euroa (taulukko 11).

Taulukko 11. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) XGBoost-koneoppimismallilla laskettuna.

Vuosi	Pimeät palkat yhteensä	Palkansaajien pimeät palkat	Osakkaiden pimeät palkat
2017	1 123	319	804
2018	1 077	306	771
2019	1 158	339	818
2020	1 350	429	921

Lähde: Verohallinto 5–8/2021.

Kuviossa 7 on havainnollistettu, kuinka XGBoost-malli arvioi tarkastettujen ja tarkastamattomien yritysten todennäköisyyttä maksaa pimeitä palkkoja. Erityisesti kuvio kertoo sen, kuinka näihin kahteen joukkoon kuuluvat yritykset jakautuvat pimeän työn todennäköisyyden suhteen. Todennäköisyyden arvo on kuviossa vaakaa-akselilla ja sen suhteen peilatut jakaumat ovat niin sanottuja tiheysfunktioita.

Tarkastettujen jakauma painottuu korostetusti kahteen ääripäähän eli pienen todennäköisyyden yrityksiin, joilla pimeän työn todennäköisyys on alle 25 prosenttia⁶¹, ja suuren todennäköisyyden yrityksiin, joilla pimeän työn todennäköisyys on yli 75 prosenttia⁶². Malli antaa

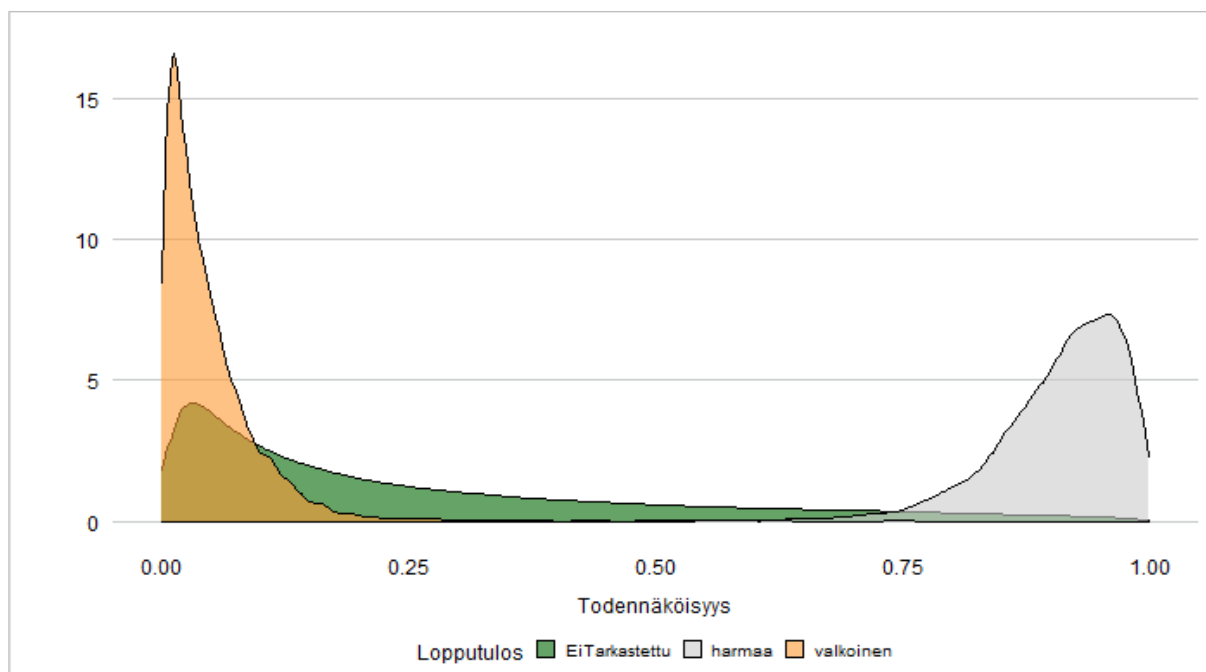
⁶¹ Tarkastettujen tiheysfunktion pinta-ala välillä 0–0.25 kuvaa sitä, kuinka suurella osalla tarkastetuista pimeän työn todennäköisyys on alle 25 prosenttia (kuviossa erottuva keltainen alue).

⁶² Tarkastettujen tiheysfunktion pinta-ala välillä 0.75–1 kuvaa sitä, kuinka suurella osalla tarkastetuista pimeän työn todennäköisyys on yli 75 prosenttia (kuviossa erottuva harmaa alue).

tarkastetuille yrityksille korostetun varman arvion riskisyydestä, koska se on nähnyt nämä tapaukset jo harjoitusaineistossa.

Tarkastamattomille yrityksille malli antaa realistisen oloisen jakauman, jossa suurimmalla osalla yrityksistä pimeiden palkkojen todennäköisyys on pieni (tarkastamattomien todennäköisyyden mediaani on 16 prosentin kohdalla). Myöhemmin luvussa 8.3 on kuvattu sellaisten yritysten ominaisuuksia, joiden XGBoost-malli arvioi suurella todennäköisyydellä maksaneen pimeitä palkkoja. Nämä ovat tarkastamattomia yrityksiä, joille malli arvioi yli 60 prosentin todennäköisyys maksaa pimeitä palkkoja. Kuviossa seitsemän ne sijoittuvat tarkastamattomien jakaumalla kohdan 0,6 oikealle puolelle.

Kuvio 7. XGBoost-mallin tuottamat pimeän työn todennäköisyysjakaumat tarkastetuille ja tarkastamattomille yrityksille.⁶³



Lähde: Verohallinto 5–8/2021.

Bayesilainen logistinen regressio

Bayesilaista lähestymistapaa voidaan hyödyntää asiakaskohtaisen pimeiden palkkojen odotetun laskemiseen. Bayesilainen logistinen regressio mahdollistaa vanhojen satunnaistarkastusten sisältämän informaation hyödyntämisen mallin rakennuksessa asettamalla niihin perustuva priorijakauma mallin vakiotermeille. Näin satunnaistarkastusten sisältämä

⁶³ Tarkastettujen ja tarkastamattomien yritysryhmien ydinestimointimenetelmällä estimoidut tiheysfunktiot. Tiheysfunktio on jatkuvan satunnaismuuttujan (tässä tapauksessa pimeän työn todennäköisyys) todennäköisyyden jakautumista kuvaava todennäköisyysfunktio. Tiheysfunktioiden arvot eivät itsessään ole todennäköisyyksiä, mutta tiheysfunktioin avulla voidaan laskea todennäköisyydet käyttämällä määrättyä integraalia. Määrätyn integraalin tulos on pinta-ala ja samalla myös kysytty todennäköisyyden arvo. Koko tiheysfunktion pinta-alan suuruus on 1, mikä tarkoittaa, että kaikkien vaihtoehtojen toteutumisen todennäköisyys on 100 prosenttia.

vinoutumaton⁶⁴ informaatio saadaan mallinnettua formaalisti osaksi ennustemallia ja mallin päätelmä pimeän työn esiintyvyydestä koko perusjoukossa tulee realistisemmaksi.

Näin arvioiden pimeiden palkkojen määrä vuosittain on noin 750–1 000 miljoonaa euroa.

Taulukko 12. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) bayesilaisella logistisella regressiolla laskettuna.

Vuosi	Pimeät palkat yhteensä	Palkansaajien pimeät palkat	Osakkaiden pimeät palkat
2017	751	219	532
2018	748	225	523
2019	855	276	579
2020	1 010	352	658

Lähde: Verohallinto 5–8/2021.

⁶⁴ Satunnaistarkastusten sisältämää informaatiota voidaan pitää lähes vinoutumattomana, kun lasketaan todennäköisyyksiä virheiden esiintymiselle yritysjoukossa. Sen sijaan keskimääräisen virheen suuruutta arvioitaessa satunnaistarkastukset antavat todennäköisesti alaspäin harhaisen kuvan.

8 Pimeän työn euromäärä sekä verojen ja muiden maksujen menetykset

Osakeyhtiöiden maksamien pimeiden palkkojen määrä Suomessa oli arviolta noin 670–850 miljoonaa euroa vuodessa. Osa menetelmistä tuotti tätä suurempia tai pienempiä tuloksia ja ne huomioiden pimeiden palkkojen vaihteluväli oli maksimissaan noin 500–1 300 miljoonaa euroa vuodessa. Suurempi vaihteluväliä sisältää kuitenkin epävarmoja tuloksia, joten tämän selvityksen perusteella vuosittainen pimeiden palkkojen määrä on todennäköisimmin ensin mainittu 670–850 miljoonaa euroa vuodessa. Pimeiden palkkojen kokonaismäärä jaettiin osakkaiden ja palkansaajien kesken. Kaikilla eri menetelmillä arvioiden osakkaiden osuus pimeiden palkkojen määrästä oli yli puolet.

Tehtyjen analyysien perusteella osakeyhtiöiden ja ulkomaisten yhteisöjen teettämästä pimeästä työstä aiheutui vuosittain arviolta noin 260–300 miljoonan euron ennakonpidätysten ja sosiaalivakuutusmaksujen menetykset ajalla 2014–2020. Todennäköisimmät arviot maksamatta jääneistä osakkaiden ennakonpidätyksistä vaihtelivat noin 120–140 miljoonan euron välillä sekä arviot palkansaajien maksamatta jääneistä ennakonpidätyksistä ja sosiaalivakuutusmaksuista noin 125–175 miljoonan euron välillä.

Pimeän palkan määrä suhteutettiin Verohallinnolle ilmoitettuihin palkkoihin vuosina 2017–2018. Pimeiden palkkojen osuus palkkasummasta oli 1,4–1,5 prosenttia eri menetelmillä mitattuna.⁶⁵ Osakkaiden pimeät palkat olivat noin 9–12 prosenttia osakkaille maksetuista palkoista ja palkansaajien pimeät palkat olivat noin 0,5–0,7 prosenttia palkansaajille maksetuista palkoista.

Arvioidut pimeiden palkkojen osuudet ovat samaa suuruusluokkaa Eurobarometri haastattelutkimuksen kanssa, jossa 3 prosenttia suomalaisista vastaajista kertoi tehneensä pimeää työtä. Nyt arvioitu pimeiden palkkojen määrä on myös samaa suuruusluokkaa, kuin Ruotsin tuoreimmassa verovajeraportissa esitetty arvio luonnollisten henkilöiden tuloverojen vajeesta (1,5 % vastaavista veroista). Sen sijaan nyt arvioitu pimeiden palkkojen määrä Suomessa on selvästi pienempi kuin Ruotsissa verotarkastustietojen pohjalta esitetty arvio, jonka mukaan pimeiden palkkojen osuus oli noin kuusi prosenttia kaikista palkoista. Näyttääkin siltä, että tässä selvityksessä on erilaisten menetelmien avulla onnistuttu torjumaan tehokkaasti riskiperusteisten verotarkastusten valikointiharhaa ja sen vaikutusta pimeän palkan estimaatteihin.

Mittaustulosten mukaan pimeiden palkkojen määrä vaihteli vuosittain. Estimaatteja ei kuitenkaan ole muodostettu aikasarjaperiaatteella. Näin ollen eri vuosien välillä nähtävät muutokset voivat johtua muistakin kuin ilmiöön liittyvistä syistä. Selvityksen keskeinen löydös onkin

⁶⁵ XGBoost-menetelmän arvio pimeiden palkkojen määrästä oli suurempi kuin muiden menetelmien tulokset, yli miljardi euroa vuodessa, eli noin 2,1 prosenttia palkkasummasta. XGBoost-menetelmän tuottama tulos arvioitiin olevan ylöspäin harhainen eli menetelmällä arvioitu pimeiden palkkojen määrä on todennäköisesti liian suuri (ks. luku 8.1.1). Sen sijaan menetelmä ei anna harhaanjohtavaa tietoa esimerkiksi pimeitä palkkoja maksavien yritysten toimialoista tai kokoluokista, joita on tarkasteltu luvussa 8.3.

pimeiden palkkojen suuruusluokka tai niiden osuus palkkasummasta, ei niiden ajallinen kehitys.

Verotarkastuksissa on havaittu pimeää palkanmaksua myös sellaisilla yrityksillä, jotka eivät kuulu Verohallinnon rekistereihin, mutta joilla on voimassa oleva y-tunnus kaupparekisterissä. Näille niin sanotuille toimimattomille yrityksille muodostettiin erillinen asiantuntija-arvio pimeästi maksettujen palkkojen määrästä. Toimimattomien yritysten arvioitiin maksaneen pimeitä palkkoja vuosittain noin 70 miljoonaa euroa ajalla 2014–2018. Toimimattomien osakeyhtiöiden pimeän työn riski arvioitiin vähäiseksi, mutta ulkomaisten yhteisöjen kohdalla riskin oletettiin olevan selvästi suurempi.

Data-analyysin perusteella hahmoteltiin myös tyypillinen pimeitä palkkoja maksanut yritys. Se on varsin pieni, tuottaen liikevaihtoa alle 400 000 euroa vuodessa. Vastaavasti myös ilmoitettujen palkkojen määrät ovat keskimäärin varsin pieniä jakaantuen suuremmalta osin osakkaille ja alle 50 prosenttia osuudelta palkansaajille. Monista laki- ja kontrollitoimista huolimatta tyyppiyritys toimii rakennusalaalla.

8.1 Pimeiden palkkojen määrä 2014–2020 eri menetelmillä laskettuna

8.1.1 Pimeät palkat yhteensä

Selvityksessä on arvioitu pimeän työn määrää usealla eri menetelmällä. Taulukossa 14 on yhteenvedo eri menetelmillä arvioidusta pimeän työn määrästä vuosina 2014–2020. Vaikka tulokset esitetään vuosittaisina estimaatteina, on tärkeää huomata, että niitä ei ole muodostettu aikasarjaperiaatteella. Näin ollen eri vuosien välillä nähtävät muutokset voivat johtua muistakin kuin ilmiöön liittyvistä syistä. Selvityksen keskeisin löydös onkin pimeiden palkkojen suuruusluokka tai niiden osuus palkkasummasta, ei niiden ajallinen kehitys.

Taulukko 13. Arviot pimeiden palkkojen määrästä vuosina 2014–2020 (miljoonaa euroa) eri menetelmillä laskettuna.

	2014	2015	2016	2017	2018	2019	2020
Substanssiperust. kaltaistaminen	991	828	738	673	720	--	--
CEM	--	--	--	780	797	801	814
DBSCAN	--	--	--	769	798	826	852
XGBoost	--	--	--	1 123	1 077	1 158	1 350
Bayesilainen logistinen regressio	--	--	--	751	748	855	1 010

Lähde: Verohallinto 5–8/2021.

Eri menetelmillä saaduista tuloksista voidaan todeta, että ne ovat pääosin samaa suuruusluokkaa eli noin 670–850 miljoonaa euroa vuodessa. Ainoastaan XGBoost-menetelmä tuotti estimaatteja, jotka olivat selvästi suurempia kuin muiden menetelmien tulokset. Lisäksi muiden menetelmien yksittäisten vuosien tulokset saattoivat poiketa mainitusta tasosta.

Substanssiperusteinen kaltaistaminen hyödyntää satunnaisverotarkastusten sisältämää informaatiota, kun se arvioi yritysten todennäköisyyttä maksaa pimeitä palkkoja. Lisäksi se pyrkii maltillistamaan riskiperusteisissa tarkastuksissa havaittujen virheiden suuruutta.⁶⁶ CEM ei

⁶⁶ Substanssiperusteinen kaltaistaminen korjaa keskimääräisen virheen harhaa trimmatulla jakaumalla. Ks. tarkemmin liite 2, luku 12.2.

hyödynnä satunnaistarkastuksia, mutta hyvin tarkka kaltaistaminen 10 muuttujan yli johti tähän nähden maltilliseen arvioon pimeiden palkkojen määrästä. Sekä DBSCAN että bayesilainen logistinen regressio hyödynsivät satunnaistarkastusten informaatiota ja siten onnistuivat pienentämään riskiperusteisten tarkastusten aiheuttamaa harhaa. XGBoost-menetelmä ei käytä satunnaistarkastuksia hyödykseen, joten se on altis riskiperusteisen valikoinnin aiheuttamalle harhalle.

Taulukossa 13 huomio kiinnittyy siihen, että kaikilla algoritmipohjaisilla menetelmillä laskettu pimeiden palkkojen määrä nousee vuonna 2020 suuremmaksi kuin muina vuosina. Vaikka trendin tarkastelu ei ole selvityksen tavoitteena, on tätä ilmiötä syytä hieman pohtia. Vuosi 2020 oli koronaviruspandemian vuoksi täysin poikkeuksellinen, ja Suomessakin käyttöön otetut laajat rajoitustoimet häiritsivät useiden yritysten toimintaa. Tämän seurauksena yritysten liikevaihto putosi ja moni yritys joutui hetkellisesti lomauttamaan koko henkilöstönsä. Tämän myötä monen yrityksen profiili muistuttaa enemmän aikaisempina vuosina verotarkastuksissa havaittuja riskiryhmiä, joiden eräinä piirteinä olivat alhainen liikevaihto ja palkat. Koska verotarkastusten perusteella päätellään tarkastamattomien yritysten riskisyys, johtaa tämä keinoitekoisesti korkeampaan riskiprofiiliin monelle tosiasiallisesti toimivalle yritykselle. Toistaiseksi on mahdotonta sanoa, kuinka paljon koronapandemian seurauksena yritykset aidosti lisäsivät harmaan talouden toimintaa sekä pimeiden palkkojen maksamista alentaakseen juoksevia kulujaan. Vuoteen 2020 kohdistuvien verotarkastusten valmistuessa asiasta on mahdollista saada lisää tietoa.

Pelkästään satunnaistarkastuksiin perustuvat substanssiperusteinen kaltaistaminen arvioi pimeiden palkkojen määräksi keskimäärin 460 miljoonaa euroa vuodessa ajalla 2014–2018. Tuloksen arvioitiin kuitenkin olevan todennäköisesti harhainen alaspäin (ks. luku 7.1.1). Lisäksi siihen tuo epävarmuutta satunnaistarkastusten vähäinen määrä.

Eri suuruiset arviot suhteutettiin palkkasummaan vuosina 2017–2018. Pimeiden palkkojen osuus maksetuista palkoista vaihteli menetelmän mukaan 1,4–1,5 prosentin välillä, pois lukien XGBoost, jonka antama tulos oli noin 2,1 prosenttia palkkasummasta.

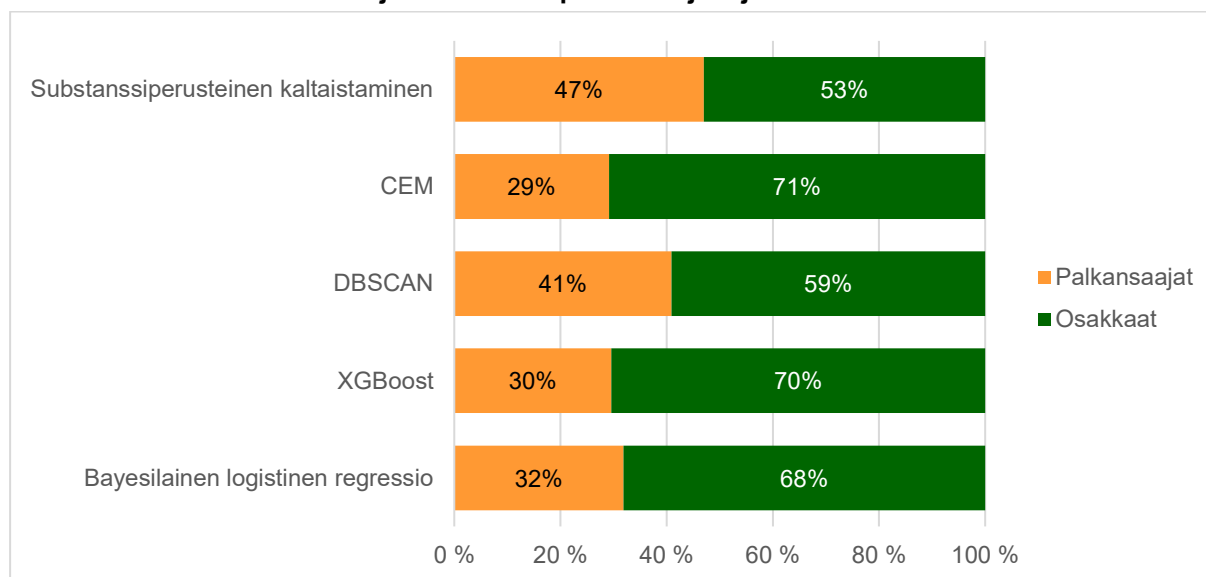
8.1.2 Osakkaiden ja palkansaajien pimeät palkat

Eri menetelmät tuottivat arviot myös osakkaille ja palkansaajille maksetuista pimeistä palkoista. Näiden suhteellinen jakautuminen on esitetty kuviossa 8 (seuraavalla sivulla). Kaikki menetelmät arvioivat osakkaiden osuuden suuremmaksi kuin palkansaajien osuuden. DBSCANin ja substanssiperusteisen kaltaistamisen perusteella ero ei ollut kovin suuri. Muut menetelmät arvioivat, että osakkaiden osuus pimeistä palkoista oli noin 70 prosenttia ja palkansaajien osuus noin 30 prosenttia.

Verotarkastuksissa sen sijaan havaittiin enemmän palkansaajien kuin osakkaiden pimeitä palkkoja. Esimerkiksi ajalla 2014–2018 verotarkastuksissa löydettyistä pimeistä palkoista 56 prosenttia kohdistui palkansaajille ja 44 prosenttia osakkaille. Samalla ajanjaksolla satunnaistarkastuksissa havaituista pimeistä palkoista 63 prosenttia kohdistui palkansaajille ja vain 37 prosenttia osakkaille. Satunnaistarkastukset tehtiin suurimmaksi osaksi yrityksiin, joiden liikevaihto oli yli 300 000 euroa vuodessa. Tämän selvityksen analyysien mukaan osakkaiden

pimeistä palkoista suurin osa maksettiin pienimmissä yrityksissä, mikä selittää pimeiden palkkojen erilaisen jakautumisen satunnaistarkastuksissa ja esitetyissä arvioissa.

Kuvio 8. Kokonaisestimaatin jakautuminen palkansaajien ja osakkaiden kesken.



Lähde: Verohallinto 5–8/2021.

Suuruusluokan hahmottamiseksi eri menetelmillä lasketut osakkaiden ja palkansaajien pimeät palkat suhteutettiin näille ryhmille maksettuihin palkkoihin vuosina 2017–2018. Osakkailla maksetut pimeät palkat olivat useimmilla mittaustavoilla noin 9–12 prosenttia osakkaiden palkkasummasta. XGBoost-menetelmän tulos poikkesi muista ja sen perusteella osakkaiden pimeät palkat muodostivat noin 17 prosentin osuuden osakkaiden palkkasummasta. Palkansaajille maksetut pimeät palkat puolestaan olivat 0,4–0,7 prosenttia palkansaajien palkkasummasta.

8.1.3 Toimimattomien yritysten pimeät palkat

Selvityksessä muodostettiin myös erillinen arvio niin sanottujen toimimattomien yritysten maksamista pimeistä palkoista. Toimimaton yritys ei kuulu Verohallinnon rekistereihin, mutta sillä on kuitenkin kaupparekisteriin merkitty y-tunnus. Verotarkastuksilla myös näillä, rekisteritietojen perusteella toimimattomilla yrityksillä, oli havaittu pimeää palkanmaksua.

Toimimattomien yritysten arvioidaan maksaneen vuosittain pimeitä palkkoja noin 70 miljoonaa euroa ajalla 2014–2018. Tästä arviolta noin puolet oli ulkomaisten yhteisöjen maksamia pimeitä palkkoja ja puolet suomalaisten osakeyhtiöiden maksamia pimeitä palkkoja.⁶⁷

Jotta toimimattomien yritysten pimeiden palkkojen suuruutta olisi helpompi hahmottaa, ne suhteutettiin perusjoukon kokoon eli jaettiin perusjoukon yritystunnusten lukumäärällä. Tällöin osakeyhtiöiden pimeät palkat olivat keskimäärin hieman alle 600 euroa yritystunnusta kohden ja ulkomaisten yhteisöjen pimeät palkat olivat keskimäärin noin 5 400 euroa yritystunnusta kohden. Vertailun vuoksi todettakoon, että toimivien osakeyhtiöiden arvioidut,

⁶⁷ Tarkastus- ja muiden aineistojen niukkuuden vuoksi, esitetty arvio on niin sanottu asiantuntija-arvio ja siihen liittyy huomattavan paljon epävarmuustekijöitä. Ks. luku 7.1.2.

pimeät palkat olivat keskimäärin noin 4 000 euroa yritystä kohden ja toimivien ulkomaisten yhteisöjen vastaava luku oli noin 12 000 euroa yritystä kohden.⁶⁸

Keskimääräiset luvut paljastavat, että pimeän työn riski arvioitiin paljon suuremmaksi toimimattomien ulkomaisten yhteisöjen keskuudessa kuin osakeyhtiöiden keskuudessa. Suhteellisesti eniten pimeää työtä arvoitiin esiintyvän sellaisten ulkomaisten yhteisöjen joukossa, jotka eivät hakeudu tunnuksen perustamisvaiheessa Verohallinnon rekistereihin. Lähes puolet arvioidusta pimeistä palkoista kohdistui tähän ryhmään (noin 35 miljoonaa euroa vuodessa). Muissa ryhmissä pimeiden palkkojen määrä arvioitiin vähäiseksi, mikä on havaittavissa myös edellä mainitusta yrityskohtaisesta virheestä (alle 600 euroa / yritystunnus).

8.2 Pimeistä palkoista aiheutuneet verojen ja muiden maksujen menetykset

Pimeästä työnteosta aiheutuvia julkisten maksujen menetyksiä voidaan arvioida prosenttiosuuksina pimeiden palkkojen määrästä. Aiemmin luvussa 2.3 on kuvattu menetettyjen verojen ja sosiaalivakuutusmaksujen laskemisessa tarvittavat prosenttiosuudet ja perustelut näille.⁶⁹ Julkisten maksujen menetykset, joiden suuruutta tässä luvussa arvioidaan, sisältävät osakkaiden ja palkansaajien ennakonpidätykset sekä palkansaajien sosiaalivakuutusmaksut. Menetettyjä osakkaiden sosiaalivakuutusmaksuja ei ole selvityksessä arvioitu (katso tarkemmin: liite 1).

Pimeästi maksetuista palkoista aiheutuneet verojen ja sosiaalivakuutusmaksujen menetykset olivat noin 260–300 miljoonaa euroa vuodessa ajalla 2017–2020. Tulos perustuu kolmen algoritmipohjaisen menetelmän estimaatteihin.⁷⁰ XGBoost-menetelmä arvioi pimeiden palkkojen määrän suuremmaksi kuin muut menetelmät. Näin ollen tähän menetelmään perustuva arvio julkisten maksujen menetyksestä oli hieman muita suurempi eli noin 390 miljoonaa euroa vuodessa ajalla 2017–2020.

Pimeiden palkkojen jakautuminen osakkaiden ja palkansaajien kesken vaikuttaa veron menetyksen arvioon, sillä palkansaajien palkoista maksettavat ennakonpidätykset ja sosiaalivakuutusmaksut ovat huomattavasti suuremmat kuin osakkaiden palkoista maksettavat ennakonpidätykset. Substanssiperusteinen kaltaistaminen arvioi palkansaajien osuuden pimeiden palkkojen määrästä olevan lähes puolet. Tällä menetelmällä arvioiden julkisia maksuja menetettiin noin 300 miljoonaa euroa vuodessa ajalla 2014–2018.

Kuviossa 9 (seuraavalla sivulla) on esitetty pimeistä palkoista aiheutuneiden julkisten maksujen suuruus eri menetelmillä mitattuna sekä kuinka tulokset jakautuivat osakkaiden ja palkansaajien ennakonpidätysten ja palkansaajien sosiaalivakuutusmaksujen kesken. Kuvioista havaitaan jo edellä mainittu seikka, eli XGBoost-menetelmällä arvio menetetyistä julkisista maksuista on hieman suurempi kuin muiden menetelmien arviot. Tätä selittää pääasiassa osakkaiden ennakonpidätykset. Kuvioista erottuu myös, että substanssiperusteisen kaltaistamisen

⁶⁸ Toimivien yritysten arvio on laskettu substanssiperusteisen kaltaistamismenetelmän pohjalta.

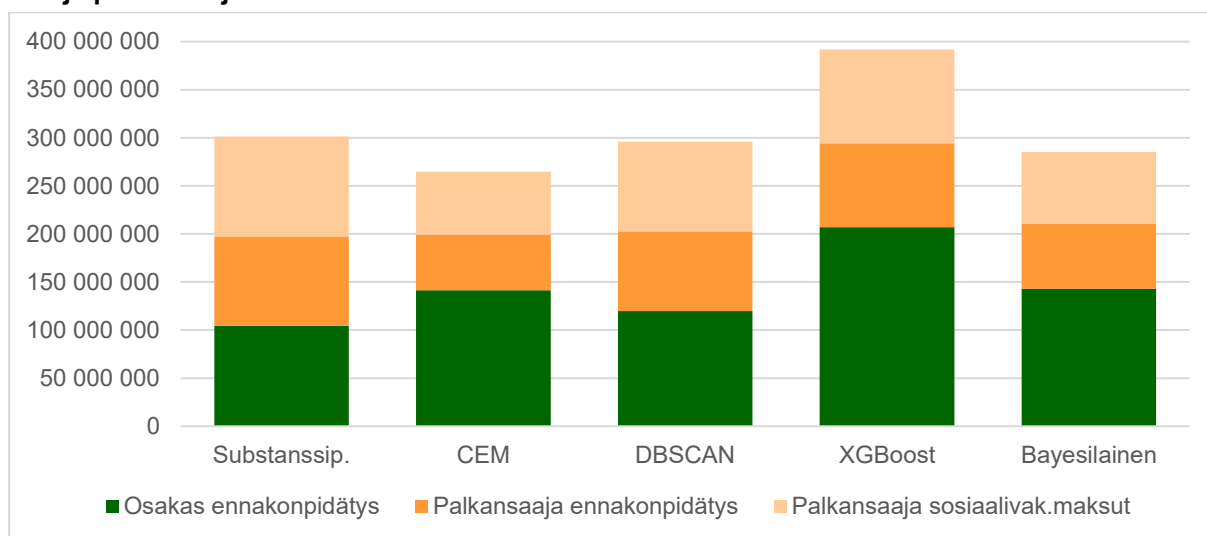
⁶⁹ Palkansaajat: menetetyt ennakonpidätykset = 25 prosenttia pimeistä palkoista; menetetyt sosiaalivakuutusmaksut 28 prosenttia pimeistä palkoista. Osakkaat: menetetyt ennakonpidätykset = 25 prosenttia pimeistä palkoista.

⁷⁰ Tulos perustuu seuraaviin menetelmiin: CEM, DBSCAN ja bayesilainen logistinen regressio.

arvio menetetyistä osakkaiden ennakonpidätyksistä on muita pienempi. Arviot maksamatta jääneistä osakkaiden ennakonpidätyksistä vaihtelivat noin 100–200 miljoonan euron välillä, mutta jos tarkastelusta poistetaan edellä mainitut ääripäät, vaihteluväli supistuu 120–140 miljoonaan euroon.

Eri menetelmien arviot palkansaajien maksamatta jääneistä ennakonpidätyksistä ja sosiaalivakuutusmaksuista liikkuvat 125–195 miljoonan euron välillä. Tässäkin substanssiperusteinen kaltaistaminen⁷¹ ja XGBoost⁷² poikkeavat muista tuloksista, ja ilman näitä vaihteluväli rajautuu 125–175 miljoonaan euroon.

Kuvio 9. Pimeästi maksetuista palkoista aiheutuneet ennakonpidätysten ja sosiaalivakuutusmaksujen menetykset (keskimääräinen vuosiarvio euroina). Menetysten jakautuminen osakkaiden ja palkansaajien kesken. Eri menetelmien tulokset.



Lähde: Verohallinto 5–8/2021.

8.3 Tyypillinen pimeän työn yritys

Tässä luvussa esitellään, millainen oli tyypillinen pimeää työtä teettänyt tai osakkaalle peiteltyä osinkoa maksanut yritys. Kuvaus perustuu suurimmaksi osaksi XGBoost-mallilla saatuihin tuloksiin. Menetelmä tuotti suurimmat euromääräiset estimaatit, joten yksittäin tarkasteltuna menetelmän arviot ovat ylöspäin harhaisia. Menetelmä kuitenkin tuotti arvokasta tarkan tason tietoa muun muassa pimeän palkan esiintyvyydestä toimialoittain niin palkansaajien kuin osakkaiden pimeistä palkoista.

XGBoost tuottaa kullekin yritykselle todennäköisyysarvion sen pimeiden palkkojen maksamisesta. Valitsemalla sopiva piste tältä todennäköisyysjanelta voimme luokitella yritykset harmaisiin ja valkoisiin. Mallin mukaan harmaalla yrityksellä pimeän palkanmaksun riski on kohonnut ja valkoisella yrityksellä puolestaan riski on vähäinen. XGBoost-malliin perustuva kuva pimeitä palkkoja maksavista yrityksistä on kuva juuri näistä kohonneen riskin harmaista yrityksistä. Luokittelun lopputulos on erehtyväisen mallin arvio kustakin yrityksestä, eikä

⁷¹ Suurempi palkansaajien osuus.

⁷² Oletetusti ylöspäin harhainen estimaatti.

täysin kuvaa todellisuutta. On todennäköistä, että malli toimii riittävän hyvällä tasolla, jotta se kuvaa pimeää palkkaa maksavien yrityksen olennaisia ominaispiirteitä.

Yrityspopulaatiosta vuosina 2017–2019 noin yksitoista prosenttia eli noin 22 000 yritystä luokiteltiin harmaiksi riskiyrityksiksi. Vuonna 2020 pimeää palkkaa maksavien yritysten lukumääräksi arvioitiin noin 26 000. Lukumäärät ovat ainoastaan suuntaa antavia, sillä ne ovat hyvin riippuvaisia subjektiivisesti valitusta todennäköisyysjakauman raja-arvosta.⁷³ Oleellisempaa kuin muodostaa tarkka lukumääräarvio, onkin rajata yritysjoukko, jonka avulla voimme muodostaa kuvan pimeän palkanmaksun riskiyrityksistä.

Luonnollisesti harmaiksi luokiteltujen yritysten joukkoon mahtuu paljon yrityksiä, jotka ovat todellisuudessa täysin valkoisia ja toisaalta moni tosiasiaassa pimeitä palkkoja maksava yritys tulee luokitelluksi valkoiseksi. Pimeän työn riskiyritykseksi tällä tavalla luokiteltu yritys on kuitenkin todennäköisemmin myös aidosti pimeitä palkkoja maksava yritys kuin oikein toimiva.

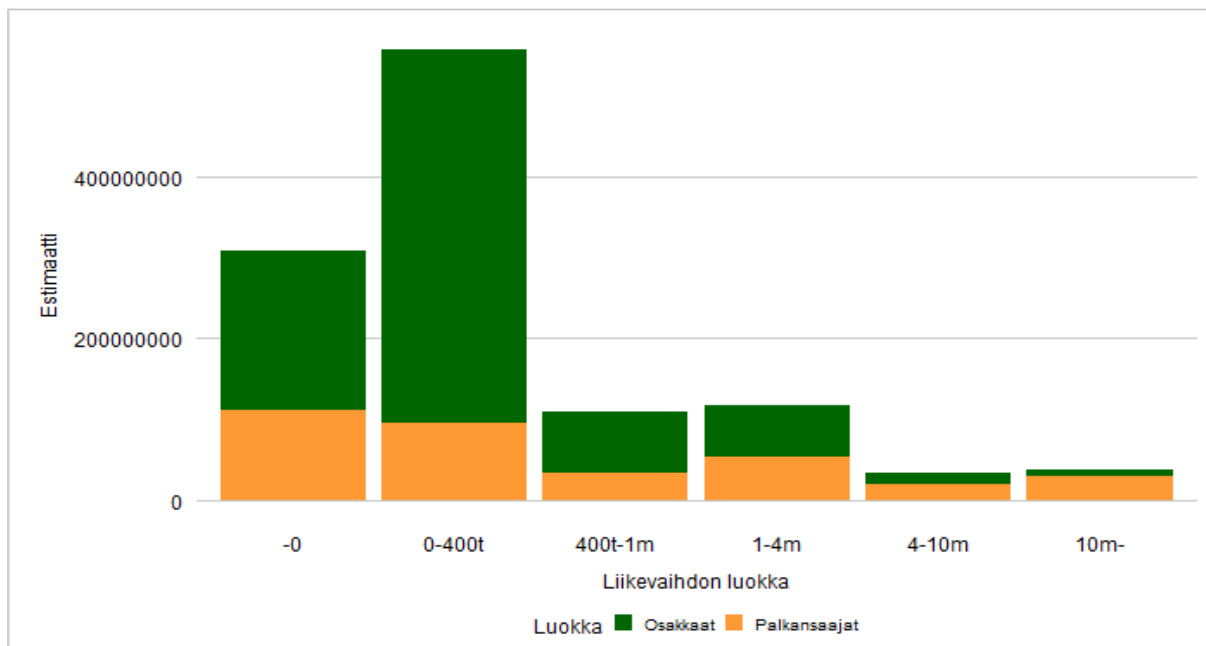
8.3.1 Pimeän työn yritysten koko liikevaihtoluokittain

Kuviosta 10 selviää arvioitujen pimeän työn yritysten jakautuminen liikevaihtoluokittain. XGBoost-menetelmän tuottamasta pimeiden palkkojen vuoden 2019 estimaatista vajaa puolet tulee yrityksiltä, joilla liikevaihtoa on alle 400 000 euroa. Myös yrityksille, joiden liikevaihto on tuntematon, on päätelty iso pimeiden palkkojen estimaatti. On selvää, että menetelmän tuottama liikevaihdottomien pimeän palkan estimaatti on todellisuutta huomattavasti suurempi⁷⁴. Alle 400 000 euron liikevaihdon yrityksillä pimeää palkkaa arvioidaan maksettavan erityisesti osakkaille itselleen. Työntekijöille maksettavien pimeiden palkkojen määrä on pienissä yrityksissä maltillisempi. Osakkaan itselleen maksaman pimeän palkan riski korostuu erityisesti pienissä osakeyhtiöissä, joissa on vain yksi osakas. Sen sijaan yli miljoonan euron liikevaihtoa tekevissä yrityksissä ei arvioida maksettavan merkittäviä määriä pimeitä palkkoja, mutta suhteellisesti isompi osuus menee näissä yrityksissä palkansaajille kuin osakkaille. Tarkastelu koskee ainoastaan vuotta 2019, mutta tulokset ovat pitkälti yleistettävissä koko ajanjaksoon 2017–2020. Liikevaihdon kokoluokittelu perustui ALV-myyntin perusteella laskettuun liikevaihtoon ja/tai veroilmoituslomakkeen 6B tietoihin.

⁷³ Valitsimme luokittelun raja-arvoksi arvon 0,6, jonka yli menevät yritykset luokitellaan harmaiksi ja alle jäävät valkoisiksi.

⁷⁴ Yritykset, joilla ei ole liikevaihtoa tai liikevaihto on pienempi kuin nolla, on haasteellinen tarkasteltavien joukko, koska se sisältää myös paljon osakeyhtiöitä, joilla ei ole toimintaa. Menetelmä laskee havaittujen tapausten esiintyvyydellä ja virheiden suuruudella arvion myös aidosti toimimattomille, ja siksi estimaatti muodostuu tässä joukossa huomattavan suureksi.

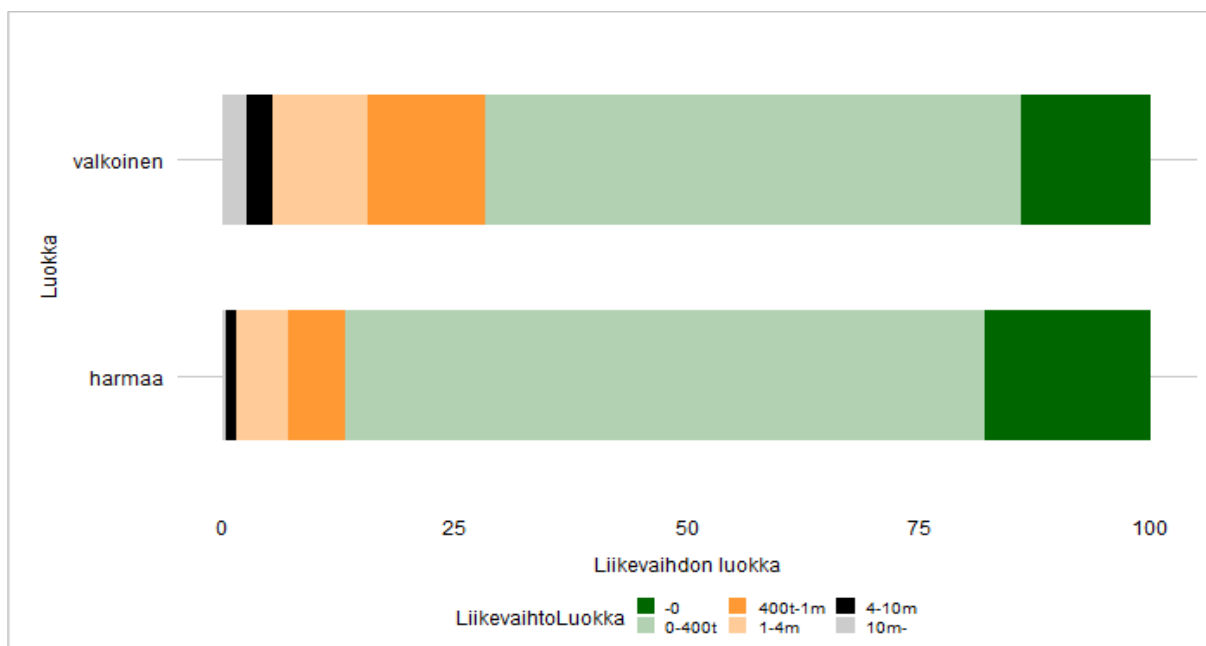
Kuvio 10. Pimeiden palkkojen estimaatin jakautuminen liikevaihtoluokkiin niitä maksaneiden yritysten mukaan vuonna 2019.



Lähde: Verohallinto 5–8/2021.

Kuviossa 11 on esitetty, kuinka harmaat ja valkoiset yritykset jakautuvat eri liikevaihtoluokkiin ja siitä havaitaan, että verrattaessa valkoisiin yrityksiin, harmaissa riskiyrityksissä korostuu pieni liikevaihto. Suhteellisesti isompi osa harmaista kuin valkoisista yrityksistä kuuluu kahden pienimpään liikevaihtoluokkaan, minkä jälkeen valkoisia yrityksiä on suhteellisesti enemmän. Erityisen iso ero on 4–10 miljoonan sekä yli kymmenen miljoonan liikevaihtoluokissa, joihin kuuluu hyvin pieni osuus harmaista yrityksistä. Valkoisista yrityksistä yli neljänneksellä on yli 400 000 euron liikevaihto, harmaista vain noin kymmenellä prosentilla.

Kuvio 11. Yritysten koko – harmaiden ja valkoisten yritysten suhteellinen jakautuminen liikevaihtoluokkiin vuonna 2019.



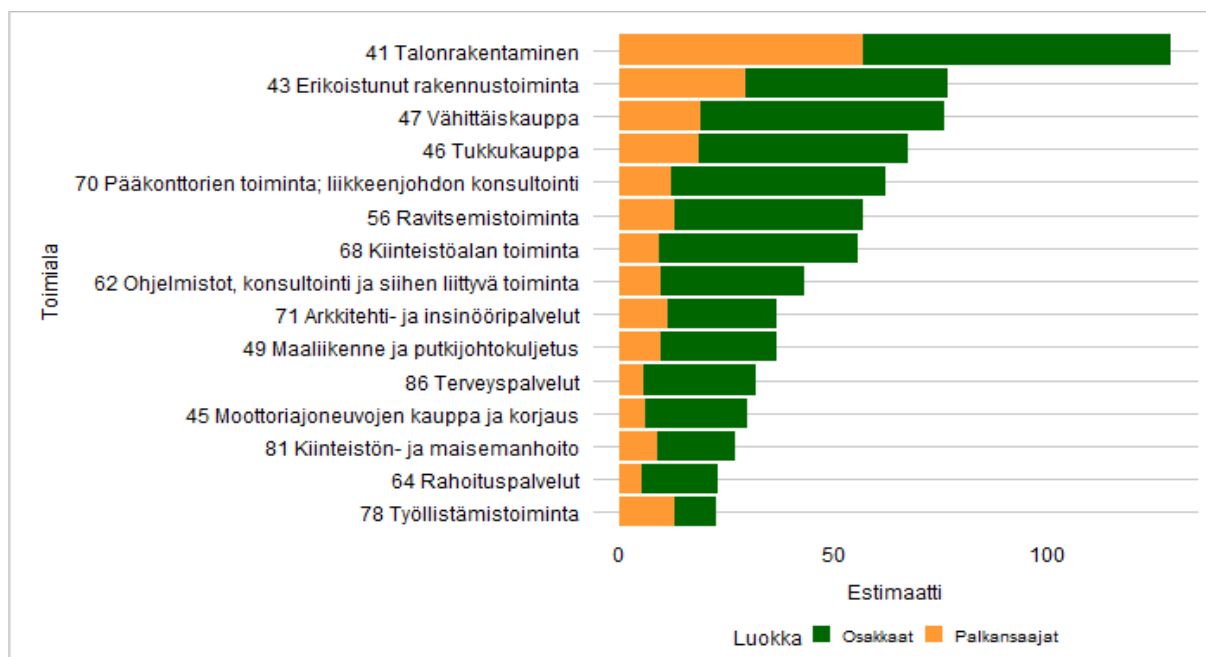
Lähde: Verohallinto 5–8/2021.

8.3.2 Pimeän työn yritykset toimialoittain

Toimialoittain kaikista suurimman euromääräisen pimeiden palkkojen arvion XGBoost-menetelmä päättää talonrakentamisen toimialalle ja toiseksi suurin estimaatti on erikoistuneen rakennustoiminnan toimialalla. Tulos on odotettu, sillä rakennusala on tunnetusti pimeän työn teettämisen suhteen ongelmallinen ala. Lisäksi rakennusala on suuri toimiala, jolla toimii paljon yrityksiä ja siten myös yhteenlaskettuna pimeiden palkkojen arvio on suuri. Talonrakentamisen toimialalla arvio koostuu lähes puoliksi osakkaiden ja työntekijöiden pimeistä palkoista, kun taas erikoistuneen rakennustoiminnan puolella korostuvat osakkaiden pimeät palkat eli peitelty osinko. Kolmas ja neljäs sija kuuluvat vähittäis- ja tukkukaupalle⁷⁵, jotka niin ikään ovat toimialoja, joissa on paljon yrityksiä. Myös liikkeenjohdon konsultoinnin, ravintola-alan sekä kiinteistöalan toimialoille estimoidaan yli 50 miljoonan summa pimeänä maksettuja palkkoja ja peiteltyjä osinkoja.

Arviot koskevat vuotta 2019, mutta ovat yleistettävissä myös muille vuosille. Kuviossa 12 on nähtävissä, että riskisimmiksi toimialoiksi nousevat suuret toimialat, joissa myös riskiyritysten esiintyvyys on suuri.

Kuvio 12. Pimeiden palkkojen ja peitelty osingon estimaatti toimialoittain (miljoonaa euroa) vuonna 2019. Toimialan pimeiden palkkojen jakautuminen osakkaiden ja palkansaajien kesken.



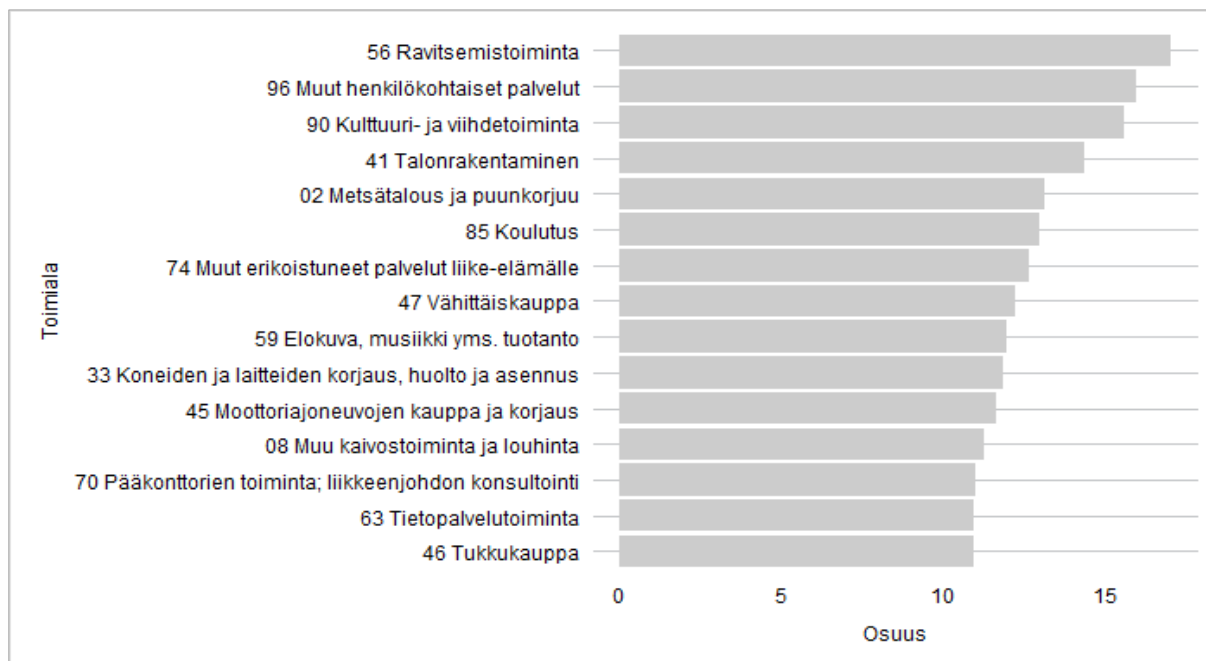
Lähde: Verohallinto 5–8/2021.

Kun tarkastellaan työntekijöiden pimeiden palkkojen ja osakkaiden pimeiden palkkojen euro-määrien sijaan harmaiden yritysten suhteellista osuutta koko toimialan yrityskannasta

⁷⁵ Harmaan talouden selvitysyksikön aiemmin laatimassa selvityksessä *Arvonlisäveron virheet - Maksetun arvonlisäveron vaje 1/2021* niin ikään puuttuvaa myyntiä arvioitiin esiintyvän osakeyhtiöissä erityisesti esiin kaupan toimialoilta (45–47) ja talonrakentamisessa (41). Luokat 46 (tukkukauppa) ja 47 (vähittäiskauppa) sisältävät kaikki muut myyntiin liittyvät toiminnot kuin ajoneuvojen kaupan ja korjaamisen. Luokkien 46 ja 47 ero perustuu asiakastyypin erilaisuuteen. Vähittäiskauppaan kuuluu uusien ja käytettyjen tavaroiden myynti pääasiassa yksityisille kuluttajille henkilökohtaiseen tai kotitalouskäyttöön.

(kuvio 13)⁷⁶, kaikista ongelmallisimpana toimialana näyttäytyy ravitsemistoiminta, jonka yrityksistä yli 15 prosenttia arvioidaan pimeiden palkkojen maksamisen suhteen riskiyrityksiksi. Yli viidentoista prosentin osuuden pääsee myös muut henkilökohtaiset palvelut⁷⁷ sekä kulttuuriala⁷⁸. Talonrakentamisen toimiala näyttäytyy myös riskiyritysten osuuden suhteen ongelmallisena toimialana, mutta erikoistunut rakennustoiminta ei suhteessa toimialan isoon kokoon nähden näyttäydy enää yhtä riskisenä eikä se mahdu 15 toimialan listalle lainkaan.

Kuvio 13. Harmaiden yritysten suhteellinen osuus toimialan yrityksistä.



Lähde: Verohallinto 5–8/2021.

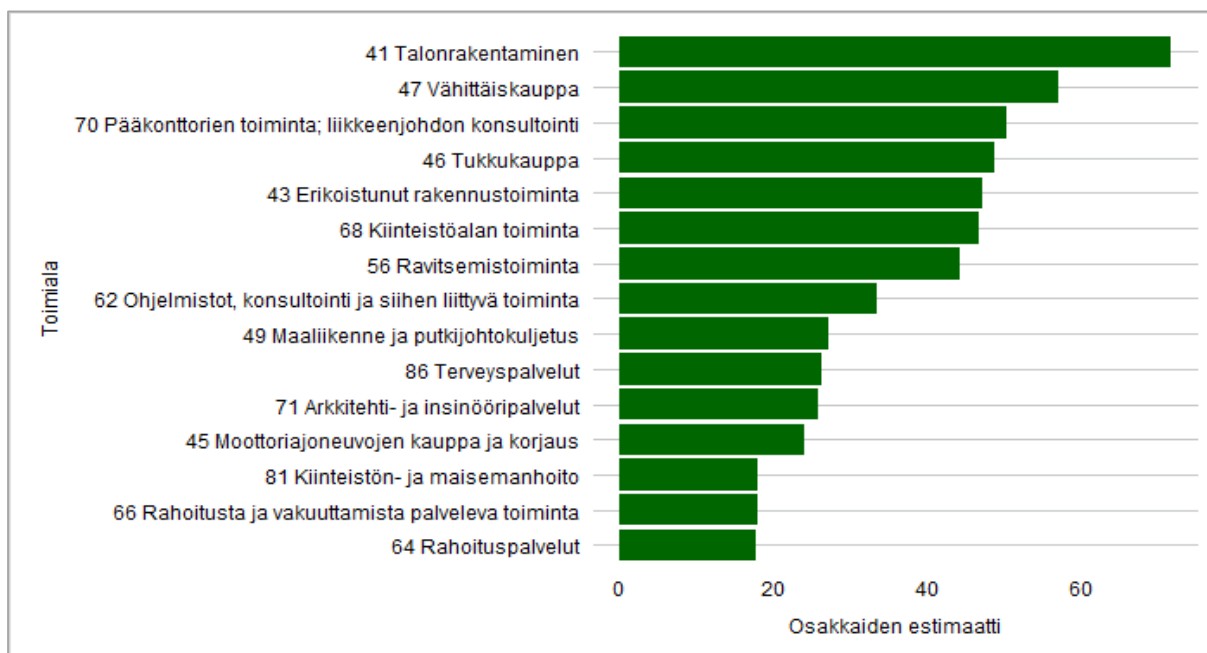
Tarkasteltaessa osakkaille estimoidun pimeän palkan euromääriä toimialoittain (kuvio 14) havaitaan, että rakennusalan yrityksissä on todennäköisesti eniten osakkaiden pimeitä palkkoja. Rakennusala on suuri toimiala, joten siksi myös arvioitu pimeän työn määrä on suuri. Myös vähittäiskauppa sekä tukkukauppa ovat suuria toimialoja, ja niissä niin ikään esiintyy estimaattien mukaan huomattavan suuret osakkaiden pimeät palkat.

⁷⁶ Tarkastelussa ei ole huomioitu toimialoja, joilla on yhteensä alle 500 yritystä, jotta kovin pieni määrä harmaita riskiyrityksiä ei nosta toimialaa esille.

⁷⁷ Toimiala 96 Muut henkilökohtaiset palvelut käsittävät muun muassa muualle luokittelemattomat palvelut, joista tunnetuimpina esimerkiksi parturit ja kampaamot sekä kauneudenhoitopalvelut.

⁷⁸ Toimiala 90 Kulttuuri- ja viihdetoiminta käsittää muun muassa kulttuuri- ja viihdepalvelujen tuottamisen: esitykset, näyttelyt, tapahtumat sekä näihin liittyvä avustavan ja promootiotoiminnan. Tähän kuuluu myös kulttuuri- ja viihdelaitosten toiminta.

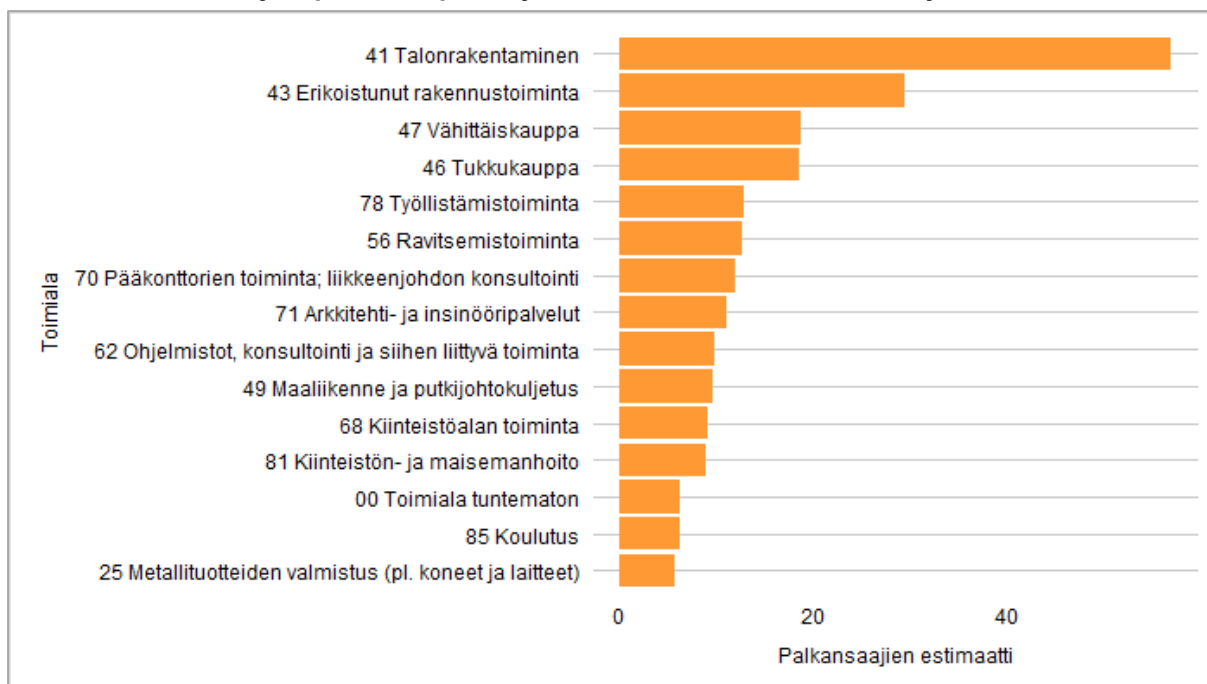
Kuvio 14. Osakkaiden peitellyn osingon (pimeän työn) estimaatti toimialoittain / miljoonaa euroa



Lähde: Verohallinto 5–8/2021.

Tutkittaessa työntekijöille maksettujen pimeiden palkkojen esiintyvyyttä ilman osakkaiden pimeitä palkkoja (kuvio 15) nousee riskisimpiin toimialoihin melko korkealle (viidenneksi) myös työllistämistoiminta (toimiala 78). Työllistämistoiminta alana käsittää muun muassa työnvälityksen, työvoiman vuokrauksen sekä muut henkilöstön hankintapalvelut. Tämä on ymmärrettävää, koska kyseinen toimiala vuokraa työntekijöitä monelle työvoimavaltaiselle alalle (rakennus-, siivous-, ravitsemusala ym.), joissa pimeää palkanmaksia tunnetusti esiintyy.

Kuvio 15. Palkansaajien pimeiden palkkojen estimaatti toimialoittain / miljoonaa euroa



Lähde: Verohallinto 5–8/2021.

9 Yhteenveto ja johtopäätökset

Tässä selvityksessä on tarkasteltu tiettävästi ensimmäistä kertaa pimeää työtä Suomessa siten, että ilmiön euromääräistä laajuutta on arvioitu kattavasti. Data-analyysia on taustoitettu pimeää työtä koskevien kotimaisten ja EU-alueella tehtyjen tutkimusten esittelyllä.

Selvityksessä tehdyn laajan data-analyysin perusteena on ollut vuosina 2014–2020 tehdyt yhteensä noin 7 500 verotarkastusta. Selvityksessä pimeän työn kokonaismäärää on arvioitu neljällä eri menetelmällä, joista ensimmäinen on ns. substanssiperusteinen kaltaistamismalli, toinen tekninen kaltaistamismalli, kolmas klusterointimalli ja neljäntenä menetelmänä on päätelty yrityskohtainen pimeiden palkkojen riskiarvio kahdella eri tavalla. Näin toimimalla pystyttiin tuottamaan hyvin kattava ja monipuolinen kokonaisarvio pimeän työn määrästä. Mukana olleet satunnaisverotarkastukset mahdollistivat varsin luotettavien arvioiden tekemisen ilman riskiperusteisten tarkastusten sisältämää kohdentamisharhaa.

Eri menetelmien tulosten perusteella voidaan arvioida, että osakeyhtiöiden palkansaajien ja yrittäjäosakkaiden pimeiden palkkojen euromäärä on Suomessa vuosittain noin 670–850 miljoonaa euroa. Menetettyjen veroeurojen määrä riippuu käytettävästä veroprosentista, mutta esimerkiksi 25 prosentin veroasteella olisi kyseessä 260–300 miljoonan euron vuotuiset ennakonpidätysten ja sosiaalivakuutusmaksujen menetykset. Voidaankin todeta, että kokonaisuutena pimeän työn aiheuttamat lakisääteisten maksujen menetykset pelkästään osakeyhtiöiden kohdalla ovat varsin merkittävät.

Tyypillinen pimeitä palkkoja maksanut yritys on varsin pieni, jonka liikevaihto on alle 400 000 euroa vuodessa. Pimeiden palkkojen määrät ovat keskimäärin varsin maltillisia jakaantuen suuremmalta osin osakkaille ja alle 50 prosenttia osuudelta palkansaajille. Monista laki- ja kontrollitoimista huolimatta pimeän työn tyyppiyritys toimii rakennusalalla. Muita pimeän työn kannalta merkittäviä toimialoja ovat esimerkiksi vähittäis- ja tukkukauppa, liikkeenjohdon konsultointi, ravintola-ala sekä kiinteistöala.

Tämä selvityksen tarkoituksena on ollut pimeän työn ilmiön laajuuden arviointi. On hyvä muistaa, että pimeä työ on vain yksi osa harmaan talouden laajasta kokonaisuudesta. Pimeää työtä tarkastelemalla ei arvioida harmaan talouden kokonaismäärää vaan vain yhtä osaa siitä. EU-maiden välisessä vertailussa pimeä työ näyttäytyy Suomessa kohtuullisen pienenä ilmiönä. Silti pimeä työ on euromääräisesti merkittävää myös Suomessa, joten keinoja sen pienentämiseksi tulee pohtia erikseen.

Selvityksen perusteella pimeän työn tehokkaan torjunnan toteuttamiseksi on tärkeää, että harmaan talouden toimenpideohjelmat jatkuisivat myös seuraavilla hallituskausilla. Pimeän työn torjuminen edellyttää jatkossakin monenlaisia toimenpiteitä. Näitä ovat esimerkiksi terveen kilpailun ja työmarkkinoiden edistäminen, harmaan talouden ennalta estämiseen liittyvät toimenpiteet, viranomaisten toimintaedellytyksien turvaaminen sekä viranomaisyhteistyön kehittäminen.

10 Lähteet

- Aluehallintovirasto, E.-S. t. (2. joulukuu 2021). *Harmaa talous & talousrikollisuus*. Noudettu osoitteesta Aluehallintovirasto valvoo ulkomaalaisten työntekoa Suomessa: <https://www.vero.fi/harmaa-talous-rikollisuus/torjunta/torjuntatilastot/ty%C3%B6nantajat-ja-tilaajavastuu/>
- Ansamäki, P. (2017). *Sujuva verotarkastus* (1. painos p.). Alma Talent Oy. Haettu 28. 9 2021
- Bjørneby, M.; Alstadsæter, A.; & Telle, K. (2018). *Collusive Tax Evasion by Employers and Employees: Evidence from a Randomized Field Experiment in Norway*. München: Munich Society for the Promotion of Economic Research - CESifo.
- Eläketurvakeskus. (2020). *Kuinka yrittäjät eläkevakuuttavat? Yrittäjien omat näkemykset ja YEL-työtulon ja yrittäjän tulojen suhde*. Eläketurvakeskus.
- Eläketurvakeskus. (ei pvm). YEL yrittäjät. Haettu 4. 11 2021 osoitteesta <https://www.etk.fi/suomen-elakejarjestelma/elaketurva/tyoelaketurvan-kattavuus-ja-vakuuttaminen/yel-yrittajat/>
- Euroopan komissio. (2020). *Special Eurobarometer 498: Undeclared Work in the European Union*. Euroopan komissio.
- Euroopan Unionin neuvosto. (1. tammikuu 2019). *Pimeän työn vastainen EU-foorumi*.
- European Platform Undeclared Work. (2017). *Member State Factsheets 2017*. Euroopan Komissio.
- Harju, Jysmä, Koivisto & Kosonen. (2021). *Kotitalousvähennyksen vaikutukset kotitalouspalvelujen kulutukseen, työllisyyteen ja harmaaseen talouteen*. Valtion taloudellinen tutkimuskeskus.
- Harmaan talouden selvitysyksikkö. (2011). *Henkilöstöpäiväkirja - Personliggare*.
- Harmaan talouden selvitysyksikkö. (2014). *Harmaan talouden tutkimuksissa käytetyistä menetelmistä ja niiden avulla saaduista tuloksista EU:n jäsenvaltioissa*. Verohallinto.
- Harmaan talouden selvitysyksikkö. (2017). *Rakentamisen tiedonantovelvollisuus urakka ja työntekijätiedoista- lainsäädännön vaikuttavuus*. Verohallinto.
- HE 114/2006 vp. (ei pvm). *Hallituksen esitys eduskunnalle laiksi tilaajan selvitysvelvollisuudesta ja vastuusta ulkopuolista työvoimaa käytettäessä*. Helsinki. Haettu elokuu 2022
- HE 134/2017 vp. (ei pvm). *Hallituksen esitys eduskunnalle laiksi tulotietojärjestelmästä ja eräksi siihen liittyviksi laeiksi*.
- HE 24/2013 vp. (ei pvm). *Hallituksen esitys eduskunnalle laeiksi työ sopimuslain 2 luvun 16 §:n sekä kirjanpitolain 2 luvun ja 8 luvun 4 §:n muuttamisesta*.

- HE 58/2011 vp. (ei pvm). *Hallituksen esitys Eduskunnalle laiksi veronumerosta ja rakennusalan veronumerorekisteristä sekä eräksi siihen liittyviksi laeiksi.*
- HE 92/2012 vp. (ei pvm). *Hallituksen esitys eduskunnalle laeiksi verotusmenettelystä annetun lain ja työturvallisuuslain muuttamisesta.*
- Hirvonen, M. (2012). *Ulkomaista työvoimaa koskevan sääntelyn toimivuus.* Poliisihallitus.
- Ilmarinen. (3. marraskuu 2021). *YEL-työtulo vakuutuksen perustana.* Noudettu osoitteesta <https://www.ilmarinen.fi/yrittaja/yel-vakuutus/yrittajan-tyotulo/>
- Lith, P. (2010). *Rakentamisen harmaa talous; Tilastonäkemys rakennusalan harmaan talouden painopistealueista.*
- Länsisaari, A. (10. 3 2020). Verotarkastus – mikä se on? Tilisanomat. Haettu 30. 9 2021 osoitteesta <https://tilisanomat.fi/verotus/verotarkastus-mika-se-on>
- Riksdagstryckeriet. (2019). *En utvärdering av personalliggarsystemet.* Tuklholma: Riksdagstryckeriet.
- Skatteverket. (2020). *Svarta arbetsinkomster: beräkningar baserade på Skatteverkets revisioner, Skatte-verket 2020.* Skatteverket.
- Skatteverket. (8. Helmikuu 2022). Skattefelsrapport 2021. Ruotsi. Haettu 17. 11 2022 osoitteesta <https://www.skatteverket.se/omoss/varverksamhet/rapporterremissvarochskrivelser/rapporter/2022.4.339cd9fe17d1714c0773a5e.html>
- Sosiaali- ja Terveysministeriö, S. (ei pvm). Sosiaalivakuutusmaksut. Helsinki, Suomi. Haettu 14. 10 2021 osoitteesta <https://stm.fi/sosiaalivakuutusmaksut>
- T-Media. (2017). *Duuri työtutkimus 2017.* T-Media.
- Työ- ja elinkeinoministeriö. (2012). *Harmaa talous ja yritysten hallinnollinen taakka majoitus- ja ravitsemisalla – työryhmän loppuraportti.* Työ- ja elinkeinoministeriö.
- Työ- ja elinkeinoministeriö. (2014). *TEM raporteja 1/20214. Selvitys tilaajavastuulain uudistamiseksi tehtyjen ehdotusten vaikutuksista.* Työ- ja elinkeinoministeriö.
- Ukko.fi. (2016). *UKKO.fi käyttäjäkysely kesä 2016.* Ukko.fi.
- Valtiontalouden tarkastusvirasto. (2021). *Tulorekisterin käyttöönotto ja vaikutukset.* Valtiontalouden tarkastusvirasto.
- Valtionvarainministeriö. (23. toukokuu 1997). He 85/1997 vp. *Hallituksen esitys eduskunnalla laiksi kotitaloustyön väliaikaisesta verotuesta.*
- Verohallinto. (22. 1 2018). Hyvä verotarkastustapa. Helsinki, Suomi. Haettu 28. 9 2021 osoitteesta https://www.vero.fi/syventavat-vero-ohjeet/ohje-hakusivu/49204/hyva_verotarkastustap/

Williams, C. C. (2019). Tackling undeclared work in the European union: an evaluation of government policy approaches. *UTMS Journal of Economics*, 135-147.

11 Luettelo kuvioista ja taulukoista

Kuviot

Kuvio 1. Suomen valtion ja kuntien verotulot ja pakolliset sosiaalivakuutusmaksut vuonna 2020, miljoonaa euroa.	19
Kuvio 2. Selvityksessä mukana olevien verotarkastusten lukumäärä ajalla 2014–2020. Satunnaistarkastukset ja muut kokonaistarkastukset.	26
Kuvio 3. Työntekijöiden palkkoja koskevat virrehavainnot verotarkastuksissa. Virheluokkien osuudet tarkastushavaintojen euromääristä.	27
Kuvio 4. Osakkaiden palkkoja koskevat virrehavainnot verotarkastuksissa. Virheluokkien osuudet tarkastushavaintojen euromääristä.	28
Kuvio 5. Pimeiden palkkojen määrä ja osuus palkkasummasta vuosittain ajalla 2014–2018.	32
Kuvio 6. Osakkaiden ja palkansaajien ilmoitetut palkat ja arvio pimeistä palkoista 2017–2018.	33
Kuvio 7. XGBoost-mallin tuottamat pimeän työn todennäköisyysjakaumat tarkastetuille ja tarkastamattomille yrityksille.	39
Kuvio 8. Kokonaisestimaatin jakautuminen palkansaajien ja osakkaiden kesken.	44
Kuvio 9. Pimeästi maksetuista palkoista aiheutuneet ennakonpidätysten ja sosiaalivakuutusmaksujen menetykset (keskimääräinen vuosiarvio euroina). Menetysten jakautuminen osakkaiden ja palkansaajien kesken. Eri menetelmien tulokset.	46
Kuvio 10. Pimeiden palkkojen estimaatin jakautuminen liikevaihtoluokkiin niitä maksaneiden yritysten mukaan vuonna 2019.	48
Kuvio 11. Yritysten koko – harmaiden ja valkoisten yritysten suhteellinen jakautuminen liikevaihtoluokkiin vuonna 2019.	48
Kuvio 12. Pimeiden palkkojen ja peitellyn osingon estimaatti toimialoittain (miljoonaa euroa) vuonna 2019. Toimialan pimeiden palkkojen jakautuminen osakkaiden ja palkansaajien kesken.	49
Kuvio 13. Harmaiden yritysten suhteellinen osuus toimialan yrityksistä.	50
Kuvio 14. Osakkaiden peitellyn osingon (pimeän työn) estimaatti toimialoittain / miljoonaa euroa.	51
Kuvio 15. Palkansaajien pimeiden palkkojen estimaatti toimialoittain / miljoonaa euroa.	51

Taulukot

Taulukko 1. Palkan sivukulut ja niiden suuruus vuonna 2021.	6
Taulukko 2. Esimerkki 40 000 euron vuosipalkkaan liittyvistä sosiaalivakuutusmaksuista, jotka työnantajan tulisi maksaa.	21
Taulukko 3. Esimerkki 40 000 euron vuosipalkkaan liittyvistä sosiaalivakuutusmaksuista, jotka työnantajan tulisi maksaa työntekijän puolesta.	22
Taulukko 4. Verotarkastuksissa havaitut virheiden lukumäärät vuosina 2014–2020.	28
Taulukko 5. Satunnaistarkastuksissa havaittujen virheiden suuruus vuosina 2014–2020.	28
Taulukko 6. Muissa kokonaistarkastuksissa havaittujen virheiden suuruus vuosina 2014–2020.	29

Taulukko 7. Perusjoukon yritysten lukumäärät 2014–2020, toimivat yritykset.....	30
Taulukko 8. Perusjoukon yritysten lukumäärät 2014–2018, toimimattomat yritykset.....	30
Taulukko 9. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) CEM-kaltaistamismenetelmällä laskettuna.....	37
Taulukko 10. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) DBSCAN-algoritmilla laskettuna.....	37
Taulukko 11. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) XGBoost-koneoppimismallilla laskettuna.....	38
Taulukko 12. Arvio pimeiden palkkojen määrästä vuosina 2017–2020 (miljoonaa euroa) bayesilaisella logistisella regressiolla laskettuna.....	40
Taulukko 13. Arviot pimeiden palkkojen määrästä vuosina 2014–2020 (miljoonaa euroa) eri menetelmillä laskettuna.....	42

12 Liitteet

Liite 1. Osakasyrittäjän eläkevakuutusmaksun määräytymisperusteet

Liite 2. Substanssiperusteinen kaltaistamismenetelmä

Liite 3. Algoritmipohjaisten kaltaistamismenetelmien tarkempi kuvaus

Liite 4. Yrityskohtaisen odottama laskennan tarkempi kuvaus

Liite 1. Osakasyrittäjän eläkevakuutusmaksun määräytymisperusteet

Tässä selvityksessä ei arvioida yrittäjien eläkemaksujen eli YEL-maksujen määrää, joka osakasyrittäjiltä jää maksamatta mahdollisen YEL-vakuuttamattomuuden ja pimeän palkan osalta. Arviota ei tehdä, koska selvityksessä ei ollut käytettävissä YEL-vakuuttamisen yrittäjäkohtaisia tietoja. YEL-vakuuttamisen puutteiden aiheuttamat yrittäjien eläkemaksujen menetykset ovat kuitenkin selvästi palkansaajia pienempiä.

Yrittäjän eläketurva muodostuu siitä, että yrittäjä itse maksaa YEL-vakuutusmaksua, joka määräytyy yrittäjän arvioiman YEL-työtulon perusteella. Yrittäjän tulee maksaa YEL-maksua, kun arvioitu työtulo on vähintään 8 063,57 euroa vuodessa (2021). Osakeyhtiössä johtavassa asemassa oleva henkilö, joka omistaa yksin yli 30 prosenttia tai perheenjäsenten kanssa yli puolet yhtiöstä tai äänivallasta, on vakuutettava YEL:n mukaan. Alle 53-vuotiaan YEL-vakuutetun vakuutusmaksun suuruus on 24,1 prosenttia työtulosta vuonna 2021 (Eläketurvakeskus).

Osakeyhtiöiden yrittäjäosakkaiden pimeiden palkkojen maksamattomien YEL-maksujen todellisen määrän arvioiminen on haastavaa, koska yrittäjä voi lain mukaan itse määritellä ns. vuotuisen työtulonsa, jonka mukaan YEL-maksu määräytyy. Periaatteessa YEL-työtulon tulee vastata työpanoksen rahallista arvoa, mutta YEL-työtulo ei tarkoita verotettavia tuloja eikä yrittäjän itselleen maksamaa palkkaa (Ilmarinen, 2021). Neljännes yrittäjistä ilmoittaa työtulokseen alle 12 600 euroa, jolloin eläkemaksu on suuruudeltaan 2 000–3 000 euroa vuodessa. Peräti puolet yrittäjistä ilmoittaa työtulonsa suuruudeksi korkeintaan 20 600 euroa. (Eläketurvakeskus, 2020)

Liite 2. Substanssiperusteinen kaltaistamismenetelmä

Kaltaistaminen (eng. *matching*) tarkoittaa menetelmiä, joilla pyritään poistamaan satunnaistamisen puutteesta johtuvia vinoumia havaintoaineistoissa. Kaltaistamisen ideana on löytää jokaiselle niin sanotulle interventioryhmän jäsenelle (tässä tapauksessa verotarkastetulle yritykselle) yksi tai useampi määriteltyjen ominaisuuksien perusteella vastaavaksi todettu havainto.

Luvussa 7.1 esitetyt tulokset laskettiin niin sanotulla substanssiperusteisella kaltaistusmenetelmällä, jossa yritykset jaettiin luokittelevilla muuttujilla pienempiin, keskenään samankaltaisten yritysten ryhmiin (ns. jälkiositus). Kaltaistusmenetelmällä voitiin pienentää etenkin riskiperusteisten tarkastusten aiheuttamaa harhaa arvioidussa pimeän työn määrässä. On huomattava, että erilaisten luokittelujen käyttö tarkentaa myös satunnaisotantaan perustuvaa estimointia.

Aineistot

Selvityksessä käytettyjä tarkastusaineistoja on kuvattu aiemmin luvussa 6.1. Taulukoissa 15 ja 16 on vielä yhteenvetoa laskennassa käytetyistä aineistoista, joita olivat osakeyhtiöiden ja ulkomaisten yhteisöjen yritysjoukot sekä verotarkastustiedot vuosina 2014–2018. Tältä ajankakselta oli käytettävissä riskiperusteisten tarkastusten lisäksi myös satunnaistarkastuksia. Yritysten kokoluokittelu perustui ALV-myyntin perusteella laskettuun liikevaihtoon. Vaihtoehtoisesti käytettiin veroilmoituslomakkeen 6B tietoa, jos ALV-myynti puuttui ja liikevaihtotieto löytyi.

Taulukko 1. Perusjoukko ja tehdyt verotarkastukset 2014–2018 yritysten koon mukaan luokiteltuna (ALV-myynti tai liikevaihto).

ALV myynti / Liikevaihto	Perusjoukko, kpl	Tarkastukset yhteensä, kpl	Tarkastettujen osuus perusjoukosta, %	Satunnais-tarkastukset, kpl	Muut kokonais-tarkastukset, kpl
Kaikki yhteensä	939 350	11 511	1,2 %	1 546	9 965
Tieto puuttuu / 0€	124 880	350	0,3 %	3	347
Alle 400t€	574 139	3 874	0,7 %	321	3 553
400t-1m€	111 830	2 757	2,5 %	565	2 192
1-4m€	86 270	2 828	3,3 %	479	2 349
4-10m€	23 029	827	3,6 %	141	686
Yli 10m€	19 202	875	4,6 %	37	838

Lähde: Verohallinto 5–6/2021.

Taulukko 2. Verotarkastuksissa löytyneet virheet 2014–2018, lukumäärä ja eurot, yritysten koon mukaan luokiteltuna (ALV-myynti tai liikevaihto).

ALV myynti / Liikevaihto	Tarkastukset yhteensä, kpl	Löytyneet virheet, kpl	Löytyneet virheet, €
Kaikki yhteensä	11 511	3 221	187 134 655
Tieto puuttuu / 0€	350	134	11 301 892
Alle 400t€	3 874	1 456	59 413 858
400t-1m€	2 757	693	41 643 390
1-4m€	2 828	656	50 085 879
4-10m€	827	149	13 716 779
Yli 10m€	875	133	10 972 857

Lähde: Verohallinto 5–6/2021.

Laskentamenetelmä – kaltaistaminen

Pimeiden palkkojen määrä laskettiin seuraavien vaiheiden avulla:

(1) Virheiden lukumäärä ositteessa = perusjoukon yritysten lukumäärä ositteessa (pl. tarkastetut) * virheiden esiintyvyys ositteessa.

(2) Pimeät palkat ositteessa = virheiden lukumäärä ositteessa * keskimääräinen virhe ositteessa.

(3) Pimeät palkat koko taloudessa = kaikkien ositteiden pimeiden palkkojen summa.

(4) Lopulliseen tulokseen lisättiin myös verotarkastuksissa havaitut pimeät palkat.

Missä

- Virheiden esiintyvyys ositteessa = verotarkastuksissa löydettyjen virheiden lukumäärä ositteessa jaettuna kaikkien tehtyjen tarkastusten lukumäärällä ositteessa.
- Keskimääräinen virhe ositteessa = verotarkastuksissa havaittujen virheiden keskiarvo ositteessa.
- Vaiheet 1–3: pimeiden palkkojen estimointi perusjoukon tarkastamattomien yritysten keskuudessa.
- Lopulliseen tulokseen lisättiin myös verotarkastuksissa havaitut pimeät palkat (vaihe 4).

Laskentamenetelmä oli teknisesti hyvin yksinkertainen, mutta eri vaiheisiin sisältyi yksityiskohtia, joita on kuvattu tarkemmin seuraavissa alaluvuissa

Ositteet ja niiden luokittelevat muuttujat – korjattu tulos ja yläraja

Korjatun tuloksen ja ylärajan laskennassa luokittelijoiksi valittiin sellaisia yrityskohtaisia muuttujia, joilla havaittiin olevan riippuvuussuhde verotarkastuksissa löydettyjen virheiden esiintyvyyteen ja/tai suuruuteen.⁷⁹ Luokittelevat muuttujat olivat ALV-ilmoituksen myynnin määrä (tai vaihtoehtoisesti liikevaihto), ulkomaalaisten palkansaajien osuus kaikista yrityksen palkansaajista sekä yrityksen maksama keskipalkka (alle 50 000 euroa ansaitsevat palkansaajat). Erilaisia kolmen luokittelijan arvoyhdistelmiä eli ositteita muodostui kaikkiaan 173 kappaletta.

Valittujen luokittelijoiden lisäksi laskentaa testattiin erilaisilla yrityksen kokoa kuvaavilla luokittelijoilla, kuten palkansaajien lukumäärä ja palkkasumma. Ne antoivat samansuuntaisia tuloksia kuin valittu myynnin määrä. ALV-myynti valittiin kokoluokittelijaksi, koska kaksi muuta luokittelevaa muuttujaa sisälsivät palkansaajiin ja palkkoihin liittyviä tietoja. ALV-myynti valittiin ensisijaiseksi kokoa kuvaavaksi muuttujaksi, koska se voidaan kohdistaa kalenterivuodelle. Veroilmoituksen liikevaihto ei aina vastaa kalenterivuotta. Kuitenkin, jos yrityksellä ei ollut ALV-myyntiä, mutta liikevaihtotieto löytyi, valittiin kokomuuttujaksi liikevaihto.

⁷⁹ "The essence of an analytical framework is that it should be useful in understanding the NOE [non-observed economy] causes, or in reflecting the various measurement options, or both." OECD 2002, Measuring the Non-Observed Economy, luku 3.3.1 kohta 3.29.

Taulukko 3. Korjatun ja suurimman tuloksen laskennoissa käytetyt luokittelevat muuttujat.

Luokitteleva muuttuja	Luokat
Yrityksen ALV-ilmoituksen perusteella laskettu myyntin määrä kalenterivuoden aikana. Jos ko. tieto puuttuu niin veroilmoituksen liikevaihto	Tieto puuttuu / negatiivinen
	Alle 400t€
	400t-1m€
	1-4m€
	4-10m€
Ulkomaalaisten palkansaajien osuus yrityksen kaikkien palkansaajien lukumäärästä	Yli 10m€
	Tieto puuttuu
	0 prosenttia
	0-10 prosenttia
	10-30 prosenttia
	30-50 prosenttia
Yrityksen palkansaajilleen maksama keskipalkka . Keskipalkassa huomioidaan vain alle 50 000 € vuodessa ansaitsevat palkansaajat	50-70 prosenttia
	Yli 70 prosenttia
	Tieto puuttuu / negatiivinen
	Alle 11 t€
	11-22t€
	22-33t€
	Yli 33t€

Lähde: Verohallinto 5–6/2021.

Osa pääjaon mukaisista ositteista laskettiin vielä tarkemmalla jaolla. Näitä olivat ositteet, joissa ero satunnaislaskennan tulokseen oli huomattava (vähintään noin 5 milj. € / vuosi) ja joissa tehtyjä tarkastuksia oli vähintään 200. Näin vaikuttavimpien ositteiden kaltaistaminen voitiin viedä vielä edellä kuvattua tarkemmalle tasolle.

Joissakin ositteissa tarkennetulla luokittelulla oli suuri vaikutus tulokseen: erityisesti ositteessa 1 tarkempi luokittelu pienensi tulosta jopa kymmeniä prosentteja. Ositteessa 31 prosentuaalinen vaikutus oli melko pieni, mutta määrällisesti vaikutus oli lähes 10 miljoonaa euroa vuodessa. Nämä olivat laskennan suurimmat ja vaikuttavimmat ositteet, joiden tarkennettu luokittelu pienensi pimeiden palkkojen arvioitua määrää keskimäärin lähes 70 miljoonaa euroa vuodessa.

Taulukko 4. Korjatun tuloksen laskennassa käytetyt luokittelijat, aliositteisiin jaetut ositteet.⁸⁰

Osite	Myynti; Ulkom. palkansaajien osuus; Keskipalkka	Perusjoukko, lkm		Lisätyt luokittelijat	Vaikutus, %
		Tarkastamaton	Tarkastettu		
1	Ei tietoa / 0€; Ei tietoa; Ei tietoa / 0€	111 078	272	oik.muoto, TA-rekisterissä	-37 %
31	Alle 400t€; Ei tietoa; Ei tietoa / 0€	215 665	1 023	oik.muoto, TA-rek., palkansaaj.lkm	-4 %
37	Alle 400t€; 0prosenttia; Alle 11t€	136 739	783	palkansaajien lkm	0 %
38	Alle 400t€; 0prosenttia; 11-22t€	93 356	774	palkansaajien lkm	+3 %
39	Alle 400t€; 0prosenttia; 22-33t€	49 578	358	TA-rekisterissä	+2 %
61	400t-1m€; Ei tietoa; Ei tietoa / 0€	9 452	203	TA-rekisterissä	-5 %
66	400t-1m€; 0prosenttia; Alle 11t€	13 374	327	TA-rekisterissä	+3 %

Lähde: Verohallinto 5–6/2021.

Ositteet ja niiden luokittelevat muuttujat – satunnaistarkastuksiin perustuva laskenta eli tuloksen alaraja

Koska satunnaistarkastuksia oli tehty paljon vähemmän kuin riskiperusteisia tarkastuksia, jouduttiin puhtaasti satunnaistarkastuksiin perustuvat laskenta tekemään muita

⁸⁰ Suurimman tuloksen laskennassa tarkempi luokittelu sovellettiin ositteisiin 1 ja 31.

karkeammalla luokittelulla. Luokittelevat muuttujat olivat palkansaajien lukumäärä ja toimiala.⁸¹ Erilaisia kahden luokittelijan arvoyhdistelmiä eli ositteita muodostui kaikkiaan 62 kappaletta. Todellisuudessa tulosta ei voitu laskea aidosti näin tarkalla tasolla, vaan se muodostettiin erityisellä painotettujen keskiarvojen menettelyllä, joka on kuvattu lähemmin seuraavassa alaluvussa.

Taulukko 5. Pienimmän tuloksen laskennoissa käytetyt luokittelevat muuttujat.

Luokitteleva muuttuja	Luokat
Palkansaajien lukumäärä	Ei palkansaajia / tieto puuttuu
	1-4
	5-9
	10-19
	20-49
	50-249
	250 tai enemmän
Toimiala – karkea yhdistelmäluokittelu	Tieto puuttuu
	Alkutuotanto (A)
	Kaivostoiminta, Teollisuus, Energia ja Ympäristönhuolto (B-E)
	Rakentaminen (F)
	Kauppa (G)
	Kuljetus (H)
	Majoitus- ja ravitsemistoiminta (I)
	Liike-elämän palvelut (J-N)
	Muut palvelut (O-U)

Vertailun vuoksi satunnaistarkastuksiin perustuva tulos laskettiin myös samoilla ositteilla kuin muut tulokset – eli muuttujina olivat myynnin määrä tai liikevaihto, ulkomaalaisten palkansaajien osuus sekä keskipalkka⁸². Tulos oli noin 441 miljoonaa euroa vuodessa ajalla 2014–2018 ja se laskettiin painotettujen keskiarvojen menettelyllä (ks. seuraava alaluku). Ongelmana tässä laskennassa oli erityisesti osite 1, jonka luokassa myynti/liikevaihto = 'Tieto puuttuu' oli tehty vain kolme satunnaistarkastusta.

Satunnaistarkastuksiin perustuvat pimeiden palkkojen määrä laskettiin myös ilman luokittelijoita ja yhden luokittelijan avulla (käyttäen viittä erilaista luokittelijaa). Suurin tulos oli 598 miljoonaa euroa vuodessa (ilman luokittelijoita) ja pienin tulos oli 434 miljoonaa euroa vuodessa (luokittelijana palkansaajien lukumäärä).

Virheiden esiintyvyys

Virheiden esiintyvyys laskettiin hieman eri tavoin korjatun ja suurimman vs. pienimmän tuloksen laskennossa. Korjatun ja suurimman tuloksen laskennoissa käytettiin satunnaistarkastusten lisäksi myös muita kokonaistarkastuksia. Riskiperusteisissa tarkastuksissa havaittu esiintyvyys kuitenkin suhteutettiin satunnaistarkastuksiin. Näin ollen voidaan sanoa, että kaikissa tuloksissa pimeän työn esiintyvyys vastaa satunnaistarkastuksissa havaittua virheiden esiintyvyyttä.

⁸¹ Myynnin määrää tai liikevaihtoa ei voitu käyttää kokomuuttujana, koska luokassa "tieto puuttuu" oli tehty vain kolme satunnaistarkastusta. Se olisi ollut ongelmallista satunnaistarkastuksiin perustuvassa laskennassa, koska pimeiden palkkojen määrän arvioitiin olleen melko suuri tässä kokoluokassa.

⁸² Yrityksen maksama keskipalkka, josta on poistettu yli 50 000 euroa vuodessa ansaitsevat palkansaajat.

Satunnaistarkastuksiin perustuvassa laskennassa esiintyvyys laskettiin erityisellä painotettujen keskiarvojen menettelyllä: satunnaistarkastusaineiston esiintyvyydet laskettiin yhtä luokitelijaa käyttäen (palkansaajien lukumäärä tai toimiala) ja tulokset painotettiin yhteen perusjoukon koon mukaan. Näin voitiin laskea esiintyvyys kaikille 62 eri ositteelle.⁸³

Muissa laskennoissa esiintyvyys laskettiin ensin riskiperusteisten tarkastusten aineistoista. Tämä tulos kuitenkin suhteutettiin satunnaistarkastusten tulokseen ja huomioitiin, millä kohdevalinta-alueilla ositteen riskiperusteiset tarkastukset oli tehty:

Laskettiin ositteille virheiden **esiintyvyydet riskiperusteisissa tarkastuksissa**.
Laskettiin ositteille niin sanotut **satunnaiskertoimet**:

(i) Eri kohdevalinta-alueiden esiintyvyydet suhteutettiin satunnaistarkastusten keskimääräiseen esiintyvyyteen. Käytännössä virheiden esiintyvyys kohdevalinta-alueella jaettiin 2,5:llä (= virheiden esiintyvyys kaikissa riskiperusteisissa tarkastuksissa jaettuna virheiden esiintyvyys satunnaistarkastuksissa).

(ii) Laskettiin eri kohdevalinta-alueiden osuus ositteen tarkastuksista. Tällä suhdeluvulla painotettiin kohdan (i) kohdevalinta-alueiden skaalatut esiintyvyydet yhteen ja näin saatiin mainittu satunnaiskerroin.

Laskettiin ositteiden satunnaistarkastuksiin ja kohdevalinta-alueisiin **skaalatut esiintyvyydet** jakamalla kohdassa (a) lasketut esiintyvyydet kohdassa (b) laskutuilla satunnaiskertomilla.

Keskimääräinen verotarkastuksissa havaittu virhe

Erilaiset arviot pimeiden palkkojen määrästä – alaraja, korjattu tulos ja yläraja – johtuivat suurimmaksi osaksi eri aineistoista ja eri tavoin lasketuista keskimääräisistä virheistä. Verotarkastuksissa havaitut virheet eli laskennassa käytetyt *aineistot* on kuvattu luvussa 6.2. Tässä luvussa keskitytään kuvaamaan erilaisia tapoja *laskea* keskimääräinen virhe.

Pimeiden palkkojen alaraja saatiin laskemalla eri ositteiden keskimääräinen virhe pelkästään satunnaistarkastusten perusteella. Satunnaistarkastusten otoskehikossa oli alipeittoa, sillä siitä puuttuivat hiljattain verotarkastetut yritykset. Verotarkastuksen kohdevalinta pyrkii löytämään tarkastettavaksi suurimmat virheet (väärin toimivat yritykset), jolloin nämä poistuvat satunnaisotannan otoskehikosta. Näin ollen, satunnaisotannan keskimääräinen virhe on oletettavasti ollut alaspäin harhainen.⁸⁴ Esimerkiksi satunnaisverotarkastuksissa ei havaittu lainkaan työntekijöitä, joiden kaikki palkat oli maksettu pimeästi. Kohdevalintatarkastuksissa tällaisia täysin pimeitä työntekijöitä kuitenkin löydettiin.

⁸³ Satunnaistarkastusten määrä ei ollut riittävän suuri, jotta esiintyvyydet olisi voitu laskea aidosti suoraan aineistosta kaikille 62 ositteelle.

⁸⁴ Tarkastustoiminta on oletettavasti vaikuttanut enemmän juuri keskimääräiseen virheeseen kuin virheiden esiintyvyyteen. Joissakin ositteissa tarkastuskattavuus oli melko korkea - tällöin myös virheiden esiintyvyys on voinut olla alaspäin harhainen satunnaistarkastuksissa.

Toinen satunnaistarkastuksiin liittyvä ongelma oli jo aiemmin mainittu niiden vähäinen määrä. Mitä pienempi satunnaistarkastusten ja niissä löytyneiden virheiden määrä oli, sitä suurempi oli sattuman vaikutus esiintyvyyteen ja keskimääräiseen virheeseen.

Pimeiden palkkojen yläraja saatiin laskemalla keskimääräinen virhe koko aineiston eli satunnais- ja riskiperusteisten tarkastusten perusteella. Riskiperusteisia tarkastuksia oli aineistossa huomattavan paljon enemmän kuin satunnaistarkastuskohteita. Ylärajan laskennassa keskimääräisen virheen oletettiin olevan ylöspäin harhainen, koska verotarkastuksen kohdevalinta on pyrkinyt poimimaan tarkastettavaksi suurimpia virheitä tehneet yritykset.

Korjattu tulos saatiin laskemalla keskimääräinen virhe koko aineistosta kuten yläraja, mutta oletettua harhaa pienennettiin niin sanotun trimmatun jakauman avulla. Käytetty trimmauksen aste oli 90 prosenttia – tämä tarkoittaa, että 5 prosenttia tarkastushavainnoista poistettiin jakauman ylä- ja alapäistä – eli suurimmat ja pienimmät virheet. Koska virheiden jakaumat olivat voimakkaasti vinoja oikealle, menettely pienensi keskimääräisiä virheitä ja siten oletettua keskimääräisen virheen harhaa.⁸⁵

Jakaumia, joista keskimääräiset virheet laskettiin, trimmattiin melko voimakkaasti. Usein käytetty menettely on 95 prosentin trimmaus, mutta tässä tapauksessa 90 prosentin trimmausta pidettiin oikeampana. Satunnaistarkastusten otoskehikossa ei siis ollut sellaisia yrityksiä, joilta oli jo aiemmin havaittu suuri virhe. Kohdevalinnassa harha kohdistuu suoraan valittuihin tarkastuksiin, jotka pyrittiin valitsemaan erilaisilla riskikriteereillä mahdollisimman osuvasti. Näin ollen voidaan olettaa, että aiheutunut harha oli suurempi jälkimmäisessä tapauksessa ja siten melko voimakas trimmaus oli perusteltu.

Substanssiperusteiden kaltaistamisen vahvuudet ja heikkoudet

Vuosia 2014–2018 koskevassa laskennassa oli mahdollista hyödyntää **satunnaisotantaan perustuvia verotarkastuksia**. Satunnaistarkastukset antoivat arvokasta vertailutietoa riskiperusteisten verotarkastusten ja niiden tulosten arviointiin.

Satunnaistarkastuksia ei kuitenkaan ollut käytettävissä niin paljon, että arvio virheiden esiintyvyydestä tai suuruudesta olisi voinut perustua pelkästään näihin. Erityisen ongelmallista oli se, että satunnaistarkastuksia oli hyvin vähän juuri niissä ositteissa, joissa pimeän työn määrä arvioitiin suurimmaksi. Toisin sanoen yritysten kokoluokassa, jossa liikevaihto jäi alle 400 000 euroon vuodessa.

Laskennan vahvuutena voidaan pitää verotarkastusaineistojen ja substanssikysymysten syvällistä tuntemusta. Esimerkiksi laskennassa hyödynnettiin tietoa virheiden esiintyvyydestä verotarkastuksen kohdevalinta-alueilla ja sitä verrattiin virheiden esiintyvyyteen satunnaistarkastuksissa. Näin voitiin muodostaa aiempaa tarkempi käsitys kohdevalinnan vaikutuksesta virheiden löytymiseen verotarkastuksissa ja tätä tietoa hyödynnettiin myös virheiden esiintyvyyden laskemisessa.

Yhtenä laskennan ongelmakohtana voidaan pitää keskimääräisen virheen arviointia. Riskiperusteisissa tarkastuksissa keskimääräiset virheet olivat suurempia kuin

⁸⁵ Ks. edellä kohta 12.2.5.

satunnaistarkastuksissa. Toisaalta satunnaistarkastusten otoskehikko ei ollut täysin objektiivisesti muodostettu, koska siitä oli poistettu hiljattain tarkastetut yritykset. Näin ollen on syytä epäillä, että satunnaistarkastusten keskimääräinen virhe oli liian pieni. Toisaalta riskiperusteisten tarkastusten keskimääräinen virhe oli todennäköisesti liian suuri. Riskiperusteisten tarkastusten harhan korjaamiseksi korjatun tuloksen laskennassa poistettiin suurimpia ja pienimpiä virrehavaintoja niin sanotun trimmauksen avulla.

Liite 3. Algoritmivetoiset kaltaistusmenetelmät

Kaltaistamista kokeiltiin kahdella eri menetelmällä puhtaan algoritmivetoisesti. Näitä menetelmiä kuvaillaan tarkemmin kahdessa seuraavassa kappaleessa. Menetelmien ollessa hyvin teknisiä ovat liitteet 3 ja 4 suunnattu pääasiassa analyysimenetelmiin perehtyneille lukijoille.

CEM-kaltaistaminen

Coarsened Exact Matching eli CEM on yksinkertainen ja tehokas kaltaistamisen menetelmä, jolla voidaan tasoittaa valikointiharhaa hakemalla kontrolliryhmästä (=tarkastamattomat yritykset) vastaavia havaintoyksilöitä kuin interventioryhmässä (=tarkastetut yritykset).⁸⁶ CEM-algoritmi toimii kahdessa vaiheessa.

Numeeristen muuttujien kategorisointi

Ensiksi CEM-algoritmissa kategorisoidaan kaikki valitut numeeriset muuttujat esimerkiksi kymmeneen eri tasaluokkaan ja annetaan kategoristen muuttujien olla sellaisenaan. Näin jokainen muuttuja muuttuu kategoriseksi, jossa kukin havainto kuuluu yhteen luokkaan kussakin muuttujassa. Tästä tulee algoritmin nimeen sana "coarsened" eli "karkeistettu". Numeeristen muuttujien kategorisointi mahdollistaa algoritmin seuraavan askeleen, jota kuvaa sana "exact" eli eksakti, tarkka.

Vastinparien etsiminen

Toisessa vaiheessa kullekin tarkastetulle yritykselle pyritään löytämään yksi tai useampi vastinpari, joka vastaa sitä täydellisesti kaikkien valittujen muuttujien ylitse. Esimerkiksi jos kaltaistus tehdään toimialan, palkollisten lukumäärän sekä liikevaihdon ylitse, etsitään toimialaan D, liikevaihto-desiiliin 6 ja palkollisten lukumäärä-desiiliin 4 kuuluvalla tarkastetulle yritykselle tarkastamattomien yritysten joukosta yritys, jonka vastaavat muuttujien arvot ovat vektorimuodossa esitettynä (D,6,4). Jos täysin eksaktia vastinparia ei löydy, paritusta ei tehdä. Jos useammalla tarkastetulla yrityksellä on täysin toisiaan vastaavat arvot, niin ne etsivät yhteisiä vastinpareja tarkastamattomien joukosta.

Kaltaisten yritysten etsintä tehdään kaikkien verovuosien ylitse eli vuoden 2018 verotarkastetun yrityksen tiedot voivat mennä niin vuosien 2017 kuin 2020 sitä muistuttaville tarkastamattomille yrityksille. Koska tarkastettujen yritysten määrä laskee vuosi vuodelta (koska tarkastukset eivät ole ehtineet valmistua), jäisi viimeisille vuosille niin pieni määrä havaintoja tarkastetuista yrityksistä, ettei vastinparejakaan löytyisi muusta populaatiosta juuri lainkaan. Tämä valinta aiheuttaa ongelmia niissä muuttujissa, joissa on selkeä ajallinen trendi yli verovuosien.

Paritus tehdään kymmenen muuttujan ylitse, jotka on valittu algoritmisesti. Muuttujien valinta tehdään vain tarkastettujen yritysten aineistosta siten, että valitaan Gini Impurity-indeksillä kymmenen tarkastuksen lopputuloksen kannalta merkityksellisintä muuttujaa. Näin yritysten

⁸⁶ Kun interventio- ja kontrolliryhmät eivät ole valikoituneet satunnaisesti, koitetaan kaltaistamismenetelmällä löytää sellaisia vastinpareja, jotka vastaavat toisiaan ja ovat siten vertailukelpoisia. Toisin sanoen, jos on esimerkiksi verotarkastettu yritys, jonka liikevaihto on 5 miljoonaa ja toimialakirjain on D, menetelmällä pyritään löytämään vastaava yritys tarkastamattomien joukosta.

kaltaistus tapahtuu muuttujien ylitse, jotka luultavimmin ennakoivat pimeää palkanmaksua. Toisaalta kaikki substanssilähtöinen asiantuntemus jää huomioimatta.

Tietojen kopiointi kontrolliryhmään

Kun tarkastetuille yrityksille on löydetty yksi tai useampi vastinpari, kopioidaan näille tarkastetun yrityksen lopputulos eli pimeiden palkkojen kokonaissumma sekä palkansaajien ja osakkaiden pimeiden palkkojen summat. Jos ryhmässä on kaksi tai useampia tarkastettuja yrityksiä, kopioidaan näiden vastinpar(e)ille tarkastustulosten lopputulosten keskiarvo.

Menetelmän hyviä ja huonoja puolia

CEM-menetelmän hyvä puoli on sen yksinkertaisuus ja läpinäkyvyys. Pimeän työn määrä päätellään ei-tarkastettujen yritysten joukossa selkeästi; kopioidaan tiedot vastinparilta sellaisenaan. CEM vie myös kaltaistamisen äärimmäisen tarkalle tasolle, sillä vastinparin pitää algoritmin nimen mukaisesti olla eksakti kopio tarkastetusta yrityksestä. Kun kaltaistamiseen valitut muuttujat on valittu niiden tarkastustuloksen ennustevoimaisuuden perusteella, voidaan perustellusti olettaa pimeää työtä teettävien tarkastettujen yritysten vastinparien olevan pimeän työn todennäköisiä riskiyrityksiä. Menetelmän hyvä ominaisuus on myös siinä, että peitellyt osingot ja pimeät palkat saadaan pääteltyä ei-tarkastetuille yrityksille suoraan tekemättä oletuksia kokonaissumman mahdollisesta jakautumisesta komponentteihinsa. Toisaalta menetelmä tekee naiiveja oletuksia kopioidessaan suoraviivaisesti pimeän työn määrän sellaisenaan tarkastetulta yritykseltä. Tällä tavoin ei saada lainkaan mallinnettua pimeän työn esiintyvyyteen tai sen määrään liittyviä epävarmuuksia. Koska kussakin vastinparissa on parhaimmillaankin vain muutamia tarkastettuja yrityksiä, ei mitään tilastollisia päätelmiä kyetä tekemään. Myöskään satunnaistarkastuksia ja niiden tarjoamaa informaatiota ei saada mitenkään hyödynnettyä. Koska täyden vastaavuuden vaatimus on varsin raju kymmenen eri muuttujan ylitse, jää moni tarkastettu yritys ilman vastinparia ja näin näiden tarkastusten pimeästä työstä tuottama informaatio jää hyödyntämättä. Tarkastuksia ei ylipäätään ole valmistunut valitulla ajanjaksolla liiallisesti, joten informaation kadottaminen ei ole tässä käyttötapauksessa toivottavaa. Lisäksi menetelmä on altis isoille muutoksille riippuen montako ja mitä muuttujia kaltaistukseen käytetään, eikä muuttujien määrän tahi koostumuksen valintaan ole objektiivisia laadun mittareita.

DBSCAN-kaltaistaminen

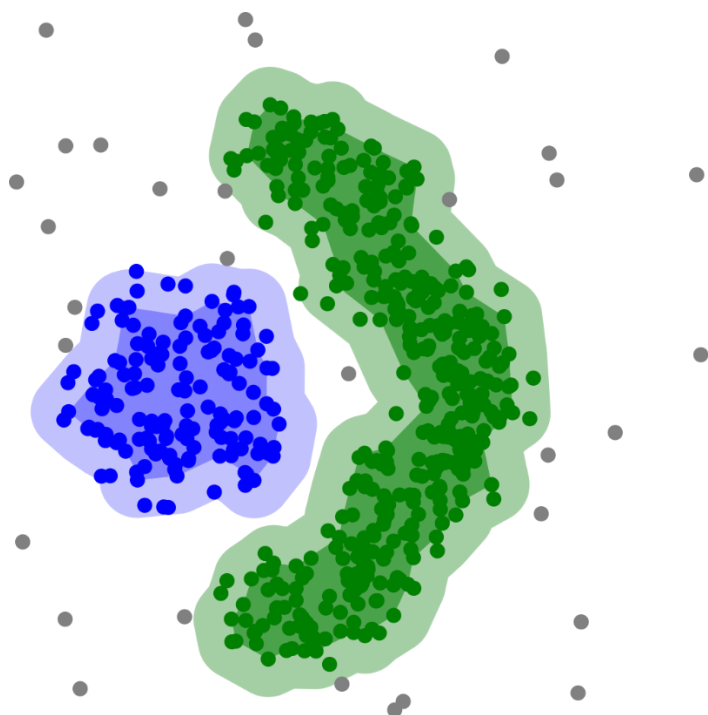
DBSCAN (=density-based spatial clustering of applications with noise) on klusterointialgoritmi, joka annettujen hyperparametrien⁸⁷ valossa päättää aineistosta itse optimaalisen määrän klustereita eli alajoukkoja. Algoritmi myös merkitsee osan havainnoista poikkeaviksi eikä aseta niitä mihinkään klusteriin. DBSCAN tarvitsee kaksi hyperparametria, kahden havainnon välisen maksimietäisyyden, jotta ne voidaan laskea samaan naapurustoon⁸⁸ sekä minimimäärän havaintoja, joita havainnolla tulee olla naapurustossaan, jotta se voidaan

⁸⁷ Hyperparametri tarkoittaa parametria, joka kontrolloi oppimisprosessia. Se voidaan käsittää algoritmin ohjausvuksi ja sen eri asennot johtavat erilaisiin oppimistuloksiin datasta. Hyperparametreja ei siis päätellä datasta

⁸⁸ Naapurusto tarkoittaa havainnon X lähimpiä muita havaintoja.

laskea "ydinpisteeksi"⁸⁹. DBSCAN perustuu nimensä mukaisesti "tiheyteen"⁹⁰ ja naapurustojen käsitteeseen. Alla oleva kuva havainnollistaa algoritmia. Kuviossa löytyy kaksi tiheää havaintojen keskittymää eli klusteria sekä joukko poikkeavia havaintoja, jotka sijaitsevat kauempana molemmista pääjoukoista.

Kuvio 1. Esimerkki DBSCAN-klusterointituloksesta



Lähde: <https://en.wikipedia.org/wiki/DBSCAN> Lisenssi: CC By-SA 3.0 Tekijä: Chire

Klusterointi

Klusterointialgoritmeille annettavat kymmenen muuttujaa valitaan numeeristen selittävien muuttujien joukosta samalla Gini Impurity-indeksillä kuin CEM-kaltaistamisessa. Sen lisäksi muuttujat skaalataan samalla asteikolle, jotta euklideiseen etäisyyteen perustuva DBSCAN-algoritmi toimisi oikein. Jos yhden muuttujan vaihteluväli kulkee nolasta yhteen ja toisen nolasta miljardiin, niin on selvää, että kahden pisteen välistä etäisyyttä mitattaisiin käytännössä vain sen muuttujan avulla, jonka suurin arvo on miljardi.

DBSCAN:in avulla saamme ryhmiteltyä yritysjoukon automaattisesti alajoukkoihin, joiden jäsenet muistuttavat valittujen muuttujien osalta toisiaan sekä poikkeaviin havaintoihin, joille ei paikkaa klustereista löydy. Näiden ryhmien sisällä kaltaistus on väljempää kuin CEM-algoritmeilla eli yritysten kesken on enemmän varianssia. Tämä mahdollistaa isommat ryhmittymät ja tilastollisen päättelyn virheen esiintyvyydelle sekä suuruudelle klusteriin kuuluvien

⁸⁹ DBSCAN: in ydinpiste eli core point on havainto, jolla on vähintään minimimäärä-parametrin (=minpts) verran havaintoja maksimietäisyyden (=eps) päässä itsestään.

⁹⁰ Tiheällä alueella tarkoitetaan aluetta avaruudessa, jossa on paljon toisilleen läheisiä havaintoja kun taas toisistaan poikkeavat havainnot ovat vähemmän tiheillä alueilla.

tarkastettujen yritysten avulla. Klustereiden ulkopuolelle jääneisiin yrityksiin ei päätellä pimeän työn riskisyyttä.

Virheen päättely bayesilaisittain

Pimeitä palkkoja maksaneiden yritysten klusterikohtaisen osuuden päättelyssä käytetään hyväksi bayesilaista tilastollista päättelyä, jossa klusterikohtaisen binomijakauman arvo p muodostuu klusteriin kuuluvien tarkastettujen yrityksen virheen esiintyvyydestä sekä vanhojen satunnaistarkastusten avulla päätellystä priorijakaumasta. Näin saadaan hyödynnettyä datan ulkopuolista informaatiota, josta on hyötyä varsinkin niiden klusterien osalta, joihin osuu vain muutama tarkastettu yritys.

Matemaattisesti pimeän työn esiintyvyys p kussakin klusterissa voidaan esittää seuraavasti:

$$p \sim \text{Binomial}(n, t)$$

$$(t) \sim \text{beta}(\alpha, \beta)$$

Priorijakauma t pimeiden palkkojen maksamisen todennäköisyydelle päätellään beta-jakaumasta, jonka parametrit α (=satunnaistarkastusten osumien lukumäärä) ja β (=satunnaistarkastusten lukumäärä) saadaan aikaisempien vuosien 2014–2016 satunnaistarkastusten datasta. Virheen esiintyvyys p taas päätellään binomijakaumana, johon vaikuttaa niin klusterin tarkastettujen yritysten joukosta havaittu virheen esiintyvyys kuin priorijakauma t . Priorijakauma asettaa ennako-odotuksia virheen esiintyvyydelle ja vaikuttaa lopputulemaan sitä enemmän, mitä vähemmän dataa eli tarkastettuja yrityksiä klusteriin kuuluu. Koska riskiperusteisesti tehdyissä tarkastuksissa pimeää työtä havaitaan huomattavasti useammin kuin satunnaistarkastuksissa, käytännössä bayesilaisen päättelyn hyödyntäminen johtaa useimmissa klustereissa virheen esiintyvyyden arvion maltillistumiseen.

Jokaiselle klusterille päätellään lisäksi gamma-jakauma, joka kertoo kyseisen klusterin virheen suuruudesta. Gamma-jakauman parametrit (skaala ja muoto) estimoidaan hyödyntäen klusterin tarkastettuja yrityksiä, joilta pimeää työtä on löydetty. Jos tällaisia yrityksiä on kaksi tai vähemmän, virheen suuruutta ei pystytä päättelemään ja se asetetaan kaikille yrityksille arvoon nolla. Virheen suuruutta mallinnetaan gamma-jakaumalla, koska se soveltuu hyvin vinoon tarkastustulosten jakaumaan paremmin kuin esimerkiksi normaalijakauma. Kun klusterille on päätelty gamma-jakauman parametrit, otetaan lopputuloksena saadusta jakaumasta satunnaisotoksia klusterin tarkastamattomille yrityksille. Satunnaislukuja jakaumasta generoimalla luonnollisesti lisäämme sattuman elementin vaikutusta lopputuloksena saataviin estimaatteihin. Toisaalta yrityksille päätellyt virheen suuruudet ovat jakaumasta otoksina otettuina realistisempia kuin keskiarvon käyttäminen koko yritys populaatiolle.

Pimeästi maksettujen palkkojen laskeminen

Kun klusterin sisällä on päätelty pimeiden palkkojen maksamisen esiintyvyys sekä kullekin yritykselle generoitu virheen suuruutta kuvaava satunnaisluku gamma-jakaumasta, nämä kerrotaan keskenään. Näin saadaan kullekin yritykselle euromääräinen virheen odottama, joka on generoitu klusterin tarkastettujen yritysten tarkastustulosten tilastollisten ominaisuuksien

perusteella. Nämä odottamat summaamalla kaikkien klusterien ylitse verovuositain saadaan verovuositaiset estimaatit.

Kunkin klusterin sisällä päätellään lisäksi, kuinka tarkastuksissa havaitut pimeät palkat jakautuvat osakkaiden ja palkansaajien kesken. Jos esimerkiksi klusterin tarkastettujen yritysten joukossa palkansaajien pimeitä palkkoja on maksettu 75 prosenttia kokonaissummasta ja loput 25 prosenttia ovat osakkaiden pimeitä palkkoja, jaetaan tällä $\frac{3}{4}$ ja $\frac{1}{4}$ jako-osuudella kokonaissumma. Näin saadaan estimaatit osakkaiden ja palkansaajien pimeiden palkkojen määrälle.

Klusteroinnin toteutustavat

DBSCAN-algoritmi voidaan ajaa aineistoon kahdella eri tavalla: siten, että klusterointi tehdään (verovuositain) koko aineistoon ja siten, että ensiksi klusterointi tehdään vain tarkastettujen aineistoon ja sitten nämä luodut klusterit projisoidaan loppudataan. Jälkimmäinen menetelmä tarkoittaa siis sitä, että ensiksi tarkastetut yritykset luokitellaan eri ryhmiin DBSCAN-algoritmillä ja tämän jälkeen näihin ryhmiin etsitään jäseniä loppupopulaatiosta. Klusterien ominaisuudet ja koostumus päätellään siis aluksi vain tarkastettujen joukosta ja kun nämä ominaisuudet ovat kertaalleen löydetty pienemmästä osajoukosta, "ennustetaan" loppudatalle heidän kuulumisensa klustereihin.

Käytännön tasolla menetelmien ero on siinä, että klusterien päättely ensiksi vain tarkastettujen yritysten joukossa takaa sen, että kaikkiin klustereihin päätyy tarkastettuja yrityksiä. Koliikon kääntöpuolella se myös lisää mihinkään klusteriin kuulumattomien poikkeavien havaintojen lukumäärää, kun tarkastamattomien yritysten joukosta löytyy paljon sellaisia, jotka eivät miltei ominaisuuksiltaan vastaa tarkastettuja. Jos klusterien koostumus päätellään alusta alkaen koko aineistosta, tällöin saadaan lopputuloksena klustereita, joihin ei osu ensimmäistäkään tarkastettua yritystä ja näin koko klusterin kaikki yritykset nähdään sataprosenttisen valkoisina eikä niille voida päätellä minkäänlaista riskisyyttä.

Näistä kahdesta menetelmästä valittiin klusterien koostumuksen päättely ensiksi vain tarkastettujen yritysten joukosta. Tällä tavalla menetelmä muistuttaa hieman enemmän CEM-kaltaistamista siten, että siinä haetaan tarkastettujen yritysten ryppäälle vastinpareja tarkastamattomien yritysten joukosta. Jos klusterit pääteltäisiin koko yritysjoukosta, ongelmaksi muotoutui lisäksi se, että niin iso yrityksistä meni yhteen massiiviseen klusteriin, ettei sen kyky kaltaistaa yrityksiä riskisyydeltään samankaltaiseen ositteeseen olisi ollut enää uskottava.

Menetelmän vahvuudet ja heikkoudet

DBSCAN:illa tehdyn kaltaistuksen hyvä puoli on, että aineisto saadaan automaattisesti jaetua sopivaan määrään sisäisesti samankaltaisia ositteita. Näiden ositteiden avulla voidaan torjua riskiperusteisesta tarkastuksesta seuraavaa vinoumaa. Ositteiden sisälle kertyy keskimäärin sen verran tarkastettuja yrityksiä, että niiden sisällä voidaan päätellä järjestelmällisesti klusteriin kuuluvien tarkastamattomien yritysten todennäköinen pimeän työn riski ja koluokka. Koska klusterien muodostukseen valitut muuttujat on valittu sellaisiksi, jotka ennustavat hyvin pimeän työn esiintyvyyttä, on syytä uskoa syntyneiden klusterien olevan mielekkäissä yhteydessä pimeään työhön. Menetelmällä voidaan myös huomioida

satunnaistarkastukset hyödyntäen bayesilaista tilastollista päättelyä. Yhtenä etuna voidaan myös pitää sitä, että osa havainnoista ajautuu poikkeavien havaintojen kaatoluokkaan eikä näille muusta yrityspopulaatiosta eroaville yritykselle edes yritetä päätellä harmaan työn suuruutta. Esimerkiksi k-means-algoritmi ei salli kaikkien klusterien ulkopuolelle jääviä rajatapauksia, jolloin riippumatta kaltaistuksen heikosta laadusta kaikille yrityksille tulisi pääteltyä riski pimeiden palkkojen maksamiseen.

Menetelmän suurin heikkous on klusterien epävakaus ja suuri vaihtelu riippuen valituista hyperparametreista sekä käytetyistä muuttujista. Tämän myötä lopputuloksissa voi tapahtua suurta heittelyä riippuen valituista hyperparametreista. Klusterointi on jossain määrin enemmän taidetta kuin tiedettä, eikä klusteroinnin laadulle ole samanlaisia objektiivisia mittareita kuin esimerkiksi regressiolle tai luokitteleville malleille. Tästä seuraa se, että lopputuloksiin liittyy paljon valitun algoritmin sisäsyntyistä epävarmuutta. Hyperparametrit pyrittiin valitsemaan siten, että lopputuloksena olisi riittävän monta järkevän kokoista klusteria, mutta riippumatta hyperparametrien kombinaatioista tuloksena oli muutama kooltaan iso klusteri ja sitten isompi tai pienempi määrä muutamien kymmenien havaintojen kokoisia pienempiä osajoukkoja. Substanssilähtöinen ositteiden teko mahdollistaisi näiden todella isojen joukkojen pilkkomisen pienempiin osiin, mutta tässä menetelmässä sitä ei lähdetty enää toteuttamaan.

Liite 4. Yrityskohtaisen odottaman laskenta

Tässä menetelmässä jokaiselle perusjoukon yritykselle päätellään erikseen kaksi suuretta, pimeiden palkkojen esiintymisen todennäköisyys sekä niiden suuruus riskin toteutuessa. Kerromalla nämä keskenään saadaan jokaiselle yritykselle pääteltyä pimeiden palkkojen arvioitu määrä. Arvio pimeiden palkkojen kokonaismäärästä perusjoukossa on yritys kohtaisten odotamien summa. Summaus voidaan tehdä jokaiselle vuodelle erikseen. Matemaattisesti menetelmä voidaan kuvata seuraavasti:

Asiakkaat: $1, \dots, N$

Pimeiden palkkojen esiintymisen todennäköisyys: $T_i \sim \text{Bernoulli}(p_i)$

Pimeiden palkkojen odotettu määrä: $M_i \sim \text{Gamma}(k_i, \theta_i)$

Pimeiden palkkojen odotusarvo: $E_i = T_i M_i$

Pimeiden palkkojen kokonaismäärä: $L^N = \sum_{i=1}^N E_i$

On hyvä huomioida, että saatua odottamaa ei voi suoraan käyttää arviona sille, kuinka paljon pimeänä teetettyä työtä löytyisi, jos yritys tarkastettaisiin. Ennemmin se kertoo, että jos 1 000 täysin vastaavaa yritystä tarkastettaisiin, kuinka paljon voitaisiin keskiarvoisesti olettaa pimeää työtä löytyvän per yritys.

Käytännössä asiakkaan riskisyyden arvio tehdään hyödyntämällä verotarkastuksia, joista tiedämme, löytyikö asiakkaalta tarkastuksessa pimeänä teetettyä työtä vai ei ja jos kyllä, kuinka paljon palkkoja on jäänyt ilmoittamatta. Menetelmässä hyödynnetään niin pimeiden palkkojen esiintyvyyttä (binäärinen 0/1 luokittelu) kuin niiden euromäärää vastemuuttujana (regressiomalli) ja rakennamme siis kaksi erillistä mallia pimeiden palkkojen odotusarvon päättämiseksi.

Yrityskohtainen pimeiden palkkojen todennäköisyys

Yritysten todennäköisyys maksaa pimeitä palkkoja päätellään koko yritys populaatiolle rakentamalla binäärisesti luokitteleva malli verotarkastettujen yritysten populaation sisällä, joka pyrkii ennustavien muuttujien avulla erottelemaan pimeää työtä teettävät yritykset muista. Vasteena toimii binäärinen 0/1-muuttuja, jossa euronkin verran pimeitä palkkoja maksaneet yritykset luokitellaan harmaiksi. Valitsemalla sellainen joukko pimeiden palkkojen esiintyvyyttä ennustavia muuttujia, jotka voidaan yhtäläisesti päätellä sekä tarkastetuille että ei-tarkastetuille yrityksille (kuten esimerkiksi Verohallinnolle ilmoitettujen palkkailmoitusten mukaan päätelty yrityksen työntekijöiden keskipalkka), voidaan myös ei-tarkastettujen yritysten riskisyyttä arvioida opitun matemaattisen mallin avulla. Rakennetuilla malleilla voidaan koko laajemman yritys kentän riskisyyttä arvioida, mutta tietyillä reunaehdoilla.

Kuten aiemmin todettiin, verotarkastukseen voi joutua joko satunnaisvalinnan kautta tai riskiperusteisesti, valtaosan ollessa jälkimmäisiä. Tästä seuraa, että aineistossa esiintyy vakavaa valikoitumisharhaa, jonka myötä tilastolliset mallit kuvittelevat pimeän työn tekemisen olevan paljon yleisempää kuin se oikeasti onkaan. Valikoitumisharhan pienentämiseksi käytetään kolmea keinoa.

Pyritään käyttämään vastemuuttujan⁹¹ ennustamiseen sellaisia muuttujia, joita ei käytetä tarkastukseen valikoimisessa apuna.

Mallinnetaan kunkin asiakkaan todennäköisyys joutua verotarkastukseen valitulla ajanjaksolla ja käytetään saatuja todennäköisyyksiä eli ns. propensiteettipisteitä yhtenä ennustavana muuttujana, jolloin tarkastukseen joutumisen valikoitumisharha tulee huomioiduksi.

Käytetään aikaisempien vuosien laajaa satunnaistarkastusaineistoa priorijakauman⁹² päättelyyn bayesilaiselle mallille.

Yrityksen maksamien pimeiden palkkojen esiintyvyydelle rakennettiin kaksi erilaista mallia. Toinen on päätöspuihin perustuva XGBoost-malli ja toinen bayesilainen logistinen regressio. Näistä jälkimmäisestä kerrotaan tarkemmin myöhemmin. Päätöspuihin perustuva XGBoost on hyvin tehokas ohjatun oppimisen algoritmi, joka perustuu siihen, että rakennetaan sarja pieniä päätöspuita, joista kukin päätöspuu koittaa korjata edellisen tekemiä virheitä. Se on epäparametrinen "musta laatikko"-algoritmi, jonka ennusteiden selittämiseksi tarvitaan erillisiä selkoälyn (=Explainable AI eli XAI) menetelmiä.

Malli toimii tarkastettujen yritysten populaation sisällä välttävän hyvin ja sen AUC on reilu 0,7. Malli pystyy tekemään eron tarkastettujen valkoisten ja harmaiden yritysten välille. Substanssin ollessa niinkin epäselvä ja vaikeasti tavoiteltava kuin pimeästi maksetut palkat, voidaan mallia pitää riittävän hyvänä mielekkään estimoinnin tekemiseen. Tietysti yleistys laajempaan joukkoon on oma asiansa, jonka validointiin voidaan käyttää vain asiantuntijaosaamista.

Yrityskohtainen pimeiden palkkojen suuruus

Pimeiden palkkojen suuruutta ennustetaan vain yhdellä mallilla, koska siinä bayesilainen lähestymistapa ja satunnaistarkastusten sisältämä informaatio ei tuota suoranaista lisäarvoa. Tiedämme, että satunnaistarkastuksissa löydetty virheen suuruus on pienempi kuin riskiperusteisesti tehdyissä tarkastuksissa, mutta bayesilainen lähestymistapa ei silti tuo lisäarvoa. Itse asiassa bayesilaisen lineaarisen regression tuottamat ennusteet olivat suurempia kuin XGBoostia ja Elastic Net -regressiota yhdistelevät ensemble-mallin, koska epäparametrinen XGBoost onnistuu ennustamaan paremmin pieniä eurosummia. Ylipäättään virheen suuruuden ennustaminen regressiomallilla on huomattavan haastavaa ja ennustevirheet isompia kuin binääristä luokittelua tehtäessä.

Regressiomalli rakennetaan tarkastettujen yritysten kohdejoukon siihen osajoukkoon, joilta on verotarkastuksessa löydetty pimeää palkanmaksua. Virheen suuruutta arvioidessa ei ole kannattavaa huomioida niitä tarkastettuja yrityksiä, joilla pimeää palkanmaksua ei ole havaittu. Jos nämä yritykset otettaisiin mukaan harjoitusaineistoon, tulisi virheen määrän arvio auttamatta epärealistisen pieneksi. Koska lopullinen arvioitu pimeän palkan määrä tapahtuu

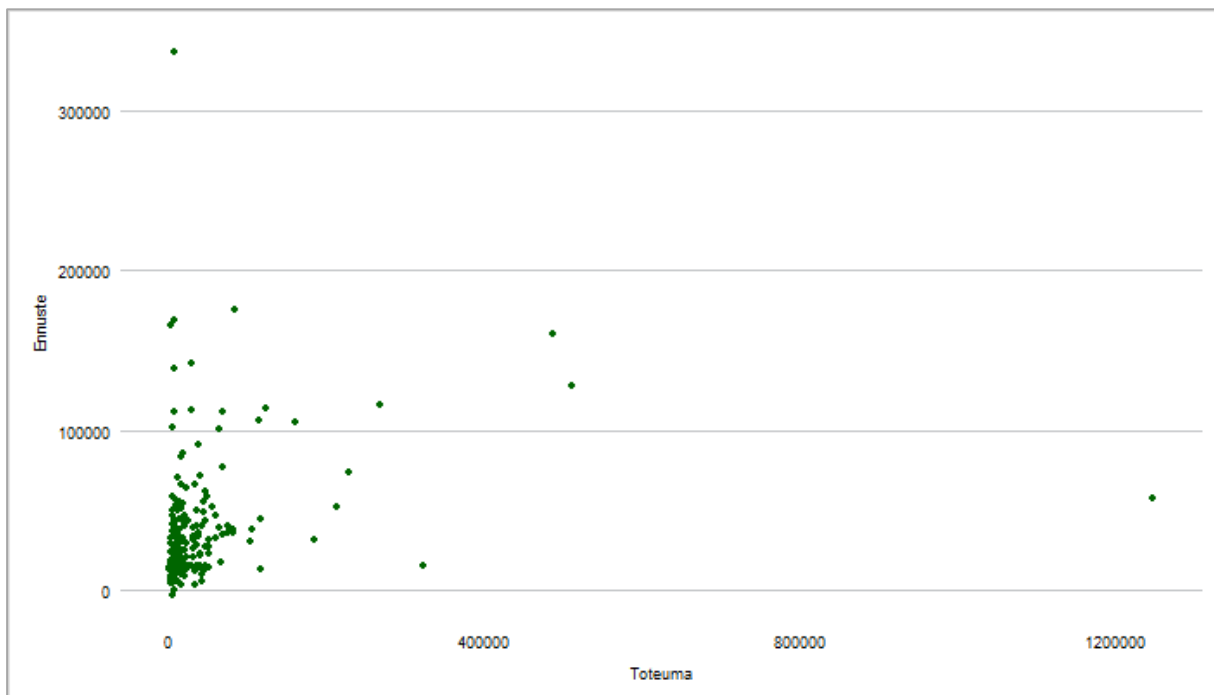
⁹¹ Vastemuuttuja on selvityksen keskeinen mielenkiinnon kohteena oleva ominaisuus, tässä tapauksessa pimeästi maksetut palkat.

⁹² Bayesilaisessa tilastotieteessä lopputulos on ns. posteriorijakauma, joka saadaan kaavalla $priori \times data$. Priorijakauma tarkoittaa käytännössä tutkijan uskoa tutkittavasta asiasta ennen kuin dataa on havaittu lainkaan. Tässä tapauksessa priorijakauma kuvaa siis selvityksen laatijoiden etukäteiskäsitystä harmaan työn esiintyvyydestä ennen kuin tarkastusdata on havaittu. Tämä etukäteiskäsitys saadaan vanhojen satunnaistarkastusten avulla.

virheen koon ja sen todennäköisyyden kertolaskuna, tulee yrityksen vähäriskisyys huomioidua jo riskin todennäköisyyden arvioissa.

Regressiomallina toimii XGBoostin ja Elastic Net -regression ensemble-malli, jonka lopullisissa arvioissa yhdistetään kummankin mallin tekemiä ennusteita. Lineaariseen regressioon pohjautuva Elastic Net tuottaa huonompia ennusteita pienempien virheiden joukossa, mutta se korjaa XGBoostin taipumusta maltillistaa kaikista suurimpien arvioiden kokoluokkaa. Pyrkimys on, että yhdistelmämallin komponentit korjaisivat toisensa puutteita ja näin lopullisista ennusteista tulisi parempia ja realistisempia. Tarkastettujen harmaiden yritysten joukosta otetulle testidatalla siivulle tehdyt ennusteet paljastavat, että regressiomalli tekee isoja virheitä aiemmin tuntemattomien yritysten kohdalla; mallin MAE (Mean Absolute Error) on noin 40 000 euroa. Kun ennustettavat yritykset sisältyvät treenidataan, MAE tippuu vain noin 11 000 euroon.

Kuvio 1. Regressiomallin antama ennuste vs. toteuma



Lähde: Verohallinto 5–8/2021.

Palkansaajien ja osakkaiden pimeät palkat

Mallit päättelivät vain asiakkaan pimeänä teetetyyn työn kokonaissummaa jakamatta sitä osakkaiden ja palkansaajien kesken peiteltyihin osinkoihin ja pimeisiin palkkoihin. Jotta kokonaissumma saadaan jaoteltuna osatekijöihin, rakennetaan erillinen koneoppimismalli, joka tarkastettujen, pimeää työtä teittäneiden yritysten populaation pohjalta päättelee kullekin yritykselle jakosuhteen pimeään työhön ja peiteltyjen osinkoihin. Malli pyrkii erottelemaan ennustavien muuttujien avulla palkansaajien pimeää palkkaa maksavat yritykset peiteltyjä osinkoja osakkaalle maksavista ja tekee sitten kullekin tarkastamattomalle yritykselle arvion, missä suhteessa sen kokonaissumma pitäisi tulkita pimeiksi palkkoiksi ja peiteltyiksi osingoksi. Mallin algoritmina toimii päätöspuu-pohjainen XGBoost.

Todennäköisyys valikoitua verotarkastukseen

Propensiteettipisteillä tarkoitetaan havainnon X todennäköisyyttä päätyä interventioryhmään, tässä tapauksessa verotarkastukseen. Propensiteettipisteitä käytetään aineiston valikoitumisharhan pienentämiseen, kun ennustetaan todennäköisyyttä sille, että yritys on maksanut pimeitä palkkoja. Propensiteettipisteet päätellään yrityksille niin, että vastemuuttujana toimii 0/1-tieto siitä, onko yritys joutunut ajanjaksolla 2017–2020 tarkastetuksi. Harjoitusaineistoon valikoidaan satunnaisotos riveistä niin, että kukin yritys esiintyy siinä vain kertaalleen satunnaiselta verovuodelta, mutta jos yritys on yhtenäkin neljästä vuodesta tarkastettu, tulee siltä valituksi tarkastettu. Tähän osajoukkoon opetetulla mallilla ennustetaan koko populaation yrityksille jokaiselle eri verovuoden riville todennäköisyys joutua verotarkastukseen. Tällä tavalla pyritään spesifisti mallintamaan verotarkastukseen joutumista ja näin valikoitumisharhasta aiheutuvaa vinoumaa.

Saaduista propensiteettipisteistä tehdään yksi uusi selittävä muuttuja, jota tarkastustulosta ennustavat mallit hyödyntävät. Tällä tavalla valikoitumisharhaa saadaan formaalisti mallinnettua ja torjuttua. Bayesilaisessa logistisessa regressiossa propensiteettipisteet toimivat vakioivana tekijänä, mutta päätöspuu-pohjaisessa XGBoostissa propensiteettipisteiden vaikutus on epäsuorempi. Propensiteettipisteitä kokeiltiin myös mallin painoina, mutta tällöin tuloksena oli heikentynyt ennusteteho.

Menetelmän vahvuudet ja heikkoudet

Asiakaskohtaisen riskisyyden päättelyn paras puoli on, että jokaiselle yritykselle saadaan yksilöllinen arvio sen pimeän työn teettämisen riskisyydestä sekä arvio pimeän palkan määrästä. Yksilötasolle viety arvio ottaa jokaisen yrityksen ominaisuudet tarkalla tasolla huomioon ja mahdollistaa myös monipuolisen aineiston jatkojalostamisen. Menetelmän etuna on lisäksi se, että se on käytetyistä menetelmistä ainoa, josta voidaan saada objektiivista palautetta sen toimivuudesta; tosin ennustetehoa voidaan testata ainoastaan tarkastettuihin yrityksiin, jolloin malli operoi omassa, pienessä, vinoutuneessa todellisuudessaan eikä ennustetehon metriikoita voida yleistää pätemään koko yrityspopulaatiossa.

Menetelmän suurin heikkous on siinä, että se on altis harjoitusaineiston vinoumille, jotka kerätyvät koko yrityspopulaatioon ennustettaessa. Siinä missä aiemmin esitellyt kaltaistamisen menetelmät ovat nimenomaisesti suunniteltu poistamaan valikoitumisharhaa, joudutaan asiakaskohtaista odottamaa laskiessa käyttämään propensiteettipisteitä ja muita kiertoteitä ongelman ratkaisemiseksi. Epäparametrissa XGBoostia käytettäessä ei voida myöskään millään tavalla hyödyntää satunnaistarkastuksia; tämän puutteen kiertämiseksi käytettiin bayesilaista logistista regressiota, josta kerrotaan seuraavassa alaluvussa tarkemmin.

Asiakaskohtaisen odottaman laskenta bayesilaisella logistisella regressiolla

Tässä alaluvussa kuvataan, kuinka yritysکوhtainen pimeiden palkkojen esiintymisen todennäköisyys laskettiin bayesilaisella logistisella regressiolla. Aikaisemmin DBSCAN-menetelmän yhteydessä käytettiin bayesilaista tilastollista päättelyä, jonka avulla saatiin hyödynnettyä aikaisempina vuosina 2014–2016 tehtyjä satunnaistarkastuksia asettamalla niiden avulla priorijakauma pimeän työn esiintyvyyden parametrille p . Bayesilaista lähestymistapaa voidaan hyödyntää myös asiakaskohtaisen pimeän työn odottaman laskemiseen.

Bayesilainen logistinen regressio mahdollistaa vanhojen satunnaistarkastusten sisältämän informaation hyödyntämisen mallin rakennuksessa vakiokertoimen priorijakauman asettamisen avulla. Logistisen regression vakiokerroin, eli asiakkaan saama pisteytys kaikkien selittävien muuttujien ollessa 0, voidaan asettaa melkoisen tiukasti noudattamaan satunnaistarkastuksissa havaittua pimeän työn osuutta μ .

Matemaattisessa muodossa bayesilaisen logistisen regression kaava on seuraava:

$$T_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) \sim (\alpha_i + \beta_1 x_{i1} + \dots + \beta_n x_{in})$$

$$\alpha_i \sim \text{Normal}\left(\frac{\log(\mu)}{1-\mu}, 0.05\right)$$

$$\beta_1 \dots \beta_n \sim \text{Normal}(0, 0.5)$$

Siis asiakkaan pimeän työn teettämisen todennäköisyys noudattaa Bernoulli-jakaumaa parametrilla p , joka taas saa arvonsa kaavan $\text{logit}(p_i) \sim (\alpha_i + \beta_1 x_{i1} + \dots + \beta_n x_{in})$ lopputuloksena, jossa α_i on vakio-termi ja $\beta_n x_{in}$ muuttujan n arvo kerrottuna muuttujan n regressiokertoimella.

Vakio-termin α_i priorijakauma noudattaa normaalijakaumaa, jonka keskiarvo tulee satunnaistarkastusten havaitusta teetetyn pimeän työn osuudesta ja jonka keskihajonnaksi asetetaan pieni arvo 0,05 sen liikkumisvaraa rajoittamaan. Yksittäisten selittävien muuttujien regressiokertoimet taas asetetaan heikosti informatiivisiksi ja ne noudattavat kaikki samaa normaalijakaumaa keskiarvolla 0 ja keskihajonnalla 0.5.

Asettamalla vakiokertoimen α_i ennako-oletus satunnaistarkastusten havaittuun pimeän työn osuuteen, voidaan mallia ohjeistaa pitämään pimeän työn esiintyvyyttä lähtökohtaisesti harvinaisempana kuin se sille esitetystä verotarkastusten aineistossa on. Kun regressiokertoimien annetaan kuitenkin asettua järkevälle vaihteluvälille, saadaan aineiston mukaan riskisille yrityksille myös korkea todennäköisyys.

Menetelmän vahvuudet ja heikkoudet

Menetelmän hyvänä puolena verrattuna päätöspuihin perustuvaan XGBoostiin on se, että satunnaistarkastusten sisältämä informaatio saadaan annettua mallille formaalilla tavalla. Näin vinoutuneen harjoitusaineiston aiheuttamia ongelmia saadaan torjuttua. Menetelmän heikkous on se, että siinä joudutaan luopumaan tehokkaasta epäparametrisesta XGBoostista ja tyytyä heikommin toimivaan logistiseen regressioon. Näin mallin ennustetehto jää heikommaksi ja arviot yksittäisten yritysten riskisyydestä heikommaksi.