

VALTIONEUVOSTON  
SELVITYS- JA TUTKIMUSTOIMINTA

Jarkko Harju, Tomi Kyyrä, Olli Kärkkäinen,  
Tuomas Matikka ja Lauri Ojala

## Työn tarjonnan mallintaminen osana talouspolitiikan vaikutusarviointia

Marraskuu 2018

Valtioneuvoston selvitys-  
ja tutkimustoiminnan  
julkaisusarja 72/2018

# KUVAILULEHTI

<b>Julkaisija ja julkaisuaika</b>	Valtioneuvoston kanslia, 28.11.2018		
<b>Tekijät</b>	Jarkko Harju, Tomi Kyyrä, Olli Kärkkäinen, Tuomas Matikka ja Lauri Ojala		
<b>Julkaisun nimi</b>	Työn tarjonnan mallintaminen osana talouspolitiikan vaikutusarviointia		
<b>Julkaisusarjan nimi ja numero</b>	Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 72/2018		
<b>Asiasanat</b>	Työn tarjonta, diskreetti työn tarjonnan malli, työn tarjonnan jousto		
<b>Julkaisuaika</b>	Marraskuu, 2018	<b>Sivuja</b> 22	<b>Kieli</b> Suomi

## Tiivistelmä

Tässä raportissa käsittelemme työn tarjonnan mallintamista rakenteellisen diskreetin valinnan mallin avulla. Käymme läpi mallin estimointiin liittyviä haasteita ja raportoimme mallin avulla saatuja alustavia tuloksia suomalaisesta aineistosta. Diskreetti työn tarjonnan malli muodostaa koherentin ja johdonmukaisen välineen tutkia talouspolitiikan vaikutuksia työllisyyteen ja sitä kautta edelleen tulonjakoon. Mallin avulla voidaan myös arvioida politiikkapäätösten hyvinvointivaikutuksia. Rakenteellinen mallintaminen mahdollistaa myös sellaisten politiikka-muutosten tarkastelun, joihin ei ole tarjolla luotettavaa kokeellista tai kvasi-kokeellista tutkimusasetelmaa. Alustavien tulosten mukaan diskreetti työn tarjonnan malli kuvaa melko hyvin havaittuja työtunteja suomalaisessa aineistossa ja tulokset ovat melko hyvin linjassa kansainvälisen tutkimuskirjallisuuden kanssa. Mallin jatkokehittämisen keskeisimpänä kohteena olisi saada SISU-mikrosimulointimalli toimimaan osana työn tarjonnan mallintamista, mikä merkittävästi laajentaisi mallin käytettävyyttä niin politiikkavalmistelussa kuin akateemisessa tutkimuksessaakin. Vasta tämä mahdollistaisi politiikkamuutosten työllisyys- ja tulonjakovaikutusten *ex ante* -arvioinnit.

Tämä julkaisu on toteutettu osana valtioneuvoston vuoden 2018 selvitys- ja tutkimussuunnitelman toimeenpanoa (tietokayttoon.fi).

Julkaisun sisällöstä vastaavat tiedon tuottajat, eikä tekstisisältö välttämättä edusta valtioneuvoston näkemystä.

# PRESENTATIONSBLAD

<b>Utgivare &amp; utgivningsdatum</b>	Statsrådets kansli, 28.11.2018		
<b>Författare</b>	Jarkko Harju, Tomi Kyyrä, Olli Kärkkäinen, Tuomas Matikka och Lauri Ojala		
<b>Publikationens namn</b>	Modellering av arbetsutbudet som en del av bedömningen av den ekonomiska politikens konsekvenser		
<b>Publikationsseriens namn och nummer</b>	Publikationsserie för statsrådets utrednings- och forskningsverksamhet 72/2018		
<b>Nyckelord</b>	Arbetsutbud, diskret modellering av arbetsutbud, flexibilitet i arbetsutbudet		
<b>Utgivningsdatum</b>	November, 2018	<b>Sidantal</b> 22	<b>Språk</b> Finska

## Sammandrag

I rapporten beskriver vi modellering av arbetsutbudet med en modell som bygger på diskreta val. Vi redogör för utmaningar med estimeringen och rapporterar preliminära resultat som erhållits med modellen utifrån data från Finland. Diskret modellering av arbetsutbudet erbjuder ett koherent och konsekvent verktyg för att analysera vilka effekter den ekonomiska politiken har på sysselsättningen och därigenom på inkomstfördelningen. Modellen kan också användas för analyser av vilka konsekvenser policybeslut får för välfärden. Med hjälp av strukturell modellering kan vi dessutom analysera sådana politikförändringar för vilka det saknas pålitliga experimentella eller kvasiexperimentella testupplägg. Enligt de preliminära resultaten beskriver den diskreta modellen av arbetsutbud de observerade arbetstimmarna i de finländska data relativt bra, och resultaten är relativt väl förenliga med internationell forskningslitteratur på området. Det viktigaste med tanke på den fortsatta utvecklingen är att integrera mikrosimuleringsmodellen SISU i modelleringen av arbetsutbudet, vilket skulle göra modellen betydligt mer användbar för både policyberedning och akademisk forskning. Det behövs för att möjliggöra förhandskonsekvensanalyser av effekterna av policyändringar på sysselsättningen och inkomstfördelningen.

Den här publikation är en del i genomförandet av statsrådets utrednings- och forskningsplan för 2018 ([tietokayttoon.fi/sv](http://tietokayttoon.fi/sv)).

De som producerar informationen ansvarar för innehållet i publikationen. Textinnehållet återspeglar inte nödvändigtvis statsrådets ståndpunkt

## DESCRIPTION

<b>Publisher and release date</b>	Prime Minister's Office, 28.11.2018		
<b>Authors</b>	Jarkko Harju, Tomi Kyyrä, Olli Kärkkäinen, Tuomas Matikka ja Lauri Ojala		
<b>Title of publication</b>	Modeling labor supply as part of economic policy evaluations		
<b>Name of series and number of publication</b>	Publications of the Government's analysis, assessment and research activities 72/2018		
<b>Keywords</b>	Labor supply, discrete choice model, labor supply elasticity		
<b>Release date</b>	November, 2018	<b>Pages</b> 22	<b>Language</b> Finnish

### Abstract

In this report we discuss how to model labor supply using Finnish micro-level data, and report some preliminary results based on a structural discrete choice model. Structural labor supply models offer a coherent tool for analyzing the labor supply impact of economic policies, and their implications on, for example, the income distribution. In addition to policy analysis and preparation, these models can be used in academic research, and they provide a method to analyze the effects of policy changes for which there is no credible quasi-experimental or experimental variation available. We find that the discrete labor supply model describes Finnish labor market conditions rather well, and preliminary results using Finnish data are broadly in line with previous literature. For future development of the model, we find that the most important step is to link the discrete labor supply model to the SISU-microsimulation model, which would enable a more straightforward policy preparation using the model, and increase the usability of the structural model in academic policy analysis.

This publication is part of the implementation of the Government Plan for Analysis, Assessment and Research for 2018 ([tietokayttoon.fi/en](https://tietokayttoon.fi/en)).

The content is the responsibility of the producers of the information and does not necessarily represent the view of the Government.



## SISÄLLYSLUETTELO

<b>1. Johdanto .....</b>	<b>1</b>
<b>2. Diskreetti työn tarjonnan malli .....</b>	<b>2</b>
2.1. Mallinnuksen perusidea .....	2
2.2. Työn tarjonta -valintojen todennäköisyydet .....	3
<b>3. Käytettävä aineisto .....</b>	<b>6</b>
3.1. Aineistolähteet .....	6
3.2. Aineiston kuvailu ja muokkaukset.....	8
<b>4. Mallin sovittaminen ja työn tarjonnan joustot .....</b>	<b>9</b>
4.1. Mallinnusvalinnat .....	9
4.2. Mallin estimointi ja työn tarjonnan jouston arviointi .....	12
4.3. Heterogeenisuustuloksia .....	15
<b>5. Työn tarjonta ja SISU-malli .....</b>	<b>18</b>
<b>6. Miten mallia voidaan tulevaisuudessa kehittää? .....</b>	<b>19</b>
<b>7. Lähteitä ja tausta-aineistoja.....</b>	<b>22</b>

# 1. JOHDANTO

Tässä raportissa tarkastellaan työkaluja, joiden avulla voidaan tutkia politiikkamuutosten vaikutuksia työn tarjontaan ja sitä kautta tulonjakoon. Raportti on osa Talouspolitiikan käyttäytymisvaikutukset – tutkimushanketta. Hankkeessa tarkastellaan politiikkamuutosten työllisyysvaikutusten arviointia simulointimallien näkökulmasta sekä työllisyysvaikutusten huomiointamisen vaikutusta tuloeromittareihin. Hankkeen ensimmäisen osan tulokset julkaistiin syyskuussa 2018 (Kärkkäinen ja Tervola 2018). Ensimmäisessä osassa kehitettiin menetelmä, jonka avulla työllisyysvaikutukset voidaan huomioida arvioitaessa talouspolitiikan tulonjakovaikutuksia mikrosimulointiympäristössä. Tarkasteluissa työn tarjonnan joustot kuitenkin otettiin annettuna. Tässä hankkeen toisessa osassa keskitytään työn tarjonnan mallintamiseen diskreetin valinnan mallin avulla.

Työn tarjonnan *ex ante* -analyysit nojaavat yleensä talusteoriasta johdettuihin rakenteellisiin malleihin, joissa yksilöt tai kotitaloudet valitsevat kulutuksensa ja työn tarjontansa annetuilla rajoitteilla (verot, tulonsiirrot yms.). Mallit voivat olla joko staattisia tai dynaamisia. Staattisissa malleissa työnteon taloudellisiin kannustimiin vaikuttavan politiikkamuutoksen jälkeen työn tarjonta sopeutuu välittömästi, eikä ajalla ole roolia. Sitä vastoin dynaamisissa malleissa mallinetaan ja simuloidaan yksilöiden käyttäytymistä ja kohtaloita yli ajan eri politiikkaskenaarioissa. Dynaamisten mallien avulla voidaan tarkastella politiikkatoimenpiteiden aiheuttamia muutoksia erikseen tietyn hetken tulojakaumassa ja elinkaarituloissa.

Tässä raportissa rajoitetaan staattisiin työn tarjonnan rakennemalleihin. Näistä kansainvälisesti sovelletuin on niin kutsuttu diskreetin valinnan malli, jossa yksilö valitsee työn tarjontansa pienestä joukosta eri vaihtoehtoja. Vaihtoehtoina voivat olla jättäytyminen työmarkkinoiden ulkopuolelle, työttömyys, osa-aikainen työskentely eri tuntimäärillä, kokoaikainen työskentely ja ylityöt normaalin kokoaikatyön lisäksi. Työn tarjonta -päätöstä tehdessään yksilö huomioi sekä vapaa-ajan että käytettävissä olevat tulot eri vaihtoehtoissa.

Tarkastelemme diskreetin työn tarjonnan mallin soveltamista suomalaiseseen rekisteriaineistoon. Vastaavaa tarkastelua ei ole Suomen osalta tehty aiemmin. Diskreetti työn tarjonnan malli muodostaa koherentin ja johdonmukaisen välineen tutkia talouspolitiikan työllisyysvaikutuksia sekä niiden vaikutusta esimerkiksi tulonjakoon. Mallin avulla voidaan tarkastella myös politiikkapäätösten hyvinvointivaikutuksia. Lisäksi rakenteellinen mallintaminen mahdollistaa sellaisten tutkimuskysymysten tarkastelun, joihin ei ole tarjolla luotettavaa kokeellista tai kvasi-kokeellista tutkimusasetelmaa.

Raportin empiiriset tarkastellut perustuvat Löfflerin, Peichelin ja Sieglochin (2018) kehittämään diskreetin valinnan rakennemalliin ja sen eri variaatioihin. Sovittamalla malli yksilötason aineistoon käytettävissä olevista tuloista ja työtunneista saadaan arviot mallin tuntemattomille parametreille. Tämän jälkeen voidaan simuloida työn tarjonta -päätöksiä halutulle populaatiolle vaihtoehtoisissa politiikkaskenaarioissa. Mallin avulla voidaan estimoida työn tarjonnan jousto palkan suhteen ja arvioida toteutettujen politiikkapäätösten työllisyysvaikutuksia tai harkinnassa olevien vaihtoehtoisten politiikkamuutosten vaikutuksia työllisyyteen. Lisäksi voidaan simuloida politiikkatoimien vaikutuksia myös muihin mielenkiinnon kohteena oleviin tulemiin, kuten tulonjakoon.

Vaikka Suomen kattavat yksilö- ja kotitalouskohtaiset rekisteriaineistot tarjoavatkin kansainvälisesti vertaillen hyvän lähtökohdan mallinnukselle, asettavat saatavilla olevat tilastoaineistot rajoitteita mallin soveltamiselle. Esimerkiksi useat tulonsiirrot ja tulot havaitaan vain

vuositasolla, kun taas työtunnit havaitaan vuoden yhtenä ajankohtana. Nämä tekijät hankaloittavat budjettirajoitteiden estimointia. Mallin soveltaminen edellyttää myös useiden mallinvalintojen tekemistä. Mallin estimoimista varten pitää muun muassa määrittää yksilön hyötyfunktio ja puuttuvat tuntipalkat pitää ennustaa työttömille.

Tässä raportissa arvioidaan, miten hyvin diskreetin valinnan malli kuvaa työn tarjontaa Suomen työmarkkinoilla ja kuinka herkkiä tulokset ovat erilaisille mallinmuutoksille ja taustaoletuksille. Tarkastelemme mallin tuottamia tuloksia koko aineistolle sekä erikseen miehille ja naisille, ja nuorille ja vanhemmille työntekijöille. Lisäksi pohdimme, mihin suuntaan rakenteellista analyysikehikkoa olisi syytä jatkossa kehittää, kun saatavilla olevien rekisteriaineistojen asettamat rajoitteet pidetään mielessä. Kehitettävä malli on dokumentoitu huolellisesti, ja sen soveltamisessa tarvittavat koodit ovat vapaasti hyödynnettävissä ja ladattavissa hankkeen Internet-sivuilta sekä SISU-mikrosimulointimallin etäkäytön käyttäjäfoorumilta. Työn yhtenä tavoitteena on laajentaa suomalaisten rekisteriaineistojen ja SISU-mallin käytettävyyttä työn tarjonnan tutkimisessa ja politiikka-arvioinnin työkaluna.

Raportin rakenne on seuraava. Luvussa 2 tarkastellaan työn tarjonnan mallintamista diskreetin valinnan mallin kehikossa sekä pohditaan tilastolliseen analyysiin liittyviä haasteita. Luvussa 3 esitellään mallin estimoinnissa käytettävät tilastoaineistot. Luvussa 4 esitellään käytännön mallinnuksen periaatteet ja keskustellaan mallinnukseen liittyvistä aineistovalinnoista ja oletuksista, sekä esitellään mallin avulla saatuja ensimmäisiä alustavia tuloksia. Luvussa 5 keskustellaan SISU-mallin käytöstä mallinnuksessa. Luvussa 6 tehdään lyhyt yhteenveto ja pohditaan mallin jatkokehittämiseen liittyviä haasteita ja mahdollisuuksia.

## 2. DISKREETTI TYÖN TARJONNAN MALLI

### 2.1. Mallinnuksen perusidea

Tässä luvussa tarjotaan lyhyt taustoitusta työn tarjonnan mallinnuksen teoriaan ja ekonometriaan. Lähtökohtana on yksilö, joka valitsee, montako tuntia hän työskentelee viikossa. Yksilön hyvinvointi riippuu kulutuksesta ja vapaa-ajasta. Kulutus määräytyy työnteosta saatujen nettotulojen ja muiden mahdollisten tulojen perusteella. Vapaa-aika muodostuu ajasta, jota ei käytetä työskentelyyn. Kulutuksen ja vapaa-ajan välillä on luonnollisesti ristiriita: molemmat lisäävät hyvinvointia, mutta lisätäkseen kulutustaan on yksilön tingittävä vapaa-ajastaan (ja päinvastoin). Yksilö valitsee kulutuksen ja vapaa-ajan yhdistelmän, joka maksimoi hänen hyvinvointinsa annetulla budjettirajoitteella.

Budjettirajoite riippuu tuntipalkasta, työtunneista sekä vero- ja tulonsiirtojärjestelmästä. Yksilö ei voi vaikuttaa tuntipalkkaansa eikä tuntipalkka riipu valittujen työtuntien määrästä. Työtunnit määräävät yksilön bruttoansiot annetulla tuntipalkalla. Verotus ja tulonsiirtojärjestelmä puolestaan vaikuttavat siihen, paljonko bruttoansioista jää käteen käytettäväksi kulutukseen. Näin ollen työtunnit määrittävät yksikäsitteisesti sekä kulutuksen että vapaa-ajan, joista yksilön hyvinvointi riippuu. Siksi yksilön valintaongelma pelkistyy optimaalisten työtuntien valintaan.

Yksilön oletetaan valitsevan työtuntinsa pienestä joukosta eri vaihtoehtoja. Yksi vaihtoehto on jättäytyä työttömäksi (tai työvoiman ulkopuolelle), jolloin vapaa-aikaa on paljon mutta rahaa kuluttamiseen vähän. Muut vaihtoehdot vastaavat osa-aikaista ja kokoaikaista työ-

kentelyä. On syytä korostaa, että mallinnuksessa työtuntien valintajoukko rajataan pieneksi lähinnä teknisten syiden vuoksi. Verrattuna malliin, jossa yksilö voi valita työaikansa mieltävaltaisen tarkasti (esim. minuutilleen), diskreetin valinnan mallin teoriakehikko on varsin yksinkertainen ja mallin sovittaminen dataan on helpompaa.

Mallinnusta varten on arvioitava käytettävissä olevat tulot eri vaihtoehdoissa. Tämä edellyttää sekä tietoa tuntipalkoista että vero- ja tulonsiirtojärjestelmän tarkkaa mallintamista. Optimaalinen työtuntien määrä vaihtelee yksilöiden välillä erilaisten lähtökohtien (ja mahdollisesti erilaisten mieltymysten) vuoksi. Esimerkiksi korkeasti koulutetun työntekijän marginaalivero lisätyötunnista on korkeampi kuin heikosti koulutetun, koska hänen tuntipalkkansa on korkeampi ja tuloverotus on progressiivista. Osalla henkilöistä on työtunneista riippumattomia pääomatuloja, osalla ei. Lisäksi tulonsiirtojärjestelmä kohtelee yksilöitä eri tavalla, koska tulonsiirrot riippuvat omien ansiotulojen ohella myös puolison tuloista, asumiskustannuksista ja lasten lukumäärästä.

Teoreettisen mallin ennustamaa työn tarjontaa, eli hyvinvoinnin maksimoivaa työtuntien määrää, verrataan yksilöiden tekemiin todellisiin valintoihin, ja mallin tuntemattomat parametrit valitaan siten, että teoreettinen malli kuvaa aineistossa havaittuja työtunteja mahdollisimman hyvin. Tätä varten on teoreettisesta mallista johdettava todennäköisyydet yksilöiden havaituille työn tarjonta -päätöksille, mikä puolestaan edellyttää useiden mallinnusvalintojen tekemistä. Seuraavaksi käymme läpi näitä valintoja varsin teknisellä tasolla. Lukija, joka ei ole mallinnuksen teknisestä puolesta kiinnostunut, voi huoletta hypätä suoraan seuraavaan lukuun.

## 2.2. Työn tarjonta -valintojen todennäköisyydet

Mallinnusta varten on muun muassa valittava hyötyfunktio. Hyötyfunktio määrittää sen, miten yksilön kulutus ja vapaa-aika tuottavat hänelle hyvinvointia. Yksi mahdollinen valinta on kvadraattinen hyötyfunktio, joka voidaan esittää muodossa

$$v(C, L) = \alpha_1 C + \alpha_2 C^2 + \beta_1 L + \beta_2 L^2 + \delta(C \cdot L),$$

jossa  $C$  on kulutus ja  $L$  on vapaa-aika. Koska tässä yksinkertaisessa mallissa yksilö ei voi säästää eikä lainata, kulutus vastaa hänen nettotulojaan:  $C = f(wh, I)$ , jossa  $w$  on tuntipalkka,  $h$  on työtunnit,  $I$  on työnteosta riippumattomat bruttotulot (esim. pääomatulot) ja  $f(\cdot)$  on funktio, joka muuntaa bruttotulot nettotuloksi. Keskeistä on pystyä mallintamaan vero- ja tulonsiirtojärjestelmä eli funktio  $f(\cdot)$  siten, että henkilön bruttotuloista saadaan muodostettua verojen ja tulonsiirtojen jälkeiset nettotulot vaihtoehtoisissa tiloissa. Apuna voidaan käyttää mikrosimulointia tai funktiota  $f(\cdot)$  voidaan approksimoida tilastollisen regressiomallin avulla.

Diskreetin valinnan mallissa keskeinen oletus on se, että yksilö valitsee työtuntinsa pienestä joukosta eri vaihtoehtoja:  $0, h_1, h_2, \dots, \bar{H}$ , jossa  $0$  vastaa työttömyyttä (tai työvoiman ulkopuolella oloa) ja  $\bar{H}$  on maksimituntimäärä, joka on käytettävissä työntekoon viikossa. Työtuntien valinta  $h$  määrittää sekä kulutuksen  $C$  että vapaa-ajan  $L$  yksikäsitteisesti. Koska  $L = \bar{H} - h$ , voidaan  $L$  korvata  $h$ :lla hyötyfunktiossa.

Yksilön valintaa mallinnettaessa hyötytasoon lisätään vielä stokastinen termi  $e_h$ , joten yksilön optimaalinen valinta maksimoi summan  $v(f(wh, I), h) + e_h$ . Stokastinen termi voi kuvata optimointivirhettä tai eri vaihtoehtoihin liittyviä havaitsemattomia, hyötyyn vaikuttavia tekijöitä. Stokastisesta termistä oletetaan, että se on identtisesti ja riippumattomasti jakautunut ja



että sen kertymäfunktio on muotoa  $\Pr(e_h \leq e) = \exp(-\exp(e))$ , jolloin työn tarjontaa annettulla tuntipalkalla kuvaa tavanomainen multinomial logit -malli.

Mallin estimointia varten tarvitaan tietoa työtunneista ja tuntipalkoista. Yksi vaihtoehto on käsitellä myös palkkaa endogeenisena muuttujana, ja mallintaa palkkojen määräytymistä simultaanisesti työtuntien kanssa. Tässä tapauksessa todennäköisyys sille, että henkilö on töissä ja hänen työtuntinsa ovat  $h$  ja palkkansa  $w$ , on

$$\Pr(H = h, W = w) = \frac{\exp\{v(f(wh, I), h)\}}{\sum_{s=0}^{\bar{H}} \exp\{v(f(ws, I), s)\}} g(w|x),$$

jossa  $g(\cdot)$  on palkan tiheysfunktio ja  $x$  on yksilön palkkaan vaikuttavien havaittujen ominaisuuksien (koulutus, ikä, sukupuoli jne.) vektori. Koska työttömien palkkaa ei havaita, pitää palkan suhteen ottaa odotusarvo, joten todennäköisyys olla työttömänä on

$$\Pr(H = 0) = E_w[\Pr(H = 0, W = w)] = \int \frac{\exp\{v(f(0, I), 0)\}}{\sum_{s=0}^{\bar{H}} \exp\{v(f(ws, I), s)\}} g(w|x) dw.$$

Integraalitermi aineiston työttömille vaikeuttaa mallin estimointia huomattavasti etenkin jos nettotulot halutaan arvioida mikrosimulointimallin avulla, koska työttömyyden todennäköisyys pitää laskea lukuisilla eri palkan arvoilla mallin estimoinnissa käytetyn iterointiprosessin sisällä. Esimerkiksi Van Soest ym. (2002) sekä Blundell ja Shephard (2012) mallintavat palkkoja ja työtunteja simultaanisesti.

Vaihtoehtoinen ja yleisemmin käytetty lähestymistapa on korvata puuttuvat palkkatiedot ennustetuilla tuntipalkoilla kaksivaiheisen menetelmän avulla. Ensimmäisessä vaiheessa estimoidaan erillinen palkkaregressio. Toisessa vaiheessa puuttuvat palkkatiedot (tai vaihtoehtoisesti kaikkien palkat) korvataan mallin sovitteella ennen työn tarjonta -päätöksen mallintamista. Koska tässä tapauksessa palkka otetaan annettuna, todennäköisyys sille, että yksilön työtunnit ovat  $h$ , on muotoa

$$\Pr(H = h | W = \hat{w}) = \frac{\exp\{v(f(\hat{w}h, I), h)\}}{\sum_{s=0}^{\bar{H}} \exp\{v(f(\hat{w}s, I), s)\}},$$

jossa  $\hat{w}$  on palkan sovite mallinnuksen ensimmäisestä vaiheesta (ks. esim. Bargain ym. 2014). Sama yhtälö pätee myös työttömyyden ( $h = 0$ ) todennäköisyydelle. Taustalla on kuitenkin implisiittinen oletus siitä, että yksilön päätösten perustana oleva odotettu palkka kyetään arvioimaan ilman ennustevirhettä. Oletus ei ole kovin realistinen, koska yksilöllä on aina enemmän tietoa oman työpanoksensa markkina-arvosta kuin tutkijalla, jonka laatima palkkaennuste nojautuu saatavilla olevan tutkimusaineiston vajavaisiin tietoihin. Ennustevirhe voidaan kuitenkin ottaa huomioon, jolloin valinnan  $h$  todennäköisyydeksi tulee

$$\Pr(H = h | W = \hat{w}) = \int \frac{\exp\{v(f((\hat{w} + u)h, I), h)\}}{\sum_{s=0}^{\bar{H}} \exp\{v(f((\hat{w} + u)s, I), s)\}} m(u) du,$$

jossa  $u$  on ennustevirhe ja  $m(\cdot)$  on ennustevirheen tiheysfunktio (ks. esim. Van Soest 1995, Thoresen ja Vattø 2015, Bargain ym. 2014 ja Löffler ym. 2018).

Mallissa oletetaan, että yksilö voi valita vapaasti haluamansa työtunnit ja että hänen hyötynsä riippuu vain kulutuksesta ja vapaa-ajasta. Molemmat ovat varsin vahvoja oletuksia. Esimerkiksi työttömyyteen voi liittyä häpeän tunne eli stigma, joka tekee siitä ikävän tilan silloinkin, kun työttömyyden aikaiset tuet tarjoavat kohtuullisen kulutustason yhdistettynä mak-

simaaliseen vapaa-aikaan. Stigman takia yksilö saattaa valita työssäkäynnin, vaikka se ei pelkän kulutuksen ja vapaa-ajan näkökulmasta olisikaan optimaalista. Toisaalta työttömäksi joudutaan usein itsestään riippumattomista syistä esimerkiksi irtisanomisen myötä. Uuden työpaikan löytäminen vie aikaa etenkin laskusuhdanteessa, jolloin avoimia työpaikkoja on vähän ja niistä kilpailee moni muukin. Siksi työttömyys tietyllä havaintohetkellä saattaa olla olosuhteiden sanelema tilapäinen tila eikä yksilön optimaalinen valinta. Näistä syistä johtuen työttömyyden ennustaminen pelkästään kulutuksen ja vapaa-ajan maksimoivana vaihtoehtona on ongelmallista.

Mallin perusversiolla on lisäksi taipumus aliarvioida työttömien osuutta ja yliarvioida osa-aikatyötä tekevien osuutta. Useimmilla toimialoilla osa-aikaisten työntekijöiden osuus on varsin pieni. Mikäli tämä olisi suoraa seurausta työntekijöiden vapaa-ajan ja kulutuksen preferensseistä, mallin pitäisi pystyä ilmiö selittämään (olettaen, että valittu hyötyfunktio approksimoi preferenssejä riittävän hyvin). Mutta osa-aikaisen työskentelyn vähyys voi johtua myös muista syistä. Osa-aikainen työ voi olla työnantajille hankalaa tai kallista järjestää, jolloin osa-aikaisia töitä on vain vähän tarjolla. Jos vain harvat työnantajat tarjoavat osa-aikaisia töitä, niiden löytäminen on vaikeaa. Lisäksi työssäkäyntiin liittyy kiinteitä kustannuksia esimerkiksi työmatkoihin kuluva ajan ja euromääräisten liikkumiskustannusten muodossa. Tällaiset kustannukset tekevät lyhyiden työpäivien tekemisestä vähemmän houkuttelevaa suhteessa työttömyyteen, mikä osaltaan vähentää osa-aikaisiin työpaikkoihin hakeutumista.

Kysyntäpuolen rajoitteiden vuoksi yksilö ei siis välttämättä pysty vapaasti valitsemaan sitä työtuntien määrää, joka maksimoisi hänen hyötytasonsa. Työttömyyteen liittyvä stigma ja työnteon kiinteät kustannukset puolestaan tarkoittavat sitä, että vain vapaa-ajasta ja kulutuksesta riippuva hyötyfunktio ei välttämättä kuvaa todellisuutta kovin hyvin. Näitä tekijöitä voidaan epäsuorasti huomioida lisäämällä malliin valintakohtaisia kiinteitä vaikutuksia (joiden voidaan sallia vaihtelevan yksilön havaittujen ominaisuuksien mukaan). Tässä tapauksessa yksilö siis valitsee vaihtoehdon, joka maksimoi summan  $v(f(wh, I), h) + \mu_h + e_h$ , jossa  $\mu_h$  on valintakohtainen vakio (ks. esim. Van Soest 1995 ja Löffler ym. 2018). Tällaisia kiinteitä vaikutuksia lisätään työn tarjonta -malleihin usein *ad hoc* -pohjalta, jolloin niille ei voida antaa yksiselitteistä tulkintaa. Tosin ne voidaan myös johtaa teoriakehikosta, jossa yksilö kohtaa kysyntärajoitteita ja tekee valintansa latenttien työpaikkojen joukosta (ks. Dagsvik ym. 2014).

Mallia voidaan edelleen joustavoittaa olettamalla, että työnteon ja vapaa-ajan preferenssit vaihtelevat yksilöiden välillä. Preferenssien heterogeenisuus voidaan huomioida sallimalla hyötyfunktion parametrien vaihdella yksilön havaittujen ja/tai havaitsemattomien tekijöiden mukaan. Esimerkiksi hyötyfunktion parametri  $\beta_1$ , joka kuvaa vapaa-ajan arvostusta, voi olla muotoa

$$\beta_1 = \beta_{10} + \beta_{11}z + \vartheta,$$

jossa  $z$  on vektori yksilön havaituista ominaisuuksista ja satunnaistermi  $\vartheta$  kuvaa havaitsemattomien ominaisuuksien vaikutusta (ks. esim. Van Soest 1995 ja Löffler ym. 2018). Satunnaistermille pitää olettaa jonkin jakauma ja eri valintojen todennäköisyyksiä johdettaessa satunnaistermien suhteen pitää ottaa odotusarvo, koska sen arvoa ei havaita. Tämä johtaa ylimääräiseen integraalitermiin todennäköisyyden yhtälössä.

Yllä on esitelty diskreetin valinnan mallin eri variaatiota ja kuvailtu niiden tuottamia todennäköisyyksiä havaituille valinnoille. Nämä todennäköisyydet muodostavat mallin uskottavuusfunktion. Koska kyseessä on epälineaarinen todennäköisyysmalli, on mallin tuntemattomat

parametrit ratkaistava maksimoimalla mallin uskottavuusfunktio jonkin iteratiivisen optimointimenetelmän avulla. Tietyissä tapauksissa mallin estimointi onnistuu yleisesti käytettyjen tilastollisten ohjelmistojen valmiiden logit-mallioptioiden avulla.

Jos valitaan hyötyfunktio, joka on lineaarinen parametriensa suhteen (kuten kvadraattinen hyötyfunktio), korvataan havaitsemattomat (tai kaikki) palkat regressiomallin soviteilla ja jätetään ennustevirheet huomiotta, työn tarjonta -pääöstä kuvaa standardi multinomial logit -malli. Tällaisen mallin parametrit on helposti estimoitavissa lähes millä tahansa tilastollisella ohjelmistolla, eikä kiinteiden vaikutusten tai havaitun heterogeenisuuden lisääminen hyötyfunktion parametreihin muuta tilannetta. Jos lisäksi oletetaan, että jotkin hyötyfunktion parametrit ovat satunnaisia, kyseessä on standardi mixed logit -malli. Myös tällaisen mallin estimointi on kohtuullisen suoraviivaista useimpien tilastollisten ohjelmistojen avulla. Tilanne vaikeutuu huomattavasti, jos palkkaa käsitellään endogeenisena muuttujana tai jos palkkojen ennustevirheet huomioidaan kaksivaiheisen menetelmän tapauksessa. Näissä tapauksissa tarvitaan erikseen räätälöityjä työkaluja, kuten Löfflerin ym. (2018) kehittämää Stata-ohjelmistoa.

Kuten yllä olevista yhtälöistä nähdään, mallin uskottavuusfunktio saattaa sisältää useita integraalitermejä, mikä vaikeuttaa estimointia. Tällaisten mallien estimoinnissa turvaututaan usein simuloidun suurimman uskottavuuden (simulated maximum likelihood) menetelmään, jonka avulla voidaan approksimoida mutkikkaita uskottavuusfunktioita, joilla ei ole suljetun muodon ratkaisua. Menetelmä on kuitenkin laskennallisesti raskas, joten pitkät estimointiajat asettavat omat rajoitteensa.

Seuraavissa luvuissa käymme läpi Löfflerin ym. (2018) kehittämän mallikehikon soveltamista suomalaiseseen aineistoon ja tarkastelemme, miten eri mallinnusvaihtoehdot sopivat työn tarjonnan arvioimiseen aineistorajoitteiden puitteissa.

## 3. KÄYTETTÄVÄ AINEISTO

### 3.1. Aineistolähteet

Jotta työn tarjonnan tarkastelu olisi rakenteellisen mallin avulla mahdollista, tarvitaan luotettava ja kattava aineisto henkilöiden työtunneista ja tuntipalkoista. Lisäksi on keskeistä pystyä mallintamaan vero- ja tulonsiirtojärjestelmä siten, että henkilön hypoteettisista bruttotuloista saadaan muodostettua verojen ja tulonsiirtojen jälkeiset nettotulot. Koska tulonsiirrot riippuvat myös puolison tuloista, lasten lukumäärästä ja asuinkustannuksista, tarvitaan tietoa myös näistä tekijöistä. Nettotulot voidaan arvioida kahdella vaihtoehdoisella tavalla: (1) käytetään verot ja tulonsiirrot mallintavaa simulointimallia eli SISU-mallia tai (2) estimoidaan nettotulojen soviteet henkilöille hyödyntämällä riittävän monipuolista aineistoa todellisista netto- ja bruttotuloista sekä verotuksen ja tulonsiirtojen kannalta olennaisista taustatekijöistä. Luvussa 4 tehtävissä tarkasteluissa keskitytään jälkimmäiseen vaihtoehtoon, ja SISU-mallin yhdistämisestä työn tarjonnan mallintamiseen keskustellaan tarkemmin luvussa 5.

Mallinnuksessa voidaan lisäksi käyttää taustamuuttujia, joiden voidaan olettaa vaikuttavan työn ja vapaa-ajan välisiin valintoihin eli preferensseihin. Näitä muuttujia ovat esimerkiksi ikä, sukupuoli ja pienten lasten lukumäärä perheessä. Lisäksi tuntipalkkojen estimointi työtömille on keskeinen osa työn tarjonnan mallintamista. Tätä varten tarvitaan tietoa sekä

tuntipalkoista että palkkoihin vaikuttavista tekijöistä, kuten koulutuksesta, iästä ja ammatista. Tuntipalkkojen estimointitarkkuuden voidaan yleisesti ottaen olettaa paranevan, mikäli käytävissä on tietoa myös henkilöiden palkka- ja työllisyshistoriasta sekä aiemmasta toimialasta. Palkkaregressiosta keskustelemme tarkemmin luvussa 4.

Työn tarjonnan mallinnusta varten olemme joutuneet yhdistämään tietoja kolmesta Tilastokeskuksen aineistosta:

- **Palkkarakennetilastosta** on poimittu tiedot tehdyistä viikkotyötunneista ja tuntipalkoista. Tilastoa varten nämä tiedot on kerätty työllistäviltä yrityksiltä loka-joulukuussa ja julkisen sektorin (kunta, valtio) toimijoilta lokakuulta. Tilastosta puuttuvat kaikki alle viiden hengen yritysten työntekijät, osa suurempien järjestäytymättömien yritysten työntekijöistä sekä ne yksityisen sektorin palkansaajat, joiden palvelussuhde on alkanut tai loppunut keskellä tilastointikuukautta.<sup>1</sup> Aineisto käsittää tilastovuodesta riippuen noin 55–75 prosenttia yksityisen sektorin työntekijöistä vuoden viimeisen neljänneksen ajalta. Palkkarakennetilasto on kattava julkisen sektorin osalta, joten julkisen sektorin työntekijät ovat aineistossa yliedustettuja. Lisäksi aineistosta luonnollisesti puuttuvat tuntipalkat tarkasteluhetkenä työttömänä olleille henkilöille. Palaamme tähän ongelmaan ja sen vaikutukseen mallintamisessa tarkemmin seuraavassa alaluvussa.
- **Mikrosimulointimallin rekisteriaineisto** on SISU-mallia varten koottu tutkimusaineisto, joka kattaa vuosittain noin 15 prosenttia Suomen asuntokunnista ja henkilöistä. Tämä aineisto muodostaa mallinnuksemme otoksen, johon yhdistämme tarvittavat lisätiedot Palkkarakenne- ja FLEED-aineistoista. SISU-aineisto on muodostettu yhdistämällä useita hallinnollisia rekistereitä ja se kuvaa henkilöiden vuoden aikana saamia palkka-, yrittäjä- ja omaisuustuloja sekä asuntokuntien saamia ja maksamia tulonsiirtoja. Näiden tulojen perusteella muodostuu henkilön ja asuntokunnan käytävissä olevat tulot. Asuntokunta muodostuu samassa asuinhuoneistossa vakituisesti asuvista henkilöistä. Henkilön budjettirajoitteen mallintamista varten poimimme tästä aineistosta henkilön omaisuustulot ja puolison nettotulot. Lisäksi aineistosta saatavia taustatietoja, kuten ikää, kotikuntaa ja siviilisäätyä, käytetään selittävinä tekijöinä tuntipalkkojen ennustamisessa.
- **FLEED-aineisto** on useista eri hallinnollisista rekistereistä muodostettu linkitetty työntekijä-työnantaja-aineisto. Se kattaa kaikki 15–70-vuotiaat Suomessa asuvat henkilöt (pl. Ahvenanmaalla asuvat). Aineisto sisältää tietoja henkilöiden perusominaisuuksista, perheestä, asumisesta, työsuhteista, työttömyysjaksoista ja koulutuksesta. Monet tiedoista ovat vuositasolla, mutta aineisto sisältää työ- ja työttömyysjaksojen alkamis- ja päättymispäivämäärät. Näiden tietojen avulla tunnistamme henkilöt, jotka olivat työttömänä tietynä ajankohtana. Lisäksi käytämme aineiston tietoja toimialasta ja työmarkkinahistoriasta, kun mallinamme tuntipalkkojen määräytymistä.

Tämän aineistokokonaisuuden avulla voimme muodostaa työn tarjonta -mallin kannalta keskeiset muuttujat sekä ottaa huomioon monia työn tarjonnan ja palkanmuodostuksen kannalta keskeisiä taustatekijöitä. Käyttämällä SISU-aineistoa mallinnuksen perusaineistona voimme melko vaivattomasti käyttää myös SISU-mallia käytävissä olevien tulojen mallintamiseen, sillä SISU-aineisto on valmiiksi siinä muodossa, jolla kyseinen simulointi voidaan toteuttaa. Tässä raportissa käytämme vuoden 2015 aineistoja.

---

<sup>1</sup> Järjestäytyneiden yritysten työntekijöiden tiedot pohjautuvat työnantajajärjestöjen kattaviin jäsenkyselyihin. Tilastokeskus täydentää tilastoa poimimalla otoksen järjestäytymättömien yritysten työntekijöistä.

## 3.2. Aineiston kuvailu ja muokkaukset

Ennen varsinaisen työn tarjonnan mallinnuksen toteuttamista on aineistoa rajattava sekä aineiston muuttujia muokattava joiltain osin. Aiempaa kirjallisuutta seuraten rajoitamme tarkastelun työikäisiin eli 20–65-vuotiaisiin henkilöihin. Lisäksi tiputamme eläkeläiset, opiskelijat, työkyvyttömyyseläkkeellä olevat sekä yrittäjät aineistosta pois. Taulukkoon 1 on koottu mallinnuksessa käytettävän aineiston keskeisten muuttujien kuvailutiedot.

**Taulukko 1. Aineiston keskeisten muuttujien kuvailu (vuosi 2015)**

	Keskiarvo	Keskihajonta
<u>Ikä</u>	44,2	11,0
<u>Sukupuoli</u> (1=nainen, 0=mies)	50 %	0.5
<u>Palkkatulot</u>	31 684	29 056
<u>Ansiotulot</u>	35 108	27 650
<u>Työtunnit per viikko</u> (Palkkarakenneaineisto)	36,7	7,6
<u>Tuntipalkka</u> (Palkkarakenneaineisto)	20,1	21,4
<u>Pääomatulot</u>	1 807	52 806
<u>Saadut tulonsiirrot</u>	4 259	6 801
<u>Saanut tulonsiirtoja</u> (1=kyllä, 0=ei)	52 %	0.50
<u>Saanut asumistukea</u> (1=kyllä, 0=ei)	9 %	0.29
<u>Saanut toimeentulotukea</u> (1=kyllä, 0=ei)	6 %	0.23
<u>On alle 7-vuotiaita lapsia</u> (1=kyllä, 0=ei)	20 %	0.40
<u>On alle 3-vuotiaita lapsia</u> (1=kyllä, 0=ei)	10 %	0.39

Rekisteripohjaisten tietojen osalta havaintojen lukumäärä on 342 937. Palkkarakenneaineiston havaintojen lukumäärä on 166 658. Havaintojen lukumääriä ei ole korotettu väestötasolle. Aineistot on rajattu 20–65-vuotiaisiin henkilöihin. Lisäksi eläkeläiset (eläkekuukaudet=12), opiskelijat (opintotuki>0), työkyvyttömät (työkyvyttömyyseläke>0) sekä yrittäjät (yritystulo>5000 euroa vuodessa) on rajattu pois aineistoista.

Tehdyt viikkotyötunnit ja tuntipalkat ovat keskeisimmät mallinnuksessa käytettävät muuttujat. Nämä muuttujat ovat osa Palkkarakennetilastoa. Kuten edellä todettiin, Palkkarakennetilastosta puuttuvat kaikki alle viiden hengen yrityksissä työskentelevät sekä osa isompien yritysten työntekijöistä, joten julkisen sektorin työntekijät ovat yliedustettuja. Lisäksi aineistosta luonnollisesti puuttuvat tuntipalkat tarkasteluhetkenä työttömille oleville henkilöille. Mallinnusta varten täytyy määritellä erikseen ne henkilöt, jotka ovat työttömiä. Työttömiksi ei voida automaattisesti luokitella kaikkia henkilöitä, joiden palkka- ja työtuntitietoja ei ole Palkkarakennetilastosta saatavilla, sillä he saattavat työskennellä sellaisessa yrityksessä, joka ei toimita tietoja tilastoon. Lisäksi henkilö saattaa olla työssä pitkiäkin ajanjaksoja vuoden aikana mutta on juuri Palkkarakennetilaston keräämisen aikaan työtön tai työvoiman ulkopuolella. Tämä hankaloittaa tarkastelua, sillä mallinnuksessa käytetään useita vuositasolla mitattuja tietoja. Tällöin työttömien taloudellinen tilanne voi aineiston perusteella muodostua liian hyväksi, mikäli määritämme työttömiksi vain Palkkarakennetilaston keräämisen aikaan työttöminä olleet. Toisaalta työttömien taustaominaisuudet ja taloudellinen tilanne voivat antaa vääristävän kuvan kaikista työttömistä, mikäli oletamme työttömiksi vain koko vuoden työttömänä olleet henkilöt.

Työttömien määrittelyyn ei siis ole olemassa itsestään selvää ratkaisua. Mallinnuksen mahdollistamiseksi hyödynnämme FLEED-aineiston työttömyysjaksotietoja. Käytämme erilaisia työttömien määrittelyjä analyysissä, ja tarkastelemme simulointitulosten herkkyyttä näiden valintojen suhteen. Ensimmäisessä vaihtoehdossa määritämme työttömiksi kaikki ne henkilöt, joille ei ole saatavilla palkka- ja työaikatietoja Palkkarakennetilastossa ja joilla on yli 300,

200 tai 100 työttömyyspäivää vuoden aikana. Toisessa vaihtoehdossa määritämme työttömiksi ne henkilöt, jotka ovat lokakuun aikana olleet työttöminä (vastaa karkeasti samaa ajanjaksoa kuin Palkkarakennetilaston tiedot). Luvussa 4 keskustelemme lyhyesti mallin herkkyydestä näiden valintojen suhteen. Keskustelemme myös muista työn tarjonnan mallintamisen mahdollistavista aineistovaihtoehdoista lyhyesti luvussa 6.

## 4. MALLIN SOVITTAMINEN JA TYÖN TARJONNAN JOUSTOT

Tässä luvussa esitellään käyttämämme mallinnusvalinnat ja muut oletukset, verrataan mallin ennustetta havaittuun työtuntien jakaumaan sekä tarkastellaan mallin avulla estimoituja ensimmäisiä ja alustavia työn tarjonnan joustoja. Lisäksi keskustellaan lyhyesti siitä, miten eri mallinnusvalinnat vaikuttavat arvioon mallin sovitteesta ja työn tarjonnan jouston koosta.

### 4.1. Mallinnusvalinnat

Rakenteellista työn tarjonnan mallia estimoitaessa on välttämätöntä tehdä useita eri oletuksia, joihin ei ole olemassa selkeästi määriteltyä oikeaa ratkaisua. Alla keskustelemme muutamista keskeisimmistä valinnoista ja niiden merkityksestä lyhyesti. Yksityiskohtaisemmin tehtyihin oletuksiin ja teknisiin mallinnusvalintoihin voi tutustua mallinnuskoodien avulla (ks. raportin lähdeluettelo).

#### Palkkojen estimointi

Jotta työn tarjonnan malli voidaan estimoida, täytyy tuntipalkat ennustaa ainakin työttömille henkilöille. Toisin sanoen tarvitaan estimaatti sille, mikä työttömän henkilön tuntipalkka olisi, mikäli hän työllistyisi. Tämä tieto tarvitaan, jotta voimme arvioida käytettävissä olevat tulot työttömille kaikissa eri mallin diskreeteissä vaihtoehdoissa. Estimoinnissa voi olla tärkeää ottaa myös valikoitumisharha huomioon eli se, että työttömien tuntipalkat ovat todennäköisesti pienempiä kuin työssä olevien havaituilta ominaisuuksiltaan samankaltaisten henkilöiden. Tätä pyrimme ottamaan huomioon käyttämällä Heckmanin (1979) kehittämää kaksivaiheista valikoitumisen huomioon ottavaa estimaattoria.

Palkkaregressiossa otamme huomioon keskeiset palkkaan yleisesti vaikuttavat tekijät, kuten iän, sukupuolen, koulutuksen, äidinkielen, lasten lukumäärän, siviilisäädyn, koulutustason ja toimialan. Kattava paneeliaineisto mahdollistaa lisäksi sen, että voimme ottaa huomioon myös edellisten vuosien palkat, työ- ja työttömyyspäivät sekä toimialat, joilla henkilö on aiemmin työskennellyt. Työhistorian kontrollointi auttaa myös valikoitumisharhan huomioon ottamisessa, sillä on odotettavaa, että heikomman työhistorian omaavien henkilöiden tuntipalkat ovat pidempään töissä olleita matalampia. Toisin sanoen työhistoriatiedot toimivat proxy-muuttujina havaitsemattomille tekijöille.

Palkkayhtälön estimoinnin jälkeen on päätettävä, käytetäänkö palkkaregressiosta saatuja palkkaennusteita kaikille henkilöille vai pelkästään työttömille, joiden palkat puuttuvat aineistosta. Molemmille valinnoille on omat perusteensa, eikä oletusvapaata yksiselitteistä valintaa ole olemassa. Havaittuja palkkoja käyttämällä saadaan luonnollisesti tarkempi kuva tämänhetkisestä palkkatasosta (annettuna mahdolliset virheet ja puutteet aineistossa). Toi-

saalta tällöin myös oletetaan, että henkilöt tekevät työn tarjonta -päätöksensä omaan nykyiseen tuntipalkkaansa perustuen, vaikka se olisikin muita samankaltaisia ja samassa työssä olevia henkilöitä suurempi tai pienempi. Käyttämällä ennustettuja palkkoja vastaavasti oletetaan, että henkilö päättää työn tarjonnastaan samankaltaisten ja samanlaisessa työssä olevien henkilöiden keskimääräisen palkkatason perusteella, mitä yksittäinen työntekijä ei välttämättä tiedä tai havaitse. Alla esiteltävässä analyysissä olemme käyttäneet kumpaakin vaihtoehtoa, ja keskustelemme näiden eroista tuloksien osalta.

On myös syytä pohtia sitä, miten ennustevirheitä käsitellään mallinnuksessa. Ennustettujen ja havaittujen tuntipalkkojen erot voivat olla hyvin merkittäviä yksittäisten henkilöiden kohdalla, mikä voi johtaa osin epäjohdonmukaisiin tuloksiin. Tämä koskee työllisten lisäksi myös työttömille ennustettuja palkkoja, joskin työttömän todellinen palkkataso ei ole havaittavissa. Ennustevirheet voidaan kuitenkin ottaa huomioon mallinnuksessa, jolloin tästä ongelmasta voidaan ainakin osittain päästä eroon (ks. tarkemmin Löffler ym. 2018). Ennustevirheiden huomioon ottaminen tekee mallin estimoimisesta kuitenkin haastavampaa, mikä näkyy erityisesti merkittävästi pidempänä estimoointiaikana. Tässä raportissa emme siksi tarkastele mallinnusvaihtoehtoa, jossa ennustevirheet olisivat eksplisiittisesti huomioitu.

#### Diskreetin valintajoukon määrittäminen

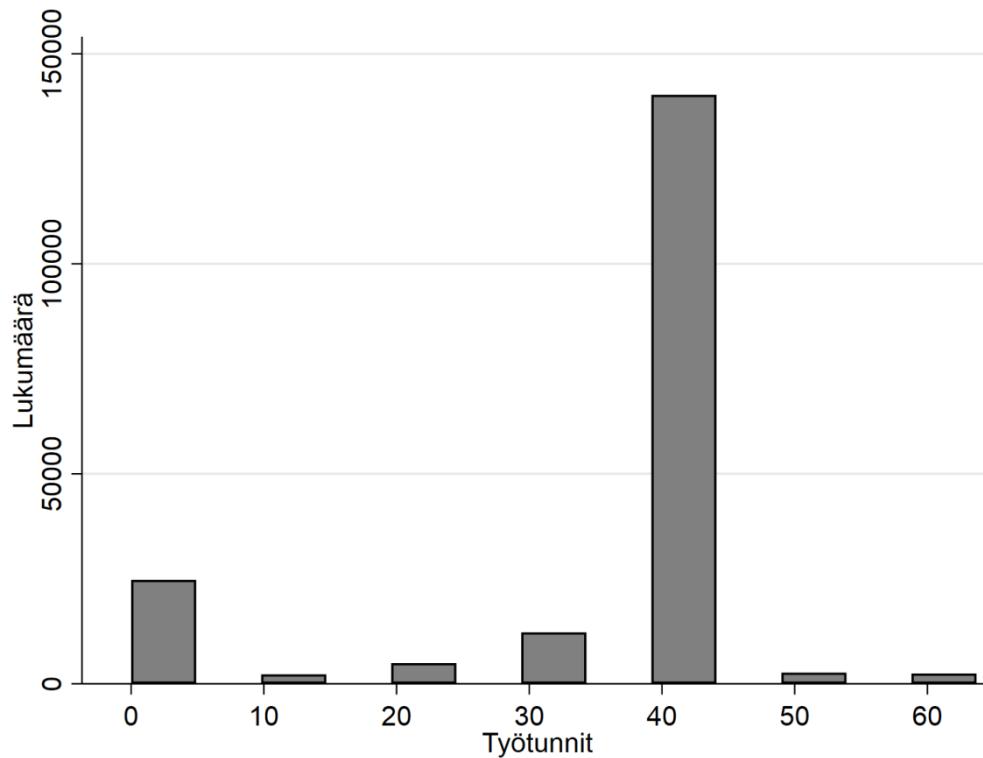
Mallin estimoimista varten täytyy myös määrittää oletettujen työtuntivaihtojen lukumäärä ja niiden väliset erot. Kirjallisuudessa on tyypillisesti käytetty jollain kriteerillä edustavaa diskreettiä valintajoukkoa. Löffler ym. (2018) käyttävät mallissaan seitsemää mahdollista valintaa: työtön, 10, 20, 30, 40, 50 tai 60 työtuntia viikkoa kohti.

Diskreetin valintajoukon määrittämiseen ei ole olemassa mitään yksiselitteistä ratkaisua. Valintajoukon määrittämisen apuna voi kuitenkin käyttää havaittua työtuntien jakaumaa. Kuviossa 1 on esitetty työtuntien jakauma palkkarakenneaineistossa mallinnuksessa käytetylle otokselle, kun mallin työttömiksi on lisätty yli 300 päivää vuodessa työttöminä olleet henkilöt, jotka puuttuvat Palkkarakenneaineiston otoksesta. Kuviossa käytetään yllä mainittua lähtöoletusta seitsemästä mahdollisesta valinnasta (yli 60 työtunnin valinnat on yhdistetty 60 tunnin luokkaan).

Työtuntien jakaumasta on havaittavissa, että tehdyt työtunnit keskittyvät melko selvästi täysiin viikkotyötuntien (36–40 tuntia/viikko) kohdalle. Osa-aikaisen työn (10–30 tuntia) tai ylitöiden (50–60 tuntia) tekeminen on melko harvinaista, muttei kuitenkaan täysin poikkeuksellista. On siis todennäköistä, että osa-aikaisen työnteon mallintaminen rakenteellisen työn tarjonta -mallin avulla on haastavaa, sillä vain harva henkilö tekee Palkkarakenneaineiston perusteella osa-aikaista työtä. Keskustelemme tästä tarkemmin myöhemmin.

Lisäksi on mahdollista, että työtunneissa ja tuntipalkoissa on enemmän tai vähemmän mitausvirheitä, sillä aineiston työtunnit perustuvat työnantajan ilmoitukseen tehdyistä tunneista. Ilmoitetut tunnit eivät välttämättä vastaa todellisuudessa tehtyjä työtunteja, jolloin myös todellinen tuntipalkka poikkeaa nimellisestä tuntipalkasta.

**Kuvio 1. Työtuntien havaittu jakauma (vuosi 2015)**



#### Käytettävissä olevien tulojen estimointi

Rakenteellinen työn tarjonnan malli edellyttää, että määritämme henkilöiden käytettävissä olevat tulot kaikissa eri diskreetin valintajoukon tiloissa. Seitsemän mahdollisen valinnan mallissa tulee siis määrittää käytettävissä olevat tulot, mikäli henkilö olisi työtön (0 työtuntia), tai tekisi 10, 20, 30, 40, 50 tai 60 työtuntia viikkoa kohti. Käytämme kattavaa rekisteripohjaista aineistoa käytettävissä olevien tulojen mallintamiseen.

Ensivaiheen regressiomallissa käytämme SISU-rekisteriaineiston havaittuja keskimääräisiä kuukausipalkka- ja pääomatuloja, perheen muita tuloja sekä muita sosiaalietuuksiin ja verotukseen vaikuttavia tekijöitä selittämään kuukausikohtaisia nettotuloja. Tämän regressiomallin estimoitujen kertoimien avulla arvioimme käytettävissä olevat tulot henkilöille kaikissa mallin diskreeteissä tiloissa käyttämällä ennustettuja ja/tai havaittuja tuntipalkkatietoja (tuntipalkka x tehdyt työtunnit).

Kokonaisuudessaan käytettävät mallit selittävät käytettävissä olevia tuloja varsin hyvin (selitysasteet ovat tyypillisesti yli 0,9). Regressiopohjaista menetelmää käyttäen on kuitenkin hyvin vaikea arvioida suunniteltujen vero- ja sosiaaliturvamuutosten vaikutuksia, koska estimoitu tilastollinen malli kuvaa ainoastaan aineistovuoden vero- ja tulonsiirtojärjestelmää. Käytettävissä olevat tulot eri tiloissa on mahdollista estimoida myös käyttämällä tuntipalkkatietoja ja SISU-mallia. Mikrosimulointimallia käyttämällä on suoraviivaisempaa arvioida todellisten tai erityisesti kuvitteellisten politiikkamuutosten vaikutuksia. Keskustelemme simulaatiopohjaisesta lähestymistavasta tarkemmin luvussa 5.



## Hyötyfunktio ja muut siihen liittyvät valinnat

Mitään yksiselitteistä sääntöä ei hyötyfunktion muodon valintaan ole olemassa. Kirjallisuudessa on tyypillisesti käytetty Box-Cox, Translog tai kvadraattista hyötyfunktiota (Löffler et al. 2018). Estimoimme mallin kaikilla näillä spesifikaatioilla ja keskustelemme tämän valinnan vaikutuksesta tuloksiin.

Kuten edellä todettiin, mallissa on lisäksi mahdollista ottaa huomioon muita tekijöitä, jotka vaikuttavat työtuntien valintaan. Tällaisia tekijöitä ovat mm. työttömyyteen tai sosiaalietuuk-sien varassa elämiseen liittyvä häpeäntunne (stigma) sekä osa-aikaisten työmahdollisuuksi-en vähyys johtuen kysyntäpuolen rajoitteista. Lisäksi käytämme spesifikaatiota, jossa huomioidaan preferenssien (eli kulutuksen ja vapaa-ajan arvostuksen) heterogeenisuutta iän, sukupuolen, lasten lukumäärän ja koulutustason suhteen. Otamme näitä tekijöitä tietyissä mallispesifikaatioissa huomioon, ja tarkastelemme niiden vaikutusta tuloksiin.

## **4.2. Mallin estimointi ja työn tarjonnan jouston arviointi**

Työn tarjonnan rakenteellinen mallintaminen on laskentateknisesti haastavaa, ja sen seurauksena mallien estimointi voi saatavilla olevan laskentakapasiteetin rajoissa kestää kauan. Tässä raportissa rajoitumme siksi käyttämään viiden prosentin otosta koko aineistosta, sekä rajoitamme logit-mallin iteraatiokierrokset maksimissaan 50 iterointiin. Lisäksi rajoitumme tässä vaiheessa mallintamaan yksilötason valintoja. Malli on kuitenkin tarvittaessa laajennettavissa kotitaloustason tarkasteluun.

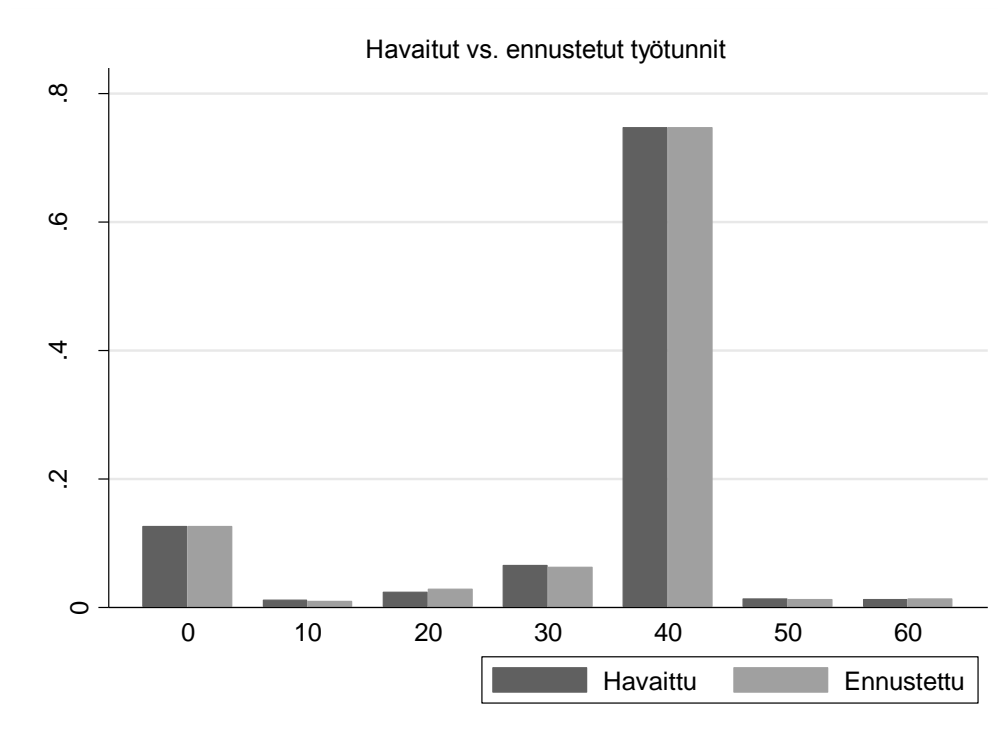
Estimoimme työn tarjonnan mallin käyttämällä Löfflerin ym. (2018) kehittämää Islogit-Stata-ohjelmaa. Tarkastelemme mallin herkkyyttä eri mallinnus- ja aineistovalinnoille estimoinnalla suuren määrän erilaisia mallispesifikaatioita. Näiden perusteella pyrimme katsomaan, mitkä spesifikaatiot tuottavat havaittuun aineistoon verrattuna realistisimmat kuvan työn tarjonnan-ta.

Mallin empiiristä sopivuutta voidaan suuntaa-antavasti arvioida esimerkiksi vertaamalla mal-lin tuottamaa työtuntien todennäköisyysjakaumaa havaittuun työtuntien jakaumaan. Yleisesti ottaen voidaan todeta, että mallin sovitteet vaihtelevat, kun käytämme eri aineisto- ja mallin-nusoletuksia. Malli vaikuttaa sopivan aineistoon paremmin erityisesti silloin, kun otamme huomioon eri valintavaihtoehtoihin liittyvät kiinteät vaikutukset sekä työn ja vapaa-ajan valin-taan liittyvän heterogeenisuuden erityyppisillä henkilöillä. Lisäksi mallin sovitteet ovat pa-rempia silloin, kun oletamme yli viisi diskreettiä valintamahdollisuutta. Sen sijaan käytetyllä hyötyfunktiolla tai palkkojen ennustamiseen liittyvillä valinnoilla ei ole yhtä merkittävää vai-kutusta mallin sovitteeseen, kun verrataan mallin ennustamia työtunteja havaittuihin työtun-teihin.

Kuviossa 2 on esitetty tyypillinen mallin mukainen estimoitu työtuntien todennäköisyysja-kauma sekä todellinen työtuntien jakauma käyttämällä kvadraattista hyötyfunktiota sekä edellä mainittuja mallin sovitetta yleisesti parantavia oletuksia. Kuvioista havaitaan, että pys-tymme mallintamaan työttömyyden sekä kokopäiväisen työnteon erittäin tarkasti. Sen sijaan osa-aika- ja ylityön kohdalla mallinnus ei ole aivan yhtä tarkkaa. Tämä johtuu todennäköi-sesti siitä, että muuta kuin kokopäiväistä työtä tekevien osuus on pieni, jolloin tämäntyyppis-ten valintojen mallintaminen on haastavaa melko yksinkertaisessa työn tarjonta -kehikossa. Toisaalta koska osa-aikatyön tekeminen on Suomessa melko harvinaista, tämä ei välttämät-tä ratkaisevasti vaikuta mallin perusteella tehtyihin laskelmiin, sillä pääosa siirtymistä työ-markkinoilla tapahtuu kokoaikaisen työn ja työttömyyden välillä. Joustava seitsemän dis-

kreetin vaihtoehdon malli kuitenkin mahdollistaa sen, että budjettirajoitteen muuttuessa myös osa-aika- tai ylityöhön siirtyminen on mallissa mahdollista.

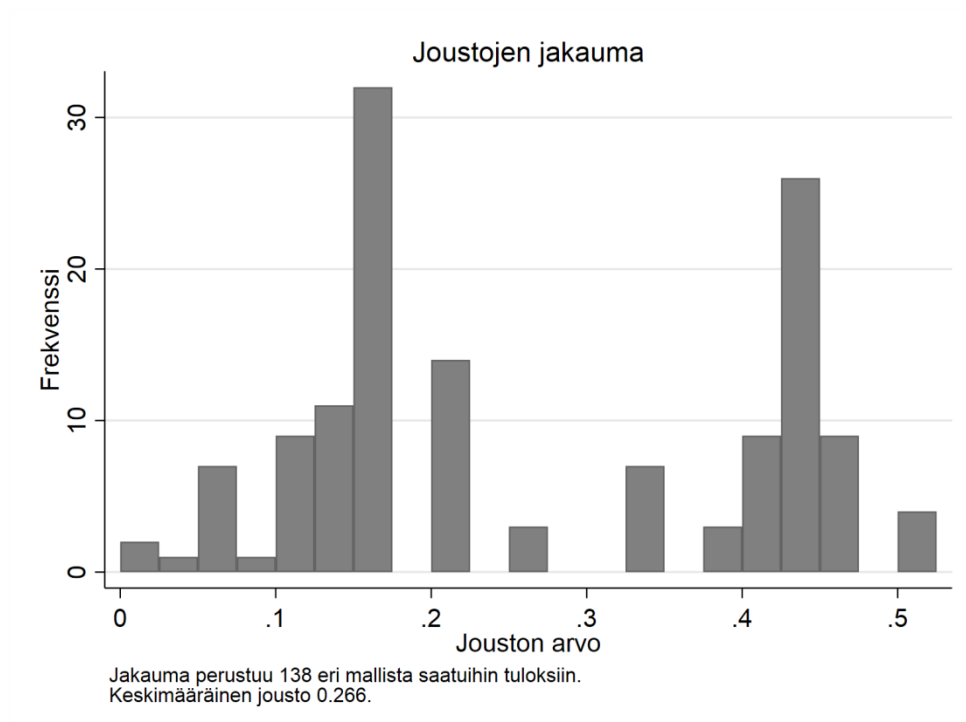
## Kuvio 2. Ennustetut ja havaitut työtuntijakaumat



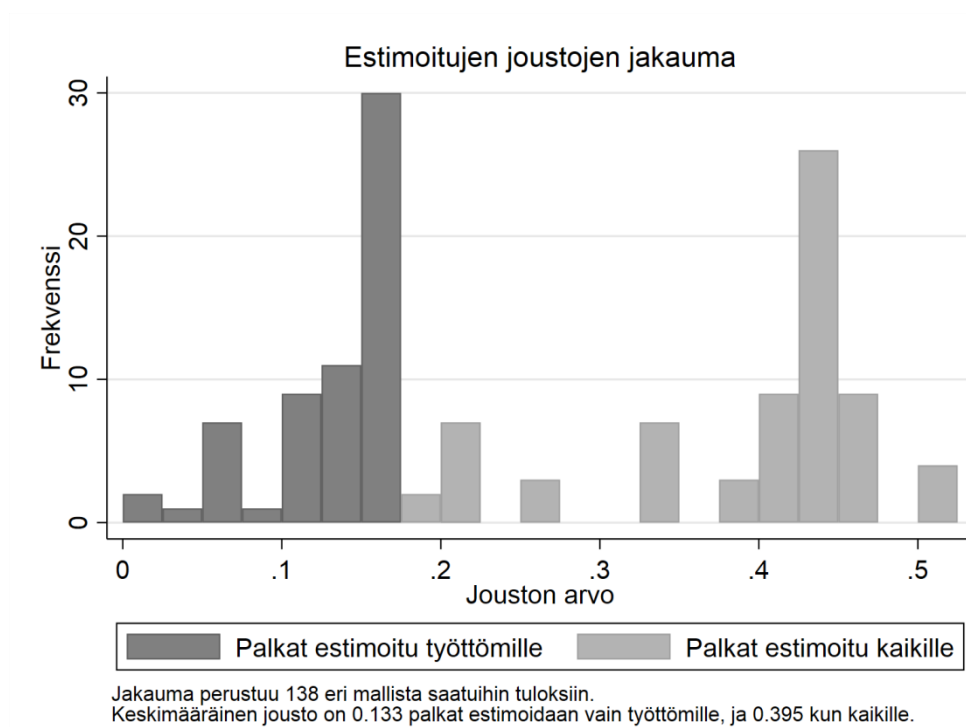
Rakenteellisen mallin avulla voimme arvioida työn tarjonnan jouston suuruutta suhteessa palkoissa tapahtuviin muutoksiin. Kuvioissa 3 on esitetty mallin tuottamia työtuntien aggregaattijoustoestimaatteja tuntipalkan suhteen eri mallinusoletuksilla. Kuviossa rajoitutaan malleihin, joissa käytetään seitsemää diskreettiä valintaa. Sen sijaan funktiomuodon, palkkojen ennustamisen, työttömien määritelmän sekä kiinteiden vaikutusten ja heterogeenisten preferenssien oletuksia vaihdellaan eri mallien välillä. Jousto on käytännössä estimoitu niin, että korotamme kaikkien henkilöiden tuntipalkkoja 10 prosenttia, ja tutkimme, miten tämä muutos vaikuttaa työn tarjontaan. Jousto saadaan täten laskettua suhteuttamalla työtunneissa tapahtuva muutos 10 prosentin muutokseen tuntipalkassa.

Kuvioista nähdään, että eri mallinvalintojen tuottamat joustoestimaatit vaihtelevat välillä 0–0,5. Eri mallinuvaihtoehdoista laskettu keskimääräinen työn tarjonnan jousto oman bruttotuntipalkan suhteen on 0,27, mikä vastaa melko hyvin aiemman tutkimuskirjallisuuden arvioita työn tarjonnan joustosta muissa maissa (ks. Löffler et al. 2018). Keskimääräinen jousto on myös mittaluokaltaan sama, jota Kärkkäinen ja Tervola (2018) käyttivät hankkeen ensimmäisen osan raportissaan lähtöoletuksena työn tarjonnan joustosta Suomessa. Haluamme kuitenkin korostaa, että arviot joustoista ovat vielä hyvin alustavia, eikä niitä tulisi tulkita kuvaavan keskimääräistä työn tarjonnan joustoa Suomessa. Tarkempi ja luotettavampi arviointi edellyttää estimoinnissa käytettävän aineiston laajentamista sekä mallin oletusten ja niiden vaikutusten perusteellisempaa tutkimista, johon ei tämän raportin puitteissa ollut vielä mahdollista ryhtyä. Alla keskustelemme muutamista keskeisistä joustoon vaikuttavista oletuksista.

**Kuvio 3. Estimoitujen joustojen jakauma eri mallinnusoletuksilla**



**Kuvio 4. Joustoestimaatit eri oletuksilla ennustetuista palkoista**



Estimoidut joustot ovat keskimäärin samansuuruisia eri hyötyfunktio-oletuksilla sekä silloin, kun mallissa otetaan kiinteät kustannukset ja heterogeisuus huomioon. Joustoestimaatit kuitenkin vaihtelevat merkittävästi, kun muutetaan oletusta palkkojen estimoinnista. Kuviossa 4 on esitetty joustoestimaatit silloin, kun palkat ennustetaan kaikille, sekä silloin, kun tuntipalkat estimoidaan vain mallin työttömille. Kuvioista nähdään, että joustot ovat selvästi pienemmät, kun palkat estimoidaan vain työttömille (keskiarvo 0,14), verrattuna siihen, jos oletetaan kaikkien palkkojen perustuvan palkkaregression soviteeseen (keskiarvo 0,40). Erot estimaateissa johtuvat todennäköisesti siitä, että soviteeseen perustuvien ennustettujen palkkojen hajonta on pienempi kuin todellisten palkkojen. Tällöin tuntipalkkojen 10 prosentin korottamisen myötä tulevat käyttäytymisvaikutukset tuottavat mekaanisesti suuremman jouston kuin silloin, kun käytetään havaittuja palkkoja työllisten osalta.

Palkkojen estimointiin liittyvään valintaan ei kuitenkaan ole olemassa selkeästi määritettävää oikeaa lähestymistapaa, ja valintaa on usein pohdittava tapauskohtaisesti. On kuitenkin hyvä huomioida, että tämä valinta todennäköisesti vaikuttaa joustoestimaattiin merkittävästi. On lisäksi oletettavaa, että joustoestimaatit muuttuisivat, jos palkkaestimoinnin ennustevirheet otettaisiin mallissa eksplisiittisesti huomioon. Tämä vaikuttaa todennäköisesti erityisesti silloin, kun palkkaestimaatteja käytetään kaikille henkilöille. Ennustevirheiden huomioimisen merkitystä ei kuitenkaan tarkastella tässä raportissa.

### 4.3. Heterogeisuus tuloksia

Tarkastelemme lyhyesti tulosten heterogeisuutta sukupuolen ja iän suhteen. Kuviossa 5 on esitetty toteutunut työtuntijakauma sekä mallilla estimoitu työtuntijakauma miehille (yllä) ja naisille (alla).<sup>2</sup> Kuvioista nähdään, että jakaumat saadaan molemmille ryhmille estimoitua melko hyvin, mutta sekä ennustetuissa että havaituissa jakaumissa on kuitenkin nähtävissä se, että osa-aikainen työnteko on naisilla odotetusti hieman yleisempää kuin miehillä.

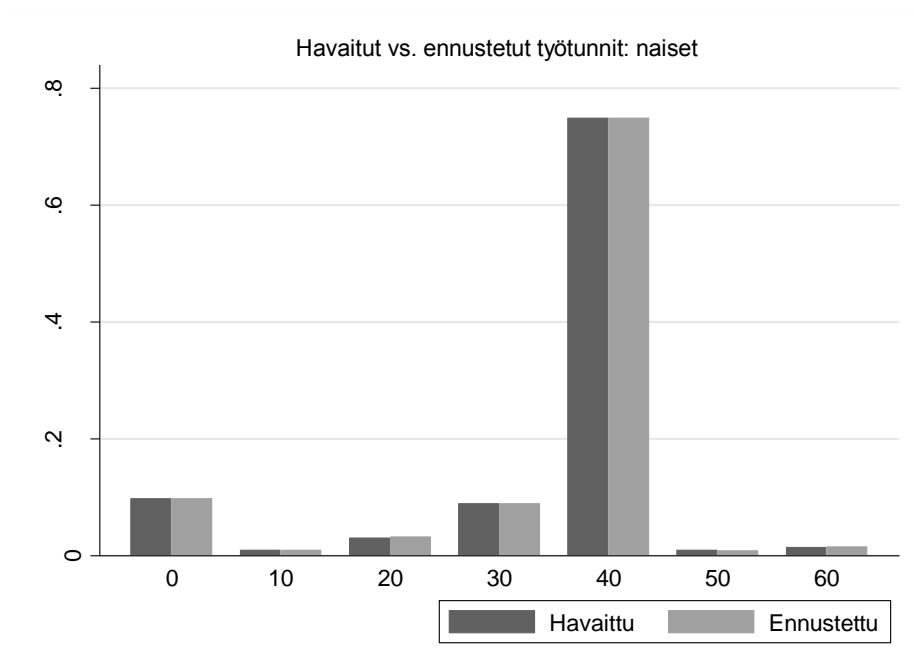
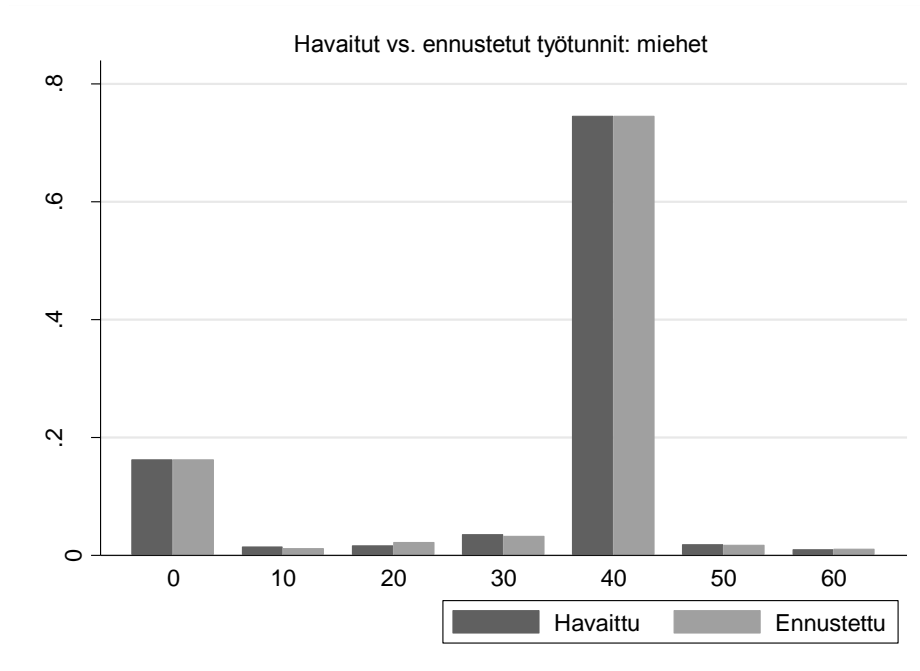
Kuviossa 6 vertaamme mallin tuottamaa työtuntijakaumaa toteutuneeseen työtuntijakaumaan yli 44-vuotiaille ja sitä nuoremmille. Kuvioista nähdään, että pystymme mallintamaan työtuntien todennäköisyysjakauman melko tarkasti myös eri ikäryhmille.

Mallin sovitteiden ohella myös työn tarjonta -joustot palkan suhteen voivat vaihdella erityyppisillä henkilöillä. Tiivistetysti voimme todeta, että työn tarjonnan joustot ovat keskimäärin suurempia naisilla (0,207) kuin miehillä (0,012), ja että yli 44-vuotiaiden henkilöiden joustot (0,113) ovat suurempia kuin tätä nuorempien (0,032). Laskenta-ajan pienentämiseksi olemme näille ryhmille estimoineet joustot käyttämällä suppeampaa joukkoa eri malleja. Tällöin koko populaation keskimääräiset joustot poikkeavat siitä, mitä edellä esitettiin. Ryhmäkohtaiset keskimääräiset joustot perustuvat malleihin, joissa on seitsemän diskreettiä valintaa, työttömiksi oletetaan yli 200 päivää vuoden aikana työttöminä olleet ja palkat ennustetaan vain työttömille. Hyötyfunktion muoto sekä oletukset kiinteiden kustannusten huomioimisesta vaihtelevat eri malleissa.

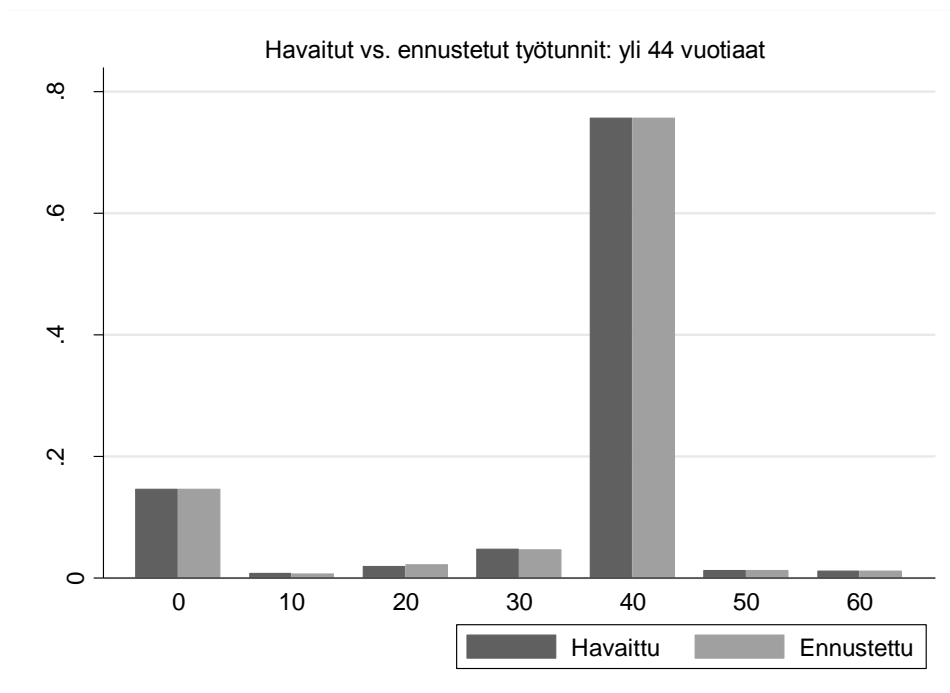
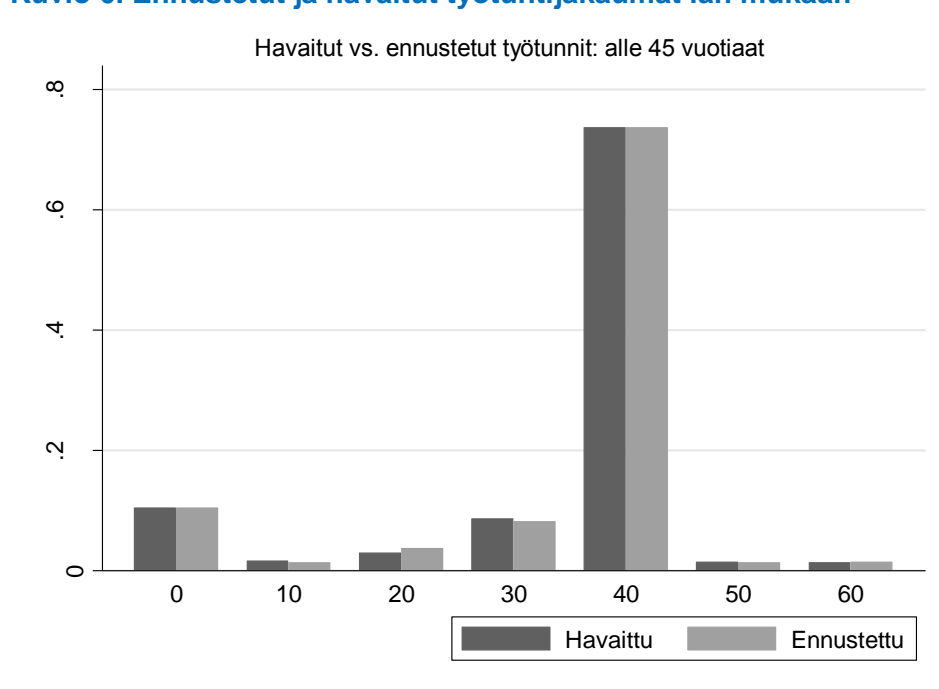
---

<sup>2</sup> Kuviossa käytetään samaa mallispesifikaatiota kuin kuviossa 2 edellä.

**Kuvio 5. Ennustetut ja havaitut työtuntijakaumat sukupuolen mukaan**



**Kuvio 6. Ennustetut ja havaitut työtuntijakaumat iän mukaan**



## 5. TYÖN TARJONTA JA SISU-MALLI

Käytettävissä olevat tulot eri työtuntivaihtoehdoissa voidaan estimoida regressiomallia hyödyntäen, kuten teimme edellisessä luvussa. Käytettävissä olevien tulojen arvioinnissa voidaan kuitenkin käyttää vaihtoehtoisesti apuna mikrosimulointimallia. Mikrosimulointimallin hyödyntämisen etuna on se, että mikrosimuloinnin avulla voidaan suoraan arvioida eri lainsäädäntömuutosten vaikutuksia käytettävissä oleviin tuloihin. Koska tämän raportin laskelmissa aineisto on rakennettu SISU-mallin rekisteriaineistoa hyödyntäen, voidaan SISU-mallia hyödyntää käytettävissä olevien tulojen simuloinnissa.

Käytettävissä olevien tulojen simulointi voidaan toteuttaa käytännössä siten, että tuntipalkkojen estimoinnin jälkeen viemme SISU-malliin tiedon henkilöiden bruttopalkasta eri työtuntivaihtoehdoissa. Tämän jälkeen korvaamme SISU-mallin tausta-aineistossa aineiston palkkatiedon estimoidulla bruttopalkalla ja nollaamme muut tulot (mm. muut palkkatulot, yrittäjätulot, eläketulot ja työttömyysturvaetuuudet). Vaihtoehdossa, jossa henkilö on työtön, palkkatulot nollataan ja henkilö määritellään koko vuodeksi työttömäksi. Työttömänä henkilölle simuloidaan joko ansiosidonnaista päivärahaa, peruspäivärahaa tai työmarkkinatukea riippuen siitä, mihin etuuteen hän on oikeutettu. Osa-aikatyövaihtoehdoissa, joissa viikkotyötunteja on 10, 20 tai 30, oletamme käytettävissä olevien tulojen simuloinnissa henkilöiden olevan oikeutettuja soviteltuun päivärahaan.<sup>3</sup> Käytettävissä olevien tulojen simulointi eri vaihtoehdoissa vastaa menetelmällisesti ja oletuksiltaan suurelta osin työllistymisveroasteiden laskennassa SISU-malliin kehitettyjä menetelmiä (ks. Kärkkäinen ja Tervola 2018).

Koska mm. asumistuen ja toimeentulotuen määrä riippuu kotitalouden kaikkien jäsenten tuloista, teimme käytettävissä olevien tulojen simuloinnin kahden tulonsaajan kotitalouksissa erikseen kummallekin puolisolalle. Käytännössä simuloimme käytettävissä olevat tulot eri vaihtoehdoissa kummallekin puolisolalle erikseen oletuksella, että toisen puolison status ei muutu.

Koska diskreetti työn tarjonnan malli estimoidaan Stata-ympäristössä ja SISU-malli toimii SAS-ympäristössä, tarvitaan mallinnuksessa molempia ohjelmistoja. Ensin Stata-ympäristössä estimoidaan tuntipalkat. Estimoitujen tuntipalkkojen avulla lasketut bruttopalkat viedään tämän jälkeen SISU-malliin ja simuloidaan käytettävissä olevat tulot eri työtuntivaihtoehdoissa. Lopulta simuloidut käytettävissä olevat tulot viedään takaisin Stata-ympäristöön ja näitä tietoja hyödyntäen estimoidaan työn tarjonnan malli.

Työn tarjonnan mallin estimoidut työllisyysvaikutukset voidaan viedä takaisin SISU-mallin tausta-aineistoon ja muokata siinä henkilöiden työllisyysstatusta estimointitulosten mukaisesti. Tätä työllisyysvaikutuksilla muokattua aineistoa hyödyntämällä voidaan simuloida lainsäädäntömuutosten tulonjakovaikutuksia siten, että laskelmissa huomioidaan myös reformien vaikutukset työn tarjontaan.

Mikrosimuloinnin hyödyntäminen on laskennallisesti haastavampaa kuin käytettävissä olevien tulojen estimointi regressiomallin avulla. Esimerkiksi seitsemällä eri diskreetillä vaihtoehdolla käytettävissä olevien tulojen laskenta vaatii 14 simulointiajoa, kun simuloinnit ajetaan puolisoille erikseen. Kun käytettävissä olevat tulot lasketaan vielä uudestaan joustojen arvioimiseksi, tarvitaan simulointiajoja yhteensä 28.

<sup>3</sup> Tämä oletus ei ole ihan sopusoinnussa taustalla olevan teorian kanssa, jonka mukaan osa-aikainen työskentely on vapaaehtoinen valinta, sillä soviteltu päiväraha on tarkoitettu vain niille osa-aikaisesti (tai pätkätyösuhteissa) työskenteleville henkilöille, jotka haluaisivat työskennellä kokoaikaisesti.

Rakensimme ja testasimme SISU-mallin ja työn tarjonnan mallin välisen yhteyden. Yhteys toimii ja työn tarjonnan malli on mahdollista estimoida SISU-simulointia hyödyntäen. Käytävissä olevien tulojen simuloinnissa tulee kuitenkin tehdä vielä enemmän testaus- ja kehitystyötä ennen kuin se voidaan sellaisenaan sisällyttää rakennemallin estimointiin. Tällä hetkellä mikrosimulointimallia hyödyntävä työn tarjonnan malli ei vielä tuota yhtä luotettavia tuloksia kuin regressiomallia hyödyntävä versio. Lisäkehitystyötä tarvitaan mm. sen määrittämiseen, milloin ja millä ehdoilla henkilö saa soviteltua päivärahaa, ja miten henkilöiden muita kuin palkkatuloja käsitellään simuloinnissa.

## 6. MITEN MALLIA VOIDAAN TULEVAISUUDESSA KEHITTÄÄ?

Tässä raportissa esittelimme diskreetin valinnan kehikkoon perustuvan työn tarjonnan mallintamisen peruseräiteitä ja sovelsimme mallia Suomen aineistoon. Alustavien tulosten perusteella voidaan todeta, että malli kuvaa melko hyvin työn tarjontaa Suomen työmarkkinoilla ja että tulokset ovat melko hyvin linjassa kansainvälisen tutkimuskirjallisuuden kanssa. Toisaalta tulokset ovat sensitiivisiä eräiden keskeisten mallinnusvalintojen suhteen. Kiinteiden vaikutusten lisääminen malliin parantaa selvästi mallin ennustaman työtuntijakauman osuvuutta. Mallista johdetut työn tarjonnan joustot riippuvat palkkojen käsittelystä. Joustot ovat selvästi suurempia silloin, kun ennustettuja palkkoja käytetään kaikille henkilöille eikä pelkästään aineiston työttömille.

Mallintamista varten jouduttiin yhdistämään tietoja useista eri rekistereistä. Vaikka yhdistetty aineisto on laaja ja monipuolinen, se ei tarjoa edustavaa otosta työvoimasta, mikä vaikeuttaa tulosten tulkintaa. On myös syytä korostaa, että aikataulullisista syistä ja laskentakapasiteetin rajoitteista johtuen rakennemallin monipuolisimmat versiot jouduttiin rajaamaan tämän raportin tarkasteluiden ulkopuolelle. Näistä syistä johtuen raportissa esitetyt tulokset ovat vasta alustavia ja suuntaa-antavia. Luotettavien työn tarjonta -vaikutusten estimointi edellyttää vielä jatkotyötä.

Mallin jatkokehittämisen ehkä keskeisimpänä kohteena on saada SISU-malli toimimaan osana työn tarjonnan mallintamista, mikä merkittävästi laajentaisi mallin käytettävyyttä niin politiikkavalmistelussa kuin akateemisessa tutkimuksessa. Vasta tämä mahdollistaisi politiikkamuutosten työllisyys- ja tulonjakovaikutusten *ex ante* -arvioinnit.

Lopuksi esitämme lyhyesti muutamia mallintamiseen ja siinä käytettäviin aineistoihin liittyviä kehittämissuhteita sekä suuntaviivoja työn tarjonnan mallintamisen kehittämiseksi tästä eteenpäin.

**Työvoimatutkimuksen hyödyntäminen.** Työvoimatutkimus on otokseen perustuva haastattelututkimus, jossa otoshenkilöiltä tiedustellaan puhelimitse heidän toiminnastaan haastatteluviikon aikana. Sen lisäksi, että tutkimuksessa kysytään tietoja tehdyistä työtunneista, siinä myös kysytään, montako työntuntia haastateltava haluaisi työskennellä ja onko henkilö etsinyt toista työtä. Vertaamalla toteutuneita ja toivottujen työtuntien määriä voitaisiin arvioida, missä määrin työvoiman kysyntä rajoittaa työtuntivalintoja. Työttömiltä puolestaan tiedustellaan heidän työhausta. Näiden vastaustensa avulla voitaisiin arvioida vapaaehtoisuuden työttömyyden yleisyyttä. Työvoimatutkimus kattaa noin 12 000 henkilöä kuukausittain ja se koostuu viidestä rotaatioryhmästä. Otoshenkilöt osallistuvat tutkimukseen viisi kertaa, joten aineisto mahdollistaisi myös pitkittäistarkastelut. Työvoimatutkimuksen käsitteet ja



määritelmät noudattavat kansainvälisiä suosituksia, mikä helpottaisi tutkimustulosten vertailtavuutta muiden maiden tuloksiin. Toisaalta aineisto ei ilmeisesti sisällä tietoja palkasta, jotka pitäisi edelleen poimia palkkarakennetilastosta, joten lopullinen otoskoko yhdistelyn jälkeen jäisi suhteellisen pieneksi.

**Palkkarakennetilastosta puuttuvien työllisten huomioiminen.** Yksi ongelma tämän raportin tarkasteluissa oli se, ettei käytetty otos työllisistä ollut edustava: kaikki julkisen sektorin työntekijät olivat mukana, mutta merkittävä osa yksityisen sektorin työntekijöistä jäi pois, koska heidän työnantajansa eivät kuuluneet Palkkarakennetilaston tietojenkeraukseen. Ongelmaan on kaksi mahdollista ratkaisua. Koska FLEED-aineiston avulla voidaan tunnistaa henkilöt, jotka olivat töissä vuoden viimeisen neljänneksen aikana mutta eivät sisällyneet Palkkarakennetilastoon, olisi mahdollista laskea eri työllisten ryhmille otospainot. Palkkarakennetilaston aliedustettujen työntekijäryhmien edustajat saisivat suuremman painon mallin estimoinnissa ja muissa tarkasteluissa. Painottamisen avulla ei kuitenkaan voida huomioida työllisten ryhmiä, joita Palkkarakennetilasto ei ollenkaan kata, eli alle viiden hengen yrityksissä työskenteleviä. Toinen tapa olisi poimia puuttuvat työlliset FLEED-aineistosta ja lisätä ne estimoinnissa käytettävään otokseen. Näiden henkilöiden puuttuville palkoille voidaan laskea sovite aivan kuten työttömienkin puuttuville palkoille. Mutta heidän kontribuutionsa uskottavuusfunktioon poikkeaa muista työllisistä, koska heidän työtuntejaan ei havaita. Periaatteessa tämän joukon huomioiminen mallinnuksessa olisi suoraviivaista, sillä todennäköisyys olla työllinen ilman tietoa työtunneista on yksi miinus todennäköisyys olla työttömänä. Vaikka teoriatasolla muutos olisikin pieni, edellyttäisi se kuitenkin estimointikoodin osittaista uudelleen kirjoittamista, koska Löfflerin ym. (2018) mallikehikko ei tällaista otosongelmaa tunnista.

**Paneeliaineiston käyttö.** Tarkasteluissa hyödynnettiin vain vuoden 2015 tietoja, jotka olivat tuoreimmat saatavilla olevat tiedot hankkeen alkaessa. Tarvittavat rekisteriaineistot ovat kuitenkin saatavilla myös joiltain aiemmilta vuosilta ja tulevat olemaan saatavilla myös myöhemmille vuosille. Luonteva laajennus olisikin yhdistää tietoja useammalta vuodelta. Koska henkilöt havaittaisiin useampana ajankohtana, pitkittäisaineisto mahdollistaisi diskreetin valinnan mallin paneeliversioiden estimoinnin. Aikadimensio edesauttaa havaitsemattomien tekijöiden mallintamista ja siten sallisi useampien satunnaiskomponenttien sisällyttämisen malliin. Tietyissä tapauksissa tällaisten paneelimallien estimointi voisi nojautua tilastollisten ohjelmistojen valmiiden paneeli mixed logit –mallien estimointiproseduureille.<sup>4</sup> Poikkileikkausaineiston tapauksessa iso osa mallin identifioivasta vaihtelusta tulee vero- ja sosiaaliturvajärjestelmästä, joka kohtelee erilaisia kotitalouksia eri tavalla. Toisaalta erityyppiset kotitaloudet (esim. sen mukaan onko perheessä pieniä lapsia vai ei) saattavat myös arvostaa vapaa-aikaa eri tavoin, eli hyötyfunktion parametrit voivat vaihdella samojen tekijöiden mukaan, jotka tuovat vaihtelua käteen jääviin tuloihin. Yhdistämällä aineistoja eri vuosilta, joiden aikana vero- tai tulonsiirtojärjestelmä on muuttunut, saataisiin käytettävissä oleviin tuloihin reformeista johtuvaa, yksilöiden preferensseistä riippumatonta vaihtelua, mikä parantaisi mallin identifikaation uskottavuutta. Tällaisen variaation hyödyntäminen ei välttämättä edellyttäisi edes diskreetin valinnan mallien paneeliversioiden estimointia, vaan pitkittäisaineistoa voidaan käsitellä poolattuna poikkileikkausaineistona ja jättää huomiotta se, että osa yksilöistä esiintyy aineistossa useammin kuin kerran.

**Tulorekisteriaineiston hyödyntäminen.** Tällä hetkellä kattavat tiedot tuloista ja tulonsiirroista on saatavilla ainoastaan vuositasolla. Tulevan tulorekisterin avulla pitäisi olla mahdollista muodostaa melko kattava kuukausikohtainen paneeliaineisto työtuloista ja maksetuista etuuksista. Tulorekisterin hyödynnettävyyttä pystytään kuitenkin arvioimaan tarkemmin vas-

<sup>4</sup> Tämä edellyttäisi hyötyfunktioita, joka on lineaarinen parametriensa suhteen, sekä palkkojen imputointia ilman ennustevirheen huomioimista.

ta kun aineistoa on olemassa sekä ansiotulojen että etuustulojen osalta riittävän pitkältä ajanjaksolta.

**Mallin SISU-linkitys ja SISU-rekisteriaineiston täydentäminen.** Työn tarjonta -mallin SISU-linkitystä ja sen käytettävyyttä osana SISU-mallin kokonaisuutta edistäisi merkittävästi se, että työn tarjonnan mallinnuksessa tarvittavat keskeiset rekisterimuuttujat olisivat osa nykyistä SISU-rekisteriaineistoa. Käytännössä tämä tarkoittaisi pääasiassa Palkkarakennetilaston työtunti- ja tuntipalkkatietoja sekä jaksotietoja tämänhetkisistä ja menneistä työ- ja työttömyysjaksoista. Viimeksi mainituista on olemassa jo jonkin verran tietoa vuoden 2016 SISU-aineistossa, eli merkittävin muutos tämän suhteen olisi Palkkarakennetilaston tietojen tuominen osittain osaksi SISU-aineistoa.

**Mallin käyttö politiikkavalmistelussa.** Työn tarjonta -mallien ja mikrosimulointimallien hyödyntäminen talouspolitiikan työllisyysvaikutusten arvioinnissa vaatii tasapainottelua tieteellisen uskottavuuden ja käytettävyyden välillä. Rakenteellisen työn tarjonnan mallien estimointi on laskentateknisesti raskasta ja etenkin mikrosimulointia hyödynnettäessä mallin estimointi vie paljon aikaa. Esimerkiksi jo pelkkä SISU-mallin simulointi koko aineistolla kestää 15–20 minuuttia. Kun simulointikertoja tarvitaan yhteensä 28, vie jo pelkkä käytettävissä olevien tulojen simulointi 7–9 tuntia laskenta-aikaa. Laskennallisesti näin raskas malli ei välttämättä sovellu kovin hyvin talouspolitiikan valmistelutyöhön, jossa yksittäisistä reformeista pitää arvioida jopa useita kymmeniä vaihtoehtoja. Toisaalta mallia voidaan soveltaa pienempiin otoksiin koko aineistosta, jolloin myös simulointiaikaa saadaan vähennettyä. Yksi vaihtoehto on käyttää rakenteellista mallia vain työn tarjonnan joustojen tuottamiseen ja tehdä politiikkasimulaatiot hankkeen ensimmäisessä osassa kehitellyn menetelmän avulla hyödyntäen rakennemallin tuottamia joustoestimaatteja.

## 7. LÄHTEITÄ JA TAUSTA-AINEISTOJA

- Bargain, O., Dolls, M., Neumann, D., Peichl, A. & Siegloch, S. (2014). "Comparing Inequality Aversion across Countries when Labor Supply Responses Differ", *International Tax and Public Finance* 21, 845-873.
- Bargain, O., Orsini, K., & Peichl, A. (2014). "Comparing Labor Supply Elasticities in Europe and the United States", *The Journal of Human Resources* 49, 723-838.
- Blundell, R. & Shephard, A. (2012). "Employment, Hours of Work and the Optimal Taxation of Low-Income Families", *Review of Economic Studies* 79, 481-510.
- Dagsvik, J. K, Jia, Z., Kornstad, T. & Thoresen, T. O. (2014). "Theoretical and Practical Arguments for Modeling Labor Supply as a Choice among Latent Jobs", *Journal of Economic Survey* 28, 134-151.
- Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161.
- Kärkkäinen, O. & Tervola, J. (2018). Talouspolitiikan vaikutukset tuloeroihin ja työllisyyteen 2015–2018, Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja, 59/2018.
- Löffler, M., Peichl, A. & Siegloch, S. (2018). The Sensitivity of Structural Labor Supply Estimations to Modeling Assumptions, ifo Working Papers 259/2018.
- Van Soest, A. (1995). "Structural Models of Family Labor Supply – A Discrete Choice Approach", *The Journal of Human Resources* 30, 63-88.
- Van Soest, A., Das, M. & Gong, X. (2002). "A Structural Labor Supply Model with Flexible Preferences", *Journal of Econometrics* 107, 345-374.
- Thoresen, T. O. & Vattø, T. E. (2015). "Validation of the Discrete Choice Labor Supply Model by Methods of the New Tax Responsiveness Literature", *Labour Economics* 37, 38-53.

MALLIKOODIT: <https://vatt.fi/tyontarjonta>

VALTIONEUVOSTON  
SELVITYS- JA TUTKIMUSTOIMINTA

[tietokayttoon.fi](http://tietokayttoon.fi)

ISSN 2342-6799 (pdf)  
ISBN 978-952-287-623-2 (pdf)

