IFAU – INSTITUTE FOR
LABOUR MARKET POLICY
EVALUATION

Strategic interactions among Swedish local governments

Karin Edmark

DISSERTATION SERIES 2007:2

Presented at the Department of Economics, Uppsala University

The Institute for Labour Market Policy Evaluation (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market policies, studies of the functioning of the labour market, the labour market effects of educational policies and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Uppsala University, September 24, 2007. The second essay has been published by IFAU as Working paper 2007:22.

ISSN 1651-4149

Doctoral dissertation presented to the Faculty of Social Sciences 2007

Abstract

EDMARK, Karin, 2007, Strategic Interactions among Swedish Local Governments, Department of Economics, Uppsala University, *Economic Studies 105*, 141 pp, ISBN 978-91-85519-12-5.

This thesis consists of four self-contained essays.

Essay 1 (with Matz Dahlberg) investigates if local governments react on the welfare benefit levels in neighbouring jurisdictions when setting their own benefit levels. We solve the simultaneity problem arising from the welfare game by utilizing a policy intervention; more specifically, we use a centrally geared exogenous placement of a highly welfare prone group (refugees) among Swedish municipalities as an instrument. The IV estimates indicate that there exists a "race-to-the-bottom" and that the effect is economically as well as statistically significant; if the neighbouring municipalities decrease their welfare benefit level by 100 SEK, a municipality decreases its benefit level with approximately 41 SEK. This result is robust to several alternative model specifications.

Essay 2 tests for strategic competition in public spending on childcare and primary education, and care for the elderly, using panel data on Swedish municipalities over 1996-2005. The high degree of decentralization in the organization of the public sector implies that Swedish data is highly suitable for this type of study. The study is not limited to interactions in the same type of expenditure, but also allows for effects across expenditures. The results give no robust support for the hypothesis that municipalities react on the spending policy of neighbouring municipalities in the decision on own spending on care of the elderly, childcare and education.

Essay 3 (with Hanna Ågren) uses data on Swedish local governments to test for strategic interaction in local tax setting. We make use of a number of indirect predictions from the theories of tax competition and yardstick competition in order to test for the presence of strategic interaction in these forms. Using such additional predictions of the theories serves a twofold purpose - first it helps us establish if the spatial coefficient is due to strategic interactions or merely reflecting spatial error correlation, and second, it helps identify the source of interaction. The analysis provides strong evidence for spatial correlation in tax rates among Swedish local governments. Moreover, we find weak evidence of tax competition effects in the setting of tax rates, while no evidence is found for yardstick competition.

Essay 4 tests for a migration response to the implementation of stricter rules for welfare benefit receipt, in the form of mandatory participation in activation programs for recipients of welfare, in Stockholm town districts. The hypothesis is that welfare benefit prone individuals will choose to live in a town district that has no program if they dislike the loss of leisure due to program participation more than they value the contents of the program, and vice versa. The results give some indications of a negative effect of the program on the outmigration of welfare prone individuals. This is however not robust to changes neither in the comparison group nor in the sample of town districts. The conclusion that can be drawn is that there are no indications that the activation programs lead to outmigration of welfare prone individuals.

Contents

Acknowledgements

Being a PhD student has its ups and downs. On the positive side, you are given an immense amount of freedom. Freedom to explore new ideas, to always learn new things, and not least, to choose your own work hours. Life as a PhD is varied, exciting, flexible and fun. On the other hand, there are moments when that freedom is less enjoyable, such as when you discover that the great idea you had was not so great after all. Or when you find that the freedom to always learn more also means that you never know enough, or that being able to choose when to work also includes the option to spend all your time in the office.

In this environment, having a supervisor as positive and inspiring as Matz Dahlberg has been worth a great deal! His encouragement and never-ending optimism made writing this thesis much more fun, and his comments and advice surely improved the quality of it a lot. I have also benefited greatly from the advice given by my co-supervisor, Sören Blomquist, and from the comments from Jørn Rattsø, Tuomas Pekkarinen and Eva Mörk. Several other persons have contributed with valuable advice - editors and referees, colleagues. Thank you!

The efficient and friendly administration staff deserves a special acknowledgement: Thank you Monica, Katarina, Eva, Ann-Sofie, Berit and Kristina for excellent administrative support, and thanks to Åke for always patiently answering my stupid computer-questions.

This PhD was enabled through funding from The Swedish Research Council, and the Hedelius foundation. Thanks to the latter, I had the opportunity to spend a semester at the STICERD at the London School of Economics. This was a rewarding experience and I am grateful for the hospitality that was shown to me there.

Without all wonderful colleagues at the department, these five years would have been a lot less memorable! I especially want to thank my office-mate Erica, who is not only a kind and clever friend, but who is also a great discussion partner for any topic - from politics to clothes to family life... The fact that we have had many similar experiences outside work during this period has made your company even more enjoyable. Thanks also to Hanna, for your sense of humor and for teaching me never to stop *thinking*, and to Cilla, for always cheering everybody up.

I also want to thank my friends outside the department - nobody mentioned, nobody forgotten - for being such great friends!

I am grateful for the unconditional love of my parents, Berit and Olle, and for having two sisters, Sara and Anna, with family, who are also very good friends. Finally, I thank my very favourite husband, Peter, and our beloved little red-headed son, Alexander, for being constant reminders of what really matters.

Karin Edmark, Uppsala, August 2007

Introduction

Ever since the seminal work of Tiebout[1], the optimal degree of decentralization has been a topic of ongoing work and discussion in the economic literature. On the one hand, decentralized decision making can be a way to better match the provision of public services with the preferences of local citizens, and the possibility of making comparisons with other local governments may lead to a more efficient public sector. On the other hand, there are also potential negative effects to consider, such as a race to the bottom in local tax rates or other forms of harmful competition.

In the recent years, there has been a trend to increase the degree of decentralization within the public sector. In Sweden, for example, a set of reforms in the early 90s increased the self-determination of the local governments, by increasing the municipalities' responsibility in areas such as primary schooling, childcare and care for the elderly. This has resulted in a situation where the local levels of government are responsible for important welfare services. Specifically, the 290 municipalities are responsible for schooling, childcare, care for the elderly, social assistance, infrastructure and environmental regulation. The main responsibility of the 21 counties is the provision of health care. The municipalities and counties have the constitutional right to self government, and have the right to set the local income tax rate, as well as a substantial degree of freedom to decide on the provision of local public services.[2]

In this perspective, it is important to critically examine the pros and cons of decentralization. One important issue is that the effects of a local policy decision are often not confined to the own jurisdiction, but that they also have consequences for surrounding jurisdictions. For example, local policy decisions can give rise to migration of residents between jurisdictions.

This thesis is an attempt to provide some answers to the question of whether such spill-over effects are important in practise. The thesis contains four self-contained essays that test for the presence of different types of spill-over effects of local decision making in Swedish local governments. The high degree of self-determination, as well as the vast range of services and goods provided by the local levels, make Sweden a particularly interesting case to study.

Essays 1-3 test for the presence of strategic behaviour in the decision making of local politicians. As we shall see in the following sections, such interactions may take different forms and may have different policy implications for the optimal degree of decentralization. Essay 4 tests for a potential source of strategic interaction, namely for a migration response to differentials in local welfare benefit policy.

The following sections first give a brief introduction to the main topic of the

[1] See Tiebout (1956).

[2] Contrary to many other countries, only income is taxed locally. Property taxes, for example, are set at the national level.

thesis; strategic interactions, followed by a summary of the results of the essays and the previous literature on this topic. Finally, the topic of Essay 4, welfare migration, is described, and the results are discussed in relation to the existing literature.

Strategic interactions among local governments

As mentioned in the previous section, three of the essays of the thesis test for strategic interactions in local decision making. The literature distinguishes between two main theoretical frameworks for these types of interactions: resource-flow models and yardstick competition models. (For more detailed overviews of the theoretical models, see e.g. Brueckner (2003) and Revelli (2006)). The models have in common that they describe a situation where the local policy maker takes the other local governments' policy decisions into account when deciding on the own local policy. However, they differ in the assumptions they make regarding the reasons for this behaviour. The following sections provide short descriptions of the two types of models.

Theoretical framework

Resource-flow models

The perhaps most well-known example of a resource-flow model is the case of tax competition. The term describes a situation where local governments strive to set tax rates lower than the surrounding jurisdictions in order to attract a mobile tax base, such as firms or individuals. This competition leads a local government to take the surrounding jurisdictions' tax rate into account in its own tax decision.

Is tax competition good or a bad from society's point of view? This depends on one's view of the workings of the public sector. In general, tax competition can be shown to put a downward pressure on tax rates. This will, ceteris paribus, drive tax rates down to below the social optimum.[3] However, it has also been argued that tax competition can work as disciplining wasteful or rent-seeking politicians. In this case, tax competition can push taxes closer to the social optimum by preventing politicians from setting tax rates too high.[4]

The resource-flow model is also relevant for other types of local policy. In particular, it has been applied to local welfare benefit policy. In this case, the mobile resource is not the tax base, but the welfare dependent population, and the idea is that local policy makers compete *not* to be the more generous jurisdiction, in order *not* to attract costly welfare dependent residents to the jurisdiction. As in the tax competition case, this will put a downward pressure on the welfare benefit generosity

[3] See e.g. Oates (2002) and Wilson (1999).

[4] See e.g. Edwards and Keen (1996) for a discussion on this.

of jurisdictions, and, ceteris paribus, result in welfare benefit levels that are lower than the social optimum.

Yardstick competition models

Whereas in the resource competition model, residents exert influence over the local policy maker by moving - or voting with their feet - in the yardstick competition model interaction stems from the political process. This model assumes that there is an information asymmetry between voters and politicians regarding the cost for public service provision. More specifically, it assumes that the local politician, but not the voters, knows the true cost of providing public services. This means that the voters do not know if for example an increase in tax rates is motivated by higher costs for service provision, or if it is due to rent-seeking behaviour on the part of the local policy maker.

In this case, decentralization can be a means to decrease the information asymmetry. The idea is that voters can infer the likelihood that the tax raise is motivated by a (regional) cost shock by comparing if taxes are also raised in the surrounding jurisdictions. That is, assuming that the surrounding jurisdictions have a similar cost structure, voters will assume that higher tax rates, for a given level of public service, is due to wasteful behaviour/rent-seeking activities by the local politician. Since a bad comparative performance is likely to decrease the probability for re-election, a local politician will take this into account and make sure not to deviate too much from the neighbours' policy decision.

In the presence of this type of information spill-over, decentralization hence has positive effects on the workings of the public sector, by helping voters to discover rent-seeking behaviour, and thereby disciplining rent-seeking politicians. This contrasts to the former tax competition model, where the welfare effect can be either negative or positive.

The pattern of interaction between neighbouring local governments will however look similar to the tax competition case, i.e. the model predicts a positive policy interaction between neighbouring jurisdictions.[5]

Empirical evidence

We have now seen that the two types of strategic interactions both give rise to the same type of interdependence between adjacent local governments: the own policy decision is affected by the neighbours' policy decision. That is, the local policy of jurisdiction i, y_i, is a function of the neighbouring jurisdictions' policies, y_{-i}, in addition to a set of jurisdiction-specific policy-relevant covariates, X_i, as illustrated in equation (1):

[5] Bordignon, Cerniglia and Revelli (2004) however show that under certain circumstances, the sign of the interaction effect can also be negative.

$$y_i = f(y_{-i}, X_i), \tag{1}$$

Furthermore, a positive interaction effect is predicted in both types of models, i.e. the neighbours' policy has a positive effect on the own policy decision. For example, lower tax rates among the neighbours will be followed by lower taxes in the home jurisdiction in the tax competition case, or a decrease in the neighbours' welfare benefit levels will limit the generosity of the own welfare benefit policy in the case of welfare competition.

In general, the hypothesis of strategic interactions is tested by running a regression of the following type:

$$y_i = \alpha + \rho y_{-i} + \beta X_i + \varepsilon_i \tag{2}$$

In equation (2) ρ measures the interaction effect, i.e. the degree with which the neighbouring local governments' policy decision affects the own policy decision. α is a constant regression term and ε_i is the regression error term.

An important issue is how to define a jurisdictions' neighbours - i.e. which jurisdictions are likely to interact with each other as described in equation (2)? In general, the definition of neighbours is based on some measure of geographical proximity, such as sharing border. There are however also other possible measures, based on factors such as migration, distance or population.[6] While the border-based definition has the advantage that it is simple and straight-forward, and it is not in itself affected by the local policy, which is an important criterion for the validity of the specification, using more elaborate definitions can yield additional insights. Essays 2 and 3 in this thesis therefore use several different neighbourhood definitions when estimating the interaction effect.

The fact that policy interaction, as described in the above equations, is a simultaneous phenomenon, means that y_{-i} is endogenous in equation (2), and OLS hence yields biased estimates of ρ. In the literature this is in general handled either by using instrumental variables for y_{-i}, or by using a maximum likelihood estimator.[7] The general approach in studies using the IV-approach is to use a set of neighbours' socioeconomic and demographic characteristics as instruments for y_{-i}.

Both IV and ML have been used in the literature. As has been shown by Kelejian and Prucha (1998), an advantage with instrumental variables estimation is that the estimates are consistent also in the presence of correlation in the error term between neighbours (spatial error correlation). Since this type of correlation is likely to be present, the essays in this thesis use instrumental variable estimation.

[6] See e.g. Revelli (2006) for a discussion of the definition of neighbours.

[7] These methods are described in detail in e.g. Anselin (1988), Revelli (2006), Kelejian and Prucha (1998), and Kelejian and Prucha (1999).

A central issue is how to differentiate between different types of strategic policy interaction, i.e. how to separate interaction driven by mobile resources from interaction stemming from yardstick competition. The fact that the policy implications of the interaction models are likely to differ, means that distinguishing between the two is important. As we shall see below, different approaches are taken in the essays of the thesis in order to identify the source of interaction.

Another important issue to consider is how well we can identify strategic interaction (of either form) from other factors that give rise to a similar pattern in the data. Several previous studies find evidence of a positive correlation in local policy between neighbouring jurisdictions. However, can we be sure that this correlation reflects strategic interaction? Or is it merely the case that adjacent local governments respond to common regional shocks with the same policy changes?

In the following sections I will discuss how these issues are handled in the essays of this thesis, as well as in the literature in general.

Welfare competition

Essay 1 tests for welfare competition among Swedish municipalities. The idea is that the municipalities will avoid being more generous than the neighbours, in fear that this will lead to an inflow of welfare prone individuals to the municipality and hence increase the costs for welfare provision. As previously mentioned, the result is a "race to the bottom" in the welfare benefit levels, and suboptimally low welfare benefit levels in all municipalities.

Table 1 shows the main result of essay 1, together with the results of similar studies. The coefficients shown in the table correspond to estimates of ρ in equation (2), i.e. they measure how a local government reacts to a one unit increase in the neighbours' welfare benefit level. ML/IV indicates if maximum likelihood or instrumental variables estimation was used to obtain the effects.

As can be seen in Table 1 positive interaction effects are estimated in all studies. However, the sizes of the effects vary, from 0.36 to 1.35. The table suggests that in particular instrumental variable estimation tends to yield quite high coefficient estimates, sometimes even unreasonably large estimates[8].

What is the reason for these high estimates? A problem with the general IV-approach - i.e. using socioeconomic and demographic characteristics as instruments for the neighbours' policy, is that these may not be exogenous with respect to neighbours policy. If it is the case that individuals sort themselves among local governments based on the local governments' welfare generosity, then the allocation of socioeco-

[8] As can be seen in Table 1, Saavedra (2000) estimates interaction coefficients above 1 when using instrumental variables estimation, which is not compatible with a stable interaction process. Figlio, Kolpin and Reid (1999) also obtain estimates over 1 in some specifications.

nomic and demographic variables are functions of the benefit level, and hence not valid as instruments.

Instead of using this approach, Essay 1 proposes and uses a policy intervention to solve the simultaneity problem arising from the welfare game. More specifically, we utilize an exogenous variation that was provided by a policy intervention in Sweden in the late 1980s and early 1990s as an instrument; a centrally geared exogenous placement of a highly welfare prone group (refugees) among Swedish municipalities. The fact that the refugee placement was not related to the welfare benefit policy of the municipalities, means that the placement of refugees solves the selection problem, and can validly be used as an instrument for the local benefit generosity. This instrument furthermore has the nice feature that it can be motivated by the theoretical model of welfare competition.

The result of Essay 1 is consistent with the hypothesis of a race to the bottom in welfare benefit levels. However, as can be seen in Table 1, the estimated interaction effect is substantially lower than what is typically found in earlier studies using instrumental variables estimation. This discrepancy can possibly be due to the endogeneity of the instruments used in the earlier studies.

Table 1: Estimates of spatial interaction in welfare benefit levels

Study^{+}	**IV**	**ML**
Figlio et al (199)	0.904*	
Fiva & Rattsø (2003)	0.81*	0.36*
Essay 1	*0.41**	
Saavedra (2000)	1.35*	0.42*

$^{+}$References for the studies are found in the Bibliography. * estimate is significantly different from zero at the 10% level of significance.

Strategic interactions in local public spending

As was described in the introduction, studies on strategic interactions among local governments in general concern local tax rates or welfare benefit policy. However, as suggested by Wilson and Gordon (2003), this type of behaviour may arise also in other policy areas. Essay 2 in the thesis tests for strategic interactions in the main expenditures of the Swedish municipalities, namely how much to spend on childcare, primary schooling and care for the elderly.

Strategic interaction in these services can arise for two reasons: the local governments may try to attract either residents or voters by providing good quality services, i.e. migration-based competition or yardstick competition. The fact that childcare, schooling and care for the elderly are services that are visible and important to residents/voters, strengthen the hypothesis that there may be scope for the local policy

maker to adjust spending on these services in order to affect the migration flows and/or the voting behaviour of residents.

The fact that the services in this study - childcare, primary schooling and care for the elderly - are targeted to different residents, families with children and elderly, respectively, suggests that the policy maker could even try to affect the demographic composition of the municipality through the spending policy of the municipality. In Essay 2, we argue that this may lead to childcare and schooling being favored on the expense of spending on care for the elderly. The idea is that having a large share of families is likely to be economically more favorable to a municipality than having a large share of elderly.

Previous studies on spatial interactions in local expenditures in general depart from either the welfare competition theory or are based on the theory of benefit spillovers.[9] These studies predominantly test for interactions in spending on goods such as infrastructure and cultural and recreational services, or aggregate expenditures (See e.g. Case, Hines and Rosen (1993), Baicker (2005), Redoano (2003), Schaltegger and Zemp (2003), and Solé-Ollé (2006))

Essay 2 adds to the literature by testing for strategic interactions in the composite expenditure policy of local governments, i.e. it focuses on the main expenditure items of the municipalities and it allows for interaction to take place both in expenditures on the same service category, and in expenditures on different categories of services. This makes sense if residents/voters care about the allocation of resources between different services, as well as how much is spent on each category.[10]

The results of Essay 3, however, do not confirm the hypothesis of strategic competition in local spending on childcare, primary education and care for the elderly. While there are some significant coefficients, especially in the regression on spending on care for the elderly, the results are not robust enough to be interpreted as evidence for strategic competition.

Strategic tax interaction

Essay 3 tests for strategic interactions in municipal tax rates. As in Essay 2, there are two potential sources for strategic interaction to occur: competition for mobile residents, or yardstick competition for votes.

The aggregate evidence from earlier empirical work indicates that spatial interac-

[9] The spill-over theory describes a situation where the benefits of local public goods and services "spill over" to surrounding jurisdictions, such as infrastructure or services that are available also to non-residents. The spillover theory is however not directly applicable to this study, since only residents of the jurisdiction are entitled to the services of the study.

[10] Two previous studies estimate strategic interactions in composite local policies: the first, Fredriksson, List and Millimet (2004), focuses on U.S. state policies to attract firms to the locality, and the second, Millimet and Rangaprasad (2007), looks at U.S. school district inputs.

tion processes are present in local tax rates, but there is no consensus on the source of interaction (tax or yardstick competition).[11] Identification of the source of interaction is however important, since the two theoretical models have different policy implications. While the theory of tax competition predicts that the tax rate in a decentralized setting will be lower than the social optimum, the yardstick competition theory in general predicts increased efficiency of the political system through better informed voters. All else equal, this implies that if yardstick competition is present, having a lot of small local governments improves the workings of the voting process, whereas if there is tax competition, central level decision making would yield a tax rate that is closer to the social optimum.

In contrast to many studies in this field, in Essay 3, we make no a priori assumptions regarding the underlying theoretical framework. Instead, we make use of additional, indirect predictions from the theories of tax competition and yardstick competition to test for the presence of strategic interaction in these forms. Specifically, we use a reform of the central government grants system, which changed the system of tax base equalization of the municipalities, to test for migration-based tax competition. The idea is that if we find the degree of interaction to be different after the reform, this can be seen as indirect evidence of tax competition. We also use two empirical implications descending from yardstick competition; namely that yardstick-type interaction is expected to be more prevalent during election years and when the political majority is weak, to test for strategic interaction in the form of yardstick competition.

The results in Essay 3 give strong evidence for spatial correlation in tax rates among Swedish local governments: a tax cut of on average 1 percentage point in neighbouring jurisdictions is correlated with a decrease of about 0.74 percentage points in own taxes. The additional tests that are used furthermore suggest that at least part of this correlation stems from a desire to attract tax base to the municipality, while no support is given for the yardstick competition hypothesis.

How does this relate to the previous literature? As previously mentioned, most previous studies estimate a positive correlation in the tax rates of neighbouring governments. However, only three other studies use some additional test to establish the source of the interaction[12], and no study tests for both migration-based competition

[11] See e.g. Heyndels and Vuchelen (1998), Besley and Case (1995), Buettner (2001), Bordignon, Cerniglia and Revelli (2003), Esteller-Moré and Solé-Ollé (2002), Solé-Ollé (2006),

Revelli (2001), Brueckner and Saavedra (2001), and Allers and Elhorst (2005), which also gives a nice overview of the literature.

[12] Besley and Case (1995) and Solé-Ollé (2006) use tests based on predicions from political features, and find evidence of yardstick competition. Esteller-Moré and Solé-Ollé (2002) uses a feature of the Canadian equalization grants system and finds evidence of tax competition.

and yardstick competition.

Welfare migration

As was previously described, Essay 1 tests for welfare competition among local governments, and finds evidence of positive interaction corresponding to such a process. As has also been described in the previous sections, the cause for welfare competition to arise is the fear of the local policy makers that having a more generous welfare benefit policy than the neighbours will attract costly welfare prone individuals to the jurisdiction.

Is the fear of welfare migration motivated? I.e. do welfare prone individuals tend to move to more generous jurisdictions? The aggregate evidence of the previous, mainly American, literature is mixed: some studies find large effects while other find no effects[13]. The results of the more recent studies however suggest that welfare generosity does affect migration, but that the effect is rather small (see e.g. McKinnish (2005) and McKinnish (2007), Gelbach (2004), and Meyer (2000)).The exception is Fiva (2007), who finds large migration effects when studying Norwegian municipalities.

However, one problem in the literature is that it is difficult to separate the effects of welfare generosity from other jurisdiction-specific factors that may affect migration. In contrast to other studies of welfare migration, Kaestner, Kaushal and Ryzin (2001), who test if the introduction of time limits, financial sanctions for non-compliance, and strict work eligibility rules in US states affected outmigration, also study the situation *after* migration, as a further test of the cause for moving. Interestingly, they find that many of those that moved from the more strict states, were employed after the move. This result may suggest an increase in the labour market mobility in the states that have implemented the stricter rules. However, an alternative explanation could be that the moves were not at all related to differences in welfare benefit policy, but rather to different employment possibilities. For example, it can be the case that jurisdictions that experience a declining economic situation are more willing to try new and stricter welfare benefit rules. This highlights the difficulties of controlling for the characteristics of all possible moving-combinations in studies of welfare migration.

In the last essay of the thesis, Essay 4, I make use of a reform of the Swedish Social Service Act in order to solve the methodological problem described above. The reform enabled the municipalities and town districts to condition benefit receipt on participation in activation programs for recipients of welfare. Essay 4 tests for a migration response to the implementation of these stricter rules in town districts in the municipality of Stockholm.

[13] See Meyer (2000) for a review of the earlier literature.

An important advantage with this data set is that all individuals live in the same municipality, i.e. the same local labour market area, which means that there is no need to control for varying labour market characteristics of the local jurisdictions. By limiting the analysis to Stockholm town districts, we hence minimize the risk of omitted variable bias due to differences in local characteristics. In addition, the fact that merely a short-distance move is needed in order to end up under a different benefit policy, makes the migration hypothesis a more plausible story.

A second advantage with this set of data is that the starting year of the activation programs differs among the town districts in our sample, which means that we can use two sources of variation to identify the effect of the program on the moving choices of welfare prone individuals. First, we can compare the moving choices of welfare prone individuals before and after the law revision, in town districts that did and did not start an activation program after the revision, i.e. a district-level difference-in-differences analysis. Second, we can add a further component to the analysis, and compare the migration effects on groups that differ in the propensity to receive welfare benefits. The idea is that the moving behaviour of individuals with a high propensity to use welfare will be affected by the programs, while individuals that are not welfare prone will not be affected. Combining this approach with the district level analysis yields a difference-in-difference-in-differences estimator. This approach gives good possibilities to control for the effects of unobserved trends that affect migration. This is an advantage, compared to most other studies of welfare migration, which rely on comparison group based difference-in-differences analysis.

The results give some indications of a negative effect of the program on the outmigration of welfare prone individuals. This is however not robust to changes in comparison group nor in the sample of town districts. The conclusion that can be drawn is that there are no indications that the activation programs lead to outmigration of welfare prone individuals. This result differs from the previous literature on welfare migration. It is possible that this difference is due to that the effects on migration actually differs between this and the previous studies. It can however also be the case that the previous evidence on welfare migration suffers from an omitted variable bias, which is not present in this study.

References

Allers, M. and Elhorst, J.: 2005, Tax mimicking and yardstick competition among local governments in the netherlands, *International Tax and Public Finance* **12**, 493–513.

Anselin, L.: 1988, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishing, Dordrecht.

Baicker, K.: 2005, The spillover effects of state spending, *Journal of Public Economics* **89**, 529–544.

Besley, T. and Case, A.: 1995, Incumbent behavior: Vote seeking, tax setting and yardstick competition, *American Economic Review* pp. 25–45.

Bordignon, M., Cerniglia, F. and Revelli, F.: 2003, In seach of yardstick competition: A spatial analysis of italian municipality property tax, *Journal of Urban Economics* **54**, 199–217.

Bordignon, M., Cerniglia, F. and Revelli, F.: 2004, Yardstick competition in intergovernmental relationships: Theory and empirical predictions, *Economics Letters* **83**, 325–333.

Brueckner, J.: 2003, Strategic interaction among governments: An overview of empirical studies, *International Regional Science Review* **26**, 175–188.

Brueckner, J. K. and Saavedra, L.: 2001, Do local governments engage in strategic property-tax competition?, *National Tax Journal* **54**, 203–229.

Buettner, T.: 2001, Local business taxation and competition for capital, *Regional Science and Urban Economics* **31**, 215–245.

Case, A., Hines, J. and Rosen, H.: 1993, Budget spillovers and fiscal policy interdependence, *Journal of Public Economics* **52**, 285–307.

Edwards, J. and Keen, M.: 1996, Tax competition and leviathan, *European Economic Review* **40**(1), 113–134.

Esteller-Moré, A. and Solé-Ollé, A.: 2002, Tax setting in a federal system: The case of personal income taxation in canada, *International Tax and Public Finance* **9**, 235–257.

Figlio, D., Kolpin, V. and Reid, W.: 1999, Do states play welfare games?, *Journal of Urban Economics* pp. 437–454.

Fiva, J. H.: 2007, Does welfare policy affect residential choices? evidence from a natural experiment. Discussion Papers No. 503, Statistics Norway Research Department.

Fredriksson, P., List, J. and Millimet, D.: 2004, Chasing the smokestack: Strategic policymaking with multiple instruments, *Regional Science and Urban Economics* **34**(4), 387–410.

Gelbach, J. B.: 2004, Migration, the lifecycle, and state benefits: How low is the bottom?, *Journal of Political Economy* **112**(5), 1091–1130.

Hernes Fiva, J. and Rattsø, J.: 2003, Welfare competition in norway. Department of Economics, Norwegian University of Science and Technology, Trondheim.

Heyndels, B. and Vuchelen, J.: 1998, Tax mimicking among belgian municipalities, *National Tax Journal* **60**, 89–101.

Kaestner, R., Kaushal, N. and Ryzin, G. V.: 2001, Migration consequences of welfare reform. NBER Working paper 8560.

Kelejian, H. H. and Prucha, I. R.: 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* **40**(2), 509–533.

Kelejian, H. and Prucha, I.: 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* **17**, 99–121.

McKinnish, T.: 2005, Importing the poor: Welfare magnetism and cross-border welfare migration, *Journal of Human Resources* **40**, 57–76.

McKinnish, T.: 2007, Cross-border welfare migration: New evidence from micro-data, *Journal of Public Economics* **91**, 437–450.

Meyer, B.: 2000, Do the poor move to receive higher welfare benefits? Unpublished Paper, Department of Economics, Northwestern University.

Millimet, D. and Rangaprasad, V.: 2007, Strategic competition amongst public schools, *Regional Science and Urban Economics* **37**(2), 199–219.

Oates, W.: 2002, Fiscal and regulatory competition: Theory and evidence, *Perspektiven der Wirtschaftspolitik* **3**, 377–390.

Redoano, M.: 2003, Fiscal interactions among european countries. Warwick Economic Research Papers No 680, Department of Economics, Warwick University.

Revelli, F.: 2001, Spatial patterns in local taxation: Tax mimicking or error mimicking?, *Applied Economics* **33**, 1101–1107.

Revelli, F.: 2006, Spatial interactions among governments, *in* E. Ahmad and G. Brosio (eds), *Handbook on Fiscal Federalism*.

Saavedra, L.: 2000, A model of welfare competition with evidence from AFCD, *Journal of Urban Economics* pp. 248–279.

Schaltegger, C. and Zemp, S.: 2003, Spatial spillovers in metropolitan areas: Evidence from swiss communes. Working Paper No 2003: 06, CREMA, Brussels.

Solé-Ollé, A.: 2006, Expenditure spillovers and fiscal interactions: Empirical evidence from local governments in spain, *Journal of Urban Economics* **59**, 32–53.

Tiebout, C. M.: 1956, A pure theory of local expenditures, *Journal of Political Economy* pp. 416–424.

Wilson, J.: 1999, Theories of tax competition, *National Tax Journal* **52**, 269–304.

Wilson, J. D. and Gordon, R. H.: 2003, Expenditure competition, *Journal of Public Economic Theory* **5**(2), 399–417.

Essay 1

Is There a "Race-to-the-Bottom" in the Setting of Welfare Benefit Levels? Evidence from a Policy Intervention[1]

1 Introduction

In the wake of the major welfare reforms that took place in the United States in the 1990s, there has been a growing interest in empirical work on strategic interactions among local governments in the setting of welfare benefit levels. The increased interest stems from a fear of a "race-to-the-bottom" (RTB) in welfare benefit levels.[2]

The most direct test of the existence of a RTB[3] in the setting of welfare benefit levels is to test for strategic interactions among local governments.[4] What one typically estimates is an equation of the form

$$B_i = \gamma \sum_{j \neq i} \omega_{ij} B_j + X_i \beta + \varepsilon_i \tag{1}$$

where B_i is the benefit level in local government i, B_j is the benefit levels in other local governments j, $j \neq i$, ω_{ij} are weights that indicate the importance attached by local government i to benefits in the other local governments, X_i is a matrix of socioeconomic and demographic characteristics for local government i with the associated parameter vector β, and ε_i is the error term. The parameter of interest, γ, represents the slope of the local government's reaction function. If γ is significantly

[1] Co-authored with Matz Dahlberg.

[2] Apart from the main reform in the U.S. in 1996, several state waivers were enacted in the first part of the 1990s. As a consequence of the reforms, there was a highly increased decentralization of responsibilities for the welfare system to the state level, implying an increased probability for strategic interaction among the states to take place.

[3] Even though it is likely to overstate the issue, we will follow the earlier literature and use the phrase "race-to-the-bottom" as a convenient shorthand description of the phenomenon of interest (as noted by Brueckner (2000), while the theory only points to a downward bias in welfare benefits caused by a concern about welfare migration, popular usage of the phrase sometimes have the meaning of a much more dire outcome).

[4] A RTB in the setting of welfare benefit levels can materialize if there are strategic interactions among local jurisdictions; if local government decision makers perceive, correctly or not, that generous benefits attract welfare migrants, this may make jurisdictions reluctant to offer generous benefits because it may increase the number of program participants and thus the total cost of providing a given level of benefits. As a result, welfare benefits may be lower than the socially desirable level.

different from zero, then strategic interaction occurs between a given local government and other local governments that have not been assigned a weight of zero.[5] The econometric problem in estimating equation (1) is that the benefit levels on the right-hand-side are endogenous variables since the benefit levels in all localities are jointly determined when strategic interactions occur.

Earlier empirical work in this area have found a positive and statistically significant estimate of γ (see, e.g., Figlio, Kolpin and Reid (1999), Saavedra (2000), and Hernes Fiva and Rattsø (2003)). The typical solution to the simultaneity problem has been to use socioeconomic and demographic characteristics of the other local governments as instruments for B_j. Using panel data on U.S. states, Figlio et al. (1999) do, for example, use the neighbor states' female unemployment rate, the neighbor states' ratio of females to employed males, and the neighbor states' average weekly wages in variety stores as instruments.[6] The problem with this approach is however that instruments based on socioeconomic and demographic characteristics are typically not randomized over local governments. Basically, it is a selection problem. If it is the case, as suggested by the welfare migration literature, that individuals sort themselves among local governments based on the local governments' welfare generosity, then the allocation of socioeconomic and demographic variables are functions of B_j. The implication is that variables like the female unemployment rate, the ratio of females to employed males, and the neighbor states' average weekly wages in variety stores are not valid as instruments.

The aim of this paper is to estimate equation (1). Our main contribution is that we propose and use a policy intervention to solve the simultaneity problem arising from the welfare game. More specifically, we utilize an exogenous variation that was provided by a policy intervention in Sweden in the late 1980s and early 1990s as an instrument; a centrally geared placement of a highly welfare prone group (refugees) among Swedish municipalities. The exogenous placement of refugees solves the selection problem.

Our IV estimates indicate that there exists a "race-to-the-bottom" and that the effect is economically as well as statistically significant; if the neighboring municipalities decrease their welfare benefit level with 100 SEK, a municipality decreases its benefit level by approximately 41 SEK. This result is robust to several alternative model specifications.

Our estimated interaction effect is substantially lower than what is typically found

[5] Another, more indirect, way of investigating if there is a RTB is to test for welfare migration. Recent work in this area include Gelbach (2004) and McKinnish (2005) and McKinnish (2007)). Excellent surveys of the earlier research on welfare migration can be found in Brueckner (2000) and Meyer (2000).

[6] The same type of instruments are also used by Hernes Fiva and Rattsø (2003) when testing for strategic interactions among Norwegian local governments.

earlier in the literature. The results in Figlio et al. (1999), for example, indicate that a state is expected to change its benefit levels by 90 cents when neighboring states change their benefit levels by one dollar. This discrepancy can be due to the endogeneity of the instruments used in the earlier studies.[7]

The remainder of the paper is organized as follows. The next section analyzes the policy intervention within a theoretical framework. Section 3 presents the policy intervention, section 4 describes the data and the empirical specification, and section 5 provides the regression results. A detailed sensitivity analysis is given in section 6 and, finally, section 7 concludes.

2 Theoretical framework

To organize our thoughts on the problem, we will use the theoretical setup presented in Brueckner (2000), building on work by Brown and Oates (1987) and Wildasin (1991). In the first three parts of this section, we present the model and use it to discuss what type of policy intervention one could use in order to solve the simultaneity problem arising from the welfare game. Then, in the last part of the section, we analyze how the proposed policy intervention affects the welfare game and examine under what circumstances it can be considered as a suitable instrument for solving the simultaneity problem.

2.1 General setup

The model economy contains two regions: A and B. In each region there are M non-poor consumers, referred to as "rich", who are immobile across states. The economy contains $2\overline{N}$ poor consumers, who work at low-paying jobs as well as receive welfare benefits from the region where they reside. The poor are assumed to be mobile across regions, with zero migration costs. There are N_A poor people in region A and $N_B = 2\overline{N} - N_A$ in region B.

The wages of the poor are determined in a competitive labor market, and thus reflect the marginal productivity of unskilled labor. Suppose that the output of region i depends on the amount N_i of unskilled labor along with other fixed factors (such as land and capital), $f(N_i)$. The wage of a region is hence equal to $w_i = f'(N_i)$ We assume that f is strictly concave, which implies that the wage falls as the unskilled labor pool grows; $w'(N_i) \equiv f''(N_i) < 0$. Wages in the two regions are then given by $w_A = w(N_A)$ and $w_B = w(N_B)$. Letting B_A and B_B denote the welfare benefits paid

[7] When we use the same approach, that is when using the neighboring municipalities covariates as instuments, we get an IV-estimate of 0.77. However, the instruments are weak in such a specification; the first-stage partial F-statistic is only equal to 2.77. This is a further indication that neighbors' covariates are not appropriate as instruments.

to the poor, the total income of a poor resident equals $w(N_A) + B_A$ in region A and $w(N_B) + B_B$ in region B.

Each region's welfare benefit level is chosen by its rich residents, who care about the well-being of the local poor (through interdependent preferences). We assume that the rich in both regions have the same utility function: $U(x_i, w_i + B_i)$, $i = A, B$, where x_i gives consumption expenditure for the rich in region i. For simplicity, the utility function is assumed to be quasi-linear, that is

$$U(x_i, w_i + B_i) = x_i + V(w_i + B_i), \qquad i = A, B \tag{2}$$

where V is increasing and strictly concave in $w_i + B_i$.

Letting y denote the income of the rich, which is assumed to be the same in both regions, the budget constraint of a rich resident is given by

$$x_i = y - \frac{N_i B_i}{M}, \qquad i = A, B \tag{3}$$

The benefit level of the region is thus chosen to maximize equation (2) with respect to the benefit level, such that the budget restriction in equation (3) holds.

We will start by briefly looking at the no mobility case, before turning to the more interesting case in which the poor are allowed to move between the regions.

2.2 The no-mobility case

In the no-mobility case, solving the maximization problem for the optimal benefit level of region A yields the following first order condition:

$$MV'(w_A + B_A) = N_A \tag{4}$$

The condition states that the rich of the region set the benefit levels so that the sum of their marginal utilities of the poor's income, is equal to the marginal cost of increasing the poor's incomes through increasing benefits. The first order condition is hence a Samuelsson condition for the provision of a public good.

2.3 The mobility case

If we allow the poor to move between the regions, the analysis becomes slightly more complicated. Solving the maximization problem in the presence of welfare migration implies that the rich of region A choose the welfare benefit level taking account of the fact that an increase in B_A raises N_A through welfare migration. The regions thus play a Nash welfare game, with the rich in region A viewing region B's welfare benefit level, B_B, as fixed in making their own choice.

In order to derive an internal migration equilibrium, i.e. in order to avoid a situation where all poor individuals move to the region with the marginally higher

benefit level, we need to put some constraint on migration. In this model setup, the assumption that wages depend negatively on the number of poor residents provides such a constraint.[8] Migration equilibrium is achieved when the total income of the poor is equalized between the two regions, i.e. when the following expression holds:

$$w(N_A) + B_A = w(N_B) + B_B \tag{5}$$

By maximizing equation (2) with respect to B_A subject to equation (3), and allowing N_A to vary, we obtain the following first order condition for the benefit level of region A:

$$MV'(w_A + B_A) = \frac{N_A + \frac{\partial N_A}{\partial B_A} B_A}{1 + w'(N_A)\frac{\partial N_A}{\partial B_A}} \tag{6}$$

By comparing equation (6) with the first order condition in the no-mobility case, equation (4), we easily see that the optimal benefit level is lower in the presence of migration. Two effects contribute to this: First, since increases in the benefit level now cause welfare migration, a marginal increase in the benefit level will increase total costs more than in the no mobility case. We call this the "cost effect". Second, benefit increases are less productive when the poor are mobile. The reason is that the induced welfare inmigration has a negative effect on the local wage, which partly offsets the increase in the local poor's income. We denote this the "wage effect".

Our main interest, however, lies in the interaction between the benefit levels of the regions, i.e. in the benefit level reaction functions. In order to simplify the derivation of these, following Brueckner (2000), we assume simple quadratic functional forms for utility and production. Specifically, we assume that $U(x_i, w_i + B_i) = x_i + \eta(w_i + B_i) - \frac{1}{2}\theta(w_i + B_i)^2$ with $\eta, \theta > 0$, and that $f(N_i) = \alpha N_i - \frac{1}{2}\beta N_i^2$, with $\alpha, \beta > 0$, which gives $w(N_i) \equiv f'(N_i) = \alpha - \beta N_i$.

By applying these functional forms and by combining the first order condition in equation (6) with the migration equilibrium constraint in equation (5), we can solve for B_A as a function of B_B:

$$B_A = \Psi + \frac{2 - M\beta\theta}{(4 + M\beta\theta)} B_B \tag{7}$$

where Ψ is a constant.

Equation (7) shows the interaction between the benefit levels of neighboring regions. By using the quadratic functional forms to solve for $B_A(N_A)$ in equation (6), we find that the following holds[9]:

$$\begin{array}{ccc} > & 0 & \\ \frac{\partial B_A}{\partial B_B} = & 0 & iff \\ < & 0 & \end{array} \quad \begin{array}{ccc} & < & 0 \\ \frac{\partial B_A}{\partial N_A} & = & 0 \\ & > & 0 \end{array}, \tag{8}$$

[8] This assumption is not crucial to the qualitative results, but could be replaced by other assumptions that constrain the migration elasticity of the poor, for example idiosyncratic moving costs or regional preferences (see for example Smith (1991) or Wheaton (2000)). The wage assumption has the advantage of being straightforward and easy to analyze.

[9] The model setup is symmetric, so the corresponding holds for B_B.

We see that, in this theoretical framework, we can expect some interaction between the benefit levels, unless we have the knife-edge case of zero-sloping reaction functions. The sign of the reaction functions depends on whether the "cost effect" or the "wage effect" dominates: if the cost effect is larger than the wage effect, we have positively sloped reaction functions; if they exactly balance, we have a zero slope; and if the wage effect is larger, the reaction functions have a negative slope.[10]

In empirical work, equation (7) is typically estimated through an equation similar to the one given in equation (1). Since the welfare benefit levels in the two regions are determined simultaneously, there is however a simultaneity problem to be solved. In order to do this, we would, generally speaking, like to have a variable that is exogenously distributed among the regions and that affects the setting of welfare benefit levels in one of the regions but that does not directly affect the corresponding levels in the neighboring region (i.e., we need a variable that shifts the reaction function of one region but not that of the other).

The theoretical analysis just laid out shows that one variable that is likely to affect the setting of welfare benefits in a region is the inflow of welfare prone individuals to that region (c.f. equation (8)). This suggests that an exogenous increase in the number of poor in a region could be used to instrument for the benefit level. If one could find a social program or a policy intervention that generates an exogenous placement of a welfare prone group in the regions, that program or policy intervention could be used as an instrument to solve the simultaneity problem arising in equation (7).

We argue that such a policy intervention existed in Sweden in the late 1980s and early 1990s, in the form of an exogenous placement of refugees. The aim of this paper is to use this policy intervention as an instrument to break the simultaneity problem arising from the welfare game. What is required for the policy intervention to be a valid instrument? To examine this, we will next analyse the policy intervention within the theoretical framework.

2.4 An analysis of a policy intervention

What effect will a policy intervention that leads to an exogenous increase in the number of welfare prone individuals in one of the regions, have on the welfare benefit levels in the two regions? In this section, we will analyze such a policy intervention within the theoretical framework presented above. We will call the individuals that

[10] It is hence the "wage effect", or the assumption that wages depend negatively on the number of poor in the region, that lies behind the possibility of negatively or zero sloped reaction functions in this model setup. This scenario is not unrealistic. We can think of other mechanisms that would yield the same result, for example including housing costs of the poor in the model, and letting these increase in the number of poor in the region.

are placed in one of the regions "refugees". In order to keep the section short and simple, we will assume positively sloping reaction functions.[11]

The refugee placement program can be described as a 3-stage game between regions A and B. We assume that only region B receives refugees, in order to derive how this affects the benefit levels of the two regions. (That is, we view region B as the "neighbor" and investigate how a change in its benefit level affects the benefit level in region A.) We furthermore assume that the refugees are poor (i.e., that they are welfare recipients). Unlike the native poor, however, we assume that the refugees do not work. This implies that the migration constraint of the refugees differs from that of the native (working) poor. We do not explicitly model any migration constraint mechanism for the refugees, but start by assuming that the refugees are immobile between the regions, and then analyse what happens if this assumption is relaxed.

In order to be able to separate between the arriving refugees and the "native poor", we change the notation of the native poor of region i to $\hat{N}_i$, and use $\mathring{N}_i$ to denote the refugees of region i.

2.4.1 Case 1: Refugees immobile between the regions

Stage 0: We start in a stable equilibrium, where the benefit levels of the regions satisfy the first order conditions in the mobility case, and the migration equilibrium of the native poor, equation (9), is fulfilled. The benefit level of region i hence satisfies:

$$MV'(w_i + B_i) = \frac{\hat{N}_i + \frac{\partial \hat{N}_i}{\partial B_i} B_i}{1 + w'(\hat{N}_i)\frac{\partial \hat{N}_i}{\partial B_i}}$$

and

$$w(\hat{N}_i) + B_i = w(2\overline{\hat{N}} - \hat{N}_i) + B_j \tag{9}$$

Stage 1: At stage one, the refugees, $\mathring{N}_B$, are placed in region B. The increase in the number of poor of the region increases the total benefit costs of the rich of region B. The inflow of refugees hence changes the budget constraint of the rich (see equation (3)) to also include the cost of the immigrants.

$$x = y - \frac{\hat{N}_B B_B}{M} - \frac{\mathring{N}_B B_B}{M} \tag{10}$$

The wage level is, however, unaffected by the refugee placement, since the refugees do not work. This also implies that the migration constraint of the native poor, equation (9), is unchanged. Assuming that the rich of a region care only about the

[11] The points to be made in this section do not rest on this assumption, but hold for the model in general.

native poor, we can rewrite the utility function of the rich in region B as[12]:

$$U(x_B, w_B + B_B) = y - \frac{\hat{N}_B B_B}{M} - \frac{\mathring{N}_B B_B}{M} + V(w(\hat{N}_B) + B_B) \tag{11}$$

How does the refugee placement affect the benefit levels of the regions? This depends on our assumption regarding the mobility of the refugees. The assumption that the refugees are immobile between the regions introduces an asymmetry in the model, since the refugees are placed only in region B. The optimal benefit level of region B after the refugee placement is derived by maximizing equation (11) with respect to the benefit level and such that equation (9) holds. The resulting first order condition for region B is given by:

$$-\frac{(\hat{N}_B + \mathring{N}_B)}{M} - \frac{B_B}{M}\frac{\partial \hat{N}_B}{\partial B_B} + V'(w(\hat{N}_B) + B_B)\left[w'(\hat{N}_B)\frac{\partial \hat{N}_B}{\partial B_B} + 1\right] = 0 = \Omega^1 \tag{12}$$

By differentiating equation (12) with respect to B_B and $\mathring{N}_B$, we obtain the effect of the refugee placement on the benefit level of region B[13]:

$$\frac{\partial B_B}{\partial \mathring{N}_B} = -\frac{\Omega^1_{\mathring{N}_B}}{\Omega^1_{B_B}} < 0$$

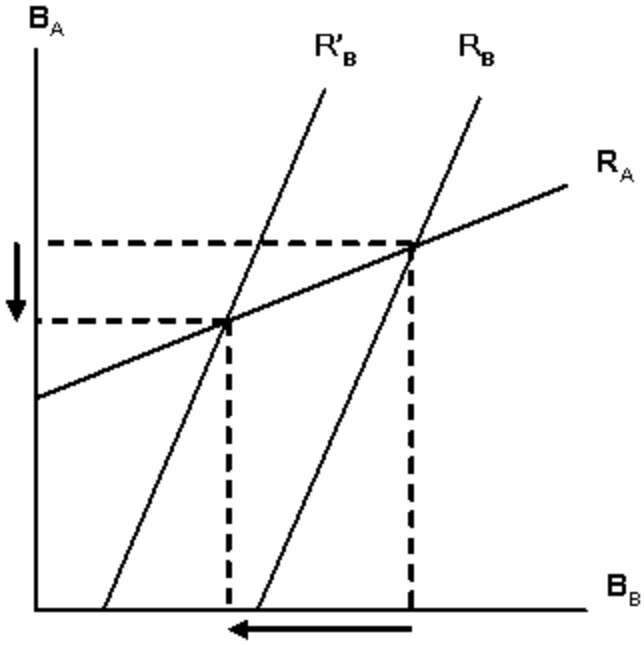
We see that the optimal benefit level of region B, given the benefit level of region A, is lower after the refugee placement. This implies a downward shift in the reaction curve of region B:s benefit level (see Figure 1).

Stage 2: At stage 2 region A responds to the decrease in B_B. Since no refugees have been placed in region A, the first order condition for the benefit provision of region A is equal to that of Stage 0. However, the benefit decrease of region B affects region A, since it makes some working poor from region B move to region A. Assuming linear reaction functions with a positive slope, the effects of the refugee placement in region B on the benefit levels of the two regions can be illustrated by the reaction functions in Figure 1.

[12] This assumption implies that the refugees only enter as a cost in the utility of the rich. This, together with the assumption that the poor immigrants do not work, ensures a negative effect on the benefit level in region B of the refugee placement.

[13] We know that $\Omega^1_{B_B} < 0$ by the assumption of strict concavity in $V(w(\hat{N}_B) + B_B)$. $\Omega^1_{\mathring{N}_B} = -\frac{1}{M} < 0$ is easily seen from equation (12).

Figure 1: Reaction functions, No refugee mobility



In Figure 1 we see that the benefit levels of both regions are lower after the refugee placement. The mechanisms are the following: In the first stage, the reaction function of region B, R_B, shifts, because the regions' optimal level of benefit is now lower, *given* the benefit level of region A. In the second stage, region A reacts to the benefit decrease of its neighbor, by lowering its benefit level. The effect on B_A is hence channeled through B_B and corresponds to a movement along the reaction function of A, R_A, in the figure. As Figure 1 shows, the decrease is larger for the benefit level of region B.

2.4.2 Case 2: Allowing the refugees to move

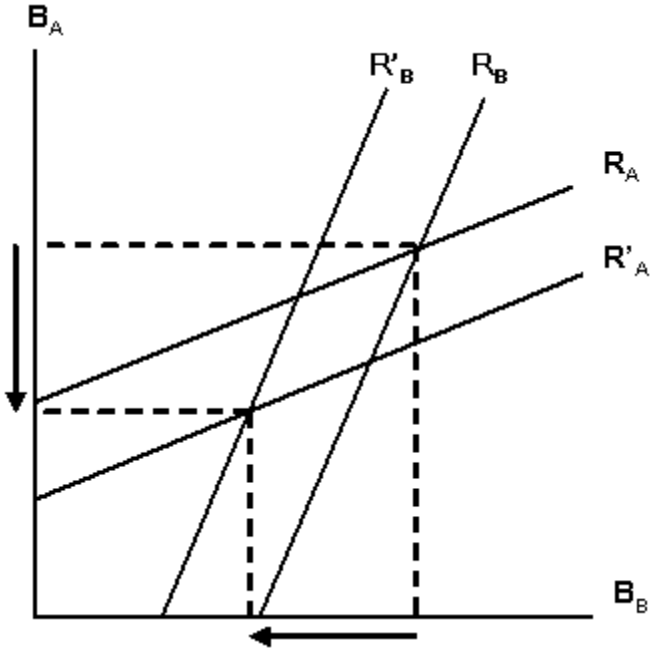
Figure 1 describes the case when the refugees are assumed to be immobile between the regions. What happens if we relax this assumption?

The assumption of immobile refugees introduced an asymmetry in our otherwise symmetric model. This asymmetry results in different optimal benefit levels of the regions. In addition, and more importantly for the empirical part of this paper, the assumption of immobile refugees assures that all the effect of the refugee placement in region B on the benefit level in region A, is transmitted through the change in the benefit level in region B. This is important for the validity of the refugee placement as an instrument for the benefit level.

If we instead assume that the refugees are perfectly mobile, we are back in a symmetric model. The intuition behind this is the following: If the migration of the refugees is perfectly elastic, it does not matter in which region they are initially placed, but they will "immediately" move to the region with the higher benefit level.

The increase in the total benefit costs will be shared equally between the regions, and we will hence see equal shifts in the reaction functions, as shown in Figure 2.

Figure 2: Reaction functions, Perfect refugee mobility



We have now illustrated how the interaction between the benefit levels of the regions work in two polar cases regarding the mobility among the placed refugees. The conclusion from this analysis is that the validity of a policy intervention, that exogenously places individuals in certain regions, as an instrument for breaking the simultaneity in the welfare game or not, crucially hinges on the migration pattern of the placed individuals.

3 The policy intervention: Exogenous placement of refugees

This section will describe the main characteristics of the refugee placement program and discuss the appropriateness of using it as an instrument for breaking the simultaneity problem in the welfare game.

3.1 Description of the refugee placement program

The system of non-voluntary placement of refugees was in place between the beginning of 1985 and the first of July 1994. The assignment of refugees to the municipalities was coordinated by The Immigration Board through municipality-wise contracts. The purpose of the program was to achieve a more even distribution of refugees over the country, or more specifically, to break the concentration of immigrants to larger towns.

Initially, only a fraction of the municipalities were contracted, but as the number of refugees soared in the late 1980s and early 1990s, so did the number of receiving municipalities. In 1991, 277 out of 286 municipalities had agreed to participate.

The original ambition was to direct the flow of immigrants toward municipalities with good future prospects in terms of labor market conditions and education possibilities. The increasing inflow of immigrants combined with the shortage of housing during the second half of the 1980s and the early 1990s meant, however, that these ambitions had to give way to the more immediate concern of available housing.

The municipalities received financial compensation, paid out by the Immigration Board, for the refugee placement. Compensation was paid out gradually, during the year of placement and the 3 following years, to compensate for the running expenses of the receiving municipality. After that period, the central government's compensation to the municipalities ended. The larger part of the expenditures consisted of welfare benefit payments. In 1991, the system was replaced by one where the municipalities were given a lump sum grant for each refugee. The grant was paid out during the year of the placement, and was estimated to cover the expenses of the municipality for about 3,5 years. In addition to the grant, the municipalities had the possibility to apply for compensation for "extraordinary costs" for the refugee placement, for example for old or disabled refugees that were in need of special care.[14]

The refugees were allowed to move immediately after the placement. Under the system with running expenses, the compensation was tied to the refugee, i.e. it was provided to the new municipality in case of migration. This was not the case under the lump sum system, where the municipality of placement received the entire sum, no matter how long the refugee stayed in the municipality. If the refugee did move within two years after placement, the new municipality also received some compensation.[15]

3.2 Using the refugee placement program as an instrument for the benefit level

We will use the policy intervention defined by the refugee placement program between 1986 and 1991 to instrument for the rival municipalities' welfare benefit levels in 1990-1994.[16] In order to motivate that the refugee placement program is an appropriate instrument, we need to discuss the exogeneity of the refugee placement program. In

[14] This system was in place until 1996. From 1996 the compensation is in the form of a lump sum, but it is paid out gradually during a 2-year period. (The Immigration Board (1997, pp22f))

[15] One previous study uses the refugee placement program as a natural experiment, Edin, Fredriksson and Åslund (2003). They study the consequences of the program placement for the labor market participation of the refugees and use data for 1987-91. The paper provides a detailed description of "the handling of a typical asylum seeker from the border to the final placement".

[16] According to Edin et al. (2003) the refugee placement program was more strictly implemented during this initial period of the program, than during the later years.

addition, we need to show that it is reasonable to believe that the placement of refugees in a municipality affected the benefit level of that municipality, but not directly the benefit levels of the rival municipalities.

Regarding the exogeneity of the program placement, what is important for our analysis, is that the refugee placement was exogenous from the point of view of the municipalities. The period we study is characterized by a couple of circumstances that we argue support this claim.[17]

First, during the time period we use, the number of refugees arriving to Sweden increased dramatically. During 1986-91 on average over 16,000 refugees arrived each year (peaking in 1989 at 24,879), compared to a yearly average of just above 5,000 during the previous six years. This made it harder for the municipalities to refuse to accept the Immigration Board's refugee placement proposals: the refugees had to be placed somewhere, and with the increasing inflow, all municipalities had to share the responsibility for this. Interviews with persons that were in charge of the placement also confirms that there was a sense of solidarity among the municipalities and that the municipalities generally accepted to participate, especially during the early years of the placement program.[18] In addition, some of the municipalities that did refuse, received a lot of negative publicity for this.

Second, refusals to accept refugee placement were in fact very rare. Only 5 out of the 281 municipalities in our data refused to receive any refugees at all during the period we study. We believe it likely that the decision to refuse refugee placement was connected to municipality-specific parameters that stay relatively fixed over time, such as ideology.

Apart from being exogenous from the municipalities' point of view, in order to be used as an instrument it is also important that the Immigration Boards's placement of refugees was not guided by certain characteristics in the municipalities that were also correlated with the welfare benefit levels. Interviews with government officials that were implementing the program suggest that if there was any factor that affected the refugee placement, it had to do with the availability of housing in the municipalities. The reason for this is that the period under study is characterized by a very tight housing market. This means that if any factor, except for fixed municipality-specific characteristics, did influence the refugee placement, it was probably the availability of housing.[19]

[17] There is very little written documentation on this topic. The information provided here is therefore based on two different sources; on the written information and on interviews with government officials that were implementing the program.

[18] It can be noted that the instruments we use are from the early years of the placement program.

[19] This claim is supported by various studies that argue that the high unemployment rates among immigrants from 1980 and onwards are partially due to the fact that housing, instead of factors such as labor market prospects, has been determining the refugee placement (see for example Edin et al., 2003).

Based on these circumstances, we argue that the refugee placement can be viewed as exogenous from the point of view of the municipalities, *conditional on housing vacancies and on municipality-specific fixed effects.*

Another criteria for using the refugee placement program as an instrument for the benefit levels, is that the program actually affects these. A first prerequisite for this is that the program leads to an increase in the number of welfare prone individuals and that this increase, in turn, implies increased costs for the municipalities.

When placed in a municipality, the refugee was supported by welfare benefits during a period of introductory Swedish courses, and after that until he/she had found other maintenance. It is therefore reasonable to expect that, during an initial period, the larger part of the refugees received social assistance. The municipalities were compensated for this, during the first part of the period we study through the compensation for running expenses during three years, and from 1991 on, through a lump sum grant that was to cover the expenses for a corresponding period. The question is whether this compensation was enough, i.e. whether the refugees had moved out of welfare when the compensation ended or not.

Looking at the data, we find it likely that the refugee placement did increase the pool of welfare dependent inhabitants in a municipality. Refugees, as well as foreign citizens in general, are overrepresented in the data on welfare recipients. Over the period 1990-1994, refugee households made up on average 11 percent, and non-Swedish citizens in general (including refugees) 26 percent, of the welfare-receiving households, while the fraction of refugees and the fraction of foreign citizens (including refugees) in the population during the same period roughly equalled 1 and 6 percent respectively.[20]

These figures may however merely represent the fact that the refugees are supported by welfare during an initial period in the country, for example during the period of mandatory introductory Swedish courses. For us to be able to use the refugee placement as an instrument, i.e. for the refugee placement to affect the costs of the receiving municipalities, we need a significant number of the refugees to stay on welfare also after the termination of the financial compensation scheme.

Franzén (2004) analyzes welfare dependency among immigrants, based on interviews conducted in 1996 with refugee immigrants that arrived in Sweden between 1980 and 1989. Of the immigrants in the sample, 24 percent are recipients of welfare benefits after 7-16 years in Sweden. In comparison, the share of welfare benefit recipients in the population in general in 1996 was below 10 percent (8.4 percent,

[20] Based on data from Statistics Sweden and the Migration Board. A person is defined as a refugee during the year of receiving a residence permit and the three following years. After that he/she is defined broadly as a foreign citizen or as a Swedish citizen if a Swedish citizenship is obtained (a refugee can obtain a Swedish citizenship at the earliest after four years).

SCB).

Hansen and Lofstrom (1999) also show that refugees as a group are less likely to move out of welfare than the native population; still after 20 years in Sweden both refugee and non-refugee immigrants show higher social assistance participation rates than statistically similar indigenous Swedes.

Our descriptive data shows that the average size of the annual refugee placement to a municipality, during the period we study, was equal to five percent of the pool of welfare recipients. Provided that many of the refugees stayed on welfare also after the compensating financial grant had run out, as suggested by the studies cited above, we can conclude that the effect on the welfare costs must have been quite substantial, at least for the municipalities in the upper part of the distribution.

This suggests that even though the municipalities were to some extent compensated for the refugee placement, we can expect some of the costs to remain after the compensation period. The fact that the municipalities were provided compensation for the first three to four years of the placement, furthermore suggests that the effect is probably lagged, and, in the baseline analysis, we will use the number of refugees placed in $t-3$ and $t-4$ as instruments.[21]

It hence seems like the refugee placement program increased the pool of welfare dependent inhabitants and also increased costs in the municipalities.[22]

3.3 Refugee migration

As noted in the theoretical framework, the appropriateness of using the refugee placement program as an instrument for the neighboring municipalities' benefit level hinges on the migration pattern of the refugees. Ideally, we wish that the refugees stay in the municipality in which they were initially placed (c.f. Figure 1). Otherwise, some of the cost effect may "spill over" directly through the migration of the refugees, which, in terms of the theoretical model in section 2, implies that some of the effect on the welfare benefit level in region A of the refugee placement in region B is a direct cost effect, and not a result of interactions on benefit levels.

[21] It can be noted that several of the Swedish municipalities that found themselves in financial trouble during the 1990s, claim that one of the main explanations to the financial situation in their municipality was due to increased costs in the wake of the refugee placement program. Also, several of the municipalities claim that they had been undercompensated in the first place, indicating that the inflow of refugees might have lead to a real cost for the municipalities earlier than three to four years after the placement. Therefore, we will, in the sensitivity analysis, examine how sensitive the baseline results are to different lags on the instruments. Since 1994 is the last year of the panel, all the observations on refugees in the baseline analysis are from the period under which the implementation of the refugee placement program was the strictest (i.e., up until 1991). This is an advantage, considering the exogeneity of the instruments.

[22] If the instruments are relevant empirically will be examined through the first-stage estimates in the results section.

How large is the risk of direct cost spill-overs between municipalities from refugee migration? This can be evaluated by looking at the migration pattern of refugees during our sample period. This information is unfortunately not directly available, but we can obtain an approximation by using sample data on the total immigration (refugees and non-refugees) to Sweden. The sample consists of approximately 20 percent of the immigrants to Sweden during 1987-89. Following Åslund (2000) and Edin et al. (2003), we exclude observations of immigrants originating from OECD-countries and a number of additional western European countries from the sample, in an attempt to remove the non-refugee immigrants from the sample. In addition, immigration of relatives of the refugees has been removed from the sample. We are left with a sample of 9,283 observations, which is indeed roughly equal to 20 percent of the total number of refugees during the period.[23]

In the data we can observe the municipality of residence for the refugees during the year of arrival, and four years after arrival. The sample hence informs us of the migration pattern of the refugees between these time periods.

Looking at some descriptive statistics, we see that 9,080 of the 9,283 refugees were still living in Sweden four years after the initial placement (i.e., 203 of the refugees had either migrated from Sweden or died). Out of the 9,080, 60.5 percent were still living in the municipality in which they were initially placed. This means that 3,589 refugees had changed municipality after four years. Where had they moved?

It turns out that it is the three big towns in Sweden (Stockholm, Malmö, and Göteborg) and their surrounding areas that are the main magnets. Out of the refugees that had changed municipality, the majority (68 percent) had moved to or within one of the counties of these three towns; the Stockholm, Malmö and Västra Götaland counties (roughly 60 percent of them had moved from counties other than these three, and approximately 40 percent had moved within or between these counties). We can conclude that the main migration flows are to these counties.

Since we use neighboring municipalities as each municipality's reference group, our instrument is especially sensitive to refugee migration between neigboring municipalities. In our sample, such migration is rare; only 624 of the refugees have moved to a neighboring municipality after four years. This is equal to 6.7 percent of the total sample, or 17 percent of those that have moved. 367 of them had moved to or within the Stockholm, Malmö and Västra Götaland counties.

Three conclusions can be drawn from these descriptive statistics. First, even after four years, the majority of the refugees (60.5 percent) is still living in the municipality in which they were initially placed. Second, only a small fraction of the total number of refugees that arrived four years earlier, had moved to a neighboring municipality.

[23] The number of granted residence permits over the period 1987-89 for refugees were 55046 (The Immigration Board).

This implies that the risk that our results suffer from bias stemming from direct "cost spillovers" between the municipalities is small. Third, out of those that after four years had migrated within Sweden, the great majority had moved to or within one of the three big city counties in Sweden: the Stockholm, Malmö and Västra Götaland counties. This means that if there is any "cost shifting" going on, it is mainly the three big cities and the surrounding areas that are bearing these costs.

The fact that the migration between neighboring municipalities seems to be small suggests that the risk of direct cost spill-overs is probably small. Furthermore, the majority of the refugees remain in the municipality of placement four years after arrival. We will therefore initially conduct the analysis without taking account of municipality-wise refugee migration. We will, however, test for the robustness of the results to secondary migration by presenting estimation results when the three migration-magnet counties are excluded.

4 Data and econometric considerations

4.1 Data

The reaction function derived in the theoretical model is estimated using data on the 280 municipalities' generosity in providing welfare benefits over the years 1990-94.[24] The reason for starting in 1990 is that we have to use the number of refugees in earlier periods as instruments (the longest lag is $t-4$; see below). Since we only have information on refugee placement from 1986, the first year in which we can use the welfare benefit levels is 1990.

There are a couple of potential candidates for measuring the benefit generosity of a municipality. One is the norm that regulates the amount of benefits that a person is eligible for, the other is the actual benefit expenditures. We choose to focus on the benefit expenditures in the baseline analysis. There are two reasons for this. First and foremost, it enables us to use a longer panel (data on expenditures is available for several years, while data on the benefit norm is only available for the years 1991, 1992 and 1994). In addition, by using benefit expenditures rather than the benefit norm, we avoid the risk of distortions based on imperfect implementation of the norm.[25] We believe that our definition of the welfare benefit level takes us closer to the "true generosity" of the municipalities.[26]

[24] Seven municipalities (Gnesta, Trosa, Nyköping, Bollebygd, Borås, Lekeberg och Örebro) were excluded since they had been involved in either secessions or mergers of municipalities in the time period 1989-1994. The municipality of Gotland was excluded since it is an island and consequently has no border-sharing neighboring municipalities.

[25] There has been a discussion in Sweden that there is an heterogenous implementation of the benefit norm, both within and between municipalities.

[26] In the sensitivity analysis, we will however examine whether the results are sensitive to this by

The typical procedure in the literature is to normalize the welfare expenditures by the number of benefit recipients. The drawback with that definition is that it is a rough measure, in the sense that it does not pick up variations in the time that a person spends on welfare - i.e. a person that is on welfare at some point during a year counts as a benefit recipient, regardless of the number of months he or she receives benefits. An alternative, and in this respect more precise, measure is to normalize the benefit expenditures by the total number of benefit months. Therefore, we have chosen to normalize by the number of benefit months, since this strikes us as a more straightforward and intuitive measure, but will in the sensitivity analysis check that the results obtained in the baseline estimations are robust to the alternative normalizing factor.

During the period we study, the municipalities were free to set their own benefit norms.[27] From Table 1 it is clear that this decentralized decision-making in the setting of welfare generosity led to a large variation in the benefit paid out; the mean benefits paid out per benefit month was 3,960 SEK, with a standard deviation of 600 SEK (and with a minimum of 2,000 and a maximum of 7,900 SEK).[28] In fact, it was the great variation between the municipalities, in particular the tendency to set the levels below the recommendations of the Board, that finally led to the introduction of a mandatory minimum level in 1998 (The National Board of Health and Welfare (1999)).

Descriptive statistics on the covariates for the years 1990-1994 are also given in Table 1. The covariates are unemployment, tax base, grants from the central government, population 19-29. Benefit level, tax base and grants are measured in 1000 SEK, while the rest of the covariates are given in percent. These are variables that have been included in similar studies and/or that we, based on Swedish welfare data, believe likely to affect the benefit expenditures. Table 1 also shows the share of vacant housing, as well as refugee placement, given as percentage points of the population. Since we have to consider lagged effects when using the number of refugees received by the municipalities to instrument for the benefit level, we will use refugee data for the years 1986-1994. For housing vacancies, which enters lagged three and four time periods in the estimations, the table shows the values for 1986-1991.

re-estimating the model using the norm instead of expenditures.

[27] There was no mandatory rule for the benefit levels, but general guidelines were provided by The National Board of Health and Welfare (Socialstyrelsen).

[28] All monetary variables are deflated to 1990 year values. There was also a significant variation in the benefit norms set by the municipalities. During the years 1991, 1992 and 1994 (which are the years for which we have information about the norms), the benefit norm averaged 112, with minimum and maximum levels at 80 and 145 and with a standard deviation of 7.8. (The norm is defined as the percentage of the basic amount and we have used the unadjusted levels for a single individual.).

Table 1: Descriptive statistics

Variable	Obs	Mean	Std.Dev	Min	Max
Benefit Level	1393	3.96	0.6	2	7.9
Unemployment	1400	4.5	2.6	0.2	12.2
Tax Base	1400	659	98	463	1366
Grants	1400	48	101	-1496	1531
Population 19-29	1400	13.9	1.8	10	23
Vacant Rentals	1647	1.3	2.3	0	29
Refugees	2512	0.3	0.3	0	5.8

The overall mean number of refugees received by the municipalities over the period 1986-1994 was 84. In the analysis we will use the number of refugees normalized by the population in the municipality. As can be seen in Table 1, the annual refugee placement to a municipality during the period equalled 0.3 percent of the population. Normalizing with the number of welfare recipients in the municipality, the corresponding number is 5.5 percent.

4.2 Econometric considerations

Before turning to the results, we will discuss how the econometric model shall be specified and how the municipalities' reference group shall be defined.

Let us start with the question of how the municipalities' reference group shall be specified. That is, which municipalities play welfare games with each other? It seems reasonable to assume that the municipalities' fear of welfare immigration from other municipalities is stronger the closer these municipalities are. Since the welfare recipients may have better information about the welfare generosity in nearby municipalities and since migration costs increase with distance, it is for example likely that welfare recipients have a stronger migration response to differences in nearby municipalities than to differences in municipalities further away.[29] This also captures the idea that geographical neighbors belong to the same media market and therefore have good information about the generosity of neighboring states.[30] We therefore define a municipality's reference group as the set of municipalities with which it shares border.

For the empirical work to be trustworthy when estimating the reaction function given in equation (1), it is important that the estimates do not suffer from bias due to unobserved correlated shocks. Is such bias likely to affect our estimates? If the if the refugee placement were a purely random process, we would not have to worry about such bias. In our case, as discussed earlier, the refugee placement can however not be considered a completely random process from the municipalities' point of view, unless we condition on housing vacancies and municipality-specifc fixed effects. This

[29] These arguments are put forward by Saavedra (2000).

[30] For example, Besley and Case (1995) also use this definition of reference group in their empirical analysis of tax interaction effects between neighboring U.S. States.

underlines the need to control for housing vacancies and municipality specific fixed effects in the regressions: If the number of vacant housing in a municipality affects the amount of refugees the municipality will receive, and if the number of vacant rentals is correlated with the welfare benefit level in the municipality, then omitting to control for the number of vacant rentals might induce a spurious correlation between the number of refugees and the benefit level.[31] Since the instruments are dated in $t-3$ and $t-4$, so are the vacant rentals variables. A similar argument applies for the fixed effects, measuring variables that can be considered as being constant for a municipality over the time period that we study, such as ideological preferences vis-à-vis immigrants.

However, if controlling for housing vacancies and fixed effects are not enough, we will take additional measures to ensure that we are estimating a causal effect of neighbors' benefit level on own benefit level. In particular, it is of crucial importance that the variation over time in the placement of refugees in neighboring jurisdictions is random with respect to any unobserved factors that might be correlated among neighboring municipalities and that directly affect the variation over time in welfare benefit levels in a given jurisdiction. In effect, we need to make sure that our estimated coefficient for strategic interaction is not merely the result of spatial-temporal correlation in the refugee placement (i.e., correlation within and between municipalities). To illustrate why this may potentially be a problem, assume that the number of refugees placed in a municipality in $t-2$ has an independent and negative effect on a municipality's welfare benefit level.[32] Suppose also that the number of refugees placed in a municipality in $t-2$ is positively correlated with the number of refugees placed in the neighboring municipalities in $t-3$. In this case, failing to control for the number of refugees placed in a municipality in $t-2$ will result in a biased estimate on the interaction coefficient with the false impression of strategic interactions.

To get an indication of whether spatial-temporal correlations might be a problem, we can look at the raw correlations. As can be seen from the main diagonal in Table (2), there is a quite strong positive correlation between contemporaneous placement of refugees in a municipality and among the municipality's neighbors (ranging from 0.61 to 0.69). However, as can be seen from the diagonals off the main diagonal, the correlation is much lower, often close to zero, between the placement of refugees in a municipality and among its neighbors at different points in time. This indicates that the spatial component is much more pronounced than the temporal one.

[31] It can however be noted that the number of vacant rentals is close to a municipality-specific fixed effect, implying that this might already be controlled for by the fixed effect specification.

[32] One reason why the number of refugees placed in $t-2$ might have an independent effect on a municipality's welfare generosity might be that the central government did not provide enough funding for the refugees or that the municipalities used up their funds too quickly.

Table 2: Correlation matrix for refugee placement

	Neighbors' Refugees (t)	Neighbors' Refugees ($t-1$)	Neighbors' Refugees ($t-2$)	Neighbors' Refugees ($t-3$)	Neighbors' Refugees ($t-4$)
Refugees (t)	0.690	0.138	-0.107	-0.081	-0.048
Refugees ($t-1$)	0.142	0.666	0.091	-0.187	-0.128
Refugees ($t-2$)	-0.098	0.094	0.644	0.034	-0.213
Refugees ($t-3$)	-0.067	-0.174	0.039	0.606	0.005
Refugees ($t-4$)	-0.041	-0.117	-0.201	0.011	0.609

To ensure that the IV-regressions only uses the variation over time in neighbor refugee placement that is orthogonal to own refugee placement, we will control for own refugee placement in t, $t-1$, $t-2$, $t-3$, and in $t-4$.

By conditioning on own refugee placement, the risk that unobserved correlated shocks influence the estimations is greatly reduced. To reduce the risk even more, we include several additional covariates that pick up correlated shocks; municipality-specific and time-specific fixed effects to control for unobserved shocks, and time-varying municipality-specific covariates, such as the unemployment rate, income variables, and demographic structure to control for observed idiosyncratic shocks.

The model to be estimated is then given by

$$B_{it} = \gamma \overline{B}_{(-i)t} + X_{it}\beta + \sum_{k=0}^{4} \delta_k R_{it-k} + \phi_1 \overline{VR}_{(-i)t-3} + \phi_2 \overline{VR}_{(-i)t-4} + \alpha_i + \lambda_t + \varepsilon_{it} \quad (13)$$

where B_{it} is the welfare benefit level in municipality i in time period t, $\overline{B}_{(-i)t}$ is the average welfare benefit level among municipality i's neighbors, X_{it} is a vector of time varying municipality-specific characteristics, R_{it-k} is own refugees received in different time periods, $\overline{VR}_{(-i)}$ is the average number of vacant rentals among the neighboring municipalities (dated in the same time period as the instruments), α_i is a municipality-specific fixed effect, λ_t is a time specific effect, and ε_{it} is an error term.

A final thing to consider is potential serial correlation in the welfare benefit level. If there is such a correlation in the error process, the resulting standard errors are inconsistently estimated and may lead to severely biased estimates in small samples (see, e.g., Kézdi (2002) and Bertrand, Duflo and Mullainathan (2004)). Therefore, we will allow the errors to be correlated over time within each municipality.[33]

Given the specification in equation (13), we believe that γ measures a causal effect.

5 Results

In this section we present our results. First, we present, for comparative reasons, the OLS results. This is followed by the results of the first stage estimates in the

[33] Technically, this is done in STATA by clustering on municipality.

two-stage procedure. Finally, we present our IV-estimates that measure the causal effect of the welfare generosity of neighboring municipalities on the welfare generosity of a certain municipality.

5.1 OLS estimates

For comparative reasons, we initially neglect simultaneity and start by estimating the model without using instruments. As can be seen from the first column in Table 3, neighbors' benefit level enters significantly and with a positive sign; if neighbors decrease their welfare benefit level with 100 SEK per benefit month, a municipality decreases its benefit level with approximately 30 SEK per benefit month.

5.2 First stage estimates

In the first stage regression in the IV approach, we run the average welfare benefit level in neighboring municipalities (the endogenous variable) on the instruments (the average number of refugees that was placed in the neighboring municipalities, measured as share of the population, in $t-3$ and $t-4$). The reduced form estimates of the endogenous variable on the instruments provide information about the relevance of the instruments. These results are presented in the second column in Table 3. As can be seen from the results, there is a significant and negative association between the number of refugees and welfare benefit generosity; the higher the average placement of refugees among the neighboring municipalities in $t-3$ and $t-4$, the lower is the average welfare benefit level among the neighboring municipalities in t. The instruments hence seem to be relevant.

Regarding the strength of the instruments, there are no indications that our instruments are weak. Staiger and Stock (1997) suggest using the F-statistic for the joint significance of the excluded instruments in the first-stage equation as a diagnostic of the power of the instruments. They argue that if the F-statistic is larger than 10, there should be no problem associated with weak instruments. Conducting partial F-tests on the excluded instruments in the first-stage regression, we get an F-statistic of 33.22.

5.3 IV estimates

The two-stage least squares (2SLS) estimates[34] are presented in the last column in Table 3. As can be seen from the first row, there is a significant and positive effect from the setting of welfare benefit levels in neighboring municipalities on the setting of the welfare benefit level in a given municipality. The point estimate indicates that if the neighboring municipalities decrease their welfare benefit level with 100 SEK, a

[34] In this paper we use IV and 2SLS synonymously.

municipality decreases its benefit level by approximately 41 SEK. The estimate for neighbors' benefit level hence provides indications of strategic interactions among the local governments in the setting of welfare benefit levels, implying that there exists a "race-to-the-bottom".[35]

Since we have an overidentified model, we can use a test for overidentifying restrictions to test for instrument validity/correct model specification. From the Hansen J-statistic, presented in the last two rows of Table 3, it is clear that we cannot reject the null hypothesis of valid instruments/correct model specification.

Next we turn to the other variables. While the number of refugees received by a municipality in t and $t-1$ leads to municipalities being more generous in their setting of welfare benefit levels (both have a positive and significant effect on the municipality's welfare benefit level in t), the number of refugees received by a municipality in $t-2$, $t-3$ and $t-4$ all have a significant and negative effect on the municipality's welfare benefit level in t. It is thus clear that the long-run effect is negative, which is in line with the main argument of this paper.[36] The municipality's tax base, unemployment rate, intergovernmental grants received from the central level, and population aged 19-29 do however not seem to have any significant impacts on the municipality's welfare generosity.

[35] The 2SLS-estimate is somewhat higher than the OLS estimate (c.f. the first and last columns in Table 3), but, using a Hausman test, it is clear that it is not significantly different from the OLS estimate (the t-statistic is 0.68).

[36] One explanation for the positive contemporaneous and one-year lag effects might be that the municipalities'are more generous in the beginning, shortly after they ave received their funding for the refugees.

Table 3: OLS-, first stage-, and 2SLS-estimates for baseline specification.

	OLS	First stage	2SLS
Dependent variable:	B_{it}	$B_{(-i)t}$	B_{it}
Neighbors' benefit level	.3***		.406**
	(.059)		(.167)
Neighbors' Refugees (t-3)		-.849***	
		(.127)	
Neighbors' Refugees (t-4)		-.743***	
		(.126)	
Refugees	.423***	.109***	.41***
	(.079)	(.025)	(.081)
Refugees (t-1)	.369***	-.012	.368***
	(.13)	(.047)	(.131)
Refugees (t-2)	-.373**	-.138***	-.352**
	(.154)	(.044)	(.15)
Refugees (t-3)	-.416***	-.135**	-.387***
	(.135)	(.058)	(.134)
Refugees (t-4)	-.544***	-.013	-.523***
	(.108)	(.053)	(.109)
Unemployment	-.009	-.031***	-.004
	(.017)	(.011)	(.019)
Tax base	.00004	-.0007*	.0001
	(.0008)	(.0005)	(.0008)
Grants	-3.97e-06	0*	6.07e-06
	(.00007)	(.00007)	(.00007)
Population 19-29	.018	.046*	.019
	(.031)	(.028)	(.032)
Neighbors' vacant rentals (t-3)	-.005	.031***	-.008
	(.016)	(.011)	(.017)
Neighbors' vacant rentals (t-4)	-.007	-.062***	-.002
	(.016)	(.011)	(.017)
Fixed effects	yes	yes	yes
Time dummies	yes	yes	yes
Obs.	1352	1358	1351
R^2	.336	.444	.333
Partial F instruments		33.22	
Hansen J-statistic			0.969
p-value for J-statistic			0.325

6 Sensitivity analysis

In this section we will conduct sensitivity analyses to check the robustness of the baseline results obtained in the previous section. In particular, we will examine how sensitive the results are to (i) different lag structures on the instrument (i.e., different lag lengths on the refugee variable), (ii) different definitions of the welfare generosity variable, and (iii) migration among refugees.

6.1 Different lag structures on the instrument

First we will investigate how sensitive the baseline results are to different lag structures on the instrument (i.e., on refugees). In the baseline estimations we used the average number of refugees in $t-3$ and $t-4$ as instruments. What happens if we use other lags or combinations of different lags? The results when we use different sets of instruments are presented in Table 4. In Table 4 we only report the coefficient for the neighboring municipalities' benefit level, implying that each cell corresponds to a separate regression. The covariates used in each regression are the same as those used in the baseline specification.

The first row in Table 4 simply replicates the baseline estimates (i.e., it shows the results when we use the number of refugees in $t-3$ and $t-4$ as instruments). When we use the instruments in t, $t-1$, $t-2$, $t-3$, and $t-4$ (in different combinations), we get significant estimates in the same order of magnitude as in the baseline estimations (c.f. the last three rows).

It can also be worth mentioning that when we test the validity of the instruments, using the Hansen J-test for overidentifying restrictions, we cannot reject the null hypothesis of valid instruments/correct model specification in any of the specifications in Table 4.

Table 4: Sensitivity analysis: Different lag lengths on the instruments. Only the coefficient of the neighboring municipalities benefit level is presented.

Lag length on instruments:	
$t-3, t-4$	.406**
	(.167)
$t-2, t-3, t-4$	.513***
	(.189)
$t-1, t-2, t-3, t-4$	.407**
	(.159)
$t, t-1, t-2, t-3, t-4$	.398***
	(.128)

6.2 Different definitions of welfare generosity

Next, we will examine how sensitive the baseline results are to alternative definitions of the welfare generosity variable. In particular, we will examine what happens if we use the benefit norm or welfare expenditures per beneficiary instead of welfare expenditures per benefit month. Starting by looking at simple correlations between the three measures of welfare generosity, it is obvious that the benefit norm is almost uncorrelated with the two expenditures measures, while the correlation between the two expenditures measures is much higher (c.f. Table 5). This indicates that the norm

is not implemented in the intended way, as discussed earlier.[37] The IV estimates for the three alternative definitions of the municipalities' welfare generosity is presented in Table 6. In Table 6 we only report the coefficient for the neighboring municipalities' benefit level, implying that each cell corresponds to a separate regression. The covariates used in each regression are the same as those used in the baseline specification.[38] When using the welfare expenditures per beneficiary, we note from the middle row that we get results that are fairly similar to the baseline estimate; 0.56 compared to 0.41 (c.f. the last two rows). The point estimates are not significantly different from each other. When using the benefit norm, we get a point estimate that is much higher than the baseline estimate, even though it is not significantly different from the other estimates. In all cases, the estimated effects are however significant, indicating that the baseline conclusion of strategic interaction among the municipalities in the setting of welfare benefit levels does not hinge on the way the welfare benefit level was defined.

Table 5: Correlations between the three measures of welfare generosity

	Benefit norm	Exp./beneficiary	Exp./benefit month
Benefit norm	1		
Expenditures/beneficiary	-0.018	1	
Expenditures/benefit month	-0.014	0.746	1

Table 6: Sensitivity analysis: Alternative definitions of the welfare benefit level. Only the coefficient of the neighboring municipalities benefit level is presented.

	(1)
Benefit norm	1.35**
	(.569)
Expenditures per beneficiary	.561***
	(.113)
Expenditures per benefit month	.406**
	(.167)

6.3 Migration among refugees

As suggested by the theoretical model, the policy intervention may not provide a valid instrument if the secondary migration of refugees (i.e., any migration that takes place after the initial placement) among municipalities is large. Our choice of instruments is motivated by the fact that refugees that arrive to a municipality are statistically likely to become recipients of welfare, and hence increase the welfare benefit costs of the municipality. This, of course, hinges on the assumption that the refugees stay in

[37] This pattern is also observed by Hernes Fiva and Rattsø (2003) on Norwegian data.

[38] A difference is that we only have information about the benefit norm for the years 1991, 1992 and 1994, which means that we have fewer observations in those estimations.

the municipality in which they were initially placed. It is hence of great importance to test the robustness of the results to such migration.

The descriptive statistics on refugee migration showed that the migration flows of the refugees during the time period we study were first and foremost directed towards the counties of the three largest towns, the Stockholm, Malmö and Västra Götaland counties. Furthermore, out of the 624 persons in the sample that had migrated to a neighboring municipality, more than half (367) had moved to municipalities in these counties. A straightforward sensitivity analysis is to re-estimate the model without these counties. If the baseline results are biased because of secondary refugee migration, excluding these observations will provide a model with more valid instruments.[39] The results of the IV-estimation, excluding the municipalities of the three "big city counties" are presented in Table 7.

Table 7: Sensitivity analysis: IV estimates when the counties of Stockholm, Malmö and Västra Götaland are excluded.

	Baseline	Excl. big cities
Neighbors' benefit level	.406**	.452**
	(.167)	(.178)
Refugees	.41***	.409***
	(.081)	(.104)
Refugees (t-1)	.368***	.367***
	(.131)	(.142)
Refugees (t-2)	-.352**	-.425**
	(.15)	(.167)
Refugees (t-3)	-.387***	-.436***
	(.134)	(.152)
Refugees (t-4)	-.523***	-.563***
	(.109)	(.124)
Unemployment	-.004	-.019
	(.019)	(.024)
Tax base	.0001	.0009
	(.0008)	(.001)
Grants	6.07e-06	.0003
	(.00007)	(.0002)
Population 19-29	.019	.021
	(.032)	(.056)
Neighbors' vacant rentals (t-3)	-.008	-.004
	(.017)	(.023)
Neighbors' vacant rentals (t-4)	-.002	.002
	(.017)	(.018)
Fixed effects	yes	yes
Time dummies	yes	yes
Obs.	1351	853
R^2	.333	.389
F-statistic (excluded instruments)	33.22	44.12
Hansen J-statistic	0.969	0.325
p-value for J-statistic	0.325	0.568

[39] It shall be noted that the municipalities are only dropped as dependent variables, i.e. they are kept when we compute the neighbors' benefit levels and characteristics. We do this to minimize the distortion of the exclusion of the observations of the big city counties.

The last column in Table 7 shows that excluding the counties that are the main "migration magnets" does not substantially change the results. The coefficient of the neighbors' benefit level is significant and, in magnitude, very close to the coefficient estimate obtained in the baseline specification (c.f. the first column in Table 7). As expected, the Hansen J-statistic is smaller (the p-value is higher) when the three migration magnets are excluded. We conclude that the baseline results do not seem to be affected by the migration of the refugees.

6.4 Some additional sensitivity analyses

In addition to the above discussed sensitivity analyses, we have checked the robustness of the baseline estimates to: (i) another functional form (using a semi-logarithmic specification), (ii) the inclusion of a lagged dependent variable on the right-hand side (estimating the model in first differences), (iii) the inclusion of lagged covariates (but no lag on the dependent variable). In none of these three cases do we reach other conclusions than those of the baseline analysis. We have also examined how sensitive the baseline results are to an alternative normalization of the refugees variable: Normalizing the number of refugees with the number of individuals on welfare in the municipality instead of normalizing it with the municipality's entire population yields almost identical results as in the baseline case.

7 Conclusions

In this paper we investigate whether local governments react on the welfare benefit levels in neighboring jurisdictions when setting their own benefit levels. The main contribution of the paper is that we suggest and use a specific policy intervention as an instrument to solve the simultaneity problem that arises from the welfare game that the local governments play; a centrally geared placement of a highly welfare prone group (refugees) among Swedish municipalities.

We argue in the paper that given that one controls for the number of vacant apartments in the municipality and for municipality-specific fixed effects, the refugee placement can be considered as exogenous. Furthermore, we show that it is theoretically appropriate to use the refugee placement program as an instrument since the refugee placement program shifts the neighboring local governments' reaction function while holding the reaction function of my local government fixed (implying that the effect on my benefit level is only channeled trough the benefit level of the neighboring jurisdictions).

In the empirical application we use panel data for Swedish municipalities. In addition to controlling for observable characteristics of the municipalities, we control for both municipality-specific and time-specific fixed effects. In the baseline analysis,

we find a significant and positive effect from the setting of welfare benefit levels in neighboring municipalities on the setting of the welfare benefit level in a given municipality. The point estimates indicate that if the neighboring municipalities decrease their welfare benefit level with 100 SEK, a municipality decreases its benefit level with approximately 41 SEK. The estimates for neighbors' benefit level hence provide indications of strategic interactions among the local governments in the setting of welfare benefit levels, implying that there exists a "race-to-the-bottom". These results seem to be robust to several alternative model specifications.

The policy intervention that we suggest and use as an instrument in this paper is not unique for Sweden. Similar programs exist in other countries, and we believe that the use of such programs can be a fruitful way of approaching the problem encountered in models of welfare competition.

Acknowledgements

We thank Per-Anders Edin, Peter Fredriksson, Ann-Sofie Kolm, Eva Mörk, Torsten Persson, Jørn Rattsø, David Strömberg, Olof Åslund, three anonymous referees, and seminar participants at the 2004 Econometric Society Summer Meeting in Providence, at the 2004 Villa Mondragone Workshop in Rome, at the 2004 CEPR Workshop on Public Policy in Stockholm, at the 2004 EEA Meeting in Madrid, at the 2004 IIPF Meeting in Milano, at the 2005 AEA Meeting in Philadelphia, at the 2005 Econometric Society Winter Meeting in Philadelphia, at Uppsala University, at Stocholm School of Economics, at Umeå University and at IIES at Stockholm University for valuable comments and discussions. We are grateful for financial support from the Swedish Research Council.

References

Åslund, O.: 2000, Immigrant settlement policies and subsequent migration. Working Paper 2000:23, Department of Economics, Uppsala University.

Bertrand, M., Duflo, E. and Mullainathan, S.: 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics* pp. 249–275.

Besley, T. and Case, A.: 1995, Incumbent behavior: Vote seeking, tax setting and yardstick competition, *American Economic Review* pp. 25–45.

Brown, C. and Oates, W.: 1987, Assistance to the poor in a federal system, *Journal of Public Economics* pp. 307–330.

Brueckner, J.: 2000, Welfare reform and the race to the bottom: Theory and evidence, *Southern Economic Journal* pp. 505–525.

Edin, P.-A., Fredriksson, P. and Åslund, O.: 2003, Ethnic enclaves and the economic success of immigrants - evidence from a natural experiment, *The Quarterly Journal of Economics* pp. 329–357.

Figlio, D., Kolpin, V. and Reid, W.: 1999, Do states play welfare games?, *Journal of Urban Economics* pp. 437–454.

Franzén, E.: 2004, *Egenförsörjning eller bidragsförsörjning? Invandrarna, arbetsmarknaden och välfärdsstaten*, SOU 2004:21, chapter Invandrare och socialbidragstagare - ett liv i ofärd, pp. 103–131.

Gelbach, J. B.: 2004, Migration, the lifecycle, and state benefits: How low is the bottom?, *Journal of Political Economy* **112**(5), 1091–1130.

Hansen, J. and Lofstrom, M.: 1999, Immigrant assimilation and welfare participation: Do immigrants assimilate into or out-of welfare? Discussion Paper No 100, IZA.

Hernes Fiva, J. and Rattsø, J.: 2003, Welfare competition in norway. Department of Economics, Norwegian University of Science and Technology, Trondheim.

Kézdi, G.: 2002, Robust standard error estimation in fixed-effects panel models. Mimeo, University of Michigan.

McKinnish, T.: 2005, Importing the poor: Welfare magnetism and cross-border welfare migration, *Journal of Human Resources* **40**, 57–76.

McKinnish, T.: 2007, Cross-border welfare migration: New evidence from micro-data, *Journal of Public Economics* **91**, 437–450.

Meyer, B.: 2000, Do the poor move to receive higher welfare benefits? Unpublished Paper, Department of Economics, Northwestern University.

Saavedra, L.: 2000, A model of welfare competition with evidence from AFCD, *Journal of Urban Economics* pp. 248–279.

Smith, P.: 1991, An empirical investigation of interstate AFDC benefit competition, *Public Choice* pp. 217–233.

Staiger, D. and Stock, J.: 1997, Instrumental variables regressions when the instruments are weak, *Econometrica* pp. 557–586.

The Immigration Board: 1997, *Individuall mångfald: Invandrarverkets utvärdering och analys av det samordnade flyktingmottagandet 1991-96*, The Immigration Board.

The National Board of Health and Welfare: 1999, Ändrad Socialtjänstlag - konsekvenser under de 18 första månaderna.

Wheaton, W.: 2000, Decentralized welfare provision: Will there be underprovision?, *Journal of Urban Economics* pp. 536–555.

Wildasin, D.: 1991, Income redistribution in a common labor market, *American Economic Review* pp. 757–754.

Essay 2

Strategic Competition in Swedish Local Spending on Childcare, Schooling and Care for the Elderly

1 Introduction

In a world where information flows and people move between regions, local policy makers do not make their decisions in isolation, but need to consider the influence of the surrounding local governments' policies. This gives rise to a situation where the local decision making is affected not only by the situation in the own jurisdiction, but also by the other jurisdictions' policy decisions.

The economic literature distinguishes between two types of strategic interaction: interaction in the form of competition for a mobile resource, and interaction based on information spill-over.[1] The first of these theories recognizes that if local residents respond to differences in local policy by moving, then local policy makers may want to adjust the local policy decision in order to attract - or avoid to attract - certain residents to the jurisdiction.[2]

In the second, information-based, theory, interaction stems from the hypothesis that the voters of a jurisdiction evaluate the performance of the local policy makers by comparison with the surrounding jurisdictions. This in turn may induce the local policy maker to mimic the neighbours' policy, in order not to look bad in the comparison and be voted out of office. The idea is that the neighbours provide a yardstick against which the voters evaluate the decisions made by the local policy maker, and the model is hence referred to as the "yardstick competition" model.[3]

Theory hence describes two mechanisms that can give rise to strategic behaviour among local policy makers: the possibility of dissatisfied residents 1) to move to another jurisdiction, or 2) to vote for another politician. In general, the literature on the former, migration-based, theory has focused on competition for a mobile tax base (tax competition), or competition to limit the inflow of costly benefit prone

[1] See e.g. Brueckner (2003) for an overview of the different theoretical models.

[2] See e.g. Wilson (1999) and Wilson and Gordon (2003) for theoretical models.

[3] See Besley and Case (1995) for the first description of the yardstick competition model in the political economy-setting.

individuals (welfare competition).[4] The second theory, yardstick competition, has predominantly been applied to local tax policy [5], although some recent studies also test for yardstick competition in local expenditures.[6]

In this paper I acknowledge that strategic behaviour may arise also in other areas of local policy, namely in the local decision on how much to spend on childcare, primary schooling and care for the elderly. In Sweden, childcare has long been a local responsibility, and in 1991-92 a series of reforms transferred the provision and financing for primary schooling and care for the elderly from the national and county levels to the municipal level.

Is the decision on how much to spend on these services likely to be affected by the threat of residents to either move from the jurisdiction or to vote the incumbent out of office? I argue that there is reason for us to believe that it might.

Let us first consider the case of competition for mobile residents. Is it likely that the local spending policy for childcare, primary schooling and care for the elderly is affected by strategic competition for residents between local governments? This naturally hinges on the assumption that there is Tiebout-migration in the sense that individuals tend to move to municipalities with high quality public service - or at least that the local policy makers believe that this is the case. There is some evidence of Tiebout-type migration in Sweden: Dahlberg and Fredriksson (2001) find a positive relationship between local public service quality and the residential choices of short-distance migrants.

The fact that the services in this study, childcare, schooling and care for the elderly do not benefit all residents, but are targeted to families with children and elderly respectively[7], furthermore means that there is scope for the local policy maker to use public service spending to attract certain demographic groups to the jurisdiction. A jurisdiction that wishes to attract more families and fewer elderly residents, may hence try be tempted to favor spending on childcare and schooling on the expense of care for the elderly, and vice versa. A local policy maker may hence use public service spending as a means to attract the desired population mix; by allocating more (than the neighbours) to the services targeted to the desirable population group, and less (than the neighbours) to the less desirable group.[8]

[4] See Brueckner (2000) and Allers and Elhorst (2005) for results of the empirical literature.

[5] See e.g. Besley and Case (1995), Bordignon, Cerniglia and Revelli (2003) and Solé-Ollé (2003).

[6] See e.g. Revelli (2006), who finds evidence of yardstick competition in the social service provision of UK local authorities.

[7] Naturally, other residents may also enjoy indirect utility of these services, however, the direct effects apply only to the users of the services.

[8] There are several reasons for why the demographic mix could matter to the local decision maker: the young and the old may differ in the income level, and hence the income tax base they provide, and they may incur different types of costs on the jurisdiction. Local labour market concerns is another potential reason.

How about the second theory - strategic interaction based on the yardstick comparison by voters? Is this type of interaction likely to be present in the services of the study? There are some factors that speak for this: Childcare, schooling and care for the elderly are services that are important and visible to a large number of the residents of a jurisdiction. They also constitute the lion's share of the municipal budget. This suggests that these services may be important in the voting decision of residents. In addition, residents are likely to be informed about the quality of the services in the own as well as in adjacent jurisdictions, which is another important prerequisite for yardstick competition. It is hence motivated to test for yardstick type interaction among local governments. In particular, we assume that the voters in a jurisdiction observe the quality of childcare, schooling and care for the elderly that they get, given the tax rate, compared to other jurisdictions, and use this comparison to evaluate whether the local policy maker does a good job or not. This will be noted by the politician, who will avoid to deviate too much from the neighbours' decisions, in order not to be punished in the coming election.

Based on the above hypotheses, this study will test for a spatial pattern in municipal spending policy on childcare, primary schooling and care for the elderly. In the baseline analysis, I will test for a spatial pattern, consistent with strategic interactions, among jurisdictions that share border. As will be discussed later, this is a simple and straightforward measure that can be motivated from both theories. As a sensitivity analysis I also use a set of neighbourhood definitions that are closely related to the respective theories, i.e. competition for mobile residents and yardstick competition.

I will test for strategic interactions in the composite expenditure policy of local governments, i.e. I allow for interaction to take place both in expenditures on the same service category, and in expenditures on different categories of services. This makes sense if residents/voters care about the allocation of resources between different services, as well as how much is spent on each category.[9] Furthermore, while the previous literature in general tests for strategic interaction in one type of expenditure, or uses aggregate expenditures, here, I test for interactions in the three main expenditure items of the municipalities.[10]

The hypothesis that the local decision maker reacts on the spending policy of the neighbouring jurisdictions is tested using data on Swedish municipal spending on childcare, primary education and care for the elderly over the period 1996-2005. I will use spending per potential user, defined as spending per individual aged 0-15 for

[9] Two previous studies estimate strategic interactions in composite local policies: the first, Fredriksson, List and Millimet (2004), focuses at U.S. state policies to attract firms to the locality, and the second, Millimet and Rangaprasad (2007), looks at U.S. school district inputs.

[10] For previous studies, see e.g. Case, Hines and Rosen (1993), Baicker (2005), Redoano (2003), Schaltegger and Zemp (2003), and Solé-Ollé (2006).

childcare and education[11]; and spending per individual aged 80 and older for care for the elderly, as a measure of quality. While it is true that increased spending does not necessarily imply higher quality, the idea here is that a politician who wants to increase the quality of a service, will probably do so by allocating more resources to the service; i.e. by increasing the spending per potential user. In addition, finding alternative and observable measures of quality is not trivial, especially for care for the elderly.

There is no Swedish study on strategic interactions in the municipal expenditures that are analyzed in this study. There are however studies that test for interactions in other expenditures. Hanes (2002) uses cross-sectional data for 1986 on the local rescue services of Swedish municipalities, and finds a negative spatial pattern, consistent with free-riding. Lundberg (2001) tests a similar hypothesis for municipal spending on recreational and cultural services over 1981-1990, and also finds support for the free-riding hypothesis. Dahlberg and Edmark (2004) find evidence of a positive spatial pattern in the welfare benefit levels of the municipalities, using a panel of 283 municipalities over 1990-1994, which is consistent with welfare competition. Finally, Aronsson, Lundberg and Wikström (2000) find evidence of vertical externalities between the county and the municipal expenditures, using Swedish panel data over 1981-86. This suggests that it is important to consider potential effects of county spending when estimating interactions between municipalities.

Identification and estimation problems abound in studies of this type. The fact that interaction is simultaneous - i.e. my neighbours' spending decision affects my decision, which in turn affects theirs and so on - invalidates the use of OLS. In this study, following Kelejian and Prucha (1998), I use instrumental variables estimation to overcome this problem. As shown by Kelejian and Prucha (1998) IV has the advantage of being unbiased in the presence of spatial error correlation. I include a set of municipality characteristics, as well as time and fixed effects to further reduce the risk of bias due to spatial error correlation. Finally, I account for dynamics by clustering on municipality.

The analysis is subject to the following sensitivity tests: First, as mentioned above, a set of alternative neighbourhood specifications is used. Second, the possibility of vertical interactions is accounted for through testing for effects of county expenditure on municipal spending policy. Third, a Cochrane-Orcutt-type transformation of the variables, suggested by Kelejian and Prucha (1998), is performed. The idea is that this can increase the efficiency of the estimations.

The results give no clear support for a spatial pattern in the local policy on childcare, primary education and care for the elderly. While there are some significant

[11] Adding spending for childcare and schooling to one category makes sense since both services are targeted to children. In addition, doing so facilitated the estimations, as discussed in section 3.

coefficients, especially in the regression on spending on care for the elderly, the results are not robust enough to draw any conclusions. Using the alternative neighbourhood definitions yielded no additional support for neither competition for mobile residents nor yardstick competition.

The disposition of the remaining study is as follows: section 2 describes the Swedish local public sector and section 3 the data used. Section 4 discusses the empirical specification and methodology, and section 5 presents the results. Finally, section 6 concludes.

2 The Swedish local public sector

The Swedish public sector is organized at three levels: municipal, county and central level. There are 290 municipalities and 21 counties. The main responsibility of the counties is the provision of health care. The municipalities have traditionally been responsible for a vast range of public services, such as social assistance, infrastructure and environmental regulation.[12] After the decentralization reforms in the early 1990s, the main responsibilities of the municipalities are in the areas of education, child care and care for the elderly.

An important prerequisite for strategic interaction to arise in these services, is that the municipalities can in fact affect the quality of the services. While there are national guidelines for the municipal provision of childcare, schooling and care for the elderly, there is also significant room for local decision making. The guidelines are most detailed when it comes to primary schooling, where national regulation[13] specifies the comprehensive goals and guiding principles, and provides the basic curricula and the minimum hours of teaching. Within this framework, there is room for the municipality to prepare an own plan for the practical organization and resource allocation. A quick look at the data on the resource allocation in the municipalities in 2005, shows important differences in for example the teacher density and expenses for teaching material.[14]

The national regulations for childcare and care for the elderly provide very general guidelines for the municipalities[15], and there is no national system for the control of the compliance with these. In the case of childcare, the municipalities are themselves

[12] Two municipalities, Malmö and Gothenburg, differ from the rest in that they were responsible for some of the services elsewhere provided by the counties until 1998-99. They are kept in the data, since excluding them did not change the results.

[13] See law 1985:1100 (Skollagen), regulation 1994:1194 (Grundskoleförordningen), and the National plan for education (Nationell skolplan Lpo 94).

[14] Per student expenses for teaching materials varies between SEK1000 (about $140) and SEK5000 (about $700), and the average number of students per teacher varies between 7 and 11.

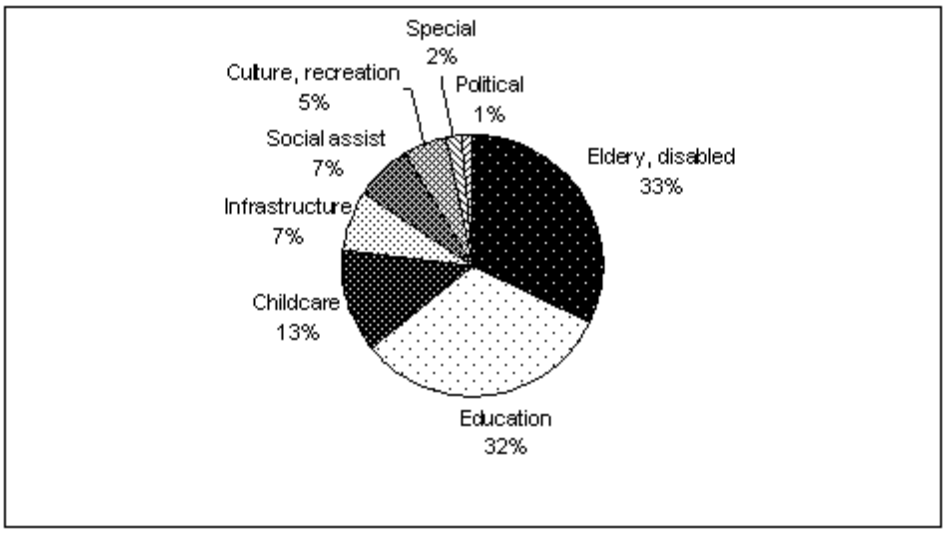
[15] For childcare see law 1985:1100 (Skollagen), and for care for the elderly, see law 2001:453 (Socialtjänstlagen).

responsible for controlling that the guidelines are fulfilled.

The local decision power is considerable also on the revenue side. The municipalities have the right to collect tax revenue in the form of a local income tax and are free to set the tax level, given that they maintain a balanced budget. The tax revenues account for around 70 percent of the total municipal revenue - the rest is made up by central government grants and user fees[16]. Until 1992 the central government grants were targeted to specific services, but since 1993 they are in general in the form of general grants that can be used freely by the municipalities.

The fact that the municipalities are responsible for both the financing and provision of a number of important services, makes Sweden a particularly interesting case for the study of spatial interactions in the policies of local governments. As is illustrated in Figure 1, spending on childcare, primary education and care for the elderly and disabled account for the main part of the municipal budget.[17] This means that the citizens and the politicians are likely to have information about the cost and quality of these services and are likely to care about the cost and quality, which are important prerequisites for the hypothesis of this study.

Figure 1: Average per Capita Municipal Spending in 2003



Note: The Figure shows the distribution of the total municipal expenditures on different spending categories, given as the municipal average per user in 2003. Source: Statistics Sweden.

As was mentioned in the previous section, an assumption for the hypothesis of migration-driven strategic competition in spending on childcare, primary education and care for the elderly, is that the demographic mix of a municipality matters economically for the local policy maker. In Sweden, however, as in many other countries, there is a system of equalization of the taxbase and of the structural costs of the

[16] This figure is from 2002, see *Kommunernas Ekonomiska Läge* (2003).

[17] It shall be noted, though, that education in Figure (1) also includes spending on secondary and adult education.

municipalities. The aim of the system is to give every municipality roughly equal conditions in structural factors such as demography, climate etc. Needless to say, this decreases the incentives for migration-based strategic competition. However, Dahlberg and Edmark (2004) find evidence of welfare competition, and Ågren and Edmark (2005) find evidence of tax competition among Swedish municipalities, using data from the same period as this study. This suggests that the equalization system may not totally eliminate the incentives for strategic behavior of this type.

Finally, Revelli (2006) argues that in a multi-tiered government structure one should consider not only horizontal (between municipalities), but also vertical (between municipalities and other levels of government) interactions. In our setting, this means that it is potentially important to include county spending in the regression equation. I will therefore also, as a robustness test, include this variable in the regression. This is furthermore motivated by the fact that Aronsson et al. (2000) find vertical externalities to be present using Swedish data during 1981-86.

3 Data

The data set of this study is a panel of 283 municipalities[18] over 1996-2005.[19][20] As stated above, I use the following variables on local public expenditures: spending on childcare, primary education, and care for the elderly. I focus on spending per potential user, and define spending on childcare and education as one category, since both of these services are targeted to children.[21] The number of potential users is defined as the number of individuals aged 0-15 for childcare and education, and as

[18] 6 of the 290 municipalities have either merged with or seceded from another municipality during the time period under study, and have hence been excluded from the sample. In addition, the municipality of Gotland has been excluded since it is an island for which it is naturally difficult to define the set of neighbors.

[19] The data on spending on childcare and care of the elderly and disabled, as well as the data on most explanatory variables, is collected from Statistics Sweden. The exception is data on unemployment, which is from the Swedish Public Employment Service (Arbetsmarknadsstyrelsen). Data on spending on primary schooling is from the Swedish National Agency for Education (Skolverket), and data on county expenditures is from The Swedish Association of Local Authorities and Regions (Sveriges Kommuner och Landsting).

[20] Using data before that period is restricted first for two reasons: First, a large part of the provision of the services in the study were not provided by the municipalities before the first years of the 1990. Second, the collection of data on primary school spending changed in 1995, which means that data from the early years of the 1990s are not comparable to the more recent years.

[21] An alternative would be to have two separate categories for childcare and primary schooling. However, when doing so I encountered problems related to weak instruments. That is, when separating spending on childcare and schooling, the set of instruments were not strong enough to separatedly identify the two first stage regressions. This suggests that a large share of the variation in the instrumet set is common for the two types of services, and that it is in this sense appropriate to estimate them together.

the number if individuals aged 80 and older for care for the elderly (and disabled). The data on municipal spending does not separate between spending on elderly and disabled, and thus also includes spending on disabled.

The analysis includes a large set of municipality-level covariates. In order to control for differences in basic economic conditions, I include the per capita municipal taxbase (taxable income), per capita central government grants[22], per capita long-term debt, unemployment, employment, and the share of the population on welfare benefits (denoted welfare in Table 1), as well as per capita county expenditures. A dummy variable, which takes the value one if the political majority is left-wing, is added to the regression in order to capture political preferences[23], and the log of the population size is included in order to capture differences in returns to scale. All covariates, except for the political dummy variable, are lagged one time period. This makes sense since the local budget is decided towards the end of the previous year, when the information available concerns the previous years' economic and demographic conditions. Finally, as suggested in the previous section, I will also, as a robustness test, add county spending as a covariate in the regressions in order to account for possible vertical interactions between county and municipal expenditures.

I also control for unobserved municipality factors that stay fixed over time by including municipality fixed effects. This is important in order to control for factors such as the size of the municipality and climate, which affect the cost of service provision.[24] In addition, the analysis includes year dummy variables.

Table 1 gives the average values for the variables over the period 1996-2005. All pecuniary variables are deflated to year 2002 monetary value.

[22] The grants variable is made up by the sum of total grants, i.e. both equalizing grants (equalizing the economic conditions across municipalities) and general grants. The negative minimum value of this variable in Table 1 is due to the fact that some municipalities end up as net payers when the equalizing grants are taken into account.

[23] We define the Left Party and the Social Democratic Party as left-wing parties.

[24] As is seen in Table 1 there are very large differences between the min and max values in spending per potential user in the cases of both childcare and education, and care for the elderly. This suggests that controlling for fixed municipality effects may be important.

Table 1: Descriptive Statistics 1996-2005

Variable	Obs	Mean	Std.Dev.	Min	Max
Spending Childcare Education	2793	59030	7586	39582	92326
Spending Care Elderly	2825	230338	45398	109790	476036
Taxbase	2830	1110	184	740	2509
Grants	2830	8018	4432	-15399	23194
Long Term Debt	2775	10282	10118	0	73482
Unemployment (%)	2830	4.6	1.9	0.9	13.8
Employment (%)	2830	44.2	3.5	29.4	54.2
Welfare (%)	2820	5.2	2.2	0.42	16.3
Population	2830	31142	58511	2553	771038
Left	2830	0.4	0.5	0	1
County Spending	2532*	18225	2626	12445	23868

*County spending only contains data for 1996-2004.

4 Empirical specification

The prediction to be tested in the empirical analysis is, as described in section 1, that the own spending policy on childcare and primary education, and on care for the elderly, is a function of the neighbouring municipalities' spending policy. Assuming linearity, the prediction can be described by the following regression equation system[25]:

$$s_t^k = \rho_e^k W s_t^e + \rho_c^k W s_t^c + X_{t-1}\beta + \epsilon_t, \qquad k = c, e. \tag{1}$$

In terms of notation, s_t^k is a vector of the per user spending on category k in period t, where c denotes childcare and education, and e care for the elderly. W is a matrix that gives positive weight to the municipalities that are defined as neighbours, i.e. a neighbour weight matrix (W is time-invariant in all specifications). Ws_t^e and Ws_t^c hence give the average of the neighbouring municipalities' spending on care for the elderly, and childcare and education, respectively. X_{t-1} is a matrix of municipality characteristics that affect the spending policy and also includes a constant term (since all municipality covariates contained in X, except for the political dummy variable, are lagged, I use the subscript $t-1$).

The hypothesis that will be tested in the empirical section is that the ρ-coefficients differ from zero, i.e. a non-zero result is consistent with the hypothesis of strategic interactions in local service spending. What can we expect regarding the signs of the coefficients? In a case with only one policy instrument, we would in general expect

[25] Similar specifications are used in Fredriksson et al. (2004), who model a situation where jurisdicitons compete for companies using a composite policy of local tax rate, environmental standards and local public spending, as well as by Millimet and Rangaprasad (2007) who test for strategic competition among school districts.

to find positive interaction coefficients, provided that all local decision makers have similar preferences.[26] However, in our present case, with two spending categories, the signs of the interaction coefficients are unknown.[27]

Since both equations in the system described in (1) include the same variables, no efficiency gains are to be made by joint estimation. The equations are therefore estimated one by one.

4.1 Definition of a municipality's neighbours

The neighbour weight matrix W needs to be defined ex ante based on exogenous factors. As discussed in the introduction, the causes for strategic interaction in the migration- based theory is the potential migration of the service-consuming residents, whereas in the yardstick competition case it is the threat to be voted out of office that gives rise to interaction. In both of these cases, a prerequisite for interaction to occur is that residents/voters, as well as policy makers, are informed about the policy of other jurisdictions. A reasonable criterion for the definition of neighbours, which is often used in the literature, is hence to let the weight-matrix reflect the geographical proximity of the jurisdictions, since information about service quality and cost is likely to be more easily available for closely situated municipalities.

A simple weight-matrix, which captures these aspects, is to define neighbours as the municipalities that share border. If we use w_{ij} to denote the elements of matrix W, i.e. w_{ij} defines the weight that municipality j has as a neighbour of i, then we can define this weight-matrix as $w_{ij} = 1$ if i and j share border and $w_{ij} = 0$ otherwise. This type of weight matrix is common in the literature on strategic interactions, and has the advantage of being exogenous in the sense that the risk of imposing the spatial pattern that we want to observe, through the definition of the weight matrix, is small.

In addition to this geographical neighbourhood definition, I define two sets of additional weighting schemes, that are closely related to the theoretical frameworks.

First, in order to better capture the information aspect, I construct a neighbour weight matrix that reflects the coverage of local news papers. In this case, we let $w_{ij} = newspaper_{ij} \cdot coverage_{ij}$, where $newspaper_{ij} = 1$ if i and j share a local newspaper, and $coverage_{ij}$ = the sum of average newspaper coverage of the local

[26] I.e., we would expect the local policy maker to mimic the neighbours' policy decision.

[27] Consider for example the situation where the objective of the policy maker is to attract more residents - of any age - to the jurisdiction. Assume also that this can be done either by increasing spending on childcare and education; on care for the elderly; or on both. A neighbour's decision to increase spending on, say childcare and education, can then be met with a strategic decision to increase own spending on either the same or the other (or both) spending categories, and can in this case hence result in interaction coefficients of either positive or negative sign.

newspapers in j and $w_{ij} = 0$ otherwise[28][29].

Second, according to the migration-based theory, it is, naturally, reasonable to assume that interaction takes place among municipalities between which migration is common. I hence let $w_{ij} = migr_{ij}$, where $migr_{ij}$ is the immigration from j to i in 1995. Under this definition, municipality j:s weight as a neighbour to i depends positively on the migration rate. In the first of the two migration based matrices, I use data on migration of all persons aged 16-65. This is intended to capture the overall migration patterns between the municipalities. However, according to our hypothesis, what really matters is the migration of those that are attracted by by good care of children and schooling, or care for elderly. I therefore let the second of the migration based weight matrices be based only on the migration of individuals with children aged 0-15. Unfortunately, we lack data over the migration flows of the elderly, and can hence not incorporate this information in the weighting scheme.[30] By using migration in 1995, which is the year before the first year of our panel, we attempt to avoid endogeneity in the definition of neighbours. Since we expect migration to be affected by the spending policies of the municipalities, it is possible that using migration in later years could give rise to a spurious relation in expenditure levels.

In all cases the weight matrices are row-standardized, i.e. they are normalized so that the individual weights of a set of neighbours sum to one. This facilitates the interpretation of the coefficients, and enables direct comparison of the coefficients from specifications using different weight matrices.

What results do we expect to obtain from the different definitions of neighbours? The use of different weighting schemes shall first and foremost be seen as a robustness test of the results. However, they can also be seen as a first indication of the type of strategic interaction. In particular, this holds for the migration-based matrices: since these correspond to the migration-based model to a higher degree, we expect interaction to be stronger in these specifications if competition for attractive residents is driving interaction. Specifically, if it is true that the municipalities compete for the desired distribution of the young and the old, we expect a stronger result when we use migration of the young to define neighbours.

[28] The data on local newspapers is from 1994, 1998 eller 2002 and is from Tidningsstatistik AB. We are grateful to Helena Svaleryd och Jonas Vlachos for having made it available to us.

[29] This type of weight matrix was also used in Ågren and Edmark (2005). We select all newspapers that are given out at least six days a week. This leaves some municipalities with no newspaper. For these we include newspapers that are given out less then six days a week. There are two newspapers that have a national coverage, Dagens Nyheter and Svenska Dagbladet. These are counted as local newspapers only for the municipalities in the Stockholm county, since they cover local news in this region.

[30] The data on inter-municipal migration comes from the data base LOUISE, and was provided by The Institute for Labor Market Policy Evaluation (IFAU).

4.2 Estimation issues

There are several issues to consider in the estimation of strategic interactions in local spending decisions. In particular, we need to minimize the risk for bias due to the simultaneity of the municipalities' policy decisions, and for bias due to spatial error correlation.

The simultaneity of the policy decision implies that using OLS to estimate equation (1) yields biased estimates (see e.g. Anselin (1988)). An alternative to OLS, which is suggested by Kelejian and Prucha (1998), is to use the neighbours' characteristics to instrument for neighbours' spending. I follow this procedure and use the neighbours' characteristics as instruments, except for the political variable describing whether the municipality is ruled by a left-wing majority. This is excluded from the instrument set since this is likely to be affected by the spending level and hence endogenous. The resulting set of instruments contain the neighbours' values of: the taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.[31] Using the lagged values of the instrumental variables makes sense not only because of the fact that the local budget decision is made towards the end of the previous year, but also since this ensures the exogeneity of the instruments in terms of there being no effect of local spending policy on the instruments.

The spatial error correlation problem can be thought of as an omitted variable problem; i.e. we want to avoid that something that is omitted from the spending equation, and that is correlated among neighbouring municipalities, affects the estimates. According to Kelejian and Prucha (1998), spatial IV regression is consistent also in the presence of spatially correlated error terms. However, in order to further minimize the risk for this type of bias, I add a set of covariates, including fixed effects and year effects. This can also be seen as a measure to strengthen the case for our instruments, since the instruments now only need to be exogenous conditional on the set of covariates. Specifically, the fact that all the variables that are used as instruments are also included as covariates means that the identifying variation that is used in the first stage of the IV-estimation is conditional on the own characteristics, i.e. only the difference between the own and the neighbours' characteristics are used for identification. This rules out any concern that the coefficients for neighbours' spending merely mirror similarities among neighbours in the variables that are used as instruments.[32]

[31] It may seem strange to include both unemployment and emplyment in the estimation, since these are likely to be correlated. However, we are interested in the prediction power of the first stage, and not the individual effects of the instruments, and we include both variables since this improves the prediction power.

[32] See e.g. Figlio, Kolpin and Reid (1999) for a discussion on this.

An alternative to using instrumental variable technique to solve the simultaneity-problem of equation (1) is to use a spatial lag maximum-likelihood estimator (see Revelli (2006) for an overview of spatial ML-models). This estimator will however not be used here, since it can be computationally demanding, especially when the number of jurisdictions is large and when the weight matrix is not symmetric in the sense that the number of neighbours differs between jurisdictions[33]. In addition to the computational burden, it can also be argued that the ML-estimator has less potential to identify the spatial process in the error term separately from spatial error correlation.

Yet another alternative, which is suggested by Fredriksson et al. (2004) and Millimet and Rangaprasad (2007), is to replace neighbours' policy variables with their lagged values. The idea is that this is a simple way to get around the simultaneity problem, since it is not particularly likely that the neighbours' past policy is affected by the own current policy, and that OLS can hence be used to estimate the effects of the neighbours' lagged policy. However, while this solves the simultaneity problem, the estimates are likely to be biased by spatial error correlation if spatial shocks are persistent.

Finally, since there is evidence that the adjustment of municipal expenditures in Sweden is sluggish (see e.g. Dahlberg and Johansson (2000)), I will need to account for dynamics in the regressions. In our setting, this implies that the residuals of equation (2) are likely to be serially correlated. I take account of this by computing standard errors that are robust for serial correlation of arbitrary form in the error term[34][35].

5 Results

This section presents the results of the regression analysis. The estimated equation is obtained by adding jurisdiction-specific fixed effects, id, and a set of yearly dummy variables, $year$, to equation (1):

$$s_t^k = \rho_e^k W s_t^e + \rho_c^k W s_t^c + X_{t-1}\beta + id + year + \epsilon_t, \qquad k = c, e. \tag{2}$$

where k denotes the two different spending categories that are included in the analysis: childcare and primary education, and care for the elderly. Following the predictions of

[33] When Kelejian and Prucha (1999) test the accuracy and time of spatial ML-computation they encounter problems when the number of cross-sectional units is 400, even though they use a symmetric weight matrix.

[34] The error covariance matrix is obtained by clustering on municipality (see Baum, Schaffer and Stillman (2003)).

[35] An alternative would be to include the lagged dependent variable in the estimations, using an Anderson-Hsiao-type estimator. This would however mean that we would lose observations from the early period of our data set, since these would be used as instruments.

the theoretical set-up, I will estimate two separate regressions, one for each spending category k, and will include the neighbours' spending for both categories as explanatory variables in all regressions.

The testable hypothesis of the theoretical set-up is that the ρ^k-coefficients differ from zero. In addition, they shall not exceed one in absolute value, since a larger interaction coefficient does not represent a stable interaction process[36].

As described in the previous section, I use the neighbours' values of the following variables to instruments for neighbours' spending: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period. The same set of instruments is used in all regressions.

5.1 Baseline regression

We start by looking at the results when using the simplest of our neighbourhood definitions, i.e. sharing border. The regressions include all municipality variables, but not county expenditures. Table 2 shows the results from the IV-estimation of equation (2). For the sake of comparison, the OLS-results are also given, although these, as discussed in section 4, are not unbiased. The results for spending on childcare and primary education are given in columns 1-2 and the results for spending on care for the elderly in columns 3-4. The coefficients for neighbours' spending per user are denoted N Childcare and Education and N Care Elderly.

[36] This restriction applies to all row-standardized neighbour weight matrices, but not to weight matrices that are not row-standardized (Anselin (1988)). Note that this restriction is not imposed on the estimations.

Table 2: Baseline regression, Border-based weight matrix

	Childcare and Education		**Care Elderly**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
N Childcare and Education	.07	.078	-.225	-.981
	(.061)	(.185)	(.284)	(.986)
N Care Elderly	-.009	.021	.249***	.677**
	(.013)	(.055)	(.073)	(.297)
Debt $t-1$	-.038*	-.041*	-.008	-.031
	(.021)	(.022)	(.124)	(.13)
Taxbase $t-1$	22.167***	20.881***	40.667	21.486
	(4.703)	(5.27)	(29.136)	(31.026)
Grants $t-1$	.998***	.968***	.91	.634
	(.142)	(.152)	(.833)	(.881)
Unemployment $t-1$	29.956	28.774	-104.71	116.822
	(159.012)	(160.997)	(794.325)	(843.852)
Employment $t-1$	315.004**	324.422**	876.471	1297.219
	(139.874)	(157.642)	(729.715)	(821.279)
Welfare Recipients $t-1$	-190.463*	-177.122*	-1105.904	-959.298
	(99.867)	(103.528)	(711.256)	(703.004)
Ln Population $t-1$	-16267.89***	-18063.58***	33459.34	-32.379
	(5036.767)	(6184.495)	(24056.27)	(31285.21)
Left	430.186	419.44	-403.919	-804.928
	(351.034)	(361.573)	(2415.638)	(2430.278)
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F		15.04		15.02
J-statistic		7.853		3.728
p-value J-statistic		0.165		0.589
Obs.	2715	2715	2746	2746

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level**,** respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

We start by looking at the IV-estimates. The signs of the coefficients for neighbours' spending are insignificant and close to zero for the regression on spending on childcare and education. The corresponding coefficients in the regression on spending on care for the elderly, are larger: A negative coefficient is estimated for neighbours' spending on childcare and education, while a positive coefficient is given for spending on care for the elderly, which suggests that the municipalities respond to changes in neighbours' spending mix with the same type of policy change. Only the latter of the coefficients is however significantly different from zero.

The first stage results are given in Table A.1, Appendix. They show that all instruments are individually significant in the regression on neighbours' spending on childcare and education, and all instruments but employment are individually

significant in the regression on neighbours' spending on care for the elderly.

The common way of testing for instrument relevance, using the F-statistic of the joint significance of the instruments in the first stage regression, is not valid when there are multiple endogenous regressors. (see e.g. Baum et al. (2003) for a description of the problem). Instead, we need to use other tests to judge whether the instrument set is relevant. Baum et al. (2003) suggests a comparison of the partial R^2 and the Shea partial R^2[37] for the instruments. This is not a formal test, but, as a rule of thumb, a large partial R^2 and a small Shea partial R^2 shall make us suspicious that the instruments are lacking sufficient prediction power to explain all the endogenous variables. This is not the case in the regressions of Table 2, where the two measures are identical down to the fourth decimal: 0.1467 for neighbours' spending on childcare and education and 0.0413 for neighbours' spending on care for the elderly.

Another test for instrument relevance is the Cragg-Donald F-statistic. This is originally a test of underidentification, but can also be used for testing for weak instruments by using the critical values computed by Stock and Yogo (2002). It shall be noted, however, that this test statistic and the related critical values are derived under the assumption of homoscedasticity, and it is not clear how well it performs when this assumption is not fulfilled. As can be seen in Table 2, the Cragg-Donald F-statistic for the baseline regression is 15[38]. This is above the critical value[39] and hence rejects the hypothesis of weak instruments.

In addition to being relevant, the instruments need to be exogenous in the sense that there shall be no direct effect of the instruments on the dependent variable, other than through their effect on the endogenous variable. The test of overidentifying restrictions, which is usually used as a test of instrument validity, does not reject the hypothesis of exogenous instruments (see the Hansen J-statistic in the table). Note however, that the validity of the full set of instruments cannot be tested, since the test of overidentifying restrictions ex ante assumes that one of the instruments is valid.

We then turn to comparing the IV-estimates with the OLS-results. How do we expect these to differ? While the simultaneity problem suggests that the OLS-coefficients will be biased upwards, in absolute value, the OLS-coefficients may also suffer from bias due to spatial error correlation, which can be positive or negative

[37] This is a partial R^2-measure which takes the intercorrelation between the instruments into account, see Shea (1997).

[38] 15.02 or 15.04, as can be seen in Table 2. The difference is due to the fact that the number of observations differs somewhat between the regressions on childcare and education and care of the elderly, and that this also affects the computation of teststatistic as I use the ivreg command in Stata.

[39] The critical value for two endogenous variables, allowing for a maximum relative bias of 10% compared to OLS, and at the 5% significance level, is 8.78. According to Stock and Yogo (2002), this value is comparable to the Staiger and Stock (1997) rule of thumb of 10 for the F-statistic in a regression with one endogenous variable.

depending on the sign of the correlation. The relation between OLS and IV hence depends on the relation between these sources of bias. Comparing the OLS- and the IV-coefficients of the interaction variables, we see that the OLS-estimates are in general smaller in absolute value than the IV-counterparts. This could be due to negative spatial error correlation. It shall however be noted that the 95%-confidence intervals for the IV-estimates in most cases well cover the OLS-coefficients.

Another interesting comparison can be made if we run the IV-regression excluding the municipality-fixed effects. The results from this specification, that are given in Table A.2, Appendix, are highly unrealistic in terms of measuring strategic interactions. The coefficient for neighbours' spending on childcare and education, in the specification in column 4, is much larger than one, which suggests that the coefficient is picking up some effect other than strategic interaction. This suggests that municipality-fixed effects may be needed to control for spatially correlated variables that stay fixed over time and that are correlated with the instrumental variables. The results furthermore indicate that the inclusion of fixed effects are important for the validity and relevance of the instruments; without fixed effects the test of overidentifying restrictions rejects the hypothesis of instrument exogeneity in the regression on spending on childcare and education. Furthermore, comparison of the Shea R^2 and partial R^2 indicates weak instruments, which suggests that identification becomes significantly weaker as fixed effects are excluded. Using deviations over time as identifying variation is therefore the proper approach.

5.2 Sensitivity analysis

5.2.1 Varying the neighbourhood definition

The regressions using the border-based definition of neighbours yielded support for an effect of neighbours' spending policy on own spending on care for the elderly, but no effect on spending on childcare and education. Are these results robust to varying the way we define neighbours? In order to test this we re-estimate equation (2) using the alternative definitions of neighbours that were described in section 4. The results for the media-based weight matrix, *Wmedia*, is given in column 2. The results for the weight-matrix based on migration of all persons aged 16-65, *Wmigr*, are shown in column 3 in Tables 3 and 4, and the results when using only migration of persons with children aged 0-15, *Wmigr015*, are shown in column 4. The results from the border-based specification, *Wborder*, are repeated in column 1 of the tables for ease of comparison.[40]

[40] Note that the instruments - i.e. the neighbours' covariates - are also weighted according to the different neighbourhood weight matrices.

Table 3: IV regression, Different neighbour weight matrices

	Childcare and Education			
	Wborder	*Wmedia*	*Wmigr*	*Wmigr015*
	(1)	(2)	(3)	(4)
N Childcare and Education	.078	-.029	.075	.013
	(.185)	(.185)	(.236)	(.223)
N Care Elderly	.021	-.074	.135	.13
	(.055)	(.061)	(.097)	(.098)
Debt $t-1$	-.041*	-.041**	-.04*	-.039*
	(.022)	(.02)	(.023)	(.023)
Taxbase $t-1$	20.881***	23.722***	18.307***	18.147***
	(5.27)	(4.925)	(5.361)	(5.156)
Grants $t-1$	.968***	1.003***	.917***	.934***
	(.152)	(.142)	(.161)	(.158)
Unemployment $t-1$	28.774	15.625	37.76	39.333
	(160.997)	(160.971)	(165.008)	(165.261)
Employment $t-1$	324.422**	331.431**	331.751**	333.674**
	(157.642)	(150.196)	(151.887)	(155.031)
Welfare Recipients $t-1$	-177.122*	-166.597	-186.649*	-183.599*
	(103.528)	(101.443)	(105.686)	(104.667)
Ln Population $t-1$	-18063.58***	-15484.34***	-22281.44***	-21739.74***
	(6184.495)	(5391.413)	(6118.678)	(6007.824)
Left	419.44	304.293	462.708	483.587
	(361.573)	(365.297)	(401.842)	(404.857)
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F	15.04	12.97	17.82	14.42
J-statistic	7.853	8.148	10.376	9.203
p-value J-statistic	0.165	0.148	0.065	0.101
Obs.	2715	2705	2715	2715

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level**,** respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

Table 4: IV regression, Different neighbour weight matrices

	Care Elderly			
	Wborder	*Wmedia*	*Wmigr*	*Wmigr015*
	(1)	(2)	(3)	(4)
N Childcare and Education	-.981	-.396	-2.172*	-1.849
	(.986)	(1.086)	(1.252)	(1.218)
N Care Elderly	.677**	.598	.699	.63
	(.297)	(.379)	(.437)	(.467)
Debt $t-1$	-.031	.053	-.005	-.006
	(.13)	(.124)	(.126)	(.125)
Taxbase $t-1$	21.486	35.635	32.177	31.227
	(31.026)	(30.967)	(30.311)	(31.621)
Grants $t-1$	.634	1.198	1.295	1.292
	(.881)	(.8)	(.849)	(.912)
Unemployment $t-1$	116.822	227.982	-42.655	22.313
	(843.852)	(819.976)	(776.779)	(779.311)
Employment $t-1$	1297.219	963.592	1084.145	1059.052
	(821.279)	(757.501)	(704.331)	(725.697)
Welfare Recipients $t-1$	-959.298	-1385.383*	-1102.433*	-1124.255*
	(703.004)	(708.034)	(650.527)	(674.115)
Ln Population $t-1$	-32.379	28137.48	27278.5	31197.52
	(31285.21)	(27332.92)	(24993.13)	(23924.9)
Left	-804.928	65.358	406.114	395.115
	(2430.278)	(2587.703)	(2385.284)	(2429.104)
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F	15.02	12.97	17.82	14.42
J-statistic	3.728	2.490	2.055	6.164
p-value J-statistic	0.589	0.778	0.842	0.291
Obs.	2746	2705	2715	2715

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

Comparing the results from the different specifications in Table 3, we see that the media- and the migration-based neighbourhood specifications yield results that are qualitatively similar to the border-based specification in the regression on spending on childcare and education: The effect of neighbours' spending policy is insignificant for both categories of spending irrespective of the definition of neighbourhood.

For the regression on spending on care for the elderly and disabled, in Table 4, the coefficient on neighbours' spending on care for the elderly turns insignificant as the alternative neighbourhood specifications are used. The coefficient on neighbours' spending on childcare and education is negative as in the border-based specification, but becomes unreasonably large, over one in absolute value, for the migration-based specifications. This is however only significant in one of the specifications, Wmigra-

tion, and then only at the 10 percent level. The results in Table 3 and 4 hence give no additional support for the theories of strategic interactions.

Regarding the validity of the instruments, the Hansen J-statistic supports the exogeneity of the instruments in all specifications, except for the migration-based specification in column 3, Table 3, when spending on childcare and education is the dependent variable. The relevance of the instruments is supported for all specifications (the Cragg-Donald F-statistic is above the critical value of 8.78, and the Shea partial R^2 is close to the partial R^2[41]).

5.2.2 Adding county expenditures

So far, we have included only municipality-specific covariates in the regressions. However, Aronsson et al. (2000) find support for the hypothesis that county expenditures and municipal spending are related. The intuition is that services provided at different levels can be either substitutes or complements for the local decision maker. Including county expenditures may therefore be important in order to correctly estimate inter-municipal interactions (see e.g. Revelli (2006)).

In general, the same endogeneity problem applies here as in the case of interactions between municipalities, i.e. if municipality spending also affects the county spending decisions, then county spending will be endogenous, although, since county is the larger unit[42], this should be a smaller problem than in the case of municipality-wise interaction. Since the aim here is merely to test the sensitivity of the results to the inclusion of the variable, we will include county expenditures without accounting for potential endogeneity. It shall however be noted that its coefficient shall not be interpreted as a causal effect.

Tables 5 and 6 show the results including county expenditures for the border-, media- and migration-based weight-matrices. For the sake of brevity, only the coefficients for neighbouring municipalities' spending and county spending are shown. (The results for all covariates are shown in Tables A.3 and A.4, Appendix).

[41] For the two migration-based specifications both the Shea partial R^2 and the partial R^2 are about 0.09 for the first stage on neighbors' spending on childcare and education, and are about 0.05-0.06 for the first stage on neighbors' spending on care of the elderly. The corresponding figures for the media-based specification are around 0.04 and 0.06, respectively.

[42] There are 290 municipalities and 21 counties in Sweden - hence on average about 14 municipalties per county.

Table 5: IV regression, Including county expenditures

	Childcare and Education			
	Wborder	*Wmedia*	*Wmigr*	*Wmigr015*
	(1)	(2)	(3)	(4)
N Childcare and Education	.098	-.017	.035	-.035
	(.212)	(.229)	(.253)	(.236)
N Care Elderly	.02	-.066	.085	.105
	(.058)	(.062)	(.106)	(.105)
County costs	.227	.068	.266	.299
	(.202)	(.218)	(.211)	(.212)
Municipality covariates	yes	yes	yes	yes
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F	12.95	7.40	11.20	9.73
J-statistic	4.619	6.950	7.482	6.676
p-value J-statistic	0.464	0.224	0.187	0.246
Obs.	2420	2411	2420	2420

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level**,** respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

Table 6: IV regression, Including county expenditures

	Care Elderly			
	Wborder	*Wmedia*	*Wmigr*	*Wmigr015*
	(1)	(2)	(3)	(4)
N Childcare and Education	-1.244	-.632	-2.441*	-2.162
	(1.168)	(1.265)	(1.378)	(1.327)
N Care Elderly	.838***	.552	1.001**	.818
	(.316)	(.369)	(.492)	(.519)
County costs	-1.386	-1.293	-1.555	-1.708
	(1.409)	(1.507)	(1.17)	(1.231)
Municipality covariates	yes	yes	yes	yes
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F	12.89	7.40	11.20	9.73
J-statistic	2.993	2.545	3.402	8.450
p-value J-statistic	0.701	0.770	0.638	0.133
Obs.	2451	2411	2420	2420

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level**,** respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

As can be seen in Tables 5 and 6, the results change somewhat when county

expenditures are included. The coefficients on neighbours' spending stay insignificant in all specifications in the regression on spending on childcare and education in Table 5. In the regression on spending on care for the elderly (Table 6), the coefficients are larger, and are over one in many specifications. In the migration-based specification, both coefficients of neighbours' spending are over one in absolute value, and significant at the 10 and 5 percent levels. This is an unreasonable result which suggests that the coefficients may be picking up the effect of some omitted variable.

The coefficient on county spending is positive in the regression on spending on childcare and education, and negative in the regression on care for the elderly, but is insignificant in all specifications.

Although, as commented earlier, the coefficient on county expenditures shall not be interpreted as a causal effect, it is nevertheless interesting to compare result in Table 6, with the findings in Aronsson et al. (2000). They find a positive relation between county and aggregate municipal expenditures, suggesting complementarity, using data over 1981-86. Since that period, the municipal responsibilities for care for the elderly have increased, due to the previously mentioned reform in 1992. An interesting topic for future research would be to test if the sign of the vertical interactions have also changed after this. Guiding from the negative, although insignificant, coefficients in Table 6, one could suspect county expenditures (which mainly consists of medical services) and municipal spending on care for the elderly to be substitutes.

The Cragg-Donald F-statistic and the Shea partial R^2 are very similar to Tables 3 and 4 of the previous section[43], supporting the instrument relevance in all specifications, except for the media-based specification, where the Cragg-Donald F-statistic falls just below the critical value and where weak instruments in this case might be a problem.

5.2.3 Transforming the variables to increase efficiency

The results obtained in the above sections over-all yield very weak evidence for strategic interactions in the spending decision on care for the elderly, childcare and education. Kelejian and Prucha (1998) however suggests that the efficiency of the estimations can be increased by using an alternative estimator, where the variables are transformed in order to take potential spatial error correlation into account. The idea is that in models of spatial interactions, we are likely to experience spatial correlation in the error term due to spatially correlated shocks, and that this correlation contains information that could be utilized in the estimation procedure. This section

[43] The Shea partial R^2 and the partial R^2 are both around 0.12 in the border-based regression on neighbors' spending on childcare and education, and 0.04 in the regression on neighbors' spending on care for the elderly. The corresponding figures for the media-based specification are around 0.03 and 0.06, and for the migration-based specifications around 0.06 and 0.05.

tests if applying this estimation procedure to the data increases the efficiency of the estimations.

The following description follows Kelejian and Prucha (1998) and Kelejian and Prucha (1999). Let us start by defining WX_t^* as the instrument set, and let $H_t = (X_{t-1}, WX_{t-1}^*)$ denote the resulting instrument matrix (that is used in the first stage regressions). Second, I assume that the error term is described by the following process:

$$\epsilon_t = \lambda W \epsilon_t + u_t, \tag{3}$$

where u_t is a vector of independently distributed error terms. That is, the error term of equation (1) is correlated with the error terms of the neighbouring municipalities.[44] The idea is to transform the variables of the second stage taking into account spatial error correlation in the form of (3). Following Kelejian and Prucha (1999), I estimate $\hat{\lambda}$ using non-linear least squares, and use the predicted coefficient to transform the variables in the following manner:

$$\tilde{Z}_t = Z_t - \hat{\lambda} W Z_t, \quad \tilde{s}_t = s_t - \hat{\lambda} W s_t, \tag{4}$$

where $Z = (Ws_t, X_{t-1})$ and s_t denotes service spending.

IV is then applied to the transformed data. The resulting estimator is the following:

$$\left(\tilde{\rho}, \tilde{\beta}'\right)'_{IV} = \left(\tilde{Z}_t' P_t \tilde{Z}_t\right)^{-1} \left(\tilde{Z}_t' P_t \tilde{s}_t\right), \qquad P_t = H_t \left(H_t' H_t\right)^{-1} H_t'. \tag{5}$$

The estimator in equation (5) is applied to the baseline regression, using the border-based neighbourhood criterion. In order to facilitate the estimations, I replace the missing values in the dataset with the municipality-wise mean over the period. Table 7 shows the results for neighbours' spending when the variables are transformed in the above described manner, *(IV transformed)*. For the sake of comparison, the results from using ordinary IV on the same dataset (with no missing values) are also shown.[45] The full set of covariates, are included in the regressions, although here only the coefficients for neighbours' spending policy are shown (the results for all coefficients can be seen in Table A.5, Appendix).

[44] We assume that the weight matrix for the spatial process in the error term is the same as that of the dependent variable.

[45] As can be seen in Table 7, the results are very similar to the results of the unbalanced panel in Table 2.

Table 7: Kelejian and Prucha IV regression, Including county expenditures, Border-based weight matrix

	Childcare and Education		**Care for Elderly**	
	IV	IV transformed	IV	IV transformed
	(1)	(2)	(3)	(4)
N Childcare and Educ	0.049	0.116	-0.851	-0.798
	[0.180]	[0.182]	[0.979]	[0.930]
N Care Elderly	0.015	0.003	0.643***	0.776***
	[0.054]	[0.052]	[0.292]	[0.250]
Municipality covariates	yes	yes	yes	yes
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
$\hat{\lambda}$		-0.392		-0.403
Obs.	2380	2380	2380	2380

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

As can be seen in Table 7, the results of the estimation on the transformed variables are very similar to the results of the regression on the untransformed variables in Table 2. Neighbours' spending has no significant effect on own spending on childcare and education, while neighbours' spending on care for the elderly has a positive significant effect on own spending on the same category. According to the results in Table 7, using the transformation suggested by Kelejian and Prucha (1998) did thus not qualitatively change our results. This is in line with recent Monte Carlo results for the estimator, which suggest that the efficiency-gains to be made from using the estimator are limited in small samples (see Kelejian, Prucha and Yuzefovich (2004)).

The NLS-estimates of $\hat{\lambda}$ are also given in the table. The negative values of the estimates suggest negative spatial error dependence.

For the alternative neighbourhood specifications, the NLS-estimation of $\hat{\lambda}$ proved unstable in many cases[46]. No results for the transformed variables are therefore given for these specifications.

6 Conclusion

To conclude, the results largely reject the hypothesis of strategic interaction in local spending on childcare, primary education and care for the elderly. While there are some significant coefficients, especially in the regression on spending on care for the elderly, the results are not robust enough to be interpreted as evidence for strategic

[46] Unrealistic values for $\hat{\lambda}$ were estimated in some cases, or the results were not robust for small changes in the starting values.

interaction. Specifically, while the border-based baseline specification for spending on care for the elderly indicate a positive effect of neighbours' spending on care for the elderly, using the alternative neighbourhood definitions yielded no additional support for the theories of strategic interaction. Furthermore, coefficients larger than one in absolute value were given in some of the alternative neighbourhood specifications.

The aggregate results hence gives no robust evidence of strategic interactions in childcare, primary schooling and care for the elderly. However, it may be that the dependent variable that is used in this study, spending (per potential user) is not a relevant measure for service quality. While alternative quality measures for the time period under study are not easily found, the Swedish Association of Local Authorities and Regions have recently started to produce open evaluations of the relative performance of the public service in all Swedish municipalities.[47], providing additional measures on the quality of local public services. Rather than establishing that strategic interactions are not an issue in the types of services of this study, the results may be due to the difficulties of capturing quality-differentials when using expenditure data, and better possibilities to test for such interactions may be given in the future.

[47] The Swedish Association of Local Authorities and Regions started to publish yearly open quality comparisons for primary schooling and care for the elderly in 2007 (see *Öppna Jämförelser 2007 - Grundskola* (2007) and *Öppna Jämförelser 2007 - Äldreomsorg* (2007)), and will, in cooperation with the The National Board of Health and Welfare, work to develop these further.

Acknowledgements

I am grateful for useful comments and suggestions from Matz Dahlberg, Sören Blomquist, Hanna Ågren, Andreas Westermark, Federico Revelli, Jon Hernes Fiva, Jørn Rattsø, Eva Mörk, Tuomas Pekkarinen, Erik Grönqvist and seminar participants at the Department of Economics, Uppsala University, the 2005 IIPF Congress in Jeju, South Korea and the 2006 EEA Congress in Vienna. Financial support from the Hedelius Foundation and from The Swedish Research Council is gratefully acknowledged.

References

Ågren, H. and Edmark, K.: 2005, On spatial interactions in local tax determination: Swedish evidence. Mimeo, Department of Economics, Uppsala University.

Allers, M. and Elhorst, J.: 2005, Tax mimicking and yardstick competition among local governments in the netherlands, *International Tax and Public Finance* **12**, 493–513.

Anselin, L.: 1988, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishing, Dordrecht.

Aronsson, T., Lundberg, J. and Wikström, M.: 2000, The impact of regional expenditures on the local decision to spend, *Regional Science and Urban Economics* **30**, 185–202.

Baicker, K.: 2005, The spillover effects of state spending, *Journal of Public Economics* **89**, 529–544.

Baum, C. F., Schaffer, M. E. and Stillman, S.: 2003, Instrumental variables and GMM: Estimation and testing, *The Stata Journal* **3**(1), 1–31.

Besley, T. and Case, A.: 1995, Incumbent behavior: Vote seeking, tax setting and yardstick competition, *American Economic Review* pp. 25–45.

Bordignon, M., Cerniglia, F. and Revelli, F.: 2003, In seach of yardstick competition: A spatial analysis of italian municipality property tax, *Journal of Urban Economics* **54**, 199–217.

Brueckner, J.: 2000, Welfare reform and the race to the bottom: Theory and evidence, *Southern Economic Journal* pp. 505–525.

Brueckner, J.: 2003, Strategic interaction among governments: An overview of empirical studies, *International Regional Science Review* **26**, 175–188.

Case, A., Hines, J. and Rosen, H.: 1993, Budget spillovers and fiscal policy interdependence, *Journal of Public Economics* **52**, 285–307.

Dahlberg, M. and Edmark, K.: 2004, Is there a "race-to-the-bottom" in the setting of welfare benefit levels? evidence from a policy intervention. Working Paper 2004:19, Department of Economics, Uppsala University.

Dahlberg, M. and Fredriksson, P.: 2001, Migration and local public services. Working Paper 2001:12, Department of Economics, Uppsala University.

Dahlberg, M. and Johansson, E.: 2000, An examination of the dynamic behavior of local governments using GMM boot-strapping methods, *Journal of Applied Econometrics* **15**, 401–416.

Figlio, D., Kolpin, V. and Reid, W.: 1999, Do states play welfare games?, *Journal of Urban Economics* pp. 437–454.

Fredriksson, P., List, J. and Millimet, D.: 2004, Chasing the smokestack: Strategic policymaking with multiple instruments, *Regional Science and Urban Economics* **34**(4), 387–410.

Hanes, N.: 2002, Spatial spillover effects in the swedish local rescue services, *Regional Studies* **36**(5), 531–539.

Kelejian, H. H. and Prucha, I. R.: 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* **40**(2), 509–533.

Kelejian, H. H., Prucha, I. R. and Yuzefovich, Y.: 2004, Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. Forthcoming in J. LeSage and R.K. Pace, Advances in Econometrics, New York: Elsevier.

Kelejian, H. and Prucha, I.: 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* **17**, 99–121.

Kommunernas Ekonomiska Läge: 2003, Svenska Kommunförbundet.

Lundberg, J.: 2001, A spatial interaction model of benefit spillovers from locally provided public services. CERUM Working Paper 35:2001, Umeå University.

Millimet, D. and Rangaprasad, V.: 2007, Strategic competition amongst public schools, *Regional Science and Urban Economics* **37**(2), 199–219.

Öppna Jämförelser 2007 - Äldreomsorg: 2007. The Swedish Association of Local Authorities and Regions (Sveriges Kommuner och Landsting).

Öppna Jämförelser 2007 - Grundskola: 2007. The Swedish Association of Local Authorities and Regions (Sveriges Kommuner och Landsting).

Redoano, M.: 2003, Fiscal interactions among european countries. Warwick Economic Research Papers No 680, Department of Economics, Warwick University.

Revelli, F.: 2006, Spatial interactions among governments, *in* E. Ahmad and G. Brosio (eds), *Handbook on Fiscal Federalism*.

Schaltegger, C. and Zemp, S.: 2003, Spatial spillovers in metropolitan areas: Evidence from swiss communes. Working Paper No 2003: 06, CREMA, Brussels.

Shea, J.: 1997, Instrument relevance in multivariate linear models: A simple measure, *The Review of Economics and Statistics* **79**(2), 348–352.

Solé-Ollé, A.: 2003, Electoral accountability and tax mimicking: The effects of electoral margins, coalititon government, and ideology, *European Journal of Political Economy* **19**, 685–713.

Solé-Ollé, A.: 2006, Expenditure spillovers and fiscal interactions: Empirical evidence from local governments in spain, *Journal of Urban Economics* **59**, 32–53.

Staiger, D. and Stock, J.: 1997, Instrumental variables regressions when the instruments are weak, *Econometrica* pp. 557–586.

Stock, J. H. and Yogo, M.: 2002, Testing for weak instruments in linear IV regression. NBER Technical Working Paper 284.

Wilson, J.: 1999, Theories of tax competition, *National Tax Journal* **52**, 269–304.

Wilson, J. D. and Gordon, R. H.: 2003, Expenditure competition, *Journal of Public Economic Theory* **5**(2), 399–417.

A Appendix

A.1 First stage results

Table A.1: First stage results Baseline IV regression, Dependent variable: neighbours spending per potential user, Border-based weight matrix

Neighbours:	**Childcare and Education**	**Care Elderly**
	1SLS	1SLS
	(1)	(2)
Debt $t-1$	.009	.054
	(.011)	(.072)
Taxbase $t-1$	2.654	22.883
	(3.369)	(15.05)
Grants $t-1$	.004	.402
	(.073)	(.376)
Unemployment $t-1$	100.966	184.084
	(87.322)	(387.003)
Employment $t-1$	66.425	-346.348
	(70.608)	(418.974)
Welfare Recipients $t-1$	-116.058**	-142.069
	(55.434)	(289.847)
Ln Population $t-1$	-173.44	20065.15
	(2841.546)	(15813.77)
Left	-369.843*	486.653
	(215.292)	(1221.218)
Neighbours' values		
Debt $t-1$	-.049**	.233**
	(.02)	(.11)
Taxbase $t-1$	21.028***	55.337*
	(4.532)	(28.296)
Grants $t-1$	.974***	2.051***
	(.118)	(.769)
Unemployment $t-1$	270.197*	-1619.824**
	(144.817)	(768.272)
Employment $t-1$	419.696***	-88.253
	(140.195)	(743.802)
Welfare Recipients $t-1$	-263.568**	-1675.133**
	(123.17)	(765.197)
Ln Population $t-1$	-23301.8***	69220.99***
	(4750.953)	(24821.87)
year effects	yes	yes
fixed effects	yes	yes
Obs.	2751	2751

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions.

A.2 Baseline IV no municipality fixed effects

Table A.2: Baseline regression without fixed effects, Border-based weight matrix

	Childcare and Education		**Care Elderly**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
N Childcare and Education	.469***	.561***	1.215**	4.613**
	(.059)	(.172)	(.525)	(1.941)
N Care Elderly	-.004	-.004	.825***	.634***
	(.008)	(.022)	(.087)	(.207)
Debt $t-1$	-.006	-.01	-.174	-.235
	(.018)	(.018)	(.159)	(.178)
Taxbase $t-1$	20.731***	18.305***	-35.956	-88.296**
	(3.543)	(4.143)	(21.975)	(35.259)
Grants $t-1$	.89***	.787***	-.839	-3.028*
	(.126)	(.152)	(.999)	(1.626)
Unemployment $t-1$	-51.244	-109.935	4655.295***	3585.592**
	(183.344)	(192.139)	(1707.959)	(1763.242)
Employment $t-1$	61.638	34.165	4191.543***	4200.267***
	(116.34)	(133.673)	(1037.598)	(1225.657)
Welfare Recipients $t-1$	-43.821	-28.145	-1158.331	-441.338
	(94.228)	(107.356)	(918.303)	(972.302)
Ln Population $t-1$	463.096	461.102	-1632.402	-1286.629
	(360.783)	(357.592)	(2392.802)	(2489.017)
Left	1131.723***	979.387**	9114.321***	4128.145
	(414.566)	(450.688)	(3500.161)	(5074.422)
year effects	yes	yes	yes	yes
fixed effects	no	no	no	no
Cragg-Donald F		35.95		36.34
J-statistic		15.296		1.352
p-value J-statistic		0.009		0.929
Obs.	2715	2715	2746	2746

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

A.3 Baseline IV including county expenditures

Table A.3: IV regression, Including county expenditures

	Childcare and Education			
	Wborder	*Wmedia*	*Wmigr*	*Wmigr015*
	(1)	(2)	(3)	(4)
N Childcare and Education	.098	-.017	.035	-.035
	(.212)	(.229)	(.253)	(.236)
N Care Elderly	.02	-.066	.085	.105
	(.058)	(.062)	(.106)	(.105)
Debt $t-1$	-.035	-.035*	-.033	-.033
	(.023)	(.02)	(.022)	(.023)
Taxbase $t-1$	19.792***	22.196***	19.109***	18.575***
	(5.615)	(5.234)	(5.448)	(5.357)
Grants $t-1$	.939***	.956***	.937***	.947***
	(.154)	(.142)	(.16)	(.158)
Unemployment $t-1$	27.881	12.871	40.194	44.367
	(163.125)	(163.763)	(165.883)	(166.766)
Employment $t-1$	217.447	213.8	217.329	216.422
	(173.037)	(167.031)	(164.727)	(168.082)
Welfare Recipients $t-1$	-174.518	-174.996	-191.724*	-189.895*
	(112.703)	(106.823)	(108.303)	(108.096)
Ln Population $t-1$	-15765.26**	-12929.23**	-18052.32***	-18288.96***
	(6940.9)	(5712.656)	(6444.467)	(6081.03)
Left	420.256	321.644	437.59	472.853
	(368.805)	(371.539)	(379.257)	(389.363)
County costs	.227	.068	.266	.299
	(.202)	(.218)	(.211)	(.212)
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F	12.95	7.40	11.20	9.73
J-statistic	4.619	6.950	7.482	6.676
p-value J-statistic	0.464	0.224	0.187	0.246
Obs.	2420	2411	2420	2420

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level**,** respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying restrictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

Table A.4: IV regression, Including county expenditures

	Care Elderly			
	Wborder	*Wmedia*	*Wmigr*	*Wmigr015*
	(1)	(2)	(3)	(4)
N Childcare and Education	-1.244	-.632	-2.441*	-2.162
	(1.168)	(1.265)	(1.378)	(1.327)
N Care Elderly	.838***	.552	1.001**	.818
	(.316)	(.369)	(.492)	(.519)
Debt $t-1$	-.071	.022	-.036	-.041
	(.137)	(.125)	(.128)	(.127)
Taxbase $t-1$	18.614	35.747	32.883	32.801
	(33.173)	(32.914)	(31.604)	(32.76)
Grants $t-1$	.548	1.148	1.226	1.243
	(.889)	(.807)	(.861)	(.918)
Unemployment $t-1$	146.513	263.845	-84.682	43.797
	(836.879)	(783.638)	(756.697)	(766.071)
Employment $t-1$	1194.01	703.632	683.364	664.378
	(963.84)	(887.585)	(843.564)	(876.37)
Welfare Recipients $t-1$	-1022.801	-1539.9**	-1312.48*	-1356.353*
	(758.867)	(739.619)	(686.248)	(715.025)
Ln Population $t-1$	-17989.67	25068.19	20427.74	28329.07
	(35275.31)	(28973.56)	(27047.76)	(25191.43)
Left	-945.967	108.453	786.898	806.25
	(2371.256)	(2426.293)	(2249.147)	(2294.201)
County costs	-1.386	-1.293	-1.555	-1.708
	(1.409)	(1.507)	(1.17)	(1.231)
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
Cragg-Donald F	12.89	7.40	11.20	9.73
J-statistic	2.993	2.545	3.402	8.450
p-value J-statistic	0.701	0.770	0.638	0.133
Obs.	2451	2411	2420	2420

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level**,** respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. The J-statistic is the test of overidentifying resrtictions. Note that the Cragg-Donald F-statistic of instrument relevance is derived under the assumption of homoscedasticity. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

A.4 Transforming the variables to increase efficiency

Table A.5: Kelejian and Prucha IV regression, Border-based weight matrix

	Childcare and Education		**Care for Elderly**	
	IV	IV transformed	IV	IV transformed
	(1)	(2)	(3)	(4)
N Childcare and Education	0.049	0.116	-0.851	-0.798
	(0.180)	(0.182)	(0.979)	(0.930)
N care for elderly	0.015	0.003	0.643**	0.776***
	(0.054)	(0.052)	(0.292)	(0.250)
Debt $t-1$	-0.045**	-0.039*	-0.061	-0.041
	(0.021)	(0.021)	(0.128)	(0.122)
Taxbase $t-1$	20.830***	21.655***	22.706	20.724
	(5.193)	(5.186)	(30.645)	(26.991)
Grants $t-1$	0.978***	0.949***	0.633	0.657
	(0.149)	(0.156)	(0.867)	(0.830)
Unemployment $t-1$	35.476	120.435	46.187	91.989
	(159.061)	(152.133)	(833.042)	(731.396)
Employment $t-1$	306.644*	325.814**	1325.677	969.841
	(157.291)	(144.750)	(820.334)	(712.725)
Welfare Recipients $t-1$	-182.854*	-254.289**	-732.091	-727.223
	(101.343)	(101.352)	(701.298)	(641.644)
Ln Population $t-1$	-18690.225***	-18371.477***	10728.280	1414.449
	(6155.095)	(6333.887)	(31432.996)	(29221.325)
Left	437.615	332.145	-1228.647	-992.321
	(383.861)	(386.621)	(2383.212)	(2206.521)
year effects	yes	yes	yes	yes
fixed effects	yes	yes	yes	yes
$\hat{\lambda}$		-0.392		-0.403
Obs.	2380	2380	2380	2380

Note: The standard errors in parenthesis are robust to heteroscedasticity and serial correlation of arbitrary form. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. The spatial weight matrix for computing neighbours' spending is row standardized in all regressions. Instruments: neighbours' values of: taxbase, central government grants, long-term debt, unemployment, employment, population (in logs) and the share of population that receive welfare benefits, all lagged one time period.

Essay 3

Identifying Strategic Interactions in Swedish Local Income Tax Policies[1]

1 Introduction

The presence of strategic interactions in the tax setting of local governments is an important issue in the organization of the public sector. Such interactions have long been investigated in theoretical economic work (see e.g. Oates (2002) or Wilson (1999) for an overview). This has resulted in two main types of theoretical frameworks: tax competition and yardstick competition. In short, the former describes a situation where local governments compete for a mobile tax base whereas in the latter, tax interaction stems from the voters' evaluation of the performance of the local politicians by comparison with the neighbouring jurisdictions, something which induces the politicians to mimic the neighbours in order to be reelected. It can be shown that both models give rise to similar spatial reaction functions, where the tax rate of a jurisdiction is a function of the tax setting behaviour of surrounding jurisdictions, denoted neighbouring jurisdictions in the following.

This study tests for strategic interaction in Swedish municipal income taxes. Swedish data is highly suitable for this type of study. The local tax, which is an income tax, constitutes the main revenue source for the local governments. Furthermore, the local governments are responsible for the provision of essential welfare services[2] and have a high degree of autonomy both when it comes to the right to decide on the provision of local public services and the right to set the local income tax rate. The degree to which citizens depend on municipal services along with the heavy reliance on tax revenues hence make tax policy a salient issue in local policy making.

The aggregate evidence from earlier empirical work indicates that spatial interaction processes are present in local tax rates, but there is no consensus on the source (tax or yardstick competition) of interaction.[3] Identification of the source of interaction is however important, since the two theoretical models have different policy

[1] Co-authored with Hanna Ågren. Forthcoming in the Journal of Urban Economics.

[2] The municipalities are responsible for the provision of services such as care of the elderly, child care and education.

[3] For previous studies, see e.g. Heyndels and Vuchelen (1998), Besley and Case (1995), Buettner (2001), Bordignon, Cerniglia and Revelli (2003), Esteller-Moré and Solé-Ollé (2002), Solé-Ollé (2006), Revelli (2001), Brueckner and Saavedra (2001), and Allers and Elhorst (2005), which also gives a nice overview of the literature.

implications. While the theory of tax competition predicts that the tax rate in a decentralized setting will be lower than the social optimum, the yardstick competition theory in general predicts increased efficiency of the political system through better informed voters.

In contrast to many studies in this field, we make no a priori assumptions regarding the underlying theoretical framework. Instead, we make use of additional, indirect predictions from the theories of tax competition and yardstick competition to test for the presence of strategic interaction in these forms[4]. Specifically, we use a reform of the central government grants system, which changed the system of tax base equalization of the municipalities, to test for tax competition. The idea is that if we find the degree of interaction to be different after the reform, this can be seen as indirect evidence of tax competition. We also use two empirical implications descending from yardstick competition; namely that yardstick-type interaction is expected to be more prevalent during election years and when the political majority is weak, to test for strategic interaction in the form of yardstick competition.

In addition, we estimate a number of alternative specifications of the tax interactions equation, where we vary the criterion for defining a municipality's reference group. By specifying a set of alternative "neighbourhood criteria", that to different degrees correspond with the theories of tax and yardstick competition respectively, we can get an additional indication of which form of interaction is taking place.

Given the difficulties in ensuring identification through standard estimation, we believe that using these forms of indirect identification strategies is a fruitful way to proceed, both to separate strategic interactions from potential bias stemming from spatial error correlation and to separate between the different underlying theories.

The main results can be summarized as follows. The analysis provides evidence of spatial dependence in the tax rates among Swedish local governments: a tax cut of on average 1 percentage point in neighbouring jurisdictions is correlated with a decrease of about 0.74 percentage points in own taxes, which is of the same magnitude as that found in similar studies on interdependence in tax setting. The result is robust to using different specifications of neighbourhood, that are derived from the tax competition and yardstick competition theories respectively, although the size of interaction varies between the specifications. We furthermore find weak evidence that tax competition, while no evidence is found for yardstick competition.

The remainder of the paper is organized as follows. Section 2 present the data and describe the institutional setting, and Section 3 discusses the methodology and the tax equation specification. In Section 4 and 5, we present the results, and, finally, Section 6 concludes.

[4] These tests will be thoroughly described in section 5.

2 Data and institutional setting

To investigate the existence of horizontal interaction in tax setting, we use data on a panel of 283 Swedish local governments during 1993-2006.[5] Before describing the data, we will briefly comment on the Swedish institutional setting. The Swedish public sector is organized into three layers of government: national, county and municipal levels. The local units are responsible for the provision of important welfare services: the Swedish municipalities supply education, child care, social assistance and care for the elderly, while medical care and public transport are organized at the county level[6]. The focus here is on the municipalities.

Personal income is also taxed at the county level. This implies that there may also be vertical interactions in the tax rates (see e.g. Revelli (2006)). This is tested by including county taxes as a covariate in the baseline regression (treating the county tax rate as exogenous). The results show county tax to have a negative (indicating substitutes) but insignificant effect on the municipal tax rate. The interaction coefficient decreases somewhat, but is not significantly different from the baseline result when county tax is not included in the regression. The county tax will hence not be included in the baseline regression.

Swedish municipalities have the constitutional right of self government. The degree of autonomy refers both to the right to decide on the provision of local public services and their right to set the local income tax rate (note that only income is taxed locally – property taxes, for example, are set at the national level). Moreover, they are not limited by borrowing constraints.[7]

Our dependent variable, the local income tax, is a proportional tax rate and generates the main source of the municipalities' own revenues: tax revenue as a fraction of total revenues amounts to about 70 percent[8]. A small proportion, 15 percent on average, consists of central government grants.

The empirical analysis includes the following set of covariates to capture the economic and demographic complexion of the municipalities: unemployment rate, the proportion of young and old (defined as the share of the population aged 0-15 and 65+ respectively), population size, and the share of the population on welfare. Sweden is commonly treated as a bipartisan electoral system with either a left-wing or a

[5] 6 municipalities, that have either seceeded from or merger with another municipality during the time under study, have been excluded from the sample. These are Bollebygd, Gnesta, Lekeberg, Nykvarn, Knivsta and Trosa. Moreover, the island of Gotland is excluded due to the obvious difficulty in identifying neighbours.

[6] In two municipalities, Malmö and Göteborg, the municipalities are also responsible for the county-level tasks.

[7] In 2000, a balanced budget rule was introduced. However, it is not clear that the introduction of a balanced budget rule has had any real effect.

[8] This figure is for 2002, see "Kommunernas ekonomiska läge", Svenska Kommunförbundet, April 2003.

right-wing majority.[9] The data includes the number of votes for the ruling party coalition (left or right) in the four local elections during the sample period: 1994, 1998, 2002 and 2006. We construct a dummy variable (left-wing), indicating the party affiliation of the majority in power, to control for systematic differences in the tax setting between left- and right-wing local governments. Finally, we include taxable income per capita and per capita central governments grants in order to control for income differences. Including taxable income may seem inappropriate, since taxable income is endogenously determined in models of tax competition, and can hence not be incorporated as an exogenous covariate. Since grants are based on taxable income, the same problem applies to this variable. However, in Sweden, the effective taxable income of a municipality for year t, is calculated based on the tax assessment in t-1. The tax assessment in t-1, in turn, refers to the incomes in t-2. This implies that the effective taxable income is not affected by the current tax rate in this case, and can hence be included in the regression as a covariate. The same applies to the grants variable.[10]

The local budget decision is taken towards the end of the previous year. This means that at the time of the tax rate decision, the local politicians have information about the economic and demographic conditions in t-1. We therefore include the lagged values of all economic and demographic covariates (except for taxable income and grants, which are based on lagged values by construction, as explained above).

Table 1 provides some descriptive statistics of the included variables. All monetary variables are in SEK and have been deflated to year 2002 price level, while the variables defined as proportions are shown as percentage points.[11]

[9] See e.g., Alesina, Roubini and Cohen (1997) and Pettersson-Lidbom (2003). Following the categorization in Peterson (1992), the left-wing parties are the Left Party and the Social Democratic Party, and the parties characterized as right-wing are the Conservative Party, the Centrist Party, the Liberal Party and the Christian Democratic party (a fifth party, New Democracy, was added in 1991).

[10] While it is true that in the tax competition theory, the effect of the taxbase is channeled solely through the interaction among jurisdictions and hence should have no independent effect on the local tax rate, we acknowledge that the taxable income can change for other reasons than migration.

[11] The table shows the municipal tax rate to average 20.8 percent over the period, with a minimum at 13 and a maximum at 31. However, the great difference between the min and max value is due to the fact that in two of the municipalities in our sample, Malmö and Göteborg, the services otherwise organized at the county level, are provided by the municipality. If these two municipalities are excluded, the maximum value for the tax level decreases to 23.6 percent. Since these municipalities are large and possibly important in terms of strategic interactions, we will keep them in the sample. The difference in tax rates for these municipalities due to additional responsibilities will be captured by the inclusion of municipality-specific fixed effects.

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std	Min	Max
Income tax rate %	3962	20.8	1.5	13.2	31.3
Taxable income	3962	1086	198	698	2509
Grants	3949	8205	4406	-15399	25050
Unemployment %	3962	5.2	2.4	0.9	13.8
Proportion young (0-15)	3962	20.0	1.9	14.0	26.8
Proportion elderly (65+)	3962	18.9	3.8	5.9	30.0
Population size (per 1000)	3962	31015	57484	2553	771038
Share on welfare	3661	5.6	2.3	0.4	16.3
Party affiliation	3962	0.4	0.5	0	1

The grants variable is defined as total grants per inhabitant and contains equalization grants as well as general grants. The negative minimum value of this variable reflects the fact that some municipalities ended up as negative grants-recipients after a reform of the intergovernmental grant system in 1996 (more about this in section 6).

As can be seen in Table 1, there are a few missing observations for grants and the share of the population on welfare. Our analysis is performed on the unbalanced panel.

3 Methodology and specification of the tax reaction function

As previously discussed, the source of horizontal interaction depends on the assumptions made about individuals' underlying behaviour. The derived reaction function will however be the same, irrespective of whether the citizenry is assumed to react to tax policy differences by moving, or if immobile voters, at the polls, punish their elected politicians by ousting them from office, namely that the tax rate of a municipality is a function not only of jurisdiction-specific characteristics, but also of the tax rates of its neighbours' tax rates. We specify the following regression equation to test this hypothesis:

$$\tau = \theta W\tau + \beta' X + year + id + \epsilon, \tag{1}$$

where τ is a vector of the municipal tax rate, W is a neighbour weight matrix which gives positive weight to the policy values of neighbouring municipalities, so that $W\tau$ gives the average tax rate of the municipalities that are defined as neighbours. X is a matrix including a rich set of municipality-specific covariates, that were specified in the previous section, and ϵ is a vector of regression error terms. The specification also includes a set of yearly dummy variables, *year*, to control for time-varying influences common to all municipalities in a certain year, and for municipality-specific fixed effects, *id*, to control for time-invariant municipality-specific factors. A non-zero coef-

ficient for the neighbours' tax rates is consistent with the theories of tax competition and yardstick competition.[12]

Due to lack of degrees of freedom, matrix W cannot be estimated, but must be defined a priori. The standard approach in the literature is to define jurisdictions that share a border as neighbours. This is a neutral and simple definition capturing the idea of interaction being more likely to take place between closely situated jurisdictions. We will also show results from alternative definitions of neighbourhood, that are more closely related to the theories of tax interaction.

There are two main methodological challenges in estimating an equation of type (1). First, we need to account for the simultaneity in tax determination, which implies that standard OLS yields inconsistent and biased estimates of θ. Second, we need to ensure that the interaction coefficient does not suffer from bias due to omitted variables/spatial error correlation. We follow Kelejian and Prucha Kelejian and Prucha (1998), and select a subset of the neighbours' covariates as instruments to obtain consistent estimates of the interaction parameter.[13]

Since we want to avoid problems related to weak instruments[14], we select a small set of instruments that has good prediction power in the first stage. For this set of instruments, we generate the average values for the neighbouring municipalities, using the same neighbour weights as to generate neighbours tax rates, (i.e. WX^*, where X^* is the subset of covariates used as instruments). These values are then used to estimate the predicted values of neighbours' tax rate in the first stage regression. The instrumental variables have been chosen so as to satisfy the criteria of having independent explanatory power in the first stage regression and high joint explanatory power (the F-statistic is 18), as well as not being rejected by the standard test of over-identifying restrictions.

The variables that we use as instruments are the neighbours' unemployment rate and neighbours' share of welfare recipients, both lagged one time period. Using the values in $t-1$ is reasonable both since the local tax decision is taken towards the end of the previous year, and since this ensures the instruments' exogeneity in the sense of there being no direct effect of the dependent variable on the instruments. The first stage estimates can be found in Table A.1, appendix.

[12] As shown by Bordignon, Cerniglia and Revelli (2004), under certain circumstances, yardstick competition can also lead to a negative correlation among the taxes of neighbouring jurisdictions. The is the case if the reelection chances of a bad government are so low that, given low taxes in neighbouring jursdictions, it will be preferable to accumulate the maximum rent in the first period in office by raising the local tax rate and hence, not be reelected.

[13] For studies using an IV approach when testing for spatial auto-correlation in taxes, see e.g. Besley and Case (1995), Buettner (2001), Esteller-Moré and Solé-Ollé (2002), Heyndels and Vuchelen (1998), and (Revelli 2001).

[14] See e.g. Staiger and Stock (1997) for a discussion on small sample over-fitting bias and problems related to weak instruments.

The estimating equation includes own municipality tax policy determinants, hence the interaction coefficient is identified using the difference in the variation between the own and neighbouring municipality characteristics that are used as instruments. Including the additional set of covariates can furthermore be seen as a means to increase the probability that our instruments are valid. The idea is that by including the municipality characteristics, we require that the instruments are valid, not unconditionally, but conditional on the covariates.

Finally, since tax rates are persistent, we, in addition to heteroscedasticity consistent standard errors, allow for serial correlation within municipality.[15]

4 Results

4.1 Baseline results

This section presents the estimation results. Column 1 of Table 2 shows the OLS results, treating $W\tau$ as exogenous, and column 2 shows the baseline IV results using neighbours' unemployment rate and the share of welfare recipients, both in $t-1$, as instruments for neighbours' tax rates. The coefficient of neighbours' tax rates has a positive and significant effect on the own tax rate in both specifications, and the effect is even larger when treating own and neighbours' tax setting as a simultaneous decision in column 2. The coefficient is 0.74, implying that an average tax decrease (increase) of one percentage point among neighbouring municipalities, induces a 0.74 percentage point decrease (increase) in the own tax rate.

The fact that the IV-estimate is higher than the OLS-correspondence may seem puzzling, considering that we expect the simultaneity bias of the OLS-estimate to be positive. A possible explanation for the lower OLS-coefficient is that it also suffers from a downward bias, due to negative spatial error correlation.[16]

The F-statistic, as commented in the previous section, supports the validity of the instruments. Moreover, the Hansen test of over-identifying restrictions does not reject the validity of the instruments. Consequently, standard testing of the instruments suggests that the estimate of the spatial coefficient can be interpreted as a genuine or substantive interaction.

In addition to these tests of instrument validity, we have re-run the tax equation using different sets of covariates and instruments as a robustness check. The general results, which are available upon request, support our baseline finding.

[15] We estimate kernel based standard errors, using the bartlett kernel and bandwidth 2 (The bandwith was chosen through the rule of thumb for the bartlett kernel ($bw = 0.75 * T^{(1/3)}$), and the results are not sensitive to varying bandwidth). See the Stata manual for details.

[16] Negative spatial error correlation was indicated by the *Moran I*-statistic, which can be found in the earlier working paper version of this paper (see Edmark and Ågren (2006)).

Table 2: Baseline estimation of the tax reaction function

	OLS	IV
	(1)	(2)
Neighbours' tax rate	0.457***	0.745***
	(.038)	(.21)
Taxable income	0.005***	0.005***
	(.0006)	(.0007)
Grants	0.00002	0.00002
	(1.00e-05)	(1.00e-05)
Unemployment rate	0.039***	0.032**
	(.012)	(.013)
Population size	-0.00009***	-0.00008***
	(.00003)	(.00003)
Proportion of young (0-15)	0.051**	0.075**
	(.026)	(.035)
Proportion of elderly (65+)	0.084***	0.111***
	(.019)	(.027)
Share of welfare recipients	0.007	0.014
	(.014)	(.015)
Left-wing	0.038	0.029
	(.044)	(.046)
F-test		18.18
Hansen J (p-value)		0.107
Obs.	3932	3932

Note: The dependent variable is the municipal tax rate. Standard errors within parenthesis are robust to heteroscedasticity and serial correlation. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Year and municipality-specific fixed effects are included in the estimations. The spatial weight matrix for computing neighbours' taxes is based on sharing border and is row standardized. The F-statistic is the test of excluded instruments obtained from the first-stage equation. Hansen J is the p-value for the Hansen test of overidentifying restrictions. Instruments: neighbours' unemployment rate and neighbours' share of welfare recipients.

4.2 Results using alternative definitions of neighbourhood

The previous section showed the results when interaction was assumed to take place among municipalities sharing border. The theories of tax interaction however provide more detailed predictions regarding the definition of neighbourhood. Comparing the results from the different specifications can give us an idea of the likely source of tax interaction, and can also serve as a robustness test. This section shows the results from a set of alternative neighbourhood specifications.

Let w_{ij} define the elements in the weight matrix W (in equation (1)), then w_{ij} describes the degree of proximity between i and j, i.e., w_{ij} is the weight that the neighbouring municipality j has for municipality i.

First, we define neighbouring municipalities according to migration flows and construct a migration weight matrix where $w_{ij} = migration_{ij}$ where $migration_{ij}$ = the average out-migration from i to j, of individuals aged 16-65 in 1995-2002 and $w_{ij} = 0$ otherwise[17]. This definition is closely related to the tax competition model which

[17] Data on migration was made available by The Institute for Labour Market Policy Evaluation,

assumes that inter-municipal interaction is driven by competition for mobile tax payers. The fact that the weight matrix is fixed over time is important. Allowing the matrix to vary with the yearly migration patterns could give rise to endogeneity in the neighbourhood definition due to effects in tax rate changes on migration.

Secondly, we want to account for the fact that information is an important aspect, in both theories of strategic tax interaction. In order to behave strategically, the decision-makers, as well as the voters or tax payers, need to be informed of the tax rates of the surrounding jurisdictions. To capture the degree of information between i and j, we construct a media weight matrix; $w_{ij} = newspaper_{ij} x coverage_{ij}$, where $newspaper_{ij} = 1$ if i and j share a local newspaper, and $coverage_{ij}$ = the sum of average newspaper coverage of the local newspapers in j and $w_{ij} = 0$ otherwise[18][19].

Finally, we take into account the possibility that having similar preferences for locally provided services can facilitate comparisons of fiscal policies across municipalities. Citizens may base their voting, or moving decision, on comparisons of fiscal policies across municipalities with similar ideological positions. If so, we should find more tax interaction in municipalities that are surrounded by municipalities with the same political majority, while municipalities with neighbours of a different political colour will be less affected by the neighbours tax setting behaviour. We test for the interaction between the dummy variables for the municipalities that fulfill these criteria, and the tax interaction effect.

In addition to these neighbourhood definitions, we also construct a weight matrix based on an arbitrary neighbourhood criterion, namely being adjacent as we rank the municipalities in alphabetical order.[20][21] The idea is that by using this arbitrary measure of neighbourliness, we can test whether our interaction coefficient at all measures a spatial pattern, or whether we would obtain the same results irrespective of how we define neighbours. Naturally, we expect to find no tax interaction as we use the alphabetic weight matrix.

The results using alternative definitions of neighbourhood are displayed in Table 3, columns 2-5. For comparison, the baseline results using the border-sharing criterion are presented in column 1.

---

and is originally from Statistics Sweden. These data were generated in Edmark (2007).

[18] The data on local newspapers is from 1994, 1998 eller 2002 and is from Tidningsstatistik AB. We are grateful to Helena Svaleryd och Jonas Vlachos for having made it available to us.

[19] We select all newspapers that are given out at least six days a week. This leaves some municipalities with no newspaper. For these we include newspapers that are given out less then six days a week. There are two newspapers that have a national coverage, Dagens Nyheter and Svenska Dagbladet. These are counted as local newspapers only for the municipalities in the Stockholm county, since they cover local news in this region.

[20] Specifically, we define the two preceding and the two following municipalities in alphabetical order as neighbours.

[21] This follows e.g. Case, Hines and Rosen (1993).

Table 3: Alternative definitions of neighborhood

	$WBorder$	$WMigration$	$WMedia$	$WPolitics$	$WAlpha$
	(1)	(2)	(3)	(4)	(5)
Tax rate of neighbours	0.745*** (.21)	0.218*** (.076)	0.777*** (.222)	0.577*** (.188)	0.627 (.442)
Neighbours same majority				-0.001 (.002)	
Neighbours different majority				-0.001 (.002)	
Controls	yes	yes	yes	yes	yes
Year effects	yes	yes	yes	yes	yes
Fixed effects	yes	yes	yes	yes	yes
F-test	18.18	52.20	29.09	-	9.99
Cragg-Donald F				9.51	
Hansen J (p-value)	0.107	0.002	0.010	0.020	0.428
Obs.	3932	3932	3918	3932	3932

Note: The dependent variable is the municipal tax rate. Standard errors within parenthesis are robust to heteroscedasticity and serial correlation. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Year and municipality-specific fixed effects are included in the estimations. For the construction of the spatial weight matrix for computing neighbours' taxes, columns 2-5, see text. The F-statistic is the test of excluded instruments obtained from the first-stage equation. Hansen J is the p-value for the Hansen test of overidentifying restrictions. Instruments: neighbours' unemployment rate and neighbours' share of welfare recipients.

As can be seen in Table 3, the coefficient on neighbours' tax rate exhibits the same pattern as in the baseline specification. The estimate is positive and significant when defining neighbours using the migration criteria, although the coefficient is smaller, as well as in the case with the media weight matrix. The results in column 4 however give no support for a stronger (weaker) interaction effect for municipalities that are surrounded only by neighbours with the same (different) political majority. The result using the alphabetical weight matrix in column 5 finally confirms that our interaction coefficient is indeed picking up something spatial; the interaction coefficient is insignificant when we use the alphabetic definition of neighbours.

We interpret the results in Table 3 as broadly supporting the hypothesis of strategic interaction in the tax rate. Using different specifications that are, in different ways, are based on the tax competition and yardstick competition theories, yields interaction coefficients that are significant, whereas our arbitrary alphabetic neighbourhood definition yields an insignificant result.

In the specification in column 4, *Wpolitics*, there are three endogenous variables, which means that the F-statistic cannot be used to test for instrument validity (see e.g. Baum, Schaffer and Stillman (2003) for a description of the problem). Instead, the Table shows the Cragg-Donald F-statistic.[22] The test statistic rejects the hypothesis

[22] This is originally a test of underidentification, but can also be used for testing for weak instruments by using the critical values computed by Stock and Yogo (2002). It shall be noted, however, that this test statistic and the related critical values are derived under the assumption of homoscedas-

of weak instruments at the five percent level of significance[23]. In addition, the partial Shea R^2 is very similar to the partial R^2 for all three instruments, which also supports the validity of the instruments (see Baum et al. (2003)).

Finally, the Hansen J-statistic rejects the null of instrument exogeneity for three of the specifications, which means that we shall interpret the results from the alternative neighbourhood definitions with caution.

5 Identification of strategic tax interaction

In this section, we exploit empirical implications that are consistent with the tax competition and the yardstick competition framework, respectively, to provide a further test of strategic interaction.[24] First, to test for tax competition, we use a reform of the equalization grants system which changed the incentives for local politicians to interact in the setting of municipal tax rates. Second, we make use of features of the electoral system to identify yardstick competition effects.

5.1 Grant reform

As previously discussed, when tax bases are mobile across regions, tax competition may have negative consequences for efficiency due to a race to the bottom in tax effort and hence, will put a downward pressure on local government spending. However, a system of equalizing grants can correct for this and lead to an efficient outcome,[25] due to the fact that the negative effect of higher tax rates on the tax base is partly compensated by higher equalizing transfers.

Sweden is viewed as an highly ambitious country regarding horizontal equity in the distribution of public services (see e.g. Rodden and Eskeland (2003)). Similarly to a number of countries, Sweden has a system of tax revenue and expenditure equalization. The purpose of tax revenue equalization is to bring per capita tax revenues in all regions close to the national average. The expenditure equalization aims at reducing the differences in structural cost conditions of public services across municipalities.

The formula for the Swedish tax revenue equalizing grant is shown in equation 2:

$$G_i = 0.95 n_i t((yg) - y_i) \tag{2}$$

ticity, and it is not clear how well it performs when this assumption is not fulfilled.

[23] The critical value for three endogenous variables, allowing for a maximum relative bias of 10% compared to OLS, and at the 5% significance level, is 6.61. According to Stock and Yogo (2002), this value is comparable to the Staiger and Stock (1997) rule of thumb of 10 for the F-statistic in a regression with one endogenous variable.

[24] For a discussion regarding identification of the theoretical model in this context, see, e.g. Brueckner (2003) and Revelli (2006).

[25] See e.g. Bucovetsky and Smart (2002) and Köthenbürger (2002).

As the equation shows, the grants for municipality i, G_i, is calculated on basis of the difference between the country-wide average taxable income, y, multiplied with a factor g, and the average taxable income in the municipality, y_i. The effective grant received by a municipality is given by multiplying this difference with 0.95 times the average tax rate in the country, t, and the number of inhabitants in the municipality, n_i.

In the 1990s, the Swedish grant system underwent a reform. Prior to 1996, the municipalities were granted 127% of the country-wide average taxable income, i.e., in terms of equation 2, $g = 1.27$. Municipalities with a tax base exceeding 127% of the country-wide average, and hence ending up with a negative equalization grant according to the grants formula, were exempt from the system. However, in practise, only three municipalities were affected by this. This implies that all municipalities were given either positive or zero equalization grants.

In 1996, two features of the equalizing grants system were changed: first, the municipalities were granted 100%, instead of 127%, of the country-average of the tax base, i.e. g changed from 1.27 to 1, and second, the new system allowed for no exemptions, but encompassed all municipalities. Under the new system, a municipality hence paid a positive contribution (negative grant) to the system, if the average taxable income in the municipality exceeded the country-wide average and paid a negative contribution (received a positive grant) if it were below the country-wide average.

We argue that this reform can be used to test for the presence of tax competition among Swedish municipalities. If local governments act strategically to attract mobile tax payers, they should react to a reform that changes the system of tax revenue equalization among municipalities.

First, the grants reform decreased the guaranteed tax base, through the reduction in g. This implies that the relative importance of the own tax base for a municipality's budget increased after the reform. As a result, the incentives for attracting/not loosing tax base may be higher after the reform, and we would hence expect to see more tax competition in the post-reform period.

Second, however, the reform made the tax base equalization clearer and more directly visible to the municipalities; first, by encompassing all municipalities and second, and by making the system financially neutral where municipalities with good conditions directly compensate those with poor conditions. As discussed in the Swedish Government Official Report, SOU 2000:120, these features of the system introduced in 1996 dramatically changed the design of the equalization system, from being a more indirect system to more direct equalization. In addition, one can argue that the new system made equalization more visible to the individual municipality by increasing the relative weight of the share of the grant that depends on the municipality-specific

tax base in the grants formula. This implies that the reform would lead to a lower degree of tax competition.

There are hence two possible effects of the grants reform on the incentives for tax competition: first, a positive effect from the increased relative importance of the own tax base, and second, a negative effect through a more transparent tax base equalization system.

In order to test if the reform of the grants system changed the degree of tax interaction neighbouring municipalities, we re-estimate the tax reaction equation (1) including neighbours' tax rate interacted with a dummy variable to indicate the pre- and post-reform periods.

The results in Table 4 indicate that interaction is about 0.1 lower after the reform. The fact that the rate of tax interaction changed after the reform, is consistent with the hypothesis of tax competition. The negative sign of the interaction coefficient, furthermore suggests that the municipalities perceived the reform as increasing the rate of tax equalization.

Regarding instrument validity, the Cragg-Donald statistic is higher than the critical value for two endogenous variables, which suggests that the instrument are valid.[26] This is also supported by the fact that the Shea R^2 and the partial R^2 for the instruments are relatively similar.[27]

[26] The critical value for two endogenous variables, allowing for a maximum relative bias of 10% compared to OLS, and at the 5% significance level, is 8.78. According to Stock and Yogo (2002), this value is comparable to the Staiger and Stock (1997) rule of thumb of 10 for the F-statistic in a regression with one endogenous variable.

[27] The Shea R^2 for Neighbours' tax rate is 0.026, while the partial R^2 is 0.035, and for Neighbours' tax rate interacted with the Reform dummy, the corresponding values are 0.194 and 0.265.

Table 4: Pre- and postreform interactions

	(1)
Neighbours' tax rate	0.412**
	(.194)
Neighbours' tax rate*Reform	-0.1*
	(.053)
Taxable income	0.005***
	(.0006)
Grants	1.00e-05
	(1.00e-05)
Unemployment rate	0.04***
	(.013)
Population size	0***
	(.00003)
Proportion of young (0-15)	0.028
	(.038)
Proportion of elderly (65+)	0.075***
	(.023)
Share of welfare recipients	0.008
	(.016)
Left-wing	0.046
	(.047)
Cragg-Donald F	18.64
Hansen J (p-value)	0.200
Obs.	3932

Note: The dependent variable is the municipal tax rate. Standard errors, which are shows in parenthesis, are robust to heteroscedasticity and serial autocorrelation. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Year and municipality-specific fixed effects are included in the estimations. The spatial weight matrix for computing neighbours' taxes is based on sharing border and is row standardized. The F-statistic is the test of excluded instruments obtained from the first-stage equation. Hansen J is the p-value for the Hansen test of overidentifying restrictions. Instruments: neighbours' unemployment rate and neighbours' share of welfare recipients.

The coefficient of neighbours' tax rate is positive and significant as in the baseline case. The coefficient is however lower than what we would expect considering the estimate of the baseline regression in Table 2. A possible explanation for this is that some of the interaction effect is picked up by the year effects for the pre-reform period.

5.2 Testing for yardstick competition effects

In this section, we will make use of two general predictions from yardstick competition theory.[28] First, since the tax policy of surrounding jurisdictions enters the tax equation though the effect on the probability of winning the election, it is likely that an incumbent with a weak political majority will pay closer attention to the neighbours' tax policy, than an incumbent with a strong majority, who is likely to win the election irrespective of neighbours' policies. Hence, we expect tax rate interaction to be stronger in municipalities where the ruling majority is weak. Second, since the

[28] Similar tests for yardstick competition effects are used in e.g. Besley and Case (1995), Bordignon et al. (2003) and Solé-Ollé (2006).

voters' evaluation of the politicians affect their voting behaviour, politicians will have stronger incentives to set taxes in line with the tax policy in neighbouring regions at the time of the local election. Hence, we expect the interaction in tax setting to be stronger during election years.[29] Table 5 displays the results. Since having a weak political majority may have an independent effect on the tax rate [30], column 2 includes a dummy variable indicating a weak political majority.

In columns 1 and 2, the tax rate of neighbours is interacted with a dummy taking on the value of one if the political majority of the municipality is weak, and zero otherwise. We define "weak" as a municipality, where either the left-wing or the right-wing block has a support equal to or less than 55 percent of the votes. According to this definition, in our sample of 3113 observations, there are 1752 municipalities where the majority is weak. The results from these specifications do not indicate any difference in interaction between municipalities with a weak and strong political majority; the coefficients for neighbours' tax interacted with our indicator for a weak majority are insignificant.

Column 3 displays the results when interacting neighbours' tax rate with an election year dummy. The result does not support the prediction that strategic interaction is stronger during election years. However, this result may be due to little variation, considering that there are only four election years in our sample period.

The Cragg-Donald statistics support the validity of the instruments, as can be seen in the table, and the Shea partial R^2 and the partial R^2 for the instruments are relatively similar for all specifications.[31]

[29] In models of rational election cycles, the prediction is that politicians signal their competence by reducing taxes (and increasing spending) in election years, see e.g. Persson and Tabellini (2000). The election dummy may pick up both the effect of lower taxes in election years and stronger incentives to interact in the tax setting, both due to stronger incentives to perform just ahead of elections. However, to the extent that the former effect is similar for all municipalities, it will be controlled for by the inclusion of year-specific effects.

[30] This result is consistent with earlier work analysing the relationship between political competition and policy outcomes. For example, recent work by Besley and Case (2003) and Besley, Persson and Sturm (2005) find that increased political competition leads to lower taxes. However, according to our definition, municipalities with a weak majority include undefined party coalitions, where neither the left- nor the right-wing bloc is in power, which is typically the case when there are strong local parties. The indicator variable for weak majorities may hence pick up information on these local parties. Since we lack information on the party affiliation of local parties, we refrain from further interpreting this result.

[31] In column 1, the Shea R^2 as well as the partial R^2 for Neighbours' tax rate lie between 0.136 and 0.020 throughout the specifications. The corresponding values for Neighbours' tax rate interacted with the Political weakness dummy are 8.814 and 0.835 in column 1, and 0.096 and 0.082 in column 2. For Neighbours' tax rate interacted with the Election year dummy, in column 3, the Shea $R^2$is 0.260 and the partial $R^2$is 0.341.

Table 5: Electoral tightness and election year

	Weak	Weak	Election year
	(1)	(2)	(3)
Neighbours' tax rate	0.743***	0.699***	0.751***
	(.207)	(.19)	(.218)
Neighbours' tax*Weak	0.001	0.083	
	(.002)	(.067)	
Neighbours' tax*Election			-0.002
			(.031)
Taxable income	0.005***	0.005***	0.005***
	(.0007)	(.0007)	(.0007)
Grants	0.00002	0.00002	0.00002
	(1.00e-05)	(1.00e-05)	(1.00e-05)
Unemployment rate	0.032**	0.03**	0.032**
	(.013)	(.013)	(.013)
Population size	-0.00008***	-0.00009***	-0.00008***
	(.00003)	(.00003)	(.00003)
Proportion of young (0-15)	0.075**	0.074**	0.076**
	(.035)	(.034)	(.035)
Proportion of elderly (65+)	0.11***	0.113***	0.111***
	(.027)	(.027)	(.028)
Share of welfare recipients	0.014	0.017	0.015
	(.015)	(.015)	(.015)
Left-wing	0.037	0.048	0.029
	(.05)	(.049)	(.046)
Weak Majority		-1.691	
		(1.398)	
Cragg-Donald F	12.77	13.60	12.24
Hansen J (p-value)	0.279	0.381	0.164
Obs.	3932	3932	3932

Note: The dependent variable is the municipal tax rate. Standard errors, which are shows in parenthesis, are robust to heteroscedasticity and serial autocorrelation. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Year and municipality-specific fixed effects are included in the estimations. The spatial weight matrix for computing neighbours' taxes is based on sharing border and is row standardized. The F-statistic is the test of excluded instruments obtained from the first-stage equation. Hansen J is the p-value for the Hansen test of overidentifying restrictions. Instruments: neighbours' unemployment rate and neighbours' share of welfare recipients.

6 Conclusions

This paper uses data on Swedish local governments to test for strategic interaction in municipal tax setting. Following a number of previous studies, the endogeneity of neighbouring local governments' taxes is handled using an IV approach where neighbours' tax rates are instrumented using a subset of neighbours' characteristics. Using an IV approach is intuitively appealing in this context, since identification becomes explicit; spatial interaction in taxes is caused by the changes in the part of neighbours' taxes attributable to the observable characteristics of neighbours, which are used as instruments.

We find evidence of a spatial pattern in tax rates: a tax cut of on average 1 percentage point in neighbouring jurisdictions is correlated with a decrease of about 0.74 percentage points in own taxes. A positive correlation between the tax changes of neighbouring municipalities is found also for the alternative specifications of neighbourhood that are derived from the tax competition and yardstick competition theories respectively, although the size of the effect varies. This suggests that there is a spatial pattern in the data that is consistent with the predictions from these theories.

However, we stress that to identify the source of the interaction, additional testing is needed. We therefore employ two tests based on empirical implications that are consistent with either the tax competition or the yardstick competition framework, to test for the source of interaction. First, to test for tax competition, we use a reform of the equalization grants system which changed the incentives for local politicians to interact in the setting of municipal tax rates. Second, we make use of features of the electoral system to identify yardstick competition effects.

Using these tests, the paper finds weak evidence supporting that the spatial correlation in taxes among Swedish local governments can be explained by incentives to attract mobile taxpayers, as suggested by the theory of tax competition. We find no support for the hypothesis that the tax interaction is driven by electoral concerns.

Our results hence underline the importance of using direct tests for strategic interactions, in addition to estimating the spatial interaction equation.

Acknowledgements

We are grateful to Matz Dahlberg, Eva Mörk, Federico Revelli, Jørn Rattsø, Jon H. Fiva and Helena Svaleryd, as well as to two anonymous referees and the editor of the Journal of Urban Economics for helpful comments and suggestions. Seminar participants at Uppsala University, Gothenburg University, University of Gävle, STICERD at the London School of Economics, "Workshop on Nordic Public Economics and Taxation" in Helsinki, 2004, the 2006 Annual meeting of EEA in Vienna, the Swedish Ministry of Finance, and at the to the Research Institute of Industrial Economics have also provided valuable discussions. Financial support from Finanspolitiska Forskningsinstitutet, Nordisk Skattevitenskapelig Forskningråd, The Swedish Research Council and from the Swedish Assocation of Local Authorities and Regions is gratefully acknowledged.

References

Alesina, A., Roubini, N. and Cohen, G.: 1997, *Political Cycles and the Macroeconomy*, MIT Press, Cambridge.

Allers, M. and Elhorst, J.: 2005, Tax mimicking and yardstick competition among local governments in the netherlands, *International Tax and Public Finance* **12**, 493–513.

Baum, C. F., Schaffer, M. E. and Stillman, S.: 2003, Instrumental variables and GMM: Estimation and testing, *The Stata Journal* **3**(1), 1–31.

Besley, T. and Case, A.: 1995, Incumbent behavior: Vote seeking, tax setting and yardstick competition, *American Economic Review* pp. 25–45.

Besley, T. and Case, A.: 2003, Political institutions and policy choices: Evidence from the united states, *Journal of Economic Literature* **41**(1), 7–73.

Besley, T., Persson, T. and Sturm, D.: 2005, Political competition and economic performance: Theory and evidence from the united states. Mimeo, Institute for International Economic Studies, Stockholm University.

Bordignon, M., Cerniglia, F. and Revelli, F.: 2003, In seach of yardstick competition: A spatial analysis of italian municipality property tax, *Journal of Urban Economics* **54**, 199–217.

Bordignon, M., Cerniglia, F. and Revelli, F.: 2004, Yardstick competition in intergovernmental relationships: Theory and empirical predictions, *Economics Letters* **83**, 325–333.

Brueckner, J.: 2003, Strategic interaction among governments: An overview of empirical studies, *International Regional Science Review* **26**, 175–188.

Brueckner, J. K. and Saavedra, L.: 2001, Do local governments engage in strategic property-tax competition?, *National Tax Journal* **54**, 203–229.

Bucovetsky, S. and Smart, M.: 2002, The efficiency consequences for local revenue equalization: Tax competition and tax distortions. CESIFO Working Paper 767.

Buettner, T.: 2001, Local business taxation and competition for capital, *Regional Science and Urban Economics* **31**, 215–245.

Case, A., Hines, J. and Rosen, H.: 1993, Budget spillovers and fiscal policy interdependence, *Journal of Public Economics* **52**, 285–307.

Edmark, K.: 2007, Strategic competition in local spending on childcare, schooling and care for the elderly - evidence from swedish municipalities. Unpublished paper, Department of Economics, Uppsala University.

Edmark, K. and Ågren, H.: 2006, Identifying strategic interactions in swedish local income tax policies. Working paper 2006:22, Department of Economics, Uppsala University.

Esteller-Moré, A. and Solé-Ollé, A.: 2002, Tax setting in a federal system: The case of personal income taxation in canada, *International Tax and Public Finance* **9**, 235–257.

Heyndels, B. and Vuchelen, J.: 1998, Tax mimicking among belgian municipalities, *National Tax Journal* **60**, 89–101.

Kelejian, H. and Prucha, I.: 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* **17**, 99–121.

Köthenbürger, M.: 2002, Tax competiton and fiscal equalization, *International Tax and Public Finance* **9**, 391–408.

Oates, W.: 2002, Fiscal and regulatory competition: Theory and evidence, *Perspektiven der Wirtschaftspolitik* **3**, 377–390.

Persson, T. and Tabellini, G.: 2000, *Political Economics: Explaining Economic Policy*, MIT Press, Cambridge.

Peterson, O.: 1992, *Kommunalpolitik*, Publica, Stockholm.

Pettersson-Lidbom, P.: 2003, Do parties matter for policy choices? a regression discontinuity approach. Working Paper 2003:15, Department of Economics, Stockholm University.

Revelli, F.: 2001, Spatial patterns in local taxation: Tax mimicking or error mimicking?, *Applied Economics* **33**, 1101–1107.

Revelli, F.: 2006, Spatial interactions among governments, *in* E. Ahmad and G. Brosio (eds), *Handbook on Fiscal Federalism.*

Rodden, J. and Eskeland, G.: 2003, Introduction and overview, *in* G. E. J Rodden and J. Litvack (eds), *Fiscal Decentralization and the Challenge of Hard Budget Constraints.*

Solé-Ollé, A.: 2006, Expenditure spillovers and fiscal interactions: Empirical evidence from local governments in spain, *Journal of Urban Economics* **59**, 32–53.

Staiger, D. and Stock, J.: 1997, Instrumental variables regressions when the instruments are weak, *Econometrica* pp. 557–586.

Stock, J. H. and Yogo, M.: 2002, Testing for weak instruments in linear IV regression. NBER Technical Working Paper 284.

Wilson, J.: 1999, Theories of tax competition, *National Tax Journal* **52**, 269–304.

A Appendix: First stage estimates

Table A.1: First stage results

	1SLS
Neighbours' share of welfare recipients	-0.061***
	(.019)
Neighbours' unemployment rate	0.099***
	(.019)
Taxable income	-0.0002
	(.0006)
Grants	-2.86e-06
	(1.00e-05)
Unemployment rate	-0.002
	(.011)
Population size	-8.62e-06**
	(3.57e-06)
Proportion of young (0-15)	-0.095***
	(.02)
Proportion of elderly (65+)	-0.097***
	(.017)
Share of welfare recipients	-0.016*
	(.009)
Left-wing	0.036
	(.029)
Obs.	3932

Note: The dependent variable is the neighbours' municipal tax rate. Standard errors within parenthesis are robust to heteroscedasticity and serial correlation. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Year and municipality-specific fixed effects are included in the estimations. The spatial weight matrix for computing neighbours' taxes is based on sharing border and is row standardized.

Essay 4

Did the Introduction of Activation Programs in Stockholm Town Districts Affect the Moving Choices of Welfare Prone Individuals?

1 Introduction

The increasing number of individuals on welfare benefits has given rise to a new trend in the welfare benefit policy of the Western world, in the form of a shift in focus from the rights to the obligations of recipients of welfare. In the US, as well as in many European countries, policies have been implemented that restrict the availability of welfare benefits, for example by introducing time limits or by conditioning benefit receipt on participation in job search or job training programs. Such policies are often described by the term "workfare", since they require that the recipients to some extent work for their welfare benefits (see e.g. Blomberg, Ekström and Rauhut (2006)).

Sweden is no exception to this trend. Here, a change in the Social Service Act in 1998 enabled local municipalities and town districts to strengthen the rules for benefit eligibility, by conditioning benefit receipt on participation in programs for job search and job training. The law has been used by several municipalities and town districts to implement a new type of labour market program for recipients of welfare benefits that are capable of working, so called activation programs. The implementation of such programs is highly decentralized: the municipalities, or, in case of larger towns, the town districts, are responsible for the decision of whether or not to start a program, as well as for the program design.

Welfare caseloads have fallen dramatically in many town districts that have implemented activation programs, something that is often interpreted as a sign of success of the activation policy. However, no study has yet confirmed that the activation programs have increased the employment rates among recipients of welfare[1], and there

[1] While there have been a number of evaluations of single activation programs, these have been descriptive in character, and have not been able to isolate program effects from for example business cycle effects.

is in fact little information on what happens to the individuals that end the program. For example, a survey on the Stockholm town district Skärholmen points out that information on the cause for ending the activation program is lacking for as much of 56% of the participants, and that only 18% state that they are employed after the program (see Thoren (2005)). Similar figures are given in Ekström (2005), who studies the same town district and also notes that the third most common category is having moved from the town district: 11% state this as the cause for ending the program.

We hence do not know if the activation programs have helped participants to become employed, or if the reduction in welfare caseloads is due to other factors such as improved labour market conditions in general or even to outmigration of welfare prone individuals from municipalities and districts that have implemented stricter activation policies.

This paper tests if the implementation of activation programs in Stockholm town districts affected the moving choices of welfare prone individuals. The fact that the implementation of such programs increases the obligations of recipients of welfare, suggests that it would make welfare prone individuals want to move from a town district that has an activation program. However, it is also possible that the services of the program are appreciated by the participants, and that the effect on migration goes the other way.

The expected aggregate effect on migration is hence unknown, and depends on how the individuals value the stricter rules and loss of leisure time against the services of the program. If a large share of the target population avoids the program by moving, the effectiveness of the program to reduce welfare benefit dependency is naturally diminished. Evaluating whether this is the case is therefore interesting from a policy perspective.

The previous literature on migration responses to local welfare benefit policy, has in general focused on the effects of differences in local welfare benefit generosity, and often use American state level data. The earlier literature gives a mixed evidence: some studies estimate large welfare migration, while other report no effects.[2]. The results from the recent, methodologically more credible, studies, however suggest that welfare generosity does affect migration, but that the effect is rather small (see e.g. McKinnish (2005) and McKinnish (2007), Gelbach (2004), and Meyer (2000)). In contrast, Fiva (2007) finds large migration effects when studying Norwegian municipalities.

The only previous paper, to my knowledge, that tests for migration responses to stricter rules for welfare benefit eligibility is Kaestner, Kaushal and Ryzin (2001), who test if the introduction of time limits, financial sanctions for non-compliance, and

[2] See Meyer (2000) and Moffitt (1992) for overviews of the early literature.

strict work eligibility rules in US states affected outmigration from the state. They compare the migration response among groups of women that differ in the propensity to receive welfare benefits, and find that the use of time limits increased outmigration among welfare prone individuals. No separate effects could however be estimated for financial sanction or work exemption policies, since the states using such policies were also using time limits.

In contrast to other studies of welfare migration, Kaestner et al. (2001) also study the situation *after* migration has taken place, as a further test of the cause for moving. Interestingly, they find that many of those that moved from the more strict states, were employed after the move. This result may suggest increased labour market mobility in the states that have implemented the stricter rules. However, an alternative explanation is that the moves were not at all related to differences in the welfare benefit policy, but rather to different employment possibilities. For example, it can be the case that jurisdictions that experience a declining economic situation are more willing implement stricter welfare benefit rules. If so, moves that look lime welfare migration may in fact be motivated by an unfavorable labor market.

This highlights the difficulties of controlling for other factors that affect migration. However, controlling for the characteristics of all possible moving-combinations is in general not feasible.

In this study, there is no need to control for varying labour market characteristics of the local jurisdictions, since all individuals live within the same municipality, i.e. in the same local labour market area. An individual who finds a job in another town district does not have to move, since one can easily commute within the municipality. For the purpose of this study, it is also important to point out that all individuals face the same welfare benefit level, since this is set at the municipal level. By limiting the analysis to Stockholm town districts, we hence minimize the risk of omitted variable bias due to differences in local characteristics.

In addition, the fact that merely a short-distance move is necessary in order to end up under a different benefit policy, makes the migration hypothesis a more plausible story. It is for example likely that individuals are better informed of the welfare benefit policies of the town districts in the vicinity, and that moving costs are lower for short-distance moves.

The fact that the starting year of the activation programs differs among the town districts in our sample, means that two sources of variation can be exploited to identify the effect of the program on the moving choices of welfare prone individuals. First, we can compare the moving choices of welfare prone individuals before and after the law revision, in town districts that did and did not start an activation program after the revision, in a district-level difference-in-differences analysis. Second, we can add a further component to the analysis, and compare the migration difference-

in-differences estimates for groups that differ in the propensity to receive welfare benefits. The idea is that the moving behaviour of individuals with a high propensity to use welfare will be affected by the programs, while individuals that are not welfare prone will not be affected. Combining this approach with the district level analysis yields a difference-in-difference-in-differences estimator. This approach gives good possibilities to control for the effects of unobserved trends that affect migration. This is an advantage, compared to most other studies of welfare migration, which rely on comparison group based difference-in-differences analysis.

The results of this study give some indications of a *negative* migration response to the activation programs among welfare prone individuals; i.e. welfare prone individuals are *less* likely to move from the town district, compared to less welfare prone groups, when there is an activation program in place. This is contrary to the positive welfare migration effects that are found in most previous studies. However, the result is not robust to changes in comparison group nor to changes in the sample of town districts. The conclusion that can be drawn from the study is that there are at least no indications that the activation programs lead to outmigration of welfare prone individuals.

The outline of the remaining paper is as follows: section 2 describes the background of the activation programs, section 3 provides a simple theoretical framework for the effects of the activation program on migration, and section 4 describes the data and the definition of comparison groups. Section 5 contains a description of the empirical specification, section 6 provides a graphical analysis, and section 7 shows the results. Finally, section 8 concludes.

2 Description of the activation programs

This section will give a short background and description of the activation programs. As was described in the introduction, the starting point for the implementation of the programs was the 1998 revision of the Social Service Act. The Act gave town districts and municipalities increased authority to demand that recipients of welfare participate in activities such as job training or other labour market related activities.[3] The law was first and foremost intended for young persons under the age of 25, but has in practise been applied to all individuals capable of working, regardless of age (see Socialstyrelsen (2005)).

The new regulation has been used by several town districts and municipalities to implement activation programs[4]. These are targeted to recipients of welfare that are

[3] See 4-5§ in the 4th chapter of the Social Service Act (SoL 2001:453).

[4] Salonen and Ulmestig (2004) estimate that there were about 800 programmes of this type in 2002, which means that a municipality often has several different types of programs. For example, there is often a special program for young persons under the age of 25.

capable of working, and generally consist of scheduled job search combined with job training. Non-compliance with the program requirements results in total or partial withdrawal of welfare benefits.

Local labour market programs for recipients of welfare existed also before the revision of the Social Service Act. What differentiates the activation programs from the previous programs is first and foremost the clear connection between program participation and receipt of welfare benefits. In addition, while the previous programs were often targeted to some subgroup of benefit recipients, such as immigrant women, the new activation programs in general encompass all recipients of welfare that are able to work.

As described in the previous section, this paper focuses on the town districts in the municipality of Stockholm. During the period under study there are 18 town districts in Stockholm[5], each of which is run by a political board. The town district is the lowest administrative unit, and is responsible for the implementation of the greater part of municipal services, including social services. The welfare benefit norm is however set at the municipal level, and is hence the same across all town districts in our sample. It is hence only whether the town district has an activation program or not that differs between the districts in our sample, and not the level of benefits received. This means that we can identify effects on migration of stricter rules for benefit eligibility separately from effects of different benefit generosity.

The information that is used to define the starting year of the activation programs in this study was gathered in a survey to the social service units in all town districts in the municipality of Stockholm. The survey contains questions on the starting year and basic contents of the activation programs, as well as on local labour market programs for recipients of welfare that were in place during the years preceding the revision of the Social Service Act.[6] The surveys were in most cases complemented with telephone interviews.[7]

Based on the survey information, we define a town district as having an activation program if it has a program: 1) that has scheduled activity daily or almost daily; 2) that encompasses all individuals capable of working; and 3) where receipt of welfare benefits is strictly conditional on program participation. (More information on the town districts of the analysis will be given in section 4.)

Since the detailed contents of the activation programs vary across town districts, it is difficult to give a detailed over-all picture of the activation programs in our sample. However, we will here give a short description of one of the most well-known cases, the activation program of the town district Skärholmen, in order to illustrate

[5] In 2007 the number of town districts was reduced to 14.

[6] The survey form can be found in the appendix.

[7] Additional information was obtained for the following town districts: Kista, Rinkeby, Spånga-Tensta, Hässelby-Vällingby, Enskede-Årsta, Farsta, Vantör, Hägersten och Skärholmen.

the contents of a typical activation program. Skärholmen was one of the first town districts to start an activation program, and has served as model for other town districts and municipalities.

The Skärholmen activation program requires participants to spend 3 hours daily in program activities, in a rotating schedule that alternates between mornings and afternoons, in order to complicate black work outside the program. The first period in the program is spent on individual job search in the facilities of the program. Each participant is assigned a personal job coach, who provides individual job search assistance. The program furthermore provides computers for job search on the internet and for writing job applications, and the participants can use telephone and envelopes and postal stamps free of charge. If the program participant fails to find a job during this period, he/she gets a job training proposal from the program officials. This can be in the street cleaning team or some other activity that is arranged within the program, or it can be at an ordinary workplace. There is no limit on the time period that an individual can participate in the program.[8]

Activation programs in other town districts are similar to the Skärholmen case in the broad design of the program. However, features such as the required attendance varies across districts. This can be seen in Table 1, which shows the starting year and minimum hours of weekly attendance of the activation programs in the town districts of our sample. The six richest town districts of the municipality are excluded from the analysis, since the share of welfare recipients in these districts is very low[9]. In addition, one town district, Skarpnäck, is excluded due to the difficulties of defining a starting year for the activation program. The sample hence consists of 11 town districts.

Table 1: Starting year and weekly required attendance of activation programs

Town District	*Activation program*	*Hours/week*
Rinkeby	1998	8
Skärholmen*	1999	15
Kista	2001	9
Farsta	2001	4
Älvsjö	2002	15
Spånga-Tensta**	2003	5
Liljeholmen	2003	15
Hägersten	2003	15
Hässelby-Vällingby	2004	8
Enskede-Årsta	2004	4
Vantör	2004	4

*The activation program in Skärholmen started on a small scale in the autumn of 1998. From 1999, the program however operated at a large scale, which is why we choose this as the starting

[8] See e.g. Ekström (2005) and Thoren (2005).

[9] These are Kungsholmen, Norrmalm, Östermalm, Maria-Gamla Stan, Katarina-Sofia and Bromma.

year.

**Spånga-Tensta had an ambitious local labour market program in place during 1997-2000, although this cannot be characterized as an activation program. We therefore test the robustness of the results for the exclusion of Spånga-Tensta throughout the analysis.

As can be seen in the table, the required hours of attendance varies significantly between the town districts, and it is therefore possible that the migration effects of the program varies between districts. We will take account of this in the analysis by, in addition to using the full sample of town districts, also estimate the migration regression for only two town districts: one with a strict activation program, and one with no program during the period. It can be added that most of the town districts that have low hours of required attendance have implemented activation programs during the last year of, or even outside of, the sample period, which means that these will be used as control groups for having no activation program in the regressions (this is the case for example for Enskede-Årsta and Vantör, where only 4 hours of weekly attendance is required).

This section has given an overview of the activation programs in the Stockholm municipality. In the next section, we will analyse their potential effects on migration.

3 Theoretical framework

In order to analyse how the utility of a recipient of welfare benefits is affected by the introduction of an activation program, we develop a simple two-period model. In the model, the individual is either unemployed and receives welfare benefits, or is employed and receives a wage.[10] The activation program is assumed to affect the utility of a recipient of welfare in two ways: first, by decreasing the leisure time available to the individual, and second, by increasing the probability of finding a job in the next time period.

We start by assuming that the utility level for an individual who is unemployed and receives welfare benefits, in the case of no activation program, depends solely on the amount of leisure time, l, and the welfare benefit level, b:

$$U^i = u(l, b), \tag{1}$$

while an individual who has a job and does not receive welfare benefits has the fol-

[10] We hence assume that a recipient of welfare benefits does not work. This is a reasonable assumption, since the activation programs are directed to unemployed individuals. In addition, during 1994-2003, among all individuals in the municipality of Stockholm, aged 18-65, the share of employed among those that received welfare benefits at least some time during the year was 31%, to be compared to 77% among the corresponding population that did not recieve welfare benefits at any time during the year.

lowing utilily level:

$$U^i = u(l-h, y). \tag{2}$$

In equation (2), the amount of leisure is reduced with the time spent working, h, which is assumed to be constant, and y is the net of tax wage income, where $y > b$ is assumed to hold. Having a job hence gives a higher income but also reduces the leisure time.

Let us assume that the individual is an unemployed recipient of welfare benefits in period one, with a probability of having a job in period two equal to p. We assume that time preferences are captured by the time discount factor ρ, and write the expected two-period utility as:

$$U^i = u_t(l, b) + \frac{1}{1+\rho}\left[p u_{t+1}(l-h, y) + (1-p)\, u_{t+1}(l, b)\right] \tag{3}$$

How will the introduction of an activation program affect the utility level of our representative individual? The program affects the individual utility in two ways: First, it reduces the leisure time in case of unemployment by g, which is the time spent in the program. Second, participation in the program increases the probability of finding a job through the job search and job training activities, so p is also a function of the time spent in the program, g:

$$U^i = u_t(l-g, b) + \frac{1}{1+\rho}\left[p\,(g)\, u_{t+1}(l-h, y) + (1-p\,(g))\, u_{t+1}(l-g, b)\right] \tag{4}$$

The effect of the program on the utility level of the individual is shown in equation (5), where equation (4) is differentiated with respect to g. In period one, there is a negative effect through the reduction in the amount of leisure time. In period two, there are however both positive and negative effects: the increased probability of finding a job and hence having a higher income has a positive effect on the utility level, whereas if the individual remains on welfare benefit the effect in period 2 is negative, as in period 1, through the reduction in leisure time. If the utility levels in case of unemployment and employment are equal, i.e. equations (1) and (2) equal, the effect is unambiguously negative, while if the utility of working is higher, the total effect can be either positive or negative. It can be noted that the more the individual values current against future utility, reflected in a higher value of ρ, the more relative weight will be given to the negative effect of the activation program on the current utility level.

$$\begin{aligned} \frac{\partial U^i}{\partial g} \quad = \quad & -\frac{\partial u_t(l-g, b)}{\partial\,(l-g)} + \\ & +\frac{1}{1+\rho}\left[p'\,(g)\, u_{t+1}(l-h, y) - \left(p'\,(g)\, u_{t+1}(l-g, b) + (1-p\,(g))\,\frac{\partial u_{t+1}(l-g, b)}{\partial\,(l-g)}\right)\right] \end{aligned} \tag{5}$$

Unless the positive and the negative effects cancel, the introduction of an activation program in the jurisdiction is hence expected to affect the utility level of the

individual, negatively or positively. Will this affect the moving pattern of welfare prone individuals? If the utility-differential between living in a town district with and without an activation program is sufficiently large to outweigh the cost of moving, it is possible that the introduction of activation programs in some of the town districts will give rise to migration of the welfare-receiving population.

We will then see more moves *from* the districts that have activation programs if the total effect on utility is *negative*, and *to* the same districts if the total effect on utility is *positive*. This is illustrated in equation (6), which shows the utility levels of town districts A and B, where B has an activation program, while A does not. The moving cost is denoted c. Importantly, since all town districts belong to the same labour market as well as have the same benefit level, so that y and b are the same in all districts, nothing else is assumed to affect the moving choices of the individuals. That is, it is only the presence of the activation program, g, that differs between U_B and U_A.

$$
\begin{aligned}
& If \quad U_B - U_A > c \quad \Longrightarrow move\ to\ district\ B,\ and\ v.v. \qquad (6)\\
= \quad & \left(u_{B,t}(l-g,b) + \frac{1}{1+\rho}\left[p\left(g\right) u_{B,t+1}(l-h,y) + \left(1-p\left(g\right)\right) u_{t+1}(l-g,b) \right] \right) \\
& - \left(u_{A,t}(l,b) + \frac{1}{1+\rho}\left[p u_{A,t+1}(l-h,y) + \left(1-p\right) u_{t+1}(l,b) \right] \right) \\
> \quad & c
\end{aligned}
$$

Can we say anything about which result is more probable in practise - migration *to* of *from* a town district that has an activation program? Blomberg et al. (2006) have studied the attitudes among activation program participants in six Stockholm town districts, five out of which are included in the analysis of this paper. Their survey results give a mixed picture: while around half of the respondents are over-all positive to the services of the programs, the beliefs in the possibilities of the program to actually help them find a job is quite low: over half of the repondents think that the possibilities of the program to help them find a job are "very small" or "quite small". Less than a third believe that the chances of getting a job have increased due to the program. Furthermore, about 40 percent state that they would not take part in the program if participation were not mandatory for benefit receipt, while 30 percent state that they would.

The attitudes among participants also seem to vary between town districts. The results in Blomberg et al. (2006) suggests that the residents in Skärholmen are most dissatisfied with the activation program. Considering the findings in Thoren (2005), one might suspect that this is due to a lack of resouces. She argues that the personal job-search coaches in the activation program of Skärholmen have too many clients and that there are too few computers available to enable efficient job search. If this is the case, there may even be negative effects of program participation on the probability

of finding a job in this case.

4 Data

The study uses individual register data, which contains information on the amount of welfare benefits received, age, sex, country of birth, education level, disposable income, family situation (civil status and number of children), and employment status for all individuals aged 18-65[11]. The data covers 10 years of pooled cross-sections during the period 1994-2003.

4.1 Town districts

We start by giving a short description of the town districts in the study, and then move on to describing the individual level data. As was mentioned in section 2, the six richest town districts are excluded from the analysis, since the share of welfare recipients in these districts is very low[12], and one town district, Skarpnäck, is excluded due to the difficulties of defining a starting year for the activation program.

Table 2 repeats the starting year of the activation program, and shows descriptive statistics for a set of socio-economic characteristics, for the remaining 11 town districts. The variables are based on our register data on individuals aged 18-65, and show the average values over the period 1994-2003. *Welfare* denotes the share of the individuals that received welfare benefits at some point during the year, *Move out* is the share that moved from the town district to some other district within the municipality of Stockholm, *Pop 18-65* is the number of individuals aged 18-65, and *Immigr* shows the share of individuals that are born outside Sweden. *Disp Income* is the average disposable income of the town district, and *Empl* denotes the share of employed.

[11] Data on individuals comes from the Institute for Labour Market Policy Evaluation.

[12] These are Kungsholmen, Norrmalm, Östermalm, Maria-Gamla Stan, Katarina-Sofia and Bromma.

Table 2: Town district characteristics

Town District	*Activation program*	*Welfare*	*Move out*	*Pop 18-65*	*Immigr*	*Disp Income*	*Empl*
Rinkeby	1998	0.35	0.07	86855	0.80	83500	0.41
Skärholmen*	1999	0.14	0.06	169228	0.42	118500	0.64
Kista	2001	0.17	0.05	174231	0.53	120600	0.63
Farsta	2001	0.11	0.05	243016	0.21	133200	0.71
Älvsjö	2002	0.05	0.06	112856	0.15	147400	0.78
Spånga-Tensta**	2003	0.18	0.06	188605	0.47	122700	0.63
Liljeholmen	2003	0.07	0.09	180212	0.18	142300	0.74
Hägersten	2003	0.06	0.07	168880	0.17	144700	0.75
Hässelby-Vällingby	2004	0.07	0.05	322649	0.20	145000	0.76
Enskede-Årsta	2004	0.06	0.08	258024	0.18	145200	0.76
Vantör	2004	0.13	0.07	192120	0.28	126700	0.69

*The activation program in Skärholmen started on a small scale in the autumn of 1998. From 1999, the program however operated at a large scale, which is why we choose this as the starting year.

**Spånga-Tensta had an ambitious local labour market program in place during 1997-2000, although this cannot be characterized as an activation program. We therefore test the robustness of the results for the exclusion of Spånga-Tensta throughout the analysis.

As can be seen in the table it was generally the poorer town districts, with high rates of welfare recipients, low employment rates and a high share of immigrant population, that started activation programs early on. This suggests that it may be important to control for district-specific factors that can have affected the decision to start a program and that are at the same time correlated with migration. As will be further discussed below, we will use several difference-in-differences based approaches that control for town districts-specific fixed effects and town districts-specific time trends. In addition, district-specific covariates for the share of immigrant population, average disposable income and employment level will be included in some of the specifications.

4.2 Target and comparison groups

As discussed the in the introduction, we will follow the previous literature on welfare migration and compare the moving choices of more and less welfare prone individuals. How shall the more welfare prone target groups and the less welfare prone comparison groups be defined? Meyer (2000) points out that one should avoid defining the target and comparison groups based on actual benefit receipt, since this can give rise to so called participation bias. This type of bias arises since the payoff of applying for welfare benefits varies with the benefit policy of the jurisdiction. In the case of our town districts, it is possible that applying for welfare benefits is less attractive in town districts that have an activation program, since benefit receipt in this case requires active participation in the program. This means that individuals that did not receive

welfare benefits in a more strict town district, may choose to apply for benefits once they are in a less strict district, even though the motives for moving there were not related to the local welfare benefit policy.

Most studies deal with this type of bias by defining target and comparison groups that differ in welfare propensity based on characteristic that are *not* affected by the welfare benefit generosity. We follow this and compare the migration responses to differences in welfare benefit policy in several groups that differ in the likelihood of being recipients of welfare benefits. The hypothesis is that more welfare-prone groups will respond to policy differentials by moving, while individuals that are comparable in every sense except being less welfare prone, will not.

Ideally, we would like to compare individuals that are similar in every sense but the likelihood to seek welfare benefits. However, if we make the comparison groups too similar, we risk to also eliminate differences in welfare-propensity. We hence face a trade-off: on the one hand we want the groups to be sufficiently similar to eliminate the risk for omitted variable bias, on the other hand, sufficiently different to capture differences in welfare-propensity. The same trade-off applies to the question of how many individual covariates that shall be included in the regressions. We want to control for all characteristics that differ between the groups and that may affect the moving decision, but not for important determinants for the likelihood of receiving welfare. Our strategy is to use several comparison groups, which differ from the welfare prone group to varying degrees, and to show results both with and without individual covariates.

As suggested by Meyer (1995) using several comparison groups can also be useful as a means to reduce the risk of bias due to unobserved group-specific trends. The idea is that if the comparison groups are sufficiently different from each other, then we can also expect them to yield different biases. Similar results from different comparison groups hence strengthen the case that the result is due to the introduction of the activation program, and not just the effect of some omitted factor. However, the fact that we want all comparison groups to be comparable to the target group, as previously discussed, naturally puts a limit to how much the comparison groups can differ.

Based on these considerations, we define a set of target and control groups, based on factors that affect the probability to receive welfare benefits, but that are not affected by the welfare benefit policy. In addition, we base our comparison groups on factors that predict long-term welfare dependency. According to our data, and to Spahic (2002), an individual is more likely to be a long-term welfare recipient if he/she is: young, foreign-born, a single mother, low-educated, or socially unstable.[13]

[13] Spahic defines long-term recipiency as receiving benefits during at least 10 months during a period of 2 years, while we look at those that receive welfare benefits during both 1996 and 1995, or

Based on this information, we define two categories of welfare prone target groups: first, being a Swedish-born, single mother (with children living at home), and second, being born in a non-Western country[14].

The comparison groups for the two sets of target groups are defined as follows: First, we compare our group of single mothers with single or cohabiting women without children, as well as with married or cohabiting mothers.[15] As in the case of single mothers, we only include Swedish-born individuals. Second, we compare individuals born in a non-Western country with individuals born in a Western country (except Sweden), and with individuals born in Sweden, respectively. Since it is plausible that the migration pattern of recent immigrants differs from other residents', we exclude those that have immigrated during the last 3 years from the sample.[16] In addition, during the first years in the country, refugees are in general entitled to compensation for participation in Swedish and introductory courses. The compensation is in about the same amount as the welfare benefit level, and is included in our data on welfare benefits. Unless we exclude recent refugees from the sample, our data will therefore overstate the likelihood that an individual born in a foreign country is a recipient of welfare.

Table 3 shows the average welfare participation rates and the average migration rates for the different groups over the period 1994-2003, using data on all individuals aged 18-65 in the 11 town districts of our sample. Column 1, *Obs*, shows the share of individuals that receive welfare benefits (of any amount), and column 2, *Welfare*, shows the share that moved from the town district, but stayed within the Stockholm municipality, during the year. As can be seen in the table, the likelihood of being a recipient of welfare benefits is clearly higher for the more welfare prone groups.

both 1996 and 1993, irrespective of the time on welfare.

[14] This category contains all countries except Europe, North America and Oceania.

[15] The reason for including cohabiting women in the former group, is that our data does not allow us to separate single women without children from cohabiting women without children.

[16] One reason for this is that the decision of where to settle may change as the information about the new country increases, and that this may lead to more moves taking place during the first years. In addition, refugees to Sweden in the early 1990s were not free to decide the municipality of placement. They were however free to move immediately after placement. We can therefore expect some adjustments in the settlements of refugees, for example moves to municipalities with a large number of nationals.

Table 3: Description data on welfare prone and comparison groups.

Comparison group	*Obs*	*Welfare*	*Move out*	*Age*	*Education*	
					Low	*High*
Single mothers	115446	0.15	0.06	38.6	0.51	0.24
Single/cohabiting women without children	264293	0.05	0.09	39.6	0.49	0.34
Married/cohabiting mothers	258828	0.04	0.04	38.7	0.47	0.38
Born in:						
Non-western country	314030	0.29	0.07	37.9	0.41	0.26
Western country	225525	0.11	0.04	45.8	0.44	0.27
Sweden	1476493	0.06	0.06	40.61	0.48	0.32

The table also shows the average share that moves from the town district during a year, *Move out*; the average age of the individuals in our sample, *Age*; and the average shares with low and high levels of education, respectively, *Low* and *High*. Low education level is defined as having finished at most secondary education, while a high education level is defined as having finished higher education. There is some variation between the groups in these variables, particularly in the education level. A set of dummy variables capturing the age- and education structure will hence be included in some of the specifications in the regression analysis. The age dummy variables are defined as one dummy for each five-year age category, and the education dummy variables equal the variables for low and high education that are given in the table.

There is also a considerable variation in sample sizes. The smallest of the groups, Single mothers, contains about 115000 observations, while the largest group, Swedish-born, contains almost 1.5 million observations. These differences naturally affect the likelihood that a significant result is obtained. Since the hypothesis to be tested is that the activation program has an effect on the target group, but not on the comparison groups, it is however comfortable that the comparison groups are in all cases but one larger than the target groups.

5 Graphical analysis

Before moving on to the regression analysis, it is interesting to look at the migration pattern of the individuals in our sample graphically. By plotting the yearly outmigration rates from town districts that started activation programs early and late, and for the different target and comparison groups, respectively, we can see if a change in the moving choices is visible for welfare prone individuals after the introduction of the activation programs.

We start by dividing the town districts into four groups: The first group consists of

the early program-starters, Rinkeby and Skärholmen, who started activation programs in 1998 and 1999. We denote this group *Td99*, since we expect to see an effect on the migration of welfare prone individuals around 1999 in these districts. The second group is denoted *Td01*, for the same reason, and consists of Kista, Farsta and Älvsjö, out of which Kista and Farsta started activation programs in 2001, and Älvsjö started a program in 2002. Finally, we construct one group, *Td03*, with the town districts that started activation programs in 2003: Spånga-Tensta, Liljeholmen, and Hägersten; and one group, *Td04*, with the town districts that started activation programs in 2004 (i.e. outside the sample period of this study): Hässelby-Vällingby, Enskede-Årsta and Vantör.

Table 4 shows the average number of observations over the 10-year period for the target and comparison groups, divided into the four town district groups that are shown in the graphs. As can be seen in the table, the number of observations varies between the groups: single mothers in town district group *Td99* has the smallest number of observations, 1,260, while the largest number of observations, 60,730, is given for Swedish-born in town district group *Td04*.

Table 4: Average yearly number of observations graphs.

Comparison group	*Town district group*			
	Td99	*Td01*	*Td03*	*Td04*
Single mothers	1260	3030	2760	4580
Single/cohabiting women without children	1680	6030	7920	10800
Married/cohabiting mothers	2120	6840	6160	10750
Born in:				
Non-western country	8060	8770	7310	7250
Western country	4140	5540	5710	7170
Sweden	11480	36850	38590	60730

Graphs 1-6, which can be found in the Appendix, show the trends in the share that moves from the town district for the four groups of town districts. As in the rest of the paper, only moves within the Stockholm municipality are included. Separate graphs are shown for each of the target and control groups.

The graphs give no clear indication of a change in the migration decisions of welfare prone individuals after the introduction of activation programs in the town district groups. On the contrary, the over-all impression is that the outmigration for the respective group follows relatively similar trends in the four town district groups.

In spite of the lacking evidence on welfare migration from the graphs, we move on to the regression analysis, where we have better possibilities to control for other factors, such as district specific trends that may affect the migration choices of individuals.

6 Estimation strategy

As discussed in the previous sections, the data contains different sources of variation - between town districts and between groups of individuals. We start by using the town district variation in a difference-in-differences analysis (DD), and then add the group variation in welfare propensity to construct a difference-in-difference-in-differences estimator (DDD).

Before the law revision of 1998, there was no activation program in any town district. After 1998, most town districts have chosen to implement activation programs, but the starting year varies between districts, as was illustrated in Table 2. This means that we can compare the migration rates before and after the implementation of activation programs in the different town districts.

It is illustrative to describe this estimation strategy in a table. Let us, for simplicity, assume that there are only two town districts and two time periods. Let us also assume that one of the districts, denoted *Program*, starts an activation program in period 2, while the other district, *No program*, does not. How is outmigration from the *Program*-district affected by the start of the activation program? One way to measure this could be to look at the difference in outmigration before and after the program start, in the town district that starts a program, i.e. $(B - A)$ in Table 5. This estimate however also captures other factors that change between the two periods, and is hence likely to give a biased measure of the program-effect.

The DD-method is based on the idea that the influence of other factors can be controlled for by comparison with a town district which is comparable in every aspect that affects outmigration, but that has not implemented the program. An unbiased estimate of the effect of the activation program on outmigration can hence be obtained by taking the difference in the migration change over the two periods between the *Program*-district and the *No Program*-district, $(B - A) - (D - C)$.

Table 5: Description Difference-in-differences

Period / Town district	*Before*	*After*	After-Before
Program	A	B	$(B - A)$
No program	C	D	$(D - C)$
Difference-in-differences (DD):			$(B - A) - (D - C)$

The DD-estimator in Table 5 hinges on the assumption that any unobserved trends in migration are the same in both town districts. By adding the comparison of groups that differ in the propensity to receive welfare benefits to the analysis, this assumption can be relaxed. This is done by taking the difference of the differences-in-differences-estimates for the target and the control group, i.e. $DDD = DD_T - DD_C$, as is illustrated in Table 6. The idea is that the non-welfare prone group

will be similarly affected as the welfare prone group by unobserved town districts factors, but unaffected by the activation program. Subtracting the outmigration rates of this group will hence control for effects of unobserved town specific trends, and the resulting estimate will measure only the effect of the activation program. The important assumption in this case is that any unobserved town district-specific factors affect the migration-decisions in both groups similarly.

Table 6: Description Difference-in-difference-in-differences

	Period / Town district	*Before*	*After*	After-Before
Target	*Program*	A_T	B_T	$(B_T - A_T)$
	No program	C_T	D_T	$(D_T - C_T)$
DD_T :				$(B_T - A_T) - (D_T - C_T)$
Control	*Program*	A_C	B_C	$(B_C - A_C)$
	No program	C_C	D_C	$(D_C - C_C)$
DD_C :				$(B_C - A_C) - (D_C - C_C)$
$DDD = DD_T - DD_C$:				$((B_T - A_T) - (D_T - C_T))$ $- ((B_C - A_C) - (D_C - C_C))$

The DD- and the DDD-estimators will be used to estimate the effect of the activation programs on outmigration from the town districts. The description above assumed only two town districts and two time periods. In this study, there are several town districts, and they start activation programs during different years. The intuition behind the DD-estimator for our data is however the same as for the two-period case; i.e. it controls for time-varying factors that affect all town districts similarly, and it controls for fixed town district characteristics. The DDD-estimator furthermore controls for town district-specific trends, through the inclusion of district-by-year fixed effects.

The resulting DD-estimation equation, corresponding to Table 5, for our pooled cross-section for the individuals in 11 town districts over 10 years, is given in equation (7):

$$probit\,(move_{ijt}) = \beta_0 + \beta_1 A_j + \beta_2 D_t + \beta_3 prog_{jt} + \beta_4 X_{ijt} + \beta_5 Z_{jt} + \varepsilon_{ijt}. \qquad (7)$$

In equation (7), the dependent variable $move_{ijt}$ is a dummy variable which equals one if individual i moves out of town district j in year t, and zero otherwise. (As was described in the previous sections, only moves within the municipality of Stockholm are included.) The main explanatory variable is the dummy variable $prog_{jt}$, which equals one if town district j has an activation program in year t, and is zero otherwise. A positive value of β_3 hence indicates that more individuals move out of the town district after the start of the program, while a negative coefficient value indicates that less individuals move out of the district when there is an activation program in place.

Fixed town district effects are denoted A_j and year effects are denoted D_t. Finally, a set of individual covariates for the age- and education level, X_{ijt}, is included, as well as a set of town district level covariates, Z_{jt}.

The DDD-estimator that adds the group based comparison to the DD-estimator and which corresponds to Table 6, is given in equation (8)[17]:

$$\begin{aligned} probit\left(move_{ijt}\right) &= \beta_0 + \beta_1 T_{ij} + \beta_2 A_j + \beta_3 D_t + \beta_4 \left(T_{ij} \cdot A_j\right) \\ &\quad + \beta_5 \left(T_{ij} \cdot D_t\right) + \beta_6 \left(A_j \cdot D_t\right) + \beta_7 prog_{jt} \\ &\quad + \beta_8 \left(T_{ij} \cdot prog_{jt}\right) + \beta_9 X_{ijt} + \beta_{10} Z_{jt} + \varepsilon_{ijt}. \end{aligned} \tag{8}$$

In equation (8), coefficient β_8 is of primary interest. It measures the extent to which the migration response to an activation program differs between the target and the control group, where the dummy variable T_{ij} is one if the individual belongs to the more welfare prone target group. As in equation (7), A_j and D_t denote town district and year specific fixed effects, X_{ijt} contains individual age- and education dummy variables, and Z_{jt} denotes a set of town district covariates. Equation (8) furthermore includes the second-order interactions between the control group dummy and the district and year fixed effects, $(T_{ij} \cdot A_j)$ and $(T_{ij} \cdot D_t)$. Finally, town-district specific year effects, $(A_j \cdot D_t)$, control for year effects that differ between the town districts.

The following section present the results of the estimations of equations (7) and (8).

7 Results

This following sections show the results from running the estimations described in the previous section. Results will be given both for the full sample of 11 town districts, and for alternative samples.

7.1 Full set of town districts

We start by estimating the DD-equation in (7) on our two groups of welfare prone individuals: single mothers and individuals born in non-Western countries, respectively, using the data set on the 11 town districts that were described in section 4.1. As discussed in the previous sections, we expect that the moving decisions of welfare prone individuals will be affected by the activation program, however, the direction of the effect will depend on whether the individual views the program as something primarily negative or positive.

It is informative to also estimate equation (7) for the control groups. Since these are less welfare prone than the target groups, we expect to see smaller effects of the

[17] Similar estimation strategies are used in e.g. Yelowitz (1995) and Ruhm (1998).

program on these groups. A different result is an indication either of miss-specification of the control- and target-groups, or of bias due to some omitted town district specific factor, which is correlated with the district's decision to start a program.

The results of the DD-estimations are shown in the first sections of Table 7 and 8. The tables show the marginal effects of the activation program on the probability to move from the district, for an individual with average characteristics.[18] The results come from separate regressions for each target and comparison group. The specification in column (1) includes the activation program dummy together with fixed town district specific effects and year effects. In specification (2), the individual age-and education dummy variables are added, and specification (3) also includes the town district specific covariates.

Table 7 and 8 also show the results from the DDD-estimation in (8).[19] As was discussed in section 6, an advantage with this approach is that the differential effects that are obtained are not affected by town district-specific trends that affect the target and control group similarly. The DDD-estimates were obtained by running separate regressions for each comparison group together with the relevant target group, in order to facilitate the interpretation of the coefficients.

For the DDD-estimator the specification including town district covariates is however dropped due to muliticollinearity. The effects of these variables are probably picked up by the town district-specific year effects.

[18] The results for the probit coefficients for the full set of covariates can be seen in Tables A.1-A.2, Appendix. Note that all tables in the Appendix show the probit coeffcient estimates, and not the marginal effects.

[19] The results for the probit coefficients for the full set of covariates can be seen in Tables A.3-A.4, Appendix.

Table 7: Probit Estimates, Marginal Effect of Activation Program on Outmigration, Target and comparison groups based on civil status and motherhood

		(1)	(2)	(3)
DD-estimates	Single mothers (SM)	-.004	-.005**	-.005*
		(.0028)	(.0025)	(.0028)
	Log pseudolikelihood	-26388.706	-24831.946	-24831.708
		n=115446	n=114278	n=114278
	Single/cohabiting women (SW)	.002	-.001	.001
		(.0024)	(.0022)	(.0023)
	Log pseudolikelihood	-81442.557	-74977.421	-74971.562
		n=264293	n=262748	n=262748
	Married mothers (MM)	-.002	-.003**	-.002
		(.0015)	(.0013)	(.0014)
	Log pseudolikelihood	-41023.44	-37508.565	-37502.588
		n=258828	n=257565	n=257565
DDD-estimates	SM-SW	-.008*	-.006	
		(.0040)	(.0037)	
	Log pseudolikelihood	-107773.38	-99880.784	
		n=379739	n=377026	
	SM-MM	-.001	-.000	
		(.0028)	(.0025)	
	Log pseudolikelihood	-67364.334	-62364.931	
		n=374274	n=371843	
	Controls:			
	individual level covariates	no	yes	yes
	town district level covariates	no	no	yes

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects and year fixed effects are included in all specifications. The DDD-specifications also includes district-by-year-effects.

Table 8: Probit Estimates, Marginal Effect of Activation Program on Outmigration, Target and comparison groups based on country of birth

		(1)	(2)	(3)
DD-estimates	*Country of birth:*			
	Non-Western country (NW)	-.001	-.002	.001
		(.0016)	(.0016)	(.0017)
	Log pseudolikelihood	-75683.204	-68675.945	-68666.048
		n=314030	n=290012	n=290012
	Western country (W)	-.002	-.002	.001
		(.0016)	(.0015)	(.0017)
	Log pseudolikelihood	-39218.682	-35883.777	-35872.489
		n=225525	n=214861	n=214861
	Sweden (S)	-.0003	-.002**	-.0004
		(.0008)	(.0007)	(.0008)
	Log pseudolikelihood	-332595.55	-307661.89	-307645.74
		n=1476493	n=1465322	n=1465322
DDD-estimates	NW-W	.003	.002	
		(.0026)	(.0025)	
	Log pseudolikelihood	-114808.89	-104535.62	
		n=539555	n=504873	
	NW-S	.001	.003	
		(.0019)	(.0018)*	
	Log pseudolikelihood	-408140.34	-376864.89	
		n=1790523	n=1755334	
	Controls:			
	individual level covariates	no	yes	yes
	town district level covariates	no	no	yes

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects and year fixed effects are included in all specifications. The DDD-specifications also includes district-by-year-effects.

The DD-results in Table 7 show a significant negative marginal effect of the activation program on outmigration of around -0.005 for single mothers, when individual characteristics are controlled for. This indicates that having an activation program reduces the probability that a single mother, with average characteristics, will move from the town district with 0.5 percentage points. This is a rather large effect, considering that the average yearly migration rate for this group is 6 percent. However, we also see a negative effect of about similar magnitude for married mothers in specification (2), although this effect is not significant as town district covariates are included. This indicates that the negative effect may not be due to the activation program, but to some other factor that affects single and married mothers alike.

Turning to the DD-estimates in Table 8, we see that negative marginal program effects are found also for the target and control groups based on country of birth, although in this case the effect is only significant in specification 2 for Swedish-born individuals. (That a significant effect is found for this group is not surprising, considering its large sample size.)

The DD-estimates hence yield negative point estimates for all groups, except for single/cohabiting women without children. Is it plausible that this is due to confounding district-specific trends that are correlated with the introduction of activation programs, and that affect outmigration in most of our target and control groups negatively? That is, can it be the case that the activation programs are started when town districts, for some reason, experience low outmigration? This is possible, although my prior beliefs would go the other way: that districts introduce stricter programs when the economic situation of the district is bad, and outmigration therefore plausibly high.

Turning to the DDD-estimates, the results in Table 7 show that the probability to move out of the town district is around 0.8 percentage points lower for single mothers if there is an activation program in place, compared to single/cohabiting women with no children. This effect is however only significant at the 10 percent level in specification (1), and turn insignificant as individual characteristics are controlled for. When comparing single and married mothers, no significant differences in the program effect are found.

The DDD-estimates in Table 8 show a positive outmigration effect of the program for individuals born in a non-Western, country compared to Swedish-born, which is marginally significant as individual covariates are included in the specification. In contrast to the result for single mothers, this indicates that welfare prone individuals are more likely to move from a town district that has an activation program, which is consistent with the hypothesis that recipients of welfare try to avoid the obligation to participate in the program by moving.

Although it could be the case that there are heterogeneous effects of the program on different groups of welfare recipients, a more prudent interpretation of the results is probably preferable. As pointed out by Meyer (1995), the fact that the results vary with the choice of comparison group, suggests that the effects may be due to some comparison group-specific and distict-specific omitted variable that is not controlled for in our estimations. One way to reduce the risk for this type of bias is to reduce the sample to districts that are as similar as possible in factors that are assumed to affect migration. In the next section, we do this by selecting only the six poorest of the town districts.

7.2 Sensitivity analysis: varying the set of town districts

7.2.1 Limiting the sample to the six poorest town districts

Table 2 in section 4 showed systematic differences in the socioeconomic variables between town districts that implemented activation programs early after the 1998 law revision, and town districts that implemented programs late during the period

under study. In order to obtain a more comparable group of districts, we select only the six poorest town districts and re-run the DDD-estimations using this sample. The resulting town districts are: Rinkeby, Skärholmen, Farsta, Kista, Spånga-Tensta and Vantör. As can be seen in Table 2, essential variation in the starting year for the activation program is kept in this sample: two of the town districts implement the program in 1998 and -99, two in 2001, and the remaining two in 2003 and 2004. The DDD-results for this sample are shown in Table 9.[20]

Table 9: Probit Estimates, Marginal Effect of Activation Program on Outmigration, Six town districts

		(1)	(2)
DDD-estimates	SM-SW	-.006	-.004
		(.0053)	(.0048)
	Log pseudolikelihood	-42151.517	-38660.239
		n=159596	n=158007
	SM-MM	-.002	.001
		(.0039)	(.0036)
	Log pseudolikelihood	-29805.839	-27764.847
		n=166058	n=164551
DDD-estimates	NW-W	-.001	-.001
		(.0030)	(.0030)
	Log pseudolikelihood	-76790.644	-69731.082
		n=373860	n=347412
	NW-S	-.001	-.000
		(.0021)	(.0020)
	Log pseudolikelihood	-187684.7	-172546.45
		n=866440	n=840647
	Controls:		
	individual level covariates	no	yes
	town district level covariates	no	no

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects, year fixed effects, and district-by-year-effects are included in all specifications.

As can be seen in Table 9, the marginal effects are fairly similar to the full-sample estimates of the previous section, but are never significantly different from zero. This can be due to the fact that the estimation power is reduced because of the smaller number of observations. It can however also be the case that the significant effects that were obtained in some specifications in the previous section, were due to some omitted factor that was related to differences in the economic situation between early and late program starters.

7.2.2 Comparing two representative town districts

The previous sections yield rather ambiguous results: some significant, albeit not too easily interpretable, effects are obtained when all eleven town districts are included.

[20] Note that the abbreviations used in the table are the same as in Table 7 and 8.

However, when only the six poorest of the town districts are analyzed, no significant effects of the program are found.

An issue to consider in light of these results, is that the design of the activation programs vary between town districts, as was commented in section 2. This means that the activation programs can differ in for example the number of hours of attendance that are required, or in the quality of the services that are offered in the program. This naturally affects the results.

In a final attempt to pin down the effects of stricter activation rules on migration, we select two districts that are as similar as possible in economic conditions, but that differ as much as possible in the strictness of the activation policy. The selected town districts are Skärholmen and Vantör. As can be seen in Table 2, these are fairly similar in all variables except the share of immigrants. However, Skärholmen can be described as having a has one of the stricter activation programs. As was described in section 2, it has among the highest requirement for the number of hours of attendance (a minimum of 3 h a day). In addition, as was discussed in section 3, studies point to a lack of resources to meet the needs of the program participants, as well as to more negative attitudes to the program among the participants compared to other town districts (see Blomberg et al. (2006) and Thoren (2005)). This suggests that recipients of welfare would probably prefer not to have to participate in the program. Vantör, on the other hand, has no activation program during the time period under study.[21] If there is a positive effect on outmigration of stricter rules in the form of activation programs, it is thus likely to turn up here.

The DDD-results for this sample of town districts are shown in Table 10.[22]

[21] As could be seen in Table 1, an activation program was started in Vantör in 2004.

[22] Note that the abbreviations used in the table are the same as in Table 7 and 8.

Table 10: Probit Estimates, Marginal Effect of Activation Program on Outmigration, Two town districts

		(1)	(2)
DDD-estimates	SM-SW	-.015	-.012
		(.0085)*	(.0078)
	Log pseudolikelihood	-18179.734	-16744.424
		n=63676	n=63188
	SM-MM	-.008	-.004
		(.0061)	(.0058)
	Log pseudolikelihood	-11671.782	-10890.232
		n=61131	n=60714
DDD-estimates	NW-W	-.013	-.014
		(.0047)***	(.0045)***
	Log pseudolikelihood	-21356.064	-19564.121
		n=108925	n=102791
	NW-S	-.007	-.006
		(.0041)*	(.0038)
	Log pseudolikelihood	-68705.251	-63300.987
		n=298062	n=291709
	Controls:		
	individual level covariates	no	yes
	town district level covariates	no	no

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects, year fixed effects, and district-by-year-effects are included in all specifications.

The results in Table 10 show no evidence of a migration effect of the activation program when we compare single mothers with single/cohabiting women with no children, or with married mothers. The differential migration effect for being a single mother, compared with a single woman with no children, is marginally significant in specification (1), but not as individual covariates are included. The coefficient for the program effect when comparing single and married mothers is never significantly different from zero.

A negative significant marginal effect is however estimated when individuals born in non-Western country and a Western country (except Sweden) are compared. This suggests that individuals of non-Western origin are about 2 percentage points less likely to move from a town district when there is an activation program in place, compared to individuals of Western origin. This is a large effect considering that the average migration rate for these groups in the full sample are 0.04 and 0.07 (see Table 3). The effect is in line with the hypothesis that welfare prone individuals value the services of the program, and hence want to stay in the town district to a higher extent when the program is in place, i.e. not the result we expected considering the negative attitudes among the activation program participants that were expressed in Blomberg et al. (2006). However, in order to make this interpretation, we would like to see a similar result for the difference in migration response when using Swedish-

born individuals as comparison group. Whereas there is a marginally significant effect in specification (1), this turns insignificant as individual covariates are included.

8 Concluding remarks

To conclude, we find some evidence indicating that the more welfare prone target groups are less likely to move from a town district that has an activation program, compared to the less welfare prone comparison groups. This result is primarily obtained in the comparison between single mothers and single/cohabiting women without children in the regression on 11 town districts, and in the comparison between non-Western-born and Western-born individuals in the regression on two town districts. This could be interpreted as evidence that welfare prone individuals are more likely to stay in town districts that have an activation program, possibly because they value the services of the program.

However, the over-all results suggest that this is too strong an interpretation of the results. While significant differences in the effects on more and less welfare prone individuals are obtained in several specifications, these vary with the choice of comparison group as well as with the sample of town districts, and it is hence probably wiser not to interpret the results as a result of the differences in local welfare benefit policy.

What we can say, however, is that the aggregate results show no evidence of a positive effect of the activation program on the outmigration of welfare prone individuals from the town districts, i.e. it does not seem that welfare prone individuals avoid the obligations to participate in the programs by moving. This is a reassuring result, which means that letting the town districts decide on the implementation and design of the activation programs has not led to harmful welfare-migration effects.

The non-significant results of this study contrast to the significant, albeit often economically small, effects found in other studies of welfare migration. What is the reason for this divergence? It is of course possible that the lack of welfare migration is specific to the sample used in this case. However, it is also possible that the difference in results is due to an omitted variable bias in the previous literature, which is not present in this study. As was explained in the introduction, the fact that this study uses variation in the welfare policy within a municipality, where that labour market conditions and other region-specific characteristics are the same for all individuals, greatly reduces the risk for omitted variable bias.

Acknowledgements

I am grateful for useful comments and suggestions from Matz Dahlberg, Sören Blomquist, Hanna Ågren, Eva Mörk, Björn Öckert, Tuomas Pekkarinen, Mikael Elinder and seminar participants at the Department of Economics, Uppsala University, and at The Institute for Labour Market Policy Evaluation (IFAU).

References

Blomberg, G., Ekström, V. and Rauhut, D.: 2006, Bidrag och motprestation - en uppföljning av arbetet med arbetslösa socialbidragstagare i sex stadsdelsområden i stockholm. FoU-rapport 2006:12.

Ekström, V.: 2005, Individens eget ansvar och samhällets stöd. en utvärdering av "skärholmsmodellen" vid jobbcentrum sydväst. FOU-rapport 2005:10, Utrednings- och statistikontoret (USK) Stockholms stad.

Fiva, J. H.: 2007, Does welfare policy affect residential choices? evidence from a natural experiment. Discussion Papers No. 503, Statistics Norway Research Department.

Gelbach, J. B.: 2004, Migration, the lifecycle, and state benefits: How low is the bottom?, *Journal of Political Economy* **112**(5), 1091–1130.

Kaestner, R., Kaushal, N. and Ryzin, G. V.: 2001, Migration consequences of welfare reform. NBER Working paper 8560.

McKinnish, T.: 2005, Importing the poor: Welfare magnetism and cross-border welfare migration, *Journal of Human Resources* **40**, 57–76.

McKinnish, T.: 2007, Cross-border welfare migration: New evidence from micro-data, *Journal of Public Economics* **91**, 437–450.

Meyer, B.: 2000, Do the poor move to receive higher welfare benefits? Unpublished Paper, Department of Economics, Northwestern University.

Meyer, B. D.: 1995, Natural and quasi-experiments in economics, *Journal of Business and Economic Statistics* **13**(2), 151–161.

Moffitt, R.: 1992, Incentive effects of the u.s. welfare system: A review, *Journal of Economic Literature* **30**(1), 1–61.

Ruhm, C. J.: 1998, The economic consequences of parental leave mandates: Lessons from europe, *The Quarterly Journal of Economics* **113**(1), 285–317.

Salonen, T. and Ulmestig, R.: 2004, Nedersta trappsteget: En studie om kommunal aktivering. Rapport 1:2004, Institutionen för vårdvetenskap och socialt arbete, Växjö universitet.

Socialstyrelsen: 2005, Hur tillämpas 4 kap. 4ğ socialtjänstlagen? krav på praktik eller annan kompetenshöjande verksamhet för ungdomar med försörjningsstöd.

Spahic, L.: 2002, En longitudinell studie över nya långvariga socialbidragstagare. Undergraduate Thesis, Department of Mathematics, Stockholm university.

Thoren, K. H.: 2005, Kommunal aktiveringspolitik: En fallstudie av det praktiska arbetet med arbetslösa socialbidragstagare. Report 2005:11, The Institute for Labour Market Policy Evaluation (IFAU).

Yelowitz, A. S.: 1995, The medicaid notch, labor supply, and welfare participation: Evidence from eligibility expansions, *The Quarterly Journal of Economics* **110**(4), 909–939.

A Appendix

A.1 Survey to the social service units of the town districts

(Note that the survey was conducted in Swedish, and that this is a translated version.)

The survey refers to information on activities for unemployed individuals, capable of working, that receive welfare benefits.

1. Does your town district currently have activation or other labour market related programs for unemployed individuals that are capable of working and that receive welfare benefits?

Yes ☐

No ☐

If no, turn to question 9 of the survey.

If yes, please name the program/programs:

2. Since which year does this program/programs exist in its current form (under the same or a different name)?

3. Does the program/s encompass all individuals, capable of working, that are unemployed and receive welfare benefits?

Yes ☐

No ☐

4. If you have responded "No" to question 3:

- How large a share of all individuals, capable of working, that are unemployed and receive welfare benefit are encompassed by the program?

- Which groups of individuals are targeted by the program?

5. Please specify how and to which extent the following activities are being used in the program/programs:

a. Job-seeking activities

b. Job training activities

c. Other assigned work (for example within the municipal services)

d. Other activities, please specify which:

6. What is the minimum number of hours of weekly attendance that is required in the program/programs?

7. Is absence systematically reported to the social service officials?

Yes ☐

No ☐

Comments:

8. Can absence (without acceptable cause) lead to rejection of the welfare benefit application?

Yes ☐

No ☐

Comments:

In the following part of the survey we ask for information on programs that were targeted to unemployed individuals that are capable of working and receive welfare benefits, before today's program/programs started.

9. Which programs have been in place under the period from 1990 until the start of today's program/programs? Under each number below, please specify the name of the program, or the main activity if a name does not exist, for example "Meeting with job counsellor". Please also specify during which years the program/activity was in place.

Program 1:

Name: _____

Time period:_____

Program 2:

Name: _____

Time period:_____

[..etc..]

Below follows a set of questions about the programs/activities that were in place before today's program/programs. Please, answer the questions about each program under the number that corresponds to the list above.

Program/Activity 1:

1. Which groups were targeted by the program/activity?

2. How large a share of all individuals, capable of working and receiving welfare benefits, were encompassed by the program/activity?

3. Please, specify to which extent the following activities were used in the program/activity:

a. Job-seeking activities

b. Job-training activities

c. Other assigned work (for example within the municipal services)

d. Other activities, please specify which:

7. Was absence systematically reported to the social service officials?

Yes ☐

No ☐

If yes, in which way:

8. Could absence (without acceptable cause) lead to rejection of the welfare benefit application?

Yes ☐

No ☐

Comments:

[The same set of questions were repeated for all programs/activities in the list.]

A.2 Graphs outmigration

Figure 1:

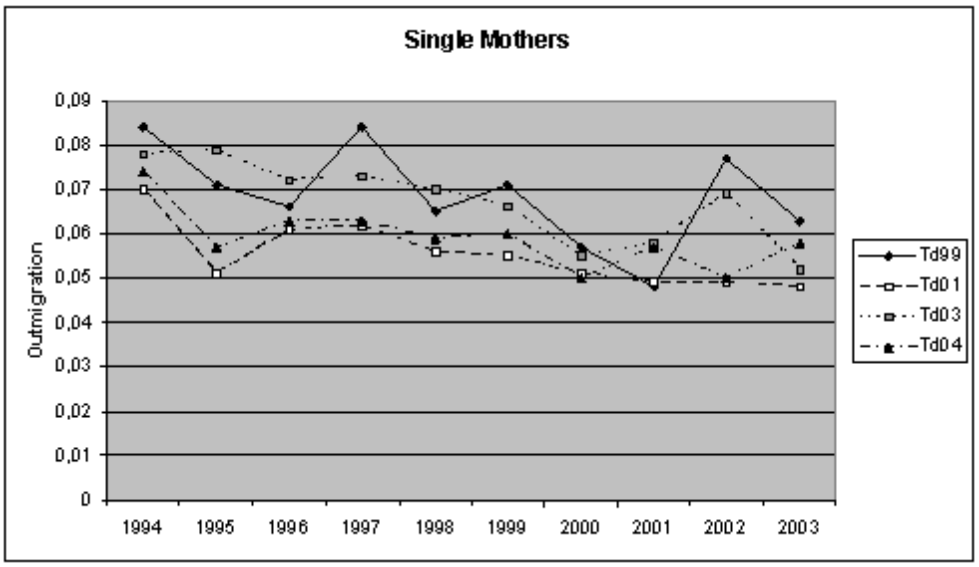


Figure 2:

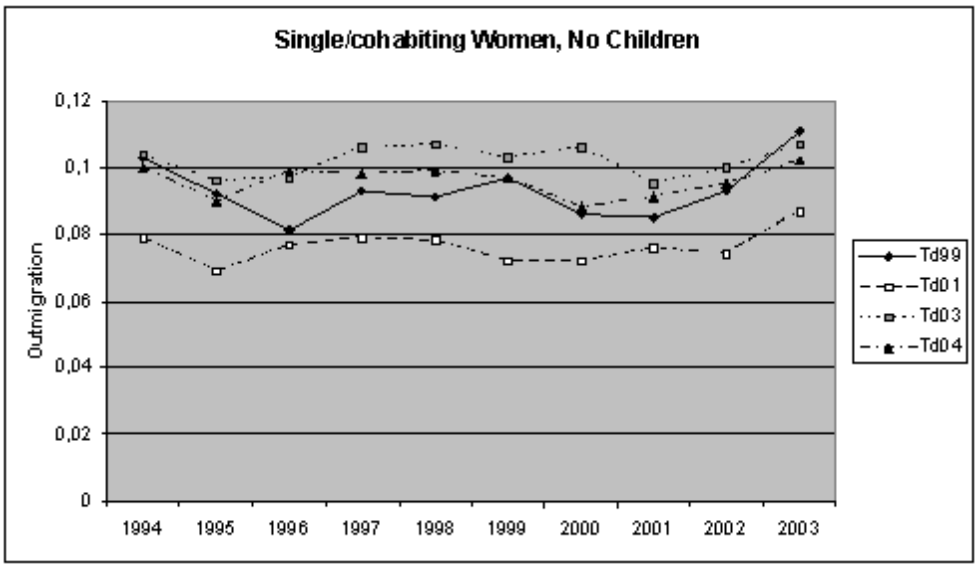


Figure 3:

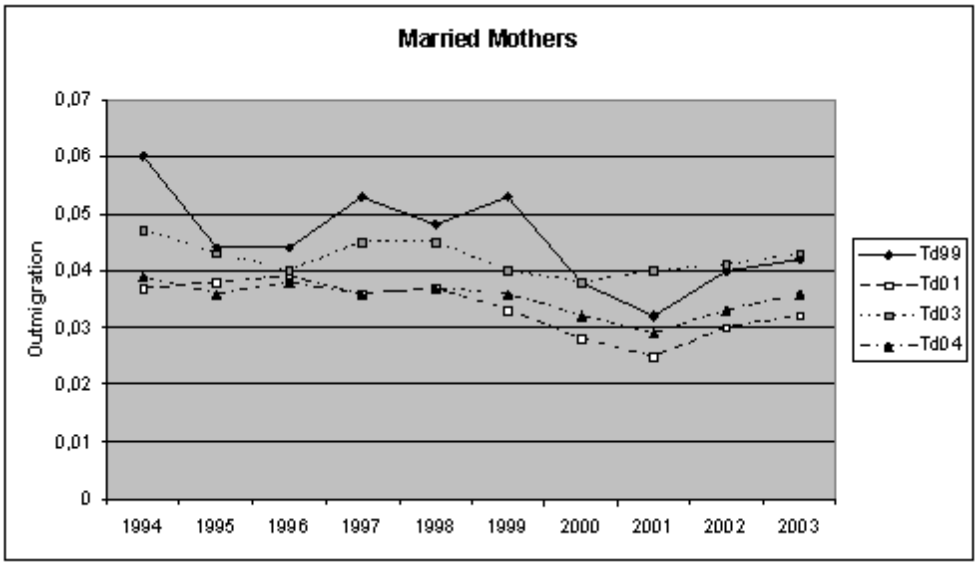


Figure 4:

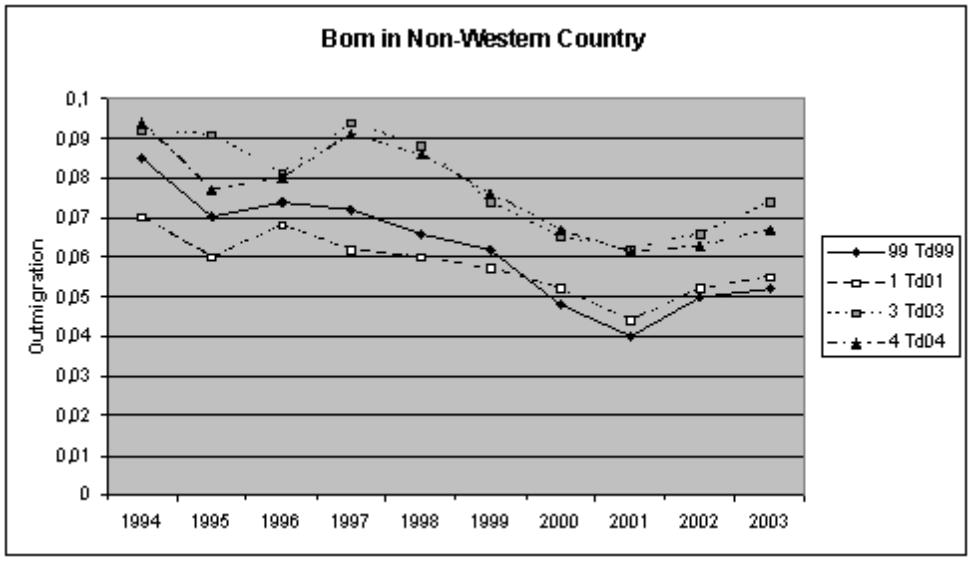


Figure 5:

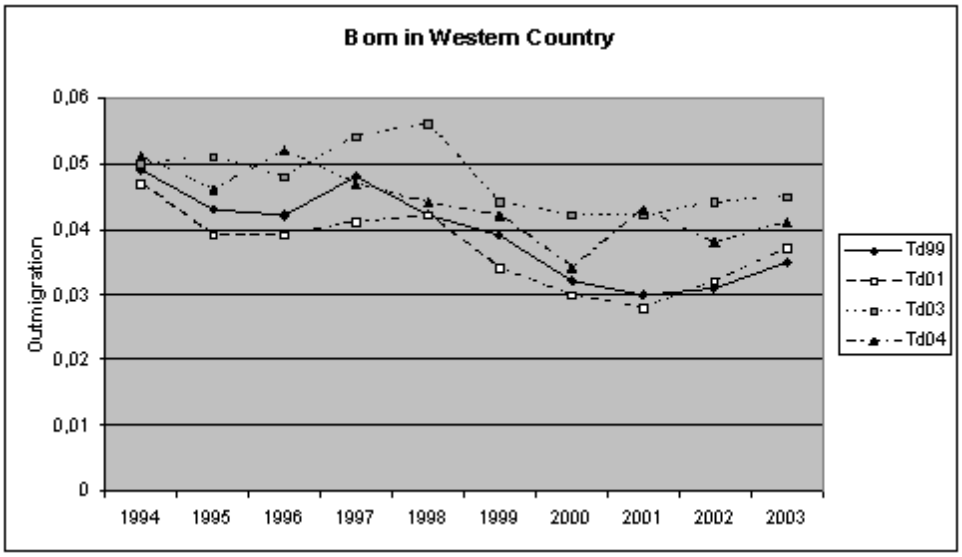
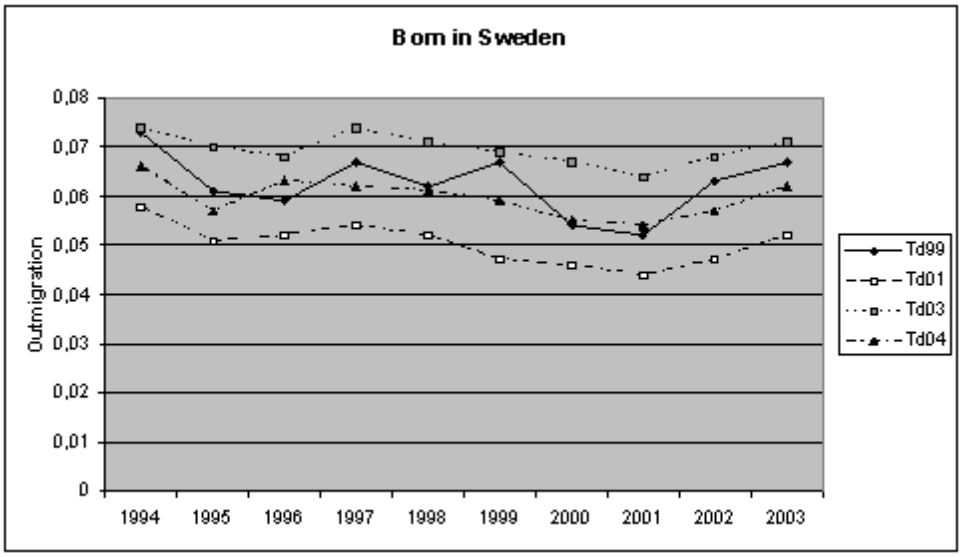


Figure 6:



A.3 Difference-in-differences full tables

Table A.1: Difference-in-Differences Probit Coefficients, Dependent variable: Outmigration

	Single mothers		**Single women**		**Married mothers**	
	(1)	(2)	(3)	(4)	(5)	(6)
Program	-.049**	-.048*	-.004	.009	-.051**	-.032
	(.025)	(.027)	(.015)	(.016)	(.021)	(.022)
Low Educ	.1***	.1***	.011	.012	.194***	.194***
	(.015)	(.015)	(.011)	(.011)	(.016)	(.016)
High Educ	.21***	.21***	.12***	.12***	.29***	.29***
	(.019)	(.019)	(.012)	(.012)	(.017)	(.017)
Age 2529	-.186***	-.186***	-.183***	-.183***	-.224***	-.223***
	(.021)	(.021)	(.01)	(.01)	(.017)	(.017)
Age 3034	-.518***	-.518***	-.397***	-.398***	-.669***	-.668***
	(.022)	(.022)	(.012)	(.012)	(.017)	(.017)
Age 3539	-.692***	-.692***	-.635***	-.636***	-.905***	-.905***
	(.022)	(.022)	(.015)	(.015)	(.018)	(.018)
Age 4044	-.763***	-.763***	-.793***	-.794***	-1.01***	-1.01***
	(.022)	(.022)	(.019)	(.019)	(.019)	(.019)
Age 4549	-.725***	-.725***	-.881***	-.882***	-.942***	-.942***
	(.022)	(.022)	(.019)	(.019)	(.019)	(.019)
Age 5054	-.713***	-.713***	-.94***	-.941***	-.955***	-.956***
	(.025)	(.025)	(.017)	(.017)	(.022)	(.022)
Age 5559	-.762***	-.762***	-1.071***	-1.071***	-.978***	-.979***
	(.033)	(.033)	(.018)	(.018)	(.031)	(.031)
Age 6065	-.938***	-.938***	-1.209***	-1.21***	-1.18***	-1.181***
	(.051)	(.051)	(.019)	(.019)	(.059)	(.059)
Town distr covariates:						
Employment		.682		-.093		-1.505*
		(1.017)		(.726)		(.83)
Income		5.60e-07		2.42e-06		5.72e-06**
		(3.12e-06)		(1.99e-06)		(2.52e-06)
Immigrants		.252		-.384		.035
		(.74)		(.496)		(.613)
Loglikelihood	-24831.946	-24831.708	-74977.421	-74971.562	-37508.565	-37502.588
Obs.	114278	114278	262748	262748	257565	257565

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects and year fixed effects are included in all specifications.

Table A.2: Difference-in-Differences Probit Coefficients, Dependent variable: Outmigration

	Non-Western		**Western**		**Sweden**	
	(1)	(2)	(3)	(4)	(5)	(6)
Program	-.013	.007	-.025	.008	-.016**	-.003
	(.013)	(.014)	(.02)	(.021)	(.007)	(.008)
Low Educ	.085***	.085***	-.001	-.001	.087***	.087***
	(.009)	(.009)	(.012)	(.012)	(.005)	(.005)
High Educ	.12***	.12***	.024*	.024*	.177***	.177***
	(.01)	(.01)	(.014)	(.014)	(.005)	(.005)
Age 2529	-.13***	-.13***	-.106***	-.107***	-.06***	-.06***
	(.014)	(.014)	(.026)	(.026)	(.006)	(.006)
Age 3034	-.253***	-.253***	-.267***	-.268***	-.361***	-.361***
	(.013)	(.013)	(.025)	(.025)	(.006)	(.006)
Age 3539	-.369***	-.369***	-.442***	-.443***	-.639***	-.639***
	(.013)	(.013)	(.025)	(.025)	(.007)	(.007)
Age 4044	-.433***	-.432***	-.553***	-.554***	-.785***	-.786***
	(.014)	(.014)	(.025)	(.025)	(.007)	(.007)
Age 4549	-.514***	-.513***	-.603***	-.605***	-.805***	-.806***
	(.016)	(.016)	(.024)	(.024)	(.008)	(.008)
Age 5054	-.574***	-.574***	-.649***	-.651***	-.821***	-.822***
	(.019)	(.019)	(.025)	(.025)	(.008)	(.008)
Age 5559	-.643***	-.643***	-.761***	-.763***	-.901***	-.901***
	(.025)	(.025)	(.026)	(.026)	(.009)	(.009)
Age 6065	-.697***	-.697***	-.901***	-.903***	-1.022***	-1.022***
	(.029)	(.029)	(.027)	(.027)	(.01)	(.01)
Town distr covariates						
Employment		.325		-.055		-.047
		(.454)		(.671)		(.32)
Income		5.99e-06***		9.06e-06***		3.47e-06***
		(1.51e-06)		(2.11e-06)		(9.28e-07)
Immigrants		.898***		1.295**		.068
		(.347)		(.518)		(.227)
Loglikelihood	-68675.945	-68666.048	-35883.777	-35872.489	-307661.89	-307645.74
Obs.	290012	290012	214861	214861	1465322	1465322

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects and year fixed effects are included in all specifications.

A.4 Difference-in-difference-in-differences full tables

Table A.3: Difference-in-Difference-in-Differences Probit Coefficients, Dependent variable: moving out of municipality, Reference group: Single mothers, 1994-2003

Omitted category:	**Single Women**		**Married Mothers**	
	(1)	(2)	(3)	(4)
Program	.021	-.052	-.045	-.134
	(.118)	(.091)	(.132)	(.097)
Program*Single Mother	-.055*	-.048	-.011	-.0003
	(.029)	(.03)	(.031)	(.033)
Single Mother	.041	-.039	.249***	.264***
	(.035)	(.037)	(.037)	(.039)
Low Educ		.055***		.152***
		(.009)		(.011)
High Educ		.169***		.248***
		(.01)		(.013)
Age 2529		-.18***		-.205***
		(.009)		(.013)
Age 3034		-.421***		-.608***
		(.01)		(.013)
Age 3539		-.657***		-.823***
		(.012)		(.014)
Age 4044		-.781***		-.913***
		(.014)		(.014)
Age 4549		-.812***		-.855***
		(.014)		(.014)
Age 5054		-.862***		-.856***
		(.014)		(.016)
Age 5559		-1.002***		-.887***
		(.016)		(.023)
Age 6065		-1.168***		-1.065***
		(.018)		(.038)
Log pseudolikelihood	-107773.38	-99880.784	-67364.334	-62364.931
Obs.	379739	377026	374274	371843

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects, year fixed effects, and district-by-year-effects are included in all specifications.

Table A.4: Difference-in-Difference-in-Differences Probit Coefficients, Dependent variable: moving out of municipality, Reference group: Individuals born in a non-Western country, 1994-2003

Omitted category:	**Western Country**		**Sweden**	
	(1)	(2)	(3)	(4)
Program	-.018	.189**	-.002	-.204***
	(.057)	(.088)	(.04)	(.048)
Program*Non-Western Country	.023	.021	.01	.029*
	(.023)	(.024)	(.016)	(.017)
Non-Western country	.217***	.085***	.064***	.053***
	(.026)	(.028)	(.018)	(.02)
Low Educ		.057***		.086***
		(.007)		(.004)
High Educ		.09***		.171***
		(.008)		(.005)
Age 2529		-.118***		-.072***
		(.012)		(.005)
Age 3034		-.25***		-.349***
		(.011)		(.005)
Age 3539		-.382***		-.587***
		(.012)		(.006)
Age 4044		-.462***		-.712***
		(.012)		(.006)
Age 4549		-.535***		-.758***
		(.013)		(.007)
Age 5054		-.588***		-.789***
		(.014)		(.007)
Age 5559		-.686***		-.87***
		(.016)		(.008)
Age 6065		-.801***		-.987***
		(.018)		(.009)
Log pseudolikelihood	-114808.89	-104535.62	-408140.34	-376864.89
Obs.	539555	504873	1790523	1755334

Note: The standard errors in parenthesis are robust to heteroscedasticity. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. Town district fixed effects, year fixed effects, and district-by-year-effects are included in all specifications.

Publication series published by the Institute for Labour Market Policy Evaluation (IFAU) – latest issues

Rapporter

2007:1 Lundin Daniela ”Subventionerade anställningar för unga – en uppföljning av allmänt anställningsstöd för 20–24-åringar”

2007:2 Lundin Daniela, Eva Mörk & Björn Öckert **”**Maxtaxan inom barnomsorgen – påverkar den hur mycket föräldrar arbetar?”

2007:3 Bergemann Annette & Gerard van den Berg ”Effekterna av aktiv arbetsmarknadspolitik för kvinnor i Europa – en översikt”

2007:4 Junestav Malin ”Socialförsäkringssystemet och arbetsmarknaden – politiska idéer, sociala normer och institutionell förändring – en historik”

2007:5 Andersson Christian ”Lärartäthet, lärarkvalitet och arbetsmarknaden för lärare”

2007:6 Larsson Laura & Caroline Runeson ”Effekten av sänkt sjukpenning för arbetslösa”

2007:7 Stenberg Anders ”Hur påverkar gymnasialt komvux löneinkomster och vidare studier?

2007:8 ”Forslund Anders & Kerstin Johansson ”Lediga jobb, arbetssökande och anställningar – den svenska matchningsfunktionen”

2007:9 Kennerberg Louise ”Hur förändras kvinnors och mäns arbetssituation när de får barn?”

2007:10 Nordin Martin ”Invandrares avkastning på utbildning i Sverige”

2007:11 Johansson Mats & Katarina Katz ”Underutnyttjad utbildning och lönegapet mellan kvinnor och män”

2007:12 Gartell Marie, Ann-Christin Jans & Helena Persson ”Utbildningens betydelse för flöden på arbetsmarknaden”

2007:13 Grönqvist Hans & Olof Åslund ”Familjestorlekens effekter på barns utbildning och arbetsliv”

2007:14 Lindqvist Linus ”Uppföljning av plusjobb”

2007:15 Sibbmark Kristina ”Avidentifierade jobbansökningar – erfarenheter från ett försök i Göteborgs stad”

2007:16 Hesselius Patrik & Malin Persson ”Incitamentseffekter och Försäkringskassans kostnader av kollektivavtalade sjukförsäkringar”

2007:17 Eriksson Stefan & Jonas Lagerström ”Diskriminering i anställningsprocessen: resultat från en Internetbaserad sökkanal”

2007:18 Eriksson Robert, Oskar Nordström Skans, Anna Sjögren & Olof Åslund ”Ungdomars och invandrades inträde på arbetsmarknaden 1985–2003”

2007:19 Agerström Jens, Rickard Carlsson & Dan-Olof Rooth ”Etnicitet och övervikt: implicita arbetsrelaterade fördomar i Sverige”

2007:20 Bennmarker Helge, Kenneth Carling & Anders Forslund ”Vem blir långtidsarbetslös?”

2007:21 Edmark Karin ”Strategisk konkurrens i kommuners utgiftsbeslut för barnomsorg, skola och äldreomsorg”

2007:22 Lindahl Erica ”Systematiska avvikelser mellan slutbetyg och provresultat – spelar elevens kön och etniska bakgrund roll?”

2007:23 Lindahl Erica ”Spelar lika kön och etnisk bakgrund på lärare och elever roll för provresultat och slutbetyg?”

2007:24 Fredriksson Peter & Björn Öckert ”Hur mycket påverkas studieresultat av resurser?

Working Papers

2007:1 de Luna Xavier & Per Johansson “Matching estimators for the effect of a treatment on survival times”

2007:2 Lundin Daniela, Eva Mörk & Björn Öckert “Do reduced child care prices make parents work more?”

2007:3 Bergemann Annette & Gerard van den Berg “Active labor market policy effects for women in Europe – a survey”

2007:4 Andersson Christian “Teacher density and student achievement in Swedish compulsory schools”

2007:5 Andersson Christian & Nina Waldenström “Teacher supply and the market for teachers”

2007:6 Andersson Christian & Nina Waldenström “Teacher certification and student achievement in Swedish compulsory schools”

2007:7 van den Berg Gerard, Maarten Lindeboom & Marta López ”Inequality in individual mortality and economic conditions earlier in life”

2007:8 Larsson Laura & Caroline Runeson “Moral hazard among the sick and unemployed: evidence from a Swedish social insurance reform”

2007:9 Stenberg Anders “Does adult education at upper secondary level influence annual wage earnings?”

2007:10 van den Berg Gerard “An economic analysis of exclusion restrictions for instrumental variable estimation”

2007:11 Forslund Anders & Kerstin Johansson "Random and stock-flow models of labour market matching – Swedish evidence"

2007:12 Nordin Martin "Immigrants' return to schooling in Sweden"

2007:13 Johansson Mats & Katarina Katz "Wage differences between women and men in Sweden – the impact of skill mismatch"

2007:14 Gartell Marie, Ann-Christin Jans & Helena Persson "The importance of education for the reallocation of labor: evidence from Swedish linked employer-employee data 1986–2002"

2007:15 Åslund Olof & Hans Grönqvist "Family size and child outcomes: Is there really no trade-off?"

2007:16 Hesselius Patrik & Malin Persson "Incentive and spill-over effects of supplementary sickness compensation"

2007:17 Engström Per & Patrik Hesselius "The information method – theory and application"

2007:18 Engström Per, Patrik Hesselius & Malin Persson "Excess use of Temporary Parental Benefit"

2007:19 Eriksson Stefan & Jonas Lagerström "Detecting discrimination in the hiring process: evidence from an Internet-based search channel"

2007:20 Agerström Jens, Rickard Carlsson & Dan-Olof Rooth "Ethnicity and obesity: evidence of implicit work performance stereotypes in Sweden"

2007:21 Uusitalo Roope & Jouko Verho "The effect of unemployment benefits on reemployment rates: evidence from the Finnish UI-benefit reform"

2007:22 Edmark Karin "Strategic competition in Swedish local spending on childcare, schooling and care for the elderly"

2007:23 Valentin Kvist Ann & Jan-Eric Gustafsson "The relation between fluid intelligence and the general factor as a function of cultural background: a test of Cattell's investment theory"

2007:24 Lindahl Erica "Comparing teachers' assessments and national test results – evidence from Sweden"

2007:25 Lindahl Erica "Gender and ethnic interactions among teachers and students – evidence from Sweden"

2007:26 Fredriksson Peter & Björn Öckert "Resources and student achievement – evidence from a Swedish policy reform"

2007:27 Adda Jérôme, Mónica Costa Dias, Costas Meghir & Barbara Sianesi "Labour market programmes and labour market outcomes: a study of the Swedish active labour market interventions"

Dissertation Series

2007:2 Edmark Karin “Interactions among Swedish local governments”

2007:1 Lundin Martin “The conditions for multi-level governance: implementation, politics, and cooperation in Swedish active labor market policy”

2006:1 Hägglund Pathric “Natural and classical experiments in Swedish labour market policy”

2006:2 Savvidou Eleni “Technology, human capital and labor demand”

2006:3 Söderström Martin “Evaluating institutional changes in education and wage policy”

2006:4 Lagerström Jonas “Discrimination, sickness absence, and labor market policy”

2006:5 Johansson Kerstin “ Empirical essays on labor-force participation, matching, and trade”