Faculty of Business, Economics and Law
School of Economics

Applied Causal Inference using Identification Robust Confidence Sets Under Sparsity

An Honours Thesis submitted to the School of Economics, The University of Queensland, in partial fulfilment of the requirements for the degree of BEcon(Honours)

Supervisors

Dr. Christiern Rose

Dr Rigissa Megalokonomou

By

Tran Hoang Phuong Linh

October 24, 2019

Approximate word count: 15,000

Acknowledgements

I would like to express my deepest gratitudes and appreciations for my supervisors, Dr. Christiern Rose and Dr Rigissa Megalokonomou. Through out my thesis progress, Christiern, with your time-management and thorough knowledge in the field, has tremendously guided me through the year and many obstacles, especially in such an uncharted territory for me. Thank you so much for your time and patience helping me understand many difficult concepts and technicalities. I have also gained much more confidence and insight on how sensible my outputs are in real-life sense thanks to Rigissa advices and words of courage. Without your help, it would not have been possible for me to get this far.

I would also like to extend my gratitudes to my parents, whose unlimited supports and kindness have helped me through tough times and kept me going.

Last but not least, it is an honour to be accompanied by a group of brilliant people, the Honour fellows and lecturers. Your optimism, enthusiasm and keen understandings on all kinds of fields have had a great influence on my own works, personalities and possibly my future route. From those, I offer my sincerest thanks.

I declare that the work presented in this Honours thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text, and that material has not been submitted, either in whole or in part, for a degree at this or any other university.

Tran Hoang Phuong Linh

October 24, 2019

Contents

1 Abstract

One of the main issues regarding the validity of a regression is when the independent variable is correlated with the error term, resulting in endogeneity problem which leads to biased estimates.. 2SLS is typically used with several assumptions such as exclusion restrictions and strong instruments. This paper applies a new method proposed in Gautier et al. (2018) (SNIV) with the idea of sparsity, which relaxes the standard 2SLS assumptions on the excluded instruments specification and instrument strength. We first consider the same exclusion restrictions as 2SLS, followed by relaxing the exact location of the excluded instruments. This is the first time SNIV has been applied to these real-life datasets, and by comparing the outcomes, the similarities and the discrepancies in the outcomes will accordingly insinuate the possible validation of the assumptions made for 2SLS.

The conclusions show, under similar assumptions of exclusion restrictions, 2SLS and SNIV yield similar results when the instruments are strong, while the former estimates and standard errors are misleading when the instruments are weak. Relaxing excluded restrictions also shed lights on the behaviour of SNIV in different datasets, including sample size and strength of the instruments.

2 Introduction

2.1 Motivation

Statistics have always been used to answer questions quantitatively. For instance, some concern estimating the consequences of a specific policy, aiding a decision maker; others might be about how much a firm should change its output to maximize profits given a shift in economic situation. One of the most crucial attentions regarding the validity of the results is the assumptions made: if these are violated, the results will be not credible. There have been many attempts to relax the assumptions while giving meaningful results with the inventions of new and creative statistical tools. Let us take the most basic OLS as an example. One of the requirements for OLS to perform well is a correct model specification. Non-parametric regression was created to relax this while assuming no particular form of the model.

This paper will be focusing on the endogeneity problem, where the independent variable is correlated with the error term, causing bias in the estimates and misleading results. Instead of the usual instrumental variables settings, new methods with weaker assumptions (to some degree) are used. Particularly, if the conventional method assuming exclusion restrictions on all of the instrument variables, the new techniques substitute this with "sparsity", which only requires a certain number of exclusion restrictions without knowing the restrictions' exact locations. It is also robust to weak IVs since there is no particular restriction on the joint distribution of the endogenous variables and the proposed instruments. Furthermore, the technique will still be working in high-dimensional settings where the usual IVs estimates perform poorly as the number of independent variables and instruments increases. However, the methods are more computationally intensive than traditional approaches such as two stage least squares by relying mostly on solving optimisation problems without a closed form solution (no first stage).

Since the method (SNIV) has not been yet applied to any real life data (only simulated data), it would be interesting to see the performances of this new method on some influential articles of choice. The main goal is to use SNIV side by side with the standard instrumental variables estimates under identical and slightly weaker assumptions. Then, by comparing the results, if the figures are similar under such weaker presumptions, it is safe to conclude the original results applying 2SLS is relatively robust. However, if there is a discrepancy between the two methods, it should be seen as a signal hinting the underlying assumptions might not be as valid as claimed to be.

2.2 Outline

For the next part, the paper will be briefly talking about endogeneity and the standard IVs estimators, with some introduction on the new methods. Section 3 tackles the methodology used, including the details on assumptions made, different types of techniques with their equivalent implementation and comparison with each other. Data implementation is in section 4 where three chosen articles are presented using both old and new methods to produce some conclusions. Section 5 raises some of the main drawbacks of the basic methodology and the proposals to adjust the problems. The paper ends with a summary and conclusions.

3 Literature review

This section will briefly go through the problem of endogeneity together with the conventional instrumental variables estimates solution, following by the "sparsity" concept in some recent studies and data selection suggestions for section 4.
Given the following model:

$$\mathbf{y} = \boldsymbol{\beta}_{endo}\mathbf{x}_{endo} + \boldsymbol{\beta}_{exo}\mathbf{x}_{exo} + \boldsymbol{\epsilon}$$

where $\mathbf{y}$ is the dependent variable.

$\mathbf{x}_{endo}$ are the endogenous variables.

$\mathbf{x}_{exo}$ are the exogenous variables.

$Cov(\mathbf{x}_{endo}, \boldsymbol{\epsilon}) \neq 0$, independent variables are partly correlated with the error term, which causes OLS estimates to be biased and gives misleading results.

The IV estimators can be used to solve the problem by using instrumental variables (IVs) called $\mathbf{z}$. For this method to work, $\mathbf{z}$ must satisfy two properties: $\mathbf{z}$ should be (i) highly correlated with the endogenous variables, and (ii) not contained in the right-hand side of the equation or satisfying exclusion restrictions. The former can be simply examined using first-stage equations while the later remains as an untestable assumption in the exactly identified case (the number of instruments equal to the number of parameters). This is where the idea of sparsity was implemented to weaken the assumption (ii) and obtain a more robust version.

Sparsity has been practised in high-dimensional statistics under the concept of regularization, which is especially useful in high-dimensional data models where the number of dimensions (or the number of independent variables) is relatively large compared to the sample size (or even larger).

Let us take LASSO (Least absolute shrinkage and selection operator, introduced by Frank and Friedman(1993) and Tibshirani (1996)) as an example. Given K independent variables and N observations, the usual multivariate statistics such as OLS fails if K is of order larger than N. LASSO, on the other hand, allows for less important regressors to be exactly zero in coefficients, and therefore still gives relatively more meaningful predictions. If it is indeed true that many coefficients in K are zero, the coefficients can be seen as being sparse: that is among K regressors, only some of the variables' coefficients are significant. "Lasso Regularization for Selection of Log-linear Models: An Application to Educational Assortative Mating" (Mauricio Bucca, Daniela R. Urbina (2019)) applies LASSO and compares the results with the conventional approach, which uses information criteria to choose among many different models. Notice that as the number of independent variables increases, the number of reasonable model specifications will also increase at a faster rate. This poses a critical problem when using conventional methods as only a handful of models can be estimated based on pure intuition and past workings, and the "best" one is chosen according to AIC and BIC. As pointed out by the papers, these usually lead to inconsistent results and complicated models unless the data is very large and non-sparse in simulated cases. LASSO, on the other hand, chooses 55 out of 300 parameters in simulated cases and at most 150 parameters out of 300 in empirical data, both give much more parsimonious model and consistent predictions. There are other articles such as Yuri Fujino, Hiroshi Murata, Chihiro Mayama, and Ryo Asaoka (2015) or Jorge A. Chan-Lau (2017) where the forecasts using conventional choosing models basing on theory can be improved by assuming sparsity among many potential regressors and let LASSO does the selection work.

It can be clearly seen that by making use of the idea of sparsity, the researchers can rely less on intuition and previous studies findings when it comes to finding appropriate regressors while still giving out sensible results. However, it should be noticed that LASSO being great at predicting outcome variables does not mean parameters inferences on these models are valid. This is especially important when it comes to instrumental variables estimates where casual inferences on the endogenous variables are the end goals. Many solutions of how to implementing sparsity to IV estimates were introduced.

Belloni, Chen, Chernozhukov, and Hansen (2012) proposed the idea of using LASSO in the first-stage (predictive-stage) equation to choose large-coefficient instrumental variables among many potential ones, assuming that the first-stage is sparse and omitting small, non-zero valid instruments from this stage will not affect the results. The predicted endogenous variables are then used as normal in the second stage to get the desired estimates. The sparsity assumption is applied only for selecting instruments and exclusion restrictions assumption still maintained.

When it comes to applying LASSO to select many potential controls, things are more complicated, and Belloni, Chernozhukov, and Hansen (2014) suggests using double selection. Particularly, sparsity is practised to sort out useful predictors for outcome variables and the endogenous variables, with the union of these set as controls variables. The article also summarises three different applications, pointing out that applying the sparsity concept improves the quality of the estimates by allowing for more variables and many of their alternate functional forms to be considered instead of just using theory and past studies to choose a small set of instruments.

There are many others such as Zhang and Zhang (2014), van de Geer, Buhlmann, and Ritov, et al. (2014). However, these articles usually assume few endogenous independent variables with strong exclusion restrictions assumption on the instruments variable, linear and sparse first stage (Eric Gautier, Chris Rose, Alexandre Tsybakov (2018)). Also notice that, even in lower dimension framework, the assumption of exclusion restrictions and strong instruments are required for the IV regressions to be valid. Bound, Jaeger, and Baker (1995) revisits Angrist and Krueger (1991) to demonstrate the potential inconsistency and even finite-sample bias coming from the weak instrumental variable estimates despite of the large sample size. Another evidence of void estimates from the use weak instruments in IV estimates is investigated in Stock and Staiger (1994), where asymptotic distribution theory is derived for the instrumental variable estimates with one endogenous variable weakly correlated with the proposed instruments. The results suggest the basic TSLS models are biased even in large sample size and needs to be modified.

From all of the above, since strong exclusion restrictions and weak instruments violations are problematic regardless of sample size, Gautier et al. (2018) proposes a new method where the these assumptions of the excluded instruments and theirs strength are relaxed, to some extent, using sparsity concept, which is also where the main methodology used in this paper derived from. Generally, it requires only part of a range of potential instruments satisfies exclusion restrictions, not all of them, and is robust to weak instruments. The details of the methods, which is known as SNIV, will be fully addressed in the next section.

Concerning data used to implement old and new techniques, there are several requirements. The articles should be cross-sectional data, meaning no panel or time series data. This is due to the results being based on independent samples where there should be no autocorrelation. For computational reasons related to SNIV, the number of variables should not be too large (particularly, less than fifty in total). These criteria are applied for simplicity purpose and can absolutely be relaxed with appropriate adjustment. We apply the approach to the following three articles: (i) Angrist and Evan (1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", (ii) Angrist, Graddy and Imbens (2000), "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish", and (iii) David Card (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". All of these articles used the conventional instrumental variables estimates and their results can be compared to those of SNIV estimates. More details are provided in section 4.

4 Methodology

The method is mainly concerned with the problem of endogeneity, where the conventional first-stage-second-stage instrument variables regressions are used. For a more detailed description formally and mathematically of the method, refer to Eric Gautier, Chris Rose, Alexandre Tsybakov, "High-Dimensional Instrumental Variables Regression and Confidence Sets", (2018). The highlighted idea is, instead of allowing for exclusion restrictions, a considerably weaker assumption (sparsity) is used. Specifically, if the former requires the exact locations of the instruments, the later only assuming the number of "true" instruments (which satisfy exclusion restrictions) among a range of potential instrumental variables. There are also other differences, which will be discussed in detail later on. Another crucial note is there is no two-stage procedure with closed-form estimates. In fact, the method relies on solving optimization problems instead and therefore will be computationally more intensive.

4.1 Model set up

The main purpose of the paper is to compare the performances of this new method with the typical instrumental variables regressions under similar assumptions. The end results are not about which method is superior; rather the differences between these estimates in many real-life data will draw some possible explanations on why the results turn out the way it is and potentially show which are more reliable in certain situations.

For simplicity purposes, data types are restrained to only cross-sectional data, meaning no panel data will be considered, and only parametric models. Let N be the number of observations, i = 1,2,...,N. The general model set up is as follow:

$$\mathbf{y}_i = \mathbf{x}_i^T \boldsymbol{\beta}_{\mathbf{x}} + \mathbf{z}_i^T \boldsymbol{\beta}_{\mathbf{z}} + \boldsymbol{\epsilon}$$
$$\text{or } \mathbf{y}_i = \mathbf{w}_i^T \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{w}_i^T = \begin{bmatrix} \mathbf{x}_i^T & \mathbf{z}_i^T \end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{x}}^T & \boldsymbol{\beta}_{\mathbf{z}}^T \end{bmatrix}$; $\mathbb{E}[\mathbf{z}_i \boldsymbol{\epsilon}_i] = 0$

$\mathbf{y}$ is the dependent variable.

$\mathbf{x}$ composes of P endogenous variables.

$\mathbf{z}$ composes of L included exogenous variables and potential instrumental variables.

$\mathbf{w}$ composes of all independent variables, K = P+L in total

The K independent variables can be partitioned into three non-overlapping and exhaustive groups:

- The included regressors: are variables typically considered as control variables and endogenous variables, has significant coefficients ($\boldsymbol{\beta}$ can be non-zero) and contributes to the explanation of independent variable. Let J_{inc} denote the indices of these variables.

- The excluded regressors: the coefficients are equal to 0 ($\boldsymbol{\beta} = 0$), which satisfy exclusion restrictions condition and If they are exogenous, they can be considered as instrumental variables. Let J_{exc} denote the indices of these variables.

- The uncertain regressors: where it is unknown whether these coefficients are excluded or included ($\boldsymbol{\beta} \neq 0$ or $= 0$). However, these are assumed to satisfy the "sparsity" condition. Let J_{unc} denote the indices of these variables. Notice that this is a new group proposed by SNIV as the 2SLS estimates only consider the other two groups.

4.2 Assumptions

One of the most important assumptions is sparsity. In this context, s sparsity means, among the uncertain regressors, there are at most s included variables (and the rest is excluded) but the exact locations of which are unidentified. In other words, in this group, it is assumed to have at least $dim(J_{unc}) - s$ exclusion restrictions, the identities of which are unknown. The larger s is, the less sparse it is. On the other hand, as s progressively gets smaller, the assumption gets stronger and we will have $dim(J_{unc})$ instrumental variables at s = 0. s sparsity assumption for J_{unc} can also be expressed as:

$$|\boldsymbol{\beta}_{J_{unc}}|_0 \leq s, |\boldsymbol{\beta}_{J_{unc}}|_0 = \sum_{i \in J_{unc}} 1(\boldsymbol{\beta}_i \neq 0)$$

For different settings, the allocations of variables among the three groups will be modified. For instance, if J_{unc} is set to be an empty set, then the assumptions will be precisely the same as that of exclusion restrictions in the normal instrumental variables regression with J_{exc} instrumental variables. This is exactly what will be done in the later section when comparing the SNIV results, which are robust to weak instruments, with those of other IVs papers, which requires strong instruments to be valid. In reality, however, the researcher might want to run a range of value of s and choose the appropriate level accordingly.

Besides relaxing exclusion restrictions assumption, since this method using regularization of high-dimensional statistics, it is entirely possible for a small number of observations and a large number of independent variables K. This is not achievable in conventional IVs estimates as the regressions falling apart or just performing poorly when K progressively getting much larger, especially when the number of regressors exceeds N. This is because when $K > N$, some of the component matrices needed in the estimates are not invertible, and therefore the formula used is invalid. Particularly, to see why this is the case, let us consider the typical two-stage least square estimates: $\boldsymbol{\beta}_{2SLS} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}$. When $k > n$, $rank(\hat{\mathbf{X}}'\hat{\mathbf{X}}) = rank(\hat{\mathbf{X}}) = min(n, k) = n < k$ and therefore the matrix is singular and cannot be inverted.

Since SNIV can only give the estimates for confidence sets, not point estimates or confidence interval, there should be a distinction between confidence intervals and confidence sets. Usually, the interpretation for the confidence interval of a specific variable is for all of the confidence intervals from all possible samples, 95% of them contain the true value of the population parameter: $\lim_{n\to\infty} Pr[\hat{\boldsymbol{\beta}} - t_{1-\alpha/2}se(\hat{\boldsymbol{\beta}}) \leq \boldsymbol{\beta} \leq \hat{\boldsymbol{\beta}} - t_{\alpha/2}se(\hat{\boldsymbol{\beta}})] = 1 - \alpha$. The confidence set can be seen as a joint probability of confidence intervals of all independent variables in the model, and therefore will be larger for the same level of significance. Thus, it is justified to compare confidence sets between different methods rather than confidence set from one method with confidence intervals from others. The following section will provide the formula for the procedures for the confidence sets of both SNIV and 2SLS.

4.3 Procedure: SNIV estimator

The method is called SNIV, which is the abbreviation for self-normalized instrumental variable. For SNIV confidence set estimates, we start with the mentioned model:

$$
\begin{aligned}
y_i &= \mathbf{x}_i^T \boldsymbol{\beta}_{\mathbf{x}} + \mathbf{z}_i^T \boldsymbol{\beta}_{\mathbf{z}} + \boldsymbol{\epsilon} \\
&= \mathbf{w}_i^T \boldsymbol{\beta} + \boldsymbol{\epsilon}
\end{aligned}
$$

with $\mathbb{E}[\mathbf{z}_i \boldsymbol{\epsilon}_i] = 0$, $|\boldsymbol{\beta}_{J_{unc}}|_0 \leq s$, and $\boldsymbol{\beta}_{J_{exc}} = 0$.

First of all, we start by defining the sparse identified set as:

$$
\mathcal{I}_s = \left\{ \boldsymbol{\beta} \in \mathbb{R}^K : \mathbb{E}[\mathbf{z}_i (y_i - \mathbf{w}_i^T \boldsymbol{\beta})] = 0_L, |\boldsymbol{\beta}_{J_{unc}}|_0 \leq s, \boldsymbol{\beta}_{J_{exc}} = 0 \right\}
$$

where the last two elements are defined by definition of sparsity and exclusion restrictions while the first element comes from the condition $\mathbb{E}[\mathbf{z}_i \boldsymbol{\epsilon}_i] = 0$. Depending on the value of s, this set could have one or many elements ($\boldsymbol{\beta}$). In general, the set is not a singleton (i.e. there might be more than one $\boldsymbol{\beta}$ in the set) and therefore we usually have partial identification, not point identification.

From here, to construct a confidence set, the concept of self-normalized sums is needed. Firstly, the confidence set can be denoted as follow:

$$
\widehat{C}_n(s) = \left\{ \boldsymbol{\beta} \in \mathbb{R}^K : \max_{l=1,\ldots,L} \left| \frac{1}{n} \frac{\mathbf{z}_l^T (\mathbf{y} - \mathbf{W}\boldsymbol{\beta})}{\sqrt{\hat{Q}_l(\boldsymbol{\beta})}} \right| \leq r_0, |\boldsymbol{\beta}_{J_{unc}}|_0 \leq s, \boldsymbol{\beta}_{J_{exc}} = 0 \right\}
$$

with $\hat{Q}_l(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_{li}^2 \boldsymbol{\epsilon}_i^2$

Most of the components are straight forward, and the only new component is the maximization constraint: $\max_{l=1,\dots,L}\left|\frac{1}{n}\frac{\mathbf{z}_l^T(\mathbf{y}-\mathbf{W}\boldsymbol{\beta})}{\sqrt{\hat{Q}_l(\boldsymbol{\beta})}}\right| \leq r_0$.

Originally, this comes from the moment condition $\mathbb{E}[\mathbf{z}_i\boldsymbol{\epsilon}_i] = 0$ or:

$$\frac{1}{n}\mathbf{Z}^T(\mathbf{y}-\mathbf{W}\boldsymbol{\beta}) = 0$$

However, with the sparsity setting, we can concentrate our attention toward a subset of the variables, for instance, those satisfy the event:

$$G_0 = \left\{\max_{l=1,\dots,L}\left|\frac{1}{n}\frac{\mathbf{z}_l^T(\mathbf{y}-\mathbf{W}\boldsymbol{\beta})}{\sqrt{\hat{Q}_l(\boldsymbol{\beta})}}\right| \leq r_0\right\}$$

which is exactly the condition mentioned before. The left hand side of the inequality is also known as self normalized sum with the choice of r_0 comes from the literature on "Moderate deviations of self-normalized sums". Technically, r_0 should be picked so that the probability of the event G_0 being true is $1-\alpha$ asymptotically, and therefore $\boldsymbol{\beta}$ belongs in the set $\widehat{C}_n(s)$ with the probability at least $1-\alpha$:

$$\lim_{n\to\infty}\inf_{\boldsymbol{\beta}\in I_s} Pr\left(\boldsymbol{\beta}\in\widehat{C}_n(s)\right) \geq 1-\alpha$$

The details of which scenario operates with which value of r_0 can be accessed in Gautier et al. (2018). In general, there are five scenario in total, and since the fourth scenario condition is satisfied most of the time, we will be using the results below (which is exactly quoted from Gautier et al. (2018)):

"**Scenario 4:** $(z_i u_i(\beta))_{i=1}^n$ are independent, $\left|\left(\left(\mathbb{E}\left[|Z_l U(\beta)|^{2+\delta}\right]\right)\left(\mathbb{E}[Z_l^2 U(\beta)^2]^{-(2+\delta)/2}\right)\right)_{l\in[L]}\right|_\infty \leq \gamma_{2+\delta}$ for $\delta\in(0,1]$ and $\gamma_{2+\delta}\geq 0$ and $L \leq \alpha/\left(2\Phi\left(-n^{1/2-1/(2+\delta)}\gamma_{2+\delta}^{-1/(2+\delta)}\right)\right)$

Under this scenario, the appropriate r_0 is:

$$r_0 = -\frac{1}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2L}\right)$$

"

Besides requiring the existence and bounds of some moments of $Z_l U(\beta)$, This scenario also assumes independence between observations, allows for conditional heteroscedasticity while does not rely on symmetry. The result also does not impose any restrictions on the joint distribution of Z and W, meaning there is no specific condition on the strength of the instruments. Thanks to this, the method enables estimations with robustness to weak instruments unlike TSLS where strong first stage is required. Another side note is, smaller r_0 will yield tighter inference.

With such construction of the confidence set, for the results to be valid, the only assumptions required are those in the scenario four and sparsity. As a result, the method is robust to weak instruments (no assumption needed on the joint distribution of the instruments and the endogenous variables) and to partial identification of $\mathcal{I}_s$.

However, working directly on the maximization constraint is not straight forward, and therefore it needs to be transformed. Specifically, the condition in the set can be rewritten in quadratic form:

$$\begin{aligned}
&\max_{l=1,\ldots,L} \left| \frac{1}{n} \frac{\mathbf{z}_l^T \boldsymbol{\epsilon}}{\sqrt{\hat{Q}_l(\boldsymbol{\beta})}} \right| \leq r_0 \\
\iff &\max_{l=1,\ldots,L} (\mathbf{z}_l^T \boldsymbol{\epsilon})^2 \leq (n r_0)^2 \hat{Q}_l(\boldsymbol{\beta}) \\
\iff &\max_{l=1,\ldots,L} a_l + \mathbf{b}_l^T \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{Q}_l \boldsymbol{\beta} \leq 0
\end{aligned}$$

where

$$a_l = (1 - nr_0^2)\mathbf{y}^T\mathbf{z}_l\mathbf{z}_l^T\mathbf{y}$$

$$\mathbf{b}_l = -2(1 - nr_0^2)\mathbf{W}^T\mathbf{z}_l\mathbf{z}_l^T\mathbf{y}$$

$$\mathbf{Q}_l = (1 - nr_0^2)\mathbf{W}^T\mathbf{z}_l\mathbf{z}_l^T\mathbf{W}$$

And therefore the confidence set can be rewritten as:

$$\hat{C}_n(s) = \left(\boldsymbol{\beta} \in \mathbb{R}^K : \max_{l=1,\dots,L} a_l + \mathbf{b}_l^T\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{Q}_l\boldsymbol{\beta} \leq 0, |\boldsymbol{\beta}_{J_{unc}}|_0 \leq s, \boldsymbol{\beta}_{J_{exc}} = 0\right) \tag{1}$$

According to this formula, we will be obtaining the confidence set by solving the following two problems: (i) min β_p s.t. $\boldsymbol{\beta} \in \widehat{C}_n(s)$, (ii) max β_p s.t. $\boldsymbol{\beta} \in \widehat{C}_n(s)$ which are lower bound and upper bound for each coefficient at a given s sparsity (β_p is an element of $\boldsymbol{\beta}$).

When computing this in a program, we replace "$|\boldsymbol{\beta}_{J_{unc}}|_0 \leq s$" by "there exists $\mathbf{m} \in R^K$: $m_k(1 - m_k) = 0, (1 - m_k)\beta_k = 0$ for $k = 1 \dots K, \sum_{k \in J_{unc}} m_k \leq s$", which is non-convex, and together with the quadratic inequality constraint, these optimisations can be seen as a non-convex quadratic program, an NP-hard problem. In practice, this can get very computationally heavy. In this paper, the solution will be determined by solving a hierarchy of semidefinite programs, which will converge to the true optimal value as the hierarchy increases. To be more specific, Gloptipoly package is used, which is "intended to solve, or at least approximate, the Generalized Problem of Moments (GPM), an infinite-dimensional optimization problem which can be viewed as an extension of the classical problem of moments" (D. Henrion, J. B. Lasserre, J. Loefberg (2009)). The hierarchy mentioned above is the degree of linear matrix inequality relaxations of the GPM or a hierarchy of semidefinite programming (SDP), called "relmin" and "relmax" for minimum and maximum relaxations. The higher the degree of the hierarchy, the smaller the relaxation, and the more precise the results will be on the optimal value. Notice that these hierarchy can be set in advance with the trade-off between easier computation and a more precise result. For most of the time, we will be using relmax = relmin = 2. The estimates using this will be at least as wide as the true confidence set and therefore, is always valid. Nevertheless, the set could be conservative (too wide). The detailed underlying theory, description and usage of Gloptipoly can be found in D. Henrion, J. B. Lasserre, J. Loefberg (2009) and J. B. Lasserre (2008). Also, the details of the code can be found in the appendix A.2.

The end results are vectors of lower bounds and upper bounds $\widehat{C}_{n,lb}(s), \widehat{C}_{n,ub}(s)$ with the confidence set as desired:

$$\lim_{n\to\infty} Pr\left(\widehat{C}_{n,lb}(s) \leq \boldsymbol{\beta} \leq \widehat{C}_{n,ub}(s)\right) \geq 1-\alpha \tag{2}$$

We will be comparing this with IV two-stage least square estimates confidence set:

$$\lim_{n\to\infty} Pr\left(\hat{\boldsymbol{\beta}}_{2SLS} + t\hat{\Sigma}^{-\frac{1}{2}}\mathbf{1}_k \leq \boldsymbol{\beta} \leq \hat{\boldsymbol{\beta}}_{2SLS} - t\hat{\Sigma}^{-\frac{1}{2}}\mathbf{1}_k\right) = 1-\alpha \tag{3}$$

where $\hat{\boldsymbol{\beta}}_{2SLS} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\beta}, \Sigma)$

$\mathbf{X}_{inc}$ are k independent variables

$\mathbf{Z}_{inst}$ are the instrumental variables

$\hat{\Sigma} = \hat{s}^2(\mathbf{X}'_{inc}\mathbf{Z}_{inst}(\mathbf{Z}'_{inst}\mathbf{Z}_{inst})^{-1}\mathbf{Z}'_{inst}\mathbf{X}_{inc})^{-1}$

$\hat{s}^2 = \frac{1}{N-k}\sum_i(y_i - \mathbf{X}_{i,inc}\hat{\boldsymbol{\beta}}_{2SLS})^2$

$\mathbf{1}_k$ is a column vector of 1 with k rows

t is the critical value derived from $Pr(\gamma \leq t) = (\alpha/2)^{1/k}$, $\gamma \sim \mathcal{N}(0, I_k)$

α is the level of significance

Table 1: Comparison between SNIV, IVs (2SLS)

	SNIV	**2SLS**
Pros	-Only need r_0 and sparsity assumption -Robust to weak instruments -Allow for a small number of observations relative to the number of regressors and instruments (n <L <K) -Allowing for partial identification	-Confidence interval are possible -Very easy to compute -Typically the narrowest
Cons	-Confidence intervals are not possible -Computationally intensive -Need to choose the appropriate s	-Need exclusion restrictions on all available instruments -Is invalid in high-dimensional settings -Need strong instruments

4.4 Summary and comparison between SNIV, IVs (2SLS)

The table 1 briefly goes through the benefits and drawbacks for each models

5 Data implementation

The three following subsections will follow a very similar structures: data description, the models specifications and the results with relevant comments. Regarding the models, it will mostly contain the following settings: (i) the original 2SLS estimates; (ii) SNIV with the same assumptions as 2SLS concerning exclusion restrictions specifications while is robust to weak instruments (which is also referred as SNIV model (A1)); and (iii) SNIV relaxing exclusion restrictions and also robust to weak instruments (which is also referred as SNIV model (A2)). To be specific, if 2SLS supposes some variables are controls while others are instruments, SNIV model (A1) will specifically setting the indexes of the endogenous and control variables to J_{inc} while the rest will go to J_{exc}, leaving J_{unc} always empty. The SNIV model (A2), on the other hand, is all about relaxing the exclusion restrictions, i.e. J_{unc} will no longer be empty. For the final model type, there are numerous ways to assign variables into the three groups, and we have decided on two different set of assumptions: the first one hypothesises all of the independent variables are in J_{unc} while the second one assuming the same except only the endogenous variable is in J_{inc}. Since all of the results from these two specifications yield similar results, the paper only reports the results from the former, with different degree of sparsity. Due to the structures of the assumptions made in SNIV (A2), SNIV (A1) assumptions are stronger than those in SNIV (A2) as the later does not specified the exact location of the exclusion restrictions given the same sparsity level.

For 2SLS, the paper shows the lower bounds and upper bounds as well as the value of the point estimates and their equivalent first-stage F statistics or Cragg-Donald statistics whenever there are more than two endogenous variables. For SNIV, the results are reported in the form of confidence sets accompanied by the estimates status. These statuses are put inside the bracket below the bounds, indicating whether or not the program is able to solve exactly (status = 1), is solved but the bounds given are bigger than the true bounds of the confidence sets (status = 0) or is not able to find feasible solution (status = -1).

Another notice coming from computational problems is the sparsity level should not be set too low as the results will have infeasible solutions (because the sparse identified set will be empty, meaning the results cannot be trusted) or too high as the confidence sets will be two large. Additionally, this paper will not implement all possible models for each dataset, only a few are used.

5.1 Data implementation: Angrist, Graddy and Imbens (2000) case

Brief data description: The article was concerned with the relationship between price and quantity of fish in Fulton fish market, which involved simultaneous equations of supply and demand function where endogeneity was a high possibility if only each equations were to be estimated separately. One of the main workings of the paper was to apply instrumental variable regressions to solve this problem. Specifically, they focused on the demand function where the dependent variable is the log quantity of whiting (pounds per day) sold at the Fulton fish market, representing demand of fish, while the variable of interest was the log average daily price of the fish (dollars per pound). Other dummies controlled variables, namely indicators for days of the week, cold and rainy on shore, were also included in all models. The proposed instrument was the weather conditions in the sea, which could be categorised into three collectively exhausted indicators depending on the measurement of wave height and wind speed: a day is classified as "stormy" if the wave is more than 4.5 feet in height and the wind speed is at more than 18 knots, "fair" if the wave height is less than 3.8 while the wind speed is less than 13 knots, "mixed" if the computed figured is in between. The dataset had 111 observations as in 111 days at the market. The summary statistics of the dependent variable and the variable of interest are provided in table 2.

Table 2: Fulton fish dataset: summary statistics (n=111)

Variable	**Mean**	**S.d.**	**Min**	**Max**
log (average daily price)	-0.1937	0.3819	-1.1077	0.6643
log (quantity)	8.5234	0.7417	6.1944	9.9814

For the instrumental variables estimator to work, the weather conditions in the sea should be an excluded from the demand function while it is included in the supply function. The author argued that supply of whiting was plausibly affected by the weather at the sea but the same condition was illogical for the demand of fish since the consumers decision to buy fish should not be affected by the ocean weather. This is the exclusion restriction conditions of the model.The instruments are also claimed to be strong, and with the rule of thumb, this implies the first stage F statistic should be greater than 10.

It should be noticed that this is the smallest dataset among three chosen articles. And in a small sample size, it might be hard to predict and explain the estimates since most of the theory based on asymptotic properties.

The models: First of all, for 2SLS, there are four different models implemented. The model (1) uses the whole dataset with the only instrument as stormy; second model uses only non-stormy days with fair or mixed as instruments; third model takes only non-fair days and has mixed or stormy as instruments; finally, the fourth specification is where all categories are included (two among three instruments to avoid dummies variable traps). Secondly, SNIV model (A1) assuming the same exclusion restrictions, for instance, the first model will list log price, dummies on weekday, cold and rain on shore as included regressors (J_{inc}) while stormy is the only excluded restriction. Since there is no elements in J_{unc}, sparsity s is equal to 0 since there is no coefficients in the uncertain set in the first place. SNIV (A2), on the other hand, specifies J_{unc} as all of the regressors including the instruments. This means there is a possibility that the "assumed instruments" might not be excluded from the model. In other words, there is a chance weather at the sea might be affecting the demand of the consumers. One possible explanation for this could be caused by the stormy weather patterns coinciding with the eating habits of the consumers, which in turn affecting the demand of whiting. As a result, SNIV (A2) has weaker assumptions relative to SNIV (A1).

The results are in the table 3. Firstly, we would expect the estimates to be negative, as implied by the law of supply and demand: for demand function, as the price of a normal good increases, the demand for that good will decrease as the buyers have less incentive to consume more of that good. The original 2SLS concludes the same things for the first and fourth models, except for the model 2 and 3 where the low observations might be the cause of the insignificant coefficients. SNIV (A1) reaches a similar conclusion that the effect of price on quantity of fish sold is negative in first and last models since the confidence sets are completely contained in the negative region while the inference is non-conclusive for the rest. Furthermore, since the only difference in the assumptions made between SNIV (A1) and 2SLS is the strength of the instruments, the similarity is not surprising considering the F-statistics for the first stage of model 1 and 4 are sufficiently high. Also, noticing that the status of SNIV (A1) are all 1, the numbers are solved exactly and are not too wide. This is not always the case, as can be seen for the SNIV (A2). Although some of the solutions' statuses in the third column are not one, the final conclusion can still be retained. To illustrate this point, the lower bound of SNIV (A2) for model 4 having status zero indicates this figure is lower than than the true value. However, since the upper bound is solved exactly and is negative (-0.0001), regardless of the value of the lower bound, the whole confidence set is smaller than zero. The same could be applied to model 4. Despite having all status at

Table 3: Fulton fish dataset: 2SLS, SNIV confidence sets estimates for log price, with log quantity as the dependent variable.

n/intruments	**2SLS**	**SNIV (A1)**	**SNIV (A2, s = 1)**	**SNIV (A2, s = 2)**
(1)n=111; **stormy**	-2.2318, -1.2228, -0.2138 F-stat = 14.61	-2.8138, NA, -0.2339 (1 1)	-2.8138 , NA , 0.0000 (0 1)	-
(2)n=79; **fair (stormy = 0)**	-1.4841 ,-0.3090, 0.8662 F-stat = 10.083	-1.6529, NA ,1.6725 (1 1)	-1.6529, NA ,0.1901 (1 0)	-
(3)n=66; **mixed (fair = 0)**	-2.9559 ,-1.2626 ,0.4308 F-stat = 5.220	-4.7403 ,NA, 0.7178 (1 1)	-0.1559 ,NA ,0.7177 (0 1)	-
(4)n=111; **stormy & mixed**	-1.7254, -0.9470, -0.1685 F-stat = 12.082	-3.3102, NA ,-0.0623 (1 1)	-1.3134, NA, -0.0003 (0 0)	-1.3086 ,NA ,0.8104 (0 0)

* The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

* All model has log(quantity) as the dependent variable and log(price) as the main endogenous variable of interest. Different specification of instruments and subsamples are defined at the first column of the table. Controlled variables included are monday, tuesday, wednesday, thursday, cold and rain on shore.

non-exact value, this simply shows the true confidence set of the effect of price on quantity sold is narrower and included in the range from -1.3134 to -0.0003, which is entirely negative.

Regarding the magnitude of the estimates, 2SLS and SNIV (A1) are very similar, but most of the time the later is wider reflecting its nature of being robust to weak instruments. SNIV (A2) with different sparsity levels also have certain characteristics, particularly the higher the sparsity level, the tighter the confidence set will be and less leaning toward zero. Model 4 SNIV (A2) demonstrates this as the sparsity at 2 (from -1.3086 to 0.8104) gives a larger and insignificant confidence set compare to the one at sparsity 1 (more sparse) (from -1.3134 to -0.0003, which does not center at zero like the former). This is reasonable as higher sparsity imposing tighter restriction and stronger assumptions, which results in a tighter confidence set. However, when the status is zero, which means the solution is not exact, higher sparsity degree will not always end up being tighter. It will become more lucid in the last datasets.

Overall, the conclusion stays the same for all of the model: the price has negative impact on the quantity of whiting sold in the Fulton fish market. And since all of the SNIV results are similar to those of 2SLS, the assumptions on exclusion restrictions of 2SLS is difficult to reject.

5.2 Data implementation: Card (1995) case

Brief data description: The article tried to estimate the effect of education on earnings, which has been researched in many papers. The main problem for this setting is the existence of many omitted variables which correlated with both wage and years of schooling, causing education to be endogenous and the conventional methods will result in an over or under-statement of the coefficients estimates. For example, elements such as family background, including financial status, parental education level, or the environment when growing up, can all influenced the decision to take more or less education. Or Griliches (1977) also suggested "ability bias", where the higher this value is, the higher the earnings will be and the more likely that person will take further education since they have the ability to do so. If any of these variable is also a direct determinant of earnings and not accounted for in the model, they could inflate/ deflate the effect of education with their own effects. Table 4 is the summary statistics for log of wage in 1976 (the dependent variable) and the years of education in 1976 (variable of interest).

Table 4: Card dataset: summary statistics

Variable	**Mean**	**S.d.**	**Min**	**Max**	**n**
log (wage in 1976)	6.2618	0.4438	4.6052	7.7849	3010
years of education in 1976	13.2253	2.7497	0.0000	18.000	3613

The author suggested using instrumental variables estimator with college proximity as an instrument, assuming this variable is not directly correlated with wage, i.e. excluded from the model. He argued that people growing up without a near access to college will have to pay more, at least by the living cost of renting instead of staying at home, compare to the people who live near college. This gap impeded the intention for further education. Furthermore, the exclusion requirement can be tested, according to the author, by regressing the income on the college distance and the education level using interaction terms of being near college and parental educational background as instruments (for more details, see Card (1995)). The conclusion was the insignificant direct effect of the instruments, which satisfied excluded restrictions. Additionally, the added controlled variables consisted of dummies for black and various of regional effects in both 1976 and 1966, specifically indicators for living in southern, SMSA and nine other regions. The paper also controls for parental educational backgrounds (14 dummies variables in total).

It is worth noticing that besides years of schooling, the article specified at most three other endogenous variables (that is experience, experience square and knowledge of the world of work), accompanied by at least three more instruments, namely age, age square and IQ score. However, since the paper only reported the coefficients for education, discussion will be restricted just on the effect of education.

The models: we start with the basic model (model 1), and the further specifications from 2 to 7 can be seen in the tables. The first specification (basic model) has three endogenous variables: education (ed76), experience (exp76) and square of experience ($exp76^2$). Being near 4-year college (near4c), age and square of age are the instruments, respectively. Other controls include indicator for black, living in southern area and living in SMSA in 1976 and in 1966, eight dummies for nine regions in 1966, and 14 indicators for parental missing education, mother's and father's education interactions and family structures at the age of 14. Model 2 is the most simple, making experience become an included exogenous variable, leaving education as the only endogenous regressor. SNIV (A1) has the same presumptions, which is reported in the second column.

SNIV (A2), again, puts all regressors in J_{unc} and sets different level of sparsity. Even the article itself have three main points why college proximity can be included in the model. For instance, area with the existence of college institutions are likely to have higher average salary and higher elementary and secondary schools quality compare to other area, which consequently leads to higher wages. Another possibility is families, with strong emphasis on education, tend to have children with higher abilities, which will obviously affecting their future earnings. The results are only reported for model 2 as this is the only model that gives feasible solutions.

The results for 2SLS and SNIV (A1) are shown in table 5 and table 6 gives SNIV (A2) results just for model 2 with three sparsity level. For the original (2SLS) estimates, the effect of education are sometimes significant (i.e. the confidence set does not include zero)

Table 5: Card 1995 dataset: 2SLS, SNIV confidence sets estimates for education in 1976, with log wage in 1976 as the dependent variable

Model	**2SLS**	**SNIV (A1)**
(1) The basic model (specified as before). (n=3010)	0.0452, 0.1324, 0.2197 F_stat= 4.010	-0.5102, NA ,1.2683 (1 1)
(2) Assuming exp76, $exp76^2$ is not endogenous. (n=3010)	-0.0231 ,0.1404, 0.3040 F_stat= 14.936	-0.0020, NA ,0.3856 (1 1)
(3) Adding kww as control variables (n =2963)	-0.0276, 0.1359, 0.2993 F_stat= 2.467	-1.2897, NA, 1.5016 (1 1)
(4) Adding kww as an endogenous variable, iq as an instrument (n=2040)	-0.1079 ,0.0893, 0.2864 F_stat=1.246	-1.5433, NA ,1.7204 (1 1)
(5) The basic model with instrument for ed76 as nearc4a instead. (n=3010)	0.0928, 0.1937, 0.2945 F_stat= 3.844	-0.9703, NA ,1.4972 (1 1)
(6) The basic model with instrument for ed76 as nearc4 and nearc2 instead. (n=3010)	0.0333 ,0.1169, 0.2005 F_stat= 3.107	-0.4972, NA, 0.9473 (1 1)
(7) The basic model with restriction on only age 14-19 in 1966. (n = 2037)	-0.0302 ,0.0944 ,0.2191 F_stat= 2.029	-1.0011, NA, 1.2687 (1 1)

* The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

Table 6: Card 1995 dataset: SNIV with relaxed exclusion restrictions results for model 2

SNIV (A2)	**s=1**	**s=2**	**s=3**
Model (2)	-0.0291, NA, 0.0000 (0 0)	-0.1018, NA ,0.4662 (0 0)	-2.7984, NA ,3.5095 (0 0)

and positive in model 1, 5 and 6. Notice that in the articles, by comparing to the conventional OLS estimates, the IVs gave much higher coefficients, explicitly 20-60 % higher and therefore, he concluded that the conventional methods undermined the benefit of education on wages. However, there is little discussion on the significance level of these coefficients. Furthermore, the first stage/Cragg-Donald statistics is not that high (all much lower than 10, except model 2), signalling very weak instruments are used. Since for 2SLS to be valid, strong instruments are required and violation of these might lead to invalid inferences of all model, except the second one. This one, however, assumes experience to be exogenous, which is not plausible as there are many factors which could cause omitted variable bias here. For instance, people with higher abilities are more likely to take further educations while it is logical to assume that earnings are increasing with ability (See Uusitalo, R (1999) for more details). With weak instruments, the outcome of 2SLS might be even worse than their conventional methods OLS in terms of bias, even with large sample size (See in the literature review section). This problem is fixed in SNIV (A1), where all of the confidence sets are from negative to positive region, implying inconclusive impact coming from education. In general, to guard against weak instruments, SNIV tends to expand the confidence sets, and in some cases where the instruments are too weak, the confidence set can be unbounded (Dufour, Jean-Marie (1997)). The good thing about this is there is no need for evaluating the strength of the instruments with some random number (rule of thumb uses 10) and the end results are valid regardless.

SNIV (A2) does not give much information. Too much sparsity might not make sense, as in model 2 there should be at least 2 variables that is included such as education and experience. Therefore, s = 1 suggesting only at most one included regressor should not be trusted. Furthermore, there is the problem with endogenous experience, only weaken the validity of the estimates. On the other hand, the two later sparsity degree give non-conclusive and progressively larger confidence sets, as expected. Therefore, although the estimates for SNIV

(A2) of the second model is available, the results do not demonstrate concrete evidences for the influence of education on wages.

To sum up, although there is several significant and relatively larger positive effects coming from education in 2SLS models, SNIV indicates the results should not be concluded as such since weak instruments are present, and might annul the legitimacy of the conclusion.

5.3 Data implementation: Angrist and Evan (1998) case

Brief data description: the article was interested in the effect of number of children on the labour supply of the mothers as well as whole families, including the fathers. Intuitively, having more children might deter the mother from working as she needed to spend more time taking care of the children. On the other hand, it was also reasonable to assume that having more children increased the financial burden of the family, leading to the mother having to work more to compensate for the extra expense. The authors try to investigate the matters using different dependent variables, which are composed of mother's labour supply, family income and father's labour supply. The main variable of interest is a dummy for having more than 2 children, which is potentially endogenous due to simultaneous causality with the dependent variables (see Goldin, 1990 p.125).

The authors proposed applying the instrumental variables estimator with indicators for same-sex children or two boys/ two girls as instruments. The idea was having two first children as either two boys or two girls encourage the family to have an extra child (see Westoff, Potter, and Sagi (1963), Iacovou, Maria, 2001), implies a positive correlation between the number of children and having same-sex of the first two children. Moreover, since the parents cannot decide the gender of their own children, the instruments can be seen as being predetermined and therefore should not be directly correlated with the mothers' labour supplies, satisfying exclusion restrictions condition. Other controlled variables included were the mother's age, age at first birth, dummies for races of the mother and whether or not the first born or the second born was a boy. The same characteristics for the fathers were used if the dependent variables represented fathers' labour supply.

There are a total of four different data sets: PUMS 1980s full sample, PUMS 1980s married sample, PUMS 1990s full sample, and PUMS 1990s married sample. Full sample uses all observation while married sample only contains those who are married. Due to the similarity between samples in 1980s and 1990s, the main discussion will focus only on samples in PUMS 1980s. The results for PUMS 1990s can be seen in the appendix (A.1)

The models: There are a total of ten different dependent variables used (6 proxies for the mothers' labour supply and 4 measurements for the counterpart ones), namely indicator for worked for pay of the mother / father (workedm/workedd), mother / father's weeks worked (weeksm1 / weeksd1), hours worked per week (hourswm / hourswd), labour's income (incomem / incomed), log family income (famincl), and log non-mom income (nonmomil). The only endogenous variable is an indicator for having more than 2 kids (morekids) with samesex or boys2, girls2 as instruments. The included exogenous variables are mother / father's age (agem1/aged1), age at first birth (agefstm / agefstd), races(blackm, hispm, othracem / blackd, hispd, othraced), boy at first birth (boy1st) and boy at second birth (boy2nd) if the instrument is samesex, otherwise excluded boy2nd. The specifications of each model can also be fully seen in the resulted tables and the summary statistics of the dependent variables and the endogenous variable is provided in table 7.

Similarly, 2SLS and SNIV (A1) has identical exclusion restrictions settings. SNIV (A2) estimates will assume all regressors (including morekids, current age and age at first birth of the mother, her dummies for races, indicator for boy first, boy second, and either samesex

Table 7: Card dataset: summary statistics for PUMS 1980s

Variable	**All women**	**Wives**	**Husbands**
workedm / workedd	0.5655	0.5282	0.9769
(=1 if worked for pay in year prior to census)	(0.4957)	(0.4992)	(0.1502)
weeksm1 / weeksd1	20.8342	19.0184	47.9571
(Weeks worked in year prior to census)	(22.2860)	(21.8674)	(10.4902)
hourswm / hourswd	18.7977	16.6985	43.4904
(Average hours worked per week)	(18.9157)	(18.3356)	(12.2865)
incomem / incomed	7,160.8	6,250	38,919
(labour earnings in year prior to census, in 1995 dollars)	(10,804)	(10,211)	(25,014)
famincl	10.3178	10.5589	
(family income in year prior to census, in 1995 dollars)	(1.3503)	(1.0436)	
nonmomil		10.3647	
(family income - Mom's labour income, in 1995 dollars)		(1.2450)	
morekids	0.4021	0.3806	0.3806
(=1 if mother had more than 2 children)	(0.4903)	(0.4855)	(0.4855)
Number of observation (n)	394840	254652	254652

or two boys, two girls) to be in J_{unc} (uncertain) and then giving different sparsity level. The reason we might want this is if the instruments, samesex and boys2/girls2, are in fact not excluded. For instance, it could be the case that having mixed gender children requires less/more attentions from their mothers, which consequently forces the mother to be at home less/more. Intuitively, family with mixed sex children will have to put in more efforts in raising the children as boys and girls are fundamentally different, which requires more knowledge compared to those with only boys or girls. There has been some evidence on how labour supply was affected by the gender of the child in the family (Pabilonia, S.W., & Ward-Batts, J. (2010) and Sun, Ang and Zhang, Chuanchuan and Hu, Xiangting (2016)) which could support this point of view.

The results for 2SLS and SNIV (A1), which are the first and second columns, are all extremely similar, if not almost identical for some of the models. This is plausible as the first stage strength of the instruments statistics are undeniably large (F-stat is from 700 to 1700). For both full and married sample, the mother's labour supply are negatively affected by the number of children in the family, demonstrated in the models using workedm, weeksm1, hourswm and incomem. However, when considering family income, non-mom income or

Table 8: PUMS 1980s full sample: 2SLS, SNIV confidence sets estimates for morekids, with various dependent variables for labour supply (n = 394840)

Dependent variable	**2SLS**	**SNIV (A1)**	**SNIV (A2) s=2**	**SNIV (A2) s=3**	**SNIV (A2) s=4**
	instruments: samesex				
workedm	-0.1782, -0.1173, -0.0564	-0.1780, -0.0561 (1 1)	-0.1770, 0.8919 (0 0)	-0.5626, 1.0573 (0 0)	-
weeksm1	-8.2631, -5.5588, -2.8545	-8.2500, -2.8433 (1 1)	0.0002, 0.0961 (1 0)	-1.5262, 1.4755 (0 0)	-
hourswm	-6.8539, -4.5468, -2.2397	-6.8460, -2.2329 (1 1)	0.0005, 1.8914 (1 0)	-1.5545, 2.1931 (0 0)	-
incomem	-3224.9, -1903.0 ,-581.0	-3218.2, -574.6 (1 1)	10.7, 4769.2 (0 0)	-1754.2, 2414.6 (0 0)	-
famincl	-0.1905, -0.0253, 0.1399 F_stat= 1674.223	-0.1905, 0.1405 (1 1)	-1.3989, 1.7252 (0 0)	-2.9485, 2.2418 (0 0)	-

Dependent variable	2SLS	SNIV (A1)	SNIV (A2) s=2	SNIV (A2) s=3	SNIV (A2) s=4
	instruments: boys2, girls2				
workedm	-0.1707, -0.1102, -0.0498	-0.1327, -0.0830 (1 1)	-0.2505, 1.8628 (0 0)	-1.7043, 2.5492 (0 0)	-1.5357, 1.7005 (0 0)
weeksm1	-7.9496, -5.2633, -2.5770	-6.8780, -3.8496 (1 1)	0.0113, 0.1973 (0 0)	-1.3937, 1.6394 (0 0)	-2.4337, 1.4451 (0 0)
hourswm	-6.6008, -4.3094 ,-2.0180	-5.7780, -2.9404 (1 1)	0.0095, 0.6623 (0 0)	-2.5855, 1.2842 (0 0)	-3.7585, 1.0980 (0 0)
incomem	-3126.2, -1813.3, -500.4	-2954.5, -528.1 (1 1)	3.0, 4735.9 (0 0)	-1674.1, 2465.2 (0 0)	-1975.0, 2543.1 (0 0)
famincl	-0.1977, -0.0337 ,0.1303 F_stat= 849.249	-0.2215, 0.1290 (1 1)	-1.5533, 1.9969 (0 0)	-2.3144, 2.3354 (0 0)	-1.4712, 2.3585 (0 0)

* The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

Table 9: PUMS 1980s married sample: 2SLS, SNIV confidence sets estimates for morekids, with various dependent variables for labour supply (n = 254652)

Dependent variable	**2SLS**	**SNIV (A1)**	**SNIV (A2) s=2**	**SNIV (A2) s=3**	**SNIV (A2) s=4**
	instruments: samesex				
workedm	-0.1848, -0.1170, -0.0492	-0.1849 ,-0.0487 (1 1)	-0.8758 ,1.8485 (0 0)	-1.4471, 2.5517 (0 0)	-
weeksm1	-8.2149, -5.2720, -2.3292	-8.2085 ,-2.3121 (1 1)	0.0066, 0.1233 (0 0)	-1.4721, 2.5956 (0 0)	-
hourswm	-7.2551, -4.7836, -2.3120	-7.2550, -2.3020 (1 1)	0.0049 ,0.2937 (0 0)	-2.1079, 0.5358 (0 0)	-
incomem	-2660.9 ,-1274.3 ,112.3	-2655.6 ,122.8 (1 1)	17.3, 2268.2 (0 0)	-1353.6 ,2210.6 (0 0)	-
famincl	-0.1870, -0.0443 ,0.0985	-0.1873 ,0.0996 (1 1)	-1.0605, 1.7238 (0 0)	-2.0272 ,2.2240 (0 0)	-
nonmomil	-0.1373 ,0.0335 ,0.2043 F_stat= 1385.684	-0.1375 ,0.2054 (1 1)	-1.1517, 1.3492 (0 0)	-1.6350 ,2.0153 (0 0)	-

Dependent variable	2SLS	SNIV (A1)	SNIV (A2) s=2	SNIV (A2) s=3	SNIV (A2) s=4
	instruments: boys2, girls2				
workedm	-0.1763, -0.1089, -0.0416	-0.1464 ,-0.0800 (1 1)	-0.4970,1.2386 (0 0)	-1.6532 ,2.1879 (0 0)	-1.9294 ,2.2648 (0 0)
weeksm1	-7.9560, -5.0336 ,-2.1112	-7.3473, -2.5647 (1 1)	0.0002, 0.2743 (1 0)	-7.7467, 5.6557 (0 0)	-8.8158 ,6.8721 (0 0)
hourswm	-7.0051 ,-4.5510 ,-2.0969	-6.2901, -2.8284 (1 1)	-0.0001 ,1.7626 (1 0)	-8.1019, 6.2874 (0 0)	-7.0310 ,7.6678 (0 0)
incomem	-2634.3, -1257.7 ,118.9	-2953.5, 674.3 (1 1)	0.7, 3506.9 (0 0)	-1663.1, 3668.5 (0 0)	-2325.4 ,2595.0 (0 0)
famincl	-0.1899, -0.0482, 0.0936	-0.2452, 0.1169 (1 1)	-1.2694 ,1.5899 (0 0)	-2.3214, 2.2933 (0 0)	-1.4687, 3.2616 (0 0)
nonmomil	-0.1467, 0.0228 ,0.1923 F_stat= 703.044	-0.1494 ,0.1808 (1 1)	-1.1492, 1.3759 (0 0)	-2.1799 ,2.4185 (0 0)	-1.9145, 1.6676 (0 0)

[*] The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

Table 10: PUMS 1980s married sample using husband characteristics: 2SLS, SNIV confidence sets estimates for morekids, with various dependent variables for labour supply (n = 254652)

Dependent variable	**2SLS**	**SNIV (A1)**	**SNIV (A2) s=2**	**SNIV (A2) s=3**	**SNIV (A2) s=4**
	instruments: samesex				
workedd	-0.0166 ,0.0040, 0.0246	-0.0011, 0.0092 (1 1)	-0.7661, 0.3872 (0 0)	-2.2215 ,2.3348 (0 0)	-
weeksd1	-0.8606, 0.5793 ,2.0192	-0.8810 ,2.0525 (1 1)	-2.1879, 0.6006 (0 0)	-2.5710 ,2.1440 (0 0)	-
hourswd	-1.1736, 0.5094, 2.1924	-1.2006 ,2.2227 (1 1)	-2.7364 ,1.5551 (0 0)	-2.8412 ,1.5701 (0 0)	-
incomed	-4743.1, -1362.1, 2018.9 F_stat= 1380.911	-4797.2, 2076.8 (1 1)	-4196.3 ,1098.3 (0 0)	-2116.3 ,4404.9 (0 0)	-

Dependent variable	2SLS	SNIV (A1)	SNIV (A2) s=2	SNIV (A2) s=3	SNIV (A2) s=4
	instruments: boys2, girls2				
workedd	-0.0192 ,0.0013, 0.0217	-0.00091, -0.00041 (1 1)	-1.2179, 0.4180 (0 0)	-2.7920, 2.1979 (0 0)	-1.8734, 1.8479 (0 0)
weeksd1	-1.0208, 0.4075 ,1.8359	-0.2603, 1.2456 (1 1)	-2.4207 ,0.6610 (0 0)	-2.9938, 1.6763 (0 0)	-2.6131, 2.1002 (0 0)
hourswd	-1.2366, 0.4336, 2.1039	-1.6643 ,2.1956 (1 1)	-2.9819 ,1.6221 (0 0)	-2.3908 ,1.9678 (0 0)	-2.1990 ,2.0669 (0 0)
incomed	-4821.8 ,-1466.3 ,1889.1 F_stat= 701.010	-6204.4 ,2363.0 (1 1)	-2333.8, 1484.0 (0 0)	-4071.5 ,3476.1 (0 0)	-3496.6 ,2674.4 (0 0)

* The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

father's labour supply, the results are inconclusive as the confidence sets spanning from negative to positive region. Generally, the conclusions are the same for 2SLS and SNIV (A1) since the only difference in assumptions made are satisfied as the models have very strong instruments. Additionally, all solution for SNIV (A1) are 1 indicating the magnitudes of the estimates are the tightest as it should be, and hence concluding 2SLS and SNIV (A1) are similar is entirely valid.

On the other spectrum, the results for SNIV (A2) relaxing exclusion restrictions, which are from the third onward column, generate much more interesting results regarding the statistics for the mothers'. Worked for pay model no longer has negative significant confidence set. Instead these sets contain zero for all sparsity levels. Other dependent variables such as weeksm1, hourswm and incomem have their model at the lowest sparsity s = 2 producing completely opposite of which is observed previously: the effect of having more children is positive on the labour supply of the mothers. (is there any paper with this kind of results, is this weird?). The rest of the confidence set on other sparsity levels are similar to those in workedm models - insignificant. This is reasonable as SNIV (A2) assumptions when s is greater or equal to 3 is weaker than those made in SNIV (A1).

This phenomenon raises some questions. First of all, why is worked for pay the only one without positive confidence sets? and secondly, why are these having the opposite sign as the original estimates? For the former question, notice that workedm is a binary variable, meaning having a linear model might not be appropriate, and therefore, the results for this dependent variable could be invalid. Another explanation is the attribute which this dependent variable describe mother's labour supply is fundamentally different from the other measures: having more children might deter the number of hour worked or salary but not necessary making the mother entirely quit the job. As a result, the conclusion for this model is different from the rest although these are all proxy for labour supply for the mother.

The second question could be explained by the assumptions of exclusion restrictions or simply due to SNIV computational problems. If it is the former one, and since SNIV (A2) has a more general assumption, the instruments chosen might not be excluded as proposed before, and should be used with cautions. For instance, there might be a possibility of the mother's age and mother's age at first birth are excluded from the models instead of samesex or the other two instruments. It could be the case that the sample only restricts itself to mothers aged 21 to 35, and therefore the supply of labour within this age bracket of women are the same, resulting in the variable being irrelevant to the models. Notice that even though ones can run the OLS or 2SLS for the coefficients of the mother's age and her age at first birth and have statistically significant coefficients accordingly (not reported in the papers, but could be run), the results could be biased and they might not actually be included regressors as indicated by the p-values or t statistics. Therefore such methods should not be used to confirm (just as references for) these suspicions, however, since the two methods yield different results, the underlying assumptions are likely to be violated for 2SLS. Moreover, having more child leading to an increase in mother's labour supply is not entirely impossible as there are articles with the same conclusions (See Maria Iacovou (2001) and Claudia Hupkau, Marion Leturcq(2016)). As mentioned before, this outcome could be due to the mother trying to work more to compensate for the expenses of the extra children.

On the contrary, another hypothesis could be due to assigning s to be equal to 2 being too small (too much sparsity), and the results are not sensible. In other words, among the regressors, the fact that there are at least three included regressors should be true. Intuitively, the three variables could be morekids; age of the mother, which could affect the workings schedule to adapt to the current situation of a certain age brackets; and even the instruments (same reasons specified in the models explanations). With this train of thought, the results from SNIV (A2) can only say more if higher sparsity gave more definite confidence set, which is not the case here as the status are all 0, meaning the set is not tight and could be either positive or negative. Subsequently, the standings of the 2SLS is still valid. However, there is no strict evidence for this case. Other models, which originally concluded as being unaffected by the number of children stay the same for all equivalent models for SNIV (A2).

Another side notes for the outcomes of SNIV (A2) might not become wider as the level of sparsity decreases. It is true that as s increases, the identified set will become larger and wider. However, this effects might not necessary transfer to the confidence sets except when the statuses are 0 or 1. Also notice that the statuses of the results here are 0 implying the solution is not exact, meaning the solutions given are not the tightest and therefore, even though the true bounds are larger for higher value of s, there is no telling the estimated confidence sets for this one with 0 status is larger since it could be relatively narrower than the higher sparsity ones but still be an inexact solutions.

6 Summary of the results

Although the discussions in three subsections of section 4 are different as the three datasets have different characteristics, there are a few similar patterns drawn from those results. First of all, 2SLS and SNIV with the same exclusion restrictions are very similar when strong instruments are used such as the first and the last datasets. This is important considering how distinct these two methods are when approaching endogeneity using instrumental variables. The results for all the datasets fortify the validity of SNIV under the assumptions of 2SLS being true in real-life datasets, and the method can be seen as a trust-worthy alternative/ substitution of 2SLS in future researches. Futhermore, when weak instruments are presented, the 2SLS will be biased and misleading while SNIV remains valid as it is robust to weak instruments, although the confidence set might produce very wide confidence sets to compensate for the lack of information. However, this is much better than having a misleading results produced by 2SLS. This can be seen the clearest for the second subsection (dataset) in section 4.

Secondly, when the exclusion restrictions are relaxed in these kind of model, things are more complicated, especially for these kind of model with only just identified amount of instruments. If it is indeed true that the exclusion restrictions are violated, the SNIV (A2) states that the exclusion restrictions are in fact controlled variables instead while the current controlled variables are actually excluded. Although this might be the circumstances sometimes, there is a possibility that the model will end up having no exclusion restrictions at all, and the end results will be unknown for the conventional 2SLS as it is now unidentified while SNIV confidence set should be very large and ranging from negative to positive numbers.

Thirdly, the higher the sparsity level, the more likely the identified set will be empty, producing infeasible solutions or just not making sense. On the contrary, too little sparsity will impose too little restraints, and the estimates will have uninformative values.

Finally, the biggest drawback for SNIV is the computational problem, especially when the exclusion restrictions assumption is relaxed. If the program is not able to solved exactly, it is really hard to make definite inferences. Angrist and Evan dataset illustrates this the most as some of the results from SNIV (A2), if solved exactly, can shred some light on the true direction of effect of having more children on the mother's labour supply, and can even show whether the exclusion restrictions in 2SLS are violated or not.

7 Extension

7.1 Computational method alternative

As can be clearly seen in the previous section, SNIV's biggest problem lies in the computational intensity. As mentioned in the methodology section, this paper uses Gloptipoly to solve the non-convex problem in SNIV confidence sets.

$$\hat{C}_n(s) = \left(\boldsymbol{\beta} \in \mathbb{R}^K : \max_{l=1,\ldots,L} a_l + \mathbf{b}_l^T \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{Q}_l \boldsymbol{\beta} \leq 0, |\boldsymbol{\beta}_{J_{unc}}|_0 \leq s, \boldsymbol{\beta}_{J_{exc}} = 0\right)$$

Under this method, it was proved that as the level of relaxation (relmax) increases, the solution will converge to the true value, and more importantly, this is a finite convergence, meaning the convergence rate is fast enough such that the level of relaxation required for the true value coinciding with our solution is not infinity. However, in many cases, it took a very long time to run the estimates even at "relmax" equal to only 3 or 4, and therefore, continuing to increase this value to get the closer-to-the-truth sets is out of the questions. This results in many confidence sets estimates providing conservative solutions, and thus, inconclusive inferences.

However, gloptipoly is not the only way to estimate the NP-hard problem in the equation above. Instead of having the standard SOS hierarchy of semi-definite programmings (SDP), another way, proposed by Weisser, Lasserre and Toh (2017), is to use sparse BSOS- "a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity". Intuitively, the method is less computational heavy in two main aspects: (i) while the former relies solely on SDP, the later is solved using a mixture of SDP and Linear programming (which is easier to solved than SDP), and (2) Sparse BSOS exploits cases in which not all constraints contain all variables. The first advantage comes from the sparse BSOS inheriting the advantages of BSOS itself, where the size of the semidefinite constraint is fixed and controlled by a certain parameter called k. Linear-program-hierarchy is obtained when k is equal to zero, leading to a simpler problem than a purely SDP problem. The second one implies the optimization is more simple with less complicated constraints. For instance, if one of the constraint is $f(x_1, x_2) = 1$ given that there are two variables, using sparsity on "there is at least of of the variable being equal to zero in this equality", the constraint will depend only on one variable such as $f(x_1) = 1$. Obviously, the later constraint will be more simple to solve compare to when both of the variables in in the restriction. Moreover, this sparsity has nothing to do with the sparsity assumption made by SNIV, and do not have to be chosen manually.

7.2 Some instruments are endogenous

Even though the basic SNIV considered in this paper has already relaxed some of the assumptions, it still assumes the instruments are exogenous. Gautier et. al (2018) extend SNIV to account for the possibility that some instruments are endogenous. Particularly, instead of having the following moment conditions:

$$\mathbb{E}[\mathbf{z}\boldsymbol{\epsilon}] = 0,$$

we assume

$$\mathbb{E}[\mathbf{z}\boldsymbol{\epsilon}] = \boldsymbol{\theta},$$

instead, where θ is sparse with the degree of $\tilde{s}$. This means among all of the instruments, there are at most $\tilde{s}$ endogenous variables. The procedures for this assumption application into the regressions can be replicated from those for the sparsity of the coefficients (the indexes of the endogenous, exogenous and unknown variables need to be provided). The modified confidence set will be:

$$\widehat{C}_n(s, \tilde{s}) = \left\{ (\boldsymbol{\beta}, \boldsymbol{\theta}) : \max_{l=1,\dots,L} \left| \frac{1}{n} \frac{\mathbf{z}_l^T(\mathbf{y} - \mathbf{W}\boldsymbol{\beta}) - n\theta_l}{\sqrt{\hat{Q}_l(\boldsymbol{\beta}, \tilde{s})}} \right| \le r_0, |\boldsymbol{\beta}_{J_{unc}}|_0 \le s, \boldsymbol{\beta}_{J_{exc}} = 0, |\boldsymbol{\theta}_{\widetilde{J_{unc}}}|_0 \le \tilde{s}, \boldsymbol{\theta}_{\widetilde{J_{exc}}} = 0 \right\}$$

with $\hat{Q}_l(\boldsymbol{\beta}, \tilde{s}) = \frac{1}{n}\sum_i (z_{li}\epsilon_{li} - \theta_l)^2$

The first element is adjusted to account for the non-zero vector $\boldsymbol{\theta}$, and the last two conditions added represents the sparsity assumptions on the number of endogenous instruments. Again, one can transform the first condition into a quadratic inequality constraint and use gloptipoly to solve the problem. This time, however, the researcher has to choose two different sparsity level for the coefficients or one can combine the two into one constraint ($|\boldsymbol{\beta}_{J_{unc}}|_0 + |\boldsymbol{\theta}_{\widetilde{J_{unc}}}|_0 \leq s$, with appropriate s value). As an example, this method can be used in the second model of the second datasets in section 5 (Card 1995 dataset). In such model, there are potentially endogenous controlled variables, namely experience and experience square, and with the specification of these variables being in the uncertain group (meaning they can be excluded or not from the model), $\tilde{s}$ can be specified as 2 to account for these potential endogeneity.

8 Conclusion

By applying the SNIV for dealing with endogeneity using instrumental variables in three different real-life datasets, the paper was able to make relative comparisons with the traditionally used 2SLS performances under different sparsity assumptions. The findings show, given the exact positions of the excluded instruments, SNIV performances are very good compared to those in 2SLS since they are robust to weak instruments and when the strong instruments are given, the confidence sets are similar. Considering that there is no dependence on the first stage for SNIV, it should be emphasized that computational intensity is a significant drawback of this method, as SNIV is technically different from the usual 2SLS where the later's analytical results can be derived and easily computed while the former depends on solving semi-definite optimisation problems. This means, under reasonable computation limitations and same assumptions, SNIV can be considered as a great alternative to 2SLS for future research projects. Furthermore, due to the model set up of SNIV, this method can be generalised in many aspects, such as the positions of exclusion restrictions or even the exogeneity of the instruments. When there is a difference in the 2SLS and SNIV methods, the underlying assumptions of exclusion restrictions might have been incorrect and the estimates should be used with cautions. There are other further extensions that has not been applied yet to real-life data, and can be considered and evaluated in the future.

9 Reference

References

[1] Eric Gautier, Alexandre Tsybakov, Christiern Rose (2018), High-dimensional instrumental variables regression and confidence sets, arXiv Cornell University, arXiv:1105.2454

[2] Frank, I., & Friedman, J. (1993). A Statistical View of Some Chemometrics Regression Tools. Technometrics, 35(2), 109-135. doi:10.23071269656

[3] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267-288. Retrieved from http://www.jstor.org/stable/2346178

[4] Mauricio Bucca, Daniela R. Urbina (2019),Lasso Regularization for Selection of Log-linear Models: An Application to Educational Assortative Mating, Sociological Methods & Research, SAGE Journals, doi10.1177/0049124119826154

[5] Yuri Fujino, Hiroshi Murata, Chihiro Mayama, Ryo Asaoka (2015), Applying "Lasso" Regression to Predict Future Visual Field Progression in Glaucoma Patients, Investigative Ophthalmology & Visual Science April 2015, Vol.56, 2334-2339. doi:10.1167/iovs.15-16445

[6] Chan-Lau, Jorge Antonio (2017), Variance Decomposition Networks: Potential Pitfalls and a Simple Solution (May 2017). IMF Working Paper No. 17/107. Available at SSRN: https://ssrn.com/abstract=3053190

[7] A. Belloni D. Chen V. Chernozhukov C. Hansen (2012), Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain, Econometrica, Volume 80, Issue 6. ttps://doi.org/10.3982/ECTA9626

[8] Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." Journal of Economic Perspectives, 28 (2): 29-50. doi: 10.1257/jep.28.2.29

[9] Cun-Hui Zhang, Stephanie S. Zhang (2013), Confidence intervals for low dimensional parameters in high dimensional linear models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), Volume 76, Issue 1. https://doi.org/10.1111/rssb.12026.

[10] van de Geer, Sara; Bühlmann, Peter; Ritov, Ya'acov; Dezeure, Ruben. On asymptotically optimal confidence regions and tests for high-dimensional models. Ann. Statist. 42 (2014), no. 3, 1166–1202. doi:10.1214/14-AOS1221. https://projecteuclid.org/euclid.aos/1403276911

[11] Adel Javanmard and Andrea Montanari (2014), Confidence Intervals and Hypothesis Testing for High-Dimensional Regression, Journal of Machine Learning Research, Volume 15, 2869-2909. http://jmlr.org/papers/v15/javanmard14a.html

[12] Joshua Angrist and William Evans, (1998), Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size, American Economic Review, 88, (3), 450-77.

[13] Angrist, J., Graddy, K., & Imbens, G. (2000). The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the

Demand for Fish. The Review of Economic Studies, 67(3), 499-527. Retrieved from http://www.jstor.org/stable/2566964

[14] David Card (1995), Using Geographic Variation in College Proximity to Estimate the Return to Schooling, National Bureau of Economic Research, Working Paper Series,number 4483. doi:10.3386/w4483

[15] Didier Henrion, Jean-Bernard Lasserre, Johan Lofberg. GloptiPoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software, Taylor & Francis, 2009, 24 (4-5), pp. 761-779. hal-00172442

[16] Lasserre, Jean B. (2008), A semidefinite programming approach to the generalized problem of moments, Mathematical Programming, Volume 112, no. 1, 65-92. issn:1436-4646, doi:10.1007/s10107-006-0085-1. https://doi.org/10.1007/s10107-006-0085-1

[17] Griliches, Zvi (1977), Estimating the Returns to Schooling: Some Econometric Problems, Econometrica, volume 45, no. 1, 1-22. https://EconPapers.repec.org/RePEc:ecm:emetrp:v:45:y:1977:i:1:p:1-22

[18] Bound, J., Jaeger, D., & Baker, R. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. Journal of the American Statistical Association, 90(430), 443-450. doi:10.2307/2291055

[19] Dufour, Jean-Marie (1997), Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, Econometrica, volume 65, no. 6, 1365-1388. https://EconPapers.repec.org/RePEc:ecm:emetrp:v:65:y:1997:i:6:p:1365-1388

[20] Goldin, Claudia Dale. Understanding the gender gap: an economic history of American women. New York: Oxford University Press. 1990. 287pp. ISBN: 978-0-19-505077-6

[21] Charles F. Westoff, Robert G. Potter, and Philip C. Sagi (1963) THE THIRD CHILD: A STUDY IN THE PREDICTION OF FERTILITY, Princeton University Press, 1963. 293 pp. Social Forces, Volume 43, Issue 1, Page 116. https://doi.org/10.1093/sf/43.1.116

[22] Iacovou, Maria (2001), Fertility and female labour supply, Institute for Social and Economic Research, ISER Working Paper Series, no. 2001-19. https://EconPapers.repec.org/RePEc:ese:iserwp:2001-19

[23] Uusitalo, R. (1999). Essays in economics of education. St. Louis: Federal Reserve Bank of St Louis. Retrieved from https://search.proquest.com/docview/1698309449?accountid=14723

[24] Pabilonia, S.W., & Ward-Batts, J. (2010). The Effect of Child Gender on Parents' Labor Supply: Responses among Natives, Immigrants, and Racial and Ethnic Subgroups.

[25] Sun, Ang and Zhang, Chuanchuan and Hu, Xiangting (2016), Boys, Girls, and Grandparents: The Impact of the Sex of Preschool-Aged Children on Family Living Arrangements and Maternal Labor Supply . Available at SSRN: https://ssrn.com/abstract=3253443 or http://dx.doi.org/10.2139/ssrn.3253443

[26] Iacovou, Maria (October 2001) 'Fertility and Female Labour Supply', Working Papers of the Institute for Social and Economic Research, paper 2001-19. Colchester: University of Essex.

[27] Claudia Hupkau, Marion Leturcq (2016), Fertility and Labor Supply: New Evidence from the UK, Collection : Documents de travail, no. 220.

[28] Angrist, J., & Krueger, A. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? The Quarterly Journal of Economics, 106(4), 979-1014. Retrieved from http://www.jstor.org/stable/2937954

[29] Staiger, Douglas and Stock, James H (1994), Instrumental Variables Regression with Weak Instruments, National Bureau of Economic Research, no. 151. Doi: 10.3386/t0151

[30] Weisser, T., Lasserre, J.B. & Toh (2018), KC. Math. Prog. Comp.10: 1. https://doi.org/10.1007/s12532-017-0121-6

10 Appendix

A.1. The 2SLS and SNIV estimates for 1990s PUMS sample (including summary statistics).

Table 11: Summary statistics for PUMS 1990s

Variable	**All women**	**Wives**	**Husbands**
workedm / workedd	0.6672	0.6701	0.9695
(=1 if worked for pay in year prior to census)	(0.4712)	(0.4702)	(0.1720)
weeksm1 / weeksd1	26.3875	26.5252	47.1569
(Weeks worked in year prior to census)	(22.8722)	(22.8206)	(11.9109)
hourswm / hourswd	22.6897	22.2617	44.2302
(Average hours worked per week)	(19.0836)	(18.8467)	(13.3508)
morekids	0.3705	0.3627	0.3627
(=1 if mother had more than 2 children)	(0.4829)	(0.4808)	(0.4808)
Number of observation (n)	380,007	301,595	301,595

Table 12: PUMS 1990s full sample: 2SLS, SNIV confidence sets estimates for morekids, with various dependent variables for labour supply (n = 380007)

Dependent variable	2SLS	SNIV (A1)	SNIV (A2) s=2	SNIV (A2) s=3	SNIV (A2) s=4
	instruments: samesex				
workedm	-0.1534 -0.0951 -0.0369	-0.1533 -0.0367	-0.0943 0.2251	-0.9369 1.8542	-
		(1 1)	(0 0)	(0 0)	
weeksm1	-8.2738 -5.4749 -2.6761	-8.2546 -2.6731	0.0035 3.2025	-6.1089 7.1058	-
		(1 1)	(0 0)	(0 0)	
hourswm	-6.4749 -4.1196 -1.7642	-6.4602 -1.7614	0.0003 0.9460	-3.5760 2.8262	-
	F_stat= 1721.328	(1 1)	(1 0)	(0 0)	
	instruments: boys2, girls2				
workedm	-0.1534 -0.0952 -0.0369	-0.1787 -0.0053	-1.7398 2.4372	-6.1007 6.4618	-7.7460 6.8818
		(1 1)	(0 0)	(0 0)	(0 0)
weeksm1	-8.2563 -5.4579 -2.6595	-8.6288 -2.2142	-0.0002 3.5834	-3.0898 6.2553	-3.4288 2.9263
		(1 1)	(0 0)	(0 0)	(0 0)
hourswm	-6.4856 -4.1307 -1.7759	-7.0029 -1.2689	0.0012 2.5955	-3.2561 3.5017	-2.3685 2.8015
	F_stat= 860.968	(1 1)	(0 0)	(0 0)	(0 0)

[*] The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

Table 13: PUMS 1990s married sample: 2SLS, SNIV confidence sets estimates for morekids, with various dependent variables for labour supply (n = 301595)

Dependent variable	**2SLS**	**SNIV (A1)**	**SNIV (A2) s=2**	**SNIV (A2) s=3**	**SNIV (A2) s=4**
	instruments: samesex				
workedm	-0.1579 -0.0996 -0.0413	-0.1577 -0.0412 (1 1)	-0.5861 1.4034 (0 0)	-3.0510 3.6542 (0 0)	-
weeksm1	-8.0337 -5.2269 -2.4201	-8.0127 -2.4201 (1 1)	0.0005 1.3393 (0 0)	-8.4404 6.9890 (0 0)	-
hourswm	-6.1990 -3.8713 -1.5436 F_stat= 1723.055	-6.1826 -1.5433 (1 1)	0.0006 0.6875 (0 0)	-5.5671 2.1994 (0 0)	-
	instruments: boys2, girls2				
workedm	-0.1579 -0.0996 -0.0413	-0.1941 -0.0035 (1 1)	-0.6628 1.1714 (0 0)	-2.4401 3.0067 (0 0)	-3.0228 2.4328 (0 0)
weeksm1	-8.0294 -5.2227 -2.4160	-8.7711 -1.6825 (1 1)	0.0006 2.6670 (0 0)	-3.2957 3.9611 (0 0)	-4.3655 5.7457 (0 0)
hourswm	-6.2006 -3.8730 -1.5454 F_stat= 861.572	-7.0383 -0.6390 (1 1)	0.0003 2.1347 (1 0)	-2.9740 4.1589 (0 0)	-3.7556 2.6200 (0 0)

[*] The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

Table 14: PUMS 1990s married sample with husband characteristics: 2SLS, SNIV confidence sets estimates for morekids, with various dependent variables for labour supply (n = 301595)

Dependent variable	**2SLS**	**SNIV (A1)**	**SNIV (A2) s=2**	**SNIV (A2) s=3**	**SNIV (A2) s=4**
	instruments: samesex				
workedd	-0.0120 0.0095 0.0310	0.0004 0.0188	-0.4681 0.5579	-3.6076 3.5249	
		(1 1)	(0 0)	(0 0)	
weeksd1	-0.6832 0.8001 2.2833	-0.6841 2.2968	-5.3624 0.0060	-4.6801 5.1635	
		(1 1)	(0 0)	(0 0)	
hourswd	-1.1862 0.4751 2.1365	-1.1926 2.1456	-5.4110 0.0272	-3.8942 5.1060	
	F_stat= 1718.927	(1 1)	(0 0)	(0 0)	
	instruments: boys2, girls2				
workedd	-0.0119 0.0095 0.0310	0.0003 0.0103	-0.4428 0.3529	-3.1823 3.5479	-2.8958 2.5754
		(1 1)	(0 0)	(0 0)	(0 0)
weeksd1	-0.6805 0.8027 2.2860	-0.9137 2.5110	-5.7229 0.0084	-4.4537 3.2509	-4.6083 2.8940
		(1 1)	(0 0)	(0 0)	(0 0)
hourswd	-1.1795 0.4818 2.1432	-0.7450 1.7337	-5.3897 0.0119	-3.9845 2.5179	-4.0428 2.6479
	F_stat= 859.490	(1 1)	(0 0)	(0 0)	(0 0)

* The results follow the format of "<lower bound> <point estimates> <upper bound>". For SNIV, the numbers under the estimates inside the brackets show the solution status of failed (-1), solved but not tight (0) or solved exactly (1).

A.2. The Matlab codes: an example for the Fulton fish dataset model 4 SNIV (A2).

The main code:

```
clear all
warning('off','all')

% MODEL
% y = w'*beta_w = x'*beta_x + z'*beta_z + u,
% E[zu] = 0,
% beta(Jexc)=0
% beta(Junc) IS SPARSE
rng(123)

realdata=1; %SET TO 1 TO USE REAL DATA, OTHERWISE SIMULATED DATA IS USED

% REAL DATA
if realdata==1
    dataFulton %THIS IS AN m FILE, THE DIRECTORY INSIDE WILL NEED TO BE CHANGED

    %4
    df = [qty, price, mon, tue, wed, thur, rainy, cold, stormy, mixed];
    df(any(isnan(df),2), :) = []; %delete nan rows

    y=df(:,1);
    n=length(y);
    x = df(:,2); %endogenous
    z = df(:,end-1:end); %continuous controls (none in this case) + instruments
    w = [x z];
```

```
    v = [ones(n,1) df(:,3:end-2)]; %constant + dummies controls
    l = size(z,2);
    k = size(w,2);
end

%F stat for first stage:
X1 = [ones(n,1) df(:,3:end-2)];
X2 = df(:,end-1:end);
Xall = [X1 X2];
Y = df(:,2);

M1 = eye(n) - X1*inv(X1'*X1)*X1';
MX = eye(n) - Xall*inv(Xall'*Xall)*Xall';

Fstat =
    (Y'*M1*X2*inv(X2'*M1*X2)*X2'*M1*Y/size(X2,2))*inv(Y'*MX*Y)*(n-size(Xall,2))


%% SOME CALCULATIONS for partial out dummies (including constant)

%partial out all possible dummies
ybix = v'*y;
ybix2 = v*inv(v'*v)*ybix;
y = y - ybix2;

wbix = v'*w;
wbix2 = v*inv(v'*v)*wbix;
w = w - wbix2;
```

```
zbix = v'*z;
zbix2 = v*inv(v'*v)*zbix;
z = z - zbix2;

%%
% now the data is only: [y, x, instruments]
Jexc=[]; %SET INDICES OF EXCLUDED REGRESSORS (beta=0)
Junc=[1:k]; %SET INDICES OF UNSURE REGRESSORS (beta SATISFIES SPARSITY)
Jinc=setdiff(1:k,union(Jexc,Junc)); %SET INDICES OF INCLUDED REGRESSORS (NO
    RESTRICTIONS ON beta)
Jexc_=[1:l]; %SET INDICES OF EXOGENOUS INSTRUMENTS (theta=0)
Junc_=[]; %SET INDICES OF POSSIBLY ENDOGENOUS INSTRUMENTS (theta IS SPARSE)
Jinc_=setdiff(1:l,union(Jexc_,Junc_));

sgrid = [1 2]; %SPARSITIES OF beta(Junc)

sgrid_=ones(size(sgrid)); %SPARSITIES OF theta(Junc_)

csind=[1]; %coefficients wishes to be estimated

kunc=length(Junc);
lunc=length(Junc_);

Jexcc=setdiff(1:k,Jexc);
kexcc=length(Jexcc);

csn=length(csind);

%% SNIV CONFIDENCE SET (GAUTIER ROSE TSYBAKOV SECTION 3.3, MODIFIED TO ALLOW
```

```
    E[Zu]=theta WITH SPARSE theta)

relmax=3;
relmin=2;
xub=10000;

sn=length(sgrid);

%r0=sqrt(2*log(2*l/0.05)/n);
r0=-(1/sqrt(n))*norminv(0.05/(2*l));

%% COMPUTE IDENTIFIED SET
a=zeros(l,1);
b=zeros(l,k+l);
Q=zeros(k+l,k+l,l);
parfor ll=1:l
    %M1=z(:,ll)*z(:,ll)'/n^2-r0^2*sparse([1:n]',[1:n]',z(:,ll).^2/n);
    M2=(r0^2-1)*z(:,ll)/n;
    M3=1-r0^2;

    a11 = y'*z(:,ll);
    A = sparse([1:n]',[1:n]',z(:,ll).^2/n);
    all  = a11*a11'/(n^2) - y'*r0^2*A*y;
    %all=y'*(z(:,ll)*z(:,ll)'/n^2-r0^2*sparse([1:n]',[1:n]',z(:,ll).^2/n))*y;

    bll=zeros(1,k+l);
    a11 = y'*z(:,ll);
    b11 = z(:,ll)'*w;
    bll(1:k) = -2*a11*b11/(n^2) + 2*r0^2*y'*A*w;
```

```
    %bll(1:k)=-2*y'*(z(:,ll)*z(:,ll)'/n^2-r0^2*sparse([1:n]',[1:n]',z(:,ll).^2/n))*w;
    bll(k+ll)=y'*M2;

    Qll =zeros(k+l,k+l);
    a11 = w'*z(:,ll);
    Qll(1:k,1:k) = a11*a11'/(n^2) - w'*r0^2*A*w;
    %Qll(1:k,1:k)=w'*(z(:,ll)*z(:,ll)'/n^2-r0^2*sparse([1:n]',[1:n]',z(:,ll).^2/n))*w;

    Qll(k+ll,1:k)=M2'*w;
    Qll(1:k,k+ll)=M2'*w;
    Qll(k+ll,k+ll)=1-r0^2;

    all=all/n;
    bll=bll/n;
    Qll=Qll/n;
    a(ll)=all;
    b(ll,:)=bll;
    Q(:,:,ll)=Qll;
end

csi=zeros(1,k+l);
csi(csind)=1;
csi(Jexc)=[];
csi=find(csi);
csn=length(csi);

progs=[csn sn 2];
nprogs=prod(progs);
```

```
CSp=NaN(nprogs,1);
STp=NaN(nprogs,1);
REp=NaN(nprogs,1);
parfor_progress(nprogs);
parfor kk=1:nprogs
    [cc,ss,bb]=ind2sub(progs,kk);
    kkk=csi(cc);
    s=sgrid(ss);
    s_=sgrid_(ss);
    [bn,st,re]=dopoly(l,k,kunc,Junc,Jexc,lunc,Junc_,Jexc_,Jinc,Jinc_,a,b,Q,kkk,s,s_,relmax,relm
    CSp(kk)=bn;
    STp(kk)=st;
    REp(kk)=re;
    parfor_progress;
end

for cc=1:csn
    for ss=1:sn
        for bb=1:2
            kk=sub2ind(progs,cc,ss,bb);
            if bb==1
                CS(cc,sn+1-ss)=CSp(kk);
                ST(cc,sn+1-ss)=STp(kk);
                RE(cc,sn+1-ss)=REp(kk);
            else
                CS(cc,sn+1+ss)=CSp(kk);
                ST(cc,sn+1+ss)=STp(kk);
                RE(cc,sn+1+ss)=REp(kk);
            end
```

```
        end
    end
end
CS(:,sn+1)=NaN;

CS
```

The supplemented codes:

dopoly.m function

```
function [bnd,st,re] =
    dopoly(l,k,kunc,Junc,Jexc,lunc,Junc_,Jexc_,Jinc,Jinc_,a,b,Q,kkk,s,s_,relmax,relmin,xub,bb)

    bet=[];
    gam=[];
    g=[];
    eps=[];
    h1=[];
    h2=[];
    i=[];
    del=[];
    h1_=[];
    h2_=[];
    i_=[];

    mset('yalmip',true)
    mset(sdpsettings('solver','mosek','verbose',0));
    %mset(sdpsettings('solver','scs','verbose',0,'scs.max_iters',5000,'scs.eps',1.35e-6));
```

```
    mpol('bet',k,1);
    mpol('gam',l,1);
    mpol('g',l,1);

    if kunc>0
        mpol('eps',kunc,1);
        mpol('h1',kunc,1);
        mpol('h2',kunc,1);
        mpol('i',kunc,1);
    end
    if lunc>0
        mpol('del',lunc,1);
        mpol('h1_',lunc,1);
        mpol('h2_',lunc,1);
        mpol('i_',lunc,1);
    end

    bet=bet(setdiff(1:k,Jexc));
    gam=gam(setdiff(1:l,Jexc_));

%     if isempty(Junc)
%         eps=[];
%     end
%     if isempty(Junc_)
%         del=[];
%     end

    junc=zeros(1,k);
    junc(Junc)=1;
```

```
junc=junc(setdiff(1:k,Jexc));
junc=find(junc);
if isempty(junc)
    junc=[];
end

junc_=zeros(1,l);
junc_(Junc_)=1;
junc_=junc_(setdiff(1:l,Jexc_));
junc_=find(junc_);
if isempty(junc_)
    junc_=[];
end

jinc=zeros(1,k);
jinc(Jinc)=1;
jinc=jinc(setdiff(1:k,Jexc));
jinc=find(jinc);
if isempty(jinc)
    jinc=[];
end

jinc_=zeros(1,l);
jinc_(Jinc_)=1;
jinc_=jinc_(setdiff(1:l,Jexc_));
jinc_=find(jinc_);
if isempty(jinc_)
    jinc_=[];
end
```

```
the=[bet;gam];
kp=[setdiff(1:k,Jexc),k+setdiff(1:l,Jexc_)];
for ll=1:l
    g(ll)=a(ll)+b(ll,kp)*the+the'*Q(kp,kp,ll)*the;
end

betunc=bet(junc);
for ll=1:kunc
    h1(ll)=betunc(ll)-xub*eps(ll);
    h2(ll)=betunc(ll)+xub*eps(ll);
end

for ll=1:kunc
    i(ll)=eps(ll)*(1-eps(ll));
end

gamunc=gam(junc_);
for ll=1:lunc
    h1_(ll)=gamunc(ll)-xub*del(ll);
    h2_(ll)=gamunc(ll)+xub*del(ll);
end

for ll=1:lunc
    i_(ll)=del(ll)*(1-del(ll));
end

K=[g<=0;bet(jinc)<=xub;bet(jinc)>=-xub;gam(jinc_)<=xub;gam(jinc_)>=-xub];
if kunc>0
```

```
        K=[K;i==0;0<=eps;eps<=1;sum(eps)<=s;h1<=0;h2>=0;(1-eps).*bet(junc)==0];
    end
    if lunc>0
        K=[K;i_==0;0<=del;del<=1;sum(del)<=s_;h1_<=0;h2_>=0;(1-del).*gam(junc_)==0];
    end

    f=the(kkk);

    if bb==1
        [bnd,st,re]=lbcompute(f,K,relmax,relmin);
    else
        [bnd,st,re]=ubcompute(f,K,relmax,relmin);
    end

end
```

---

lbcompute.m function (nested in the dopoly function)

```
function [lb,stlb,relb] = lbcompute(f,K,relmax,relmin)
    rel=relmin;
    status=-1;
    while status<1 && rel<=relmax
        P=msdp(min(f),K,rel);
        [status,obj]=msol(P);
        rel=rel+1;
    end
    lb=obj;
    stlb=status;
    relb=rel-1;
```

```
end
```

ubcompute.m function (nested in the dopoly function)

```
function [ub,stub,reub] = ubcompute(f,K,relmax,relmin)
    rel=relmin;
    status=-1;
    while status<1 && rel<=relmax
        P=msdp(max(f),K,rel);
        [status,obj]=msol(P);
        rel=rel+1;
    end
    ub=obj;
    stub=status;
    reub=rel-1;
end
```

$parfor_progress.m$to print out the results and keeping track of the solutions:

```
function percent = parfor_progress(N)
%PARFOR_PROGRESS Progress monitor (progress bar) that works with parfor.
%   PARFOR_PROGRESS works by creating a file called parfor_progress.txt in
%   your working directory, and then keeping track of the parfor loop's
%   progress within that file. This workaround is necessary because parfor
%   workers cannot communicate with one another so there is no simple way
%   to know which iterations have finished and which haven't.
%
%   PARFOR_PROGRESS(N) initializes the progress monitor for a set of N
%   upcoming calculations.
%
```

```
%   PARFOR_PROGRESS updates the progress inside your parfor loop and
%   displays an updated progress bar.
%
%   PARFOR_PROGRESS(0) deletes parfor_progress.txt and finalizes progress
%   bar.
%
%   To suppress output from any of these functions, just ask for a return
%   variable from the function calls, like PERCENT = PARFOR_PROGRESS which
%   returns the percentage of completion.
%
%   Example:
%
%      N = 100;
%      parfor_progress(N);
%      parfor i=1:N
%         pause(rand); % Replace with real code
%         parfor_progress;
%      end
%      parfor_progress(0);
%
%   See also PARFOR.

% By Jeremy Scheff - jdscheff@gmail.com - http://www.jeremyscheff.com/

%error(narginchk(0, 1, nargin, 'struct'));

if nargin < 1
    N = -1;
end
```

```
percent = 0;
w = 50; % Width of progress bar

if N > 0
    f = fopen('parfor_progress.txt', 'w');
    if f<0
        error('Do you have write permissions for %s?', pwd);
    end
    fprintf(f, '%d\n', N); % Save N at the top of progress.txt
    fclose(f);

    if nargout == 0
        disp(['  0%[>', repmat(' ', 1, w), ']']);
    end
elseif N == 0
    delete('parfor_progress.txt');
    percent = 100;

    if nargout == 0
        disp([repmat(char(8), 1, (w+9)), char(10), '100%[', repmat('=', 1, w+1),
            ']']);
    end
else
    if ~exist('parfor_progress.txt', 'file')
        error('parfor_progress.txt not found. Run PARFOR_PROGRESS(N) before
            PARFOR_PROGRESS to initialize parfor_progress.txt.');
    end
```

```
    f = fopen('parfor_progress.txt', 'a');
    fprintf(f, '1\n');
    fclose(f);

    f = fopen('parfor_progress.txt', 'r');
    progress = fscanf(f, '%d');
    fclose(f);
    percent = (length(progress)-1)/progress(1)*100;

    if nargout == 0
        perc = sprintf('%3.0f%%', percent); % 4 characters wide, percentage
        disp([repmat(char(8), 1, (w+9)), char(10), perc, '[', repmat('=', 1,
            round(percent*w/100)), '>', repmat(' ', 1, w - round(percent*w/100)),
            ']']);
    end
end
```

And other minor functions:

```
function Y = VChooseK(X, K)

function x=vec(X)
    x=reshape(X,[],1);
end
```