NORTHWESTERN UNIVERSITY

Social Interactions and Labor Market Outcomes of War Veterans

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Ron A. Laschever

EVANSTON, ILLINOIS

June 2007

ABSTRACT

Social Interactions and Labor Market Outcomes of War Veterans

Ron A. Laschever

Social networks play an important role in the labor market. Various surveys document that 30-60% of jobs are found through friends or relatives. To better understand how networks operate in the labor market, I examine how networks that were formed involuntarily as a result of the American Civil War and the First World War draft affect the postwar labor market outcomes of veterans in 1880, 1900, and 1930.

My study uses two data sets. The first, contains new data on 1,295 drafted American Infantrymen who served together overseas during World War I, and was formed by matching military service records, prewar draft records, and postwar information from the 1930 Census, as well as information on up to sixty of a veteran's nearest neighbors in 1930. The second, collected by Fogel et al. (2000), matches 35,570 Civil War veterans to postwar censuses. I exploit the time-series feature of the Union army sample and eliminate all unobserved individual and group-level fixed effects.

For both samples, the military unit's overall unemployment rate has a negative and statistically significant effect on a veteran's own likelihood of employment. The findings

are consistent with a model in which information about job vacancies is communicated through the network. Both samples are fairly representative of the white working-age male population, therefore contributing to the external validity of the results.

I introduce a new framework which allows one to further decompose the social effect into its two components, the endogenous ("the effect of others' outcomes"), and the contextual ("the effect of others' characteristics"). I show that the two effects are separately identified, provided that some people belong to more than one group. I apply the framework using two types of reference groups for each veteran, those who had served in his unit and his neighbors. I find the endogenous effect to be much stronger than the contextual effect, indicating the presence of a large social multiplier: a change in an individual's employment propagates through the network and affects the employment of others. The framework is also applicable in other settings, since in many cases individuals are potentially affected by multiple types of reference groups.

Acknowledgements

This work and my studies at Northwestern have benefited from the help and support of numerous, perhaps countless, people.

First, and foremost, I am grateful to my committee members, Joseph Ferrie, Dale Mortensen, and Christopher Taber for their guidance. I am indebted to my advisor Dale Mortensen for all his help. During the early and crucial stages, as I was facing the transition from student to researcher, Dale was always willing to meet, discuss my half-baked ideas, and comment on every draft of my work. His endless patience and his belief that my proposed data collection effort was feasible have been vital. I am grateful to Joe Ferrie for all his suggestions and support. His past and current data collection efforts have inspired me to undertake my own data collection endeavor. I have benefited greatly from Joe's remarks as well as his encouraging outlook. Last, and not least, I would like to thank Chris Taber for his invaluable help. Chris was always willing to meet and discuss my progress to-date. I have benefited immensely from our conversations. I thank Chris for never willing to settle for a less-than-satisfactory answer and always challenging me to come up with better solutions.

I am also indebted to two other members of the faculty, Burt Weisbrod and Elie Tamer. Burt Weisbrod provided me with a job for three years as his research assistant, an opportunity from which I have learned a lot about research. In addition, I was fortunate that Burt had shared with me his wisdom and insights about both life and research. I will always cherish our conversations and the great advice he gave me on various occasions. I thank Elie Tamer for all his support and helpful comments. Elie's encouragement and attitude were extremely valuable at an important juncture.

Many Northwestern faculty have provided me with valuable comments and I thank them for their time, patience and assistance. They include Raquel Bernal, Greg Duncan, Luojia Hu, Hilarie Lieb, Charles Manski, Rosa Matzkin, Bruce Meyer, Joel Mokyr, and Martin Zelder. I have also benefited from the help and comments of Ran Abramitzky, Michael Coates, Ann Janda, Francesca Molinari, and Eugene Orlov. I thank Lindsay Larsen and Karen Muth for dedicated and excellent assistance with the data collection effort. I thank Evelyn Asch for her encouragement and her willingness to carefully read several earlier drafts of my work.

I gratefully acknowledge financial support through Northwestern University's Graduate School Research Grant and the Center for Population Economics at the University of Chicago Graduate School of Business (subgrant NIH P01AG10120).

During my years at Northwestern and Evanston I met many wonderful people. I thank Susan Lee, Mas, the Joneses, Adam, Danny, Dimitar, Etienne, Eugene, Hanna, Itai, Maria, Ofer, Ran and the rest of my classmates, for their support and friendship. I also thank my out-of-town friends Assif, Claire, Dana, Greg, Hagai, Iddo, Marc, and Oded for their patience and emotional support during my graduate school journey.

This work was completed in Champaign, Illinois. I thank all the people I have met here for welcoming me and helping me acclimate. I would especially like to thank Craig, Darren, George, Hyun, Rich, Roger, Vered, and Yair.

I have been fortunate to have a supporting and loving family that has always stood by me. My late grandparents Karl, Eva, and Norman, my grandmother Ruth, my uncles and aunts, Gadi, Batia, Yoram, Avital, Rick, Bobby, Margo, John, Cathy, and John have been a source of inspiration and support. I hope this work will answer their question regarding when I intended to graduate. I thank my brother Dan for always willing to lend a helping hand. First and last, I thank my parents, Mark and Miriam, for all they have done for me.

To Mark and Miriam

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

During the past couple of decades, there has been a growing interest in social networks among economists. Economists have examined the effect of social interactions in a wide range of areas.[1] In the labor market, various surveys have documented the importance of the "informal" channel, that is finding jobs through friends and relatives. Ioannides and Loury (2004) summarize a number of surveys which find that 30-60 percent of jobs (in various industries and of various statuses) are found through the "informal" channel.[2]

This dissertation has twin goals. First, I introduce and illustrate an application of a new methodology for decomposing the social effect into its two components, the *endogenous* and the *contextual* effects. The second part focuses on the role social networks play in the labor market. The empirical part of the dissertation seeks to better understand how social networks affect labor market outcomes by examining groups that were formed involuntarily due to a quasi-random event. I examine the postwar outcomes of two groups of veterans, Civil War veterans and World War I veterans. I also make use of a new panel dataset I constructed of American war veterans who were drafted and served together during World War I.

[1] These range from theoretical work on network formation and network games (see Jackson (2004) for a survey of networks games) to the measurement of peer effects in such settings as welfare take-up (Bertrand et al., 2000), drug use among college students (Duncan et al., 2003), recidivism (Bayer et al., 2004), etc.
[2] The importance of the "informal" market has been documented as early as 1932. For example, De Schweinitz (1932) finds that 45% of workers in the hosiery industry in Philadelphia obtained their job through friends and relatives.

In Chapter 2, I introduce a new framework, Multiple Reference Groups, which allows one to separately identify the two components of the social effect. The components are commonly referred to in the literature as the *contextual* (or exogenous), and *endogenous* effects. Informally, the *endogenous* effect measures the effect of a statistic of the group outcomes (say, the average unemployment rate of a group), and the *contextual* effect is the effect of the group characteristics (for instance, the racial composition and average age of group members). Manski (1993) was the first to introduce the "reflection problem." Informally, this refers to the inability to separately identify these two types of effects.[3] In other words, one cannot identify whether some group characteristics have a direct effect on an individual (the contextual effect) or are "reflected" and mistakenly attributed to the effect of the group members' outcomes (the endogenous effect).

My methodological contribution can be used to separately identify the two effects if some people are influenced by more than one reference group. I further show how to estimate the two effects by explicitly solving for the two effects in the setting I consider. This allows for a comparison of the relative importance of the two effects. The magnitude of the two "types" of social effects determine the extent to which a change in one's outcome affects others in the group. This has important policy implications for determining the benefit of virtually any program, be it welfare, job training, or bussing of school children.

Since in many cases individuals are potentially affected by multiple circles of influence (such as neighborhood, family, friends from high school, friends from college, etc.), the framework can be applied in other settings. In addition, I provide a straightforward way to compute the various components of the social effect. Estimations of functions of

[3]Manski (1993) actually discusses two types of "reflection problem." I present both in Section 2.2.

the parameters of the model can be obtained using linear regression. In the structure I consider in the empirical part, the endogenous social effect can be easily calculated from the ratio of two coefficients obtained from a linear regression.

In the empirical part of my dissertation (Chapters 3 and 4) I focus on the effect of social interactions on the labor market outcomes of war veterans. The settings I consider allow me to address some of the critical issues faced by many empirical studies of social influence and peer effects. The three primary advantages of the settings I consider are that groups were formed due to an exogenous shock, that I observe all members of the groups (and these groups are well defined), and that I observe labor market outcomes of interest, such as employment.

There are two main problems which most empirical studies in the area of social networks face that my dissertation addresses. First, my datasets contain information on actual ties between agents. In the settings I consider, I observe the actual group memberships, and all members of the group. Further, the groups I examine had all experienced battle and were likely to forge meaningful ties. While the standard economic datasets include a wealth of information on labor market outcomes, they lack information on group membership.[4] Surveys which include additional information on the channel through which a job was obtained (for example, a neighbor as opposed to an employment agency), can only be used to test various predictions or highlight the importance of a certain channel. Without additional information on the actual group members who caused the outcome,

[4]One approach taken to overcome this problem and use the standard datasets is the use of some proxy of the relevant group. For instance, Bertrand et al. (2000) look at those who speak the same foreign language in the census, and Bayer et al. (2005) consider those who reside in the same census block to be the reference group.

one cannot hope to further uncover the mechanisms through which social interactions operate.

The second advantage of my empirical setting addresses a more substantive issue that goes beyond data limitations. I examine groups that were formed involuntarily due to an exogenous shock, the American Civil War or America's decision to enter World War I and its need to quickly raise a large army. This allows me to use far less restrictive assumptions. In contrast, in most instances, groups or social networks are formed endogenously. This can lead to many potential problems in inference, an issue that is recognized by almost every empirical study. For example, consider a case in which individuals with a higher unobserved ability (to the researcher) choose to become members of groups with higher observed group characteristics (say, average level of education). A straightforward estimation of the effect of the group characteristic will lead to biased results. As emphasized by Moffitt (2001), in the case of group interactions, correcting for this selection is even more challenging than the usual selection bias in the non-group case.

Realizing the importance of having randomly assigned groups, researchers in recent years examined social interactions in various settings in which groups were randomly assigned (for example, Sacerdote (2001), Zimmerman (2003), and Duncan et al., 2003). However, these studies focus on populations or outcomes which are somewhat specialized, be it recidivism among ex-cons, grades or drinking among college students, or the outcomes of welfare recipients. The samples I examine represent an important segment of the labor market, namely, working-age white males. Because the samples I study are fairly representative of the entire working-age male population, one may be more inclined to use the findings to address policy issues that affect the general population.

In Chapter 3, I focus on a sample of World War I veterans. I construct a new dataset of American men who were drafted and served together during World War I (1917-1919), and use it to examine the effect of networks formed during the war on postwar (1930) likelihood of employment. In the 1930 census, I find that a group's unemployment rate has an economically and statistically significant effect on a veteran's own likelihood of being employed. For instance, the magnitude of the effect can be summarized as follows. All else equal, a 1- percentage-point increase in his peers' unemployment rate decreases a veteran's likelihood of employment by 0.3-0.4 percentage points. I then provide robustness checks to address various concerns. For example, I examine alternative specifications of the "correct" reference group, and find that larger groups, such as battalions (which consist of four military companies) have no statistically significant effect. I also find the employment outcomes of other military companies within the same regiment to have no statistically significant effect. I show that the company's group effect persists after controlling for the prewar place of residence of the group's members by exploiting the variation in the group's composition of prewar locations.

I conclude Chapter 3 with an empirical application of the Multiple Reference Groups method introduced in Chapter 2. The method is illustrated by considering my sample of World War I veterans. For each of the veterans, the two groups of reference are the veterans who had served with him during World War I, and a group of his closest (in terms of distance) neighbors. The statistically significant results suggest that the endogenous effect is much larger when compared with the contextual effect. For various characteristics, the results suggest that at most, 20% of the total social effect on employment is due to the contextual effect.

The final chapter of the dissertation, Chapter 4, examines the labor market experiences of Civil War veterans of the Union army. The contribution of this chapter is twofold. First, as in the World War I case discussed above, the Union army sample provides an unusual circumstance under which networks were formed, namely a large-scale war, coupled with a rich data set which provides information on all members of a reference group, as well as labor market outcomes over time. Second, the time-series nature of the sample, that is, the fact that I observe the employment outcomes of the men during several periods after the war, allows me to purge out any effect that is due to an individual or group-level unobserved effect, provided that these unobserved effects are constant over time.

I find evidence of a statistically and economically significant peer effect among the Union army veterans. For example, in the 1900 census, the marginal effect of a 1-percentage-point increase in one's peers' long-term unemployment rate (defined as six or more months of unemployment in the past year), all else equal, increases one's probability of being long-term unemployed by an additional 0.2 percentage points. The statistically significant effect persists after correcting for the simultaneity generated by the peer effects, and controlling for personal characteristics such as age, marital status, occupation and macroeconomic conditions.

I conclude Chapter 4 by exploiting the time-series nature of the sample to purge out any individual and unobserved-group-level fixed effect. The analysis performed using the Union army sample illustrates the advantage a time-series dataset provides, as it allows one to estimate the social interaction regardless of the source or nature of the unobserved group and individual-level characteristics.[5]

[5]One still needs to assume that this unobserved effect is constant across the periods used.

This dissertation makes several contributions to the study of social interactions from a policy standpoint. The main methodological contribution is providing a framework which allows one to separately identify the contextual and endogenous effects. The method is applicable in any setting in which agents belong to more than one reference group. For example, consider a principal of a school who must decide on how to assign children to classrooms. Assume he or she observes the characteristics of the children. For a given pool of children, an assignment to groups could lead to different results, depending on the magnitude of the two effects. Understanding the relative importance of the effects can help when deciding on assignments to group. This dissertation also illustrates the advantage a time-series dataset can provide. For instance, consider a case in which policy makers are collecting data to study the effect of social interactions for the purpose of program evaluation. It is important to give serious consideration to collecting data over several time periods (time series). While the collection of time series data will be more costly, the results of the evaluation are likely to be more credible.

The empirical findings in this dissertation suggest that social interactions play an important role in the labor market even when groups are involuntarily formed. Two features of the samples examined make the results more likely to extend to other settings of interest. First, in both cases (Civil War and World War I), the sample is fairly representative of the white working-age male population. Second, the groups I examine are fairly heterogenous in their makeup. This could be of special importance if one is trying to understand the effect of forming heterogenous groups, such as the "bussing" of school kids.

Finally, my findings suggest that in the labor market there exists a strong endogenous effect. In other words, every time an individual becomes employed, an effect is propagated through their network, thereby affecting the employment likelihood of their group members. The existence of a strong endogenous effect implies that any public policy that targets unemployment is underestimating the benefits of the program if it fails to account for the "social multiplier" effect.

CHAPTER 2

Identification and Estimation of Social Interactions Using Multiple Reference Groups

2.1. Introduction

This chapter examines some of the identification and estimation issues surrounding social effects. I decompose the social effect by introducing a new framework, Multiple Reference Groups, which allows one to separately identify the two components of the social effect. The components are commonly referred to in the literature as the *contextual* (or exogenous), and *endogenous* effects. Informally, the *endogenous* effect measures the effect of a statistic of the group outcomes (for example, the average unemployment rate of a group), and the *contextual* effect is the effect of the group characteristics (for example, the race and average age of group members).

Further, I show that identification holds even in the case where some group types have perfectly correlated unobservables. I illustrate an application of the method by considering a sample of American World War I veterans and their neighbors in 1930 in Chapter 3.

Manski (1993) was the first to introduce what is known as the "reflection problem." Informally, one type of this problem refers to the inability to separately identify endogenous and contextual effect.[1] In other words, one cannot identify whether some group characteristics have a direct effect on an individual (the contextual effect) or are "reflected" and mistakenly attributed to the effect of the group members' outcomes (the endogenous effect).

My methodological contribution can be used to separately identify the two effects if some people are influenced by more than one reference group. The intuition behind the result is as follows. Consider two subgroups of people that do not know each other, but do share a common intermediate group of contacts (the intermediate group are those who belong to more than one group). One would not expect the characteristics of one subgroup to directly affect the outcomes of those in another subgroup to which they are not directly tied. If, in fact, I do find an effect, it is evidence for the existence of an endogenous effect that the intermediaries had propagated. However, the extent to which the endogenous effect is propagated also depends on the contextual effect. I further show how to estimate the two effects by explicitly solving for the two effects in the setting I consider. This allows one to compare the relative importance of the two effects.

The magnitude of the two "types" of social effects determine the extent to which a change in one's outcome affects others in the group. This has important policy implications for determining the benefit of virtually any program, be it welfare, job training, or bussing of school children. Since in many cases individuals are potentially affected by multiple circles of influence (such as neighborhood, family, friends from high school, friends from college, etc.), the framework can be applied in other settings.

[1] Manski (1993) actually discusses two types of "reflection problem." I present both in Section 2.2.

To give an example of the importance of the decomposition, as in Chapter 1, consider a principal of a school who must decide how to assign children to classrooms. Assume he or she observes the characteristics of the children. For a given pool of children, an assignment to groups could lead to different results, depending on the magnitude of the two effects. If the principal knew the relative importance of the two effects, he or she could optimally assign the children to classes in order to maximize some objective, such as helping the weaker students, or maximizing the average test score, or any other goal. Moreover, a strong endogenous effect suggests that any program which targets outcomes, for instance, helping one find a job, will have a "spillover" effect, such as increasing the employment likelihood of someone else in the network.

This chapter is related to a growing interest in the identification of social interactions. Manski (1993) was the first to formally consider many of the issues which are particular to the identification of social interactions. Concurrently with my work, Cohen-Cole (2006) showed identification for a setting in which agents are affected by more than one group in a linear-in-means model, but did not consider estimation. Bramoullé et al. (2006) show identification for various types of network structures in a linear-in-means setting, and characterize necessary and sufficient conditions for structures to be identified. Lee (2006) shows identification and discusses estimation in the linear-in-means case when there is a large variation in group size.

The contribution of this chapter is two-fold. First, I show that the existence of multiple reference groups can aid in identification even in cases where group characteristics are perfectly correlated with the average of members' characteristics. Second, I provide a straightforward way to compute the various components of the social effect. Estimations of functions of the parameters of the model can be obtained using linear regression. In the structure I consider as an example, the endogenous social effect can be easily calculated from the ratio of two coefficients obtained from a linear regression.

Though in this chapter I focus on the linear-in-means case, as the bulk of the literature has used that specification for estimation, identification is by no means a result of functional form. Brock and Durlauf (2001a, 2004) show identification for the case in which the dependent variable is binary by exploiting the non-linearity of that setting. For the functional forms they consider, the Multiple Reference Groups framework can be used by relaxing some of their assumptions (though requiring additional information on multiple group membership).

The rest of the chapter is organized as follows. Section 2.2 introduces the standard linear-in-means single group case and some of the issues regarding identification. Section 2.3 introduces the Multiple Reference Groups framework and Section 2.4 illustrates how to adapt the identification result to a concrete structure that will be estimated in Chapter 3. Section 2.5 concludes and suggests some possible extensions.

2.2. The Linear-In-Means Single Group Case

In this section I introduce some of the basic notation as well as several issues regarding the identification of social influences. I discuss the three most commonly raised issues in regards to the identification of group or social effects, but in no way is this a complete treatment. The interested reader is referred to Brock and Durlauf (2001b) and Moffitt (2001) for a more thorough discussion of these issues.

Assume there are $g = 1..G$ groups, each with n_g members $i = 1, 2..n_g$. In many types of models, the econometric specification is written as:

$$y_{i,g} = h[\alpha + x'_{i,g}\beta + Z'_g\gamma + \rho \cdot m(\overrightarrow{y}_{-i,g}) + \epsilon_{i,g}] \tag{2.1}$$

where each individual, indexed by i, g has an outcome of interest y, say the binary outcome of being employed or unemployed, a vector of covariates x which affect the likelihood of employment, such as age, occupation, and local labor market conditions, and an error term ϵ_i, a scalar capturing the individual unobservable characteristics and shocks to his or her employment prospects. In addition, each individual's job prospects might depend on the group's characteristics summarized by the vector Z_g, and the outcomes of all other members in the group $\overrightarrow{y}_{-i,g}$. γ is often referred to in the literature as the contextual (or exogenous) effect, and ρ as the endogenous effect. It is possible, and in most instances, quite likely, that Z_g depends on the characteristics of others. For instance, Z_g might just be the average of the group characteristic, $Z_g = \frac{1}{n_g}\sum_{i \in g} x_i$, for example, the average age among group members.

This specification already incorporates the assumption that the various arguments are separable and linearly additive in the function $h[\cdot]$, and that all of the group's endogenous effect is aggregated or summarized through the function $m : \mathbb{R}^{G-1} \to \mathbb{R}^1$. For instance, when considering social norms, one could motivate the form $m = E[y_j | j \in g]$, that is, one's utility depends on the expectation of others' behavior. However, depending on the setting, one could motivate other statistics, such as $m = \min[y_j | j \in g]$. For example, in a classroom setting where the most disruptive kid would have the strongest effect on others' learning.

Equation (2.1) in some form or another is used in almost every empirical study of peer effects.

In the empirical parts of Chapter 3 and Chapter 4, and in the majority of empirical studies, m_g will be replaced with $\bar{y}_g = \frac{1}{n_g}\sum_{i\in g} y_i$, the average outcome among group members, or the same measure, self excluded, $\bar{y}_{g,-i} = \frac{1}{n_g-1}\sum\limits_{j\in g, j\neq i} y_j$. In Sections 3.4 and 4.4, I discuss how I operationalize the above specification in the empirical part of my work.

For both measures, in the case of small group sizes, the use of either measure ($\bar{y}_g$ or $\bar{y}_{g,-i}$) introduces problems in the coherency of the model, depending on the functional form used (See Heckman (1978) for example). In the linear-in-means model considered below, coherency is not a problem.

The first two types of problems regarding identification were first formalized by Manski (1993) and are known as the "reflection problem." Consider again equation (2.1). For simplicity, and with no loss of generality, assume that all variables are of dimension one. Further, the endogenous effect depends on the expected outcome among group members $m_g^e = E[y_g]$. The basic linear-in-means model, where $h[t] = t$ can be written as:

$$y_{i,g} = \alpha + \beta x_{i,g} + \gamma Z_g + \rho m_g^e + \epsilon_{i,g} \tag{2.2}$$

Next, assume that:

Assumption 1. $E[\epsilon_i \,|\, x_i, Z_g, i \in g] = 0$

Then taking expectation of both sides of equation (2.2) one obtains:

$$m_g^e = \alpha + \beta E[x_i \,|\, i \in g] + \gamma Z_g + \rho m_g^e \tag{2.3}$$

Rearranging terms and substituting equation (2.3) back into equation (2.2) one can derive:

$$y_{i,g} = \frac{\alpha}{1-\rho} + \beta x_{i,g} + \frac{\beta\rho}{1-\rho} E[x_i \,|\, i \in g] + \frac{\gamma}{1-\rho} Z_g + \epsilon_{i,g} \tag{2.4}$$

Seemingly, based on Assumption 1, one could estimate the unknown coefficients by using ordinary-least-squares. However, if group attributes depend on the characteristics of its members (for example, say $E[x_i \,|\, i \in g] = Z_g$), then ρ and γ, the endogenous and exogenous effects, cannot be separately identified.[2] (See the remark following Proposition 1 in Manski, 1993.) Manski coined the term "the reflection problem" to describe this problem. Note that this result was derived even though it was assumed that the error terms are independent of group and own characteristics, nor was it assumed that the errors were correlated among group members.

The second type of "reflection" discussed in Manski (1993) is one in which the unobserved errors are correlated across group members and depend on the group attributes. Similar to the above case, one cannot separately identify whether the observed effect of the group outcomes is due to the "endogenous" effect or whether it is just a reflection of the group's unobservables. This concern will be addressed in Section 2.3.2, where I show

[2] One could, however, identify the existence of a social effect, namely whether: $\frac{\beta\rho+\gamma}{1-\rho} \neq 0$.

that identification is possible even in the case that the unobserved errors are correlated with group characteristics in some of the groups.

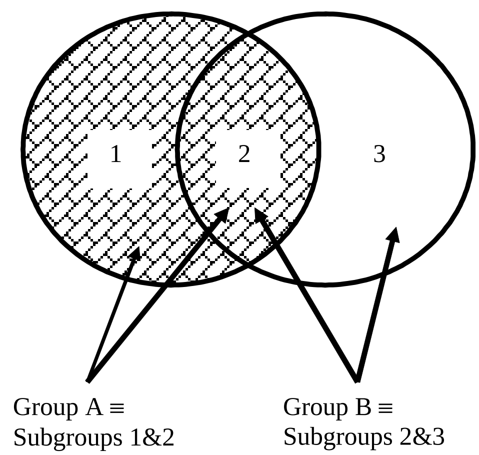
The third concern which is common to many group settings is that often group formation is endogenous. If the unobserved characteristics of members influence their choice of groups, and if that choice is based on the observed (or unobserved) characteristics of the group, then the usual selection problems arise. This selection problem is, of course, not unique to group settings, but is usually harder to address. In addition, while the selection problem is likely to result in the second type of reflection problem discussed above (correlated unobservables), even if group membership is exogenous, correlated unobservables might still exist if the researcher cannot control for all of the group characteristics or if there are measurement errors (see Moffitt, 2001). The military groups considered in the empirical application in Chapter 3 were formed involuntarily, and are consistent with random assignment.

2.3. Identification Using Multiple Reference Groups

I now introduce a framework in which the endogenous and contextual effects can be separately identified, even if group attributes are perfectly correlated with the characteristics of the group members. I prove identification for the linear-in-means case. Also, I show that even if the error terms in one of the groups are correlated with the group's characteristics identification is still possible.

The main requirement for identification is that (some) individuals belong to more than one reference group. It is not required that all individuals belong to more than one group, nor that the econometrician observes all group memberships for all individuals.

Figure 2.1: Group Subdivision Notation



2.3.1. Linear-In-Means Case

To illustrate, I start out by focusing on the linear-in-means case, and focus on the most basic possible structure, one in which at least some individuals have more than one reference group. This is the structure depicted in Figure 2.1. There are two groups, Group A and Group B, which intersect. I further label the partition that this structure creates. The intersection of the two groups, Subgroup 2, contains all the members which belong to both reference groups A and B. The two other subgroups, Subgroup 1 and 3, consist of members which belong only to Group A or Group B, respectively. It is possible that those in Subgroups 1 and 3 belong to additional groups which are unobserved by the econometrician. It is important to note that, in general, identification does not depend on observing those additional group memberships.[3] To illustrate the structure using the

[3]Note that according to the assumed structure the unobserved group membership of those in Subgroup 1 cannot be those in Subgroup 3, since by definition, those in subgroup 2 are those members who belong both to subgroups 1 and 3.

empirical application that will be introduced in Section 2.4, Group A consists of all members who had served in a certain military unit, and Group B consists of all those residing in a certain neighborhood block.

The above structure is for expositional purposes only, and the focus of this chapter is not to characterize the set of all structures which allow one to separately identify the contextual and endogenous effects. However, there are two necessary conditions which the structure must meet. The first is the existence of a non-empty intersection, that is Subgroup 2 cannot be empty.[4] The second needed condition is that there exists some form of exclusion, that is that one group is not a subset of the other (or in order words, both Subgroups 1 and 3 are non-empty).[5] As remarked above, one should not literally think of those members in Subgroup 3 as belonging to only one group, but rather that the other group they belong to is not Subgroup 1. The fact that I do not observe the other memberships of those in Subgroups 1 and 3, does not impede identification (this will be made clear in the following sections).

[4]If Subgroup 2 were empty for all cases observed by the econometrician, then this is equivalent to the one-group case discussed in Section 2.2 that cannot be separately identified.

[5]If Subgroup 1 or 3 are empty this means that those individuals in subgroup 2 belong to two groups, one of which is entirely contained in the other. If the effects of all groups to which a member belongs are assumed to be the same then this again reduces to the one-group case. If one allows for the effects to be different for the two groups a member belongs to, it is straightforward to show that separately identifying the two effects is not possible.

Allowing the contextual effect to vary across group types,[6] the linear-in-means model for this structure can be written as:

$$y_{i,g=2} = \alpha + \beta x_{i,g=2} + \gamma_A Z_{groupA(1\&2)} + \gamma_B Z_{groupB(2\&3)} + \rho \cdot m_{groups1\&2\&3} + \epsilon_{i,g=2} \tag{2.5}$$

$$y_{i,g=1} = \alpha + \beta x_{i,g=1} + \gamma_0 Z_{groupA(1\&2)} + \rho \cdot m_{groups1\&2} + \epsilon_{i,g=1}$$

$$y_{i,g=3} = \alpha + \beta x_{i,g=3} + \gamma_0 Z_{groupB(2\&3)} + \rho \cdot m_{groups2\&3} + \epsilon_{i,g=3}$$

In the above, α and β are the individual effects, γ_0 is the contextual effect for those who belong to only one group, γ_A and γ_B are the contextual effects for those who belong to two groups, and ρ is the endogenous effect which is affected by the average outcome of all those in the groups the individual belongs to. Allowing for a different contextual effect for those who only belong to one group (or for those for which information on only one group is available to the econometrician), allows for a more general case. As remarked above, identification does hold if all individuals belong to multiple groups. In which case $\gamma_0 = \gamma_A$ for those in Group A, and $\gamma_0 = \gamma_B$ for those in Group B.

As can be seen from equation (2.5), each type of individual has a different reference group, denoted by the index of m and Z. From a structural standpoint, the last two equations (for $y_{i,g=1}$ and $y_{i,g=3}$) are equivalent. However, both are included here, since it is important to keep in mind that the covariates of Group 3 enter Group 1 due to the simultaneous nature of the structure. Note that it is possible that those in Group 1, or 3,

[6] This assumption embeds the special case in which the contextual effect is assumed to be the same for all groups.

are influenced by additional groups that the econometrician does not observe. Identification is still possible as long as those additional unobserved groups are uncorrelated. This point can be seen more clearly by examining the assumptions and proof of Proposition 1.

Next, for notational convenience and with no loss of generality, assume that the three subgroups 1-3 are of the same size and define: $m_t = E[y|Subgroup\ t]$, the expected value of the outcome y for the subgroup t, for the three groups. For example, $m_1 = E[y_{i,g=1}]$.

Using the above notation, rewrite equation (2.5) as:

$$\begin{aligned}
y_{i,g=2} &= \alpha + \beta x_{i,g=2} + \gamma_A Z_{g=1\&2} + \gamma_B Z_{g=2\&3} + \frac{\rho}{3}[m_1 + m_2 + m_3] + \epsilon_{i,g=2} \\
y_{i,g=1} &= \alpha + \beta x_{i,g=1} + \gamma_0 Z_{g=1\&2} + \frac{\rho}{2}[m_1 + m_2] + \epsilon_{i,g=1} \\
y_{i,g=3} &= \alpha + \beta x_{i,g=3} + \gamma_0 Z_{g=2\&3} + \frac{\rho}{2}[m_2 + m_3] + \epsilon_{i,g=3}
\end{aligned} \tag{2.6}$$

Assuming, as in Section 2.2, that $E[\epsilon_i|x, Z] = 0$, one could take expectations in both sides of (2.6) to obtain:

$$m_2 = \alpha + \beta E[X]_2 + \gamma_A Z_{g=1\&2} + \gamma_B Z_{g=2\&3} + \frac{\rho}{3}[m_1 + m_2 + m_3] \tag{2.7}$$

where $E[X]_q$ is defined as $E[x_{i,g}|g = q]$, and in the empirical setting $\bar{X}_q = \frac{1}{n_{g=q}} \sum\limits_{i \in g=q} x_i$. Similarly, one could obtain the same for m_1 and m_3. From equation (2.7) one can better see the entwined simultaneity. Assume the contextual effect is perfectly correlated with the average group characteristics $E[X]_g$. Recall from Section 2.2 that in the case of a single reference group, one cannot separately identify the endogenous and contextual effects. In contrast, in the above case the system is fully identified.

Proposition 1. *In the above model (equation 2.5), under the assumptions:*

(i) $E[\epsilon_{t,g}|x_t, Z_g; t \in g] = 0$

(ii) $(1, E[x|g = 1], E[x|g = 2], E[x|g = 3])$ *are linearly independent for some groups.*

In the case of perfectly correlated contextual effects: $Z_g = E[x_t \,|\, t \in g] \;\; \forall g, t,$ *the parameters* $\alpha, \beta, \gamma_0, \gamma_A, \gamma_B,$ *and* ρ *are globally identified.*

Proof. Consider two sets of parameters $(\alpha, \beta, \gamma_0, \gamma_A, \gamma_B, \rho)$ and $(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}_0, \tilde{\gamma}_A, \tilde{\gamma}_B, \tilde{\rho})$, and the system of equations implied by the observational equivalence.

The system of interest is:

$$
\begin{aligned}
(2.8) \quad y_{i,g=2} &= \alpha + \beta x_{i,g=2} + \gamma_A Z_{g=1\&2} + \gamma_B Z_{g=2\&3} + \frac{\rho}{3}(m_1 + m_2 + m_3) + \epsilon_{i,g=2} \\
y_{i,g=1} &= \alpha + \beta x_{i,g=1} + \gamma_0 Z_{g=1\&2} + \frac{\rho}{2}(m_1 + m_2) + \epsilon_{i,g=1} \\
y_{i,g=3} &= \alpha + \beta x_{i,g=3} + \gamma_0 Z_{g=2\&3} + \frac{\rho}{2}(m_2 + m_3) + \epsilon_{i,g=3}
\end{aligned}
$$

Taking expectation with respect to x_t of each equation by using assumption (i) and since $Z_g = E[x_t| \, t \in g] \;\; \forall g$ we obtain:

$$
\begin{aligned}
m_2 &= \alpha + \beta E[X]_2 + \frac{\gamma_A}{2}(E[X]_1 + E[X]_2) + \frac{\gamma_B}{2}(E[X]_2 + E[X]_3) + \frac{\rho}{3}(m_1 + m_2 + m_3) \\
m_1 &= \alpha + \beta E[X]_1 + \frac{\gamma_0}{2}(E[X]_1 + E[X]_2) + \frac{\rho}{2}(m_1 + m_2) \\
m_3 &= \alpha + \beta E[X]_3 + \frac{\gamma_0}{2}(E[X]_2 + E[X]_3) + \frac{\rho}{2}(m_2 + m_3)
\end{aligned}
$$

where $E[X]_q$ is defined as $E[x_{i,g=q}]$ and $m_q = E[y|subgroup\ q]$

First, it is straightforward to establish that if $\rho = 0$, we are done. Assuming $\rho \neq 0$, from the above system of equations it follows that if $\rho = \tilde{\rho}$, then

$\alpha = \tilde{\alpha}, \beta = \tilde{\beta}, \gamma_0 = \tilde{\gamma}_0, \gamma_A = \tilde{\gamma}_A, \gamma_B = \tilde{\gamma}_B$ by assumption (ii). Hence, it remains to show that ρ is identified under the assumption $\rho \neq 0$.

One can then solve the system as a linear projection of (m_1, m_2, m_3) on $(1, E[X]_1, E[X]_2, E[X]_3)$:

$$
\begin{aligned}
m_1 =& \theta_0 + \theta_1 E[X]_1 + \theta_2 E[X]_2 + \theta_3 E[X]_3 \\
m_2 =& \theta_0 + \theta_4 E[X]_1 + \theta_5 E[X]_2 + \theta_6 E[X]_3 \\
m_3 =& \theta_0 + \theta_3 E[X]_1 + \theta_2 E[X]_2 + \theta_1 E[X]_3
\end{aligned} \tag{2.9}
$$

The parameters $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ are all functions of the original parameters of interest $(\alpha, \beta, \gamma_0, \gamma_A, \gamma_B, \rho)$ and are identified by assumption (ii), that is $\theta_t = \tilde{\theta}_t, \forall t = 0..6$. One can then substitute equation (2.9) back into equation (2.8). The equation for group 3 is symmetrical to the case $g = 1$.

$$
\begin{aligned}
y_{i,g=1} \quad = \quad & \frac{\alpha}{(1-\rho)} + \beta x_{i,g=1} + \\
& \frac{-12\gamma_0 - 6\rho\gamma_A + 3\rho^2\gamma_A - 12\rho\beta + 10\rho\gamma_0 + 6\rho^2\beta + 2\rho^3\beta}{-2(12 - 16\rho + 3\rho^2 + \rho^3)} E[X]_1 + \\
& \frac{3\rho\gamma_B + 6\rho\beta - 2\rho\gamma_0 + 3\rho\gamma_A + 6\gamma_0}{-2(\rho^2 + 5\rho - 6)} E[X]_2 + \\
& \frac{\rho(-3\rho\gamma_B + 2\rho\gamma_0 + 4\rho\beta + 6\gamma_B)}{2(12 - 16\rho + 3\rho^2 + \rho^3)} E[X]_3 + \epsilon_{i,g=1}
\end{aligned} \tag{2.10}
$$

$$
\begin{aligned}
y_{i,g=2} \;\; = \;\; & \frac{\alpha}{(1-\rho)} + \beta x_{i,g=2} + \frac{-3\rho\gamma_A + 2\rho\gamma_0 + 4\rho\beta + 6\gamma_A}{-2(\rho^2+5\rho-6)} E[X]_1 + \\
& \frac{2\rho^2\beta + 4\rho\gamma_0 - 3\rho\gamma_A - 3\rho\gamma_B + 4\rho\beta + 6\gamma_B + 6\gamma_A}{-2(\rho^2+5\rho-6)} E[X]_2 + \\
& \frac{-3\rho\gamma_B + 2\rho\gamma_0 + 4\rho\beta + 6\gamma_B}{-2(\rho^2+5\rho-6)} E[X]_3 + \epsilon_{i,g=2}
\end{aligned}
$$

While the expressions for each θ_t are somewhat intricate, one could directly compute the endogenous effect ρ.

consider the ratio:

$$\theta_6/\theta_3 = 1 - \frac{2}{\rho}.$$

(recall $\rho \neq 0$). Since $\theta_6/\theta_3 = \tilde{\theta}_6/\tilde{\theta}_3$ it follows that $\rho = \tilde{\rho}$. □

Remark 1 (Intuition for the result of Proposition 1)**.** *The intuition for the above result is as follows. Consider those in Subgroup 3. Even though they are not directly linked to those in Subgroup 1, both their characteristics and their outcomes (contextual and endogenous effects) affect those in Subgroup 1. However, these effects are propagated to those in Subgroup 1 via an intermediate, those in Subgroup 2 (since those in Subgroups 1 & 3 are not directly linked). In turn, this implies that any effect of those in Subgroup 3 on those in Subgroup 1 must be due to the* ***endogenous effect.***

This point is made explicit by examining the proof, in which the effect of X_3 is different for those in Subgroups 1 and 2 (as captured by the coefficients θ_3 and θ_6). It is the ratio of the two (θ_6/θ_3) which ultimately leads to the identification of the endogenous effect ρ.

Note that once the endogenous effect is identified, then the contextual effect is identified, since the sum of the effects is identified even in the single group case.

The proof of the above proposition also demonstrates that in the linear-in-means structure considered, calculating the endogenous effect is straightforward. It amounts to running a least-squares regression of the outcomes on the covariates and the three subgroups averages of those covariates ($\frac{1}{n_{g=q}} \sum_{i \in g=q} x_i$; $q = 1, 2, 3$). The endogenous effect can then be computed as a function of the coefficients. In addition, if the covariate x is a vector, then a testable implication is that the ratio of the coefficients should be the same across the various components of the covariates. For example, in the above proof, the ratio θ_6/θ_3 should be the same for all components of the vector $E[\vec{X}]_3$.

Equation (2.10) also illustrates the fact that if $\rho = 0$ then the coefficient θ_3 (for the variable $E[X]_3$) would be zero. The reduced form allows to test for the existence of an endogenous social effect. This can be done with far less restrictive assumptions that those used in the full specification, such as equation (2.8). This is further illustrated in Section 2.4.

2.3.2. Correlated Unobservables

Correlated unobservables are likely one of the biggest concerns faced by any empirical study of social interactions. The correlation could arise when group membership is endogenous. Though as discussed in Section 2.2, it could arise for other reasons. For instance, in the case of neighborhoods, people with higher ability might choose their neighborhood based on the education level and racial composition of their neighbors. Even if the error term is uncorrelated with the group characteristics Z_g, if people choose based on the individual characteristics of others then the results will be, in general, biased.

In this section, following Manski (1993), I consider correlated unobservables to mean that the individual unobservable term is correlated with the group average characteristics $E[X]_g$.

Definition: Correlated Unobservables: $E[\epsilon_i \mid i \in g] = s \cdot E[X]_g$ for some $s \neq 0$

In the single-group case this can be seen when examining the reduced form equation. To simplify, consider a case in which it is known a priori that there was no contextual effect, that is $\gamma = 0$. If the individual unobservable is correlated with the group average characteristics $E[X]_g$ (say $E[\epsilon_i \mid i \in g] = sE[X]_g$), then looking at the reduced form:

$$y_{i,g} = \frac{\alpha}{1-\rho} + \beta x_{i,g} + \frac{\beta\rho}{1-\rho} E[X]_g + \epsilon_{i,g}$$

it is straightforward to see one cannot hope to identify ρ without any further information (such as a valid instrument).

In contrast, in the case of Multiple Reference Groups, if the errors of one of the group types were uncorrelated, then the model is identified even when the errors of the other group type are correlated.

Consider again the basic Multiple Reference Groups linear-in-means model:

$$
\begin{aligned}
y_{i,g=2} =& \alpha + \beta x_{i,g} + \gamma_A Z_{g=1\&2} + \gamma_B Z_{g=2\&3} + \frac{\rho}{3}[m_1 + m_2 + m_3] + \epsilon_{i,g=2} \\
y_{i,g=1} =& \alpha + \beta x_{i,g} + \gamma_0 Z_{g=1\&2} + \frac{\rho}{2}[m_1 + m_2] + \epsilon_{i,g=1} \\
y_{i,g=3} =& \alpha + \beta x_{i,g} + \gamma_0 Z_{g=2\&3} + \frac{\rho}{2}[m_2 + m_3] + \epsilon_{i,g=3}
\end{aligned} \tag{2.12}
$$

Even if $E[\epsilon_t \mid t \in g] = sE[X]_g$ for some types of subgroups the model is identified.

Proposition 2. *In the above model, under the assumptions:*

(i) $E[\epsilon_{i,g}|x_i, X_g, Z_g; g = 1] = E[\epsilon_{i,g}|x_i, X_g, Z_g; g = 2] = sE[X]_{g=1\&2}$

and $E[\epsilon_{i,g}|x_i, X_g, Z_g; g = 3] = 0$ *(unobserved correlation in one of the groups)*

(ii)$(1, E[x|g = 1], E[x|g = 2], E[x|g = 3])$ *are linearly independent for some groups.*

Then in the case of perfectly correlated contextual effects: $Z_g = E[x_t \mid t \in g]$

the parameters $\alpha, \beta, \gamma_0, \gamma_A, \gamma_B, \rho,$*and* s *are globally identified.*

Proof. The proof is very similar to that of Proposition 1.

Consider the projection of (m_1, m_2, m_3) on $(1, E[X]_1, E[X]_2, E[X]_3)$:

$$
\begin{aligned}
m_1 =& \theta_0 + \theta_1 E[X]_1 + \theta_2 E[X]_2 + \theta_3 E[X]_3 \\
m_2 =& \theta_0 + \theta_4 E[X]_1 + \theta_5 E[X]_2 + \theta_6 E[X]_3 \\
m_3 =& \theta_0 + \theta_3 E[X]_1 + \theta_2 E[X]_2 + \theta_1 E[X]_3
\end{aligned} \tag{2.13}
$$

Here too, the parameters $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ are all functions of the original parameters of interest $\alpha, \beta, \gamma_0, \gamma_A, \gamma_B, \rho, s$ and s are identified, that is $(\theta_t = \overline{\theta}_t, \forall t = 0..6)$. The added complication is that some of them include an extra term (s) because of the correlation. However, since the individual error terms are uncorrelated with the group average of one of the subgroups, namely Subgroup 3 (by assumption (i) $E[\epsilon_{i,g}|x_i, X_g, Z_g; g = 3] = 0$), the coefficients θ_3, and θ_6 are in turn, the same as in the uncorrelated case. Hence, as in the uncorrelated case $\theta_6/\theta_3 = 1 - \frac{2}{\rho}$ and so $\rho = \overline{\rho}$. The other parts of the proof are the same as the proof of Proposition 1 with one additional complication, the term variable s. Once ρ is identified $(\rho = \tilde{\rho})$ the reduced form of Subgroups 1 and 2 has an additional term $(s - \tilde{s})$ which in the single-group case would prevent the effects from being identified

$$(\alpha - \alpha) + (\beta - \tilde{\beta})E[X]_1 + \tfrac{(\gamma_0 - \tilde{\gamma}_0) + (s - \tilde{s})}{2}(E[X]_1 + E[X]_2) = 0$$

$$(\alpha - \tilde{\alpha}) + (\beta - \tilde{\beta})E[X]_2 + \tfrac{(\gamma_A - \tilde{\gamma}_A) + (s - \tilde{s})}{2}(E[X]_1 + E[X]_2) + \tfrac{(\gamma_B - \tilde{\gamma}_B)}{2}(E[X]_2 + E[X]_3) = 0$$

However, this term is identified by looking at the reduced form of Subgroup 3, which does not contain the correlation term:

$$(\alpha - \tilde{\alpha}) + (\beta - \tilde{\beta})E[X]_3 + \tfrac{(\gamma_0 - \tilde{\gamma}_0)}{2}(E[X]_2 + E[X]_3) = 0 \qquad \square$$

It is useful to consider whether the assumptions and structure of the above proposition could in fact occur in reality. One such scenario could arise if those in Subgroup 3 are randomly assigned and their unobservable error terms are not correlated with the group characteristics. Some of the group members (those in Subgroup 2) then go on to choose an additional subgroup (Subgroup 1) based on their own unobserved component and the characteristics of Subgroup 1. Such a scenario would potentially yield correlation of unobservables in one of the type of groups, as assumed in the above proposition.

2.3.2.1. Group Size and Functional Form. The Multiple Reference Groups framework is identified regardless of the group sizes (recall from Section 2.3.1 that there must be some overlap between the groups, and that one group cannot be a subset of the other).

However, estimation does depend on the size of the groups, and more importantly, on the relative size of the various subgroups. For instance, if the intersection of two groups consists of only one member (Subgroup 2 is of size 1), then the effect of Subgroup 3 propagated through Subgroup 2 is likely to be small, and therefore empirically difficult to detect. In the next section, I present a way around the problem that in the dataset used in Chapter 3 the size of the intersection groups (i.e., those who had served in the same unit and live in the same block) is relatively small.

In regards to functional form, it is important to note that the underlying source of identification is not the functional form, such as the linear-in-means model considered in the previous section, but rather the group structure (i.e. the availability of multiple reference groups for at least some).

In the case for which the outcome is binary, Brock and Durlauf (2001, 2004) show identification for various settings, taking advantage of the non-linearity induced by the functional form. I have considered identification using a different set of assumptions by exploiting the Multiple Reference Groups structure. For instance, it is possible to show identification for the case in which the functional form is logit, corresponding to the setup considered in Brock and Durlauf (2001a). It remains a topic for future research to examine what classes of functional forms are identified when some members belong to multiple reference groups.

2.4. An Application of the Multiple Reference Groups Framework

I first modify the framework presented in Section 2.3.1 to address the fact that in the application estimated in Chapter 3, the size of the intersection subgroup (those who had served in the same unit *and* live in the same neighborhood block) is relatively small when compared with the sizes of the military units and neighborhood blocks. In addition, to illustrate that identification does not come from assuming that the endogenous effect is the same for the different types of reference groups, I allow for both the contextual and the endogenous effects to be different for the two types of reference groups (for example, the army unit, and the neighborhoods).

2.4.1. Setup

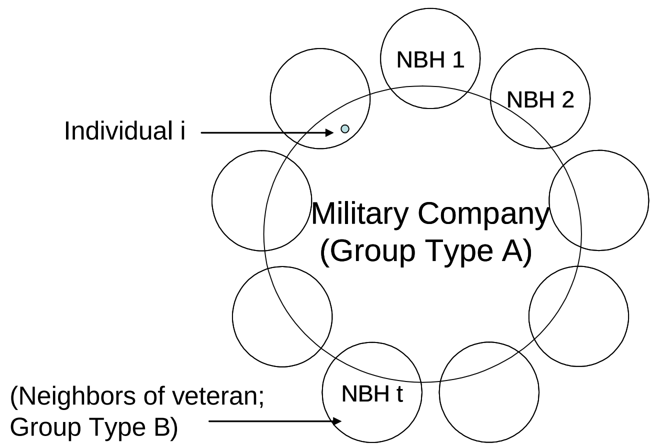
The basic model of interest is:

$$y_{i,g=k} = \alpha + \beta x_{i,g=k} + \gamma Z_{-i,k} + \rho \cdot m_{-i,k} + \epsilon_{i,k}$$

To reduce the use of notation, throughout this section I already incorporate the case in which group characteristics are perfectly correlated with the average individual characteristics $Z_k = E[x_g|g = k]$. The use of the $-i$ notation denotes the fact that the specifications in this section incorporate the "self-excluded" social effect, both for the endogenous effect ρ, and the exogenous effect γ.

The group structure assumed is that the econometrician observes the members of the two reference groups which influence the agent. To illustrate using the empirical setting considered in Chapter 3, group A are those in the military unit, and group B are those in one's neighborhood block. Further, the size of the intersection between each group A and group B is very small, say of size one. Groups of type A and B could be of different sizes, as will be shown below. As discussed in Section 2.3, it is important to note that those in group B may have additional reference groups that influence them that the econometrician does not observe. As long as these additional unobserved groups are not correlated across

Figure 2.2: Veterans and Neighbors Multiple Reference Group Diagram



groups of type B, the estimates will be consistent. This point will be further illustrated in the empirical application. The group structure is illustrated in Figure 2.2.

We can now write the model of interest as a system of equations for groups of Type A and B:

$$
\begin{aligned}
(2.14)\quad y_{i,g=A} &= \alpha + \beta x_{i,A} + \gamma_A E[X]_{-i,A} + \gamma_B E[X]_{-i,B} + \rho_A m_{-i,A} + \rho_B m_{-i,B} + \epsilon_{i,A} \\
y_{j,g=B-A} &= \alpha + \beta x_{j,B-A} + \gamma_0 E[X]_{-j,B-A} + \rho_0 \cdot m_{-j,B-A} + \epsilon_{j,B-A}
\end{aligned}
$$

where $B - A$ are those in Group B that do not belong to Group A, and as in the previous section, the notation used is: $E[X]_{-i,g} = E[x_t | t \in g; t \notin i]$, and $m_{-j,g} = E[y_t | t \in g; t \notin j]$. This specification incorporates the self-excluded social effect for both the endogenous effect ρ, and the exogenous effect γ. Note that this specification allows for a different exogenous effect for different groups. It nests the more restrictive case $\gamma_A = \gamma_B = \gamma_0$, $\rho_A = \rho_B = \rho_0$.

In addition to the structure, the main assumption is:

Assumption 2. $E[\epsilon_t \mid X] = 0$

For the empirical part, I will use the analog of the expectation operator.

Definition

$\bar{X}_{-i,A} = \frac{1}{n_A - 1} \sum\limits_{j \in \text{Group } A, j \neq i} x_j$

$\bar{X}_{-i,B} = \frac{1}{n_B - 1} \sum\limits_{j \in \text{Group } B \text{ of } i, j \neq i} x_j$

and $\overline{\overline{X}}^{A}_{-i,B} = \frac{1}{\text{Size of Group } A} \sum_{s \in A} \bar{X}_{-s,B}$

$\overline{\overline{X}}^{A}_{-i,B}$ is Group A's average of the average neighborhood characteristic (self-excluded) of all the neighborhoods in which the members of Group A reside. In the empirical application, for each unit (Group A) this amounts to the (unweighted) average of all the neighborhoods (Group B) of the unit members.

Consider the following reduced form:

$$y_{i,g} = \alpha + \beta x_{i,g} + \delta_1 \bar{X}_{-i,A} + \delta_2 \bar{X}_{-i,B} + \delta_3 \overline{\overline{X}}^{A}_{-i,B} + \epsilon_{i,g} \quad (i = 1..n_g; g = 1..G) \tag{2.15}$$

where δ_2 is the coefficient in the reduced form of the effect of the average neighborhood characteristic and δ_3 is the effect of the average of the average neighborhood for that same characteristic.

In the empirical application (Section 3.6) I exploit the fact that the likelihood of being a World War I veteran in 1930 is small enough to address any concern of a bias due to unobserved links among other veterans. Hence, in the specification considered here I assume that there are no systematic unobserved ties between type-B groups.

Before turning to the result, it is useful to examine the above reduced form. As discussed in further detail in the next section, $\delta_3 \neq 0$ suggests that there exists an endogenous effect. For example, say $\overline{\overline{X}}_B^A$ corresponds to the average across the military company of the average age of neighborhood residents. The fact that the characteristics of such a large group have an effect on company member i to which they are not directly tied suggests that they are, in fact, influencing the other company members, which in turn affect member i. Since the above equation includes a control for the characteristics of all those who are directly tied to member i ($\bar{X}_{-i,A}$ and $\bar{X}_{-i,B}$) then it follows that any effect detected by δ_3 is due to the endogenous effect.

Result 2.a

In the above model (equation 2.14), assuming $E[\epsilon_i|x_i, \bar{X}_{-i,A}, \bar{X}_{-i,B}, \overline{\overline{X}}_{-i,B}^A] = 0$ *the coefficients of the reduced-form Equation (2.15) are:*

$$\delta_2 = \frac{(\rho_B \cdot \beta \cdot n_B + \rho_B \cdot \gamma_0 \cdot n_B - \rho_B \cdot \gamma_0 - \rho_0 \cdot n_B \cdot \gamma_B + \gamma_B \cdot \rho_0 + n_B \cdot \gamma_B)}{(\rho_0 + n_B - \rho_0 \cdot n_B)}$$

$$\delta_3 = function(\rho_{A,}\rho_B, \rho_0, \beta, \gamma_0, \gamma_A, \gamma_B, n_B)$$

$$\rho_A = \frac{(\rho_B \cdot \rho_0 \cdot \frac{\delta_2}{\delta_3} + \rho_B \cdot \rho_0 + \rho_0 \cdot n_B - \rho_0 - n_B)}{(\frac{\delta_2}{\delta_3} + 1)(\rho_0 \cdot n_B - \rho_0 - n_B)}$$

and

$$\lim_{n_B \to \infty} \rho_A = \frac{1}{1 + \frac{\delta_2}{\delta_3}}$$

Result 2.b

Further, if it is assumed that the endogenous effects are the same for all types of groups ($\rho_A = \rho_B = \rho_0$) then:

$$\frac{\delta_2}{\delta_3} = \frac{((n_B)^2 - (n_B)^2\rho + n_A n_B \rho - n_B \rho^2 - 2n_A n_B \rho + n_B \rho + n_A n_B - n_A \rho^2 + n_A \rho)}{\rho \cdot (n_A n_B + n_B \rho - n_A n_B \rho + n_A \rho)}$$

where n_A, n_B are the sizes of group type A and B.[7]

The result is derived similarly to the proof of Proposition 1, by solving the system of equations and examining the coefficients of the reduced form in equation (2.15).

Similarly to Section 2.3.1, the intuition behind the result is as follows. The effect of all others' neighborhoods on one's own outcomes can only be due to the endogenous effect being propagated through a person's group (military unit) members.

2.4.2. Decomposition into the Contextual and Endogenous Effects

In this section, I show one way of comparing the endogenous and contextual effects. Note that one cannot directly compare the two since they measure the effect of different variables (say the effect of average age, and the effect of the average unemployment rate). Instead, I consider a decomposition of the marginal effect.

Consider again the specification:

$$y_{i,g=A} = \alpha + \beta x_{i,A} + \gamma_A E[X]_{-i,A} + \gamma_B E[X]_{-i,B} + \rho_A m_{-i,A} + \rho_B m_{-i,B} + \epsilon_{i,A} \tag{2.16}$$

[7]This assumes that the sizes are the same for each group of Type A and B. It is straightforward to allow for a different size for each specific group $g = 1..G$, which will end up producing a weighted average of the sizes.

For example, consider the measure $\overline{X}_{-i,A}$ (the empirical analogue of $E[X]_{-i,A}$). For large enough groups of size B, the marginal effect can be written as:

$$\lim_{n_B \to \infty} \frac{\partial y_i}{\partial \overline{X}_{-i,A}} = \gamma_A + \frac{\rho_A}{1-\rho_A} \cdot [\beta + \gamma_A] \tag{2.17}$$

The first part of this effect (γ_A) is the (direct) contextual effect.
The second part, $\frac{\rho_A}{1-\rho_A} \cdot [\beta + \gamma_A]$, is due to the endogenous effect (note that it equals zero if there is no endogenous effect, i.e. $\rho_A = 0$). The second part of the effect depends on the contextual effect and the own effect (β).

Finally, in order to estimate the effect, consider the following reduced form specification:

$$y_{i,g} = \alpha + \beta x_{i,g} + \delta_1 \bar{X}_{-i,A} + \delta_2 \bar{X}_{-i,B} + \delta_3 \overline{\overline{X}}^{A}_{-i,B} + \epsilon_{i,g} \quad (i = 1..n_g; g = 1..G) \tag{2.18}$$

Note that this specification only includes exogenous variables on the right hand side, and it is possible to estimate the coefficients $\alpha, \beta, \delta_1, \delta_2$, and δ_3 in equation (2.18) using ordinary-least-squares. In the above reduced form, for large enough n_B, $\delta_1 = \gamma_A + \frac{\rho_A}{1-\rho_A} \cdot [\beta + \gamma_A]$ and since $\lim_{n_B \to \infty} \rho_A = \frac{1}{1+\frac{\delta_2}{\delta_3}}$ we can estimate γ_A. I provide an empirical demonstration of the decomposition in Section 3.6.

2.5. Conclusion

In this chapter I examined the case in which some members belong to multiple reference groups and proved that the endogenous and contextual effects are separately identified in the linear-in-means case. This result is in contrast to Manski (1993) who shows that in the single group case in the linear-in-means setting the two effects are not separately identified. Moreover, the result holds even if one of the groups has perfectly correlated unobservables. A situation of perfectly correlated unobservables could arise if selection into groups is voluntary. Hence, it is important to allow for such a situation as it frequently arises in many datasets.

I then considered a specific case in which the intersection between the reference groups is very small. This structure will be estimated in the next chapter. I demonstrated how one could separately identify and compute the endogenous component of the social effect, as well as how to compare the relative strength of the components.

As briefly mentioned in Section 2.3.2.1, some possible extensions are examining other functional forms beyond the linear-in-means case, as well as characterizing the "power" of the estimates as a function of the group sizes.

It would be of interest to apply this framework to other settings in which social influences are believed to play a role. For example, consider a case in which a researcher observes the outcomes of high schoolers who come from different junior high schools.

CHAPTER 3

Social Interactions and Labor Market Outcomes of World War I Veterans

3.1. Introduction

In the labor market, various surveys have documented the importance of the "informal" channel, that is finding jobs through friends and relatives.[1] The goal of this chapter is to better understand how social networks affect labor market outcomes by examining groups which were formed involuntarily due to a quasi-random event. In addition, it illustrates an application of the new methodology introduced in Chapter 2 for decomposing the social effect into its two components, the *endogenous* and the *contextual* effects.

I construct a new dataset of American men who were drafted and served together during World War I (1917-1919), and use it to examine the effect of networks formed during the war on postwar (1930) likelihood of employment. The setting I consider allows me to address some of the critical issues faced by many empirical studies of social influence and peer effects. The three primary advantages of the setting I consider are that groups were formed due to an exogenous shock, that I observe all members of the groups (and these groups are well defined), and that I observe labor market outcomes of interest, such as employment.

[1]Ioannides and Loury (2004) summarize a number of surveys which find that 30-60 percent of jobs (in various industries and of various statuses) are found through the "informal" channel. As early as 1932, De Schweinitz (1932) finds that 45% of workers in the hosiery industry in Philadelphia obtained their job through friends and relatives.

In most instances, groups and social networks are formed endogenously. This can lead to many potential problems in inference, an issue which is recognized by almost every empirical study. For example, consider a case in which individuals with a higher unobserved ability (to the researcher) choose to become members of groups with higher observed group characteristic (say, average level of education), then a straightforward estimation of the effect of the group characteristic will lead to biased results. As emphasized by Moffitt (2001), in the case of group interactions, correcting for this selection is even more challenging than the usual selection bias in the individual case. In contrast, I examine groups that were formed involuntarily due to an exogenous shock, America's decision to enter the war and its need to quickly raise a large army.[2]

Researchers in recent years examined social interactions in various settings in which groups were randomly assigned (for example, Sacerdote (2001), Zimmerman (2003), and Duncan et al., 2003). However, these studies focus on populations or outcomes which are somewhat specialized, be it recidivism among ex-cons, grades or drinking among college students, or the outcomes of welfare recipients. The sample I examine represents an important segment of the labor market, namely, white males who were in their thirties and forties in 1930.

The other advantage of the setting I consider is the fact that I observe the actual group memberships, and all members of the group. Further, the groups I examine had all experienced battle overseas and were likely to forge meaningful ties. While the standard economic datasets include a wealth of information on labor market outcomes, they

[2]While the World War I draft was a "natural experiment" (See Meyer (1995) for an overview of "natural experiments"), note that in this paper I only examine those who had served (those "treated").

lack information on group membership.[3] In those cases where surveys include additional information on the channel through which a job was obtained (for example, a neighbor, as opposed to an employment agency), ultimately, they can only be used to test various predictions or highlight the importance of a certain channel. Without additional information on the actual group members which led to the outcome, one cannot hope to further uncover the mechanism through which social interactions operate.

In the 1930 census, I find that a group's unemployment rate has an economically and statistically significant effect on a veteran's own likelihood of being employed. The magnitude of the effect can be summarized as follows. All else equal, a 1-percentage-point increase in his peers' unemployment rate decreases a veteran's likelihood of employment by 0.3-0.4 percentage points. (Over 93% of the sample members are employed in the 1930 census; the unemployment rate among the entire male population was less than 10%.)

I then provide robustness checks to address various concerns. For example, I examine what is the "correct" reference group, and find that larger groups, such as battalions (which consist of four military companies) have no statistically significant effect. I also find the employment outcomes of other military companies within the same regiment to have no statistically significant effect. I show that the company's group effect persists after controlling for prewar place of residence of the group's members by exploiting the variation in the groups' composition of prewar locations.

In Sections 3.6 and 3.6.1, I further decompose the social effect by using the new Multiple Reference Groups framework introduced in Chapter 2 which allows one to separately

[3]One approach taken to overcome this problem and use the standard datasets is the use of some proxy of the relevant group. For instance, Bertrand et al. (2000) look at those who speak the same foreign language in the census, and Bayer et al. (2005) consider those who reside in the same census block to be the reference group.

identify the two components of the social effect. I illustrate an application of my method by considering a newly constructed sample of World War I veterans. For each of the veterans, the two groups of reference are the veterans who had served with him during World War I, and a group of his closest (in terms of distance) neighbors. The statistically significant results suggest that the endogenous effect is much larger than the contextual effect. For various characteristics, the results suggest that at most 20% of the total social effect on employment is due to the contextual effect.

There are several possible types of explanations as to why one's likelihood of employment might be affected by one's peers. These include referrals, social norms or "stigma" effects, information transmission, etc. Some of these channels will be discussed in more detail in Section 3.2.1.

The various explanations do not necessarily contradict one another, and at times the distinctions are somewhat arbitrary. Nevertheless, this chapter is motivated by the role networks play in information transmission, some reasons for which are given in Section 3.4.

The rest of the chapter is organized as follows. In Section 3.1.1, I review the relevant literature. Section 3.2 motivates the relation between social networks and labor market outcomes, such as employment. It includes a simple model that predicts that the peers' unemployment rate would affect one's own likelihood of employment. I then present the data used in this chapter in Section 3.3 and provide some of the institutional background on the draft and the army's structure during World War I. The empirical strategy is discussed in Section 3.4. It serves as an introduction to the notation and issues involving the estimation of peer effects, and discusses the advantages to using the military company as a reference group. The results are in Section 3.5. The decomposition of the social effect into its two components is presented in Section 3.6. Section 3.7 concludes and suggests some possible extensions.

3.1.1. Related Literature

Job search methods are traditionally categorized in the literature into "formal" and "informal" channels. The "formal" market consists of job posting, placement agencies, etc.[4] The "informal" market refers to jobs found through personal contacts, such as friends and relatives. One of the most commonly used frameworks for analyzing the "formal" labor market is the Mortensen-Pissarides search-equilibrium framework (Mortensen and Pissarides, 1999). In this framework, one's likelihood of employment only depends on others indirectly, through the likelihood of obtaining a match which in turn is a function of the number of vacancies and job seekers ("market tightness"). While I control for characteristics that are likely to affect one's likelihood of finding a job through the "formal" market, this chapter focuses on the "informal" channel.

The importance of the "informal" channel as a source of jobs has been documented by various surveys. Contacts have been shown to be of importance across occupations, skill levels, and countries. Ioannides and Loury (2004) provide a comprehensive overview of many of these findings. Bewley (1999, 368) lists 24 studies that were published between the years 1932-1990. The percent of jobs or job offers obtained through friends and relatives ranges from 18% to 78%, and is between 30% and 60% in most of those studies.

[4]For example, Maryland, as early as 1902, legislated (Chapter 365, Acts of 1902) the operation of a free state employment agency (State of Maryland Department of Labor, Licensing, and Regulation).

The recent work of Bayer et al. (2005) contributes to a better understanding of the referral aspect of networks. Using micro-level census data for Boston, they find that those who live on the same block are more than 50% more likely to work together, than those living in nearby blocks. Their findings substantiate my use of block-neighbors as the choice of the second reference group (the first being the military unit), as discussed in Section 3.6. One limitation of my data is that I do not observe the place of employment. On the other hand, Bayer et al. (2005) point out that their estimates are likely to downward bias the importance of networks, since block-neighbors are not the only source of jobs. My data is better suited to examine cases in which agents pass on information about other job openings, not necessarily in their own firm.

One aspect I do not address in this chapter is that of job quality (conditional on job type). For instance, one could think of a wage as a proxy for job quality. The 1930 Census did not collect information about income and wages. Various theoretical models predict different effects of networks on wages. For instance, Mortensen and Vishwanath (1994) predict a positive effect whereas Bentolila et al. (2004) predict a negative effect. This is further corroborated by Ioannides and Loury (2004), who report that various studies find different signs for the effect of networks on wages.

Group formation in my data is most closely related to the literature on random assignment to groups. For instance, Sacerdote (2001), Zimmerman (2003), and Duncan et al. (2003) consider random roommate assignments for various colleges. Bayer et al. (2004) consider cell-mate assignment for prisoners, and Kling et al. (2005), among others, have examined the Moving to Opportunity program which assigned low income families to socio-economically stronger neighborhoods. One common feature of all those settings is that the "treated" population is very homogenous in its nature. In contrast, the members of my sample come from a diverse background.

There is a relatively large literature in economics which examines the military. For instance, Angrist (1990) uses the Vietnam War draft to examine the effect of veteran status on earnings. However, the majority of these studies examine the "treatment effect" of military service (or type of military service). Note that I focus only on those who had served (and who have had very similar military experiences). In exploiting the strength of camaraderie in the military, my work is closest in spirit to that of Costa and Kahn (2003). They examine how group characteristics of units in the Civil War affect such measures as desertion and arrest. De Paula (2005) uses the same dataset to model and estimate the decision to desert within the framework of a continuous-time game and finds the endogenous effect to be of importance. These papers provide further evidence on the importance of ties formed during war. I use the same data to examine the postwar outcomes of the Civil War veterans in Chapter 4. However, the advantage of examining units formed during World War I is that they were drawn from a far larger geographic base.[5]

3.2. The Effect of Social Networks on Labor Market Outcomes

3.2.1. Social Effect Mechanisms in the Labor Market

The importance of social networks as a source of jobs has been documented in many surveys. As mentioned in Section 3.1.1 various surveys report that anywhere from 18% to 70% of jobs are found through networks. However, there are several possible mechanisms

[5] As discussed in Chapter 4, most of the units during the American Civil War were based on people coming from the same town or nearby towns.

through which networks actually affect one's labor market outcomes. Some possible mechanisms through which networks may operate are information transmission, peer effect or social norms, and referrals or references.

Information transmission has been the explanation on which some of the recent literature has focused on. Boorman (1975) considers a framework in which agents learn of job openings and either take them, if unemployed, or pass them on to their contacts, if they already have a job. This succinct framework is sufficient for generating the prediction that the unemployment level of one's group has an adverse effect on one's own likelihood of employment. I present a simple model that yields the same result in Section 3.2.2. This type of mechanism is used by Calvó-Armengol and Jackson (2004) to model a system which delivers persistence in unemployment levels among groups. The striking feature of their model is that very small differences in the initial conditions can lead over time to large differences between groups. The informational aspect of networks was used by Topa (2001) to explain the clustering of unemployment within Chicago neighborhoods. Whereas Topa uses a probabilistic approach for the likelihood of a contact (which allows for a "spillover" of information across census tracts), in my data I am able to observe actual contacts.

Note that in the informational case networks can be of value even if all agents are identical and their utility does not depend on that of others. Networks can reduce the search frictions in the labor market via transmission of information.

While my work is motivated by the role networks play as a source of information about jobs, it is important to note that there are other mechanisms, not necessarily contradicting. One such mechanism is that of peer pressure or social norms. For example, those unemployed might enjoy consuming leisure more if their peers are unemployed. Alternatively, there might be an associated shame in work (or in unemployment) within a peer group. One could derive utility functions in which an agent seeks to confirm to the behavior of the group (e.g., Akerlof, 1997). Stutzer and Lalive (2006) find that the differences across communities in the belief of what the appropriate level of employment benefits is are correlated with the duration of unemployment. However, as further discussed in Section 3.4, the nature of both the World War I sample used in this chapter, and the Union army sample used in Chapter 4 make this less likely to be an important mechanism. Given that the men in both samples are white working-age males suggests it is more likely that unemployment is involuntary. However, social norms or peer effects may play a role in the retirement decision of member of the Union army sample, or in the choice of occupation.

In the case of referrals or references, employers might use their current workers as a source of information to reduce the uncertainty regarding the quality of a prospective employee (see Montgomery (1991) for example). Greenwald (1986) presents a model that considers the moral hazard that a worker faces when deciding to refer a potential candidate. It is plausible that veterans acted as references for each other, or hired those who had served with them, as they had first hand knowledge of the ability and trustworthiness of their comrades. Given that I do not observe in the data the employer of each veteran, I cannot directly estimate the importance of this mechanism. My findings, that networks play an important role, could be in part due to the strong bonds that were formed, or the knowledge that veterans gained about their fellow comrades' abilities. This does not contradict the informational role networks play, but rather complements it. Without further data or further assumptions, one cannot separate out the two types of effects. For instance, empirically, a network in which information passes along very efficiently could be equivalent to a network which is less efficient, but in which every job tip is passed along with a strong recommendation for hiring.

3.2.2. A Simple Model of Networks and Unemployment

In this section, I present a very simple specification to illustrate how others' employment might affect one's own employment likelihood. This model is based on the notion that a social network provides one with a source of information regarding job openings. This mechanism of transmitting job information was first modeled formally by Boorman (1975). In Section 3.2.1 I briefly discuss other channels through which networks may operate and affect labor market outcomes.

Consider a simple specification. Assume there are $g = 1..G$ groups, each with n_g members $i = 1, 2..n_g$. Each individual, indexed by g, i has an outcome of interest y, say the binary outcome of being employed or unemployed, covariates x which affect the likelihood of employment, such as age, and an error term ϵ_i, a scalar capturing the individual unobservable characteristics and shocks to one's employment prospects.

Assume the following:

Assumption 3. *(i) The groups are exhaustive and mutually exclusive- Each agent belongs to one group and only one group.*

(ii) Symmetry and equal centrality- Each member is connected or influenced by all members of the group in the same way.

(iii) Utility from work- For all agents: $U(employment) > U(unemployment)$*.*

(iv) "Limited altruism"-

$U(employment) - U(unemployment) > U(helping\ a\ group\ member\ find\ a\ job) > 0$

The first assumption is made for estimation convenience. The second assumption could be relaxed, without loss of generality, to allow various weighting. For instance, one could weigh contacts by their geographic distance, a plausible assumption when considering transportation and communication costs in the beginning of the 20th century. Assumption 3.*iii* is conditional on being in the labor force. It is meant to capture the fact that unemployment is not a voluntary state. Implicitly, I assume that agents seek to maximize utility. If assumption 3.*iv* did not hold, we would either have no mutual help, or there would never be a stable outcome as each agent cares about its peers more than it cares about his or her own utility.

Timing:

Period $t = -\tau$: Each individual is assigned to one group of size n_g.

Period $t = 0$: Each individual is in one of two states $y = 0$ (unemployed) or $y = 1$ (employed), possibly as a function of events occurring during periods -τ until 0.

Unemployed individuals, of which there are u_g in the beginning of the period, search for a job and find one with the probability: $1 > s(x_i, \varepsilon_i) > 0$. x could contain such measures as age, occupation, and local labor market conditions. If an unemployed person finds a job, they take it (by assumptions 3.*iii* and 3.*iv*).

Concurrently, each employed individual might learn of one job opening with a constant probability $1 > a > 0$. They then (randomly) pass that offer to one unemployed individual (by assumptions 3.*iii* and 3.*iv*).

The probability in the beginning of period $t = 1$ that an individual is still unemployed is:

$$\Pr(y_{i,t=1} = 0 | y_{i,t=0} = 0) = [1 - s(x_{i,0}, \varepsilon_{i,0})] \cdot [1 - a \cdot \frac{n_{g,t=0} - u_{g,t=0}}{u_{g,t=0}}]$$

and the comparative static result is that:

$$\frac{\partial \Pr(y_{i,t=1} = 0 \mid y_{i,t=0} = 0)}{\partial(\frac{u_{g,t=0}}{n_{g,t=0}})} \geq 0$$

(with strict inequality if $s < 1$ and $a > 0$)

In the above framework, all else equal, an increase in the group's unemployment rate would decrease the likelihood that a member of the group finds a job.

Though succinct, the above model captures the availability of "two markets" that each individual faces. Those are referred to in the literature as the "formal" and "informal" markets. The first, the formal market, is captured by characteristics in x and the probability s. The informal market is affected by the parameters n_g and u_g. Even in this simple framework, the outcomes of a group are path-dependent in the sense that the current unemployment outcomes depend on the previous period's unemployment rate among group members. They also depend on the individual level shocks and the formal labor market conditions.

The model presented above is a mechanical one, as agents do not truly face a decision problem. This excludes the case in which an individual's choice depends on the group's (expected) choice. While this is a potentially important effect in many cases of social influence, it is less likely to be the case once conditioning on labor force participation, as

I will do in the empirical part. It may be the case that people would choose not to work or retire based on their peers' behavior. However, conditional on being in the labor force, I treat unemployment as an involuntary outcome. In a richer framework, where search effort is endogenous, one could further incorporate one's reservation rule to depend on those of his or her peers.

A major advantage of this "mechanical approach" is that it mitigates the problem of multiple equilibria and the need to make any assumptions regarding expectations and equilibrium, such as rational expectations. In contrast, if people face a decision which depends on their expectations of their peers' actions, for the model to be consistent they all must have the "correct" expectation of the actions their peer take. Further, when choice is involved, one could face instances in which there are multiple equilibria consistent with the model. If estimation is performed using a cross-section of groups, this could lead to biased results since the econometrician must, in addition to estimation, be able to decide which equilibrium is each group observed to be playing.

3.3. Data Collection and Description

This chapter uses a new dataset which I constructed from various sources. It consists of United States infantrymen who had served together during World War I in the 313th Infantry Regiment, Seventy-ninth Division. The core sample of men (*n=1,295*) were all drafted and had fought overseas. I focus on the military company as the individual's reference group and examine *all* those in his unit. There are two main features which are likely to make these groups a significant reference group. The first is the sense of affiliation, or unit pride. Second, all of the units I examine consist of men who trained together, shipped to Europe, and fought together overseas. Ties forged during battle are likely to be meaningful. The choice of company as reference group is further motivated in Section 3.4.

In addition to their military service records, the men were linked to two additional data sources, the 1930 United States Census of Population, and their prewar draft registration card. Finally, for each of the men linked to the Census of 1930, information about up to 60 of their nearest neighbors was collected. Tables 3.1 and 3.2 provide some summary statistics for the various samples.

The linked dataset allows one to observe those who had served together in the same military company (companies consisted of 100-200 men), observe their, and their neighbors' postwar outcomes in the 1930 census, while controlling for their prewar place of residence and occupation. The 1930 Census includes information on labor market outcomes, such as employment, occupation, and industry, housing market information, such as ownership and housing values, and various demographics, such as age, race, parents' place of birth, and immigration information. The military service records provide information on place of residence prior to enlistment, place and date of birth, ranks and promotions, citations and court martials, whether wounded, and the (military) company affiliation within the regiment. The draft registration records were used to obtain information on the men's occupation prior to enlistment. The data collection procedure is further explained in Section 3.3.2.

Table 3.1: World War I Veterans Sample Summary Statistics

313th Infantry Regiment Veterans Sample (*n=1,295*+)

Variable	**Mean (Std. Dev.)**
Age (in 1918)	25.67 (2.87)
Wounded during World War I (all levels, conditional on survival)	15.36%
Sustained severe wound during World War I (conditional on survival)	2.70%
Occupational income score* (prewar)	25.96 (8.85)
Occupational income score** (postwar)	27.89 (8.92)
Unemployed in 1930**	6.85%
In the labor force (1930)**	96.02%
Home owner (1930)**	62.34%
Married (1930)**	68.13%

Notes:
+ Size of the sample of those who survived the war.
* Score ranges from 3 to 80. Measure available only for those linked to the prewar draft records.
** Postwar measures are calculated using the sample linked to the 1930 census.
All men in the veteran sample are white, as the army in WWI was segregated.
Standard deviation in parentheses.

Table 3.2: Neighbors of World War I Veterans Sample Summary Statistics

Neighbors of Veterans in the 1930 Census (*n=31,678*)	
Variable	**Mean**
Age (in 1930)	30.26
Veteran Status (any war)	5.10%
Male	49.55%
Unemployed in 1930 Census	7.62%
Unemployment among males in the 1930 Census	8.14%
Married (1930)	45.46%

Note: Sample of those who were neighbors (defined as residing in the same block) of the veterans in 1930

3.3.1. Sample Design

The sample design, that is the choice of units to examine, focused on minimizing the possibility for a selection bias. There are several levels to that choice, choice of war (or era), choice of type of division, and last, choice of units.

The decision to focus on veterans of World War I was motivated by several factors and presents several advantages compared to other wars in which the United States took part. World War I is the first time a "modern draft" (Chambers, 1987) was instituted in the United States, most of the American army during World War I was comprised of drafted men, the American experience during World War I was a relatively easy one, and publicly available micro-level data exists for both the military experience as well as the postwar experiences of the men.

On April 1st 1917, prior to the draft, the Army's size totaled 281,880 men. By November 11, 1918, it had grown almost fifteen-fold to 4,185,220 (Maryland War Records Commission, 1933). Most of this increase, 2,810,296 men, came from those drafted. The World War I draft was the first American draft in which a systematic and comprehensive approach was taken (Chambers (1987) and United States War Office Provost Marshall reports, 1919 and 1920). The draft was administrated nationally, though the actual registration and selection followed state quotas and was administrated by local boards (*ibid*). The World War I draft was far more inclusive compared both to earlier experiences, such as the Civil War, and later wars, such as the Vietnam War. There were far less cases of "draft dodging" and exemptions based on socio-economic factors. I describe the draft in more detail in Section 3.3.1.1.

To be able to focus on the case in which groups are involuntarily formed, and to minimize the selection bias, I have chosen to look only at those units formed by the draft. World War I units differed greatly in the ways in which they were formed. Some units were part of the regular army, while others consisted of those drafted. The army divisions during World War I were formed in three primary ways. The Regular Army, divisions which were composed of (former) National Guard units, and the National Army which consisted of inducted (drafted) soldiers.[6] The drafted soldiers made up the divisions numbered Seventy-six and up. My sample is drawn from the Seventy-ninth Division.

The main advantage of looking at those drafted, in addition to being the largest portion of the army, is the involuntarily nature of the assignment to groups. This provides a rare opportunity to examine a case in which networks were formed due to an exogenous shock, President Wilson's decision to enter the war.

The second advantage of looking at World War I, is that while World War I's casualty rate for the European armies (and civilian population) was of great magnitude, the United States had relatively less casualties, as it joined the war very late (see Table 3.3). The United States military World War I's death toll, in both percent and absolute terms, was smaller than World War II (and certainly the Civil War). Table 3.4 presents a comparison for some of the major wars the United States participated in. World War I was also a relatively short war (less than 2 years for the American troops). For example, the men in my sample participated in battles (or held defensive lines) for a total of less than 60 days. Their experience is considered typical for an American combat unit during World War I.

The next level of choice was the division on which to focus. Divisions were largely separated across the different components of army (Regular Army, National Guard, and the National Army). Within the National Army (the drafted men), some of the divisions were based on men from a number of states, and others were based on the population at

[6]The regular army divisions (numbered 1-25) were the "professional" army. When the United States entered the war, there was a need for an enormous number of soldiers, a demand which could not be met by the existing regular army. The National Guard units were already organized groups, many with actual military experience (for example, in the Mexican border fighting in 1916) and a long legacy (Civil War, and the American-Spanish War). A decision was made to use them instead of recruiting and assembling new units. These units were "federalized" and were used to form divisions in the 26-75 number range. However, this still was not enough to meet the demand for soldiers, and a national draft was enacted.

Table 3.3: Cross-Country Comparison of Deaths in Battle During World War I

Russia	1,700,000
Germany	1,600,000
France	1,385,300
Britain	900,000
Austria	800,000
Italy	364,000
Turkey	250,000
Serbia and Montenegro	125,000
Belgium	102,000
Romania	100,000
Bulgaria	100,000
United States	**50,300**
Greece	7,000
Portugal	2,000

Source: Maryland War Records Commission (1933).

Table 3.4: United States Military Death Rates in Major Wars

	Died in battle	Percent died during service
Civil War (Union army only)	6.34%	29.20%
World War I	1.13%	6.76%
World War II	1.79%	6.60%
Korean War	0.58%	2.43%
Vietnam War	0.54%	2.42%

Source: Author's calculations based on The World Almanac and Book of Facts (2003).

large. For instance, the Seventy-ninth Division primarily included men from the District of Columbia, Maryland, and Pennsylvania. However, note that these divisions were not based on any existing units, and were newly formed during 1918. As divisions were being organized, men were constantly being transferred between divisions during the first few months prior to shipping overseas. Each division had, on average, more than twice as many men pass through it as appeared on the final roster when the division shipped to Europe.

Similarly, within the division, some regiments were formed with the intention of having men come from the same state or states. For example, the sample used in this chapter, the men of the 313th Infantry, were largely drawn from Maryland. In fact, they were sometimes nicknamed "Baltimore's Own," though only about 40% of them were actually from Baltimore.[7] I chose to focus on a regiment that was largely drawn from one state for two main reasons. The first was to circumvent some of the issues regarding the draft mechanism. Since all men were from the same state, they were all selected using similar criteria, and according to the same probability.[8]

Second, in the case in which men are all from the same region, there is a higher likelihood that the bonds formed during the war would be of use postwar. For instance, a farmer who had lived in a village in Maryland, would be more likely to move to Baltimore to take up a job found through his fellow unit member than he would if that job opportunity was in Washington state. However, as shown in the next section, there was enough variation within each group that the men of each unit did not all come from the same place in Maryland.

The sample I have selected belonged to the Infantry Branch. One might be concerned that men were selected into different military occupations based on their ability or skills.

[7]In 1917, Baltimore's population was approximately 700,000 and the size of the next largest town in Maryland was about 40,000. Approximately half of Maryland's population lived in Baltimore. (In the 1920 Census, Baltimore City's population stood at 733,826 and Maryland's at 1,449,661.) Sources: Gibbons (1998) and Forstall (1996).

[8]Conditional on being a civilian, white, male, citizen, and with no major health problems, the actual likelihood of being drafted was determined by a lottery, and hence, was random. However, the probability slightly varied across the states, as different states had different quotas and different pool sizes.

It is important to note that the military during World War I was far less specialized in its occupations, training, or needed skills than it is today or even during the Second World War. There was little, if any, screening or testing on which assignment was based. The bulk of the fighting men were infantrymen. The infantry was by far the largest branch with over 1,000,000 men in 1918; the next largest branch was field artillery with almost 400,000 men. In comparison, there were only 30,000 men in the Cavalry (Maryland War History Commission, 1933). In addition, the training of the infantry did not vary much from that of, say those in the artillery. The skills acquired during the war, especially those transferable to the civilian labor force, were fairly homogenous across military professions, and quite limited. In my sample, men had trained in the United States for only a few months before shipping to France.

Last, a distinction can be made according to whether or not the divisions were sent oversees and participated in battles or not (as well as a more continuous measure of the number of days spent oversees and number of days spent in battles). However, this is not a concern in terms of sample selection since the magnitude of the war experience does not depend the manner or type of division. All three types of divisions (National Guard, Regular Army, and National Army) were deployed overseas.

3.3.1.1. The World War I Draft and the Seventy-ninth Division. In this section I provide a brief overview of the United States World War I draft as well as the experience of the Seventy-ninth Division from which my sample is drawn.

As mentioned above, within a period of less than 20 months the United States army had grown almost fifteen-fold and numbered over 4 million men by the end of the war

(Maryland War Records Commission, 1933). The majority of this increase, 2,810,296 men, came from those drafted.

My discussion of the draft mechanism draws on Chambers (1987) and the United States War Office Provost Marshall reports (1919 and 1920). On May 18, 1917, the Selective Service Act was passed. The draft was administrated nationally, though the actual registration and selection followed state quotas and was administrated by local boards. District boards were established by the president (roughly corresponding to the Federal Judicial Districts). Each district board was in charge of the local boards. Local boards usually followed county lines. In large counties and cities there was a local draft board for every 30,000 people. For example, the city of Baltimore was covered by 24 local boards.

Registration was mandatory for all males, regardless of citizenship status. There were three main registration drives. The first registration took place on June 1917 for all men ages 21 to 30. The second registration (June 1918) added all those who turned 21 since the first registration. The last registration took place on September 1918 and included all men aged 18 to 45. In total, 23,908,576 men were registered in a period of less than two years (War Office Provost Marshal Final Report, 1920). In my sample, all but a few men were registered during the first registration period.

Each registrant was assigned a number. Numbers where then randomly drawn in Washington D.C. The first numbers were actually drawn by President Wilson who was blindfolded. Each state then called the men corresponding to the drawn numbers. In Section 3.3.3 I show that men's assignment to companies is consistent with random assignment.

According to the Provost Marshal's report, of the more than 23 million men registered only 1.4% had attempted to desert or avoid their call, of which about half were apprehended and punished (*ibid*). While this number could be downward biased, as people could attempt more subtle ways of avoiding the draft (such as faking a medical condition), by in large the World War I draft suffered from little public opposition with the exception of very localized riots (Chambers, 1987). The draft was also more inclusive with respect to socio-economic background. For instance, unlike the Civil War draft, the World War I draft did not allow conscripts to purchase substitutes (*ibid*). In contrast, consider the Vietnam War era in which the ratio of conscientious objectors to Selective Service call-ups was 18%.[9]

As mentioned in Section 3.3.1, the United States army had an easier experience than that of other nations. Table 3.5 provides a brief summary of the timeline the men in my sample had experienced. The men in the sample had spent less than two years together, of which less than one year was spent overseas. The total actual combat experience was less than two months.

I conclude this section with Table 3.6. The table provides summary statistics of the experience of the 313th Infantry Regiment, the regiment on which my sample is based.

3.3.2. Data Collection and Linking Procedures

The core component of the data is the unit affiliation of each person at the (military) company level. Using unit rosters and service records, the list of men in each unit was

[9] According to my calculations based on McAdam and Su (2002).

Table 3.5: Timeline of the Experience of the Seventy-ninth Division

May 18, 1917	Congress authorizes the formation of the National Army. (draft legislation)
August 25-29, 1917	Brigadier-General Kuhn and officers arrive at camp Meade, Maryland.
September 19, 1917	First group of selected (drafted) men arrive at Camp Meade.
September 1917- June 1918	Organization of units and training.
July 8, 1918	Division commences overseas movement.
July-September 1918	Training in France.
September 12, 1918	First units of division entered the line.
September 1918 – November 1918	Participate in various battles. (Held defensive sectors for 34 days and offensive or active sectors for 19 days.)
November 11, 1918	Armistice signed.
May 1919	Division sails home from France.

Sources: Barber (1922) and Thorn (1920).

created.[10] In addition, these sources provide information on the date of birth, place of birth, address prior to enlistment, and the service record during the war. The service records include information on dates served in each unit (including enlistment and discharge date), campaigns in which the soldier participated, wounds, citations, disciplinary action, ranks and promotion dates, and a characterization of the type of discharge.

[10]After World War I, each state was required by Congress to establish a War History Commission. The Commissions were provided with the service records by the Federal Government. For instance, in the case of Maryland, the records of the 70,000 who had served during WWI were published by the Commission (Maryland War Records Commission, 1933). I have obtained the records for a number of states. In addition, many units published their own history.

Table 3.6: The War Experience of the 313th Infantry Regiment

Infantry Battalions of the 313th during World War I	
Days in service	300-700
Days overseas	~300 days
Days spent in battle (both defensive and offensive campaigns)	~50 days
Died as a result of battle	9.87%
Died from disease during war	2.01%
Died in an accident during war	0.16%

Source: Author's calculations from collected veterans sample and Thorn (1930)

These sources provide a very detailed account of the experience of each unit (and the men within the unit). Since men were moved and transferred between units, a soldier might have been affiliated with a few units. The minimal criteria I used for including a veteran in my sample was to actually have shipped with the unit overseas (the units examined are all infantry units which had fought overseas). If a soldier had spent only a few days at a unit it is unlikely that this would be a sufficient amount of time to allow for any meaningful ties to form.

Next, every person in the units chosen was searched for in two sources, the 1917/8 draft records, and the 1930 United States Census of Population.[11] The last step of the data collection process included obtaining the 1930 census information of a veteran's neighbors. The census schedules from which the information was collected were recorded

[11]Both the draft registration records and the 1930 census are publicly available through the United States National Archives.

in the order in which the census taker had visited the households. Hence, finding one's neighbors in the 1930 census is straightforward (the schedules also include the street and number address for the larger towns and cities).

Each match between a veteran and his (potential) 1930 census information was assigned one of 20 categories. For instance, one of the categories include those for which a veteran is found to be living at the same exact house and street in 1917 and 1930, and has the "correct" age in 1930. On the other extreme, imagine a situation in which the names are spelled slightly differently, and the age in 1930 is not consistent with the age reported in 1917. Categories were then assigned quality codes and ranked (quality wise).

As in all historical sources, matching people across datasets presents some complications. However, relative to other researchers who had linked individuals across historical censuses, in my case there was more additional relevant-to-linking information on each individual. For example, many of the military records had a full date of birth (day, month, and year), which helped identify someone in the draft records (the draft records have a full date of birth). In the 1930 census, there was a question regarding participation in World War I, which, again, greatly aided in linking. Other information which proved useful was place of birth, as well the house number and street address (for those cases in which a person had stayed in the same exact address). These factors account for my relatively high match rates.

The linkage rates for each of the sources when considering only the "high quality" matches is 64% for the 1930 census and 75% for the draft records.[12] These linkage rates

[12]The first number is in fact downward biased, since I did not account for the fact that some of the men had died or immigrated from the US by 1930.

Table 3.7: Reasons for Non-Linkage to the 1930 Census

"Extremely" common name	6.5%
Two people with "too similar to decide" information	8.5%
Not found	85.0%

are comparable to the rates reported by Viechnicki (2003) who examined a sample of Union Army veterans that were linked to the censuses of 1850 through 1910.[13]

Table 3.7 provides a breakdown of the reasons veterans were not matched to the 1930 census. In 85% of the cases, the reason for not linking a veteran is that no possible match is found (names are searched for in the entire United States census). The two most likely reasons a person is not found in the 1930 census are either death, or the person was missed by the census takers. The records were also searched according to the Soundex system, which allows for variation in spelling. The extremely common name reason (6.5% of non-linked cases) corresponds to a case in which there were over 20 people with the same name in the 1930 census. Finally, 8.5% of the non-linked records were a case in which two people were indistinguishable from one another (same name, same birth date and birth state, etc.).

The only statistically significant variable (and only in some specifications) in predicting non-linking is age, though the marginal effect is small (less than half of a percentage point per an additional year of age). This is consistent with mortality rates increasing in age.

In addition, to insure higher reliability, both the linking and transcribing of the 1930 Census data were independently done by at least two people. All discrepancies were

[13]The Union Army Study is a monumental data collection effort led by Fogel et al. (2000). Its goal was to study and better understand aging and mortality. It contains the members of 303 companies in the American Civil War Union army. The sample was then linked to the census of 1860, 1870, 1880, 1900, and 1910, as well as other pension records. I use this sample in Chapter 4.

resolved by a third person. In the case of linking, only one discrepancy was found. In about 10% of the cases, one linker found a match which the other had missed (or was too cautious to link). This is partially a reflection of the fact that linkers were instructed to err on the side of not linking as opposed to linking the wrong person.

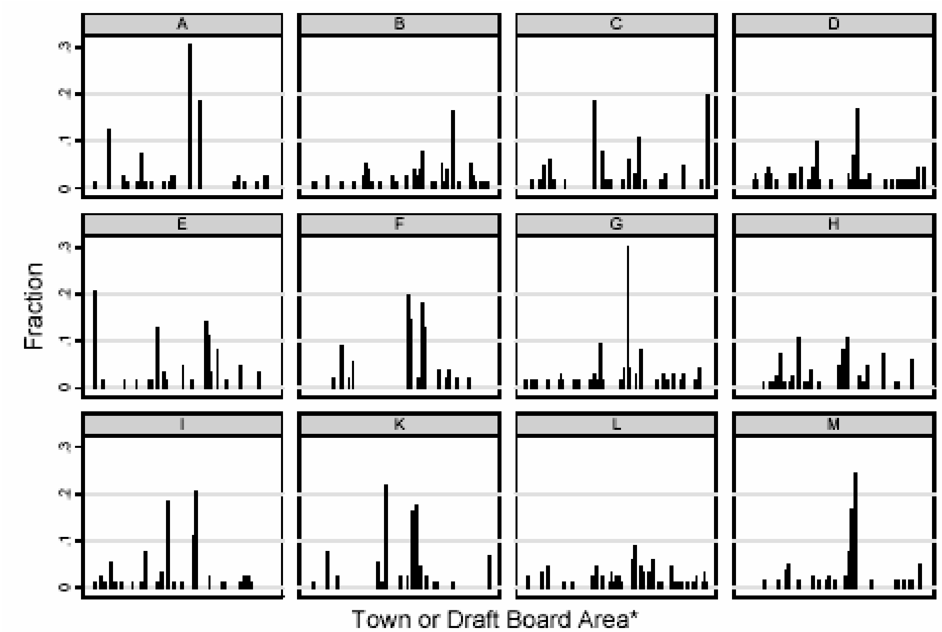
3.3.3. Assignment to Groups and Sample Representativeness

Since I focus on units which were largely drawn from the same geographic region, one might be concerned that assignment within the regiment at the company level might be correlated with the draftees' place of residence. I show that any correlation with place of residence was limited to a portion of the company. Figure 3.1 presents the distribution across all towns for all those who were from Maryland, for each of the 12 companies in the 313th Infantry Regiment. In the case of Baltimore, I used the draft board areas (draft boards were assigned geographically, and in denser areas, to every 30,000 people). Note that in no case does the fraction of men from one town exceed 30%. In two-thirds of the cases, the largest fraction does not exceed 20%.

To illustrate, consider all the men who had lived in the last 5 blocks (2500-3000) of Woodbrook Avenue, Baltimore, Maryland.[14] In total, 25 men from those 5 blocks served during World War I, of whom 14 were drafted. Of the 14 drafted men, only 5 of them ended up in the 313th Infantry Regiment (the others were not even in the same division). Of the 5 who ended up in the 313th infantry, one served in the Headquarters Company, one served in company G, and two served in company I (the fifth never shipped overseas and was transferred to a different unit). Within company I, while it is likely the two

[14] I arbitrarily chose Woodbrook Ave. since it was the last street in the alphabet from which men from the 313th Infantry Regiment had come.

Figure 3.1: Within-Company Distribution of Town of Residence Prior to Enlistment



Notes:
Sample: Those in the 12 Infantry Companies of the 313th Infantry who were from Maryland (*n=941*)
Fraction of company members from each town reported on the vertical axis. (Town location on axis is the same across companies)
*194 towns in Maryland.
Due to the size of Baltimore (20 times larger than the next town), Baltimore was partitioned into draft boards (draft board size was approximately 30,000 people).

knew each other (as they lived two blocks from each other before enlistment), only 8 others in that company were from the same draft board that covered an area of 30,000 people. The rest of the company consisted of 50 men from Baltimore,[15] and another 52 from one of 14 towns in Maryland, and yet another 5 were from Ohio. Those who had come from small towns and villages were even less likely to have prior connections with

[15]In 1917, Baltimore's population was approximately 700,000 and the size of the next largest town in Maryland was about 40,000. Approximately half of Maryland's population lived in Baltimore. (In the 1920 Census, Baltimore City's population stood at 733,826 and Maryland's at 1,449,661.) Sources: Gibbons (1998) and Forstall (1996).

others in their companies. For instance, in the same Company I, there were only 3 men from Hagerstown (Washington County), the third largest city in Maryland at the time.

Since the nature of the assignment is crucial, I further investigate its properties. Table 3.8 reports the results of a chi-square test for the null hypothesis that the number of people with a certain characteristic were randomly assigned among the companies. I examine all prewar variables for which there are on average more than 5 outcomes for each company. These variables include age, place of birth, parents' place of birth, and prewar marital status. The last two variables, killed or wounded during war, and died as a result of battle, are included as evidence on the similarity of wartime experience these companies had experienced. For none of the 17 variables is the null rejected at the $p<.1$ level, suggesting that these characteristics are consistent with random assignment. Table 3.9 reports the results for two variables which have multiple possible values, the occupational income score prior to enlistment, and age. The occupational income score ranges from 3 to 80, and is a measure of how well-paying a job is.[16] A Kruskal-Wallis rank-sum test does not reject the null hypothesis that the distributions are consistent with random assignment for those two measures (p-values of 0.47 and 0.33 respectively).

In Table 3.10, I examine the representativeness of the sample by comparing the occupational income score of my sample prior to enlistment, to two other samples based on the 1920 Census. The first of the two is for all white males in the same age bracket, 21-30, from Maryland. The second is the same demographic group from the entire United States. The companies and the 1920 sample population seem very similar. This is not all

[16]This measure is based on the median wage of an occupation in 1950. While the measure is not a perfect proxy, the purpose here is to examine the distribution. The results would still hold even if the scores of some occupations had changed over time.

Table 3.8: Chi-Square Test of Characteristics Consistent with Random Assignment to Companies

313th Infantry Companies Sample in 1917/8	
Variable (binary)	p-value*
Inducted in Maryland	0.467
Not married before enlistment	0.547
Prewar occupational score greater than 31 (the cutoff for the top quartile for the entire population in the census of 1920) **	0.575
Born in Maryland	0.660
At least one parent born outside of the US	0.567
Both parents born in the US	0.431
Born in the year 1888	0.140
Born in the year 1889	0.468
Born in the year 1890	0.710
Born in the year 1891	0.413
Born in the year 1892	0.174
Born in the year 1893	0.450
Born in the year 1894	0.211
Born in the year 1895	0.743
Born in the year 1896	0.111
War Experience	
Killed or wounded during war	0.528
Died as a result of battle	0.192

Notes:
* Null hypothesis tested: assignment to companies consistent with random assignment.
** Information only available for those linked to draft records.

Table 3.9: Kruskal-Wallis Test of Characteristics Consistent with Random Assignment to Companies

313th Infantry Companies Sample in 1917/8	
Variable	p-value*
Birth year	0.47
Prewar occupational income score** (for those born before 1895, that is, beyond college age)	0.33

Notes:
This is a nonparametric (rank sum) test for “equivalence” of distribution.
* Null hypothesis tested: distribution is consistent with random assignment.
** Information only available for those linked to draft records.

Table 3.10: Comparison of Occupational Income Score for Sample and 1920 Census

	Mean (Std. Dev.)	Min	Max	First Quartile	Median	Third Quartile
My sample of veterans in 1917/8 (prewar)	25.96 (8.85)	4	80	22	25	32
All white males from Maryland ages 21-30 (1920 Census)	24.88 (10.23)	4	80	20	24	32
All white males ages 21-30 (1920 Census)	23.42 (10.21)	3	80	14	23	30

Notes: The last two rows were calculated from the 1920 Integrated Public Use Microdata Series (Ruggles et al., 2004).
Occupational income score ranges from 3 to 80 (highest income).
Standard deviation in parentheses.

that surprising considering that the draft included a large part of the United States male population in that age bracket.

3.4. Empirical Strategy

3.4.1. Econometric Specification

In Section 3.1.1 I survey some studies which had documented the importance of the "informal" channel, that is jobs found through friends and relatives. Section 3.2 discusses some of the possible mechanisms through which networks may affect labor market outcomes. The primary goal of this chapter is not to model why peer effects might be influential, various authors have suggested different rationales, and one such example was presented in Section 3.2.2, but rather to better understand empirically the channels through which social effects operate in the labor market.

In Section 2.2 I discuss some of the issues surrounding identification of social interactions. In this section, I discuss the specifications that are estimated in this chapter, as well as the assumptions behind those specifications.

Using the notation from Section 2.2, assume there are $g = 1..G$ groups, each with n_g members $i = 1, 2..n_g$. The econometric specification discussed in Section 2.2 is written as:

$$y_{i,g} = h[\alpha + x_{i,g}^{'}\beta + Z_g^{'}\gamma + \rho \cdot m(\overrightarrow{y}_{-i,g}) + \epsilon_{i,g}] \tag{3.1}$$

Equation (3.1) in some form or another is used in virtually every empirical study of peer effects. The identification of this model is discussed in Section 2.2. To illustrate using the empirical specification, each individual, indexed by i, g has an outcome of interest y, say the binary outcome of being employed or unemployed, a vector of covariates x which affect the likelihood of employment, such as age, occupation, and local labor market conditions, and an error term ϵ_i, a scalar capturing the individual unobservable characteristics and shocks to his or her employment prospects. In addition, each individual's job prospects might depend on the groups characteristics summarized by the vector Z_g, and the outcomes of all other members in the group $\overrightarrow{y}_{-i,g}$. Some mechanisms that would predict that networks affect employment outcomes were presents in Section 3.2. These would correspond to a finding that either γ or ρ (or both) are different than zero.[17]

γ is often referred to in the literature as the contextual (or exogenous) effect, and ρ as the endogenous effect (see Section 2.2 for more details). It is possible, and in most instances quite likely that Z_g depends on the characteristics of others. For instance, Z_g might just be the average of the group characteristic, $Z_g = \frac{1}{n_g}\sum_{i \in g} x_i$. This specification already incorporates several assumptions and restrictions. The specification in equation (3.1) assumes that the various arguments are separable and linearly additive in the function $h[\cdot]$, and that all of the group's endogenous effect is aggregated or summarized through the function $m : \mathbb{R}^{G-1} \rightarrow \mathbb{R}^1$. For instance, when considering social norms, one could motivate the form $m = E[y_i | i \in g]$, that is, one's utility depends on the expectation of other's behavior. Alternatively, in a more mechanical model, such as the one presented in Section 3.2.2, where networks just provide information on job availability, if those employed are more likely to have such information, again, one's own probability of employment will depend on the expected outcome among others.

The aggregation function, $m : \mathbb{R}^{G-1} \rightarrow \mathbb{R}^1$, imposes the restriction that all group members have the same effect. Alternatively, one could introduce a weighting scheme into the function. However, since the data I use do not provide any information on the

[17]This is assuming the model is identified, and that one can separately identify the two, as discussed in Section 2.2.

strength of ties between different group members, I weight equally every member in the group. As shown in Section 2.4, using the Multiple Reference Groups method, one can relax this restriction and allow different types of groups to have different types of effect on an individual. For example, one can allow the effect of the war comrades to be different than that of one's neighbors.

In the empirical part of the chapter, m_g will be replaced with $\bar{y}_g = \frac{1}{n_g}\sum_{i\in g} y_i$, the average outcome among group member, or the same measure, self excluded, $\bar{y}_{g,-i} = \frac{1}{n_g-1}\sum\limits_{j\in g, j\neq i} y_j$. As group size increases, from an empirical standpoint, there is little difference between the two measures.[18] Also, note that if $\rho \neq 0$ then $\bar{y}_{g,-i}$ does contain y_i in it (since $y_2, y_3..y_{i-1,}y_{i+1}..y_{n_g}$ each depend on $y_{i,g}$). Hence, using the self-excluded group average instead of the group average does not eliminate the bias.

For both measures, in the case of small group sizes, the use of either measure ($\bar{y}_g$ or $\bar{y}_{g,-i}$) introduces problems in the coherency of the model, depending on the functional form used. (See Heckman (1978) for example.) The issue of coherency is related to whether or not individuals are affected by a latent group statistic. The groups I consider in my data are large enough to assume that the sample analog is a measure of the expected value and that the members do not make their choices by calculating or anticipating the unemployment outcomes of each and every member of the group.

[18]However, if group size exhibits a large variation one could use the different group sizes to separately identify the endogenous and contextual effects (Lee, 2006). In which case for the purpose of identification, the exclusion or inclusion of own-outcome in the group average could be crucial for identification, and identification fails if own-outcome is included (see the discussion in Bramoullé et al., 2006).

3.4.2. Reference Groups Formed During War

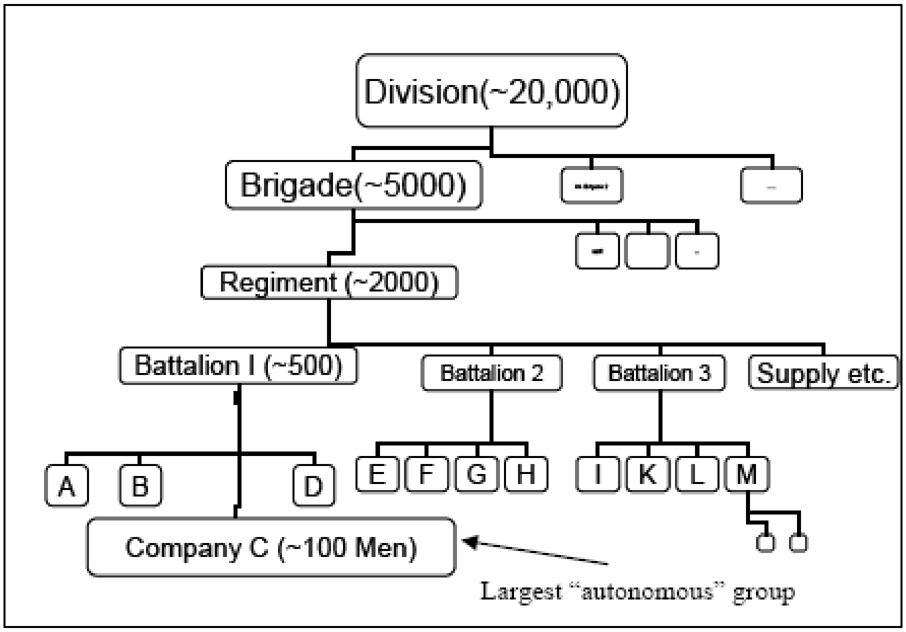
The military group used as the reference group in most of this chapter is the military company in which each veteran served during World War I. There are several advantages which groups formed during military service provide. These include the circumstance and manner in which the groups were formed, scope and size of reference group, and strength of ties.

From a methodological point of view, the major advantage of examining a military setting is the way in which companies were formed. Unlike most other settings, formation of networks in this case was involuntarily and due to an exogenous shock.[19] Further, the nature of the experience was likely to create strong bonds. All the men examined served overseas and participated in one or more campaigns during the war. The men did not only spend all of their time with each other, but also depended on one another and had to develop the ability to work as a team. At times their lives depended on the actions of their comrades. Unit spirit and pride were also encouraged by the military, as a way of building unit cohesion.

From the various levels of hierarchy, I have chosen the military company. Figure 3.2 illustrates the organizational structure of a typical infantry division in World War I. In the infantry, each company is composed of 4 platoons. The full strength size of a company was 100-200 men. Every four companies made up a battalion. There were 12 infantry companies (3 battalions) in each infantry regiment, in addition to other support elements such as the Headquarters and Supply company. A division consisted of many regiments,

[19] In Section 3.1.1 I discuss some studies which examine groups formed involuntarily.

Figure 3.2: Infantry Division Organizational Chart



Notes: Number of men in parentheses.
Sizes are approximate and are for illustration of relative magnitude.

of which four were infantry regiments. (A full strength division had more than 20,000 men in it.)

During World War I, for the 313th Infantry Regiment I examine, the company level was almost always the largest autonomous unit.[20] While it was possible for different companies in a battalion to be assigned to different locations or tasks, the companies themselves were almost never divided and were always assigned as a group to a task. In addition, the company viewed itself as a cohesive unit of reference. For instance, after the war many companies published their history. Similarly, the most detailed level in the army's records is the company level. For instance, the monthly unit rosters and daily

[20] This is based on the detailed account of Thorn (1920).

reports are all at the company level, and the administrative records contain no reference to a more detailed level, such as the platoon (there are 4 platoons in a company).[21]

In addition to companies being cohesive units, from the point of view of network analysis they have an appealing size property. Much attention has focused on the importance of "weak ties." For instance, in the case of labor markets, Granovetter (1974, 1995) finds that it is weak ties, such as a neighbor, as opposed to strong ties, such as a brother, which are responsible for most of the jobs found. Networks at the company level contain relatively large number of individuals, increasing the probability that while not all veterans become best friends, they would still be of value as a "weak tie," and hence more likely to have an effect significant enough to be picked up. Companies are a small enough group to allow its members to know each other. Yet, they are large enough to allow for a substantial number of potentially useful ties. In Section 3.5.3, I look at a larger, less finer, group level (the battalion, which includes several military companies), and find that it does not have a statistically significant effect after controlling for own company. This further suggests that the military company is an appropriate reference group to examine.

3.4.3. Postwar Interactions Among Veterans

In the sample I consider it is certain that the men had direct contact for a period of at least one year. The outcomes I analyze are drawn from the men's 1930 census information, and are therefore a decade after the men had first met. I do not have any direct information on

[21] I was able to view the actual monthly rosters and daily reports through a Freedom of Information Act request. The records are stored in the National Personnel Records Center which is part of the United States National Archives and Records Administration.

any actual interactions that occurred among the veterans once the war was over. While this does not affect the validity of the analysis and the results, it is important to examine whether social interactions were plausible.

The geographical nature of the sample suggests that interactions were feasible. For instance, about half of the sample lived in Baltimore in 1930. Hence, for a large part of the sample, the contacts they had made during World War I, were available, at least geographically, if the veterans wanted to make use of them. Place of residence in 1930, like any migration decision is endogenous. Therefore, one must be cautious of using geographical weights for the networks measure. In most of the chapter, the outcomes of all group members are equally weighted.

The second type of indirect evidence of social interactions during the postwar years is the existence of unit-specific veteran associations. Many units held reunions and published yearly newsletters and directories.[22] For the units I examine in my sample I am aware of several formal organizations and associations.

The Seventy-ninth Division, to which all the men in my sample belonged to, had established an Association. It was officially established on May 7, 1919, as the unit was preparing to return home (Barber 1922, 510). It held its first convention in Baltimore on September 3, 1921, and its headquarters were in Philadelphia. The stated purpose of the association was: "To perpetuate the achievements of the Seventy-ninth Division, in history and in tradition; to promote fellowship among comrades who gave their all for their country."(*ibid*) The president of the association, H. Harrison Smith, also acknowledged that: "Many of the Regiments, Battalions, and even Companies within the Division have

[22]The Maryland Historical Society have in their collection a photograph taken in 1923 at the Reunion Banquet of the 313th Infantry Regiment, the regiment used in this chapter.

formed their separate organizations or associations for just this very purpose." Smith had set forth a plan with ten goals, of which the third was: "Creating an efficiency in the matter of bringing up to date and keeping current the rosters of the different individual units."(*ibid*)

In the opening paragraph of the introduction of the published division history, the commander of the division, Major General Joseph E. Kuhn wrote: "This history has been prepared primarily for you in order to preserve the ties of comradeshiop formed during strenuous days of training at home and stirring incidents of campaign abroad."(Barber 1922, 7)

The formal veteran organizations and the postwar geographic proximity of the veterans suggest that social networks formed during the war, could have been utilized during the postwar years.

3.5. Results

Recall that the goal is to estimate a model of the form:

$$y_{i,g} = h(\alpha + x_{i,g}^{'}\beta + Z_g^{'}\gamma + \rho \cdot m_g^e + \epsilon_{i,g}) \tag{3.2}$$

where m_g^e is replaced with $\bar{y}$ or $\bar{y}_{-i}$, as described in Section 3.4.

I start by presenting the results for the "naive" case under the assumption that there is no exogenous effect ($\gamma = 0$). I then address the simultaneity problem by using two approaches. First, in Section 3.5.2, I consider a reduced-form specification which uses far less assumptions. I show that the group effects are statistically significant, though the magnitude of the results do not necessarily have a straightforward economic interpretation.

I then instrument the group's average outcome. Those results are in Section 3.5.5. For robustness purposes, Section 3.5.3 tests whether an alternative sample and alternative reference group specifications have a statistically significant effect. The results suggest the company-level group specification is a meaningful specification to consider. Various controls for the place of residence prior to enlistment are presented in Section 3.5.4. Even when examining only those who had very few others in their company come from the same place, the company's unemployment rate is still found to be statistically significant. In Section 3.6, I present an application of the Multiple Reference Groups framework which allows one to estimate both the endogenous and contextual effect.

3.5.1. The "Naive" Estimator

I first consider the base case in which it is assumed there are no contextual effects, that is $\gamma = 0$. This is the approach taken by many empirical studies, as it circumvents the need to deal with some of the identification issues presented in the previous sections.[23]

Even under the assumption that the unobserved term is uncorrelated with individual characteristics, the coefficient ρ is biased for finite group sizes.[24] However, it is of use to examine this base specification as a point of departure.

After controlling for personal characteristics, as well as occupation and county fixed effects, the group's unemployment rate has a statistically and economically significant effect. In the case of the veterans sample, an increase of 1 percentage point in one's (military) company unemployment rate, decreases one's probability of being employed by almost 0.4 percentage points (the average employment rate in the veteran's sample is

[23]Many times, there is little justification given as to why one might believe there is no contextual effect.
[24]For example, in the linear-in-means case $\text{cov}(\epsilon_i, \bar{y}) = \frac{1}{n_g(1-\rho)} Var(\epsilon_i)$.

93%). The marginal effects reported are calculated at the average rate, and are based on the point estimates.

Table 3.11 reports the results of one's own likelihood of being employed in the 1930 Census as a function of various control variables. The functional form used in these tables is the probit. Similar results (not reported) hold for the logit functional form. In Table 3.11, I present the results for the baseline specification for the sample of World War I veterans. The last columns include the "naive" specification in which the average unemployment rate among company members and those living on the same block are included without controlling for the simultaneity. The next sections present various alternatives for addressing the issue of simultaneity.

I now turn to a more detailed description of the results. Columns 1-4 contain only one's own covariates, and columns 5-7 examine the effect of various measures of unemployment among group members in addition to one's own covariates. In the case of the World War I veterans, as all were in their thirties and forties in 1930, I have included a dummy variable of being over 40 instead of age, though the results don't change much if age and age-squared are used instead. The variable wounded denotes whether the veteran was wounded during the war. Most of the wounds were classified as "slightly wounded," and the effect of having been wounded on employment does not significantly depend on the severity of the wound. All the men in the veterans sample were enlisted men, and I include controls for their rank (the ranks are private, private first class, corporal, sergeant, staff sergeant and supply or mess sergeant).

Column 1 includes one's age, whether head of household (which in the veterans sample is highly correlated with being married, $r = .86$), and controls for rank and being wounded

Table 3.11: World War I Veterans' Employment Including the Effect of Peers' Employment in 1930

World War I Veterans Sample in 1930; Probit Specification; Dependent variable- employment							
	1	2	3	4	5	6	7
Unemployment rate among company members						-9.811*** (1.171) [-0.391]	-10.323*** (3.012) [-0.377]
Unemployment rate among neighbors							-4.049*** (1.041) [-0.148]
Percent of company members not working					-7.464*** (1.282) [-0.298]		
Is over 40	-0.581*** (0.218) [-0.093]	-0.722*** (0.259) [-0.065]	-0.707** (0.322) [-0.048]	-1.128** (0.500) [-0.236]	-0.776*** (0.239) [-0.06]	-0.784*** (0.237) [-0.061]	-0.865*** (0.280) [-0.067]
Head of household	0.159 (0.183) [0.019]	0.171 (0.209) [0.009]	0.418* (0.247) [0.019]	0.425 (0.325) [0.052]	0.211 (0.237) [0.009]	0.202 (0.236) [0.009]	0.3 (0.225) [0.013]
Corporal Rank	0.656** (0.301) [0.055]				0.609** (0.290) [0.017]	0.584** (0.290) [0.017]	0.493 (0.344) [0.014]
Sgt. Rank	0.149 (0.245) [0.015]						
Mess or Supply Sgt.	-0.146 (0.412) [-0.018]						
Wounded during WWI	-0.073 (0.254) [-0.009]						
Lives in Baltimore		-0.076 (0.196) [-0.004]					
Constant	1.399	6.908	12.158	6.819	7.961	8.075	8.225
Rank, and whether wounded controls		Yes	Yes	Yes	Yes	Yes	Yes
Occupation controls		Category	Category	Every Occupation	Category	Category	
County fixed effects				Yes			
Observations	563	517	300	170	529	529	494
Pseudo R-squared	0.06	0.17	0.20	0.31	0.20	0.21	0.28

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)

during the war. Being over 40 reduces one's likelihood of employment by 9 percentage points. Household heads are slightly more likely to be employed, though this effect is rarely statistically significant in any of the specifications.

In regards to rank, those who had reached a higher rank were more likely to be employed. However, this effect reaches its peak for the rank of corporal and decreases for sergeant. Further, the effect for those who were Supply or Mess Sergeants is negative (though not statistically significant). In all specifications, the largest, and usually the only statistically significant effect is for corporals. Those who had reached the rank of corporal during the war are 1-5 percentage points (depending on the specification) more likely to be employed in the 1930 Census. This is consistent with promotions not being random, but rather being correlated with some ability, which possibly could be transferred to the civilian labor market.

Since all men in the sample participated in battle, I examine whether being wounded has an important impact on the employment likelihood in 1930. Those wounded are slightly less likely to be employed, though the effect is less than one percentage point. In the specifications I have tried, being wounded does not have a statistically significant effect on employment. This is true even when looking at those severely wounded. Severe wounds do not even seem to affect one's likelihood of being in the labor force (or even of survival until 1930). Note that there are several possible reasons for this finding. First, the percent of those in the sample who were severely wounded is less than 3%. Those slightly wounded might not be affected by the wound. Second, there might be a selection bias involved, and we only observe those who had survived, and therefore recovered, whereas those who didn't survive (or recover) would not be in the sample (or the labor force).

Column 2 includes occupation category (one of twelve) as well as a dummy variable for whether one resided in Baltimore in 1930. The inclusion of occupation category has a lot of explanatory power. Being a Baltimorean is not statistically significant, perhaps since while there were more opportunities in the city, farmers were less likely to be reported as unemployed. Being over 40 is still statistically significant when introducing the occupation controls. Column 3 includes county fixed effects and column 4 includes a dummy for each of the more than 100 occupations (as opposed to categories).

In column 5, I include the rate of those not working (be it unemployed or not in the labor force). This measure is meant to capture one possible explanation, that those not working can provide less information on job openings. The effect is statistically significant at the $p<.01$ level. Column 6 reports a similar specification using the unemployment rate instead of the non-working rate. The results are the same, and the point estimate of the marginal effect is slightly higher. The point estimates imply that an increase of 1 percentage point in one's (military) company unemployment rate decreases one's probability of being employed by almost 0.4 percentage points. The last column (7), includes both the neighborhood and the unit unemployment rates concurrently. Both are found to be statistically significant at the $p<.01$ level, and have the expected sign. The point estimate of the unit's marginal effect is more than twice as large. An increase of one point in the company's (neighborhood's) unemployment rate, decrease one's own likelihood of employment by 0.377 points (0.148 points for the neighborhood effect). I conclude by noting that from a practical point of view, one might be inclined to believe that the inclusion of both effects might alleviate some of the concerns regarding the bias.

Since one might be concerned that the sample of veterans is very specialized, I present a similar specification for a more general sample. Table 3.12 examines the neighborhood sample, which is a more heterogenous sample, and I include controls for race, sex, veteran status, etc. The results are similar across the two samples. For the neighborhood sample I also include a specification in which only males are included. This avoids some of the issues regarding women's labor supply. The results remain the same, as in 1930, most women were not part of the formal labor force.

In the case of the neighborhood sample, the neighborhood-block's unemployment rate (columns 6-7) has a statistically significant effect. In column 6, I consider the unemployment rate among both males and females. Column 7 considers the unemployment rate only among men. In both cases, the point estimates of the marginal effect are quite high. One concern is that since the neighborhood block is smaller in size than the company considered in the previous table, one's own outcome determines a higher proportion of the group average $\bar{y}$. To address this concern, I have estimated the same specifications using $\bar{y}_{-i}$, that is the average among all others, self excluded. The point estimates are lower for this specification but are of the same magnitude.

3.5.2. Reduced Form Estimates

The results of the "naive" estimator suggest that networks play an important role in determining one's own likelihood of being unemployed. Before addressing some of the issues associated with the "naive" specification, it is of use to examine the reduced-form estimates which include only one's own covariates and other's (exogenous) covariates. This

Table 3.12: Neighbors of World War I Veterans Employment Including the Effect of Peers' Employment in 1930

Neighborhood Sample in the Census of 1930; Probit Specification; Dependent variable- employment							
	1	2	3	4	5	6	7
Block unemployment rate (only males)							-5.145*** (0.203) [-0.478]
Block unemployment rate						-5.216*** (0.217) [-0.502]	
Age	0.022*** (0.007) [0.003]				0.023** (0.009) [0.003]	0.016* (0.009) [0.002]	0.016* (0.009) [0.001]
Age-squared (divided by 100)	-0.036*** (0.007) [-0.004]				-0.007*** (0.001) [-0.007]	-0.031*** (0.009) [-0.003]	-0.032*** (0.010) [-0.002]
War veteran	0.0005 (0.057) [0.0007]	0.008 (0.060) [0.001]	0.001 (0.061) [0.0007]	-0.003 (0.065) [0.0008]	-0.082 (0.071) [-0.015]	-0.030 (0.067) [-0.003]	-0.043 (0.068) [-0.004]
Head of household	0.196*** (0.054) [0.027]	0.183*** (0.054) [0.025]	0.170*** (0.065) [0.025]	0.081 (0.069) [0.01]	0.233*** (0.084) [0.046]	0.127* (0.076) [0.013]	0.130* (0.077) [0.012]
Is married	0.143*** (0.051) [0.020]	0.148*** (0.051) [0.021]	0.216*** (0.063) [0.033]	0.220*** (0.067) [0.029]	0.167** (0.081) [0.033]	0.156** (0.073) [0.016]	0.141* (0.074) [0.014]
Is male	-0.272*** (0.051) [-0.033]	-0.281*** (0.051) [-0.034]					
Age Dummies (25-34 Omitted)							
Is under 25		-0.073 (0.058) [-0.010]	-0.089 (0.068) [-0.013]	-0.077 (0.070) [-0.010]			
Age 35 to 44		-0.056 (0.056) [-0.008]	-0.097 (0.061) [-0.014]	-0.106* (0.064) [-0.014]			
Age 45 to 54		-0.149** (0.061) [-0.022]	-0.192*** (0.067) [-0.030]	-0.223*** (0.070) [-0.031]			
Age 55 to 64		-0.262*** (0.069) [-0.042]	-0.313*** (0.074) [-0.053]	-0.354*** (0.078) [-0.055]			
Age 65 to 74		-0.665*** (0.082) [-0.137]	-0.661*** (0.090) [-0.139]	-0.691*** (0.096) [-0.134]			
Over 75		-0.617*** (0.140) [-0.126]	-0.698*** (0.148) [-0.153]	-0.677*** (0.177) [-0.134]			
Neighborhood block Dummy Variable					Yes		
Race and occupation Dummy Variables	Race only	Race only	Race only	Yes	Yes	Yes	Yes
Constant	1.247	1.606	1.318	2.022	0.692	2.273	2.311
Observations	11037	11085	8806	8598	5187	8554	8554
Pseudo R-squared	0.02	0.02	0.03	0.07	0.11	0.20	0.22

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)

specification addresses the two aforementioned concerns, the omission of the contextual effect, and bias due to the simultaneity.

The results for this specification are reported in Table 3.13. I find that the characteristics of the military unit, such as the percent of unit member over 40, and the neighborhood-block characteristics, such as marital rate and the average age of those in the labor force have a statistically significant effect. I interpret the fact that I find a statistically significant effect for some of the coefficients as further evidence for the existence of a peer effect in determining unemployment.

The reduced form specification can be written as:

$$y_{i,g} = h(\alpha + x_{i,g}^{'}\beta + \bar{X}_g^{'}\gamma_{1r} + Z_g^{'}\gamma_{2r} + \epsilon_{i,g}) \tag{3.3}$$

where $\bar{X}_g = \frac{1}{n_g}\sum\limits_{i \in g} x_i$. The distinction between the group's average of individual characteristics $\bar{X}_g$, and Z_g, the group's characteristics is somewhat arbitrary, and mainly included for the consistency with the notation used in other sections. The coefficients γ_r are indexed with r as they are only a measure of whether a social influence exists. If there is no endogenous effect their magnitude can be interpreted as the contextual effect.[25]

In Table 3.13, column 1 includes in addition to one's own characteristics three group averages, the percent of those over 40 years old, the percent married, and the percent of the company members who were wounded during the war. The percent of company members who are over 40 has a statistically significant effect at the p<.05 level. The point estimate suggests that a 10-percentage-points increase in the rate of those over 40 would decrease one's own likelihood of being employed by two percentage points. The company's

[25]For example, in the linear-in-means case, if there is no endogenous effect ($\rho = 0$), then it must follow that $\gamma_{1r} = 0$, and γ_{2r} is the contextual effect. This can be seen by examining equation (2.4).

Table 3.13: Reduced-Form Peer Effect Estimates of Veterans' and Neighbors' Employment in 1930

Veterans and Neighbors in the Census of 1930; Probit Specification; Dependent variable- employment						
	Veteran Sample				Neighborhoods Sample	
	1	2	3	4	5	6
Over 40 years old	-0.781***	-0.661***	-0.739***	-0.794***		
	(0.227)	(0.229)	(0.238)	(0.228)		
Age					0.029***	0.011
					(0.008)	(0.008)
Age-squared (divided by 100)					-0.041***	-0.023***
					(0.008)	(0.008)
Is married	0.175	0.238				0.282***
	(0.267)	(0.264)				(0.048)
Average age of neighborhood block members who are in the labor force			-0.045		0.065**	0.060*
			(0.148)		(0.031)	(0.031)
Average of Age^2 among neighborhood block			0.001		-0.001*	-0.001*
			(0.002)		(0.0004)	(0.000)
wounded during WWI	-0.137	-0.113				
	(0.267)	(0.271)				
% in company over 40	-3.593**	-4.543**	-4.215**	-4.122***		
	(1.425)	(1.953)	(1.897)	(1.347)		
Company marital rate	2.194	1.881				
	(1.769)	(2.218)				
% of co. wounded	0.02	0.048				
	(1.334)	(1.432)				
Company average of the average neighborhood block age				-1.977***		
				(0.458)		
Company average of the average neighborhood block age-squared				0.022***		
				(0.005)		
Head of household (HH)			0.326			
			(0.231)			
% in co. who are HH			2.056			
			(3.979)			
Own occupation professional, technical, managerial, clerical, or sales			1.470***			
			(0.351)			
Own occupation craftsmen, operatives, or laborers			0.379			
			(0.278)			
% in co. who are professional, technical, managerial, clerical, or Salesmen			2.000			
			(2.367)			
% in co. who are craftsmen, operatives, or laborers			3.328			
			(2.823)			
% in neighborhood who are professional, technical, managerial, clerical, or salesmen			-0.811*			
			(0.423)			
% in Neighborhood who are craftsmen, operatives, or laborers			-1.141**			
			(0.532)			
% in co. who are farmers		-6.366**				
		(2.713)				
% in co. who are clerical and salesmen		-3.423				
		(3.476)				
% in co. who are craftsmen		-0.526				
		(3.116)				
% in co. who are operatives, service workers, and laborers		-3.664				
		(2.708)				
Occupation Category (12 categories)	Yes			Yes	Yes	Yes
Observations	529	550	523	529	8,869	8,869
Pseudo R-squared	0.16	0.14	0.15	0.17	0.06	0.07

Notes: Robust standard errors in parentheses (assume correlation within company)
significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
Samples: Columns 1-4, WWI veterans linked to the census. Columns 5-6, neighbors of the WWI sample.
Excluded category in column 2- Professional, technical and managerial. Excluded category in column 3- farmers.

Table 3.14: Marginal Effects of the Reduced-Form Peer Effect Estimates of Veterans' and Neighbors' Employment

(Marginal effects corresponding to the preceding table)

Veterans and Neighbors in the Census of 1930; Probit Specification; Dependent variable- employment

	Veteran Sample				Neighborhoods Sample	
	1	2	3	4	5	6
Over 40 years old	-0.079***	-0.085***	-0.106***	-0.074***		
Age					0.004***	0.001
Age-squared (divided by 100)					-0.005***	-0.002***
Is married	0.011	0.022				0.038***
Average age of neighborhood block members who are in the labor force			-0.004		0.008**	0.007*
Average of Age^2 among neighborhood block			0.0002		-0.010*	-0.0001*
wounded during WWI	-0.008	-0.010				
% in company over 40	-0.199**	-0.377**	-0.371**	-0.203***		
Company marital rate	0.121	0.156				
% of co. wounded	0.001	0.004				
Company average of the average neighborhood block age				-0.097***		
Company average of the average neighborhood block age^2				0.001***		
Head of household (HH)			0.033			
% in co. who are HH			0.181			
Own occupation professional, technical, managerial, clerical, or sales			0.121***			
Own occupation craftsmen, operatives, or laborers			0.034			
% in co. who are professional, technical, managerial, clerical, or salesmen			0.176			
% in co. who are craftsmen, operatives, or laborers			0.293			
% in neighborhood who are professional, technical, managerial, clerical, or salesmen			-0.071*			
% in Neighborhood who are craftsmen, operatives, or laborers			-0.101**			
% in co. who are farmers		-0.528**				
% in co. who are clerical and salesmen		-0.284				
% in co. who are craftsmen		-0.044				
% in co. who are operatives, service workers, and laborers		-0.304				
Occupation Category (12 categories)	Yes			Yes	Yes	Yes
Observations	529	550	523	529	8,869	8,869
Pseudo R-squared	0.16	0.14	0.15	0.17	0.06	0.07

Notes: Marginal effects [dF/dx] evaluated at the mean value of x.
Significant at 10%; ** significant at 5%; *** significant at 1%
Samples: Columns 1-4, WWI veterans linked to the census. Columns 5-6, neighbors of the WWI sample.
Excluded category in column 2- Professional, technical and managerial. Excluded category in column 3- farmers.

marital rate has a positive effect on one's own employment prospects, but the standard errors are very large. This is not all that surprising since in most of the specifications I have tried, one's own marital status is not found to be a statistically significant predictor of employment. Similarly, the company's injury rate during the war is not found to have a statistically significant effect. (In Section 3.5.1 I have discussed more extensively the fact that having been wounded during the war does not seem to have a statistically significant effect.)

Column 2 adds the percent of others in one of five occupational categories. The only one found to be statistically significant is the percent of company members who are farmers, which adversely affects one's own likelihood of employment. Possibly this could be since most were not farmers, and so farmers were a less useful source of job information for the non-farmers. Column 3 includes, in addition to the company's occupations, the percent of those in one of three occupational categories for those in the neighborhood block. The only statistically significant results are for the occupation choice of others in one's own neighborhood block. One way to interpret these results are that the neighborhood effects are stronger. On the other hand, selection into neighborhoods is not exogenous and could be influenced or correlated with the type of occupation.

Columns 5-6 repeat the reduced form specification, and examine the likelihood of employment for all of the neighborhood sample members which consists of all the neighbors of the veterans. The average age and age-squared of one's neighbors are found to have a statistically significant effect.

In Section 3.5.5 I will consider an instrument which I include here directly. Column 4 examines the likelihood of employment as a function of one's own characteristics (including occupational categories fixed effects), and the average age and age-squared across all neighborhoods in which the company members reside. The age of all other neighbors of the company members is found to have a statistically significant effect. While it is difficult to attach any meaningful interpretation to the magnitude of the estimates, it does suggest that the company members (as captured by their neighbors' ages) have a statistically significant effect. While one could argue that the average age of members in one's own block is endogenous, it would be unlikely that one could affect the choice of neighborhood for all other company members.

Due to the exogenous nature of the assignment to groups it is less likely that the individual unobserved error is correlated with the average of other's individual characteristics, such as age. This assumption primarily depends on the nature and the time in which most of the unobserved error term ϵ was determined.[26]

The main disadvantage of the reduced form approach is its relatively weak explanatory power when compared to the naive approach. Often times, the available covariates, as in my case, do not account for the majority of the observed variation. The less explanatory power the covariates have, the less likely they are to be a good proxy for other's outcomes. This could be further exacerbated in the case in which the model is non-linear, as in the probit estimates presented in this section. Even if some linear combination of covariates have a strong explanatory power for each individual, a linear combination of the means

[26] At one extreme, the error might be solely due to events which occurred during 1918/9, in which case it is possible that those assigned to companies in which the average age was higher gained some higher/lower unobserved ability. On the other hand, if most of the shock is due to events and circumstances occurring in 1930 (such as being laid off), or were built over a long period (say during the person's entire life since the 1890s), then the assumption is likely to hold.

of the covariates may have far less explanatory power in explaining the group's average outcome.

3.5.3. Alternative Specifications of the Reference Group

In this Section, I examine whether my findings of the importance of one's military company are robust to alternative specifications. First, I show that when examining a larger group (the battalion), which subsumes the military company, the results no longer hold. Further, when both the battalion-level and the company-level outcomes are included, only the latter is found to have an effect. Similarly, other companies within the same regiment are not found to affect one's outcomes.

I then consider a random sample of young white males consisting of the same location composition, that is males who live on the same blocks, but who have not necessarily served together during World War I. I find that the same measure of group unemployment which is found to be influential in the case of those who served together, is not significant in the case where similar males, are "arbitrarily assigned" to groups.

The first two columns of Table 3.15 examine the case in which the reference group is defined as one level higher within the army hierarchy, that of the battalion (see Figure 3.2). Each battalion consists of four companies, and hence this group is four times larger. As can be seen in column 1, the effect of the battalion's nonworking rate is not statistically significant. The measure is very noisy, as it contains the military company members which were found to be statistically significant in the previous sections. Once the nonworking rate of the company members is included in column 2, the point estimate of the battalion measure is very close to zero.

Table 3.15: Alternative Group Specifications

Veterans Sample in 1930; Probit Specification; Dependent variable- employment				
	(1)	(2)	(3)	(4)
Nonworking rate among **battalion** members	-5.703 (5.608) [-0.535]	-0.053 (6.005) [-0.004]		
Nonworking rate among members of **another company within the same regiment**			0.438 (1.979) [0.042]	0.575 (1.933) [0.048]
Nonworking rate among company members		-6.532*** (2.190) [-0.548]		-6.559*** (2.095) [-0.551]
Is over 40	-0.659*** (0.220) [-0.094]	-0.681*** (0.224) [-0.09]	-0.681*** (0.219) [-0.099]	-0.685*** (0.224) [-0.091]
Wounded during WWI	-0.04 (0.247) [-0.004]	-0.075 (0.253) [-0.007]	-0.053 (0.247) [-0.005]	-0.066 (0.254) [-0.006]
Corporal rank	0.538* (0.286) [0.039]	0.538* (0.291) [0.034]	0.514* (0.284) [0.038]	0.536* (0.290) [0.034]
Occupation category (12 categories) controls	Yes	Yes	Yes	Yes
Constant	2.380	2.570	1.912	2.531
Observations	604	604	604	604
Pseudo R-squared	0.13	0.16	0.13	0.16

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)

Columns 3 and 4 consider the effect of another company from the same regiment, though from a different battalion.[27] I find no statistically significant effect for the other company, while one's own company is found to have a large effect. In column 4, the point estimates suggest that an increase of one-percentage-point in one's own company's nonworking rate, decrease one's own likelihood of employment by half-of-a-percentage point.

In Table 3.16, I report the results for various specifications in which the unemployment rate of a pseudo company is included. The purpose of this exercise is to show that the company unemployment rate is not just a proxy for a group of people who happen to have some distribution of locations in the 1930 census. The pseudo company was constructed as follows. For each of the veterans, I randomly picked one white male in the same age bracket who lived on the same street as the veteran in the 1930 Census. We would expect the unemployment rate among this fabricated company to have a very small direct effect, if any.[28] This expectation is confirmed by the results of Table 3.16. The first column includes the base specification without the effect of the pseudo company. The other 3 columns (2-4) include the unemployment rate among the pseudo company members. In all three cases, the measure is not statistically significant. The standard errors are quite large, since this measure introduces a lot of "noise" into the estimation as they are a proxy for local labor market conditions, but based only on a single observation for each neighborhood.

[27] For each company, the "corresponding" company was selected by looking at a company which was alphabetically at the same position in the next battalion. For example, in the case of Company A (1st Battalion), the corresponding company is the first company in the 2nd Battalion (Company E).

[28] Of course, the members of the pseudo company each possibly affect the one veteran on their street, and are collectively a proxy for the local labor market conditions, though a very "noisy" one, since they are based only on one observation.

Table 3.16: The Effect of a Pseudo Company on Veterans' Employment in 1930

Veterans Sample in 1930; Probit Specification; Dependent variable- employment				
	(1)	(2)	(3)	(4)
Unemployment rate among **pseudo** company members		1.196 (1.976) [0.064]	-0.042 (2.014) [-0.002]	0.954 (2.164) [0.036]
Unemployment rate among **actual** company members	-7.284*** (2.155) [-0.295]		-7.293*** (2.196) [-0.296]	-7.390*** (2.338) [-0.279]
Unemployment rate among neighbors				-4.090*** (1.051) [-0.155]
Over 40	-0.792*** (0.261) [-0.063]	-0.759*** (0.254) [-0.073]	-0.791*** (0.261) [-0.063]	-0.874*** (0.278) [-0.07]
Wounded during WWI	-0.158 (0.290) [-0.007]	-0.16 (0.280) [-0.01]	-0.157 (0.292) [-0.007]	-0.251 (0.313) [-0.012]
Corporal Rank	0.606* (0.339) [0.017]	0.532* (0.322) [0.021]	0.607* (0.341) [0.017]	0.504 (0.352) [0.014]
Constant	8.171	6.885	8.174	8.354
Occupation category (12 categories) Fixed Effects	Yes	Yes	Yes	Yes
Observations	529	529	529	494
Pseudo R-squared	0.20	0.16	0.20	0.27

Notes: The group outcome was "artificially" constructed replacing each veteran's outcome by the outcome of a randomly chosen white male in the same age bracket living on the same street as the veteran in the 1930 census.
Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)

3.5.4. Controlling for Prewar Place of Residence

In Section 3.3.3 I discussed the correlation between one's prewar place of residence and the assignment to group. Recall that in no case did soldiers from one town (or the in the case of the large cities, the area equivalent to a voting precinct) constitute more than 30% of a company, and in most instances, less than 20%. Moreover, many members of each company were one of only a handful from the same town, or even state if they were not from Maryland.

In this section, I address the concern that the nature of assignment to groups might lead to some unobserved correlations among company members due to the fact that some of them originated from the same area as some of the other company members. The results of this exercise are summarized in Table 3.17 and provide further evidence that the peer effect should not be attributed to correlations due to prewar place of residence.

The first column includes a fixed effect for each town (or draft board areas, which were 30,000 people large, for the big cities). The effect of the company's unemployment rate is statistically significant at the p<.01 level. The point estimate is very large, but also has a wide range since it is not estimated very accurately. This is due to the sample size, as many towns were very small. The upper end of the 95% confidence interval of the marginal effect is -0.27, which is a little less than the point estimates found in other sections.

In column 2, I restrict the sample to those who had no more than 8 other men in their company originate from the same town (based on their home address prior to enlistment). Column 3 is the same for those who had no more than 3 other men in their company originate from the same town. In both cases the unemployment rate of all others in

Table 3.17: World War I Veterans' Employment Controlling for Prewar Location

Veterans Sample in 1930; Probit Specification; Dependent variable- employment						
	1	2	3	4	5	6
Sample Used:	All Veterans	Those who were from towns[&] for which **less than 10** men ended up in same company	Those who were from towns[&] for which **less than 5** men ended up in same company	Those who were from towns[&] for which **less than 10** men ended up in same **regiment**	Those who were from towns[&] for which **less than 5** men ended up in same **regiment**	Within the regiment, less than half of the same town[&] members ended up in the same company
Unemployment rate among company members	-24.807*** (7.238) [-1.467]	-11.881*** (3.238) [-0.437]	-13.208*** (3.816) [-0.542]	-12.439*** (3.803) [-0.5]	-6.114 (4.751) [-0.387]	-14.599** (6.123) [-0.603]
Over 40	-0.61 (0.395) [-0.056]	-0.778** (0.340) [-0.057]	-0.682* (0.414) [-0.051]	-0.757* (0.432) [-0.059]	-0.579 (0.450) [-0.056]	-0.351 (0.555) [-0.019]
Corporal rank during WWI	1.230** (0.622) [0.045]	0.380 (0.359) [0.011]	0.629 (0.523) [0.017]	0.199 (0.471) [0.007]		0.753 (0.572) [0.02]
Wounded during WWI	0.366 (0.569) [0.017]	-0.143 (0.338) [-0.006]	-0.090 (0.387) [-0.004]	-0.070 (0.403) [-0.003]	0.078 (0.534) [0.005]	
Constant	7.287	8.568	8.591	8.334	2.332	8.404
Town/Draft Board [&] Controls	**Yes**					
Occupation Category (12 categories) controls	Yes	Yes	Yes	Yes		Yes
Constant	7.287	8.568	8.591	8.334	2.332	8.404
Observations	202	383	233	266	177	108
Pseudo R-squared	0.39	0.21	0.22	0.19	0.07	0.23

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
[&] Large cities were subdivided according to draft boards (roughly the size of a voting precinct)

the company has a statistically significant effect, even though almost all of the company members were from a different town than those examined in these columns. Columns 4 further restricts the sample to those from really small towns, that is those who had no more than 8 other men serve in the entire regiment. Column 5 is the same for the case of 3 others. Once again, the company's unemployment rate is found to be statistically significant in the case of column 4. In column 5 the effect has the expected sign, but the standard errors are large probably due to the size of the sample.

The results in columns 2-5 are inherently based on those from small towns. Column 6 incorporates those from the large cities by looking at a slightly different sample. For each town I had calculated the number of men in the regiment from that town. I then focus only on those men in companies in which less than half of the men from that town ended up in a particular company. For instance, consider the area designated draft board 15 in Baltimore. In total, 20 men from that area had served in the regiment. But in company C only 3, that is less than half, were from that area. On the other hand, there were 13 men in company I from that area, and they were not included in this regression (however, they were used for computing the group's unemployment rate). Even though the sample size is quite small, the effect of the company's unemployment rate is statistically significant and is found to adversely affect a veteran's likelihood of being employed.

3.5.5. Instrumental Variable Estimates

As discussed in the previous sections, even under the assumption of no contextual effect, the estimate of the endogenous variable is likely to be biased due to the simultaneity problem. In the empirical social influence literature there are two approaches to addressing this problem. The first is the lag spatial model approach (e.g. Anselin ,2001). However, as sample size and the size of the group increases, estimation becomes very unreliable in my experience, since one needs to estimate a system of equations with as many equations as the size of the group.

A second widely-used approach is to instrument the group's average outcome. In this section I present the results for the instrumental variable estimates and discuss some of the problems which are unique to instrumental variable estimation in the presence of social influences in addition to the usual concerns regarding the choice and validity of the instrument.

In order for an instrument to be valid, it needs to be correlated with the instrumented variable, in this case, the percent of those not working in each company, *and* it must be independent of the individual's error term. Taking advantage of the rich dataset, I use an exogenous characteristic of all of the neighborhoods in which the veterans of a company reside in.

In Table 3.18, I present an example using as an instrument for the unit's non-working rate, the average of the average labor force age across all blocks in which the company members reside ($\frac{1}{unit\ size}\sum_{b\in blocks}(\frac{1}{block\ size}\sum_{i\in block} age_{i,b})$). Similarly, I use the average of the average marital rate (not reported). The statistically significant results suggest that the magnitude of the effect is approximately the same as that of the "naive" estimator. The marginal point estimate for the group's effect is 0.35. An increase of 1 percentage point in one's (military) company non-working rate decreases one's probability of being employed by a little over a third of a percentage point.

Table 3.18: Instrumental Variables Estimates of the Effect of Peers' Employment on Veterans' Employment in 1930

Veterans Sample in 1930; Dependent variable- employment			
	(1)	(2)	(3)
2nd Stage Specification	**Probit (See above note>)**	**Probit (See note below>)**	**Linear Probability Model**
Estimated using	MLE	2-stage (Newey, 1987)	2-stage-least-squares
IV(s) used:	Company average of the average neighborhood age	Company average of the average neighborhood age	Company average of the average: i. neighborhood age ii. neighborhood age^2, iii. neighborhood marital rate.
Percent of company members not working	-8.479* (4.663) [-.354]	-8.493* (4.761)	-0.785*** (0.303)
Over 40	-0.850*** (0.270) [-.072]	-0.851*** (0.269)	-0.122*** (0.036)
Corporal rank	0.625* (0.346) [.018]	0.626* (0.346)	0.042 (0.028)
Constant	8.514	8.083	1.064
Observations	489	489	489
Occupation category (12 categories) controls	Yes	Yes	Yes
Wald Chi-Square test of exogeneity&	**0.07**	**0.07**	
Exogeneity p-value	**0.79**	**0.79**	

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
& The Null hypothesis is that in the 2nd stage, the coefficient for the first-stage error is equal to zero.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
> The actual coefficient estimates for columns (1) and (2) entail very restrictive assumptions. However, the test for exogeneity (the last two rows in bold) is valid even under less restrictive assumptions.

Table 3.18 reports the results of an instrumental variable estimation. The first two columns examine the case in which the dependent variable, whether employed or not, is estimated using the probit functional form and the endogenous variable, the non-working rate among company members, is instrumented using the aforementioned instrument, the average of the average labor force age across all blocks in which the company members reside.

The functional form for the "first stage" is a linear regression. The results of the first stage regression (not reported) suggest that instrument used is a fairly good predictor of the endogenous variable. The R-squared of the first stage is .20, and the t-statistic for the instrument is 9.72. I provide a brief discussion of whether the instrument is a valid one later in this section. Column 1 estimates the model using maximum-likelihood estimate. Column 2 is the same exact model, using the two-stage approach suggested by Newey (1987). The advantage of the second estimation method is that the likelihood function is less computationally intensive and more likely to converge. However, as can be seen, the results are very similar for the two methods, probably since there is only one endogenous variable. The coefficients (and the marginal effects) obtained using the second method (two-step) estimation are scaled and do not have a meaningful interpretation (though the ratios of two marginal effects would). I focus on the results obtained in column 1.

The marginal effect of the company's non-working rate, is similar to that found in Section 3.5.1. The point estimates imply that an increase of 1 percentage point in one's (military) company unemployment rate, decreases one's probability of being employed by 0.35 percentage points. The point estimate suggest that for a veteran, the effect of being over 40 is twice as large as belonging to a company in which the non-working rate is 10 percentage points higher.

The Wald tests for exogeneity which tests whether the first-stage's residual is correlated with that of the second-stage residual are rejected. The p-value for the test is 0.79, suggesting that the null hypothesis that the instrumented variable, the company's non-working rate, is indeed endogenous, can be rejected.[29] This in turn implies that the results reported in Section 3.5.1, using the "naive" estimator, are a good approximation. Recall, that if the error terms of each veteran are uncorrelated with any of the group characteristics, and if groups are large enough, then one should not expect the company's non-working rate to be "endogenous."

The 2-stage model, in which the second stage is a probit specification, assumes a very strong condition, namely that the distribution of the error term is homogenous across agents. Unlike the case of the linear model, if this assumption is, in fact, violated, the estimates are not only inefficient, but inconsistent. However, the test for exogeneity discussed above is consistent even when allowing for an heterogenous error term

In addition to the aforementioned need for strong assumptions, in the case of instrumental variable estimates when the endogenous variable is the group average outcome, a 2-stage model in which the second stage is probit, will, in general, lead to a logically inconsistent model (for instance, this is a point mentioned by Krauth, 2006). To address these concerns, in column 3 I present the results for a 2-stage least-squares estimate. The point estimates are considerably higher, though the 95% confidence interval does contain the point estimates found in column 1.

[29]Rivers and Vuong (1988) found the Wald test for exogeneity for their proposed 2-step CMLE, which is similar to the Newey(1987) used in column 2, to perform well in the finite samples they had considered.

In regards to the validity of the instrument, while it is possible that an individual chooses to live at a block based on the average age of its members, the instruments I used are based on the average across all blocks in which the members of each military company reside. These instruments are based on the characteristics of hundreds of others, and are likely to be uncorrelated with one's individual unobserved characteristics.

The more crucial assumption is that the average age of the neighborhood does not have a direct contextual effect, but only through the endogenous effect, the group's average unemployment rate. This assumption cannot be tested. In the next section, I make use of the Multiple Reference Groups framework (Chapter 2) which allows for both endogenous and contextual effects.

3.6. Estimation of the Endogenous Effect Using Multiple Reference Groups

In this section, I first present the results of the reduced form specification corresponding to equation (2.15) discussed in Section 2.4:

$$y_{i,g} = \alpha + \beta x_{i,g} + \delta_1 \bar{X}_{-i,A} + \delta_2 \bar{X}_{-i,B} + \delta_3 \overline{\overline{X}}^{A}_{-i,B} + \epsilon_{i,g} \quad (i = 1..n_g; g = 1..G) \tag{3.4}$$

The empirical exercise examines the employment outcomes of the veterans as recorded by the 1930 census. Each individual's outcome of interest y is the binary outcome of being employed or unemployed.

In the most reduced form specification, I find that some of the characteristics, such as the average of the average neighborhood-level, have a statistically significant effect on one's likelihood of employment (corresponding to $\delta_3 \neq 0$). Recall that these variables are constructed as an average of the characteristics of hundreds of others, and are likely to be uncorrelated with one's individual unobserved characteristics.

The results for a linear probability specification of one's likelihood of employment are reported in Table 3.19. The independent variables are occupation and age for various group levels (own, neighborhood, military unit, and average of neighborhoods across military unit). The variables corresponding to $\overline{\overline{X}}^{A}_{-i,B}$, the average across the unit of the average neighborhood level, are those labeled *v., vi., xiii, and xiv* in Table 3.19. Except for *xiv*, the coefficients are statistically significant at the p<.05 level. However, one should be cautious of any interpretation of these results beyond the fact that they suggest the existence of an endogenous effect.

Next, I solve for the coefficient of interest ρ_A, the endogenous effect of one's military company. To do so, I use the reduced form estimated in Table 3.20. This is the same specification as Table 3.19, corresponding to equation (3.4), with one difference. In this regression, I only include those who live on a block in which no other World War I veteran resides. This is the case for a little over a quarter of the sample. However, note that the company average ($\bar{X}_A$) and the average of neighborhoods across the company ($\overline{\overline{X}}^{A}_{B}$) are computed using all members .

The reason for focusing on this subsample is to address any concern for potential bias due to some unobserved affiliation of neighbors who are themselves World War I veterans (and served in units which are not in my sample). To illustrate, consider two veterans in my sample who had served together. Next, assume that each of the veterans has a neighbor who had served in the Navy, on the same ship. The coefficient of the neighborhood-block effect might be biased because the two possible paths through which

Table 3.19: Evidence of an Endogenous Social Effect using Reduced-Form Estimates

Veterans Sample in 1930; OLS Specification; Dependent variable- employment		
(label)		
i.	Is over 40 years old	-0.118** (0.053)
ii.	Percent of company members over 40	-0.329 (0.333)
iii.	Average age of those in the labor force in neighborhood (self excluded)	0.006 (0.015)
iv.	Average of age-square of those in the labor force in neighborhood (self excluded)	-0.00007 (0.00018)
v.	Company average of the average age in neighborhood of those in the labor force	-0.094*** (0.024)
vi.	Company average of the average age-squared in neighborhood of those in the labor force	0.001** (0.000)
vii.	Own occupation professional, technical, managerial, clerical, or sales	0.180* (0.087)
viii.	Own occupation craftsmen, operatives, or laborers	0.082 (0.071)
ix.	% in co. who are professional, technical, managerial, clerical, or salesmen	1.015*** (0.311)
x.	% in co. who are craftsmen, operatives, or laborers	0.807*** (0.203)
xi.	% in neighborhood who are professional, technical, managerial, clerical, or salesmen	-0.134 (0.076)
xii.	% in neighborhood who are craftsmen, operatives, or laborers	-0.170** (0.061)
xiii.	Company average of the average % in neighborhood who are professional, technical, managerial, clerical, or salesmen (average of the above averages)	-0.952*** (0.271)
xiv.	Company average of the average % in neighborhood who are craftsmen, operatives, or laborers (average of the above averages)	-0.368 (0.304)
	Constant	3.039
	Observations	496
	R-squared	0.09

Notes: Robust standard errors in parentheses
(assume correlation within company and independence across companies)
* significant at 10%; ** significant at 5%; *** significant at 1%
Omitted occupation is farming.
Sample: Veterans linked to the 1930 Census
Those blocks with less than 10 people in the labor force were omitted.

Table 3.20: Auxiliary OLS Regression Results for Computing the Endogenous Social Effect of Peers' Employment

Veterans Sample in 1930; OLS Specification; Dependent variable- employment		
(label)		
i.	Is over 40 years old	-0.109 (0.111)
ii.	Percent of company members over 40	-0.668 (0.560)
iii.	Average age of those in the labor force in neighborhood (self excluded)	0.025 (0.046)
iv.	Average of age-square of those in the labor force in neighborhood (self excluded)	-0.000 (0.001)
v.	Company average of the average age in neighborhood of those in the labor force	-0.191** (0.071)
vi.	Company average of the average age-squared in neighborhood of those in the labor force	0.002** (0.001)
vii.	Own occupation professional, technical, managerial, clerical, or sales	0.217** (0.097)
viii.	Own occupation craftsmen, operatives, or laborers	0.029 (0.060)
ix.	% in co. who are professional, technical, managerial, clerical, or salesmen	3.177*** (0.485)
x.	% in co. who are craftsmen, operatives, or laborers	2.793*** (0.490)
xi.	*% in neighborhood who are professional, technical, managerial, clerical, or salesmen*	-0.336** (0.115)
xii.	**% in neighborhood who are craftsmen, operatives, or laborers**	-0.295** (0.126)
xiii.	*Company average of the average % in neighborhood who are Professional, technical, managerial, Clerical, or Salesmen (average of the above averages)*	-2.718*** (0.322)
xiv.	**Company average of the average % in neighborhood who are craftsmen, operatives, or laborers (average of the above averages)**	-1.015** (0.405)
	Constant	3.911
	Observations	145
	R-squared	0.20

Notes: Robust standard errors in parentheses (assume correlation within company and independence across companies) * significant at 10%; ** significant at 5%; *** significant at 1%
Omitted occupation is farming.
Sample: Veterans linked to the 1930 Census. The regression includes as a dependent variable only those for which no other WWI veteran lives in neighborhood block (to prevent biased estimates, as explained in text). However, the company and neighborhood values (such as average age of neighborhood members) were computed using the entire sample. Also, those blocks with less than 10 people in the labor force were omitted.
+ For example, the endogenous effect of the military unit is a function of the ratio of *xii/xiv* or *xi/xiii* (where the roman numerals are the labels of the coefficients in the above table).

the effect passes cannot be distinguished. (The two possible paths are the comrade's neighbor to own neighbor to the veteran considered, or from the comrade's neighbor to the veteran's comrade to the veteran considered.) This concern is addressed by including only those who have no neighbors who are veterans.[30]

As shown in Section 2.4, for large enough groups, the endogenous effect of the military unit is given by $\rho_A = \frac{1}{1+\frac{\delta_2}{\delta_3}}$. For calculating the endogenous effect, note that either of the pairs, *xi* (δ_2) and *xiii* (δ_3), or *xii* (δ_2) and *xiv* (δ_3), could be used to estimate the effect. In fact, one testable implication of the model is that this ratio should be the same.

Using the delta method to compute the standard errors (reported here in parentheses), the endogenous effect of the military company based on the coefficients of Table 3.20 is:

$\rho_A = 0.88$ (0.03) with a t-value of 26.09. Using the other set of coefficients the results are $\rho_A = 0.77$ (0.11) with a t-value of 6.62. The effect is relatively higher than the estimates obtained in the previous sections. This could possibly be due to the small sample size. Using the coefficients of Table 3.19, which is a much larger sample (though includes also those neighborhoods in which there are other veterans). The endogenous effect is $\rho_A = 0.68$, and a value such as $\rho_A = .31$ is still within the 95% confidence interval.

Using the value $\rho_A = .8$, the interpretation of the point estimate of the military endogenous effect is that a 1-percentage-point increase among others' unemployment (the military company members), decreases one's own likelihood of employment by 0.8 percentage points.

[30]There are many other ways in which people across blocks might be connected. However, the only two effects considered here are that of the neighborhood and that of one's World War I military unit. The results will, in general, not be biased if these unobserved networks and connections across neighborhoods are not systematic.

3.6.1. Decomposition into the Contextual and Endogenous Effects

The endogenous effect estimated in the previous section was found to be substantial in size. In this section, I provide a comparison of the endogenous and contextual effects. Note that one cannot directly compare the two since they measure the effect of different variables (say the effect of average age, and the effect of the average unemployment rate). Instead, I consider a decomposition of the marginal effect.

Throughout this section I will use as an example the marginal effect of the "percent in company who are professional, technical, managerial, clerical, or salesmen." I denote this average as $\overline{X}_A$ (this corresponds to label *ix* in Tables 3.19 and 3.20). For large enough groups of size B, the marginal effect can be written as:

$$\lim_{n_B \to \infty} \frac{\partial y_i}{\partial \overline{X}_A} = \gamma_A + \frac{\rho_A}{1-\rho_A} \cdot [\beta + \gamma_A] \tag{3.5}$$

The first part of this effect (γ_A) is the (direct) contextual effect.
The second part, $\frac{\rho_A}{1-\rho_A} \cdot [\beta + \gamma_A]$, is due to the endogenous effect (note that it equals zero if there is no endogenous effect, i.e. $\rho_A = 0$). The second part of the effect depends on β, the own effect, and on the contextual effect.

Using the coefficients from Table 3.20, the 95% confidence interval for γ_A is $[-.15, .09]$. In fact, one cannot reject the null that $\gamma_A = 0$ (p-value=0.57). In contrast, the large endogenous effect implies a large multiplier. For instance, using $\rho_A = .7$ then $\frac{\rho_A}{1-\rho_A} > 2$. Hence, if β were equal to zero, more than twice of the marginal effect is due to the endogenous effect. Since β is so much larger than the contextual effect $\beta = 0.217$ (with a standard error of 0.097), the endogenous effect is even more important (relative to the case in which $\beta = 0$).

Similar results hold when considering the marginal effect of other covariates. These suggest that the endogenous effect is more important than the contextual effect in the case of employment. Taken literally, the results imply that it is the employment status of others which matters for finding jobs, not the characteristics of others (such as whether they are professionals). This is consistent with the models of job-information-transmission where the employment of others affects one's own likelihood of employment (e.g. Calvó-Armengol and Jackson , 2004).

3.7. Conclusion

In this chapter, I examined how social ties formed during World War I affect a veteran's likelihood of employment in the 1930 census. I introduced a new data set which links American infantrymen to the 1930 census, as well as their prewar draft registration records. I observe all members of the (military) company. In addition, the data contains information on the neighbors of these veterans in the 1930 census. By considering a setting in which groups were formed due to a "natural experiment," I was able to circumvent the problems that typically arise due to endogenous group formation. I find that these networks, forged during war, have a statistically significant effect which is relatively large. In future work I hope to examine additional types of outcomes, such as the migration decision.

In the last part of the chapter, I examined the case in which some members belong to multiple reference groups and illustrated an application of the Multiple Reference Groups method I introduced in Chapter 2. I found that a large part of the total effect is attributed to the endogenous component. My finding suggests that it is the actual employment of others which matters, not the characteristics of others. This finding is consistent with the informational channel being an important one.

Since almost always networks are endogenously formed, it would be hard to imagine an incentive scheme or government intervention that would radically change people's choice of association. Government intervention could help to strengthen and encourage the formation of contacts among those already likely to associate, and by doing so provide them with a network and its associated benefits. This could be the motivation for strengthening associations for minorities, women in business, etc. By better understanding how networks operate, we can better design such programs.

Social networks formed during war allow for a rare opportunity to examine the results of what would otherwise be a difficult "social experiment" to carry out. One could possibly extrapolate the results of this "experiment" to discuss the merits of various programs which seek to integrate people from various socio-economic backgrounds (such as the "bussing" of school children).

Finally, evidence supporting the existence of a (large) endogenous effect imply that any public policy that targets unemployment is underestimating the benefits of the program if it fails to account for the "social multiplier" effect.

CHAPTER 4

Social Interactions and Labor Market Outcomes of Union Army Veterans

4.1. Introduction

In Chapter 3, the importance of the "informal" channel in the labor market was extensively discussed. For example, contacts have been shown to be of importance across occupations, skill levels, and countries. Ioannides and Loury (2004) provide a comprehensive overview of many of these findings. Bewley (1999, p. 368) lists 24 studies that were published between the years 1932-1990. The percent of jobs or job offers obtained through friends and relatives ranges from 18% to 78% and is between 30% and 60% in most of those studies. The large social effect found in Chapter 3 for the sample of World War I veterans is further evidence of the importance of the "informal" channel.

This chapter investigates empirically the role of social interactions for a different era, the experiences of Civil War veterans in 1880 and 1900. This chapter has two main goals. First, if provides further evidence of the importance of networks in a different era from that considered in Chapter 3 using a very large and detailed dataset. Second, it illustrates the advantage a time-series panel data structure can provide, as it allows for controlling for any effect that is due to an individual or group-level unobserved effect, provided that these unobserved effects are constant over time.

Using a panel of members who served together in the Civil War Union army, I examine the effect of networks formed during the war on postwar employment outcomes. Using networks formed during military service, especially at times of war, to identify network effects has two main advantages. First, the mapping of the network among agents is absent from most economic datasets. My dataset contains information on actual (past) ties between agents.[1] In the other extreme, many studies in the field of sociology have a very detailed description of the network. However, this is either for a very small sample that does not allow for econometric analysis or does not contain information on outcomes relevant to the labor market.

The second advantage of networks formed during war is the nature of these networks. Specifically, these assignments are made based on attributes which are not directly related to the outcome or covariates of interest, create fairly strong bonds, and are of large scale. The rich dataset further allows for controlling for the strength of the war experience that the various units experienced.

In Section 3.2.1, I discussed various mechanisms through which networks may operate in the labor market. In this chapter, I will focus on the "informational" role networks play, such as the case in which network members provide each other with job contacts and leads. While there are many instances in which norms or peer effects, such as a "stigma effect" might play an important role, in the case of unemployment and job search I emphasize the "informational" role for two reasons. First, the sample I examine contains only white

[1] This is in contrast to the approach taken by many researches interested in economic outcomes (such as labor market outcomes or welfare program use), who had come up with various proxies for the probability of the existence of a network. For example, Bertrand et al. (2000) use Census data and proxy for networks by whether individuals speak the same foreign language at home. In a recent paper, Bayer et al. (2005) examine the effect of networks at the census block level on the likelihood of working together. However, this provides only a partial view of one's network, namely one's nearby neighbor.

males in 1880 and 1900. This group is fairly representative of a socio-economic segment of the population which is likely interested in working. Second, most of the empirical part focuses on those in the labor force, and so conditions on those who are interested in working. However, for other outcomes, such as the decision to retire, peer effects may very well play an important role.

The contribution of this chapter is twofold. First, I use a detailed dataset to examine the effect of networks that were formed involuntarily on labor market outcomes. In addition, the group examined, while not representative of the population, represents a substantive segment of the labor market.

Second, the time-series nature of the sample I examine allows me to purge out any effect that is due to an individual or group-level unobserved effect, provided that these unobserved effects are constant over time. Since I examine postwar outcomes of middle-aged men, this is likely to be a plausible assumption.

I find evidence of a statistically and economically significant peer effect among the Union army veterans. For example, in the 1900 census, the marginal effect of a 1-percentage-point increase in one's peers' long-term unemployment rate (defined as six or more months of unemployment in the past year), all else equal, increases one's probability of being long-term unemployed by an additional 0.2 percentage points. The statistically significant effect persists after correcting for the simultaneity generated by the peer effects, and controlling for personal characteristics such as age, marital status, occupation and macroeconomic conditions.

This chapter uses the notation and concepts regarding identification of social interactions that were introduced in Section 2.2. Section 4.2.2 presents the advantage that a panel data setting provides. I derive the result for a two-period linear-in-means case corresponding to the structure of the available data. Section 4.3 describes the Union army data used in this chapter. Section 4.4 details the empirical strategy. It shows the advantages of using the military company as a reference group. I also provide a discussion of why contextual effects are less likely to play a role in the setting studied in this chapter. The results are in Sections 4.5 and 4.6. Section 4.7 concludes and includes a brief discussion of the external validity of the results.

4.2. Econometric Framework

4.2.1. Base Specification

In Section 3.2 I provide a discussion of the possible mechanisms through which others may affect one's labor market outcomes. In Section 3.1.1 I survey some studies which have documented the importance of the "informal" channel, that is jobs found through friends and relatives. Section 2.2 provides a discussion of some of the issues surrounding identification of social interactions.

The first part of the empirical investigation in this chapter will focus on the goal of estimating a specification of the form:

$$y_{i,g} = h[\alpha + x'_{i,g}\beta + Z'_g\gamma + \rho \cdot m(\overrightarrow{y}_{-i,g}) + \epsilon_{i,g}] \tag{4.1}$$

where I use the same notation as that of Section 2.2. There are $g = 1..G$ groups, each with n_g members $i = 1, 2..n_g$, each individual, indexed by i, g has an outcome of interest y, say the binary outcome of being employed or unemployed, a vector of covariates x which affect the likelihood of employment, such as age, occupation, and local labor market conditions, and an error term ϵ_i, a scalar capturing the individual unobservable characteristics and shocks to his or her employment prospects. In addition, each individual's job prospects might depend on the groups characteristics summarized by the vector Z_g, and the outcomes of all other members in the group $\overrightarrow{y}_{-i,g}$.

Section 3.4.1 provides a detailed discussion of this specification, as well as the assumptions behind those specifications.

In the empirical part of the chapter, m_g will be replaced with $\bar{y}_g = \frac{1}{n_g}\sum_{i\in g} y_i$, the average outcome among group member, or the same measure, self excluded, $\bar{y}_{g,-i} = \frac{1}{n_g-1}\sum\limits_{j\in g, j\neq i} y_j$. Recall that as group size increases, from an empirical standpoint, there is little difference between the two measures. Similarly, the group characteristics will be estimated by the average of the group characteristic, $Z_g = \frac{1}{n_g}\sum_{i\in g} x_i$, for example, the average age among group members. As discussed in Section 3.4.1, using the group average imposes the restriction that all group members have the same effect. Alternatively, one could introduce a weighting scheme into the function. However, since the Union army sample I use does not provide any information on the strength of ties between different group members, I weight equally every member of the group.

4.2.2. Identification and Estimation Using Time-Series Data

The time-series nature of the available sample provides a powerful advantage for overcoming one of the main obstacles most empirical work in the area of social interactions face. By observing individuals during more than one time period, one can overcome the foremost criticism most empirical studies of social interactions must address- that an unobserved group characteristic is, in fact, behind the findings that some observed group characteristics have a statistically significant effect. In each of the groups (military units), all men (subject to being found by those compiling the data) are observed multiple times over potentially more than half a century.

I present here the derivation for the case in which there are two post-assignment time periods $t = 1, 2$, since the variables I am interested in are available in two post-assignment time periods. However, this could easily be extended to include more time periods. Graham and Hahn (2004) examine a number of specifications and show identification in panel data settings and provide a GMM framework to estimate some of the effects. I consider a simple two-period linear setting and show that the coefficients of interest can be derived from the coefficients of an ordinary-least-squares estimation.

Consider the following linear specification:

$$\begin{aligned} y_{i,g,t=1} &= \alpha + x_{i,g,t=1} \cdot \beta + \rho m_{g,t=1} + \mu_g + \theta_i + \epsilon_{i,g,t=1} \\ y_{i,g,t=2} &= \alpha + x_{i,g,t=2} \cdot \beta + \rho m_{g,t=2} + \mu_g + \theta_i + \epsilon_{i,g,t=2} \end{aligned}$$

where $m_g = E[y|Subgroup\ g]$, x is a vector of covariates and ρ, the endogenous group effect, is the main parameter of interest. However, this system of equations also includes μ_g an unobserved group effect and an unobserved individual-level effect θ_i. The advantage of the time series setting is that both of the unobserved group and individual effect $(\mu_g+\theta_i)$ can be differenced (or averaged) out. As implied by the fact that these effects do not have a time script, this specification assumes that these two components do not vary over time, just across groups and individuals (fixed effects specification).

For simplicity, and to focus on purging out the unobservable group effect, it is assumed that there is no contextual effect ($\gamma = 0$). However, as discussed in Section 2.2, the sum of the contextual and endogenous effect is identified from the reduced form specification.

Differencing the two equations one obtains:

$$\Delta y_{i,g} = \Delta x_{i,g} \cdot \beta + \rho \Delta m_g + \Delta \epsilon_{i,g}$$

where $\Delta z = z_1 - z_2$ is the difference operator. Taking expectations with respect to x and under the assumption that $E[\Delta \epsilon_{i,g} | \Delta x_{i,g}] = 0$, one obtains:

$$\Delta m_g = \Delta E x_g \cdot \beta + \rho \Delta m_g$$

Rearranging terms and substituting back into the above equation we arrive at a specification which no longer contains the unobserved group effect:

$$\Delta y_{i,g} = \Delta x_{i,g} \cdot \beta + \frac{\rho}{1-\rho} \Delta E x_g \cdot \beta + \Delta \epsilon_{i,g} \tag{4.2}$$

Under the assumption that $E[\Delta \epsilon_{i,g} | \Delta x_{i,g}, \Delta \overline{X}_g]$, which will be discussed in the empirical part, the above specification can be consistently estimated.

One can derive the effect of interest, ρ, by applying the delta method to the coefficients from the above regression. Specifically, in the regression:

$$\Delta y_{i,g} = \Delta x_{i,g} \cdot \delta_1 + \Delta \overline{X}_g \cdot \delta_2 + \nu_{i,g} \tag{4.3}$$

For any component j of the vector x_i (and the corresponding component of the vector $\Delta\overline{X}_g$) ρ is a function of the jth component of the coefficients δ_1 and δ_2.
Hence $\rho = (1 + \frac{\delta_1^j}{\delta_2^j})^{-1}$.

Moreover, if the assumptions and specification of the model hold then one testable implication is that this result holds for any component of the vector x. The testable implication is then:
$\frac{\delta_1^j}{\delta_2^j} = \frac{\delta_1^h}{\delta_2^h}$ for all $j, h = 1..K$ components (where K is the size of the vector x).

4.3. Data

The data in this chapter are primarily based on the Union Army Study, a monumental data collection effort led by Fogel et al. (2000). It contains the members of 303 companies in the American Civil War Union Army that were randomly drawn from a sample of over 20,000 companies whose records are stored at the National Archives in Washington, D.C. The companies were all part of volunteer white infantry regiments and represent all of the participating states except for Rhode Island. The base sample consists of the military records of the 35,570 individuals in those companies. The sample was then linked to the censuses of 1860, 1870, 1880, 1900, and 1910, as well as other pension records.

In this chapter, I will only make use of the linked files of the 1880 and 1900 censuses as well as information from the pension files. I do not use the other census years since the 1870 census does not contain employment status (just occupation), and by the 1910 census the majority of the sample belongs to an age group which is highly likely to be retired or deceased. Note that the 1880 file has yet to be officially released and in assigning

Table 4.1: Summary Statistics for the Union Army Sample Linked to the 1880 Census

(*n=5,798*)

Variable	**Mean**	**Std. Dev.**
Age (in 1880)	42.43	8.08
Gainfully employed (in the labor force)	97.51%	15.56%
Months unemployed in the past year (for those gainfully employed)	0.13	0.99
Long-term unemployment rate (among those in the labor force)	1.44%	11.95%

and imputing values, I have attempted to be consistent with the 1900 data definitions and practices.

Of the original sample of more than 35,000 soldiers, 11,543 are linked to the 1900 census and almost 6,000 are linked to the census of 1880. According to Fogel et al. (2000), those linked do not statistically differ from those which the study was not able to link. Note that those linked to the 1900 and 1880 censuses are not a representative sample of the population in 1900 and 1880 respectively. Tables 4.1 and 4.2 provide summary statistics of the sample linked to the 1880 census and 1900 census. Table 4.3 provides summary statistics at the company level.

Each record has a match quality code denoting the likelihood that the 1900 census record is actually that of the veteran. I excluded 477 observations which were in the bottom two categories (out of 4). None of the results changed when the low quality link observations were used. I also excluded one observation for which there was no age information.

Table 4.2: Summary Statistics for the Union Army Sample Linked to the 1900 Census

(*n=11,544*)

Variable	**Mean**	**Std. Dev.**
Age (in 1900)	60.03	6.29
Gainfully employed (in the labor force)	68.65%	46.4%
Months unemployed in the past year (for those gainfully employed)	1.18	2.74
Long-term unemployment rate (among those in the labor force)	6.00%	23.76%
Literacy rate	92.65%	26.09%
Percent who reside in 1900 in the same state in which the company was formed in the 1860s	39.49%	48.88%

In addition, I have used the 1880 and 1900 census Integrated Public Use Microdata Series (Ruggles et al., 2004) to calculate the county and state unemployment rate, as well as the distribution of occupations among white males of the same age group.

4.4. Empirical Strategy

4.4.1. Military Company as Reference Group

Throughout this chapter, the reference group used is one's military company during the Civil War. There are several advantages which groups formed during military service provide (see also the discussion in Section 3.4.2). These include the circumstance and

Table 4.3: Summary Statistics for the Company-Level Characteristics of the Union Army Sample

Union army military companies in the sample (*n=303*)

Variable	**Mean**	**Std. Dev.**
Roster size of company (1860's)	117.12	36.30
Size of company in 1900 (number linked to the census)	36.51	14.76
Company participation rate in 1900 (in the labor force)	69.62%	30.16%
Company long-term unemployment rate in 1900 (among those in the labor force)	6.13%	8.71%
Long-term unemployment rate in 1900 (among those in the labor force)	6.00%	23.76%
Percent of company members killed during the Civil War	27.65%	27.59%

Note: Each company is equally weighted.

manner in which the groups were formed, scope and size of reference group, and strength of ties.

Unlike most other studies, the formation of networks in this case, while not completely exogenous in nature, is not entirely endogenous, and selection into companies is based on factors which are less likely to be correlated with the observed and unobserved variables used in this study. Companies were primarily formed based on geographical considerations, usually at the town or neighboring-towns level. Hence, those in small communities had little or no choice as to which company to join. Those in larger cities might have had a choice of a few companies. However, while those who had a choice might join with a few of their friends, it is not likely they would be able to coordinate the formation of a military company with over 100 other men.

Other factors which affected the company assignment were health (passing a high enough threshold), income (the very rich could buy their way out), and bounty size (some recruits enlisted in places where bounties were higher). Nonetheless, the Civil War presents a unique opportunity to study the outcomes and dynamics of groups that were formed by a strong external force, namely war. While not a "natural experiment" per se, the selection into groups could not have possibly been done for the purpose of improving labor market outcomes fifteen or thirty-five years later. Typically, network formation is costly, and an agent must make a decision how much to invest in the network. However, in this case, while in some sense networks were extremely costly, the cost was both fixed, and involuntary conditional on having served in the army.

Second, the groups formed during the war are likely to be significant, especially when sharing the experience of battle. Further, part of the essence of the groups was based on unit pride. This means that similar to other exclusive groups, membership was enough to create an affiliation, even if there was no direct contact among members.

There has been much attention focused on the importance of "weak ties." For instance, Granovetter (1974, 1995) finds that it is weak ties, such as a neighbor, as opposed to strong ties, such as a brother, which are responsible for most of the jobs found. Networks at the, for instance, company level, contain a relatively large number of individuals. This increases the likelihood that they would fall into the "weak tie" category and are therefore more likely to have an effect large enough to be estimated. Companies are a small enough group to allow its members to know each other. Yet, they are large enough to allow for a substantial number of potential ties.

Last, the high participation rate in the Civil War contributes not only to the external validity of the results, but also provides a large variability within group. While it is true that there is some truncation in the tails of the distribution of the men's characteristics (for example, the very rich are less likely to be part of the company), the company includes a large portion of the population. I argue that even the population of military-eligible men in a rural area exhibits more variance in their characteristics than the social networks we typically observe.

4.4.2. Employment Outcomes

Most of the results in this chapter will focus on examining the effect of social networks on the likelihood of being long-term unemployed. Following Margo (1991) and Moen (1994), I define long-term unemployment as being unemployed for 6 months or more during the past 12 months. This binary measure could be extended to look at the actual number of months unemployed.

There are several technical advantages to looking at long-term unemployment, as opposed to other unemployment measures such as length of spell. First, when looking at a one-year period over which the "months unemployed" question is answered in the 1880 or 1900 census, it is more likely that individuals' unemployment spells would overlap with each other. Second, by using the long-term unemployment measure the results are less likely to be driven by seasonal unemployment spells (especially for farmers). Last, if there is a cost associated with using the network for job search, the network is more likely to be utilized when the spell is longer.

When looking at long-term unemployment, the sample is restricted to those in the labor force (or gainfully employed, as defined in the late 1800s and early 1900s). To address this concern, the results include specifications in which the sample is restricted to those under 65. Similar results hold even for younger age groups. I also performed some analysis of how the decision to enter the labor force is affected by others decision. However, the results are only suggestive, as one would have to carefully account for the fact that the decision to enter the labor force might be influenced by the likelihood of finding a job conditional on being in the labor force. It would be of interest to examine in more detail the decision to retire, using additional information on pension benefits. Costa (1995) and Kanjanapipatkul (2003) have looked at the retirement decision of Union Army veterans using pension data but have not examined the effect of one's peers on the decision to retire.

4.4.3. Contextual Effects

In most of the topics examined by the literature on social influence, it is very likely that contextual effects play an important role. For instance, for children's achievements in school, the socio-economic background of the other students or the quality of the teacher can play an important role. In welfare use, the norms shared by the group could have a big influence on the propensity to choose welfare, and so on.

The time-series estimates control for any group attributes that remain constant over time, as explained in Section 4.2.2. However, for those specifications in which I use a single cross-section it is useful to consider how the group contextual effects would operate directly. I argue that these effects are not likely to be strong for the employment outcomes examined. Consider three possible contextual company measures which one might believe could affect employment: unit spirit, size of company, and war-induced stress or health status.

While it is certainly possible that different units enjoyed different levels of group cohesion, or unit spirit, it is less likely this would have any direct effect on employment 35 years later. Stronger bonds mean that the network would be more valuable, say others are more likely to help one get a job. Differences across companies in unit spirit might translate into different coefficients for the endogenous effect but would have no direct effect on one's likelihood of finding a job. Similarly, larger networks, should, if at all, only affect the way in which the networks operate. In the framework presented in Section 3.2.2 the size of the network should not matter, only the unemployment rate within the network. Intuitively, if both the network is larger and there are more unemployed, under a symmetric structure, the additional job offers generated by the bigger network must be shared among more unemployed. In Section 4.5.4 I empirically examine the effect of network size.

The results in Chapter 3 suggest that the endogenous effect is much stronger (see Section 3.6.1) than the contextual effect and one cannot reject that the contextual effect is equal to zero. One should be cautious about extrapolating from the results of another era and sample. Nonetheless, this could further motivate the assumption used in Section 4.5.1 that the contextual effect is zero.

Different companies have had different experiences during the war. Some suffered great losses, others might have been established later and therefore existed for a shorter period. Some might have had better or more influential commanders, etc. One possible proxy for the stress experienced by the company, is the percent of company members killed during the war (similarly, one could look at length of service, battles participated in, etc.). Any effect this might have had on the physical well being of the company members assumes that one's health depends on the number of others wounded or killed. Further, if such an effect did exist, it would be captured by the individual's health measures or reflected through the decision to participate in the labor force. The only area in which a contextual effect might exist is if war-related stress causes a change in behavior or mental well-being which adversely affects employment. I control for this issue in Section 4.5.3.

4.5. Results

Most of the results in this section will use the long-term unemployment measure, for reasons discussed in Section 4.4.2. The results in this section make use of the census of 1900 part of the sample, since the 1900 census is a little more modern. In addition, the unemployment rate was much higher in 1900 compared to 1880. As summarized by Fairlie and Sundstrom (1999) the unemployment rate for white males ages 16-64 was 3.8% in 1880 and 6.1% in 1900. A similar, and sharper trend can be seen in Tables 4.1 and 4.2 which summarize the unemployment rate for the Veterans sample in 1880 and 1900. The higher unemployment rate in 1900 means that the likelihood of being long-term unemployed is higher, and in turn the estimates are less likely to suffer from problems due to an event being very rare. Section 4.6, which makes use of the panel structure will also make use of the 1880 census information.

I first present a baseline specification which does not include any network effects. I then introduce network effects, followed by instrumental variable specifications to account for the simultaneity problem. Some robustness specifications include looking at other proxies for network outcomes, examining only those who migrated, younger men of the sample, and a "reduced form" specification which only examines the effect of other's covariates. In Section 4.6, I estimate peer effects by taking advantage of the time-series nature of the sample.

With the exception of the time series estimates, almost all of the results presented here are based on a probit specification , due to the binary nature of the outcome. The econometric specification can be written as:

$$\Pr(y_{i,g} = 1 | x_{i,g}, network_g) = \Phi(\alpha + x_{i,g}^{'}\beta + Z_g^{'}\gamma + \rho \cdot m_g^e) \tag{4.4}$$

where y is equal to one if the individual is long-term unemployed and zero otherwise (or equal to one if gainfully employed and zero otherwise), Φ is the normal cumulative distribution function, m_g^e is some measure of the network's outcome, for example $\bar{y}_{g,-i} = \frac{1}{n_g-1} \sum\limits_{j \in g, j \neq i}$, the long-term unemployment rate among all other company members self excluded ; $x_{i,g}$ is a vector of various personal characteristics and macroeconomic conditions which are likely to affect employment. These controls include, age, marital status, literacy, occupation and state fixed effects, or state unemployment rate for white men of the same age. The vector Z_g contains the group contextual effects which have been discussed above.

4.5.1. Base Specifications

Table 4.4 contains the base specification of equation (4.4), assuming there are no network effects by constraining $\gamma = 0$ and $\rho = 0$. Column 1 examines the 7,624 men who are gainfully employed and linked to the census of 1900. Only age is used as a regressor. The marginal effect of an additional year of age on the likelihood of being long-term unemployed is 0.004. The average long-term unemployment rate is 6%, and so at the average age (59 years), 1 additional year of age corresponds to an increase of 0.4 percentage points in the likelihood of long-term unemployment. Note that this result is conditional on being in the labor force, and that those older are less likely to be in the labor force. In column 2 occupation categories are included. Similar results hold when more specific occupation categories are used, as I do in some of the tables. The most notable thing is that farmers are far less likely to be long-term unemployed. This is consistent with the findings of other studies (Costa, 1995). Being literate has the expected effect of decreasing the likelihood of unemployment. The marital status fixed effects are quite noisy, and throughout, they are often not statistically significant, suggesting that marital status might not affect the likelihood of employment. However, it does affect the likelihood of participating in the labor force.

Column 3 adds state fixed effects which control for any state specific macroeconomic conditions. It also allows for correlation in the error terms within companies (though the error terms are assumed to be independent of the regressors, and the error terms of members of other companies). Column 4 includes company fixed effects. The marginal effects of the other coefficients remain very similar.

Table 4.4: Estimating Union Army Veterans' Long-Term Employment in 1900 Assuming No Peer Effects

Union army veterans sample linked to the Census of 1900
Probit specification; Dependent variable- Long-term unemployment

	(1)	(2)	(3)	(4)
Constant	-3.504	-2.637	-7.614	-12.161
Age	0.032***	0.033***	0.033***	0.034***
	(0.004)	(0.004)	(0.004)	(0.004)
	[0.004]	[0.003]	[0.003]	[0.003]
Occupation dummy- farmer		-0.794***	-0.791***	-0.875***
		(0.066)	(0.072)	(0.084)
		[-0.069]	[-0.068]	[-0.074]
Occupation dummy- Professional or proprietor		-0.629***	-0.621***	-0.677***
		(0.077)	(0.076)	(0.088)
		[-0.043]	[-0.042]	[-0.044]
Occupation dummy- Artisan		0.128**	0.122*	0.139*
		(0.064)	(0.067)	(0.076)
		[0.013]	[0.012]	[0.013]
Literate (=1 if yes, 0 otherwise)		-0.251**	-0.218**	-0.228*
		(0.099)	(0.105)	(0.126)
		[-0.028]	[-0.023]	[-0.024]
Marital status fixed effects	No	Yes	Yes	Yes
State of residence fixed effects	No	No	Yes	Yes
Military company fixed effects	No	No	No	Yes
Standard errors adjusted for clustering on company	No	No	Yes	Yes
Pseudo R^2	.0216	.1086	.1196	.1755
Observations	7624	7607	7499	6337

Notes: Standard errors in parentheses
(Those adjusted for clustering assume correlation within company and independence across companies)
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
Long-term unemployment defined as unemployed for more than 6 months in the past year.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
For the occupation categories, laborer is the omitted category.
Sample: All those linked to the 1900 census who are gainfully employed.

Before examining any group (network) characteristics, it is of use to examine whether or not membership in companies actually has any predictive power regarding employment. I examine the joint significance of an F-test of a specification that includes a dummy variable for each company. Table 4.5 contains two such specifications, one for 1880 and the other for 1900. In each of the years, after controlling for age, one of twelve occupation categories, and state of residence, membership in a company is found to have a statistically significant effect on the likelihood of being long-term unemployed. Any company fixed effect captures any characteristic of the company including group level statistics such the long-term unemployment rate among group members, and so these results should only serve as a motivation, rather than be of any direct interest.

In Table 4.6, network effects are included (though the contextual effect is constrained to $\gamma = 0$). The measure used is the average long-term unemployment rate among all members, self excluded. The long-term unemployment rate of others has a statistically significant effect in almost all specifications. Column 1 provides the base specification with networks omitted. In column 2, the group's long-term unemployment rate is included. The effect is statistically significant at the $p<0.1$ level. At the mean, where the average long-term unemployment rate is 5.97%, a 1-percentage-point increase in the long-term unemployment rate among company members would increase an individual's likelihood of being unemployed by 0.08 percentage points. In column 3 state fixed effects are added. The social effect now has large standard errors. This may be due to the fact that for smaller states there are not many companies in the sample, and so the network effect is captured by the state fixed effect. Instead of state fixed effects, columns 4-6 use the state long-term unemployment rate for males age 50-90, to capture the state macroeconomic

Table 4.5: Testing the Statistical Significance of the Military Company in Determining Employment in 1880 and 1900

Union army veterans sample linked to the Censuses of 1880 and 1900 OLS regressions; Dependent variable- Long-term unemployment		
Year	**1880**	**1900**
Controls	-Age -One of 12 occupation categories -only those gainfully employed -state of residence	-Age -One of 12 occupation categories -only those gainfully employed -state of residence
(joint) F-Value on the significance of the company dummy variable	**1.12**	**1.16**
P-value for the above null	0.08	0.03
Observations	5300	6778

Note: Long-term unemployment defined as unemployed for more than 6 months in the past year.

conditions.[2] The social effect coefficient is once again statistically significant at the $p<0.1$ level or better.

Columns 5 and 6 show that the result is robust when looking at a sub-sample. In column 5, only those who are under the age of 65 are used in the estimation (though the company long-term unemployment rate is calculated using all sample members). Those under age 65 are more likely to be participating in the labor force. Similar results are obtained when restricting the estimation to those under the age of 60. The even more surprising result is in column 6, where the sample is restricted to only those who currently

[2]For each state, I calculated the long-term unemployment rate of males ages 50-90, using the 1900 census sample available from the Integrated Public Use Microdata Series (Ruggles et al., 2004).

Table 4.6: Estimating Union Army Veterans' Long-Term Employment in 1900 with Peer Effects

Union army veterans sample linked to the Census of 1900
Probit specification; Dependent variable- Long-term unemployment

	(1)	(2)	(3)	(4)	(5) Under 65	(6) Only movers
Constant	-3.041	-2.691	-7.617	-2.829	-2.926	-3.444
Long-term unemployment rate among members of company (self excluded)		0.899** (0.436) [0.082]	0.604 (0.460) [0.054]	0.880** (0.436) [0.080]	0.819* ++ (0.488) [0.069]	1.210* ++ (0.709) [0.091]
Age	0.034*** (0.004) [0.003]	0.033*** (0.004) [0.003]	0.033*** (0.004) [0.003]	0.033*** (0.004) [0.003]	0.034*** (0.007) [0.003]	0.036*** (0.007) [0.003]
Occupation dummy- Farmer	-0.807*** (0.065) [-0.071]	-0.784*** (0.066) [-0.068]	-0.785*** (0.068) [-0.067]	-0.780*** (0.066) [-0.068]	-0.762*** (0.074) [-0.060]	-0.641*** (0.114) [-0.046]
Occupation dummy- Professional or proprietor	-0.641*** (0.077) [-0.045]	-0.624*** (0.078) [-0.043]	-0.617*** (0.079) [-0.042]	-0.622*** (0.078) [-0.043]	-0.558*** (0.084) [-0.036]	-0.569*** (0.131) [-0.034]
Occupation dummy- Artisan	0.119* (0.064) [0.012]	0.131** (0.064) [0.013]	0.125* (0.065) [0.012]	0.131** (0.064) [0.013]	0.130* (0.071) [0.012]	0.134 (0.115) [0.011]
Literate (=1 if yes, 0 otherwise)	-0.247** (0.099) [-0.028]	-0.251** (0.100) [-0.028]	-0.218** (0.102) [-0.023]	-0.251** (0.100) [-0.028]	-0.285** (0.112) [-0.030]	-0.325* (0.188) [-0.032]
Marital status fixed effects		Yes	Yes	Yes	Yes	Yes
State of residence fixed effects		No	Yes	No	No	No
long-term unemployment state rate among whites 50-90 (from IPUMS sample)				3.218* (1.802) [0.293]	2.792 (2.008) [0.234]	3.060 (2.328) [0.231]
Pseudo R^2	.1054	.1097	.1209	.1106	.0907	.0932
Observations	7624	7597	7489	7588	6502	2976

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
For the occupation categories, laborer is the omitted category.
Long-term unemployment defined as unemployed for more than 6 months in the past year.
Sample: All those linked to the 1900 census who are gainfully employed.
Column (5): Only those under 65; Column (6): Only those residing in a different state than enlistment state.
++ In the sub-samples, the network measure is based on the full sample

reside in a state different than the state in which the company was formed. Roughly 40% of the sample had moved across state lines. This suggests that the transmission of information extended past state lines, or that the migration decision was influenced by that of one's peers.

The peer effect persists when using different measures of employment for the company, as can be seen in Table 4.7. All specifications include occupation, marital, and state fixed effects. Column 1 is the base specification. In column 2 instead of the long-term unemployment rate among those in the labor force, I use as a denominator all those in the company, whether or not in the labor force. This is meant to capture the fact that those not in the labor force who learn of jobs might pass them on to those searching for a job. Using the notation of Section 3.2.2, this corresponds to n_g being equal to the company size, where as the previous results treated n_g as the number of those in the labor force. Using this measure, the effect is statistically significant and much larger. At the mean, a 1-percentage-point increase in this company's new unemployment measure, would lead to an increase of 0.6 percentage points in one's likelihood of long-term unemployment.

In columns 3 and 4, those not in the labor-force are not treated as potential information providers but rather as competing for the same information. For instance, those not in the labor force might be discouraged unemployed men who have stopped searching. The size of the effect is comparable to the previous results, where the long-term unemployment rate was used. Column 4 conditions only on those who are under 65 (though the company rate is calculated using the entire sample). The results for the restricted sample are similar to those of the full sample.

Table 4.7: The Effect of Alternative Group-Employment Measures on Veterans' Long-Term Employment in 1900

Union army veterans sample linked to the Census of 1900
Probit specification; Dependent variable- Long-term unemployment

	(1)	(2)	(3)	(4) Under 65
constant	-7.614	-7.654	-7.735	-7.838
Percent long-term unemployed among all company members (both in and not in labor force)		8.666*** (0.709) [0.654]		
Percent who are long-term unemployed or not in labor force (among all company members)			0.824*** (0.202) [0.070]	0.692*** (0.221) [0.054]
Age	0.033*** (0.004) [0.003]	0.032*** (0.004) [0.002]	0.034*** (0.004) [0.003]	0.036*** (0.008) [0.003]
Occupation dummy- Farmer	-0.791*** (0.068) [-0.068]	-0.770*** (0.072) [-0.057]	-0.776*** (0.070) [-0.064]	-0.769*** (0.080) [-0.058]
Occupation dummy- Professional or proprietor	-0.621*** (0.079) [-0.042]	-0.659*** (0.087) [-0.036]	-0.643*** (0.084) [-0.041]	-0.563*** (0.091) [-0.034]
Occupation Dummy- Artisan	0.122* (0.065) [0.012]	0.120* (0.070) [0.010]	0.123* (0.068) [0.011]	0.132* (0.076) [0.011]
Literate (=1 if yes, 0 otherwise)	-0.218** (0.102) [-0.023]	-0.238** (0.108) [-0.022]	-0.219** (0.105) [-0.022]	-0.239** (0.120) [-0.023]
Marital and State of residence fixed effects	Yes	Yes	Yes	Yes
Pseudo R^2	.1196	.1698	.1263	.1049
Observations	7499	7046	7046	6011

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
For the occupation categories, laborer is the omitted category.
Long-term unemployment defined as unemployed for more than 6 months in the past year.
Sample: All those linked to the 1900 census who are gainfully employed.
Column (2)-(4): Only companies larger than 20 included.
Column (4): Only those under 65.
In all cases, the network measure is based on the entire sample
(e.g. In Column 2, long-term unemployment rate is measured by including all those not in the labor force)

4.5.2. Instrumenting for Simultaneity

Even under the assumption of no contextual effect, the estimate of the effect of the endogenous variable is, in principal, likely to be biased due to the simultaneity problem. In this Section, I address the fact that equation (4.4) is actually a system of simultaneous equations, as each outcome is also part of the other's reference group outcome. My experience has been that while using Maximum-Likelihood one could estimate the model, and this is a the most efficient estimation in theory, in actuality, the non-linear-optimization issues lead to questionable results in some cases. Part of the reason being, the company size is rather large and one is therefore faced with a large system of equations. A similar approach is the lag spatial model approach (such as Anselin, 2001). However, as sample size and the size of the group increases, estimation becomes very unreliable in my experience, since one needs to estimate a system of equations with as many equations as the size of the group.

Instead, the results presented here are based on two-stage least-squares estimation, which depending on the specification might not be efficient but is far less computationally intensive. In this section, I instrument the group's average long-term unemployment outcome and discuss some of the problems which are unique to instrumental-variable estimation in the presence of social influences in addition to the usual concerns regarding the choice and validity of the instrument.

Table 4.8 presents the results for a linear probability specification:

$$y_{i,g} = \alpha + x_{i,g}\beta + \frac{1}{n_g - 1} \sum_{j \in g, j \neq i} y_j + \epsilon_{i,g} \tag{4.5}$$

Table 4.8: Instrumental Variables Estimates of the Effect of Peers' Employment on Veterans' Long-Term Employment in 1900

Union army veterans sample linked to the Census of 1900
Linear probability specification; Dependent variable- Long-term unemployment

	(1)	(2)	(3)	(4)
IVs used	No	No	Average age among co. members (self excluded)	Average age and percent in each occupation among co. members (self excluded)
Constant	-0.220	-0.225	-0.230	-0.229
Long-term unemployment rate among members of company (self excluded)		0.123** (0.063)	0.291 (0.268)	0.251** (0.121)
Age	0.005*** (0.001)	0.005*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
Literate (=1 if yes, 0 otherwise)	-0.014 (0.015)	-0.014 (0.015)	-0.015 (0.015)	-0.015 (0.015)
Long-term unemployment state rate among whites 50-90 (from IPUMS sample)	0.302* (0.170)	0.297* (0.165)	0.276 (0.168)	0.280* (0.161)
Occupation fixed effects (10 categories)	Yes	Yes	Yes	Yes
Standard errors adjusted for clustering on company	Yes	Yes	Yes	Yes
Hansen J statistic (over identification test) Null: Valid instruments				5.785 (P-val = 0.76) Null is not rejected
Observations	6797	6789	6789	6787
R-squared	0.06	0.06		

Notes: Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Long-term unemployment defined as unemployed for more than 6 months in the past year.
Sample: All those linked to the 1900 census who are gainfully employed and have an occupation category.

While less desirable since the outcome variable is binary, as will be seen below, linearity provides for a better fit in the second stage of the estimation. Moreover, this specification is easily estimated using the usual linear two-stage estimator. Last, the linear procedure allows me to perform an over-identification test for the validity of the instruments, as explained below. Columns 1 and 2 present the base specification, with the latter including the peer effect. The results are similar to those obtained when using the probit specification.

The challenge is to instrument $\overline{y}_{g,-i} = \frac{1}{n_g - 1} \sum\limits_{j \in g, j \neq i} y_j$ as a way to overcome the simultaneity problem. The first instrumental variable considered is presented in column 3. The instrumental variable used is the average age among company members (self excluded). This is not likely to be correlated with the error term and is statistically significant in determining the average unemployment rate. However, it is a weak instrument and the standard errors of the coefficient are very large.

Ideally, one would want to first fit each y: $\widehat{y} = x\widehat{\beta}$ using only the covariates and then instrument $\overline{y}_{g,-i}$ with $\overline{\widehat{y}}_{g,-i}$, the average of these fitted values. If other company members' x's are uncorrelated with one's own this should yield a valid instrument. Next, note that if the first stage of the two-stage-least-squares procedure is linear then the average of the fitted values is almost equivalent to the fit of the average. This is the measure I use in column 4. The instruments I use are the average age of company members (self excluded) and the percent of company members in each occupation category (there are 10 categories). The instruments are: $\overline{x}_1, \overline{x}_2, ...\overline{x}_{10}$ where $\overline{x}_1$is the average age, and $\overline{x}_2, ...\overline{x}_{10}$ are the percent in each occupation (with one occupation omitted). This amounts to having occupation fixed effects for each individual if I were to fit each individual separately.

Using these instrumental variables the coefficient on the company's long-term unemployment rate is now statistically significant and quite large. The point estimate suggests that a 1-percentage-point increase in the company's long-term unemployment rate would cause an increase of 0.25 percentage points in one's likelihood of being long-term unemployed. Though, the lower bound of the confidence interval suggests the effect might not be as strong. This effect is after controlling for age, literacy, occupation category, and the long-term unemployment rate in one's state of residence.

The instruments are "strong," but are they valid? Since I include an occupation fixed effect for each individual, I assert that $\overline{x}_2, ...\overline{x}_{10}$ are not likely to be correlated with the unobserved error term directly. However, one might be concerned that the occupation itself is endogenous and might be affected by the choice of occupation of one's peers. Taken with caution, the use of ten instruments allows me to perform an over identification test, from which I can not reject that the instruments are valid. The P-value is 0.76.

Table 4.9 uses the same approach discussed above, but now the second stage uses a probit specification, which provides a better fit. However, this specification should be interpreted with caution since in addition to the usual needed assumptions for a valid instrument, a two-stage model in which the second stage is probit, might lead to a logically inconsistent model (see Krauth, 2006).

When only average age in company is used as an instrument (column 2), the standard errors for the network effect are very large, suggesting the instrument is "weak." Once "percent in categories" is included in columns 3-4, the effect is now statistically significant at the $p<0.1$ level. The magnitude is very similar to that of the linear probability case, discussed above. While the point estimates are considerably larger when instrumented,

Table 4.9: Instrumental Variables with Second-Stage Probit Estimates of Veterans' Long-Term Employment in 1900

Union army veterans sample linked to the Census of 1900
Second Stage Probit specification; Dependent variable- Long-term unemployment

	(1)	(2)	(3)	(4)
Instrumental variables used	No	Average age among co. members (self excluded)	Average age and percent in each occupation among co. members (self excluded)	Average age and percent in each occupation among co. members (self excluded)
Long-term unemployment rate among members of company (self excluded)	1.000** (0.450) [0.086]	4.034 (3.256) [0.346]	2.482* (1.294) [0.214]	2.477* (1.434) [0.209]
Age	0.036*** (0.004) [0.003]	0.035*** (0.004) [0.003]	0.036*** (0.004) [0.003]	0.036*** (0.004) [0.003]
Literacy and Occupation fixed effects (10 categories)	Yes	Yes	Yes	Yes
Long-term unemployment state rate among whites ages 50-90 (from IPUMS sample)	3.227* (1.908) [0.277]	2.848 (1.954) [0.244]	3.026 (1.921) [0.261]	
State of residence fixed effects	No	No	No	Yes
Constant	-3.886	-3.988	-3.931	-8.697
Pseudo R-squared	.1251			
Observations	6789	6789	6787	6698

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Long-term unemployment defined as unemployed for more than 6 months in the past year.
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
Sample: All those linked to the 1900 census who are gainfully employed and reported an occupation

the standard errors are large and one cannot reject that the magnitude of the effect is similar to the one presented in Tables 4.6 and 4.7.

4.5.3. Reduced Form and Robustness Checks

This section presents various robustness specifications. As discussed in Section 2.2, by examining the effect of others (exogenous) covariates, one could identify the composite effect (composed of the endogenous and contextual effects) with far less restrictive assumptions. Some results are presented in Table 4.10. This reduced-form specification can be written as:

$$y_{i,g} = h(\alpha + x_{i,g}^{'}\beta + \bar{X}_{g}^{'}\gamma_{1r} + Z_{g}^{'}\gamma_{2r} + \epsilon_{i,g})$$

The distinction between the group's average of individual characteristics $\bar{X}_g$, and Z_g, the group's characteristics is somewhat arbitrary, and mainly included for the consistency of the notation in other sections. The coefficients γ_r are indexed with r as they are only a measure of whether a social influence exists. If there is no endogenous effect their magnitude can be interpreted as the contextual effect.[3]

In all specifications of Table 4.10, the effect of $\bar{X}_{g,-i} = \frac{1}{n_g - 1} \sum\limits_{j \in g, j \neq i} x_j$ is examined. In column 1, only own age is included the covariates. Not surprisingly, the fit is very small, and the network measure is not statistically significant from zero. The effect becomes positive, and statistically significant at the p<0.1 level when additional controls are added, as seen in column 2. The magnitude of the point estimate of the effect of the average company age is similar to the marginal effect of one's own age. The effect persists when conditioning only on those who reside in a different state than the state in which

[3] For example, in the linear-in-means case if there is no endogenous effect ($\rho = 0$), then it must follow that $\gamma_{1r} = 0$, and γ_{2r} is the contextual effect.

Table 4.10: Reduced-Form Estimates of the Effect of Peers on Veterans' Long-Term Employment in 1900

Union army veterans sample linked to the Census of 1900
Probit specification; Dependent variable- Long-term unemployment

	(1)	(2)	(3)- Those moved	(4)	(5) Under 65
Average age of company members++ (self excluded)	0.010 (0.013) [0.001]	0.024* (0.013) [0.002]	0.056*** (0.021) [0.004]	0.015 (0.014) [0.001]	0.030** (0.015) [0.002]
Age	0.032*** (0.004) [0.004]	0.033*** (0.004) [0.003]	0.034*** (0.008) [0.002]	0.033*** (0.004) [0.003]	0.032*** (0.008) [0.003]
Literacy		-0.213** (0.106) [-0.023]	-0.163 (0.194) [-0.014]	-0.248** (0.104) [-0.028]	-0.251** (0.118) [-0.025]
Occupation dummy variable- Farmer		-0.806*** (0.072) [-0.070]	-0.667*** (0.125) [-0.046]	-0.809*** (0.071) [-0.071]	-0.804*** (0.081) [-0.062]
Occupation dummy- Professional or proprietor		-0.635*** (0.076) [-0.043]	-0.594*** (0.139) [-0.034]	-0.637*** (0.075) [-0.044]	-0.591*** (0.081) [-0.037]
Occupation dummy- Artisan		0.114* (0.067) [0.011]	0.116 (0.120) [0.009]	0.125* (0.067)	0.112 (0.077) [0.010]
Percent of company who are farmers++				0.095 (0.251) [0.009]	
Percent of company who are Professional or proprietors++				0.008 (0.305) [0.001]	
Percent of company who are Artisans++				0.084 (0.326) [0.008]	
Constant	-4.088	-8.796	-11.133	-3.952	-9.087
State of residence controls	No	Yes	Yes	No	No
Standard errors adjusted for clustering on company	No	Yes	Yes	Yes	Yes
Pseudo R^2	.0218	.1169	.1223	.1058	.1012
Observations	7622	7514	2857	7614	6170

Notes: Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
Those adjusted for clustering assume correlation within company and independence across companies
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
For the occupation categories, laborer is the omitted category.
Long-term unemployment defined as unemployed for more than 6 months in the past year
Sample: All those linked to the 1900 census who are gainfully employed
Column 3: Only those under 65; Column 5: Only those residing in a different state than enlistment state.
++ In the sub-samples, the network measure is based on the full sample

their company was formed (column 3), or those under 65 (column 5). In both cases, the network measures are calculated using the entire sample. In column 4 additional network measures are added, the average rate of participation in four broad occupational categories. However, the results are not statistically significant and one might be concerned that others' choice of occupation is not exogenous.

I conclude the robustness checks with the inclusion of company characteristics which could potentially affect the unemployment likelihood, as discussed in Section 4.4. These are reported in Table 4.11. All specifications in Table 4.11 include age, and literacy, occupation, and state of residence fixed effects. The long-term unemployment rate among company members is instrumented.

The measure used in column 1 is the percent of men killed during the Civil War. It was calculated as the number killed divided by the original company size. The measure was truncated at 100% (as some companies had replacements during the war and had suffered more loses than the initial roster size). This measure is a proxy for the stress suffered during the war. The effect is not statistically significant and the standard errors are quite small. The point estimate of the marginal effect implies that an increase of 50 percentage points in the death rate, would only increase the probability of being long-term unemployed by an additional 0.3 percentage points. In column 2, I use as a measure of stress suffered the total number of men killed, instead of percent. I find similar results. If in fact the two measures are part of the contextual effect, then it is possible that they enter through the endogenous effect, as the model is then mis-specified. Nonetheless, the finding that the two are not statistically significant and are very small suggests that the contextual effect is of second-order compared to the endogenous effect.

Table 4.11: Instrumental Variables Estimates of the Effect of Peers' Employment with Additional Company-Level Controls

Union army veterans sample linked to the Census of 1900
Probit specification in second stage; Dependent variable- Long-term unemployment

	(1)	(2)	(3)
IV	Average age and percent in each occupation among co. members (self excluded)	Average age and percent in each occupation among co. members (self excluded)	Average age and percent in each occupation among co. members (self excluded)
Long-term unemployment rate among members of company (self excluded)	2.413 (1.489) [0.204]	2.817* (1.510) [0.236]	3.389** (1.479) [0.283]
Percent killed during the Civil War+	**0.071** (0.114) [0.006]		
Number of company members killed during the Civil War		**-0.00007** (0.00095) [-0.00006]	
Company state of origin (1860's) fixed effects	No	No	**Yes**
Constant	-8.775	-8.943	-8.265
Age	0.036*** (0.004) [0.003]	0.036*** (0.004) [0.003]	0.036*** (0.004) [0.003]
Literacy and occupation fixed effects (10 categories)	Yes	Yes	Yes
State of residence fixed effects	Yes	Yes	Yes
Pseudo R^2	.1353	.1330	.1419
Observations	6698	6520	6614

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
Long-term unemployment defined as unemployed for more than 6 months in the past year
Sample: All those linked to the 1900 census who are gainfully employed.
+ percent killed was calculated as number killed divided by the original company size truncated from above at 100% (some companies had replacements during the war)

Column 3 includes fixed effects to control for company's state of origin, that is, where the company was formed (and the veteran grew up, most likely). The results remain largely unchanged. For instance, this can be seen when comparing to the results reported in column 4 of Table 4.9.

4.5.4. Estimating Network Characteristics and Their Effect

In the model presented in Section 3.2.2, it was assumed that the network had constant returns to scale. However, it is possible that the "structural" attributes of the network matter. For instance, larger networks might be more or less efficient. This is related to the assumption used in some of the above specifications in which the characteristics of the network, the contextual effects, enter only through the coefficient on the endogenous effect. The size of the company, for example, is a contextual attribute of the network. However, it is likely only to affect how the network operates and not have a direct effect on the probability of getting a job.

In Table 4.12, I present the estimation results for two sub-samples. One sub-sample contains those companies which are in the first quartile of company size. The other contains the fourth quartile of company size. If in fact the size of the company should not matter, then the coefficients for the effect of others' long-term unemployment rate (first row), should be the same. While the results are not statistically significant for the larger companies, the upper bound of the 95% confidence interval for the coefficient of the large companies sub-sample is smaller than the lower bound of the confidence interval for the small companies' coefficient.

Table 4.12: Strength of the Peer Effect for Different Group Sizes

Union army veterans sample linked to the Census of 1900 Probit specification; Dependent variable- Long-term unemployment		
	(1)	(2)
Sample consists of only those in size category	Smallest quartile (smallest companies)	Largest Quartile (largest companies)
Long-term unemployment rate among members of company (self excluded)	6.824*** (2.025) [0.849]	2.662 (5.141) [0.388]
Age	0.022 (0.014) [0.003]	0.036*** (0.008) [0.003]
Constant	-2.179	-6.157
State of residence, occupation (10 categories), and literacy fixed effects	Yes	Yes
IV used	Average age and percent in each occupation among co. members (self excluded)	Average age and percent in each occupation among co. members (self excluded)
Observations	685	2118

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Marginal effects [dF/dx] at the mean value of x in square brackets.
For binary (dummy) variables, the marginal effect is calculated for the change: F(x=1)-F(x=0)
Long-term unemployment defined as unemployed for more than 6 months in the past year
Sample: All those linked to the 1900 census who are gainfully employed and reported an occupation

I do not provide any statistical test (since the result is not significant for the large companies), but this result may hint at the difference of the effect between large and small companies. Namely, large networks, seem to exhibit, all else equal, a smaller effect.

4.6. Differencing Out the Unobserved Group Effect

The time-series nature of the sample provides an opportunity to address one of the primary concerns most empirical work in the area of social interactions face. As explained in Section 4.2.2, one can purge out any fixed (over time) unobserved group effect. In addition, any individual-level fixed (over time) unobserved characteristic would also be differenced out. Recall from Section 4.2.2 that these two fixed unobserved components, μ_g and θ_i, are allowed to possibly be correlated with any of the other variables of the model.

The basic equation estimated in this section is of the form:

$$\Delta y_{i,g} = \Delta x_{i,g} \cdot \delta_1 + \Delta \overline{X}_g \cdot \delta_2 + \Delta \varepsilon_{i,g} \tag{4.6}$$

where $\Delta Z_{i,g} = Z_{i,g,year=1900} - Z_{i,g,year=1880}$ and recall that δ_2 can be interpreted as the composite group effect $\delta_2 = \dfrac{\rho\beta + \gamma}{1-\rho}$, or under the assumption that there is no contextual effect, $\delta_2 = \dfrac{\rho\beta}{1-\rho}$. In the latter case, one could then solve for the endogenous group effect by looking at the components of the coefficients δ_1 and δ_2.

For any component j of the coefficients of the vectors $\Delta x_{i,g}$ and $\Delta \overline{X}_g$:

$$\rho = (1 + \frac{\delta_1^j}{\delta_2^j})^{-1}$$

The main assumption needed for a consistent estimate is:

Assumption 4. $E[\Delta\varepsilon_{i,g} | \Delta x_{i,g}, \Delta\overline{X}_g] = 0$.

This assumption requires that any change in the transitory shock between the two time periods (1880 and 1900) must be uncorrelated with the change in observable characteristics, both at the individual and group level. While this assumption might seem very restrictive at first, it is important to note that the change in transitory shock $\Delta\varepsilon_{i,g}$, is purged of any fixed group or individual unobserved effect.

Hence, any individual or group characteristic which has been stabilized by 1880 and does not change its effect in 1900 would not present a problem for the validity of the results. This would be true even if these unobservables were correlated with any of the observables or the transitory shock. For instance, any group experience forged during the Civil War or any individual level ability that was formed during the veteran's childhood and teen years would be removed, even if those effects were correlated with, for example, age or the percent of company members who ended up working in a certain occupation in 1880.

Table 4.13 reports the results for the specification in equation (4.6). It examines how changes between 1880 and 1900 in individual and group characteristics affect the change in long-term unemployment (defined as being unemployed for more than 6 months in the past year) between 1880 and 1900.

Note that any variable, whether at the individual level or the company level, that is the same during those two data points is not included in the regression. Similarly, any variable, such as age, which has a constant difference between the two periods for every person in the sample is omitted (as it would be just absorbed by the constant term). In contrast, the difference in the variable age-squared across time would depend on one's age, and hence is included. Each variable is included at the individual level and the

Table 4.13: Changes in Long-Term Employment between 1880 and 1900 Controlling for Individual and Group Unobservables

Union Army Veterans linked to the Censuses of 1880 and 1900; OLS Regressions
Dependent variable- Change in long-term unemployment ($y_{1900} - y_{1880}$)

	(1)	(2)	(3) Only Non Farmers+
($\Delta Z = Z_{1900} - Z_{1880}$)			
Change in age-squared divided by 100	0.013*** (0.002)	0.013*** (0.003)	0.019*** (0.005)
change in county long-term unemployment rate	0.412*** (0.112)	0.447*** (0.119)	0.355 (0.222)
Change in farmer (yes/no) status	**-0.027** (0.012)**		
Change in average of (age-squared divided by 100) across company	-0.003 (0.003)	-0.003 (0.003)	-0.005 (0.004)
Change in average of county long-term unemployment rate across company	0.036 (0.263)	0.008 (0.270)	0.260 (0.468)
Change in % of company members who are farmers	**0.092** (0.041)**		0.106* (0.064)
Change in occupational income score		-0.000 (0.001)	
Change in average occupational income score across company		-0.002 (0.002)	
Constant	-0.160	-0.162	-0.224
Observations	2533	2446	1101
R-squared	0.02	0.02	0.03

Notes: Robust standard errors in parentheses adjusted for clustering- assume correlation within company and independence across companies
* significant at 10%; ** significant at 5%; *** significant at 1%
Sample: All those linked to the 1880 and 1900 censuses who are gainfully employed (in both years). The group averages were computed using all those linked to the sample year.
Long-term unemployment defined as unemployed for more than 6 months in the past year.
+ Only those who were not farmers in 1880 and 1900 were included, though the company measures (including percent of farmers) were calculated using the entire sample.
Those variables which are constant over time or change in a fixed linear way (such as age and age twenty years later) were not included in the regression.

company-average level. Specifications in which the company measure ($\overline{X}_{-i,g}$) excludes the individual yield similar results.

Examining the results in column 1, at the individual level, the coefficients on the change in age-squared, the county unemployment rates,[4] and the change in being a farmer are all statistically significant at the $p<.05$ level or better. The signs of the coefficients are as one would expect. For instance, those who were older in 1880 (and 1900) are more likely to be long-term unemployed, as are those who experience an increase in the unemployment rate in their county. Those who become farmers are less likely to become long-term unemployed. Of the group level variables, the only one that is statistically significant is the percent of company members who are farmers.

Using the coefficients on change in farmer status for the individual and the corresponding coefficient for the change in percent who are farmers (both marked in bold in Table 4.13), one can compute ρ, the "endogenous" effects of others' unemployment, as explained above: $\rho = (1 + \frac{\delta_1}{\delta_2})^{-1}$. Using the delta method the effect is found to be very large and bounded away from zero. In fact, the lower bound of the 95% confidence interval of the estimate of ρ is 0.8. This lower bound varies across specifications and is not measured very accurately. However, in all the specifications I have tried, the lower bound exceeds 0.4.

Other specifications included occupational categories (besides farming), marital status, and location (at the state level). The results seem to be similar across these specification. More generally, all of the specifications suffer from a low level of explanatory power (as implied by the low R-squared values). In most of the cases, the individual-level unobserved effect (which these regressions difference out) accounts for over half of the unexplained variance.

[4]For each county, I calculated the long-term unemployment rate of males ages 50-90, using the 1880 and 1900 census samples available from the Integrated Public Use Microdata Series (Ruggles et al., 2004).

In column 2 of Table 4.13, a similar specification is examined. Instead of looking at occupation categories, or a farmer indicator, as in column 1, this specification uses the individual occupational income score and the corresponding group average. The occupational income score ranges from 3 to 80, and is a measure of how well-paying a job is (higher scores correspond to better paying jobs).[5] This variable is not found to be statistically significant at either the individual or the group level. This result seems to be consistent across many of the specifications. It implies that long-term unemployment is more dependent on the sector of occupation, such as farming, rather than on the socio-economic status of the job.

Column 3 repeats the specification of column 1 only for those who were not farmers during both 1880 and 1900. The one striking result is that the percent of farmers in one's company is still found to be statistically significant. This can be viewed as additional evidence that the employment status of others matters as captured by this reduced-form specification.

To address the fact that there are only 3 possible values for the dependent variable when looking at the difference of a binary variable, such as long-term unemployment, in Table 4.14 I examine a potentially more suitable functional form, a multinomial logit. However, this specification does not difference out the unobserved effects correctly, and hence the coefficients are likely to be biased. Still, it provides for a better fit and the size and magnitude of the coefficients correspond to what one would expect. This specification provides additional evidence for the existence of peer effects. For instance, the change in the average county long-term unemployment rate among all company members

[5]This measure is based on the median wage of an occupation in 1950.

Table 4.14: Multinomial Logit Estimates of Changes in Long-Term Employment between 1880 and 1900

Union Army Veterans linked to the Census of 1880 and 1900
Dependent variable- Change in long-term unemployment ($y_{1900} - y_{1880}$)

	Changed From Long-Term Employed to Long-Term Unemployed ($y_{1900} - y_{1880} = -1$)	Changed From Long-Term Unemployed to Long-Term Employed ($y_{1900} - y_{1880} = 1$)
$(\Delta Z = Z_{1900} - Z_{1880})$		
Change in age-squared divided by 100	-0.013 (0.109)	0.193*** (0.032)
Change in county long-term unemployment rate	-23.021*** (5.604)	4.303* (2.207)
Change in farmer (yes/no) status	0.794 (0.522)	-0.364** (0.183)
Change in average of age-squared divided by 100 across company	0.099 (0.147)	-0.026 (0.051)
Change in average of county long-term unemployment across company	-15.296 (12.206)	-2.061 (5.378)
Change in % of company members who are farmers	0.359 (1.986)	1.601** (0.703)
Constant	-5.825	-6.331
Observations	2533	
Pseudo R-squared	0.06	

Comparison Group: Those who had the same long-term employment status.

Notes: (Only 4 individuals were Long-term unemployed in both years.)
Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Long-term unemployment defined as unemployed for more than 6 months in the past year.
Sample: All those linked to the 1880 and 1900 censuses who are gainfully employed (in both years). The group averages were computed using all those linked to the sample year.

Table 4.15: Changes in Length of Unemployment Spells between 1880 and 1900 Controlling for Individual and Group Unobservables

Union Army Veterans linked to the Census of 1880 and 1900; OLS Regressions
Dependent variable- Change in number of months unemployed in the last year ($y_{1900} - y_{1880}$)

	(1)	(2) Only Non Farmers+
$(\Delta Z = Z_{1900} - Z_{1880})$		
Change in age-squared divided by 100	0.135*** (0.029)	0.255*** (0.056)
Change in average of (age-squared divided by 100) across company	-0.034 (0.038)	-0.033 (0.056)
Change in county long-term unemployment rate	3.244** (1.312)	2.118 (2.539)
Change in average of county long-term unemployment rate across company	-4.772 (3.246)	-2.237 (5.337)
Change in farmer (yes/no) status	-0.607*** (0.133)	
Change in % of company members who are farmers	0.971** (0.412)	1.097 (0.674)
Constant	-0.846	-2.601
Observations	2442	1024
R-squared	0.03	0.03

Robust standard errors in parentheses adjusted for clustering- assume correlation within company and independence across companies
* significant at 10%; ** significant at 5%; *** significant at 1%
Sample: All those linked to the 1880 and 1900 censuses who are gainfully employed (in both years). The group averages were computed using all those linked to the sample year.
In two cases, over 12 months of unemployment reported were imputed as 12 months.
+ Only those who were not farmers in 1880 and 1900 were included, though the company measures (including percent of farmers) were calculated using the entire sample.
Those variables which are constant over time or change in a fixed linear way (such as age and age twenty years later) were not included in the regression.

has a statistically significant effect on the likelihood that an individual "switches" from unemployment to employment (or vice versa). I find similar results when I examine a logit fixed effects specification which correctly accounts for the unobserved fixed effect.

Table 4.15 examines a slightly different dependent variable. Here, the dependent variable is the difference between the two years of the actual number of months unemployed. Hence the possible range of this measure is -24 to 24. The results in the two columns (the second one omits farmers) are very similar to the results discussed above.

4.7. Conclusion

This chapter provides evidence on the positive effect of one's peer group unemployment rate on one's own likelihood of employment using a sample of Civil War veterans. The effect is substantial in size and is statistically significant even after controlling for various factors. Moreover, taking advantage of the time-series feature of the sample, I purge out any individual or group-level unobserved fixed effect and find this result holds. This finding corresponds to the predictions of a simple model presented in Section 3.2. The results are also consistent with the findings presented in Chapter 3 regarding the outcomes of veterans from a different era, World War I.

From a methodological standpoint, this chapter is instrumental in illustrating the advantages to observing a time series panel of individuals, as well as their groups. This allows one to address some of the major concerns regarding identification and estimation of peer effects and social interactions.

The American Civil War presents a unique opportunity to study a large scale assignment to networks. Assignment to companies was based on variables which are less likely to be correlated with the outcomes and covariates used to examine employment more than 10 and 30 years later in 1880 and 1900. Three features of companies formed during war lend themselves to the study of networks and their effect. Each soldier has a very distinct reference group. Moreover, this reference group is potentially significant in its strength. People who have served together in war are likely to form strong bonds. At the same time, these networks are large enough in size to make them meaningful and useful for searching for jobs. Complementing the group assignment feature, the unique data set provides a rare opportunity to observe not only group membership but also post-assignment information on economic outcomes in the labor market, such as employment.

A large part of the male population participated in the Civil War. Hence, I argue that it is not the characteristics of the members of my sample that are unique, but rather the way in which the networks were formed.

In order for the findings to be extended to networks formed in other circumstances, ties formed during military service cannot differ substantially from those formed in other types of settings. This depends on what the underlying mechanism through which networks formed during military service operate. Is it strong bonds formed among a small group of men? The importance of "weak ties" (Granovetter, 1975) might suggest otherwise. Is there an extra emotional value, such as unit pride, which increases the strength of ties beyond what would otherwise be the tie between two people who met? If so, would this type of affiliation be all that different from that experienced by the alumni of a college? Answers to such questions will help in determining the external validity of the results.

Often, research in the area of social networks contributes little to the discussion of policy. Since networks are almost always endogenously formed, it would be hard to imagine an incentive scheme or government intervention that would radically change people's choice of association. Government intervention could help to strengthen and encourage the formation of contacts among those already likely to associate, and by doing so provide them with a network and its associated benefits. This could be the motivation for strengthening associations for minorities, women in business, etc. In Europe, there have been programs to support networking among people of low socio-economic background. By better understanding how networks operate, we can better design such programs.

Social networks formed during war allow for a rare opportunity to examine the results of what would otherwise be a difficult "social experiment" to carry out. One could use the results of this "experiment" to discuss the merits of various programs which seek to integrate people from different socio-economic backgrounds (such as the "bussing" of school children).

References

Akerlof, George A. "Social Distance and Social Decisions," *Econometrica*, *65*:1005–1027 (1997).

Angrist, Joshua D. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, *80*:313–336 (1990).

Anselin, Luc. "Spatial Econometrics." *A Companion to Theoretical Econometrics*, edited by B. Baltagi. 310–330. Oxford: Blackwell, 2001.

Barber, J. Frank, editor. *The History of the Seventy-Ninth Division, A.E.F. During the World War: 1917-1919*, Lancaster, Pa.: Steinman and Steinman, Second Edition 1922.

Bayer, Patrick, et al. "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections," *Working Paper* (2004).

Bayer, Patrick, et al. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *NBER working paper series no.11019* (2005).

Bentolila, Samuel, et al. "Social Contacts and Occupational Choice," *CEMFI Working Paper 0406* (2004).

Bertrand, Marianne, et al. "Network effects and welfare cultures," *The Quarterly Journal of Economics*, *115*(3):1019 (2000).

Bewley, Truman F. *Why Wages Don't Fall During a Recession*. Cambridge: Harvard University Press, 1999.

Boorman, Scott A. "A Combinatiorial Optimization Model for Transmission of Job Information Through Contact Networks," *The Bell Journal of Economics*, *6*(1):216 (1975).

Bramoullé, Yann, et al. "Identification of Peer Effects Through Social Networks," *Mimeo*, *Laval University* (2006).

Brock, William A. and Steven N. Durlauf. "Discrete Choice with Social Interactions," *Review of Economic Studies*, *68*:235–260 (2001a).

Brock, William A. and Steven N. Durlauf. "Interactions-Based Models." *Handbook of Econometrics* edited by J.J. Heckman and E. Leamer, 3298–3380, Elsevier Science, 2001b.

Brock, William A. and Steven N. Durlauf. "Identification of Binary Choice Models with Social Interactions," *Working Paper* (2004).

Calvó-Armengol, Antoni and Matthew O. Jackson. "The Effects of Social Networks on Employment and Inequality," *The American Economic Review*, *94*(3):426 (2004).

Chambers, John Whiteclay. *To Raise an Army: The Draft Comes to Modern America*. New York: The Free Press, 1987.

Cohen-Cole, Ethan. "Multiple Groups Identification in the Linear-in-Means Model," *Economics Letters*, *92* (2006).

Costa, Dora and Matthew Kahn. "Cowards and Heroes: Group Loyalty in the American Civil War," *Quarterly Journal of Economics*, *118*:519–548 (2003).

Costa, Dora L. "Pensions and Retirement: Evidence from Union Army Veterans," *The Quarterly Journal of Economics*, *110*(2):297–319 (1995).

De Schweinitz, Dorothea. *How Workers Find Jobs: A Study of Four Thousand Hosiery Workers in Philadelphia*. Philadelphia: University of Philadelphia Press, 1932.

Dickinson, John. *The Building of an Army*. New York: The Century Co., 1922.

Duncan, Greg J., et al. "Peer Effects in Drug Use and Sex Among College Students," *Journal of Abnormal Child Psychology*, *33*(3):375–385 (2005).

Fairlie, Robert W. and William A. Sundstrom. "The Emergence, Persistence, and Recent Widening of the Racial Unemployment Gap," *Industrial and Labor Relations Review*, *52(2)*:252–270 (1999).

Fogel, Robert W., et al. *Public Use Tape on the Aging of Veterans of the Union Army: Military, Pension, and Medical Records, 1820-1940 [Computer File]*. Center for Population Economics, University of Chicago, 2000.

Forstall, Richard. *Population of States and Counties of the United States: 1790 to 1990*. Washington, D.C.: U.S. Government Printing Office, Bureau of the Census, 1996.

Gibson, Campbell. "Population of the 100 Largest Cities and Other Urban Places in the United States: 1790-1990," *United States Bureau of the Census, Population Division Working Papers No. 27* (1998).

Graham, Brian and Jinyong Hahn. "Identification and Estimation of the Linear-in-Means Model of Social Interactions," *Economics Letters*, *1*:1–6 (2005).

Granovetter, Mark. *Getting A Job: A Study of Contacts and Careers 2nd Edition*. The University of Chicago Press, 1974, 1995.

Greenwald, Bruce C. "Adverse Selection in the Labour Market," *The Review of Economic Studies*, *53*:325–347 (1986).

Heckman, James J. "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, *46*:931–959 (1978).

Ioannides, Yannis M. and Linda Datcher Loury. "Job Information Networks, Neighborhood Effects and Inequality," *Journal of Economic Literature*, *42 (4)*:1056–1093 (2004).

Jackson, Matthew O. "A Survey of Models of Network Formation: Stability and Efficiency." *Group Formation in Economics; Networks, Clubs and Coalitions* edited by Gabrielle Demange and Myrna Wooders, chapter 1, Cambridge U.K.: Cambridge University Press, 2004.

Kanjanapipatkul, Tayatat. "Pensions and Labor Force Participation of Civil War Veterans." *Health and Labor Force Participation over the Life Cycle: Evidence from the Past*, edited by Dora L. Costa. 231–52. University of Chicago Press for NBER, 2003.

Kling, Jeffrey, et al. "Experimental Analysis of Neighborhood Effects," *Econometrica*, *forthcoming* (2005).

Krauth, Brian. "Peer Effects and Selection Effects on Youth Smoking in California," *Journal of Business and Economic Statistics*, *forthcoming* (2006).

Lee, Lung-Fei. "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics*, *forthcoming* (2006).

Manski, Charles F. "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, *60*(3):531–542 (1993).

Margo, Robert. "The Labor Force Participation of Older Americans in 1900: Further Results," *NBER working paper series no. 27* (1991).

Maryland War Records Commission. *Maryland in the World War 1917-1919: Military and Naval Service Records*, *I-II*, Baltimore: Twentieth Century Press, 1933.

McAdam, Doug and Yang Su. "The War at Home: Antiwar Protests and Congressional Voting, 1965 to 1973," *American Sociological Review*, *67*:696–721 (2002).

Meyer, Bruce. "Natural and Quasi- Experiments in Economics," *Journal of Business and Economic Statistics*, *13*:151–162 (1995).

Moen, Jon R. "The Unemployment and Retirement of Older Men," *Historical Methods*, *27*(1):40 (1994).

Moffitt, Robert A. "Policy Interventions, Low-Level Equilibria, and Social Interactions." *Social Dynamics* edited by S. Durlauf and H. P. Young, 45, Washington, D.C.: Brookings Institution Press, 2001.

Montgomery, James D. "Social Networks and Labor-Market Outcomes: Towards an Economic Analysis," *The American Economic Review*, *81, Issue 5*:1408–1418 (1991).

Mortensen, Dale and Tara Vishwanath. "Personal Contacts and Earnings. It Is Who You Know!," *Labour Economics*, *1*:187–201 (1994).

Mortensen, Dale T. and Christopher A. Pissarides. "New Developments in Models of Search in the Labor Market." *Handbook of Labor Economics Vol.* edited by Orley Ashenfelter and David Card, 2567–2627, Amsterdam: Elsevier Science, 1999.

Newey, Whitney K. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Journal of Econometrics*, *36*:231–250 (1987).

Paula, Aureo De. "Inference in a Synchronization Game with Social Interactions," *Working Paper* (2005).

Rivers, Douglas and Quang H. Vuong. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, *39*:347–366 (1988).

Ruggles, Steven, et al. *Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.

Sacerdote, Bruce. "Peer effects with random assignment: Results for Dartmouth roommates," *Quarterly Journal of Economics*, *116*(2):681 (2001).

State of Maryland Department of Labor, Licensing and Regulation, "Origin." http://www.msa.md.gov/msa/mdmanual/20dllr/html/dllrf.html.

Stutzer, Alois and Rafael Lalive. "The Role of Social Work in Job Searching and Subjective Well-Being," *Journal of the European Economic Association*, *forthcoming* (2006).

Thorn, Henry C. *History of the 313th U.S. Infantry*. Wynkoop Hallenbeck Crawford Co., 1920.

Topa, Giorgio. "Social Interactions, Local Spillovers and Unemployment," *The Review of Economic Studies*, *68*:261–295 (2001).

United States Provost Marshal General's Bureau. *Selective Service Regulations Prescribed by the President under the Authority Vested in Him by the Terms of the Selective Service Law*, 1918.

United States War Office. *Second Report of the Provost Marshal General to the Secretary of War on the Operations of the Selective Service System to December 20, 1918*, Washington, D.C.: Govenment Printing Office, 1919.

United States War Office. *Final Report of the Provost Marshal General to the Secretary of War on the Operations of the Selective Service System to July 15, 1919*, Washington D.C.: Government Printing Office, 1920.

Viechnicki, Peter. "Properties and Availability of the Union Army Life-Cycle Sample." *Health and Labor Force Participation over the Life Cycle: Evidence from the Past*, edited by Dora L. Costa. 323–334. The University of Chicago Press, 2003.

The World Almanac and Book of Facts. *Casualties in Principal Wars of the U.S.*, New York, NY: World Almanac Books, 2003.

Zimmerman, David J. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, *85*:9–23 (2003).