IFAU – INSTITUTE FOR
LABOUR MARKET POLICY
EVALUATION

Cluster sample inference using sensitivity analysis:
the case with few groups

Johan Vikström

WORKING PAPER 2009:15

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

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

ISSN 1651-1166

Cluster sample inference using sensitivity analysis: the case with few groups[a]

by

Johan Vikström[b]

11th June, 2009

Abstract
This paper re-examines inference for cluster samples. Sensitivity analysis is proposed as a new method to perform inference when the number of groups is small. Based on estimations using disaggregated data, the sensitivity of the standard errors with respect to the variance of the cluster effects can be examined in order to distinguish a causal effect from random shocks. The method even handles just-identified models. One important example of a just-identified model is the two groups and two time periods difference-in-differences setting. The method allows for different types of correlation over time and between groups in the cluster effects.

Keywords: Cluster-correlation; Difference-in-Differences; Sensitivity analysis.
JEL-codes: C12; C21; C23.

[a]I would like to thank Per Johansson, Michael Svarer, Nikolay Angelov, Xaiver de Luna, Gerard van den Berg, Bas van der Klaauw, Per Petterson-Lidbom, and seminar participants at VU-Amsterdam, IFAU-Uppsala, Örebro University, Stockholm University, and the RTN Microdata meeting in London for helpful comments. All remaining errors are mine. This paper was initiated when I visited University of Southern California supported by the Tom Hedelius Foundation and written in part when I was employed by VU Amsterdam. The financial support of the Swedish Council of Working Life and Social Research FAS (dnr 2004-2005) is acknowledged

[b]IFAU and Uppsala University, johan.vikstrom@ifau.uu.se

Table of contens

1 Introduction

In many studies the analysis sample consists of observations from a number of groups, for example families, regions, municipalities, or schools. These cluster samples impose inference problems, as the outcomes for the individuals within the groups usually cannot be assumed to be independent. Moulton (1990) shows that such intra-group correlation may severely bias the standard errors. This clustering problem occurs in many difference-in-differences (DID) settings, where one usually use variation between groups and over time to estimate the effect of a policy on outcomes at the individual level. As such the DID methodology is compelling, since it has the possibility of offering transparent evidence, which is also reflected in the exploding number of studies using the approach, for surveys see e.g. Meyer (1995) and Angrist & Krueger (2000). Many of these studies use data from only a small number of groups, such as data for men and women, a couple of states, or data from only a few schools or villages. For more examples see e.g. Ashenfelter & Card (1985), Meyer et al. (1995), Card & Krueger (1994), Gruber & Poterba (1994), Eissa & Liebman (1996), Imbens et al. (2001), Eberts et al. (2002), and Finkelstein (2002). The purpose of this paper is to provide a new method of performing inference when the number of groups is small, as is the case in these studies.

The importance of performing correct inference is also reflected in the growing number of studies addressing the inference problem.[1] One key insight from this literature is that the number of groups is important when deciding how to address the clustering problem. If the analysis sample consists of data from a larger number of groups, several solutions to the inference problem are available; the cluster formula developed by Liang & Zeger (1986), different bootstrap procedures (see e.g. Cameron et al. (2008)), or parametric methods (see e.g. Moulton (1990)). As expected however several Monte Carlo studies show that these methods perform rather poorly if the number of groups is small.[2]

[1] See e.g. Moulton 1986, 1990, Arrelano (1987), Bell & McCaffrey (2002), Wooldridge 2003, 2006 Bertrand et al. (2004), Kezdi (2004), Conley & Taber (2005), Donald & Lang (2007), Hansen2007*a*, 2007*b*, Ibragimov & Muller (2007), Abadie et al. (2007) and Cameron et al. (2008). Related studies are Abadie (2005) and Athey & Imbens (2006) which study semi-parametric and non-parametric DID estimation.

[2] See e.g. Bertrand et al. (2004), Donald & Lang (2007), Cameron et al. (2008), Ibragimov & Muller (2007), and Hansen (2007*a*).

To address this problem Donald & Lang (2007) introduce a between estimator based on data aggregated at group level.[3] They show that under certain assumptions, the aggregated error term is i.i.d normal and standard normal inference can be applied even if the sample consists of data from a small number of groups. Their method works as long as the number of groups is not too small. Since their method is based on aggregated data their inference will be conservative in the absence of within group correlation, or if the within group correlation is small. In the limit case when the model is just-identified, i.e. when the number of aggregated observations equals the number of variables varying at group level it not possible to perform Donald & Lang (2007).[4] An important example of a just-identified model is the two groups and two time periods DID setting. Another alternative is the two-stage minimum distance approach suggested by Wooldridge (2006). One important by-product of this approach is a simple test for the presence of within cluster correlation. However, as for the Donald & Lang (2007) approach the test does not work if the model is just-identified, as it is then based on a chi-square statistic with zero degrees of freedom. A final alternative is to use bias corrected standard errors as suggested by Bell & McCaffrey (2002). The method has two limitations; it does not work if the number of groups becomes too small or if the model includes a dummy variable taking the value one for exactly one cluster and zero otherwise.

As a response this paper proposes to use sensitivity analysis as a new method of performing inference when the number of groups is small. Design sensitivity analysis has traditionally been used to test whether an estimate is sensitive to different kinds of selectivity bias: see e.g. Cornfield et al. (1959) and Bross (1966), further see e.g. Rosenbaum & Rubin (1983), Lin et al. (1998), Copas & Eguchi (2001), Imbens (2003), Rosenbaum (2004) and de Luna & Lundin (n.d.). In these papers sensitivity analysis is performed with respect to the unconfoundedness assumption or with respect to the assumption of random missing data. If these assumptions hold, the usual estimators are unbiased and the sensitivity analysis amounts to assessing how far one can deviate from for example the unconfoundedness assumption before changing the estimate by some pre-specified

[3] Under certain assumptions the aggregation can be made on group-time level, instead of group-level.

[4] The inference is then based on a t-statistic with zero degrees of freedom.

amount.

My sensitivity analysis approach is similar, but nevertheless different in spirit. Under the assumption of no within group correlation standard normal i.i.d. inference based on disaggregated data is applicable. If this assumption is violated any standard errors based on the assumption of no within group correlation will be biased downwards. It is shown that under certain assumptions this bias can be expressed in terms of a few parameters, called sensitivity parameters. In the basic case the variance is expressed in terms of a single sensitivity parameter, defined as the ratio between the variance of the group common error term creating within cluster correlation, and the variance of the individual error term. The sensitivity analysis then amounts to assessing how much one can deviate from the assumption of no within group correlation before changing the standard error estimate by some pre-specified amount. That is to investigate how sensitive the standard errors are to within group correlation. The test can also be inverted in order to calculate a cut-off value, where higher values of the sensitivity parameter or simply larger variance of the group common shocks renders a certain estimate insignificant. If this cut-off value is unreasonably large one can be confident that the null hypothesis of no effect can be rejected. Optimally one could use information from other sources, for instance data from other countries, other time periods, or for another outcome, in order to assess the reasonable size of the sensitivity parameter. The approach proposed in this paper is therefore similar to standard sensitivity analysis, since it also assesses how much one can deviate from an important assumption, but it is also different in spirit since it is performed with respect to bias in the standard errors and not with respect to bias in the point estimate.

One key question is of course how to assess whether the sensitivity cut-off value is unreasonably large, that is how to assess the reasonable size of the within group correlation. I believe that this has to be done on a case by case basis. However, one advantage with the approach here is that the basic sensitivity parameter is defined as a ratio between two variances. It gives a sensitivity parameter with a clear economic interpretation, which of course is a basic condition for an informative sensitivity analysis. The next step is the discussion about a reasonable size of the sensitivity parameter. In order to shed more light on this issue two applications are provided. The sensitivity analysis method is applied to data analyzed in Meyer et al. (1995) on the effects of an increase in disability benefits on the duration of the period spent out of work and to Eissa & Liebman (1996) on the effects of an expansion in the earned income tax credit on labor supply. In both these studies key regressions are based on just-identified models. The sensitivity analyses indicate that the conclusion from the first study that the treatment effect is significant is not sensitive to departure from the independence (no-cluster) assumption, whereas the results of the second study are sensitive to the same departure and its conclusion cannot therefore be trusted. It demonstrates that the sensitivity analysis approach is indeed helpful for determining the validity of treatment effects.

By introducing sensitivity analysis in this way, this paper contributes in several ways. The method is applicable when the analysis sample consists of data from only a small number of groups. It even handles just-identified models. As no other method is applicable in the just-identified case it is the best application of the sensitivity analysis method. If the model is not just-identified but the number of groups is still small, the Monte Carlo study in this paper show that the sensitivity analysis method offers an attractive alternative compared to other commonly used methods. The method is also able to handle different types of correlation in the cluster effects, most importantly correlation within the group over time and multi-way clustering. This is done by introducing several sensitivity parameters.

The paper is structured as follows. Section 2 presents the basic model and analyzes the asymptotic bias (asymptotic in the number of disaggregated observations) of the OLS standard errors. Section 3 introduces the basic sensitivity analysis approach. Section 4 extends these basic results to more general settings. It is shown that different assumptions about the cluster effects lead to different types of sensitivity analyses. Section 5 presents Monte Carlo estimates on the performance of the sensitivity analysis method. The method is also compared to other commonly used methods of performing inference. Section 6 presents the two applications, and Section 7 concludes.

2 Basic model and bias in the regular OLS standard errors

Consider a standard time-series/cross section model. Take a linear model for the outcome y for individual i in time period t in group g as

$$y_{igt} = x'_{igt}\beta + e_{igt} \tag{1}$$

$$e_{igt} = c_{gt} + \varepsilon_{igt}$$

Here ε_{igt} is an individual time specific error, c_{gt} is a cluster effect which varies across groups and time, and x_{igt} the regressors. Of course individuals can represent any disaggregated unit. The regressors may or may not include fixed group effects and/or fixed time effects. This model covers a wide range of different models, including a "simple" cross-section, with data from for instance a couple of schools or villages. Another important example is the heavily used standard DID model. In a regression framework, a usual DID model is

$$y_{igt} = \alpha_g + \alpha_t + bD_{gt} + c_{gt} + \varepsilon_{igt}, \tag{2}$$

including fixed time, α_t, and fixed group effects, α_g, and where D_{gt} is an indicator function taking the value one if the intervention of interest is implemented in group g at time point t and zero otherwise. The treatment effect is hence identified through the variation between groups and over time. In this setting c_{gt} can be given a specific interpretation as any group-time specific shocks.[5]

Define $N = \sum^G \sum^T n_{gt}$, where G is the number of groups, T is the number of time periods, and n_{gt} is the number of individual observations for group g in time period t. If $\mathbb{E}[e_{igt}|x_{igt}] = 0$, the ordinary least square (OLS) estimate of β

$$\widehat{\beta} = (X'X)^{-1}X'Y \tag{3}$$

[5] c_{gt} also captures any differences in the group mean due to changes in the composition of the group over time. If n_{gt} is large this problem is mitigated.

is an unbiased estimate of β. Here Y is a N-vector collecting all y_{igt}, X is a $N \times K$ matrix containing the observations of the independent variables, and accordingly β a K-vector of the coefficients of interest.

Next consider inference. Assume that

$$\mathbb{E}(ee') = \sigma^2 C,$$

where e is a N-vector collecting all e_{igt}, and $\sigma^2 \equiv 1/N\, tr(ee')$, and C is a positive-definite matrix that captures the correlation in the error terms between the individuals. The true covariance matrix is then

$$V = \sigma^2 (X'X)^{-1} X'CX(X'X)^{-1}, \tag{4}$$

which can be compared with the regular OLS covariance matrix formula

$$\hat{V} = \hat{\sigma}^2 (X'X)^{-1}. \tag{5}$$

The asymptotic bias (asymptotic in the number of individuals (N)) of the regular standard errors has been analyzed extensively: see e.g. Greenwald (1983): other contributions are Campbell (1977), Kloek (1981) and Holt & Scott (1982). To be clear here we mean asymptotic in the number of individuals (N). Following equation (9)-(11) in Greenwald (1983) and some algebraic manipulations[6] gives the asymptotic bias in the estimated covariance matrix which can be expressed as

$$\mathbb{E}(\hat{V}) - V = \tag{6}$$

$$\sigma^2 \Big(\frac{tr[(X'X)^{-1}X'(I-C)X]}{N-K} (X'X)^{-1} + (X'X)^{-1}X'(I-C)X(X'X)^{-1} \Big).$$

Hence if $C = I$, that is the identity matrix, the estimated covariance matrix is an unbiased estimate of the true covariance matrix, and the estimated standard errors are unbiased. It

[6]Notice that V and n are defined in a different way here compared to Greenwald (1983). The expression follows from substituting equation (10) and (11) in Greenwald (1983) into equation (9) in Greenwald (1983), breaking out σ^2 and simplifying.

holds if $c_{gt} = 0$ for all g and all t, and if ε_{igt} is i.i.d. This general formula incorporates the two main reasons for bias in the standard errors into one expression. They are: (*i*) the cluster correlation problem caused by the presence of c_{gt}, highlighted by Moulton (1990), and (*ii*) the policy autocorrelation problem caused by correlation over time in c_{gt}, highlighted by Bertrand et al. (2004). The exact size of these problems depend on the case specific shape of C.

For the model in equation (1) the bias is negative, i.e. V is larger than $\mathbb{E}(\hat{V})$. It should also be noted that the bias consist of two distinct parts. Thus in the case of cluster effects in form of within group correlation the OLS standard errors underestimate the true standard errors. First, the OLS estimator of the error variance $\hat{\sigma}^2$, is neither an unbiased nor a consistent estimator of the true error variance σ^2, if the error covariance matrix does not satisfy the OLS assumptions. Second, and more obvious, even if the error variance is known, the standard errors are biased since the coefficient covariance matrix is misspecified.

3 Sensitivity analysis for cluster samples

The discussion in the previous section reveals that whether or not $c_{gt} = 0$ is crucial for how to perform inference. If $c_{gt} = 0$ regular OLS inference can be performed, possibly with control for heteroscedasticity. If $c_{gt} \neq 0$ on the other hand the regular OLS standard errors will be severely biased. As shown by Donald & Lang (2007) this has very important implications when the number of groups is small. They introduce a between estimator based on data aggregated at group level. It creates an all or nothing situation; under the assumption of $c_{gt} = 0$ there are apparently narrow confidence intervals based on individual data, and under the assumption of $c_{gt} \neq 0$ there are apparently very wide confidence intervals based on aggregated data. Needless to say arguing that $c_{gt} = 0$ will almost always be very difficult, whereas arguing that the variance of c_{gt} is small is reasonable in many applications. In the end it is the size variance of the within group correlation that matters. This is the key idea behind the new method proposed in this paper.

Formally, the starting point for the sensitivity analysis method is the general formula for the bias in the regular OLS standard errors presented in equation (6). This expression is based on derivations in Greenwald (1983), which among other expressions uses

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2(N - tr[(X'X)^{-1}X'CX])/(N-K).^7 \tag{7}$$

Combining this expression with the definition of V in equation (4) and the definition of $\hat{V}$ in equation (5), and noting that $\mathbb{E}(\hat{V}) = \mathbb{E}(\hat{\sigma}^2)(X'X)$ gives

$$V = \frac{N-K}{N - tr[(X'X)^{-1}X'CX]}(X'X)^{-1}X'CX\mathbb{E}(\hat{V}). \tag{8}$$

Further plim$(\hat{V}) = \mathbb{E}(\hat{V})$ and thus $\mathbb{E}(\hat{V})$ can be consistently (in terms of number of individuals) estimated by $\hat{V}$.

Starting with this equation the idea behind the sensitivity analysis is straightforward. Faced with a cluster sample with data from only a small number of groups, we can use disaggregated data and estimate β using OLS. Then estimate $\hat{V}$ in equation (4) as if there were no cluster effects. Then notice that $\hat{V}$ only gives correct standard errors if $c_{gt} = 0$ for all g and t. However, as a sensitivity analysis we can use the expression above and express the bias in the covariance matrix in terms of different so called sensitivity parameters, and assess how large they have to be in order to change the variance of a parameter estimate by a certain amount: that is, if the results are insensitive to departures from the assumption of no within group correlation, it indicates that the results can be trusted. As shown below the exact specification of the sensitivity parameters will depend on the assumptions which can be imposed on C.

Let us start with the simplest case. If ε is homoscedastic and if $\mathbb{E}(c_{gt}c_{g't}) = 0$ for all t and all $g \neq g'$, and $\mathbb{E}(c_{gt}c_{gt'}) = 0$ for all g and all $t \neq t'$, then the full error term, $e_{igt} = c_{gt} + \varepsilon_{igt}$, is homoscedastic[8], equi-correlated within the group-time cell and uncorrelated between the group-time cells. Further assume $n_{gt} = n$ and $x_{igt} = x_{gt}$, that is, the

[7] See derivations of equation (A.3) in Greenwald (1983).

[8] The sensitivity analysis throughout this paper is made under the homoscedasticity assumption. The assumption makes it possible to write the bias in terms of single parameters. If one suspect heteroscedasticity, one approach is to use standard errors robust to heteroscedasticity in the spirit of White (1980), and use this covariance matrix instead of $\hat{V}$. The sensitivity analysis based on this specification will then be conservative.

regressors are constant within each group, and constant group size. This special case has been analyzed by Kloek (1981).[9] He shows that under these assumptions equation (8) reduces to

$$V = \mathbb{E}(\hat{V})\tau\frac{nGT-K}{nGT-K\tau} \tag{9}$$

with

$$\tau = 1+(n-1)\frac{\sigma_c^2}{\sigma_c^2+\sigma_\varepsilon^2}. \tag{10}$$

Here σ_c^2 is the variance of c, and σ_ε^2 the variance of ε. Expressing the ratio between these two variances as $\sigma_c^2 = \gamma\sigma_\varepsilon^2$ gives

$$V = \mathbb{E}(\hat{V})\left(1+(n-1)\frac{\gamma}{1+\gamma}\right)\frac{nGT-K}{nGT-K(1+(n-1)\frac{\gamma}{1+\gamma})} \tag{11}$$

In other words the bias in the covariance matrix is expressed in terms of observables and a single unknown parameter γ, which is interpreted as the relation between the variance of the group-time error term and the variance of the individual error term.[10]

Using standard textbook results; if $\gamma = 0$, that is if there is no within group correlation, and $\sum n_{jt}$ is large

$$t = \frac{\hat{\beta}_a}{\sqrt{\mathbb{E}(\hat{V}_{aa})}} \overset{a}{\sim} N(0,1), \tag{12}$$

where $\hat{\beta}_a$ is the ath element of $\hat{\beta}$, and $\hat{V}_{aa}$ the element in the ath column and ath row of $\hat{V}$. Furthermore if $\gamma \neq 0$ and known, $c_{jt} \sim N(0,\sigma_c^2)$[11], and $\sum n_{jt}$ is large

$$t = \frac{\hat{\beta}_a}{\sqrt{V_{aa}}} = \frac{\hat{\beta}_a}{\sqrt{\mathbb{E}(\hat{V}_{aa})(1+(n-1)\frac{\gamma}{1+\gamma})\frac{nGT-K}{nGT-K(1+(n-1)\frac{\gamma}{1+\gamma})}}} \overset{a}{\sim} N(0,1). \tag{13}$$

[9] Kloek (1981) analyzes the one dimensional case with only a group dimension and no time dimension. A group-time version of his proof is presented in Appendix.

[10] Actually γ is only potentially unknown. If the number of groups is larger σ_c^2 can be consistently estimated using the between group variation, and σ_ε^2 can be consistently estimated using the within group variation, and this gives p.

[11] The normality assumption can be replaced by any other distributional assumption, for instance a uniform distribution. However this will complicate the sensitivity analysis, since the combined error term will have a mixed distribution.

It is then possible to use γ as a sensitivity parameter. After estimating $\hat{\beta}_a$ and consistently estimating $\mathbb{E}(\hat{V}_{aa})$ by $\hat{V}_{aa}$ using the disaggregated data, the sensitivity analysis then amounts to assessing how much γ has to deviate from zero in order to change the standard errors by a pre-specified amount. The sensitivity analysis method is applicable as long as the model is identified. In the present case with variables constant within each group-time cell, this holds if $GT \geq K$, i.e. if the number of group-time cells is larger than or equal to the number of explanatory variables. In other words our sensitivity analysis method even handles just-identified models, for instance the two groups and two time periods DID setting. As no other method is applicable in the just-identified case it is the best application of the sensitivity analysis method. If the model is not just-identified but the number of groups is still small the sensitivity analysis method offers an alternative to other commonly used methods such as the Donald & Lang (2007) approach.

The test can also be inverted in order to calculate the γ value which corresponds to a specific p-value. One could for example be interested in the γ cut-off value which renders the estimated treatment effect statistically insignificant at $\alpha\%$ level. This follows from setting $t = Z_{1-\alpha/2}$ and solve for γ in the equation (13) above

$$\gamma_{c,a} = \frac{(\hat{\beta}_a^2 - Z_{1-\alpha/2}^2 \hat{V}_{aa})(nGT - K)}{(nZ_{1-\alpha/2}^2 \hat{V}_{aa})(nGT - K) - \hat{\beta}_a^2 (nGT - nK)}. \tag{14}$$

Here Z_υ is the υ quantile of the standard normal distribution. Note that we have replaced $\mathbb{E}(\hat{V}_{aa})$ with $\hat{V}_{aa}$ as it is consistently estimated by $\hat{V}_{aa}$. Furthermore, note that $\gamma_{c,a}$ depends on n, the number of observations for each group. This dependence comes both from $\hat{V}$ which decreases as n increases and also directly as n enters the expression for $\gamma_{c,a}$. Taken together these two effects means that $\gamma_{c,a}$ increases as n goes from being rather small to moderately large: however as n becomes large this effect flattens out, and $\gamma_{c,a}$ is basically constant for large n.

If $\gamma_{c,a}$ is unreasonably large, one could be confident that the null-hypothesis about zero effect could be rejected. The key question then becomes: what is unreasonably large? At the end of the day, as with all sensitivity analyses, some judgment has to be made. Since the true γ may vary a lot between different applications, we believe that the assessment has to be done on a case by case basis. However, the sensitivity analysis presented here avoids the common sensitivity analysis pitfall. That is, that one is left with a sensitivity parameter which is hard to interpret and thus hard to relate to economic conditions. Here the basic sensitivity parameter, γ, is defined as the ratio between two variances, which makes it both easier to interpret and easier to discuss. Optimally one could also use information from other sources to make the discussion more informative for instance data from another country, other time periods, or for another outcome. In some cases it may also be beneficial to re-scale γ. The two applications presented in Section 6 using data from Meyer et al. (1995) and Eissa & Liebman (1996) further exemplify how γ can be interpreted.

If either the assumption of either $n_{gt} = n$ or $x_{igt} = x_{gt}$ is relaxed the sensitivity analysis is still straightforward. Note that the general formula for the bias presented in equation (8) nevertheless holds. In the basic case with $n_{gt} = n$ or $x_{igt} = x_{gt}$ this expression could be simplified considerably. In general under assumption $\mathbb{E}(c_{gt}c_{g't}) = 0$, assumption $\mathbb{E}(c_{gt}c_{gt'}) = 0$, and with the model specified as in equation (1), C has the familiar block-diagonal structure

$$C = \begin{bmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{GT} \end{bmatrix} \tag{15}$$

with $C_{GT} = [(1 - \frac{\gamma}{1+\gamma})I_{gt} + \frac{\gamma}{1+\gamma}J_{gt}]$. Here I_{GT} is an n_{gt} times n_{gt} identity matrix, and J_{gt} is an n_{gt} times n_{gt} matrix of ones: $\gamma_{c,a}$ is then found by numerically solving for γ in

$$Z_{1-\alpha/2} = \frac{\hat{\beta}_a}{\sqrt{V_{aa}}}, \tag{16}$$

with V defined as in equation (8) and C defined as in equation (15) above. From calculations I note that in general $\gamma_{c,a}$ is quite insensitive to violations of $n_{gt} = n$, except when some groups are very large and others are very small.

4 Extended sensitivity analysis

4.1 Correlation over time in the cluster effects

The sensitivity analysis presented in the previous section is applicable under a number of assumptions about c_{gt}. Most notably $\mathbb{E}(c_{gt}c_{g't}) = 0$ for all t and all $g \neq g'$, and $\mathbb{E}(c_{gt}c_{gt'}) = 0$ for all g and all $t \neq t'$. In many studies $\mathbb{E}(c_{gt}c_{gt'}) = 0$ for all g is a restrictive assumption. In a model with fixed group and fixed time effects, c_{gt} captures any group-time shocks. Consider a study on the effects of minimum wages on employment using variation across regions and over time. The group-time shocks then capture all regional specific shocks in employment. If present they are most likely correlated over time. This problem, often refereed to as the policy autocorrelation problem, was highlighted by Bertrand et al. (2004).

This subsection therefore relax the assumption that $\mathbb{E}(c_{gt}c_{gt'}) = 0$: instead we assume an AR(1) structure for c_{gt}

$$c_{gt} = \kappa c_{gt-1} + d_{gt}, \tag{17}$$

where d_{gt} is assumed to be a white noise series with mean zero and variance σ_d^2. Further, assume that $|\kappa| < 1$. I make the natural extension of the basic sensitivity analysis and define $\sigma_d^2 = \gamma\sigma_\varepsilon^2$. It gives two sensitivity parameters, γ and κ, instead of the single sensitivity parameter γ. Then if $\kappa = 0$ the basic sensitivity analysis is applicable. To be clear, κ is interpreted as the first-order autocorrelation coefficient for c_{gt}, and γ as the relation between the variance of the group-time specific shock and the variance of the unobserved heterogeneity.

Consider the case with repeated cross-section data. Assume that data on n_{gt} individuals from group g in time period t are available. The general formula presented in equation (8) for the covariance matrix still hold. However, since c_{gt} is allowed to follow an arbitrary AR(1) process, C will obviously differ from the basic sensitivity analysis. In order to express C in terms of κ and γ we use the well know properties of an AR(1) process. It turns that out if $n_{gt} = n$ and $x_{igt} = x_{gt}$ holds, there is a simple expression for the relation between V and $\hat{V}$

$$V_{aa} \approx \mathbb{E}(\hat{V}_{aa})(1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2}+n\frac{\gamma}{1+\gamma-\kappa^2}H_{aa}) \tag{18}$$

where H_{aa} is the element in the ath column and ath row of H given by

$$H = (\sum_g \sum_t x_{gt}x'_{gt})^{-1} \sum_g \sum_t \sum_{t' \neq t} (\kappa^{|t-t'|} x_{gt}x'_{gt'})$$

The proof can be found in Appendix.

Based on this simple expression for the bias in the regular OLS standard errors, one can assess the sensitivity of the standard errors with respect to both the autocorrelation and the variance of the group-time specific shocks. As for the basic sensitivity analysis one may be interested in the cut-off value which renders an interesting estimate insignificant. In this case with two sensitivity parameters a natural way to proceed is to solve for γ for a range of values of κ. Let us that interest lies in the effect of variable a, then the cut-off value for γ is

$$\gamma_{c,a} = \frac{(\hat{\beta}_a^2 - Z_{\alpha/2}^2 \hat{V}_{aa})(1-\kappa^2)}{(nZ_{\alpha/2}^2 \hat{V}_{aa})(1+H_{aa}) - \hat{\beta}_a^2}. \tag{19}$$

Again note that $\mathbb{E}(\hat{V}_{aa})$ is replaced with $\hat{V}_{aa}$ as it is consistently estimated by $\hat{V}_{aa}$. If the combinations of $\gamma_{c,a}$ and κ values are unreasonable large, one could be confident in that the null hypothesis about zero effect should be rejected. Also note that $\gamma_{c,a}$ can either increase or decrease with κ, as H_{aa} can either increase or decrease with κ.

If either $n_{gt} = n$ or $x_{igt} = x_{gt}$ do not hold it is not possible to obtain a closed from solution for $\gamma_{c,a}$. But using numerical methods, it is possible to solve for γ in

$$Z_{1-\alpha/2} = \frac{\hat{\beta}_a}{\sqrt{V_{aa}}}, \tag{20}$$

for a range of values of κ and the desired significance level. Here V is defined in equation (8), and C is defined in equation $(A.13)$ presented in appendix.

4.2 Multi-way clustering

Consider an application where we have data from a number of regions and where the region is defined as the group. In the sensitivity analysis presented so far, the assumption of $\mathbb{E}(c_{gt}c_{g't}) = 0$ is crucial. In other words it is assumed that the outcomes for individuals within a region are correlated and that there is no correlation between individuals on different sides of the border between two different regions. Most likely this will be violated in many applications. Here this assumption is relaxed in the situation with cross-section data. Assume that the groups can be divided into group clusters containing one or more groups. Dropping the time dimension, the outcome y for individual i in group g in group-cluster s is

$$y_{igs} = x_{igs}\beta + c_{gs} + \varepsilon_{igs}. \tag{21}$$

Retain the definition of γ from the basic sensitivity analysis as $\sigma_c^2 = \gamma\sigma_\varepsilon^2$. γ is then again interpreted as the relation between the variance of the group-time shocks and the variance of the individual unobserved heterogeneity. Further assume that if $s \neq s'$ then $E(c_{gs}c_{g's'}) = 0$ and if $s = s'$ then $E(c_{gs}c_{g's'}) = \xi\sigma_c^2$.[12] ξ should be interpreted as the relation between the inter-group correlation and the intra-group correlation for groups in the same cluster of groups. This means that it will be far below one in many applications.

Note that the general expression for the covariance matrix presented in equation (8) holds. If the above assumptions hold, and if $n_g = n$ and $x_{igt} = x_{gt}$ hold, the derivations in the appendix show that there is a simple relation between V_{aa} and $\hat{V}_{aa}$

$$V_{aa} \approx \hat{V}_{aa}\left(1 + (n-1)\frac{\gamma}{1+\gamma} + n\frac{\gamma}{1+\gamma}\xi M_{aa}\right) \tag{22}$$

where M_{aa} is the element in the ath column and ath row of M given by

$$M = \left(\sum_s\sum_g x_{gs}x'_{gs}\right)^{-1}\sum_s\sum_g\sum_{g'\neq g}(x_{gs}x'_{g's}).$$

[12] It is obviously possible to also allow for an time-dimension, which generally gives sensitivity analysis in three parameters, which would measure the variance, the autocorrelation respectively the between group correlation in the cluster effects.

Again there are two sensitivity parameters, γ and ξ. As in the previous case one can proceed to solve for $\gamma_{c,a}$ for a range of values of ξ. Let us that the interest lies in the effect of variable a: then

$$\gamma_{c,a} = \frac{\hat{\beta}_a^2 - Z_{\alpha/2}^2 \hat{V}_{aa}}{(nZ_{\alpha/2}^2 \hat{V}_{aa})(1+\xi M_{aa}) - \hat{\beta}_a^2}. \tag{23}$$

If these combinations of $\gamma_{c,a}$ and ξ values are unreasonable large, one could be confident that the null hypothesis about zero effect should be rejected. One could also interpret the division of the groups into group clusters as a sensitivity analysis. The standard errors may be sensitive to some divisions but not to others. Note that introducing multi-way clustering in the way done here increases the standard errors, and thus $\gamma_{c,a}$ decreases with ξ.

If either $n_{gt} = n$ or $x_{igt} = x_{gt}$ do not hold it is not possible to obtain a closed from solution for $\gamma_{c,a}$. But it is possible to solve for γ in

$$Z_{1-\alpha/2} = \frac{\hat{\beta}_a}{\sqrt{V_{aa}}}, \tag{24}$$

for a range of values of ξ and the desired significance level. Here V is defined in equation (8), and C is defined in equation (A.25) presented in the appendix.

5 Monte Carlo evidence

This section provides Monte Carlo estimates of the performance of the proposed sensitivity analysis method. The small sample properties of the method and the sensitive of method is to the choice of reasonable γ are investigated. The sensitivity analysis method is also compared to other commonly used inference methods. I consider a DID set up. The treatment is assumed to vary at group-time level, and the interest lies in estimating the effect of this treatment on individual outcomes.

Assume that the underlying model is

$$y_{igt} = c_{gt} + \varepsilon_{igt}.$$

The group error term, c_{gt}, and the individual error term, ε_{igt}, are both independent normals with variance σ_c^2 and σ_ε^2. Take $\sigma_c^2 = 0.1$ and $\sigma_\varepsilon^2 = 1$. Note that the result is not sensitive to this choice. I experiment with different numbers of of groups (G) and different number of time periods (T). Data are generated with a constant group-time cell size, $n_{gt} = n$. In all experiments 50,000 simulations are performed.

I estimate models of the form

$$y_{igt} = \alpha_g + \alpha_t + bD_{gt} + c_{gt} + \varepsilon_{igt}.$$

This represents a general DID setting, with fixed group effects, α_g, fixed time effect, α_t, and a treatment indicator variable, D_{gt}, taking the value one if the treatment is imposed in group g at time point t. b is then the treatment effect. The treatment status is randomly assigned. In the basic case we take two time periods ($T = 2$) and two groups ($G = 2$). The treatment status is then assigned to one of the groups ($G_1 = 1$), and they experience the treatment in the second period. Besides the basic case, other combinations of T,G and D are considered [13]. To be precise, the basic model with $T = 2$, $G = 2$ and $G_1 = 1$ includes two group dummies, one time dummy for the second period, and one treatment dummy taking the value one in the second period for group two. The models for other combinations of T,G and D follow in the same way.

5.1 Small sample properties

As shown in Section 3, the sensitivity analysis method can be used to derive a cut-off value, γ_c. This value can be seen as a test-statistic. If one is confident that this value is unreasonably large one should reject the null-hypothesis of zero effect. In other words the critical value is decided by the researchers knowledge about reasonable values of γ.

If the researcher knows the true relation between σ_c^2 and σ_ε^2, referred to as $\gamma_t = \sigma_c^2/\sigma_\varepsilon^2$, then theoretically if N is large a test for $b = 0$ using γ_c as a test-statistic and using γ_t as the critical value should have the correct size. This should hold for any combination of $T \geq 2, G \geq 2$ and $G > G_1$. This subsection confirms this property. I also examine the small sample properties of this approach. To this end the approach is somewhat mod-

[13] If $T > 2$ the treatment occurs after $T/2 - 0.5$ if T is a odd number and after $T/2$ if T is an even number.

Table 1: Monte Carlo results for the small sample properties of the sensitivity analysis method.

Group Size (n)	$G=2,\ T=2$ $G_1=1$	$G=3,\ T=2$ $G_1=1$	$G=3,\ T=3$ $G_1=1$	$G=5,\ T=5$ $G_1=2$
10	0.0503	0.0492	0.0495	0.0504
20	0.0496	0.0489	0.0512	0.0500
50	0.0505	0.0516	0.0494	0.0500
100	0.0484	0.0519	0.0501	0.0505
1000	0.0505	0.0504	0.0495	0.0494

Notes: Monte Carlo results for the treatment parameter which enters the model with a true coefficient of $b=0$. The model and the data generating process is described in detail in the text. Each cell in the table reports the rejection rate for 5% level tests using the sensitivity analysis γ_c as test-statistic, and γ_t as critical value. Test based on a $t_{nGT-G-T}$. T is the number of time periods, G the number of groups, and G_1 the number of groups who receives the treatment. The number of simulations is 50,000.

ified. Asymptotically (in N) the sensitivity analysis method can be based on a normal distribution, regardless of the distribution of the individual error, ε. If N is small but ε is normally distributed the analysis should be based on a t-distribution with $nGT-G-T$ degrees of freedom. This follows since the t-statistic reported in equation (12) has an exact t-distribution instead of a normal distribution.

Table 1 present the results from this exercise. Each cell of *Table 1* represents the rejection rate under the specific combination of n,T,G,D, and γ_t. As apparent from the table, the sensitivity analysis method works as intended for all sample sizes. It confirms that the derived properties of the sensitivity analysis method are correct. This is not surprising since the sensitivity analysis is based on OLS estimates with well established properties. It does not however give evidence for an inferential method in a strict statistical sense as the exact value of the used critical value γ_t is not known in practice. In practice reasonable values of γ_t have to be assessed, for instance using other data sources.

5.2 Robustness and comparison with other inference methods

The researcher may have information through other data sources, or for other outcomes, which enables a closer prediction of γ_t. However information that enables an exact estimate of γ_t is not likely to be available. The second experiment therefore test the robustness of the results with respect to assessing an incorrect γ_t. Distinguish between the true ratio between the two error variances, γ_t and the ratio that the researcher thinks is the correct one, γ_r. If $\gamma_c > \gamma_r$ the sensitivity analysis suggests rejecting the null-hypothesis of zero effect. If $\gamma_t > \gamma_r$ this leads to over-rejection of the null-hypothesis. Here the severity of this problem is tested.

As a comparison the sensitivity analysis method is contrasted with other methods commonly used to perform inference. The other methods include OLS estimates without any adjustment of the standard errors, labeled OLS regular. Furthermore, OLS estimates with the commonly used Eicker-White heteroscedasticiy robust standard errors for grouped data. I either ”cluster” at group level or ”cluster” at the group-time level, i.e. variance matrices which is robust to within group correlation, and robust to within group-time cell correlation, respectively. These inference methods are labeled cluster group and cluster group-time. They are by far the most common ways of correcting for the use of individual data and outcomes that vary only on at group level. The general cluster formula is

$$\hat{V}_{cluster} = \frac{N-1}{N-K}\frac{C}{C-1}\Big(\sum_{c=1}^{C} X_c'X_c\Big)\Big(\sum_{c=1}^{C} X_c'\hat{u}_c\hat{u}_c'X_c\Big)\Big(\sum_{c=1}^{C} X_c'X_c\Big)$$

where c indicates the cluster and C the number of clusters. Further, $\hat{u}_c$ is a vector containing the OLS residuals, and X_c is a matrix containing the observations of the independent variables for the individuals in cluster c. Also note that a degrees of freedom correction is used. The tests are based on a t_{C-1}, i.e. t_{G-1} for clustering at the group level, and t_{GT-1} for clustering at the group-time level.

The two-step estimator suggested by Donald & Lang (2007) is also considered. In the present case with explanatory variables which vary only at group-level, and in the absence of correlation over time in c_{gt}, the first step is aggregation at the group-time level. This gives

$$\bar{y}_{gt} = \alpha_g + \alpha_t + \beta X_{gt} + c_{gt} + \bar{\varepsilon}_{gt},$$

where $\bar{y}_{gt}$ and $\bar{\varepsilon}_{gt}$ are the group-time averages of y_{igt} and e_{igt}. The second step amounts to estimating this model using OLS. In the present case when both the error terms are independent normals the resulting t-statistic for the hypothesis test of $b = 0$ has a t-distribution with $GT - K$ degrees of freedom. The number of variables, K, is here $G + T$.

The upper panel of *Table 2* presents the results for the sensitivity analysis method, and the lower panel presents the results for the other four methods. Size is for 5% level tests for the treatment parameter which enters the model with a true coefficient of $b = 0$. Power is 5% level test versus the alternative that $b = 0.1$. In this analysis we take $n = 200$. Before interpreting these results note that the power should be compared for tests with the same nominal size. Furthermore, the terminology size and power for the sensitivity analysis method is not entirely correct from a statistical point of view. The sensitivity analysis gives a ”test-statistic” as a cut-off value, γ_c, but the cut-off value is decided by the researchers assessment of a reasonable size off γ. In that sense it is a test, which makes it reasonable to report the size and power. Also note that the main point of this section is to explore how sensitive the results are to the assessment of a reasonable size off γ.

First, consider the performance of the other methods commonly used to perform inference. The results for the OLS estimates using regular standard errors and the two cluster formulas confirm what has been found in earlier studies, see e.g. Bertrand et al. (2004), Donald & Lang (2007), Cameron et al. (2007), and Hansen (2007*a*). The regular uncorrected OLS estimates have large size distortions. The rejection rate for 5% level tests is 0.256 with $G = 2$, $T = 2$, $G = 1$. The two OLS cluster estimators also suffer from large size distortions. As expected these methods behave poorly if the number of groups is small: after all they were designed for the case with a large number of groups. If the number of groups is only moderately small, say $G = 5$ and $T = 5$, these tests perform somewhat better.[14]

Next, consider the performance of the Donald & Lang (2007) two step estimator. If the model is just-identified as in the case with $G = 2$ and $T = 2$ the test of the null hypothesis should be done using a t-statistic with zero degrees of freedom. In other words it is not possible to use this test for just-identified models. Next consider how the two step estimator performs if the groups become somewhat larger, but are still very small. The results in Column 2, 3 and 4 show that the DL estimator has correct size if the model is not just-identified. This confirm the results in Donald & Lang (2007). However, if the number of groups is very small (Column 2 and 3) the power of the DL estimator is low.

Let us compare these results with the results for the sensitivity analysis method, which

[14]Notice that this experiment is set up with no correlation between the groups or over time in c_{gt}. If that were the case we could expect these cluster estimators to perform even worse. The size distortions for $G = 5$ and $T = 5$ would then be likely to also be very large.

Table 2: Monte Carlo results for the sensitivity analysis method when the true relation between the variance of the group-time error and the individual error is unknown.

	$G=2$, $T=2$ $G_1=1$		$G=3$, $T=2$ $G_1=1$		$G=3$, $T=3$ $G_1=1$		$G=5$, $T=5$ $G_1=2$	
	Size	Power	Size	Power	Size	Power	Size	Power
Sensitivity analysis								
$\gamma_t=0.010$								
$\gamma_r=0.005$	0.111	0.671	0.112	0.778	0.108	0.870	0.112	1.000
$\gamma_r=0.008$	0.069	0.587	0.068	0.701	0.066	0.816	0.069	0.999
$\gamma_r=0.009$	0.057	0.562	0.058	0.677	0.058	0.794	0.059	0.999
$\gamma_r=0.010$	0.050	0.531	0.048	0.657	0.050	0.779	0.050	0.998
$\gamma_r=0.011$	0.044	0.504	0.041	0.627	0.044	0.758	0.044	0.998
$\gamma_r=0.012$	0.038	0.481	0.036	0.609	0.038	0.739	0.037	0.998
$\gamma_r=0.015$	0.024	0.411	0.024	0.539	0.024	0.675	0.024	0.996
$\gamma_r=0.020$	0.011	0.314	0.012	0.431	0.011	0.575	0.012	0.992
OLS regular	0.256	0.817	0.257	0.889	0.260	0.945	0.257	1.000
Cluster group-time	1.000	0.978	0.562	0.934	0.370	0.944	0.133	0.973
Cluster group	1.000	0.978	0.276	0.684	0.280	0.750	0.107	0.996
DL two step	n.a.	n.a.	0.051	0.148	0.051	0.462	0.050	0.998

Notes: Monte Carlo results for simulated data. The model and the data generating process is described in detail in the text. γ_t is the true relation between the variance of the group-time error and the individual error, and γ_r the assessed relation between these two variance. Further T is the number of time periods, G the number of groups, G_1 the number of groups who receives the treatment, and n_{gt} the sample size for each group-time cell. Size is for 5% level tests for the treatment parameter which enters the model with a true coefficient of $b=0$. Power is 5% level test versus the alternative that $b=0.1$. The number of simulations is 50,000.

uses γ_c as the test-statistic and γ_r as the critical value.[15] First, consider the results when $\gamma_t=\gamma_r$, i.e. the researcher is able to correctly assess the size of within group correlation. As before the test has the correct size. Since the size of the test for the sensitivity analysis method and the DL method are the same for the results in Column 2-4, the power estimates are comparable. The results show that the power is higher in the sensitivity analysis method. If the number of groups is very small, as in Column 2 and 3, the difference is large. For example if $G=3$, $T=2$, and $G=1$ the power is 0.657 for the sensitivity analysis compared to 0.148 for the DL two step estimator. If the number of groups becomes somewhat larger as in Column 4 the difference is smaller. In this case the Donald & Lang (2007) two step estimator is likely to be preferable to sensitivity analysis. Also note that even if $G=2$, $T=2$, $G_1=1$ the power of the sensitivity analysis test is high.

The previous comparison was based on the assumption that the researcher is able to

[15] I will use the terminology ”size” and ”test” here even though the sensitivity analysis method is not a statistical test, as the critical value is assessed and not calculated.

assess the correct value of γ. In practice this is unreasonable. It is therefore also interesting to see what happens if γ_t not equal to γ_r, i.e. when the researcher is unable to exactly infer γ. These results are also presented in *Table 2*. These results show that the sensitivity analysis method performs well if the difference between γ_r and γ_t is rather small. For example, the rejection rate for 5% level tests is 0.069 if $\gamma_r = 0.008$ and $\gamma_t = 0.010$. This is only a small over-rejection of the null-hypothesis. However if the difference between γ_r and γ_t becomes large, there are as expected substantial size distortions.

To summarize, the Monte Carlo simulations have confirmed that the derived properties of the sensitivity analysis method are correct for both large and small sample sizes. They further show that existent inference methods run into problem when the number of groups is very small. Finally, the results show that the sensitivity analysis method is applicable, even if the number of groups is very small, as long as the size of the within group correlation can be reasonably assessed. The two applications provided in the next section show that this can often can be done.

6 Applications[16]

6.1 Application 1: Disability benefits

Meyer et al. (1995)[17] (MVD) study the effects of an increase in disability benefits (workers compensation) in the state of Kentucky. Workers compensation programs in the USA are run by the individual states. Here we describe some of the main features of the system in Kentucky. A detailed description is found in MVD. The key components are payments for medical care and cash benefits for work related injuries. MVD focus on temporary benefits, the most common cash benefit. Workers are covered as soon as they start a job. The insurance is provided by private insurers and self-insurers. The insurance fees that

[16] This section presents two different applications. In order to focus on the application of the sensitivity analysis approach we re-examine some basic results from the two studies. I should however point out that a more elaborated analysis is performed in both studies is. It includes estimating for different sample, different outcomes and including additional control variables. However the basic regressions re-examined here constitute an important part of both studies.

[17] This data has also been reanalyzed by Athey & Imbens (2006). They consider non-parametric estimation, and inference under the assumption of no cluster effects. Meyer et al. (1995) also consider a similar reform in Michigan.

employers pay are experience rated. If eligible the workers can collect benefits after a seven day waiting period, but benefits for these days can be collected retroactively if the duration of the claim exceeds two weeks. The claim duration is decided mainly by the employee and his or her doctor, and there is no maximum claim duration.

The replacement rate in Kentucky before 1980 was $66\frac{2}{3}\%$ and the benefits could be collected up to a maximum of $131 per week. The reform as of July 15, 1980, analyzed by MVD increased the maximum level to $217 per week: a 66% increase or 52% over one year in real terms.[18] The replacement rate was left unchanged. Thus workers with previous high earnings (over the new maximum level) experience a 66% increase in their benefits, while the benefits for workers with previous low earnings (below the old ceiling) were unchanged. This creates a natural treatment group (high earners) and a natural control group (low earners). MVD analyze the effect of the increase using a DID estimator, which contrasts the difference in injury duration between before and after the reform for the treatment group and the control group.

The upper panel of *Table 3* restates MVD's results for the outcome mean log injury duration, taken from their Table 4.[19] Column 1-4 present the pre-period and post-period averages for the treatment and control group, Column 5 and 6 the difference between the pre- and post-period for the two groups, and Column 7 present the DID estimate. The DID estimate of the treatment effect is statistically significant and suggests that the increased benefits increased the injury duration by about 19%. MVD ignores the cluster-sample issue and use regular OLS standard errors. Thus their standard errors are biased downwards if there are any cluster effects. It is also not possible to perform Donald & Lang (2007) inference, since the model is just-identified.[20] It is also clear that MVD study an interesting question, and we ultimately want to learn something from the reform in Kentucky. The study by MVD is therefore a good example where sensitivity analysis should be applied.

[18] For calculations see Meyer et al. (1995) p 325.

[19] The terminology ”mean” is not totally accurate. The outcome used by MVD is censored after 42 months. However, at this duration only about 0.5% of the cases are still open. MVD therefore sets all ongoing spells to 42 months. Meyer et al. (1995) also consider other outcome variables and note that their results are quite sensitive to the choice of specification. Here, the focus is on their preferred outcome.

[20] The model includes four variables: a constant, a group dummy, a time dummy and a group time interaction.

Table 3: Sensitivity analysis estimates for application 1 on disability benefits

	Treated (High earnings)		Non-Treated (Low earnings)		Differences		DID
	Pre period [1]	Post Period [2]	Pre period 3]	Post Period [4]	[2-1] [5]	[4-3] [6]	[5-6] [7]
Log duration	1.38	1.58	1.13	1.13	0.20	0.01	0.19
	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)	(0.07)
Sample Size	1,233	1,161	1,705	1,527			
Sensitivity Analysis:							
γ_c - 5 % [10%]					0.0026	-	0.00067
					[0.0041]		[0.00127]
$\sqrt{\gamma_c} * \sigma_\varepsilon$: - 5 % [10%]					0.0629	-	0.0335
					[0.0787]		[0.0461]

Notes: The results in the upper panel are taken from Meyer et al. (1995), their standard errors in parentheses. The outcome is mean log duration, censored after 42 months. The sensitivity analysis results in the lower panel is own calculations. γ_c is calculated by numerically solving for γ in equation (16), for the specified significance level.

Let us start with the basic sensitivity analysis, applicable under the most restrictive assumptions, namely that the cluster-effects (group-time specific shocks) are uncorrelated between the groups as well as uncorrelated over time. The sensitivity analysis presented in Section 3 is then applicable. The γ_c values for 5% level (10-% in brackets) under these assumptions are reported in the lower panel of *Table 3*. We report cut-off values for both the difference estimates as well as the DID estimate.[21] The 5% level cut-off value for the DID estimate is 0.00067. The meaning of this estimate is that the variance of the group-time shocks is allowed to be 0.00067 times the variance of the unobserved individual heterogeneity before the treatment effect is rendered insignificant. At first glance it may seem difficult to assess whether this is a unreasonably large value. *Table 3* therefore also reports these values recalculated into cut-off standard deviations for the group-time shocks ($\sqrt{\gamma_c}\sigma_\varepsilon$). These cut-off values show that the standard deviation of the group-shocks is allowed to be 0.034 on 5% level (0.046 10% level). Column 1 and Column 3 show that the mean of the outcome log injury duration are 1.38 and 1.13 for the treatment group and the control group before the reform. Compared to these means the allowed standard deviation of the shocks is quite large. Furthermore, Column 6 show that the change in injury duration in the control group between the two time periods is 0.01. Even if if does

[21] Notice that no cut-off values are reported for the control group since the difference for this group is already insignificant using the regular standard errors.

not offer conclusive evidence, it suggests that the variance of the group-time shocks is small. Taken together it is therefore fair to say that there is a statistically significant effect on the injury duration.

Next consider an extended sensitivity analysis, which allows for correlation over time in the group-time shocks. In order to take this into account, replace the assumption of no autocorrelation in the cluster effects with an assumption of first order autocorrelation in these shocks. This gives two sensitivity parameters, γ and κ, measuring the size of the cluster effects and the correlation over time in these cluster effects. Since MVD work with repeated cross-section data we can directly apply the results in subsection 4.1. The results from this exercise are presented in *Figure 1*, displaying cut-off values at 10% level for standard deviation of the group-specific time for a range of κ values. In this case with two time periods, a positive autocorrelation in the group-time shocks increases the cut-off values for γ. This extended sensitivity analysis therefore ultimately strengthening the conclusion that there is a statistical significant effect on the injury duration from an increase in disability benefits.

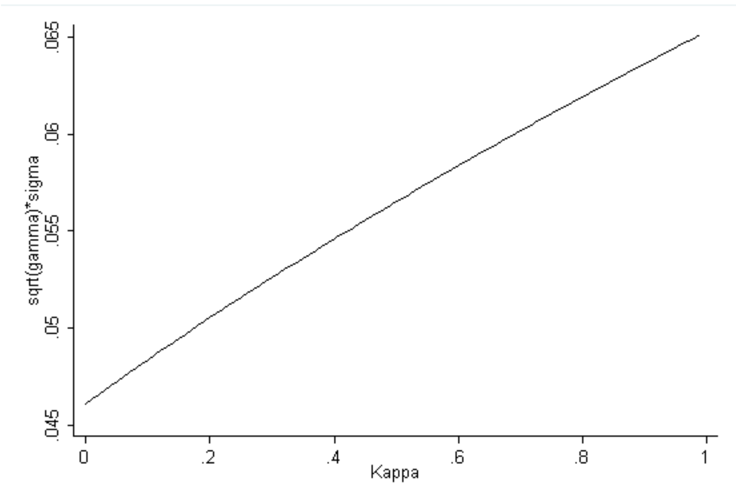

Figure 1: Two parameter sensitivity analysis for the DID estimates in Meyer et al. (1995). Autocorrelation in group-time shocks and allowed standard deviation of the group-time shocks.

6.2 Application 2: Earned income tax credit

Eissa & Liebman (1996)(EL) study the impact of an expansion of the Earned income tax credit (EITC) in the USA on the labor force participation of single women with children. EITC was introduced in 1975. Currently a taxpayer needs to meet three requirements in order to be eligible for the tax credit. The taxpayer need to have positive earned income, the gross income must be below a specified amount, and finally the taxpayer needs to have a qualifying child.[22] The amount of the credit is decided by the taxpayers earned income. The credit is phased in at a certain rate for low incomes, then stays constant within a certain income bracket, and is phased out at a certain rate for higher earnings. High earners are therefore not entitled to any EITC tax credit.

EL study the effects of the 1987 expansion of EITC in USA on labor supply. The reform changed EITC in several ways. The main changes were increases in the subsidy rate for the phase-in of the credit, an increase in the maximum income to which the subsidy rate is applied, and a reduction in the phaseout rate. This resulted in an increase in the maximum credit from $550 to $851, and made taxpayers with income between $11,000 and $15,432 eligible for the tax credit. All these changes made EITC more generous and the treatment consist of the whole change in the budget constraint. Obviously the reform only changes the incentives for those eligible for the tax credit. One key requirement is the presence of a qualifying child in the family. A natural treatment group is then single women with children, and a natural control group is single women without children. However, some single women with children are high income earners and thus are most likely to be unaffected by the EITC reform. EL therefore further divides the sample by education level. Here we report the results for all single women and single women with less than high-school education, from now on referred to as low educated.

EL use CPS data to estimate the treatment effect. Their outcome variable is an indicator variable taking the value one if the annual hours worked is positive. Similarly to MVD they use a DID approach, which contrast the differences between the post- and pre-reform period labor supply for the treatment and the control group. The main results from their analysis are presented in the upper panel of *Table 4*, taken from Table 2 in EL. The results from the DID analysis, presented in Column 7, suggest a positive and statistically significant effect of the EITC expansion in both specifications. If all single women are used, EL estimates that the expansion increased the labor force participation with 2.4 percentage points (4.1 percentage points for low educated single women).

[22] A qualifying child is defined as a child, grandchild, stepchild, or foster child of the taxpayer.

Table 4: Sensitivity analysis estimates for application 2 on earned income tax credit

Sample	Treated (with children) Pre period [1]	Post Period [2]	Non-Treated (without children) Pre period 3]	Post Period [4]	Differences [2-1] [5]	[4-3] [6]	DID [5-6] [7]
All	0.729	0.753	0.952	0.952	0.024	0.000	0.024
	(0.004)	(0.004)	(0.001)	(0.001)	(0.006)	(0.002)	(0.006)
Low education	0.479	0.497	0.784	0.761	0.018	-0.023	0.041
	(0.010)	(0.010)	(0.010)	(0.009)	(0.014)	(0.013)	(0.019)
Sample Size							
All	20,810		46,287				
Low education	5396		3958				
Sensitivity Analysis:							
γ_c - 5 % [10%]							
All					0.00030	-	0.00022
					[0.00048]		[0.00034]
Low education					-	-	0.00005
							[0.00031]
$\sqrt{\gamma_c} * \sigma_\varepsilon$: - 5 % [10%]							
All					0.0075	-	0.0053
					[0.0094]		[0.0066]
Low education					-	-	0.0043
							[0.0080]

Notes: The results in the upper panel are taken from Eissa & Liebman (1996), their standard errors in parentheses. The outcome is an indicator variable taking the value one is hours worked is positive, and zero otherwise. Two different samples, all single women and single women with less than high school. The sensitivity analysis results in the lower panel is own calculations. γ_c is calculated by numerically solving for γ in equation (16), for the specified significance level. The calculations are made under the assumption that the sample size is the same before and after the reform in the two groups.

The inference issues are very similar to those of the MVD study. In the presence of any group-time effects the standard errors presented by EL are biased downwards. We have two DID models, which both are just-identified, making sensitivity analysis an attractive alternative. I first consider sensitivity analysis under assumption of no autocorrelation in the group-time shocks, and then we allow for first order autocorrelation in these shocks. The results from the basic sensitivity analysis is presented in the lower panel of *Table 4*. The 5 percent, γ_c, cut-off value for the two DID estimates is 0.00022 for the full sample and 0.00005 for the sample of low educated mothers. It implies that the variance of the group-time shocks is allowed to be 0.0002 and 0.00005 times the variance of the unobserved individual heterogeneity. It further means that the standard deviation of the group-time shocks is allowed to be about 0.005 for the full sample and about 0.004 for the smaller sample of low educated mothers. In other words even very small shocks render the treatment effect insignificant. It can be compared with the mean labor force participation before the reform, which was 0.73 for all single women with children and 0.48 for low educated single mothers. Single women with children are after all a quite different group compared to single women without children. We can therefore expect quite large group-time specific shocks. Furthermore, there is a large drop of 0.023 in the labor force participation for the control group of low educated single women without children. It therefore seems unreasonable to believe that the variance of shocks is smaller than the variance implied by the cut-off values.

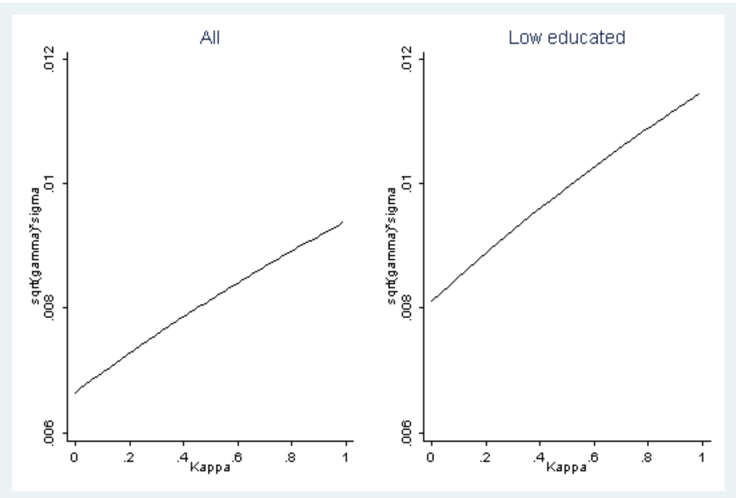

Figure 2: Two parameter sensitivity analysis for the DID estimates in Eissa & Liebman (1996). Left panel: the full sample of single women and right panel: the sample of low educated single women. Autocorrelation in group-time shocks and allowed standard deviation of the group-time shocks.

Next consider allowing for first order autocorrelation in the group-time effects. As in the previous application we use the results in subsection 4.1 for repeated cross-section data. The cut-off standard deviation of the group shocks at 10% level is displayed for a range of κ values in *Figure 2*. The left graph display the cut-off values for the full sample and the right graph displays the cut-off values for the smaller sample of low educated mothers. Introducing autocorrelation in the two group two time period case increases the allowed variance of the group specific shocks. However, the variance is still only allowed to be very small before the estimates are rendered insignificant. We therefore conclude based on the estimates presented, that there is no conclusive evidence of any important labor supply effects from the EITC expansion in 1987.

7 Conclusions

Many policy analyses rely on variation at the group level to estimate the effect of a policy at the individual level. A key example used throughout this paper is the difference-in-differences estimator. The grouped structure of the data introduces correlation between the individual outcomes. This clustering problem has been addressed in a number of different studies. In this paper I have introduced a new method of perform inference when faced with data from only a small number of groups. The proposed sensitivity analysis approach is even able to handle just-identified models, including the often used two group two time period difference-in-differences setting. Consider for example having data for men and women, for two cities or for a couple of villages.

The key feature of the proposed sensitivity analysis approach is that all focus is placed on the size of the cluster effects, or simply the size of the within group correlation. Previously in the applied literature a lot of discussion concerned no within group correlation against non-zero correlation, since these two alternatives imply completely different ways to perform inference. This is a less fruitful discussion. In the end it is the size of the cluster effects which matters. In some cases it is simply not likely to believe that an estimated treatment effect is solely driven by random shocks, since it would require these shocks to have a very large variance. The sensitivity analysis formalizes this discussion by assessing how sensitive the standard errors are to within-group correlation.

In order to demonstrate that our method really can distinguish a causal effect from random shocks the paper offered two different applications. In both applications key regressions are based on just-identified models. The sensitivity analysis results indicate that the conclusion from the first study that the treatment effect is significant is not sensitive to departure from the independence (no-cluster) assumption, whereas the results of the second study are sensitive to the same departure and its conclusion cannot therefore be trusted. More precisely in one of the applications it is not likely that the group effects are so large in comparison with the individual variation that it would render the estimated treatment effect insignificant. In the second application even small within group correlation renders the treatment effect insignificant.

Besides offering a new method of performing inference, this paper contributes by introducing a new type of sensitivity analysis. Previously in the sensitivity analysis literature, the sensitivity of the point estimate has been investigated. This paper shows that sensitivity analysis with respect to bias in the standard errors may be equally important. This opens a new area for future sensitivity analysis research.

References

Abadie, A. (2005), 'Semiparametric difference-in-differences estimators', *Review of Economic Studies* **72**, 1–19.

Abadie, A., Diamond, D. & Hainmuller, J. (2007), Synthetic control methods for comparitive case studies: Estimating the effect of california's tobacco control program. NBER Working Paper No. T0335.

Angrist, J. & Krueger, A. (2000), Empirical strategies in labor economics, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Amesterdam, Elsevier.

Arrelano, M. (1987), 'Computing robust standard errors for within-groups estimators', *Oxford Bulletin of Economcis and Statistics* **49**, 431–434.

Ashenfelter, O. & Card, D. (1985), 'Using the longitudinal structure of earnings to estimate the effect of training programs', *Review of Economics and Statistics* **67**, 648–660.

Athey, S. & Imbens, G. (2006), 'Identification and inference in nonlinear difference-in-differences models', *Econometrica* **74**(2), 431–497.

Bell, R. & McCaffrey, D. (2002), 'Bias reduction in standard errors for linear regression with multi-stage samples', *Survey Methodology* **28**, 169–179.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimators?', *Quarterly Journal of Economics* **1**, 249–275.

Bross, I. (1966), 'Spurious effect from extraneous variables', *Journal of chronic diseases* **19**, 637–647.

Cameron, C., Gelbach, J. & Miller, D. (2007), Robust inference with multi-way cluseter-ing. Mimeo, Department of Economics, University of California at Davis.

Cameron, C., Gelbach, J. & Miller, D. (2008), 'Boostrap-based improvements for inference with clustered erros', *Review of Economics and Statistics* **90**, 414–427.

Campbell, C. (1977), Properties of ordinart and weighted least squares estimators for two stage samples, *in* ‘Proceedings of the Social Statistics Section’, pp. 800–805.

Card, D. & Krueger, A. (1994), ‘Minimum wages and employment: A case of the fast food industry in new jersey and pennsylvania’, *American Economic Review* **84**, 772–784.

Conley, T. & Taber, C. (2005), Inference with difference in differences with a small number of policy changes. NBER Technical Working Paper 312.

Copas, J. & Eguchi, S. (2001), ‘Local sensitivity approximations for selectivity bias’, *J. R. Statist. Soc. B* **83**, 871–895.

Cornfield, J., Haenzel, W., Hammond, E., Lilenfeld, A., Shimkin, A. & Wynder, E. (1959), ‘Smoking and lung cancer: Recent evidence and discussion of some questions’, *J. Nat. Cancer Inst.* **22**, 173–203.

de Luna, X. & Lundin, M. (n.d.), Sensitivity analysis of the unconfoundedness assumption un observational studies. IFAU working paper, forthcoming.

Donald, S. & Lang, K. (2007), ‘Inference with difference-in-differences and other panel data’, *Review of Economics and Statistics* **89**(2), 221–233.

Eberts, R., Hollenbeck, K. & J, S. (2002), ‘Teacher performance incentives and student outcomes’, *Journal of Human Resources* **37**, 913–927.

Eissa, N. & Liebman, J. (1996), ‘Labor supply response to the earned income tax credit’, *Quarterly Journal of Economics* **111**(2), 605–637.

Finkelstein, A. (2002), ‘The effect of tax subsidies to employer-provided supplementary health insurance: Evidence from canada’, *Journal of Public Economics* **84**, 305–339.

Greenwald, B. (1983), ‘A general analysis of bias in the estimated standard errors of least squares coefficients’, *Journal of Econometrics* **22**, 323–338.

Gruber, J. & Poterba, J. (1994), 'Tax incentives and the decision to purchase health insurance', *Quarterly Journal of Economics* **84**, 305–339.

Hansen, C. (2007*a*), 'Asymptotic properties of a robust variance matrix estimator for panel data when t is large', *Journal of Econometrics* **141**, 597–620.

Hansen, C. (2007*b*), 'Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects', *Journal of Econometrics* **140**, 670–694.

Holt, D. & Scott, A. (1982), 'The effect of two-stage sampling on ordinary least squares methods', *Journal of the American Statistical Association* **380**, 848–854.

Ibragimov, R. & Muller, U. (2007), T-statistic based correlation and heterogeneity robust inference. Harvard Insitute of Economic Rsearch, Discussion Paper Number 2129.

Imbens, G. (2003), 'Sensitivity to exogenity assumptions in program evaluation', *American Economic Review* **76**, 126–132.

Imbens, G., Rubin, D. & Sacerdote, B. (2001), 'Estimating the effect of unearned income on labor earnings, savings and consumption: Evidence from a survey of lottery players', *American Economic Review* **91**, 778–794.

Kezdi, G. (2004), 'Robust standard error estimation in fixed-effects panel models', *Hungarian Statistical Review* **Special English Volume 9**, 95–116.

Kloek, T. (1981), 'Ols estimation in a model where a microvariable is explained by aggregates and contemporaneous disturbances are equicorrelated', *Econometrica* **1**, 205–207.

Liang, K.-Y. & Zeger, S. (1986), 'Longitudinal data analysis using generlized linear models', *Biometrika* **73**, 13–22.

Lin, D., Psaty, B. & Kronmal, R. (1998), 'Assessing the sensitivity of regression results to unmeasured counfunders in observational studies', *Biometrics* **54**, 948–963.

Meyer, B. (1995), ‘Natural and quasi-experiments in economics’, *Journal of Buisness and Economic Statistics* **13**, 151–161.

Meyer, B., Viscusi, W. & Durbin, D. (1995), ‘Workers’ compensation and injury duration: Evidence from a natural experiment’, *American Economic Review* **85**(3), 322–340.

Moulton, B. (1986), ‘Random group effects and the precision of regression estimates’, *Journal of Econometrics* **32**, 385–397.

Moulton, B. (1990), ‘An illustration of a pitfall in estimating the effects of aggregate variables on micro units’, *Review of Economics and Statistics* **72**, 334–338.

Rosenbaum, P. (2004), ‘Design sensitivity in observational design’, *Biometrika* **91**(1), 153–164.

Rosenbaum, P. & Rubin, D. (1983), ‘Assesing sensitivity to an unobserved binary covariate in an observational study with binary outcome’, *Journal of the Royal Statistical Society, Series B* **45**(2), 212–218.

White, H. (1980), ‘A heteroscedasticity-consistent covariance matrix estimator and a direct test of heteroscedasticity’, *Econometrica* **48**, 184–200.

Wooldridge, J. (2003), ‘Cluster-sample methods in applied econometrics’, *American Economic Review* **93**, 133–138.

Wooldridge, J. (2006), Cluster-sample methods in applied econometrics an extended analysis. Mimeo Department of Economics, Michigan State University.

Appendix

Derivation of equation 9.

Start with equation (8)

$$V = \frac{N-K}{N-tr[(X'X)^{-1}X'CX]}(X'X)^{-1}X'CX\mathbb{E}(\hat{V}).$$

First, consider $(X'X)^{-1}X'CX$. Under assumption of $x_{igt} = x_{gt}$ we have

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_G \end{bmatrix} \qquad x_g = \begin{bmatrix} l_{g1}x'_{g1} \\ l_{g2}x'_{g2} \\ \vdots \\ l_{gT}x'_{gT} \end{bmatrix}$$

, where l_{gt} is a column vector of n_{gt} ones, G is the number of groups and T is the number of time periods. If $\mathbb{E}(c_{gt}c_{g't}) = 0$ for all t and all $g \neq g'$, and $\mathbb{E}(c_{gt}c_{gt'}) = 0$ for all g and all $t \neq t'$, we further have

$$C = \begin{bmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_G \end{bmatrix} C_g = \begin{bmatrix} C_{g1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{gT} \end{bmatrix}$$

with

$$C_{gt} = \begin{bmatrix} 1 & p & \dots & p \\ p & 1 & & \ddots \\ \vdots & & \ddots & p \\ p & \dots & p & 1 \end{bmatrix} = [(1-p)I_{gt} + pl_{gt}l'_{gt}]$$

Here I_{gt} is a unit matrix of order n_{gt}, and $p \equiv \frac{\sigma_c^2}{\sigma_c^2+\sigma_\varepsilon^2}$. It follows that

$$X'X = \sum_g \sum_t n_{gt}x_{gt}x'_{gt} \tag{A.1}$$

and

$$X'CX = \sum_g \sum_t x_{gt} l'_{gt} C_{gt} l_{gt} x'_{gt}. \tag{A.2}$$

and

$$x_{gt} l'_{gt} C_{gt} l_{gt} x'_{gt} = x_{gt} l'_{gt} \begin{bmatrix} 1+(n_{gt}-1)p \\ 1+(n_{gt}-1)p \\ \vdots \\ 1+(n_{gt}-1)p \end{bmatrix} x'_{gt} = x_{gt} n_{gt} [1+(n_{gt}-1)p] x'_{gt} \tag{A.3}$$

Combining equation (A.1), (A.2) and (A.3) gives

$$X'XX'CX = (\sum_g \sum_t n_{gt} x_{gt} x'_{gt})^{-1} \sum_g \sum_t n_{gt} \tau_{gt} x_{gt} x'_{gt} \tag{A.4}$$

with

$$\tau_{gt} = 1 + (n_{gt} - 1)p.$$

Imposing $n_{gt} = n$ we have equation (A.4) as

$$X'XX'CX = \tau I_K \tag{A.5}$$

with

$$\tau = 1 + (n-1)p.$$

Next consider $\frac{N-K}{N-tr[(X'X)^{-1}X'CX]}$: using the result in equation (A.5) gives

$$tr[(X'X)^{-1}X'CX] = K\tau. \tag{A.6}$$

Substituting equation (A.5) and equation (A.6) into equation (8) and imposing $n_{gt} = n$ (then $N = nGT$) gives

$$V = \mathbb{E}(\hat{V})\tau \frac{nGT - K}{nGT - K\tau},$$

i.e. equation (9).

Derivation of equation 18.

Again start with equation (8)

$$V = \frac{N-K}{N-tr[(X'X)^{-1}X'CX]}(X'X)^{-1}X'CX\mathbb{E}(\hat{V}).$$

First, consider $(X'X)^{-1}X'CX$. Remember that C is defined as

$$\mathbb{E}(ee') = \sigma^2 C,$$

where e is a vector collecting all $e_{igt} = c_{gt} + \varepsilon_{igt}$, and $\sigma^2 \equiv 1/Ntr(ee')$. In order to express C in terms of κ and γ use the well know properties of an AR(1) process (under the assumption of $|\kappa| < 1$), and the definition of $\sigma_d^2 \equiv \gamma\sigma_\varepsilon^2$ from section 4.1. This gives

$$\sigma_c^2 = \mathbb{E}(c_{gt}c_{gt}) = \frac{\sigma_d^2}{1-\kappa^2} = \frac{\gamma\sigma_\varepsilon^2}{1-\kappa^2} \tag{A.7}$$

and if $t \neq t'$

$$Cov(c_{gt}c_{gt'}) = \mathbb{E}(c_{gt}c_{gt'}) = \kappa^{|t-t'|}\frac{\sigma_d^2}{1-\kappa^2} = \kappa^{|t-t'|}\frac{\gamma\sigma_\varepsilon^2}{1-\kappa^2}. \tag{A.8}$$

Thus if $i = j$

$$\mathbb{E}(e_{igt}e_{jgt}) = \sigma^2 = \sigma_c^2 + \sigma_\varepsilon^2 = \frac{\gamma\sigma_\varepsilon^2}{1-\kappa^2} + \sigma_\varepsilon^2 = \sigma_\varepsilon^2\frac{1+\gamma-\kappa^2}{1-\kappa^2}. \tag{A.9}$$

Further, using (A.7) and (A.9), if $i \neq j$

$$\mathbb{E}(e_{igt}e_{jgt}) = \sigma_c^2 = \frac{\gamma\sigma_\varepsilon^2}{1-\kappa^2} = \sigma^2\frac{\gamma}{1+\gamma-\kappa^2} \tag{A.10}$$

and using (A.8) and (A.9), if $t \neq t'$

$$\mathbb{E}(e_{igt}e_{jgt'}) = \kappa^{|t-t'|}\frac{\sigma_d^2}{1-\kappa^2} = \sigma^2\kappa^{|t-t'|}\frac{\gamma}{1+\gamma-\kappa^2}. \tag{A.11}$$

and under assumption $\mathbb{E}(c_{gt}c_{gt'}) = 0$, if $g \neq g'$

$$\mathbb{E}(e_{igt}e_{jg't'}) = 0 \tag{A.12}$$

Then under (A.9), (A.10), (A.11) and (A.12)

$$C = \begin{bmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_G \end{bmatrix} C_g = \begin{bmatrix} C_{11} & \dots & C_{T1} \\ \vdots & \ddots & \vdots \\ C_{1T} & \dots & C_{TT} \end{bmatrix} \tag{A.13}$$

with if $t = t'$

$$C_{tt'} = [(1 - \frac{\gamma}{1+\gamma-\kappa^2})I_{gt} + \frac{\gamma}{1+\gamma-\kappa^2} l_{gt} l'_{gt}] \tag{A.14}$$

and if $t \neq t'$

$$C_{tt'} = \kappa^{|t-t'|} \frac{\gamma}{1+\gamma-\kappa^2} l_{gt} l'_{gs}. \tag{A.15}$$

Define $p_c = \frac{\gamma}{1+\gamma-\kappa^2}$. Then, using equation (A.14), if $t = t'$

$$x_{gt} l'_{gt} C_{tt} l_{gt} x'_{gt} = x_{gt} l'_{gt} \begin{bmatrix} 1+(n_{gt}-1)p_c \\ 1+(n_{gt}-1)p_c \\ \vdots \\ 1+(n_{gt}-1)p_c \end{bmatrix} x'_{gt} = x_{gt} n_{gt} [1+(n_{gt}-1)p_c] x'_{gt}, \tag{A.16}$$

and, using equation (A.15), if $t \neq t'$

$$x_{gt} l'_{gt} C_{tt'} l_{gt'} x'_{gt'} = \kappa^{|t-t'|} x_{gt} l'_{gt} \begin{bmatrix} n_{gt'} p_c \\ n_{gt'} p_c \\ \vdots \\ n_{gt'} p_c \end{bmatrix} x'_{gt'} = \kappa^{|t-t'|} x_{gt} n_{gt} p_c n_{gt'} x'_{gt'} \tag{A.17}$$

Using equation (A.1), (A.16) and (A.17) gives

$$X'XX'CX = (\sum_g \sum_t n_{gt} x_{gt} x'_{gt})^{-1} \tag{A.18}$$

$$(\sum_g \sum_t \sum_{t' \neq t} (\kappa^{|t-t'|} n_{gt} n_{gt'} x_{gt} x'_{gt'}) + n_{gt}[1+(n_{gt}-1)p_c] x_{gt} x'_{gt})$$

Imposing $n_{gt} = n$, substituting for $p_c = \frac{\gamma}{1+\gamma-\kappa^2}$ and simplifying we have equation (A.18) as

$$X'XX'CX = (1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2})I_K + n\frac{\gamma}{1+\gamma-\kappa^2}H \qquad \text{(A.19)}$$

with

$$H = (\sum_g \sum_t x_{gt} x'_{gt})^{-1} \sum_g \sum_t \sum_{t' \neq t} (\kappa^{|t-t'|} x_{gt} x'_{gt'}).$$

Next consider $\frac{N-K}{N-tr[(X'X)^{-1}X'CX]}$: using the results in (A.19) gives

$$tr[(X'X)^{-1}X'CX] = K(1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2}) + n\frac{\gamma}{1+\gamma-\kappa^2}\sum_{a=1}^{K} H_{aa}. \qquad \text{(A.20)}$$

where H_{aa} is the element in the ath column and ath row of H.

Substituting equation (A.19) and equation (A.20) into equation (8) and noting that under $n_{gt} = n$ we have $N = nGT$ gives

$$V_{aa} = \mathbb{E}(\hat{V}_{aa}) \frac{nGT - K}{nGT - K(1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2}) + n\frac{\gamma}{1+\gamma-\kappa^2}\sum_{a=1}^{K} H_{aa}}$$

$$(1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2}) + n\frac{\gamma}{1+\gamma-\kappa^2}H_{aa}$$

Both the first and the second part of this expression, the two sources of bias in the standard errors are greater than one. However, it will be highly dominated by $(1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2})I_K + n\frac{\gamma}{1+\gamma-\kappa^2}H_{aa}$. Thus we have

$$V_{aa} \approx \mathbb{E}(\hat{V}_{aa})(1+(n-1)\frac{\gamma}{1+\gamma-\kappa^2}) + n\frac{\gamma}{1+\gamma-\kappa^2}H_{aa}.$$

i.e. equation (18).

Derivation of equation 22.

Again start with equation (8)

$$V = \frac{N-K}{N-tr[(X'X)^{-1}X'CX]}(X'X)^{-1}X'CX\mathbb{E}(\hat{V}).$$

Using the definition $\sigma_c^2 \equiv \gamma\sigma_\varepsilon^2$ from section 4.1, if $i = j$ we have

$$\mathbb{E}(e_{igs}e_{jgs}) = \sigma^2 = \sigma_c^2 + \sigma_\varepsilon^2 = \sigma_\varepsilon^2(1+\gamma). \tag{A.21}$$

Using this and under assumption $\mathbb{E}(c_{gt}c_{g't}) = 0$ for all t, and the multiway clustering assumptions if $s \neq s'$ then $E(c_{gs}c_{g's'}) = 0$ and if $s = s'$ then $E(c_{gs}c_{g's'}) = \xi\sigma_c^2$, it gives if $i \neq j$

$$\mathbb{E}(e_{igs}e_{jgs}) = \sigma_c^2 = \gamma\sigma_\varepsilon^2 = \sigma^2\frac{\gamma}{1+\gamma} \tag{A.22}$$

and if $i \neq j$ and $g = g'$ holds

$$\mathbb{E}(e_{igs}e_{jg's}) = \xi\sigma_c^2 = \xi\gamma\sigma_\varepsilon^2 = \sigma^2\xi\frac{\gamma}{1+\gamma} \tag{A.23}$$

and if $s \neq s'$

$$\mathbb{E}(e_{igs}e_{jg's}) = 0. \tag{A.24}$$

Thus under (A.21), (A.22), (A.23) and (A.24)

$$C = \begin{bmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_S \end{bmatrix} C_s = \begin{bmatrix} C_{11} & \dots & C_{G_s1} \\ \vdots & \ddots & \vdots \\ C_{1G_s} & \dots & C_{G_sG_s} \end{bmatrix} \tag{A.25}$$

with if $g = g'$

$$C_{gg'} = [(1 - \frac{\gamma}{1+\gamma})I_g + \frac{\gamma}{1+\gamma}l_g l_s] \tag{A.26}$$

and if $g \neq g'$

$$C_{gg'} = \xi\frac{\gamma}{1+\gamma}l_g l_g'. \tag{A.27}$$

Here G_s is the number of groups belonging to group-cluster s. Retain the definition $p =$

$\frac{\gamma}{1+\gamma}$, and define l_{gs} as a column vector of n_{gs} ones. Then, using equation (A.26), if $g = g'$

$$x_{gs}l'_{gs}C_{gg}l_{gs}x'_{gs} = x_{gs}l'_{gs}\begin{bmatrix}1+(n_{gs}-1)p\\1+(n_{gs}-1)p\\\vdots\\1+(n_{gs}-1)p\end{bmatrix}x'_{gs} = x_{gs}n_{gs}[1+(1-n_{gs})p]x'_{gs}, \tag{A.28}$$

and, using equation (A.27), if $g \neq g'$

$$x_{gs}l'_{gs}C_{gg'}l_{g's}x'_{g's} = \xi x_{gs}l'_{gs}\begin{bmatrix}n_{g's}p\\n_{g's}p\\\vdots\\n_{g's}p\end{bmatrix}x'_{g's} = \xi x_{gt}n_{gs}p_c n_{g's}x'_{g's} \tag{A.29}$$

Using equation (A.1), (A.28) and (A.29) gives

$$X'XX'CX = (\sum_s\sum_g n_{gs}x_{gs}x'_{gs})^{-1} \tag{A.30}$$

$$(\sum_s\sum_g\sum_{g'\neq g}(\xi n_{gs}n_{g's}x_{gs}x'_{g's}) + n_{gs}[1+(1-n_{gs})p]x_{gs}x'_{gs})$$

Imposing $n_{gs} = n$, substituting for $p = \frac{\gamma}{1+\gamma}$ and simplifying we have equation (A.30) as

$$X'XX'CX = (1+(n-1)\frac{\gamma}{1+\gamma})I_K + n\frac{\gamma}{1+\gamma}\xi M, \tag{A.31}$$

with

$$M = (\sum_s\sum_g x_{gs}x'_{gs})^{-1}\sum_s\sum_g\sum_{g'\neq g}(x_{gs}x'_{g's}).$$

Next consider $\frac{N-K}{N-tr[(X'X)^{-1}X'CX]}$, using the results in (A.31) gives

$$tr[(X'X)^{-1}X'CX] = K(1+(n-1)\frac{\gamma}{1+\gamma}) + n\frac{\gamma}{1+\gamma}\xi\sum_{a=1}^{K}M_{aa}. \tag{A.32}$$

where M_{aa} is the element in the ath column and ath row of M.

Substituting equation (A.31) and equation (A.32) into equation 8 and noting that under $n_{gt} = n$ we have $N = nGT$ gives

$$V_{aa} = \mathbb{E}(\hat{V}_{aa}) \frac{nGT - K}{nGT - K(1 + (n-1)\frac{\gamma}{1+\gamma}) + n\frac{\gamma}{1+\gamma}\xi \sum_{a=1}^{K} M_{aa}}$$

$$(1 + (n-1)\frac{\gamma}{1+\gamma} + n\frac{\gamma}{1+\gamma}\xi M_{aa}).$$

Both the first and the second part of this expression, the two sources of bias in the standard errors, are greater than one. However, it will be highly dominated by $(1 + (n-1)\frac{\gamma}{1+\gamma} + n\frac{\gamma}{1+\gamma}\xi M_{aa})$. Thus we have

$$V_{aa} \approx \hat{V}_{aa}(1 + (n-1)\frac{\gamma}{1+\gamma} + n\frac{\gamma}{1+\gamma}\xi M_{aa})$$

i.e equation (22).

Publication series published by the Institute for Labour Market Policy Evaluation (IFAU) – latest issues

Rapporter/Reports

2009:1 Hartman Laura, Per Johansson, Staffan Khan and Erica Lindahl, ”Uppföljning och utvärdering av Sjukvårdsmiljarden”

2009:2 Chirico Gabriella and Martin Nilsson ”Samverkan för att minska sjukskrivningar – en studie av åtgärder inom Sjukvårdsmiljarden”

2009:3 Rantakeisu Ulla ”Klass, kön och platsanvisning. Om ungdomars och arbetsförmedlares möte på arbetsförmedlingen”

2009:4 Dahlberg Matz, Karin Edmark, Jörgen Hansen and Eva Mörk ”Fattigdom i folkhemmet – från socialbidrag till självförsörjning”

2009:5 Pettersson-Lidbom Per and Peter Skogman Thoursie ”Kan täta födelseintervaller mellan syskon försämra deras chanser till utbildning?”

2009:6 Grönqvist Hans ”Effekter av att subventionera p-piller för tonåringar på barnafödande, utbildning och arbetsmarknad”

2009:7 Hall Caroline ”Förlängningen av yrkesutbildningarna på gymnasiet: effekter på utbildningsavhopp, utbildningsnivå och inkomster”

2009:8 Gartell Marie ”Har arbetslöshet i samband med examen från högskolan långsiktiga effekter?”

2009:9 Kennerberg Louise ”Hur försörjer sig nyanlända invandrare som inte deltar i sfi?”

2009:10 Lindvall Lars ”Bostadsområde, ekonomiska incitament och gymnasieval”

2009:11 Vikström Johan “Hur påverkade arbetsgivaransvaret i sjukförsäkringen lönebildningen?”

2009:12 Liu Qian and Oskar Nordström Skans ”Föräldraledighetens effekter på barnens skolresultat

2009:13 Engström Per, Hans Goine, Per Johansson and Edward Palmer ”Påverkas sjukskrivning och sjukfrånvaro av information om förstärkt granskning av läkarnas sjukskrivning?”

2009:14 Goine Hans, Elsy Söderberg, Per Engström and Edward Palmer ”Effekter av information om förstärkt granskning av medicinska underlag”

Working papers

2009:1 Crépon Bruno, Marc Ferracci, Grégory Jolivet and Gerard J. van den Berg “Active labor market policy effects in a dynamic setting”

2009:2 Hesselius Patrik, Per Johansson and Peter Nilsson “Sick of your colleagues’ absence?”

2009:3 Engström Per, Patrik Hesselius and Bertil Holmlund “Vacancy referrals, job search and the duration of unemployment: a randomized experiment”

2009:4 Horny Guillaume, Rute Mendes and Gerard J. van den Berg ”Job durations with worker and firm specific effects: MCMC estimation with longitudinal employer-employee data”

2009:5 Bergemann Annette and Regina T. Riphahn “Female labor supply and parental leave benefits – the causal effect of paying higher transfers for a shorter period of time”

2009:6 Pekkarinen Tuomas, Roope Uusitalo and Sari Kerr “School tracking and development of cognitive skills”

2009:7 Pettersson-Lidbom Per and Peter Skogman Thoursie “Does child spacing affect childrens’s outcomes? Evidence from a Swedish reform”

2009:8 Grönqvist Hans “Putting teenagers on the pill: the consequences of subsidized contraception”

2009:9 Hall Caroline “Does making upper secondary school more comprehensive affect dropout rates, educational attainment and earnings? Evidence from a Swedish pilot scheme”

2009:10 Gartell Marie “Unemployment and subsequent earnings for Swedish college graduates: a study of scarring effects”

2009:11 Lindvall Lars “Neighbourhoods, economic incentives and post compulsory education choices”

2009:12 de Luna Xavier and Mathias Lundin “Sensitivity analysis of the unconfoundedness assumption in observational studies”

2009:13 Vikström Johan “The effect of employer incentives in social insurance on individual wages”

2009:14 Liu Qian and Oskar Nordström Skans ”The duration of paid parental leave and children’s scholastic performance”

2009:15 Vikström Johan “Cluster sample inference using sensitivity analysis: the case with few groups”

Dissertation series

2009:1 Lindahl Erica “Empirical studies of public policies within the primary school and the sickness insurance”

2009:2 Grönqvist Hans “Essays in labor and demographic economics”