Stockholm School of Economics
Department of Economics
5350 Master's thesis in economics
Academic year 2019-2020

Occupations under threat:

Estimating the impact of automation on the Swedish labour market

Ludvig Holmér (23503)

Abstract

Building upon research which estimates the risk of automation for different occupational categories labour market outcomes of workers switching occupation due to automation is estimated using a data driven occupational mobility model. Models of this kind are constructed using network theory, agent based modelling and, a large data set, and have proven to be useful in analysing flows in the labour market to understand aggregate phenomena such as the Beveridge curve. In order to determine worker behaviour in the model, an occupational mobility network is constructed - which contains the probabilities of transition between ordered pairs of occupations - using all transitions that occurred in Sweden between 2016 and 2017. In addition, the model is calibrated using the Swedish Beveridge curve. A shock, based on estimates of occupation specific automation probability is implemented. In this way, the findings of del Rio-Chanona, Mealy, Beguerisse-Díaz, Lafond, and Farmer (2019) are replicated using Swedish data instead of US data. Which means that it is shown that the structure of the occupational mobility network may affect unemployment levels in certain occupations more than the shock itself. Meaning that some occupations with a high probability of automation are unaffected, while others, with low probability of automation are adversely affected.

JEL codes: C63, J63, J64
Keywords: Labour market, Matching, Automation, Computational economics, Agent based Economics, Network Theory

Supervisor: Andreea Enache
Date submitted: 18/5 - 20
Date examined: 25/5 - 20
Discussant: David Hakula
Examiner: Magnus Johannesson

Contents

List of Figures

1 Introduction

The impact of automation on the labour market has been a topic of discussion within economics almost since the fields inception. Lately, the discussion has garnered widespread interest as the efficiency and applicability of machine learning algorithms and other automation methods have proven increasingly successful. A literature devoted to analysing the prospects of these automation tools within different occupations has received much attention in recent years (Frey & Osborne, 2017) (Brynjolfsson, Mitchell, & Rock, 2018) (Brynjolfsson, Rock, & Syverson, 2017). Because this work is successful in gaining knowledge about which occupations are more, and which are less, suited for automation, it may be used in a *data driven labour flow network model* to assess the network effects of automation on unemployment. A model of this kind was created by del Rio-Chanona et al. (2019), who used it to analyse US data. In addition to providing insight into the network effects of automation on unemployment, the model is able to shed light on some of the mechanics which underlie the relationship between the unemployment rate and the vacancy rate. This empirical relationship is a well known macroeconomic stylised fact called the Beveridge curve.

In this thesis, the results of del Rio-Chanona et al. (2019) will be replicated using Swedish data. In doing so, the Swedish labour market is analysed through the same lens as that of the US labour market, which enables direct comparison of the results. In addition, the behaviour of the model under a completely new set of data is investigated. This provides more evidence regarding the relationship between the model and the Beveridge curve. Furthermore, in constructing the model, an *occupational mobility network* is built. This is a network of occupations, with directed edges from one occupation to another weighted by the *transition probability*, which is the probability that a worker in one occupation finds a job in another. The concept was first introduced by Mealy, del Rio-Chanona, and Farmer (2018) who showed that predictions made by their network correlated significantly with transition rates. The network affects how workers move between occupations. This is important since workers in occupations, affected by decreased demand as a result of automation, will have to look for jobs elsewhere. In this way, the analysis conducted here provides a more complete picture of labour market outcomes due to automation than looking at each occupation in isolation. This makes analysis of this kind even more useful for individuals making choices about their future career as well as policymakers.

The skill required to conduct an automatable task might be useful in conducting other types of non-automatable tasks. The model does not explicitly cover skill but it is endogenous to the network, which occupation you are employed in says a lot about the skills you have. For example, estimates by Frey and Osborne (2017)[1] claim that the tasks of Office assistants and secretaries are more likely than those of Construction and manufacturing supervisors to be computerised. If this is indeed the case and the library assistant find themselves out of a job, the skills they possess might allow them to find work in a different occupation. Perhaps one where the automation threat is less consequential. In contrast, the Construction supervisor is relatively safe from automation. However, workers from different occupations that require similar skills but which are under more direct automation pressure might flee to the relative safety of Construction supervisors. In their working paper del Rio-Chanona et al. (2019) found that their model validates these claims.

[1] translated to the Swedish occupational classification system, SSYK, see section 4

The analysis is conducted in two steps; first the model parameters are calibrated with time-series of unemployment and vacancy rates, i.e. the Beveridge curve, in Sweden. The parameters are related to the worker separating and vacancy creating processes that exist in the labour market (del Rio-Chanona et al., 2019). Then the calibrated parameters are used to simulate the behaviour of the market under an automation shock spanning several decades. Each occupation is affected differently by the shock based on the results of Frey and Osborne (2017) - where the *probability of computerisation* is estimated for each occupation. The shock does not decrease aggregate demand for labour, rather it reallocates it. Occupations with a low probability of computerisation see an increase in demand, whereas occupations with a high probability of computerisation see a decrease in demand. This means that labour flows between and within occupations are affected by this exogenous shock. In replicating the work of del Rio-Chanona et al. (2019), we are able to draw conclusions within two areas of interest:

- Claims relating to labour markets generally: To what extent is the model able to replicate the Swedish Beveridge curve? What can the differences in optimal parameters between Sweden and the US tell us about the differences in the labour markets?
- Claims relating to the consequences of automation: What does the model say about the performance of the Swedish labour market under an automation shock? Are the phenomena found by del Rio-Chanona et al. (2019) also present in Sweden?

In addition, del Rio-Chanona et al. (2019) make claims related to the mechanics that underpin the Beveridge curve. Testing their model with significantly different data provides additional evidence regarding these claims. Furthermore, some of the aforementioned conclusion may be used to instruct policy. Specifically relating to the structure of optimal job training programmes. In addition, the results are compared to a null model of the labour market without skill restrictions. When conducting the analysis with this base, the results change in ways expected by theory. This shows that when workers' movements between occupations are restricted by skill, there are significant frictions which may create bottlenecks and result in higher unemployment.

In order for the model to be viable, a number of assumptions of various strength are made. The consequences of these assumptions on the results are explored in the discussion but brief descriptions of key issues encountered when making the thesis are provided here. It is assumed that workers are perfectly geographically mobile (del Rio-Chanona et al., 2019). This assumption allows workers to apply to any open vacancy which they have the prerequisite position for. However, it is not an assumption that holds in Sweden - which is sparsely populated outside of the cities. This is, of course, an issue for del Rio-Chanona et al. (2019) as well, which they solve by decreasing the scope of the model from the entire US to a representative city. This solution comes with its problems since national data is more readily available than local data, something that is true in Sweden as well. To this end, the unit of analysis is chosen to be a representative city which has worker flows between occupations of the same scale as those for the entire country (del Rio-Chanona et al., 2019). In the US there is a plethora of big cities, making the idea of a representative city more palatable, which is not the case in Sweden. Therefore, the results presented in section 5 are not based on a representative city. Rather, the implications of the geographical mobility assumption are explored in 6 where a potential solution is mentioned.

Furthermore, the estimates of computerisation probabilities of occupations from Frey and Osborne (2017) is used, where occupations are classified using the American SOC (Standard for Occupation Classification). However, Sweden uses its own classification system (SSYK). Fortunately, there exists cross-walks between different classification systems occupations exist. First, a crosswalk between SOC to ISCO (International Standard Classification of Occupations) is used. Then the official crosswalk between ISCO and SSYK provided by Statistics Sweden is used for the final step. However, the mapping is not one to one which means that assumptions are needed regarding the transformation of the automation shock between systems. Furthermore, there are no worker nor firm optimisation behaviour with respect to how workers choose vacancies to apply for. This may be incorporated into the model by adding wages, a feature that exists in the model by Axtell, Guerrero, and López (2019). However, the focus of this thesis is to replicate the findings of del Rio-Chanona et al. (2019) using different data which means that this feature is not something that will be incorporated here.

Similar to del Rio-Chanona et al. (2019) we find that certain occupations have a counter-intuitive change in unemployment due to the automation shock relative to their computerisation probabilities. As is shown throughout the thesis, the computerisation probabilities of an occupations neighbours[2] affects the labour market outcomes of the occupation. This means that certain occupations with a low probability of computerisation, might still be aversely affected by automation. While, some occupations with a high probability of computerisation will not be as affected by automation as might be expected. This effect likely holds - despite problems encountered when calibrating the model.

2 Literature Review

2.1 Network models of the labour market

This sort of model comes from a large body of literature that has showed the usefulness of analysing labour market flows with networks and *agent-based* models (Schmutte, 2014) (Dworkin, 2019) (Nimczik, 2017) (Michael & Richiardi, 2018) (Alabdulkareem et al., 2018) (Guerrero & Axtell, 2013) (Lopez, Guerrero, & Axtell, 2015) (Neffke, Otto, & Weyh, 2017) (Diodato & Weterings, 2015) (Axtell et al., 2019) (Goudet, Kant, & Ballot, 2017) (Jackson & Kanik, 2019). In fact, the concept originated with Granovetter (1977) who showed the important effect that distant acquaintances in social networks have in spreading information about vacancies. While the network analysed in this thesis is not a social network, a similar concept underlies both ideas: workers are constrained in what vacancies they apply for. The works referenced above tend to use networks to study these constraints or frictions that exist in the labour market. Workers are bound by their geography, education and skills which prevent them from freely moving between firms, occupations or industries. In order to get a grasp of how these frictions play out in the labour market, data is used to construct networks depicting different aspects of labour market frictions.

For example, Axtell et al. (2019) uses employer-employee matched records to construct a network of firms. In their work, each node is a firm and a weighted, directed edge exist

[2]nodes with an edge to the node in question, see 7.1

from firm i to firm j corresponding to the flow of workers from i to j. Accounting for this network, they model application decisions of unemployed workers and hiring policies of firms. This enables them to create a dynamic model where workers transition in and out of employment in different firms. Their work finds that hiring behaviour may generate bottlenecks and uneven concentrations of unemployment around specific companies. Specifically, they show that these behaviours correlate through the network (hiring behaviour of a firm affects the behaviour of firms around it) and that the network topology (see Network terms glossary) induces a large part of frictional unemployment (Axtell et al., 2019). An effect of this kind is often called a network effect, where each node is affected by, and affects, the nodes around them, a result of this kind is presented in section 5. Mealy et al. (2018) take a different approach to the construction of the network, looking at occupations rather than firms. In their network, nodes are specific occupations and edges do not correspond to worker flows directly, although they present a network of this kind as well. In addition, they develop a measure of similarities between each pair of occupations using descriptions of discrete work activities that are undertaken at these occupations. This measure serves as the weighted, undirected edge between occupations and correlates to a high degree with the actual labour flows between occupations, meaning that it is a good predictor of occupational transitions (Mealy et al., 2018). The work of del Rio-Chanona et al. (2019) combine these different approaches by creating a dynamic model of worker flows - similar to that of Axtell et al. (2019) - between occupations.

The two main components of the model by del Rio-Chanona et al. (2019) are the *occupational mobility network* and the dynamic model where workers are *random walker agents* similar to those of Axtell et al. (2019). Each node in the *occupational mobility network* is an officially classified occupation category and there exists an edge from node i to node j if workers have transitioned from occupation i to j within a period of time. The network is constructed using empirical data and may be combined with the estimates from Frey and Osborne (2017) by drawing the network and colouring the occupations (or nodes) based on the computerisation probability estimate of that occupation. This network allows us to get an idea of how the computerisation probability is distributed among different categories of occupations and how these occupations relate to each other. Analysing these networks allows for insight into the credibility of the claims raised in section 1 but does not provide a full answer. However, there is a literature which uses labour flow networks of different kinds to build a dynamic model where workers flow from nodes to nodes through the edges in the network (Jackson & Kanik, 2019) (Dworkin, 2019) (Axtell et al., 2019). In the particular model proposed by del Rio-Chanona et al. (2019) workers are separated (let go from their job) and vacancies are created within each occupation. In the next period separated workers apply to a vacancy within an occupation that neighbours (is connected to) the occupation the worker got separated from. Data then drives the calibration of the model's parameters. Using the calibrated model and the results from Frey and Osborne (2017) an *automation shock*, happening over several decades, is introduced in the model. In this way the model provides a way of measuring unemployment dynamics at the occupational level due to shifts in labour demand following automation. It also allows us to examine to what extent the network structure of the labour market affects the efficiency of labour allocation (del Rio-Chanona et al., 2019). In this way, the authors demonstrate that the network structure affects labour market outcomes due to an automation shock to a large extent. The specifics of this model is formally developed in section 3.

2.2 The effect of automation on labour markets

Both within and, but perhaps especially, outside of academia there has been a flurry of interest in the automation of labour over the last decade. It should of course be noted that economists have studied automation arguably since the inception of the field. However, today there is a notion among many that we are standing on the precipice of another technological revolution. But what do the experts have to say about this? The answer is found in the work of Frey and Osborne (2017), who used a list of descriptions of work tasks related to different occupations and asked experts what probability they would place on a specific task being *computerised* in the coming decades. They managed to compile the probability of computerisation of all 722 occupations specified by the four digit SOC code (each containing a variety of specific tasks classified in O*Net[3]). Since interviewing experts about each of these tasks would be too time consuming for everyone involved they instead took a sample of the tasks and used advanced machine learning methods to estimate the remaining tasks (Frey & Osborne, 2017). A similar approach has been conducted by Brynjolfsson et al. (2017), Brynjolfsson et al. (2018) - where they constructed a measure of *suitability for machine learning*. While del Rio-Chanona et al. (2019) implements shocks based on both of the aforementioned estimates - their main result is based on Frey and Osborne (2017) and this is the only one implemented here.

3 Theory

In this section, the theoretical framework of the data driven occupational mobility model is developed. The model may be viewed as having two, interconnected components - an occupational mobility network and an agent-based model. Below each of these are described separately and the formal model (which combines the two) is derived. As we shall see, there is a way to solve the model deterministicly using only the parameters and the edges of the network - which is much more computationally efficient than the agent-based simulation. For brevity, the derivations of the deterministic solution are not included here but may be found in the Supplemental Information of del Rio-Chanona et al. (2019).

3.1 Occupational mobility network

A network, or graph, is defined by an *adjacency matrix*, which is a matrix that has the same set of labels on both its columns and rows. Each label then defines a *node* and there exists an *edge* (or *link*) from one node to another if the corresponding value in the adjacency matrix is non-zero. Formally speaking, if A is an adjacency matrix with elements A_{ij} then i and j are nodes and an edge between them exist if $A_{ij} \neq 0$ (if negative links are allowed). If there exists an edge between two nodes, we call these nodes neighbours. Neighbours are an important concept in Network theory since they describe the local surroundings of a node. In the context of an occupational mobility network, neighbouring nodes have labour flows between them. As will be shown, an occupation's neighbours are important in determining the unemployment outcome of the occupation as a result of the automation shock.

The adjacency matrix used to define Sweden's occupational mobility network is con-

[3] Occupational Information Network

structed from the occupational transitions between 2016 and 2017. That is, the matrix contains information about the number of workers who transitioned between occupations. In the raw data acquired from Statistics Sweden[4] the rows and columns are labelled with 3-digit SSYK codes. In this way, each code defines a distinct occupation which is a node in the network. The elements of the matrix are labour flows from the row occupation to the column occupation between 2016 and 2017. Depending on the structure of the *adjacency matrix* the resulting network may have different attributes. In our case we have a *directed graph*, where there might exist an edge from i to j but not from j to i. In other words, the matrix is not symmetric around the diagonal. In addition the graph has *self-loops*, which means that each node points to itself. This is the case since many (most) workers stayed within their occupation between 2016 and 2017. As shown in the section below, these attributes of the network allow us to define a simple agent based model where the automation shock is implemented.

So far labour flows have been discussed in absolute terms, which is how the raw data is constructed, but in the occupational mobility network proposed by Mealy et al. (2018), edges are weighted by transition probabilities. Therefore, we introduce the adjacency matrix A which defines the occupational mobility network:

$$A_{ij} = \frac{T_{ij}}{\sum_k T_{ik}} \tag{3.1}$$

Where T is the matrix containing raw data on number of workers transitioning between occupations described above. In this way, we estimate the transition probabilities between all occupations. Transition probabilities are used to shape the behaviour of the agents which populate the model. As we will show in the coming sections, they are the foundation used to determine labour flows between occupations. In their work, Mealy et al. (2018) showed that their occupational mobility network performed well at predicting occupational transition rates in the US. Since a similar analysis has not been conducted in Sweden - we assume that the transition probabilities estimated with A_{ij} correlate sufficiently with transition rates in Sweden. Having transition data over several years would, of course, make this assumption more palatable but transitions over one year is sufficient to conduct the analysis. Furthermore, the coming sections show how additional data is used to calibrate other aspects of the model.

In Figure 1 and 2, the occupational mobility networks constructed using Swedish data are presented. Since the data used to construct the networks do not have a spatial component - the position of each node is calculated using an algorithm. The layout of the first network in each figure is based on a spring algorithm, where each edge acts like a spring pulling the nodes together. The nodes are initialised at random points in the xy plane and then their positions change in discrete time steps based on the forces applied by the springs. In this way, nodes that are more interconnected tend to be grouped closer to one another, which allows some structure of the network to emerge. It should be noted that since there are over 2000 edges in the network and only 143 nodes, the spring algorithm does not converge to a steady state, where all the forces are balanced, quickly.

Nevertheless, looking at Figures we see that nodes of the same colour tend to be close to each other. This is not trivial since the spring algorithm does not take into account

[4]SCB: Statistiska Centralbyrån

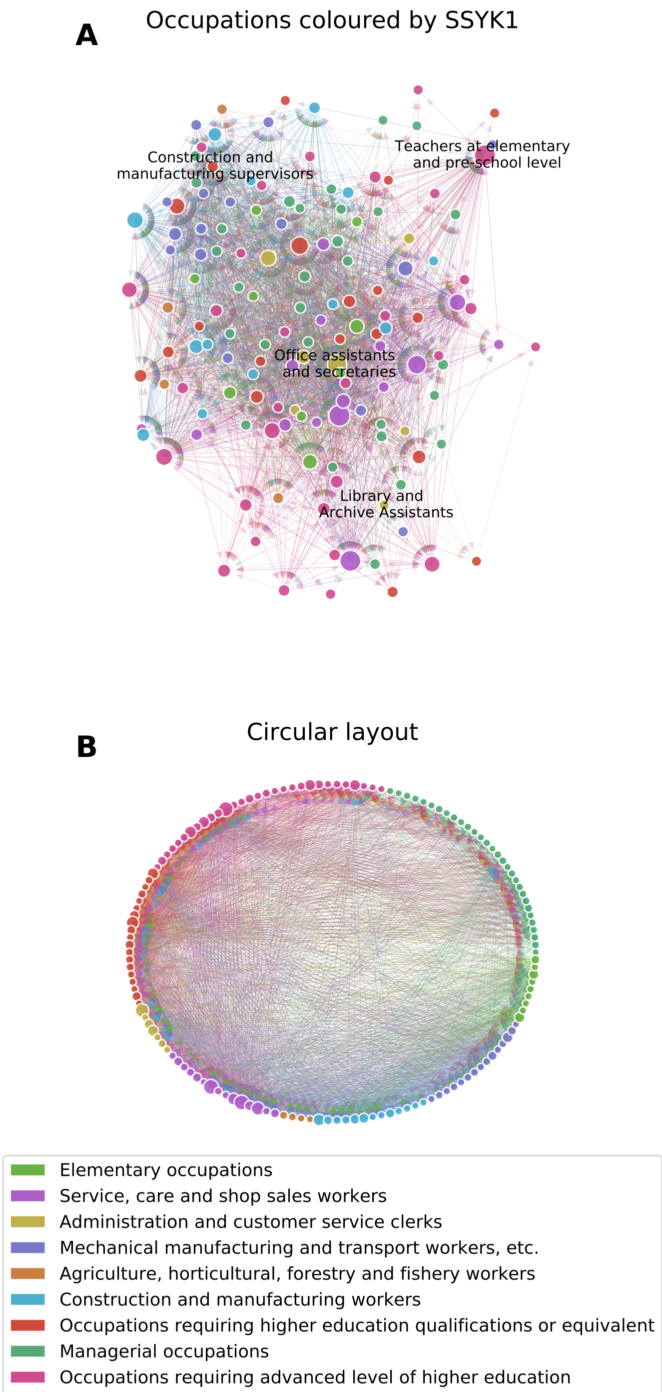

Figure 1: *The Swedish occupational mobility network using two different algorithms for deciding the positions of each node. The size of the node is proportional to the amount of workers who stayed in the occupation and the colour of each node and its edges is based on the first level SSYK code. Some nodes are labelled with the description of the category they represent.* ***A****: Spring layout, where each edge pulls on the target node with a force inversely proportional to the edge weight.* ***B****: Circle layout, where each node is placed on a circle, the first being the lowest SSYK value and the last being the highest SSYK value. Node sizes for the circle layout are half as large.*

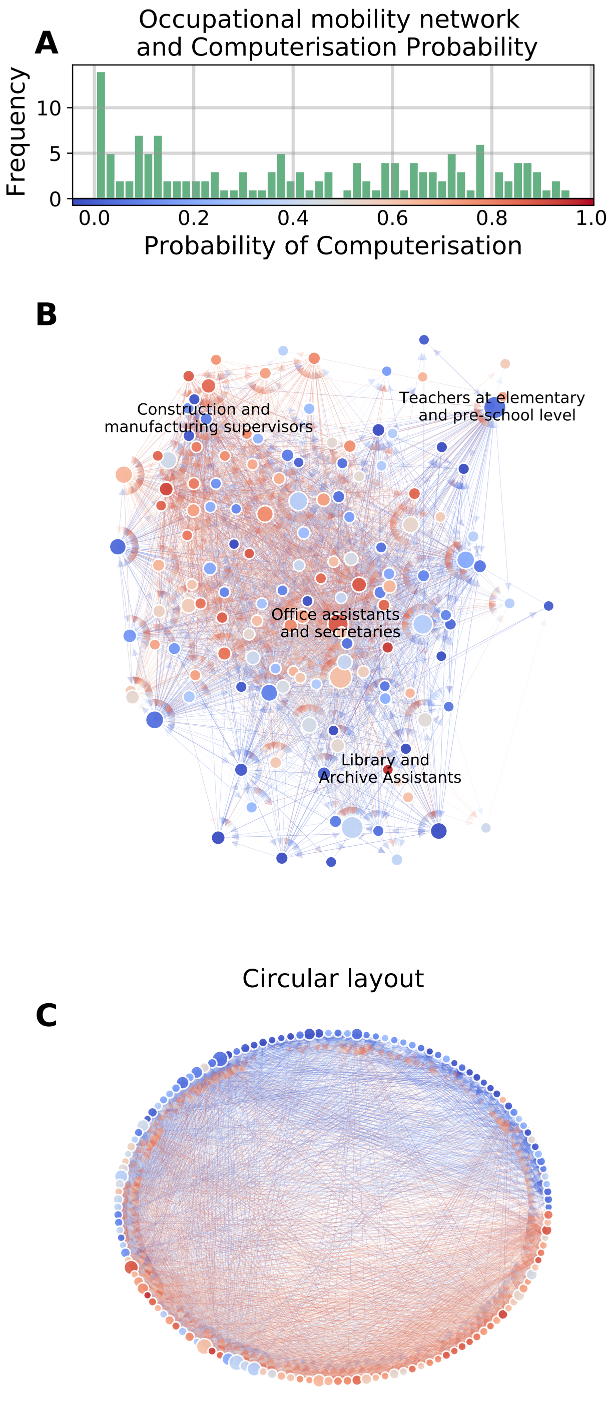

Figure 2: *The Swedish occupational mobility network and computerisation probability distribution. Some nodes are labelled with the description of the category they represent.* ***A****: The frequency distribution of computerisation probability values.* ***B****: The occupational network using the same spring layout algorithm from Figure 1* ***A*** *but with nodes and edges coloured by their probability of computerisation.* ***C****: Circle layout of Figure 2* ***B****, again nodes coloured by probability of computerisation.*

that some nodes belong to the same occupational category or have similar levels of computerisation probability. Rather, it shows that transitioning workers tend to move to an occupation in the same higher order category as the one they came from (Figure 1 **B**). Similarly, transitioning workers tend to move to occupations with a similar level of computerisation probability (Figure 2 **B**, **C**). However, there are exceptions which will be shown to play an important role in our main result. The networks are also presented using a circular layout, where each node is placed on the unit circle, ordered by SSYK code. This type of layout can be useful for networks with large number of edges relative to the number of nodes since it is easier to see the edges between nodes. For example, workers in Elementary occupations transition to many different categories of occupations (Construction and manufacturing, Service, care and shop sales workers but also occupations requiring higher education qualifications or equivalent).

Four nodes are labelled in the graphs, we are briefly going to discuss two of them: Construction and manufacturing supervisors, and Office assistants and secretaries. The former has a low probability of computerisation and the latter a high. However, the Construction and manufacturing supervisor node is surrounded by a lot of red nodes, whereas the Office assistants and secretaries are surrounded by blue nodes (Figure 2 **B**). This means that there are labour flows between Construction supervisors and other occupations with high probability of computerisation. Conversely there are labour flows between Office assistant and occupations with low probability of computerisation. As is shown in section 5, this means that they will be affected differently by the automation shock relative to other occupations with similar computerisation probabilities. The Construction supervisor occupation observes an increase in unemployment after the shock - even though the labour demand increases for this occupation. The explanation is that workers in neighbouring occupations which are affected by decreased labour demand as a result of automation will all apply to the vacancies in Construction supervision. This means that it is harder for unemployed Construction workers to find jobs there, in addition, there are few available jobs in neighbouring occupations as well which means that the workers stay unemployed. Conversely, the Office assistants are able to apply to the vacancies in neighbouring occupations which observe an increase in labour demand as a result of the shock. Which means that despite a lot of Library assistant losing their jobs - relatively many of them are able to find jobs elsewhere. As mentioned, this is described in section 5. However, before we get there, the coming sections describe the theoretical model and the methods used to achieve these results.

3.2 Agent based model

The ABM framework allows us to introduce rules for workers which, combined with data, produce labour market outcomes. To that end there are two types of entities in the model, workers and vacancies. In addition, both entities belong to one distinct occupation and have two different states. Workers are either employed or unemployed and vacancies are either open or close, where both always belong to an occupation. Unemployed workers are able to apply to open vacancies within their occupation, i, or within a neighbouring occupation j, where there exists an edge from i to j. In this way - workers traverse the occupational mobility network in a guided, stochastic fashion. Workers move stochastically since they are only allowed to apply to one vacancy, which they do at random. However, the probability at which they apply to vacancies are weighted by the transition probabilities T_{ij} defined above. This means that workers in occupation i are more likely to apply

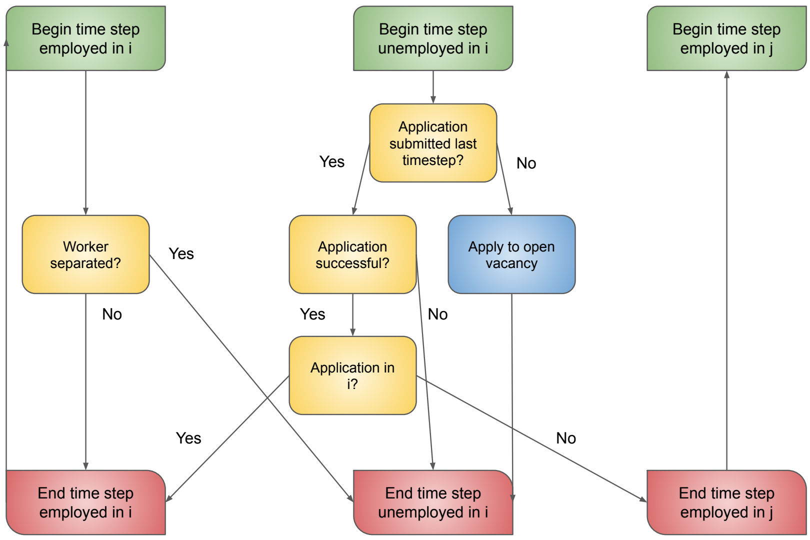

Figure 3: *Flow chart of worker behaviour during simulation*

to a vacancy in occupations j when the *transition probability*, T_{ij}, is higher. When the unemployed workers have made their application, each vacancy accepts one application with uniform probability. Vacancies that do not receive applications remain open until the next period (Axtell et al., 2019) (del Rio-Chanona et al., 2019).

Thus far we have taken the states of the entities, employed or unemployed for workers and open or closed for vacancies, for given. But these are determined within the model by a combination of parameters and an initial, empirical, employment distribution across occupations. In turn, the parameters are calibrated such that the model outputs a Beveridge curve which is similar to the empirical one. The transition probabilities and the rules followed by workers and vacancies outlined above determine what choices the workers are able to make in the model. In addition, it is required in order to determine who will be hired where. But it does not tell us who will lose their job or in which occupations vacancies are available. These aspects of the model are set by separate stochastic processes. As will be explained in formal detail below, workers transition into unemployment and vacancies are opened following a binomial stochastic process (del Rio-Chanona et al., 2019).

3.3 Dynamic labour market model

In the replicated model worker flows in the network are described by discrete time stochastic processes for employment, unemployment and vacancies within each occupation. The following assumptions are made:

1. Workers are perfectly geographically mobile
2. Wage pressure is neglected
3. The set of possible occupations are fixed

In addition, a worker's occupation is the one she was last employed in. We begin formalising the model by letting $e_{i,t}$, $u_{i,t}$ and $v_{i,t}$ be employment, unemployment and vacancies respectively in occupation i at time t. In addition, let $d_{i,t} = e_{i,t} + v_{i,t}$ denote the realised demand for labour in occupation i at time t. As mentioned, these are modelled as stochastic processes:

$$e_{i,t+1} = e_{i,t} - \underbrace{\omega_{i,t+1}}_{\text{separated workers}} + \underbrace{\sum_j f_{ji,t+1}}_{\text{hired workers}} \tag{3.2}$$

$$u_{i,t+1} = u_{i,t} + \underbrace{\omega_{i,t+1}}_{\text{separated workers}} - \underbrace{\sum_j f_{ij,t+1}}_{\text{transitioning workers}} \tag{3.3}$$

$$v_{i,t+1} = v_{i,t} + \underbrace{\nu_{i,t+1}}_{\text{opened vacancies}} - \underbrace{\sum_j f_{ji,t+1}}_{\text{hired workers}} \tag{3.4}$$

Where $\omega_{i,t}$ are the number of workers separated at time t in occupation i, $\nu_{i,t}$ the number of vacancies opened at time t in occupation i and $f_{ij,t}$ are the number of workers from occupation i hired in occupation j at time t. In this way, all workers are accounted for. The above set of master equations define the dynamics of the model and are the basis of what occurs at each time step. As will be shown, they characterise a non-equilibrium model where vacancies are opened and workers are separated within each occupation and at each time step, in order for the realised demand, $d_{i,t}$, to reach the target demand, $d^{\dagger}_{i,t}$ of the occupation. In addition, the equations are used to find an approximate, deterministic solution to the model. Furthermore, the number of separated workers $\omega_{i,t}$ as well as the number of opened vacancies, $\nu_{i,t}$ are modelled as independent binomial processes:

$$\omega_{i,t+1} \sim Bin\left(e_{i,t}, \pi_{u,i,t}\right) \tag{3.5}$$

$$\nu_{i,t+1} \sim Bin\left(e_{i,t}, \pi_{v,i,t}\right) \tag{3.6}$$

Where $Bin(n, p)$ denotes the binomial distribution with n trials and success probability p. Thus $\pi_{u,i,t}$ and $\pi_{v,i,t}$ are the probability that a worker employed in occupation i at time t gets separates and that for each worker in occupation i at time t a vacancy opens, respectively. We break each of these into two random processes: a *state-dependent* process and a *state-independent* (or *spontaneous*) process. The separation, state dependent process is defined such that workers are more likely to get separated if the realised demand is higher than the target demand. Conversely, the vacancy state dependent process is defined such that vacancies are more likely to be opened if the realised demand is lower than the target demand. In this way, the state dependent processes are adjusting the realised demand towards the target demand. However, in the spontaneous processes, workers are separated and vacancies are opened at random. Therefore, the occupations will tend to not be in demand equilibrium and even if at some point they there, the spontaneous processes will shift them out of equilibrium.

In order for workers to flow through the network there has to be a match between an unemployed worker and an open vacancy. A worker is unemployed in occupation i, and will apply to *one* vacancy in occupation j that neighbours i *and* has at least one open vacancy. When all the unemployed workers have applied, each open vacancy that was applied for chooses one worker at random who fills the position and closes the vacancy. If a vacancy is not applied for, it remains open. These rules enable us to calculate the probability ,$q_{ij,t+1}$, that an unemployed worker in occupation i applies to a vacancy in occupation j. This is done by taking the number of vacancies open in j, multiplied by the probability that the worker applies to j, divided by the total number of available weighted options:

$$q_{ij,t+1} = \frac{v_{j,t}A_{ij}}{\sum_l v_{l,t}A_{il}} \tag{3.7}$$

Therefore the expected number of applications submitted from occupation i to occupation j is the probability $q_{ij,t+1}$ multiplied by the number of unemployed workers in $u_{i,t}$:

$$E\left[s_{ij,t+1}|u_{i,t}\right] = u_{i,t}q_{ij,t+1} \tag{3.8}$$

This is because each worker sends one application and because the random variables $s_{ij,t+1}$ follow a multinomial distribution with $u_{i,t}$ trials and probabilities $q_{ij,t+1}$ for fixed i and $j = 1, ..., n$, where n is the number of nodes in the network. All vacancies which receive applications hire one worker but, as mentioned above, some may not receive applications and will then be left open.

This is the framework that is used to model worker flows across the network. Formally, we let δ_u and δ_ν denote the overall, spontaneous probability that a worker is separated and a vacancy is opened respectively. Where overall means that the probability is the same, regardless of occupation or time step (which is another way of expressing state independent). As mentioned, fluctuations in labour demand also influence the state of workers and vacancies. This is governed by a *state-dependent* process, where workers are more likely to be separated if realised demand, $d_{i,t}$, (current number of employed workers plus number of open vacancies) is higher than the target demand, $d^{\dagger}_{i,t}$. Conversely, vacancies are more likely to be opened if current demand is lower than the target demand. Formally, let $\alpha_{u,i,t}$ and $\alpha_{\nu,i,t}$ denote the state-dependent probability that a worker is separated or a vacancy is opened due to an imbalance between current and target demand. As opposed to the *spontaneous* processes, these probabilities depend on both the occupation, and the time. Then, the probability that a worker is not separated is $(1 - \delta_u)(1 - \alpha_{u,i,t})$, since the worker does not get separated by the *spontaneous* process nor the *state-dependent* process. Similarly, the probability that a vacancy is not opened is $(1 - \delta_\nu)(1 - \alpha_{\nu,i,t})$. This allows us to express the success probabilities of the binomial processes above as:

$$\pi_{u,i,t} = 1 - (1 - \delta_u)(1 - \alpha_{u,i,t} = \delta_u + \alpha_{u,i,t} - \delta_u\alpha_{u,i,t} \tag{3.9}$$
$$\pi_{\nu,i,t} = \delta_\nu + \alpha_{\nu,i,t} - \delta_\nu\alpha_{\nu,i,t} \tag{3.10}$$

Which are the probabilities that a worker gets separated and a vacancy opened. As mentioned, the probabilities denoted α are state dependent. Specifically they depend on the imbalance between realised and target demand in occupation i and at time t. Furthermore, these probabilities are constructed such that they minimise this imbalance. The imbalance is given by the difference between the *target demand* and the *realised* *demand* at the time and occupation. Target demand, $d^{\dagger}_{i,t}$, is the desired quantity of labour in occupation i at time t. This is not endogenous to the model, making it the channel that is used to implement the automation shock over time, more on this in section 4.2. In contrast, realised demand is internal and the sum of employed workers and vacancies in occupation i at time t:

$$d_{i,t} = e_{i,t} + \nu_{i,t} \tag{3.11}$$

In order for $\alpha_{u,i,t}$ and $\alpha_{\nu,i,t}$ to fill the roles given above, they must satisfy the following conditions:

1. If there is no labour supply and demand imbalance, i.e. $d^{\dagger}_{i,t} = d_{i,t}$, then no adjustment should be made and $\alpha_{u,i,t} = \alpha\nu, i, t = 0$.
2. When realised demand is higher than target demand, $\alpha_{u,i,t} > 0$ such that more workers are separated to decrease realised demand. Conversely, when realised demand is lower than target demand, $\alpha_{\nu,i,t} > 0$ such that more vacancies are opened to increase realised demand. Therefore $\alpha_{u,i,t}$ is an increasing function of $d_{i,t} - d^{\dagger}_{i,t}$ and $\alpha_{\nu,i,t}$ is an increasing function of $d^{\dagger}_{i,t} - d_{i,t}$.
3. $\alpha_{u,i,t}$ and $\alpha_{\nu,i,t}$ are probabilities and thus lie in the interval $[0, 1]$.

There are many functional forms of $\alpha_{u,i,t}$ and $\alpha_{\nu,i,t}$ that satisfy the above conditions. However, del Rio-Chanona et al. (2019) assume that the supply and demand equilibrate linearly with respect to the imbalance, which gives the following form:

$$\alpha_{u,i,t} = \gamma_u \frac{\max\left\{0, d_{i,t} - d^{\dagger}_{i,t}\right\}}{e_{i,t}} \tag{3.12}$$

$$\alpha_{v,i,t} = \gamma_v \frac{\max\left\{0, d^{\dagger}_{i,t} - d_{i,t}\right\}}{e_{i,t}} \tag{3.13}$$

Where the max function ensures that condition 3) is satisfied. γ_u and γ_ν are parameters that determine the speed of adjustment and lie in the interval $[0, 1]$. Where a value of 1 means maximum adjustment speed and a value of 0 means no adjustment at all. Here del Rio-Chanona et al. (2019) is again followed and the parameters are such that $\gamma_u = \gamma_\nu$. This is because we only observe the aggregate data and are therefore unable to calibrate them separately. In order to calibrate the parameters $\delta_u, \delta_\nu, \gamma$ and τ (the length of the time step in weeks) data on vacancy rates and unemployment, the Beveridge curve, for Sweden is used. This is because the model outputs the number of employed and unemployed workers as well as the number of vacancies every time step, which allows us to plot the model's progression on the Beveridge curve. The results of the calibration are detailed in section 5.1

3.3.1 Deterministic Solution of the model

Using the law of large numbers and multivariate Taylor expansion, it is possible to solve the model deterministicly at each time step. This is useful since even though the rules of how workers and vacancies behave are simple, it is computationally costly. This is mostly due to the large number of possible choices that unemployed workers have (del Rio-Chanona et al., 2019). In addition to saving computational time, the output of the deterministic approximation is easier to analyse, which is significantly useful for exploring the parameter space and calibrating the model. This approximation is built upon analysing the system's behaviour in terms of expected values.

The purpose of the approximation is to find expressions for the expected values of employment, unemployment and vacancies for each occupation and time. To keep the notation impact let:

$$\bar{u}_{i,t+1} \equiv E[u_{i,t+1} \mid \boldsymbol{u}_{i,t}, \boldsymbol{v}_{i,t}, \boldsymbol{e}_{i,t}] \tag{3.14}$$

The master equations are reduced to a $3n$ dimensional deterministic dynamical system of equations given by:

$$\bar{e}_{i,t+1} = \bar{e}_{i,t} - \underbrace{\left(\delta_u \bar{e}_{i,t} + (1-\delta_u)\,\gamma_u \max\left\{0, \bar{d}_{i,t} - d^{\dagger}_{i,t}\right\}\right)}_{\text{separated workers}} + \underbrace{\sum_j \bar{f}_{ji,t+1}}_{\text{hired workers}} \tag{3.15}$$

$$\bar{u}_{i,t+1} = \bar{u}_{i,t} + \underbrace{\left(\delta_u \bar{e}_{i,t} + (1-\delta_u)\,\gamma_u \max\left\{0, \bar{d}_{i,t} - d^{\dagger}_{i,t}\right\}\right)}_{\text{separated workers}} - \underbrace{\sum_j \bar{f}_{ij,t+1}}_{\text{transitioning workers}} \tag{3.16}$$

$$\bar{v}_{i,t+1} = \bar{v}_{i,t} + \underbrace{\left(\delta_v \bar{e}_{i,t} + (1-\delta_v)\,\gamma_v \max\left\{0, d^{\dagger}_{i,t} - \bar{d}_{i,t}\right\}\right)}_{\text{opened vacancies}} - \underbrace{\sum_j \bar{f}_{ji,t+1}}_{\text{hired workers}} \tag{3.17}$$

A large part of the derivations are to show that the expected flows between occupations, $\bar{f}_{ij,t+1}$, may be written in terms of the adjacency matrix and expected values of the state variables. The results are shown below:

$$\bar{f}_{ij,t+1} = \frac{\bar{u}_{i,t}\bar{v}^2_{j,t}A_{ij}\left(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}\sum_k \bar{v}_{k,t}A_{ik}} \tag{3.18}$$

$$\bar{s}_{j,t+1} = \sum_i \frac{\bar{u}_{i,t}\bar{v}_{j,t}A_{ij}}{\sum_k \bar{v}_{k,t}A_{ik}} \tag{3.19}$$

The relative error of the approximation is:

$$\left|\frac{E\left[f_{ij,t+1}|\mathbf{u}_t, \mathbf{v}_t; A\right] - \bar{f}_{ij,t+1}}{E\left[f_{ij,t+1}|\mathbf{u}_t, \mathbf{v}_t; A\right]}\right| < \frac{c}{L+c} \tag{3.20}$$

Where c is a constant and L is the size of the labour force. This means that given a set of time series for the target labour demand $d^{\dagger}_{i,t}$ and a set of initial conditions, the above equations determine the expected employment, unemployment, and vacancies as a function of time. In principle, this framework may be used to study all countries where data is available and any type of occupational labour demand shock. The full derivation of the deterministic, expected solution is available in the Supplemental Information of del Rio-Chanona et al. (2019).

3.4 The Beveridge Curve

The Beveridge curve is a well known macroeconomic stylised fact, stating the relationship between the vacancy rate and unemployment (Diamond, 1982) (Beveridge, 2014). The idea is that when more vacancies open, unemployment goes down as workers gain more opportunities in the job market. Inversely, when fewer vacancies are opened, unemployment goes up as workers have access to fewer job opportunities. Figure 4 shows the Beveridge curve for Sweden from Q1 2004 to Q4 2019.

There are three features of the Beveridge curve which have been shown to exist in many economies (Diamond, 1982) (Diamond & Şahin, 2015) (Bouvet, 2012):

1. The curve may shift away or toward the origin
2. Unemployment and vacancy rates both tend to move downward along the curve during recessions and tend to move upward along the curve during recovery
3. The curve tends to shift outwards during a recovery, meaning that the curve would cycle counter clockwise

There are several models that are able to explain the first two features of the Beveridge curve. A prominent one is the Diamond-Mortensen-Pissarides model, which explains shifts in the Beveridge curve with structural changes such as skill mismatches (Diamond, 1982) (Pissarides, 2011). In addition, the first feature was explained by Axtell et al. (2019) (which was mentioned in section 2), who also showed that changes in network structure may shift the Beveridge curve towards or away from the origin. However, the third feature is not fully understood yet. It has been argued that this is because of increased frictions in the matching process of the economy and therefore a result of a structural change (Diamond & Şahin, 2015). While others argue that the counter-clockwise movement of the curve could be independent of structural change and is instead due to the business cycle (Pissarides, 1985)(Mortensen, 1999). There is a difference in the flexibility of the variables in these models, where vacancies arise immediately following the decisions of firms but unemployment only decrease as a result of a match in the labour market. Because of the lag in the hiring process relative to the vacancy creating process - the vacancy rate recovers faster than the unemployment rate which gives you a counter clockwise motion of the Beveridge curve. The model presented above supports this hypothesis since a simulated business cycle results in counter clockwise motion in the Beveridge curve produced by the simulation (del Rio-Chanona et al., 2019). In fact, as is explained in section 5.1 the model parameters are calibrated in order for the output of the model given a simulated business cycle to match that of the Beveridge curve during an observed business cycle. However, it should be noted that most of this research is conducted on the US Beveridge curve. The behaviour of the Swedish Beveridge curve is described in the following section 3.5.

3.5 The Beveridge curve in Sweden

As is explained in section 4.3, the Swedish Beveridge curve presented here is based on seasonally adjusted, quarterly data from 2004 to 2019. In addition, the curve is coloured by whether the economy is in a recession or a recovery. Where a a quarter is defined as being in a recession if the output gap in GDP is lower than the previous quarter. As is seen in the figure, the curve cycles counter clockwise from 2004 to 2014. However, in 2012, the curve cycles clockwise and, in the most recent periods, the curve looks like it might exhibit clockwise movement. In addition, there are a few periods in which there has been movement upwards along the curve during a recession and movement downwards along the curve during a recovery. However, the features mentioned above are generally exhibited in the Swedish Beveridge curve (Jonsson & Theobald, 2019).

Since the model is calibrated on this empirical data, we need to find a suitable period for which to calibrate the model. To this end we will use the most recent business cycle, starting in 2008Q3 (where the recession starts in the figure) and ending in 2016Q3 (where the recovery ends in the figure). At this point it looks like the curve has traversed a bit

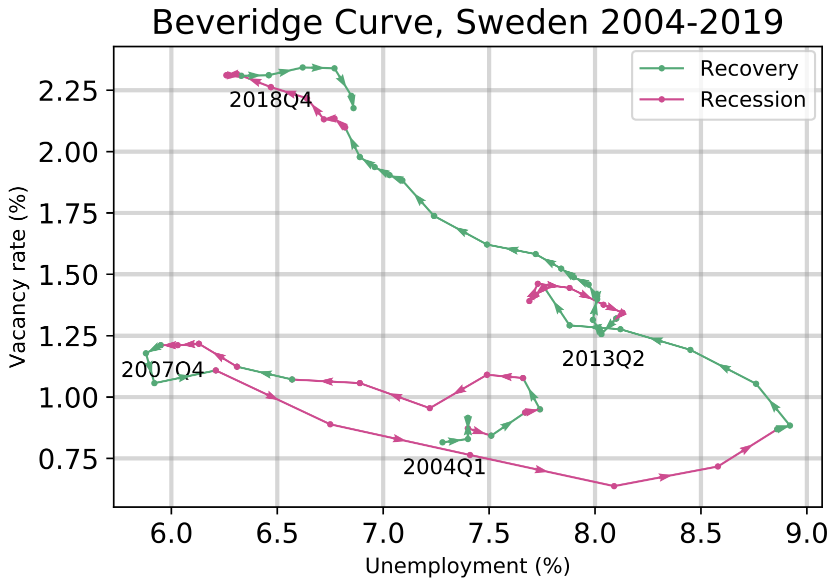

Figure 4: *Sweden's Beveridge curve using seasonally adjusted quarterly data from 2004 to 2019*

less than four fifths of the period, which means that we set the period to 10.25 years. The assumption about phase do not affect the results (del Rio-Chanona et al., 2019). It is out of the scope of the thesis to try to explain what happened in the small recession of 2012, where the curve exhibits 'abnormal' behaviour.

4 Materials & Methods

In this section, the practical aspects of implementing an automation shock into the model are developed. In addition, the data and its origins is described. Furthermore, the model is coded in Python where a novel package of functions, written by the author, is used to conduct the analysis. More details regarding the code is found in section 7.2.

4.1 Labour Reallocation

Using the calibrated values for the *spontaneous* probabilities δ_u and δ_ν as well as the rate of adjustment γ and the occupational employment in 2016 as a starting point, the shock is implemented into the model. As mentioned previously, the shock is determined by the *computerisation probabilities* developed in Frey and Osborne (2017). These probabilities determine the post-shock target demands,$d_i^\dagger$ of each occupation, which need to be specified in the model. The key assumption here is that these probabilities specify the fraction of total hours worked in an occupation that are automated post shock, i.e. the fraction of total hours worked that do need require human labour. In addition, working hours are reduced for all workers, meaning that the total number of jobs stay constant, i.e there is no change in aggregate demand before, during and after the shock. Furthermore, the labour force (total number of employed and unemployed workers), denoted by L is assumed to stay constant during the simulation (del Rio-Chanona et al., 2019). Now, let x_0 denote the number of hours worked for the average worker in a year. Then the hours of work demanded by each occupation at the start of the simulation is given by the elements of the vector $\boldsymbol{h}_0$:

$$\mathbf{h}_0 = x_0 \mathbf{e}_0 \tag{4.1}$$

Where $\boldsymbol{e}_0$ again is a vector containing the number of workers in each occupation at the start of the simulation. The vector containing the computerisation probabilities of each occupation is denoted by $\boldsymbol{p}$. As described above, this allows us to give an expression for the post-shock human labour hours demanded by each occupation, $\boldsymbol{h}_{t*}$:

$$\mathbf{h}_{t*} = \mathbf{h}_0 \cdot (\mathbf{1} - \mathbf{p}) \tag{4.2}$$

Here, $\cdot$ refers to the element-wise multiplication of vectors and $\mathbf{1}$ denotes the vector of ones. Furthermore, $\boldsymbol{t}^*$ refers to the time at which the automation shock has subsided. Then the post shock, average amount of hours worked per year is:

$$x_{t*} = \frac{\sum_i^n h_{i,t*}}{L} \tag{4.3}$$

Where the post shock, aggregate hours of work is split equally between occupations. To calculate the post shock target demand of the occupations, $\boldsymbol{d}^{\dagger}_{t*}$, it is assumed that automation has no impact on aggregate labour demand unemployment. This allows the hours of labour demanded by occupations to be split equally among workers:

$$\mathbf{d}^{\dagger}_{t*} \equiv \mathbf{d}^{\dagger} = \mathbf{h}_{t*} \frac{1}{x_{t*}} \tag{4.4}$$

As shown in del Rio-Chanona et al. (2019) it is possible to specify the model such that there is an aggregate increase or decrease in the number of jobs (labour demand). Their findings were expected: if there is an aggregate increase in the number of jobs - the changes in unemployment rate as a result of the shock are lower. More jobs simply means that more people find jobs. Conversely, the changes in unemployment rate when there is an aggregate decrease in the number of jobs are higher - less people are able to find jobs as a result of the shock.

4.2 Time dependent shock

Within the innovation literature it is suggested that adoption of technologies begins at an exponential rate but over time decays to a logarithmic rate, which is described by a sigmoidal function (Stoneman, 2001). This gives the shape of the automation shock. The computerisation probabilities estimated by Frey and Osborne (2017) are said to cover 'some unspecified number of years, perhaps a decade or two'. This is probably because it is harder to estimate the progress of technology within an exact time frame. del Rio-Chanona et al. (2019) assumes that the shock happens within 30 years, where the bulk happens within 10 years and also explore different alternatives. In addition, they assume that the initial target demand is the steady state demand and over time it reaches the post-shock reallocated demand $\boldsymbol{d}^{\dagger}$. This yields the following expression for the dynamics of the target demand:

$$d^{\dagger}_{i,t} = \begin{cases} d_{i,0} & \text{if } t < t_s \\ d_{i,0} + \frac{d^{\dagger}_i - d_{i,0}}{1+e^{k(t-t_0)}} & \text{if } t \geqslant t_s \end{cases} \tag{4.5}$$

Where t_0 is the time at which the target demand is in the middle of the initial and post shock values and t_s is the time at which the shock begins. The midpoint of the curve occurs 15 years after the shock begins: $t_0 = t_s + 15$. In addition, the growth rate of the curve, k, is set to 0.79, which guarantees that the target demand equals the post-shock reallocate demand up to a 0.0001 tolerance.

As with the calibration, the model is first initialised so that it converges to the steady state unemployment rate and to the employment distribution of occupations in 2016. When the steady state is reached, the shock is implemented as explained above. del Rio-Chanona et al. (2019) show that when the time span of the automation shock is increased, the change in unemployment rate become lower. And conversely that when the time span is decreased the changes become lower. This is because there is more time for workers to leave occupations which are becoming computerised - there is more time to respond to the automation shock.

4.3 Data

Data plays a central role in this thesis and is used for different purposes. Before we dive deeper into the mechanics of the analysis let us take a moment to describe what the data is and where it is used. First we have empirical data regarding the Swedish labour market, this includes:

1. Occupational transitions between 2016 and 2017
2. Aggregated, quarterly, seasonally and calendar adjusted unemployment data from 2004 to 2020
3. Aggregated, quarterly, seasonally and calendar adjusted employment data from 2004 to 2020
4. Aggregated, quarterly seasonally adjusted vacancy data from 2004 to 2020
5. Aggregated, monthly, seasonally adjusted and smoothed hours worked from 2014 to 2018
6. Occupational employment data from 2014 to 2018

While some of this data is not available directly on Statistics Sweden's website, it has been collected through correspondence with different departments of Statistics Sweden. This data is used for three different purposes. 1) is used for constructing Sweden's occupational mobility network. Using 2), 3) and 4), the Beveridge curve for Sweden is constructed, which is used to calibrate the dynamical model. Employment per occupation, 5), is used as an equilibrium starting point both for the calibration of the model and the automation shock analysis that follows. Finally, hours worked is used to implement the automation shock, where automation plays out by reducing hours worked in some occupations (more on this below). In addition, the results of Frey and Osborne (2017) is used to construct the automation shock that the analysis is built upon. The calibration is independent of this shock, which means that once the model is calibrated a number of different shocks may be applied to it. Conversely the aggregate calibration data (2), 3) and 4)) is not used in the analysis. The above is enough data to conduct the analysis, which is the primary goal of this work. However, for future work a source of data that is left unexplored is that of the Swedish Public Employment Service[5], which publishes forecasts on occupations based on SSYK as well as geographical data on employment. As is discussed in section 6, this could be useful in extending the model in order to relax the assumption of perfect worker

[5] Arbetsförmedlingen

mobility.

4.3.1 Occupational classification with SSYK

1) and 5) is data grouped by the Swedish occupation classification system, SSYK[6]. Each occupation is defined by a 4 digit code, where each digit is a sub-category of the digit before it. The first digit is a number between 0 and 9 and therefore defines 10 broad categories of occupation. For example, the digit 1 defines the category "Managerial Occupations", 11 and 15 then define further sub-categories of "Managerial Occupations" (managerial occupations within politics and healthcare respectively). In this way, the more digits that are used to define an occupation, the more precise that definition becomes and more occupations are defined. The occupational employment and labour flow data received from Statistics Sweden uses 3 digits. This presents some challenges when translating the automation shock, which are discussed in the next section. In addition, The data that is used to define Sweden's occupational mobility network is different than that used by Mealy et al. (2018) to define the US's occupational mobility network (OMN). Statistics Sweden has complete data over all transitions for the entire work force, whereas Mealy et al. (2018) uses survey data on peoples occupations taken at different times to define their network. That data only describes the flow between occupations and not how many people stayed in an occupation (Mealy et al., 2018) (del Rio-Chanona et al., 2019). Therefore they approximate an aggregate share of employees who stayed in their occupation, which is the same for all occupations. Something that we do not have to do in Sweden's case. However, the US Current Population Survey (CPS) span several years, which, of course, is important, especially when dealing with labour flows which have considerable variance.

There are two categories of occupations that had no transitions to nor from other occupations, this is probably because only transitions which happened between 2016 and 2017 are in the data. These occupations are labelled Therapists in alternative medicine and Fishery workers and had 117 and 604 employees respectively in 2018. In addition, there is another occupation, labelled Square and market vendors, which had no incoming transitions from other occupations. In 2018, there were 190 employees within this occupation. Including these occupations in the analysis does not change the results to a large extent. However, the choice was made to remove the nodes which only has self-loops (the first two that were mentioned). This is because even though the occupations do not contain large numbers of workers, they are affected by the model. Each time step some of them are separated and, if vacancies exist, they will be filled. However, the workers are only able to apply to vacancies within their occupation. In addition, they are the only workers able to apply to those vacancies. This means that occupations with a high Computerisation Probability and no transitions to, such as Fishery workers that have a value of 72%, will have a large number of unemployed workers after the shock. It is unlikely that no Fishery workers would find a different job as a result of being let go due to automation, which the model would imply if the occupation was left in. But since the occupation labelled Square and market vendors has outgoing transitions, workers which are unemployed here are able to apply to vacancies in a different occupation (labelled Shop staff). This means that these workers are still able to enter the occupational mobility network. However, the post-shock demand may still be calculated for all occupations which is found together with other data in the Appendix 7.3.

[6] Standard för Svensk Yrkesklassificering

	Unemployment rate	Vacancy rate
Count	64	64
Mean	7.31	1.44
Std	0.77	0.51
Min	5.88	0.64
25%	6.74	1.06
50%	7.40	1.33
75%	7.89	1.89
Max	8.92	2.34

Table 2: *Summary statistics for Beveridge curve data*

4.3.2 Data Processing

As stated above, the computerisation probabilities derived by Frey and Osborne (2017) are given using the Standard Occupation Classification (SOC) system that is used to classify occupations in the US. As mentioned above, Sweden uses a different system for classification, SSYK. Therefore we need to translate this data from SOC to SSYK. This is a fairly common task and official crosswalks (translation keys) exists between SOC and the ISCO (International Standard Classification of Occupations). In addition, Statistics Sweden has released a crosswalk between ISCO and SSYK. These keys were used to move the Frey and Osborne (2017) computerisation probabilities from SOC to SSYK. The resulting network is illustrated in Figure 2. However, the data from Frey and Osborne (2017) is more fine grained (contains more occupations) than the labour flow data I have for Sweden. This is because the SOC codes that are the basis for the estimates produced by Frey and Osborne (2017) are equivalent to the 4 digit SSYK codes. Therefore, the shock was translated to 4 digit SSYK first, and then an average shock was calculated based on the 4 digit codes which share the first 3 digits. For example, the SSYK code 333 corresponds to the nurse occupation, but since there are many types of nurses (emergency nurses - SSYK: 2226, allergy nurses - SSYK: 2221, etc), each of which might have a different computerisation probability, the only way to proceed with the analysis is to somehow map all of the computerisation probabilities of different nurses into one value for the entire category. The most straightforward way of doing this is to take the average of all the computerisation probabilities - which is what is done here. A way to improve upon this would be to take an average weighted by the employment share of each type of nurse - however, that data is not publicly available.

5 Results

In this section the results of the replication of the data driven occupational mobility model described in section 3.3 are presented. In addition, the analysis is synthesised into two figures (7 and 8).

The model is calibrated using the Beveridge curve by implementing a simulated business cycle. The fluctuations are generated by oscillating aggregate target demand, D_t, around its initial value with a sine wave. The amplitude along with the other parameters, $\delta_u, \delta_\nu, \tau$, are calibrated to match the empirical Beveridge curve during the most recent Swedish business cycle.

5.1 Replication of Beveridge curve feature reproduction

The aim of the model calibration is for the Beveridge curve produced by the model given a set of parameters to match the empirical Beveridge curve. This is done by imposing a simulated aggregated demand shock on the model. The demand shock is given by a sine curve:

$$D_t = \sum_i d^\dagger_{i,0}(1 + a\sin(\frac{t}{2\pi T})) \tag{5.1}$$

Where D_t is the aggregate target demand and $d^\dagger_{i,t}$ the target demand in occupation i, both at time t. a is the amplitude of the demand shock as a share of the initial target demand and T is the period of the shock. In calibrating the model, the intersection and the union of the area enclosed by the empirical curve, A_e, and the simulated curve, A_m, are compared. The minimising function used is:

$$\min_{a,\delta_u,\delta_\nu,\gamma,\tau} 1 - \frac{A_e \cap A_m}{A_e \cup A_m} \tag{5.2}$$

The model is initialised with the occupational employment distribution in 2016 where the economy is assumed to be in equilibrium. This means that the initial target demand of each occupation is equal to the number of employed workers in said occupation in 2016. Then, the dynamics described in section 3.3 begins, with workers being separated and vacancies created according to the spontaneous processes characterised by the parameters δ_u and $delta_\nu$. Note that the state dependent processes, characterised by $\alpha_{u,i,t}$ and $\alpha_{\nu,i,t}$, do not affect the system at first since the target demand is equal to the realised demand by the equilibrium assumption. However, as workers begin to become separated and vacancies open, the system leaves exits the equilibrium and the state dependent processes begin to take effect. The system continues to evolve in this manner until a new steady state is reached. At this point, the model is not in equilibrium, rather the state dependent and spontaneous processes cancel out. Meaning that the aggregate number of hired workers is the same as the aggregate number of separated workers. During this time, the target demand of each occupation has remained at its initial level matching the employment distribution of 2016. When the model has reached this state, it makes little to no movement on the Beveridge curve. Until the simulated business cycle described above is introduced. Then the target demand of each occupation begin to move in tandem according to the sine function. This throws the system out of its steady state and it begins a cycle on the Beveridge curve. Since the sine function returns to its original point after a period, T, the system also tracks the same pattern over again as the initial period passes and the next cycle begins.

Two examples of what this looks like are presented in Figure 6. As is clear from the graphs - the calibration was not very successful. Before we go into the details of how the model behaves under different sets of parameters, the result of the replication with respect to shifts in the Beveridge curve is discussed. In del Rio-Chanona et al. (2019) are able to show that their model reproduces the three features of the Beveridge curve mentioned in section 3.4. They show that structural changes such as a decrease in the efficiency of worker-vacancy matching, cause the Beveridge curve to shift with respect to

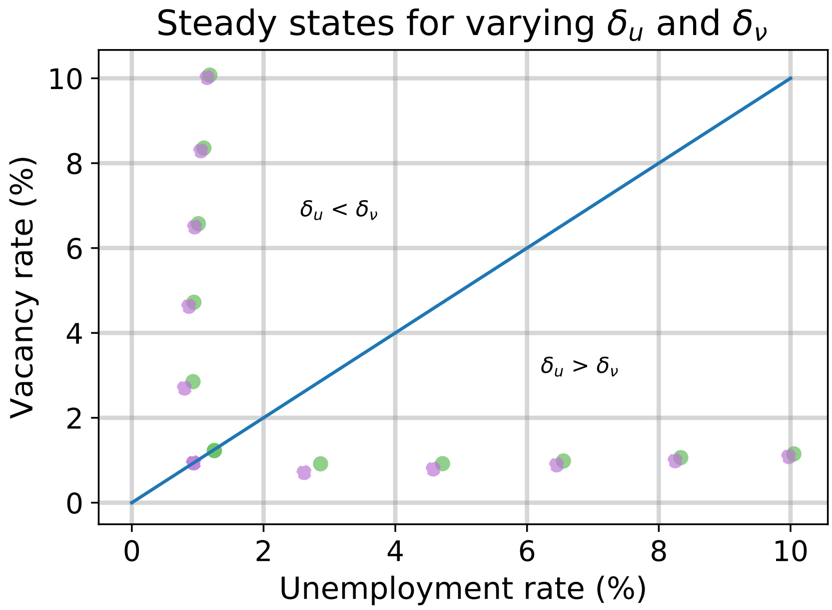

Figure 5: *Steady state of the model for different values of δ_u and δ_ν. Green dots come from the empirical network and purple dots come from the complete network.*

the origin by changing the topology of the occupational mobility network and keeping the demand constant. Specifically, they replace the empirical network with a complete one where $A_{ij} = \frac{1}{n}$, meaning that each occupation is linked to every other occupation with equal weight. Such a network corresponds to the null hypothesis of no skill restrictions, where every worker may apply to every occupation. The points that lie on the diagonal correspond to the case where $\delta_u = \delta_\nu = 0.006$, each step to the right of this correspond to an increase of δ_u by 0.001 and a decrease of δ_ν by 0.001. Similarly, each step to the left of the diagonal correspond to a decrease of δ_u by 0.001 and an increase of δ_ν by 0.001. As shown in the figure - removing skill restrictions shifts the steady state downwards towards the origin, but not by much. This behaviour is expected when removing matching frictions because vacancies are more likely to be applied for since any unemployed worker may apply for any vacancy. This shows the first feature mentioned in section 3.4 (del Rio-Chanona et al., 2019). In addition, the simulated Beveridge curve moves downward and out during periods of decreasing aggregate labour demand and upwards and in during periods of increasing aggregate labour demand (Figure 6). Furthermore, the model exhibits the counter clockwise behaviour described in section 3.4, moving upwards along the line in a recovery and downwards in a recession. Similarly to del Rio-Chanona et al. (2019), this feature is present when $\delta_u > \delta_\nu$ and the model exhibits clockwise motion in the opposite case. Thus the model is able to reproduce the dynamics of the Beveridge curve (del Rio-Chanona et al., 2019). The calibrated values of the model using Swedish data are substantially different than the calibrated values found by del Rio-Chanona et al. (2019) using US data, which is to be expected. However, the behaviour of the model given the truth value of $\delta_u > \delta_\nu$ is similar. But the similarity between the generated Beveridge curve and the empirical Beveridge curve is not ideal - the reasons for this are explained below and the consequence of this with respect to the results of the simulation are discussed in section 6.

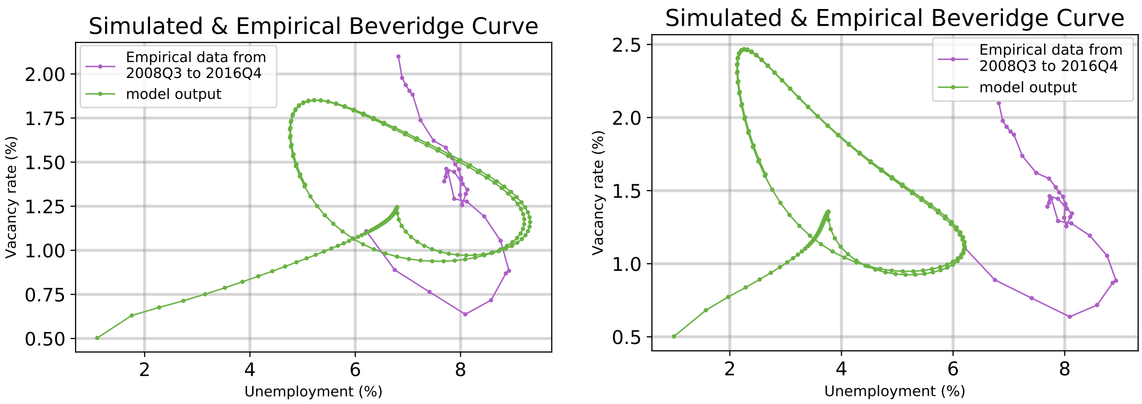

Figure 6: *Left figure shows the calibration which the results are based on and the right figure is an example of a set of parameters which give a good shape but a incorrect location.*

Parameter	Example	Sweden	US
δ_u	0.01	0.011	0.016
δ_ν	0.005	0.005	0.012
γ	0.2	0.1	0.16
τ	8	8	6.67

Table 3: *Calibrated values for the two figures using Swedish data, and the calibrated values of del Rio-Chanona et al. (2019)*

Because of the large parameter space that determine the relevant output of the model, the calibration of the model is not trivial. Initially, hundreds of sets of parameter values were drawn from normal distributions and the sets which yielded a curve in the right neighbourhood were used to calibrate the normal distributions for each parameter. However, there is to an extent a guiding theory, as each parameter controls a specific part of the simulation. We say that the output of the model calibration, measured in terms of the Beveridge curve, has a shape and a location. Each parameter affects the features of the output in a different way. As mentioned - we are dealing with a non-equilibrium model. The values of the parameters determine the steady state of the model del Rio-Chanona et al. (2019). If δ_u and δ_ν are both zero - the state dependent processes will never kick in since the model will remain at the initial equilibrium and therefore no adjustments are needed - resulting in 0 % unemployment rate and vacancy rate. However, the state dependent processes would kick in as soon as the realised demand would differ from the target demand. This means that δ_u and δ_ν do not control unemployment and vacancies independently, rather both of them affect both dimensions of the Beveridge curve.

This proved to be a persistent problem especially since the unemployment rate and the vacancy rate both depend on employment. Meaning that the vacancy rate increases as the spontaneous probability of separation, δ_u, increases. This is because the size of the work force is fixed - an increase in unemployment directly leads to a decrease in employment which is part of the denominator of the vacancy rate. Since Sweden had a vacancy rate close to 0.5% and an unemployment rate of about 8% during the second quarter of 2009, this means that a good balance between δ_u and δ_ν is hard to achieve. I was able to find a good shape that fit the empirical Beveridge curve (Figure 6, but at the wrong location and, conversely, a good location but with the wrong shape. This is because the curve moves upwards and away from the origin for increasing δ_u. Decreasing δ_ν in response to this movement makes the generated curve lose its desired shape since the recovery after the financial crisis is steep. In addition, too large values for γ (which would create more vacancies in the recovery) means that vacancies are created too quickly and unemployment decreases by too much. However, given enough data data from the model, it is possible that a machine learning algorithm trained on this data would be able to find better results. This process was started with the use of an evolutionary algorithm, which helped to find the curves presented here.

5.2 Unemployment under automation shock

Despite challenges during the calibration process, results of the kind presented in del Rio-Chanona et al. (2019) are found below. The figures are based on 5 agent based simulations for the occupational mobility network and a deterministic simulation of the complete network[7] are presented. The complete network corresponds to the null case of no skill frictions. For both cases the model starts at the unemployment level matching that of 2016, around 6.6%. However, similarly to del Rio-Chanona et al. (2019) we find that the network structure is important in determining the unemployment outcomes, where the structure of the occupational network is less efficient in dealing with the reallocated demand due to the automation shock (Figure 8 **B)**. In addition, the change in target demand as a result of the automation shock are presented for two occupations which we discussed in sections 1 and 3.1.

The peak unemployment for the occupational mobility network is about 7.5% which is more than 0.5% higher than that of the complete network. The peak is reached after a little more than half of the shock duration has transpired, where it then starts to decay. In addition, the peak is about 0.9% higher than the pre-shock unemployment rate (which corresponds to the level in 2016) for the empirical network. In the complete network, the shock results in an increase of about 0.5%, which means that the change in unemployment due to the shock is almost twice as large for the empirical network. We again focus on the changes from the initial level (0.9% and 0.5%) and compare with the effect of the shock in the US from del Rio-Chanona et al. (2019). There, the unemployment rate increased from the initial level of about 5.3% to the peak of 6.7% before it decays, meaning that the US empirical network is more affected by the shock than the Swedish network. This also holds for the complete network, which in the US increased from about 4.1% to 4.7%.

This means that for the US case, the change in unemployment rate for the empirical network is also about twice as large as that of the complete network. This might be a coincidence but it would be interesting to analyse the differences between the Swedish occupational mobility network and that of the US and see if it is possible to predict this. There are only two differences between the empirical networks and the complete ones, the number of edges (which is n^2 for the complete network where n is the number of nodes) and the edge weights. It is likely that the difference in unemployment rate change due to the shock between the empirical networks and the complete networks is related to the number of edges being added, although testing this is outside the scope of the thesis.

[7] Agent based simulations for the complete network are very computationally costly, which means that the occupational mobility network results were prioritised

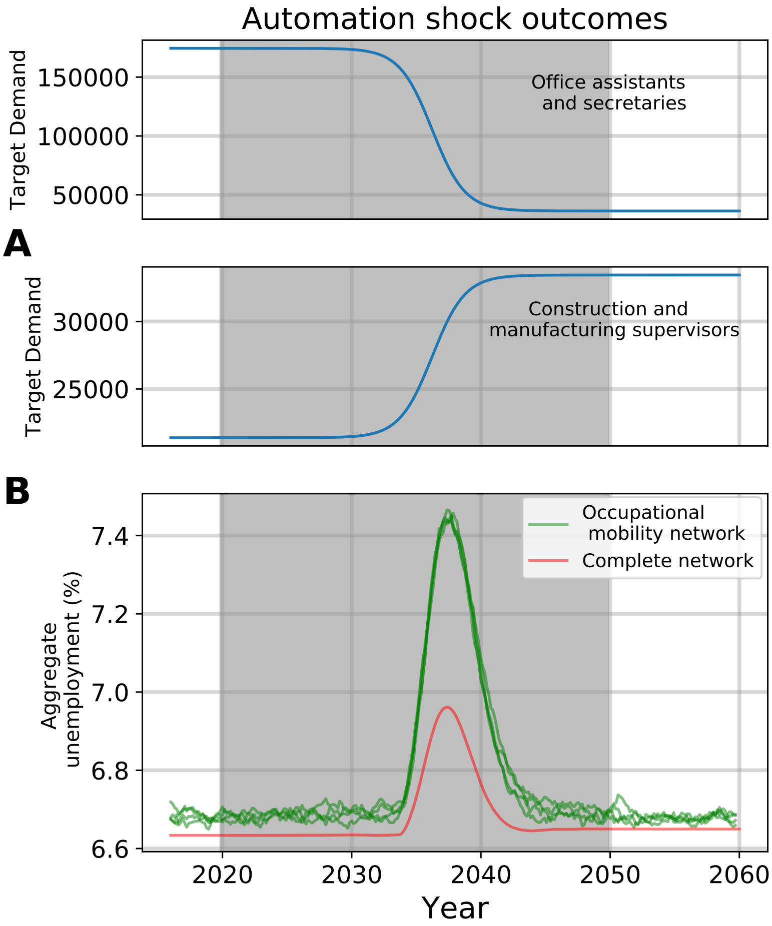

Figure 7: *Outcomes of the simulation shock for deterministic solution and agent-based simulation, over time. The shaded area corresponds to the shock period.* ***A****: Changes in target demand for two occupations, Office assistants and secretaries as well as Construction and manufacturing supervisors.* ***B****: Aggregate unemployment over time as a result of the shock - where dotted lines are deterministic results and full lines are agent-based simulation results.*

The next, and final, result is unemployment change for individual occupations as a result of the automation shock. Following del Rio-Chanona et al. (2019) unemployment rates for individual occupations are calculated as the average unemployment rate over the period:

$$u_{i,\text{ average}}(T) = \frac{100}{T} \frac{\sum_{t \in T} u_{i,t}}{\sum_{t \in T}(u_{i,t} + e_{i,t})} \tag{5.3}$$

For each occupation, the unemployment change is compared to the computerisation probability, where the size of the dot is proportional to the average employment of the occupation during the shock (Figure 8). Results for both the empirical network and the complete network are presented. A key aspect of the figure is that the only thing affecting the unemployment change of the occupations in the complete network (purple dots) is their computerisation probability. Which is demonstrated by the almost[8] perfect relation-

[8]Some of the purple dots are above the 'line' that the others are on, which should not be the case. This

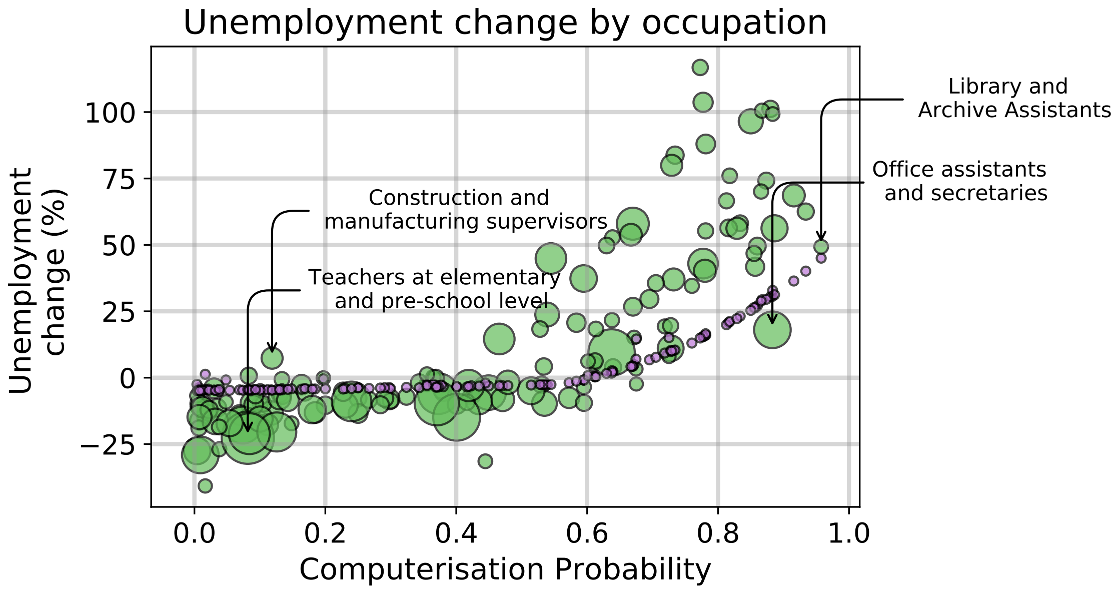

Figure 8: *Change in average unemployment due to automation shock. Each point is an occupation where the green ones are based on the occupational mobility network and purple ones based on the complete network. For the occupational mobility network, the size of the point is proportional to average employment during simulation.*

ship between unemployment change and computerisation probability. This is because the occupations in the complete network is unaffected by network effects since every node is connected to every other node. Therefore all the deviations from the purple 'line' are due to the structure of the occupational mobility network. This means that, as have been mentioned previously, some occupations are more adversely affected by the automation shock than that expected by their computerisation probability, for example the Construction and manufacturing supervisors. In addition, other occupations are less affected by the automation shock than that predicted by the computerisation probability, for example the Office assistants and secretaries. These results are completely in line with those of del Rio-Chanona et al. (2019). In addition, the effects are large, however, as is further discussed in section 6, we should be careful in interpreting these results. Partly due to the state of the calibration and partly due to idealised nature of the model. Nevertheless, the change in unemployment rate presented here are of the same order of magnitude as those presented in del Rio-Chanona et al. (2019), where the minimum is about -25, and the maximum over 100. Which is the same as the results presented here. Available results for all occupations are presented in the Appendix 7.3

6 Discussion

There are a number of issues which hinders certain conclusions to be drawn. However, most of the results from del Rio-Chanona et al. (2019) were successfully replicated using Swedish data, but it remains to be seen if the model, as it is currently formulated, can be satisfactorily calibrated using Swedish data. The model, implemented on top of Sweden's occupational mobility network, behaves in the manner described by del Rio-Chanona et al. (2019) and exhibit similar features with respect to the parameter values of δ_u and δ_ν.

is probably because the complete network data is from only one simulation.

But since the model calibration was relatively unsuccessful, it is hard say if this provides evidence for or against the theory that the counter cyclical nature of the Beveridge curve is due to business cycle dynamics. To this end, we refrain from drawing any particular conclusion with respect to the Beveridge curve debate.

As expected based on the differences in the empirical Beveridge curves of Sweden and the US, the calibrated values for δ_u and δ_ν are much smaller in the case of Sweden. This means that the Swedish labour market is less dynamic than the that of the US, and vacancies are created and workers are separated to a lower extent. This is likely due to the rigidity associated with employing people in Sweden.

The second conclusion which is hindered by the calibration, is that of the magnitudes of the demonstrated effects. Below it is argued that the main effect presented here still holds, despite the calibration. However, it is certainly not argued that the effect is of the magnitude presented here. In addition, at various points, certain assumptions have been made, the implications of which are discussed below.

A strong assumption is that workers are perfectly geographically mobile - which they most certainly are not in a country with low population density such as Sweden. This is important since, currently, the only thing that hinders workers from applying to vacancies are their position in the network (in which occupation they are employed). To address this the transition probabilities between occupations should incorporate some spatial element. This would require workers and vacancies to have a geographical location and the probability that a worker applies to a vacancy should be inversely proportional to this distance. It is possible that the data required to augment the analysis in this way is available from the Swedish Public Employment Service. If workers are perfectly geographically mobile this means that a friction in the network is omitted. If workers would be unable or unwilling to apply to vacancies far away, less occupations would be applied for and more workers would be unemployed for longer. Omitting this friction therefore skews the aggregate unemployment in Figure 7 **B** downwards.

Another assumption is that wage pressure is neglected - this means that workers do not factor in potential wage differences between occupations when choosing which vacancy to apply for. It is possible to implement wages into a model of the kind used here, as is done by Axtell et al. (2019). This is another example of an omitted friction in the model, since workers might choose to stay unemployed rather than apply for a vacancy in a lower paying occupation. Again, omitting this friction skews the aggregate unemployment in Figure 7 **B** downwards. In addition, the automation shock implemented here decreases the amount of hours worked by every worker while keeping the aggregate demand constant. This is done since amount of hours worked has a historical decreasing trend whereas unemployment does not have a trend. However, Boppart and Krusell (2020) shows that falling hours worked are due to stronger income effects on labour/leisure choice and not solely automation. Since there are no worker nor firm optimisation behaviour in the model - this fact can not be accounted for. If wages were present in the model - workers would be able to optimise their labour/leisure decisions based on economic theory. Furthermore, the set of jobs is constant and automation is completely exogenous. This makes the model descriptive in the sense that it makes predictions not through firm nor worker optimisation but rather through pure empirical observations. It can never answer the question of why there is automation - only estimate the effects of it.

This not a criticism of the model, rather, there are some benefits because of this. Since the shock is exogenous, any type of occupation specific labour demand shock may be implemented. For example, given estimates of how occupations are affected by the current CoVid-19 pandemic, it is possible to estimate the future unemployment outcomes as a result of this. Which could be useful in analysing the effect of continued lock-down. This shock would have to be a bit different, since the changes in labour demand across occupations would be more or less temporary - rather than permanent. This result in particular, could be useful when choosing who to target with job-retraining programmes.

The main result from del Rio-Chanona et al. (2019) is that certain occupations with high probability of computerisation are more or less unaffected by the shock - while other occupations with low probability are more affected than expected. As we have seen throughout the paper, this result is found to be true in Sweden's case as well. Despite the aforementioned issues, we argue here that this result still holds. This is because we do not observe the effect in the null case of the complete network - where changes in unemployment for individual occupations are only due to the computerisation probability of that occupation. This means that the structure of the occupational mobility network affects labour market outcomes. However, the problems mentioned above certainly make the magnitude of the presented results dubious. Furthermore, these results are not predictions of what the observed levels of unemployment will look like in the coming decades. This is because of the large number of variables that affect unemployment, for example a global pandemic. The model presented here gives us an estimate of how one of these variables,[9] automation, may affect occupational employment in a, seemingly, counter-intuitive way.

References

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. *Science advances*, *4*(7), eaao6030.
2.1

Axtell, R. L., Guerrero, O. A., & López, E. (2019). Frictional unemployment on labor flow networks. *Journal of Economic Behavior & Organization*, *160*, 184–201.
1, 2.1, 3.2, 3.4, 6

Beveridge, W. H. (2014). *Full employment in a free society (works of william h. beveridge): A report*. Routledge.
3.4

Boppart, T., & Krusell, P. (2020). Labor supply in the past, present, and future: a balanced-growth perspective. *Journal of Political Economy*, *128*(1), 000–000.
6

Bouvet, F. (2012). The beveridge curve in europe: new evidence using national and regional data. *Applied Economics*, *44*(27), 3585–3604.
3.4

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? In *Aea papers and proceedings* (Vol. 108, pp. 43–47).

[9] which itself consist of many variables

1, 2.2

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (Tech. Rep.). National Bureau of Economic Research.
1, 2.2

del Rio-Chanona, R. M., Mealy, P., Beguerisse-Díaz, M., Lafond, F., & Farmer, J. D. (2019). Automation and occupational mobility: A data-driven network model. *arXiv preprint arXiv:1906.04086*.
(document), 1, 2.1, 2.2, 3, 3.2, 3.2, 3.3, 3.3, 3.3.1, 3.3.1, 3.4, 3.5, 4.1, 4.1, 4.2, 4.2, 4.3.1, 5.1, 3, 5.1, 5.2, 5.2, 5.2, 6

Diamond, P. A. (1982). Aggregate demand management in search equilibrium. *Journal of political Economy*, *90*(5), 881–894.
3.4, 3.4

Diamond, P. A., & Şahin, A. (2015). Shifts in the beveridge curve. *Research in Economics*, *69*(1), 18–25.
3.4, 3.4

Diodato, D., & Weterings, A. B. (2015). The resilience of regional labour markets to economic shocks: Exploring the role of interactions among firms and workers. *Journal of Economic Geography*, *15*(4), 723–742.
2.1

Dworkin, J. D. (2019). Network-driven differences in mobility and optimal transitions among automatable jobs. *Royal Society open science*, *6*(7), 182124.
2.1

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, *114*, 254–280.
1, 2.1, 2.2, 4.1, 4.2, 4.3, 4.3.2

Goudet, O., Kant, J.-D., & Ballot, G. (2017). Worksim: A calibrated agent-based model of the labor market accounting for workers' stocks and gross flows. *Computational Economics*, *50*(1), 21–68.
2.1

Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347–367). Elsevier.
2.1

Guerrero, O. A., & Axtell, R. L. (2013). Employment growth through labor flow networks. *PloS one*, *8*(5).
2.1

Jackson, M. O., & Kanik, Z. (2019). How automation that substitutes for labor affects production networks, growth, and income inequality. *Growth, and Income Inequality (September 19, 2019)*.
2.1

Jonsson, M., & Theobald, E. (2019). A changed labour market–effects on prices and wages, the phillips curve and the beveridge curve. *Sveriges Riksbank Economic Review*, *1*.
3.5

Lopez, E., Guerrero, O. A., & Axtell, R. (2015). The network picture of labor flow. *Available at SSRN 2626179*.
2.1

Mealy, P., del Rio-Chanona, R. M., & Farmer, J. D. (2018). What you do at work matters: New lenses on labour. *Available at SSRN: https://ssrn.com/abstract=3143064 or http://dx.doi.org/10.2139/ssrn.3143064*.

1, 2.1, 3.1, 3.1, 4.3.1
Michael, N., & Richiardi, M. G. (2018). Agent based models of the labor market. *LABORatorio R. Revelli working papers series, 125.*.
2.1
Mortensen, D. T. (1999). Equilibrium unemployment dynamics. *International Economic Review*, *40*(4), 889–914.
3.4
Neffke, F. M., Otto, A., & Weyh, A. (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization*, *142*, 275–292.
2.1
Nimczik, J. S. (2017). Job mobility networks and endogenous labor markets. Kiel, Hamburg: ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft. Retrieved from http://hdl.handle.net/10419/168147
2.1
Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *The American Economic Review*, *75*(4), 676–690.
3.4
Pissarides, C. A. (2011). Equilibrium in the labor market with search frictions. *American Economic Review*, *101*(4), 1092–1105.
3.4
Schmutte, I. M. (2014). Free to move? a network analytic approach for learning the limits to job mobility. *Labour Economics*, *29*, 49–61.
2.1
Stoneman, P. (2001). *The economics of technological diffusion.* Wiley-Blackwell.
4.2

7 Appendix

7.1 Network term glossary

Adjacency matrix - A matrix having the same set labels in its columns and rows. It contains the data that is used to construct the graph (or network). Often denoted A.
Node - Each node is defined by the column and row labels in the adjacency matrix.
Edge - Edges between nodes are defined by the cell values, A_{ij}, in the adjacency matrix. For undirected graphs $A_{ij} = A_{ji}$ meaning that A is symmetric in the diagonal. For directed graphs this does not have to be the case, meaning that the edge from i to j (A_{ij}), is not necessarily the same as the edge from j to i (A_{ji}).
Neighbour - A node is a neighbour to another node if there exists an edge between them.
Degree - The number of edges connected to a node. For undirected graphs: degree = in-degree = out-degree.
In-degree - The number of edges going to a node (for directed graphs only).
Out-degree - The number of edges going from a node (for directed graphs only).
Density - The number of actual edges in the graph divided by the number of possible edges in the graph.
Topology/structure - The set of nodes and edges that are present in the graph. Defined by the adjacency matrix.
Complete network - A specific network topology where all possible edges between nodes exist. Always has density 1. May be directed or undirected.

7.2 The model in python code

As mentioned in section 4 the model is coded in Python. All of the data, the package and the jupyter notebooks which compile different parts of the results may be found here. Most of the code is commented, where more attention is given to the functions that are used to execute the model. In order to run everything a few more packages, in addition to python itself, are needed. These are:

- pandas - A data handling package. One of the most well-known and widely used data science packages in Python.
- numpy - A data handling and processing package which implements features from MatLab into Python. In addition, it is optimised meaning that it data handling is fast.
- networkx - A data handling, processing and illustration package for networks. Comes with a large amount of useful built in functions for handling networks.
- matplotlib - A data illustration package which implements graphing tools from matlab into Python. Allows for highly customisable graphs.
- DEAP - Distributed Evolutionary Algorithms in Python. This package was used for the calibration and implements evolutionary algorithms.
- Shapely - Used for calculating union and intersection of curve areas in the calibration.

In addition, jupyter notebook is required to run the notebooks. To install packages and jupyter notebook I would recommend using Pip. Once all the packages are installed, it should be straightforward to run the code from the notebooks.

7.3 Occupational data and results

Below is a table with individual results for occupations.

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
171	Hotel and conference managers	0.00390	1471	2661	4.84	-6.80
221	Medical Doctors	0.00420	38873	72700	7.29	-27.54
152	Managers of social and curative work	0.00700	4812	8973	5.72	-18.95
224	Psychologists and psychotherapists	0.00700	8588	16082	5.11	-10.84
151	Healthcare managers	0.00730	11718	22159	5.71	-16.14
141	Managers in elementary, high-school and adult education	0.00730	10198	19066	5.32	-9.70

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
233	High school teacher	0.00780	29858	53149	6.01	-14.78
222	Nurses	0.00900	83956	147208	8.48	-29.06
227	Naprapaths, physiotherapists, occupational therapists and other	0.01244	19896	35686	5.31	-11.06
142	Managers in preschool activities	0.01500	4463	8304	5.51	-14.26
154	Managers and leaders of faith communities	0.01655	708	1063	3.77	-40.75
133	R&D managers	0.01750	6232	10997	5.37	-10.89
226	Dentists	0.02150	6081	11049	5.08	-13.26
228	Other specialists within healthcare	0.02235	9730	18480	5.42	-12.03
137	Production and manufacturing managers	0.03000	16575	29162	5.68	-4.01
231	Teachers in higher education	0.03200	35359	62343	6.26	-16.45
131	Information technology managers	0.03500	10801	19648	5.72	-11.32
267	Priests and deacons	0.03789	3795	6363	5.36	-18.55
225	Veterinarians	0.03800	2308	4311	5.14	-26.94
149	Other managers within education	0.04820	1094	2085	4.71	-9.15
266	Social secretaries and curators	0.05375	38285	77525	6.60	-17.09
121	Chief Financial Officers	0.06900	16864	30110	6.08	-14.07
214	Advanced engineering occupations	0.07375	88418	160228	9.66	-17.84
234	Teachers at elementary and pre-school level	0.08153	189597	312531	10.48	-22.59
172	Restaurant and kitchen managers	0.08300	7460	12657	4.92	0.72
153	Managers within elderly care	0.08365	9716	15921	5.54	-14.73
251	IT-architects, system developers and others	0.08389	111725	209359	9.84	-21.07

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
112	Assorted Chief Officers	0.08750	24598	35305	5.63	-9.62
217	Designers	0.09064	14227	24732	5.56	-10.19
344	Traffic teachers and instructors	0.09331	6211	10148	5.20	-6.70
235	Other teachers with theoretical, special competence	0.10025	35506	58538	6.56	-15.57
111	Politicians and higher officials	0.10050	3889	4813	5.10	-9.75
161	Managers in banking, finance and insurance	0.10540	6203	8833	5.29	-9.85
223	Other nurses	0.11167	21623	34603	6.05	-17.65
125	Sales and marketing managers	0.11700	29866	50433	6.62	-13.33
312	Construction and manufacturing supervisors	0.11867	21382	39234	6.19	7.35
242	Organisational developers and HR-specialists	0.12581	103534	175997	9.52	-20.42
341	Treatment assistants and pastors	0.13000	25709	32620	5.68	-7.72
213	Biologists, pharmacologists and specialists within agriculture and forestry	0.13052	6585	10823	5.08	-6.34
124	Information, communication and PR managers	0.13371	3789	6763	4.85	-0.64
232	Teacher in occupational subjects	0.13440	10093	15539	5.22	-6.64
342	Athletes and leisure coaches	0.14278	23871	41081	5.84	-8.45
212	Mathematicians, actuaries and statisticians	0.14840	1929	3434	5.83	-17.18
136	Construction, plant and mine operational managers	0.16367	17035	27194	5.94	-2.63

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
179	Other service managers	0.16875	7020	10639	5.22	-5.63
351	Operation, support and network technicians	0.18000	46483	67954	7.03	-12.00
261	Lawyers	0.18429	21037	32702	5.95	-13.12
174	Managers within wellness, sport and leisure	0.19700	1304	2128	4.66	-0.13
173	Managers within trade	0.20000	11092	14938	5.62	-10.39
264	Authors, journalists and translators	0.22725	14173	20159	5.17	-5.30
243	Marketing and public relations	0.23060	37141	53258	6.48	-10.38
531	Childcare and student assistants	0.24000	115195	170351	8.20	-8.92
159	Other managers within societal services	0.25000	16208	22941	6.01	-13.43
134	Architecture and engineering managers	0.25000	9858	15352	5.65	-7.70
218	Specialists within environmental- and health protection	0.26678	7643	11285	5.43	-8.32
262	Museum superintendents and librarians	0.28411	10496	13819	5.42	-10.18
216	Architects and surveyors	0.29060	11405	15104	5.53	-7.72
129	Other administration and service managers	0.29765	20973	29789	6.02	-8.09
211	Physicists, chemists and similar	0.29964	6301	7939	5.23	-8.73
122	Personnel and HR officers	0.32393	7876	10604	5.45	-7.17
514	Beauty and body therapists	0.34342	9692	12382	5.03	-1.85
135	Real estate and administration managers	0.35500	3742	4588	4.91	-2.38

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
516	Other service personnel	0.35526	3309	4528	4.99	1.22
345	Chefs and sous chefs	0.36500	3647	4523	4.95	0.19
511	Cabin crew, train personnel and guides	0.36888	8050	9801	5.07	-0.02
534	Caretakers, carers and personal assistants	0.37022	161192	180960	8.71	-9.62
332	Insurance advisers, business sales and purchasing	0.37277	134197	152627	9.94	-6.01
315	Pilots, ship and machine officers	0.37713	5295	5879	5.10	-5.03
265	Artists, musicians and actors	0.38005	9150	10875	5.17	-5.70
532	Assistant nurses	0.40056	186915	193910	9.73	-14.89
336	Police officers	0.40097	15947	17274	5.58	-10.12
343	Photographers, decorators and entertainment artists	0.41637	7466	7330	5.64	-9.20
533	Nursing assistants	0.41900	76546	78746	6.57	-2.94
123	Administration and planning managers	0.42312	10816	11075	5.58	-6.87
335	Tax and social security officers	0.43500	47554	43944	6.66	-8.91
324	Animal nurses and others	0.44450	2127	2716	4.75	-31.46
311	Engineers and technicians	0.44948	98649	104777	9.46	-5.63
911	Cleaners and home service staff	0.46571	75409	74187	6.22	14.58
333	Intermediares and others	0.47090	36325	30075	6.41	-8.22
541	Other security proffessions	0.47885	37697	35662	6.14	-1.73
331	Bankers and accountants	0.51447	56687	49669	7.28	-4.91
933	Harbor workers and ramp staff	0.52800	7970	8369	5.30	18.32
132	Purchasing, logistics and transport managers	0.53031	10954	9506	5.64	-2.63

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
742	Electronics repairers and communication electricians	0.53373	10613	8792	5.23	4.22
241	Accountants, financial analysts and fund managers	0.53477	46540	41951	7.24	-9.68
741	Installation and industrial electricians	0.53894	41852	35680	6.32	23.77
833	Truck and bus driver	0.54467	80359	65453	7.39	44.87
321	Biomedical analysts, dental technicians and laboratory engineers	0.57239	27193	20502	5.90	-7.57
832	Car, motorcycle and bicycle driver	0.58362	18638	13851	5.30	20.70
723	Vehicle mechanics and repairers	0.59464	58964	44638	6.98	37.35
535	Dental nurses	0.59500	10308	7521	5.55	-9.48
325	Dental hygienists	0.59500	3773	2741	4.94	-3.48
753	Tailors, wallpapers and leather craftsmen	0.60135	2466	1675	4.71	6.17
352	Image, sound and lighting technicians	0.61083	3891	2988	4.92	5.20
612	Animal breeders and caretakers	0.61200	8068	5804	5.13	6.28
912	Washers, window cleaners and other cleaning workers	0.61333	5638	4858	5.39	18.33
961	Recycling workers	0.62967	8875	5889	4.96	49.77
522	Shop staff	0.63730	223919	145078	10.06	9.50
731	Fine mechanics and craftsmen	0.63787	4108	2632	4.99	21.65
831	Locomotive driver and yard staff	0.63867	5635	3587	4.88	52.74
334	Legal secretary, chief secretary and department secretary	0.63950	11529	11221	5.61	3.90
712	Ceiling fittings, flooring and plumbing workers	0.66692	35869	22332	6.20	53.82

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
711	Carpenter, mason and construction workers	0.66976	100639	63380	8.74	57.97
611	Plant growers in agriculture and gardening	0.67000	17675	10718	5.25	26.72
421	Croupiers and debt collectors	0.67167	1736	1366	4.82	15.17
443	Elected official	0.67457	869	553	5.26	3.29
138	Forestry and agriculture managers	0.67513	607	435	4.44	-2.36
819	Operating technicians and process supervisors	0.69513	19093	10528	5.66	29.62
932	Hand packers and other factory workers	0.70500	11184	5886	5.05	35.60
621	Forestry workers	0.71857	3590	1812	4.96	19.31
835	Sailors and deckhands	0.72500	1087	679	5.12	8.00
732	Prepress technicians, printers and bookbinders	0.72750	8179	3708	5.21	19.57
422	Travel agents, customer service personell and receptionists	0.72760	62331	31356	7.09	11.08
834	Machine drivers	0.72908	34912	17949	6.03	79.93
512	Cooks and cold bar attendants	0.73200	39641	19701	5.70	36.99
817	Process operators, wood and paper industry	0.73429	16605	7890	5.38	83.77
613	Plant breeders and animal breeders, mixed operation	0.76000	3946	1563	5.03	34.58
811	The ore processing profession and well drillers	0.77273	8934	3430	5.37	116.79
432	Warehouse workers and transport managers	0.77711	93679	40162	7.70	42.96

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
721	Casters, welders and sheet metal blowers	0.77714	26927	10641	5.61	103.67
515	Cleaning managers and property managers	0.78000	42803	17710	6.10	40.25
761	Butchers, bakers and food processors	0.78100	8315	3284	4.89	55.26
713	Painters, lacquers and chimney sweeps	0.78125	24293	10162	5.47	87.98
818	Other process and machine operators	0.81285	8333	2573	5.52	66.57
816	Machine operators, food industry	0.81600	15692	5170	5.07	56.39
931	Rough workers in construction and civil engineering	0.81778	6510	2374	5.62	76.00
962	Newspaper distributors, custodians and other service workers	0.82885	39224	12012	5.79	56.24
752	Surface finishers, wood and furniture carpenters	0.83375	11180	3290	5.58	58.15
722	Blacksmiths and toolmakers	0.85000	59479	13830	6.66	96.51
524	Event sales and telemarketing	0.85500	9241	2213	5.68	46.78
513	Head waitors, waitors and bartenders	0.85667	23619	6649	5.56	41.77
442	Mail and post workers	0.86000	16842	4228	5.48	49.56
815	Machine operators, textile, laundry and leather industry	0.86571	5287	1290	5.20	70.09
813	Machine operators, chemical and pharmaceutical products	0.86714	5179	1239	5.16	100.52
814	Machine operators, rubber, plastic and stationery industry	0.87368	13379	3170	5.43	74.18

Continued on next page

SSYK	Description	Computerisation Probability	Employment 2016	Post-shock demand	Unemployment 2016 (%)	Unemployment change (%)
812	Process and machine operators, at steel and metal plants	0.88000	14526	3726	5.37	101.20
411	Office assistants and secretaries	0.88297	174494	35557	10.59	18.02
921	Berry pickers and planters	0.88333	3140	648	4.33	99.26
941	Fast food staff, kitchen and restaurant assistants	0.88625	77340	15887	6.00	56.25
821	Installers	0.91571	49282	9216	6.87	68.62
523	Cashiers and others	0.93400	15101	1562	5.68	62.54
441	Library and Archive Assistants	0.95750	3463	249	4.97	49.29