Relaxing the Exclusion Restriction in Shift-Share Instrumental Variable Estimation

Nicolas Apfel*

July 2, 2019

Abstract

The widely used shift-share instrument is generated by summing the products of regional shares and aggregate shifts. All products must fulfill the exclusion restriction, for the instrument to be valid. I propose applying methods which can preselect invalid products when either more than half or the largest group of products is valid. I discuss extensions of these methods for fixed effects models. I illustrate the procedures with three applications: a simulation study, the labor market effect of Chinese import competition and the effect of immigration to the US. My results help explain why previous studies have found low effects of immigration.

Keywords: Shift-Share instrument, Causal inference, Invalid instruments, Lasso

JEL classification: C36, C52, F22, F66

*I thank David Dorn, Benjamin Elsner, Helmut Farbmacher, Joachim Winter and seminar participants at the Ludwig-Maximilian University, TU Munich, DAGStat 2019, IAAEU Trier, the University of Regensburg and the Ammersee Workshop for helpful comments. I thank Rodrigo Adão, Michal Kolesár and Eduardo Morales for sharing their data and code. I acknowledge funding through the International Doctoral Program “Evidence-Based Economics” of the Elite Network of Bavaria. This paper has received the Best Paper Award at the 11th Workshop on Labor Economics at the IAAEU Trier.

Address: University of Regensburg, Department of Economics, RW(L) 4.04, E-Mail: nicolas.apfel@ur.de, Phone: +49 (0)941 943-2716
1. Introduction

The use of instrumental variables has been a common remedy to omitted variable bias and reverse causation in applied economics. Shift-share variables are a specific class of IVs which are broadly applied in labor, development, international economics and other fields. For example, when a researcher is interested in the effect of immigration on unemployment, the main empirical problem is that migrants self-select into more prosperous labor market regions, hence creating a positive correlation between outcome and exposure. A shift-share strategy uses the share of immigrants from a certain origin country from an earlier point in time and multiplies it with the relative inflow at a later point in time. This exercise is repeated for different countries of origin and the products are then summed up, yielding a single instrument. The shifts and shares used in the construction are different for each given application. In international economics, for example, the group-dimension commonly used is industries instead of countries.

Even though a single IV is used, the exclusion restriction in this case needs to hold for each origin country. Shifts or shares should not be directly correlated with the outcome variable through unobservable shocks or longterm effects. The exclusion restriction in shift-share estimation is very restrictive because it must hold for all origin countries and typically a researcher cannot a priori exclude structural relationships between instruments and outcome variables for all countries.

The main contribution of this paper is to propose an adjusted shift-share estimator which consistently estimates effects in the presence of sets of shift-share products for which the exclusion restriction is not fulfilled in general. To the best of my knowledge, this is the first paper proposing methods which allow for consistent estimation in this setting. A small but steadily growing literature in econometrics and epidemiology has made efforts to provide estimators that are able to consistently estimate the effect of an exposure in the presence of endogenous instruments after selecting the invalid instruments. These estimators have oracle properties. That means that asymptotically they perform as well as if the researcher knew the identity of invalid IVs.

Because in this setting potentially there is a large number of instruments, this field has attracted practitioners and theorists who are fluent in machine learning methods. With the help of the adaptive Least Absolute Shrinkage and Selection Operator (adaLasso Windmeijer, Farbmacher, Davies, and Smith, 2018) and the Confidence Interval Method (Windmeijer, Liang, Hartwig, and Bowden, 2019), I propose a two-step procedure: in the first-step, invalid instruments are selected. In the second step, I reconstruct the shift-share instrument without the instruments chosen as invalid.

Depending on the estimator used in the selection step, key assumptions for oracle properties are that only the majority of components or a plurality (the largest group) need to be valid. Under these assumptions, the modified shift-share estimator will also yield consistent estimates. In the migration setting, the majority assumption would mean that for more than 50% of the origin countries, long-term adjustments or correlation with unobserved shocks can be ruled out, while for the rest it is allowed. The plurality assumption means that the largest group of origin country-specific shift-share products is valid. A group is defined as a set of IVs which, if taken
separately, yield estimators which converge to the same numerical value. These new assumptions constitute a significant relaxation of the classical exclusion restriction, according to which all products have to be valid. I also provide an ado-file in Stata with which the methods can be easily applied by practitioners.

Many machine learning methods lend themselves mostly to predictive tasks (see Mullainathan and Spiess, 2017). However, my paper provides a remedy for a commonly seen endogeneity problem, which threatens the reliability of causal inference in applied work.

To show the implications and generality of the presented methods in practice, I apply them to three empirical examples. First, I apply the adjusted shift-share estimators on a panel data set in a Monte Carlo simulation. In this setting, the data-generating process is known, and hence the performance of the estimator can be evaluated. The results suggest that in economic applications with panel data the new method should be combined with first-differencing to eliminate fixed effects.

Second, I reestimate the effect of Chinese import exposure on employment (Autor, Dorn, and Hanson, 2013), because this work is exemplary for a long series of papers in international economics which use the shift-share IV. In this example, the effect size is robust to leaving potentially invalid instruments out of the construction of the shift-share IV. Up to 48 out of 397 industries are chosen as invalid. Overall, the adjustments only yield slight changes of the effect of import competition. The industries chosen as invalid most often are also those the authors worry about in their discussion. These are also some of the industries that are most sensible to misspecification according to Goldsmith-Pinkham, Sorkin, and Swift (2018).

Third, I estimate the impact of migration on labor market outcomes. I use this application because empirical work in migration economics relies heavily on the use of the shift-share instrument. For some of the estimations, the estimated effects on employment are negative and double in magnitude, whereas the estimated effects on wages of the high-skilled become significant and the coefficient size increases. In this application, a maximum of 31 out of 57 origin countries are chosen as invalid. Notably, Mexico and the Philippines are chosen as invalid. Mexico is the largest source country and the two countries are the ones with the highest sensitivity-to-misspecification as calculated by Goldsmith-Pinkham, Sorkin, and Swift (2018).

The remainder of this paper is structured as follows. Section 2 presents the exclusion restriction in more detail. Section 3 introduces methods which allow for the presence of invalid instruments in IV estimation. In Section 4, I present my approach to adjusted shift-share estimation, discuss the application of the new methods in a fixed effects model and present the adjusted shift-share estimator. I also discuss weak instruments, the power of the Hansen-Sargan test and whether invalid IVs should be included as controls. Section 5 applies the methods to a Monte Carlo simulation and two case studies. Section 6 concludes.
2. Shift-share instrumental variables

Consider a linear model with a constant treatment effect \( \beta \):

\[
y_{lt} = x_{lt} \beta + \epsilon_{lt},
\]

where \( l \) indicates the location (usually the region) and \( t \) the time period.\(^1\) \( y_{lt} \) is some outcome variable, \( x_{lt} \) a local treatment and \( \epsilon_{lt} \) an idiosyncratic error term which can be interpreted as unobservable shocks. For example, the outcome variable is employment growth in a certain region and year and the independent variable is growth of the immigrant share. To keep things simple, I abstract from covariates.

2.1. The instrument

The shift-share instrument is defined as the sum of local growth rates, disaggregated by industry or origin group, weighed by the share of each group:

\[
x_{lt} \equiv \sum_{j=1}^{J} s_{jlt} \cdot g_{jlt},
\]

where \( j \) indicates a group (e.g. the origin country of migrants), \( s_{jlt} \) is the group-specific share in a certain region and \( g_{jlt} \) is the region-specific growth-rate (or shift) of that group at time \( t \). For example, \( s_{\text{Mexico,CA,2019}} \) is the share of Mexicans in California in 2019 and \( g_{\text{Mexico,CA,2019}} \) is the inflow of migrants from Mexico to California in 2019. These shifts and shares are available for \( J \) origin countries.

In many settings, the local growth rates or the initial shares are correlated with unobserved shocks or have a direct effect on the outcome variable. In the migration context, Mexican migrants may have chosen to settle down in California precisely because of the good labor market prospects. Part of the correlation that would be measured in a least-squares regression would therefore be due to migrant self-selection into regions.

A shift-share approach takes the structure of the explanatory variable and replaces its components by shares and shifts which are presumably unrelated with the outcome variable: e.g., the share of Mexicans in California relative to Mexicans in the US is replaced with the same share, at a certain base period \( t^0 \) earlier in time (e.g. 1990), while the growth rate of Mexican immigrants in California is replaced by its equivalent at the aggregate (e.g. US-) level. The resulting shift-share IV is

\[
z_{lt} = \sum_{j=1}^{J} s_{jlt^0} \cdot g_{jt},
\]

where \( z_{jlt} \equiv s_{jlt^0} \cdot g_{jlt} \). \( z_{lt} \) is then used to instrument for \( x_{lt} \). The motivation for the relevance of this instrument is that today’s migrants settle in regions where they find communities of earlier migrants from their same country of origin and hence past and present settlement is correlated.

\(^1\)The constant treatment effect assumption will be relaxed in subsection 4.3.
2.2. Exclusion restriction

The key assumption for IVs is the exclusion restriction, also known as exogeneity or validity.\(^2\) The key orthogonality condition for consistency of the shift-share estimator is stated in Borusyak, Hull, and Jaravel (2018)\(^3\):

\[
\text{Cov}(z_l, \varepsilon_l) \overset{P}{\to} 0
\]

(3)

where \(\overset{P}{\to}\) denotes convergence in probability. This is a necessary condition for consistency of the shift-share IV estimator. There are two sufficient conditions for Equation 3 to be fulfilled: Either shares or shocks are exogenous. Goldsmith-Pinkham, Sorkin, and Swift (2018) state the key identifying assumptions in terms of shares:

**Assumption 1. Strict exclusion restriction**

\[
E[\varepsilon_l|s_{lj}] = 0 \quad \forall j \text{ where } g_j \neq 0
\]

Under Assumption 1 and instrument relevance, the shift-share IV estimator is consistent (Assumption 1 and Proposition 2 in Goldsmith-Pinkham, Sorkin, and Swift, 2018).

Note that this exclusion restriction is strict in the sense that it must hold for all \(J\) industries, origin countries or whichever group is used. What looks like a single exclusion restriction is in fact a set of \(J\) exclusion restrictions. Therefore, the researcher needs to feel comfortable defending the exclusion restriction for Mexicans, Cubans, Peruvians, Swedes, Syrians, Micronesians and all other origin groups used in the construction of the IV. In the migration example, the reasoning behind the use of lagged shares in the instrument is that if the lag chosen is long enough, there is no direct correlation between settlement patterns of Mexicans in 1990 and outcome variables in 2019. In other words, if the share of immigrants is sufficiently lagged, regions to which people migrated in the past are not those that prosper today. The only allowed channel is that earlier settlement of Mexicans in California attracts Mexicans in 2019.

The second way in which equation 3 could be fulfilled is that shocks are exogenous. This way to think about the exclusion restriction also motivates Autor, Dorn, and Hanson (2013), who use Chinese imports to other countries than the US. This alternative way to think about the exogeneity of the shift-share IV is formalized in Borusyak, Hull, and Jaravel’s (2018) Assumptions A1 and A2 (p.8).

2.3. Violations of the exclusion restriction

Why should the researcher worry about the validity of the shift-share instrument? The validity of instruments might be violated because neither shocks nor shares are valid. Jaeger, Ruist, and Stuhler (2018) warn that the exclusion restriction could be violated in migration applications, when there is a direct effect of lagged shares on the outcome variable of interest, because of long-term adjustment processes which have begun in the past. For example, the Marielitos who settled in Miami in the 1980s may have triggered responses such as native labor or capital adjustments which were still ongoing in 2000. Borjas (1999) mentions another way in which

\(^2\) These terms will be used synonymously in the following.

\(^3\) The time indices are dropped for simplicity.
the exclusion restriction could be violated. If past productivity shocks are correlated with past settlement and the former are serially correlated, this creates correlation between today’s outcome variable and the instrument. If Mexicans migrate to California for precisely the same reasons as the Mexicans in 1980, the exclusion restriction can not be upheld. When it is argued that the validity stems from industry-level shifts, there still might be unobserved shocks in $\varepsilon_{lt}$ that drive both instrument and the outcome variable. A few examples for this kind of violation of the exclusion restriction are presented in the empirical section.

In practice, it is very difficult, if not impossible to credibly uphold the strict exclusion restriction. While building an intuition about which components of shift-share products are valid might be feasible, arguing that none of the shifts or shares had a long-term effect or are correlated with unobservable shocks is a very restrictive assumption. Thinking about whether there have been adjustment effects of capital to the inflow of Mexicans in 1990 in the US and how long these have taken, makes it clear how difficult and hypothetical such an argument is destined to be. This holds true especially in settings in which a large number of shift-share elements is used. Such detailed knowledge about the structural mechanisms at work is only available for very few countries, if any.

Goldsmith-Pinkham, Sorkin, and Swift (2018) propose computing sensitivity-to-misspecification weights, which indicate by how much percent the bias from using the shift-share IV with endogenous instruments changes if the bias from a certain instrument increases by one percent. The authors point out that one should argue prudently for the validity of shares associated with large weights. While these weights indicate the relative importance with which an individual instrument contributes to the bias of the shift-share IV estimator. The latter can still be considerable in absolute terms, even if only IVs associated with low weights have a large bias. Therefore, it doesn’t suffice to argue for the validity of the share associated with the largest weights to make a case for a low bias in absolute terms.

In the following, I propose methods which allow to identify the invalid instruments. I then rebuild an adjusted shift-share instrumental variable estimator which uses only shift-share products which fulfill the exclusion restriction and therefore attains consistency in a setting in which Equation 3 is violated.

3. Selection of invalid IVs

In this section, I present two papers from an emerging literature that allow for invalid instruments when estimating causal effects. Even though the literature which tries to deal with invalid instruments encompasses other papers, the following two allow the researcher to be agnostic about the identity of invalid IVs. The presented methods have been developed primarily for the

---

4 Andrews (1999) proposes a procedure with information criteria based on the J-test for overidentifying restrictions. This moment selection criterion requires to search over all possible subsets, which makes it computationally infeasible, when the number of IVs becomes moderately large. Gautier, Tsybakov, and Rose (2011) and Caner, Han, and Lee (2018) also allow for the presence of invalid instruments; however, a set of valid instruments must be known a priori. Kolesář, Chetty, Friedman, Glaeser, and Imbens (2015) show that even when many instruments are invalid, under the assumption of uncorrelatedness between first stage effects and direct effects of instruments on the outcome, there is a consistent modification of the two-stage least squares (2SLS) estimator. However, this no-correlation assumption is rather strict and therefore I confine my attention...
use in epidemiology and more specifically for Mendelian Randomization, where genetic markers are used as IVs when estimating the effect of an exposure on an outcome.\(^5\) The problem with this kind of analysis is that usually genetic markers are correlated with the outcome through multiple channels.

### 3.1. Setup

Consider the model in Equation 4, which is the same as in Equation 1, augmented by the direct effects of IVs on the outcome:

\[
y_l = x_l \beta + Z'_l \alpha + \varepsilon_l. \tag{4}
\]

\(Z_l\) is a matrix of \(J\) industry-specific instruments and \(\alpha\) is a \(J \times 1\) parameter vector which indicates which instrument is endogenous. If a vector in \(Z_l\) is associated with a zero entry in \(\alpha\), this means the instrument is valid. If the entry is non-zero, the instrument has a direct effect on the outcome and is hence invalid. Moreover, let \(Z_I\) be the matrix of invalid instruments with \(I = \{j : \alpha_j = 0\}\) and \(\hat{I}\) the set of instruments selected as valid. Accordingly, let \(Z_V\) be the matrix of valid instruments with \(V = \{j : \alpha_j = 0\}\) and \(\hat{V}\) the set of instruments selected as valid. Then \(|Z_I|\) is the number of invalid and \(|Z_V|\) the number of valid instruments.

The advantageous properties of the methods that will be leveraged in the context of shift-share estimation are the so-called “oracle properties”. Oracle properties mean consistent selection of invalid IVs and convergence in distribution to the asymptotic distribution of the ideal estimator that uses the model under perfect knowledge about the identity of invalid IVs.

The oracle properties are:

- Consistent selection of invalid IVs: \(\lim_{n \to \infty} P(\hat{I} = I) = 1\)
- Convergence in distribution: \(\sqrt{n}(\hat{\beta} - \beta_0) = N(0, \sigma^2_{or})\),

where \(\sigma^2_{or}\) is the variance of the oracle estimator.

### 3.2. Adaptive Lasso

In the following, I summarize the key ideas of the methods used and summarize the assumptions under which they have oracle properties in the shift-share setting.

Windmeijer, Farbmacher, Davies, and Smith (2018, WFDS) show that the Lasso proposed in Kang, Zhang, Cai, and Small (2016) does not consistently select endogenous instruments if valid and invalid instruments are correlated or instruments have unequal strength. They instead propose the adaptive Lasso which chooses invalid instruments to then apply 2SLS with the instruments which have been assigned coefficients \(\alpha_j = 0\). An initial consistent estimate used for weighting is given by the median of IV estimates for exactly identified models (Han, 2008). This estimator is consistent when less than 50% of the instruments are invalid and the key requirement for the adaLasso to have oracle properties is that it uses an initial consistent

---

\(^5\)For example, Windmeijer, Farbmacher, Davies, and Smith (2018) estimate the effect of the Body Mass Index on diastolic blood pressure.
estimate. The key assumption for the adaLasso to have oracle properties hence also is that the number of valid instruments exceeds one half.

**Assumption 2. Majority exclusion restriction**

$$|V| > J/2$$

According to Theorem 1 and Proposition 3 in WFDS, the adaptive Lasso has oracle properties when Assumption 2 is fulfilled. As compared to Assumption 1, this assumption is already a considerable relaxation. Note that the estimator having oracle properties does not depend on the different strength or the correlation of instruments.

The result of adaLasso is dependent on the penalty parameter $\lambda$, which determines how many parameters are shrunk to zero. WFDS propose two main options for the choice of $\lambda$. The first option is to choose $\lambda$ via cross-validation, so that the squared $L^2$-norm used in equation (17) is minimized. A second possibility, which WFDS favor is the use of the Hansen-J statistic in a stopping rule, testing at each adaLasso step. More details on the method are presented in Appendix A.1.

### 3.3. Confidence Interval Method

Windmeijer, Liang, Hartwig, and Bowden (2019) develop the Confidence Interval Method (CIM) which further relaxes the majority assumption. This method builds on Guo, Kang, Cai, and Small’s (2018) two-stage hard thresholding (TSHT) which still has oracle properties when the majority assumption is violated.6

The plurality condition in assumption 2 in Windmeijer, Liang, Hartwig, and Bowden (2018) states that the group of valid instruments is larger than any other group, where a group is defined as a set of IVs associated with an estimate which asymptotically deviates from the true $\beta$ by the same constant $c = \frac{\alpha_j}{\gamma_j}$. For the valid group $c = 0$. Formally

**Assumption 3. Plurality exclusion restriction**

$$|V| > \max_{c \neq 0} \{j : \frac{\alpha_j}{\gamma_j} = c\}$$

The Confidence Interval Method sets a critical value $\psi$ and calculates confidence intervals (CI) for each just-identified model estimate. These confidence intervals are then ordered by their lower endpoints and these are compared to the upper ones of each CI preceding it in order. If the upper endpoint of the $k$-th interval is larger than the lower endpoint of the $j$-th interval, the estimates are said to belong to the same group. The largest group corresponds to the set of estimates with the most overlapping confidence intervals. $cil_j$ and $ciu_j$ denote lower and upper endpoints of the CI from using the $j$-th instrument. For instruments $j = 1, ..., J$, $n_{i,j} = \sum_{k=1}^{J-1} 1(ci_{ik} > cil_j)$

---

6I restrict my attention to the Confidence Interval Method, because TSHT preselects individually strong IVs. However, in the shift-share setting, the composite instrument needs to be strong, but not necessarily all of its components. Moreover, Windmeijer, Liang, Hartwig, and Bowden (2019) show that the CIM outperforms TSHT in many settings in terms of the correct selection of invalid IVs.
is the number of overlapping intervals, when comparing from instrument \( j \) downwards. The key assumption for the confidence interval method having oracle properties is assumption 3.

Windmeijer, Liang, Hartwig, and Bowden (2018) treat the critical value \( \psi \) as a tuning parameter. The result is dependent on its value. For a large value of \( \psi \), all CIs will overlap and hence all variables will be chosen as valid. Gradually decreasing the value of \( \psi \) narrows the confidence intervals down, and drives the number of IVs chosen as valid down. The Hansen-Sargan test of overidentifying restrictions can be performed to choose an optimal level of \( \psi \), analogously to the choice of \( \lambda \) in adaLasso. After having pre-established a significance level of the test, the confidence interval comparisons are performed and \( \psi \) gradually decreased, until the \( H_0 \) is not rejected any more.

When comparing adaLasso and CIM, it becomes clear that the key condition for adaLasso to have oracle properties is more strict than the one needed for CIM. Why should a researcher then rely on adaLasso? First, adaLasso is slightly more established than CIM. Second, in the Monte Carlo exercise of section 5.1, when the majority condition holds, the CIM tends to select too many IVs as invalid. Therefore, I recommend using both methods and to concentrate on the results of CIM, when many IVs are chosen as invalid, suggesting a violation of the majority assumption.

4. Shift-share IV with invalid products

In the previous section, new methods for IV with many invalid instruments have been presented. In this section, I show how these methods can be applied to yield a modified shift-share instrumental variable estimator which is robust to the presence of endogenous shift-share products. I then extend the methods to a setting with fixed effects and present an extension of the plurality assumption to the setting with heterogeneous effects. Finally, I discuss low power of the Hansen-Sargan test, weak instrument issues and discuss whether invalid shift-share products should be controlled for.

4.1. A procedure to select invalid shift-share products

When using shift-share instruments, the shift-share products are summed, and the model is exactly identified. As discussed in section 2 all of the products need to be valid by themselves. Therefore, each product can also be used on its own in an overidentified model. Instead of using 2SLS assuming that all products are valid, the relaxed exclusion restrictions become that the majority or plurality of shift-share products are valid instruments. If the majority or plurality assumptions hold for shift-share products, the adaLasso or Confidence Interval method can be applied, respectively.

In short, the proposed procedure works as follows:

1. Multiply shifts and shares, and yield matrix \( Z \) of shift-share products with elements

\[
z_{ljt} = s_{ljt} \cdot g_{jt}, \quad j \in \{1, \ldots, J\}
\]

2. Run adaLasso or CIM of \( y_l \) on \( x_{lt} \) using \( Z \) as instrument matrix
3. Use IVs chosen as valid (associated with zero elements of $\alpha$) for the construction of the corrected instrument:

$$\sum_{j \in \hat{V}} s_{ljt} \cdot g_{jt}$$  \(5\)

4. Estimate shift-share IV$^7$

Consistency of the proposed method follows straightforwardly from already given proofs. AdaLasso has oracle properties when the majority of instruments is valid (see proposition 3 in WFDS), CIM has oracle properties when the plurality condition holds (see Theorem 2 in Windmeijer, Liang, Hartwig, and Bowden, 2018). Asymptotically, valid instruments are selected as valid. Hence, the instruments used for the construction of the corrected shift-share IV fulfill the key orthogonality assumption in equation (3), i.e. $E(z_{ljt} \varepsilon_t) \overset{P}{\rightarrow} 0$ for $j \in \hat{V}$ from which it directly follows that $E\left[\sum_{j \in \hat{V}} z_{ljt} \varepsilon_t\right] \overset{P}{\rightarrow} 0$. The resulting shift-share IV is hence consistent, according to Proposition 2.1 in Goldsmith-Pinkham, Sorkin, and Swift (2018).

Since in step 4 some products can’t be used for the construction of the adjusted estimator, the shares don’t sum to one. Therefore, the incomplete shares case discussed in Borusyak, Hull, and Jaravel (2018) (section 3.2) applies. As indicated there, for missing shares, the corresponding shares should be included, but with shifts of zero.$^8$

4.2. Fixed effects

In economic applications, the researcher often has panel data and would like to allow for unit-specific intercepts which are correlated with both the outcome and the treatment. To eliminate these fixed-effects, first-differencing, or demeaning the data beforehand is necessary. These transformations can be combined with IV techniques as summarized in Wooldridge (2010). In the following applications I do the same before using the new adaLasso and CI methods.

Consider the following fixed-effects model

$$y_{lt} = x_{lt}\beta + C_l + \varepsilon_{lt}$$  \(6\)

with $l = 1, \ldots, L$ and $t = 1, \ldots, T$, and where $C_l$ denotes unobserved, regional heterogeneity.

The first-differencing equation that eliminates fixed effects

$$\Delta Y_{lt} = \Delta D_{lt} + Z_{lt}\alpha + \Delta \varepsilon_{lt}$$  \(7\)

is augmented by $Z_{lt}$ (in matrix $Z$). The first stage equation is

$$\Delta D_{lt} = Z_{lt}\gamma + \nu_{lt}.$$  \(8\)

$I$ use the standard errors proposed in Adão, Kolesár, and Morales (2018). The industry-level regressions detailed in Borusyak, Hull, and Jaravel (2018) are used.

$^8$Note, that the commands presented in the Appendix do not account for corrected standard errors yet and hence Borusyak, Hull, and Jaravel’s (2018) ssaggregate should be used complementarily.
The key assumptions for consistency of a procedure which combines first-differencing with the proposed instrument selection procedure are a standard rank condition and the following exogeneity condition:

**Assumption 4. FD exclusion restriction**

\[ E(Z^j_t \Delta \varepsilon_{it}) = 0 \text{ for } t = 2, \ldots, T. \]  

(9)

This is a standard assumption, as in Assumption FDIV.1 in Wooldridge (2010, p. 362). The majority condition holds for the fixed effects model if assumption 4 holds for the majority of instruments \( j \) and the plurality condition holds if the largest group of instruments fulfills the assumption. First-differencing is efficient as compared with demeaning when the error term follows a random walk, because then homoscedasticity and no serial correlation are given. The Monte-Carlo exercise in section 5.1 provides supportive evidence for the consistency of the methods in combination with first-differencing.

### 4.3. Shift-share instrumental variable selection with heterogeneity

The methods proposed in previous sections rely on the constant treatment effects assumption. Goldsmith-Pinkham, Sorkin, and Swift (2018) and Borusyak, Hull, and Jaravel (2018) allow for location-specific coefficients \( \beta_l \). Under Goldsmith-Pinkham, Sorkin, and Swift’s (2018) Assumption 3 (monotonicity), each instrument estimates a weighted average of location-specific effects. The shift-share estimate is a weighted combination of \( \hat{\beta}_j \) IV estimates from using each group-specific IV at a time. The LATE-like interpretation of the shift-share estimator is therefore a weighted average of industry specific weighted averages. Because the weighted averages may differ, the group-specific estimates \( \hat{\beta}_j \) are also different. As Goldsmith-Pinkham, Sorkin, and Swift (2018) point out, the shift-share IV does not have a LATE-like interpretation in presence of negative Rotemberg weights.

Can the selection methods proposed before deal with heterogeneous effects? One practical solution is to perform the analogous analyses for clusters of industries, inside of which the constant effect assumption is believed to hold. If the constant effect assumption is more generally violated, another approach is needed.

WFDS do not make mention of a setting with heterogeneous effects and uphold the constant treatment effect assumption throughout the paper. Therefore, I further assume that adaLasso does not allow for heterogeneous treatment effects. However, Windmeijer, Liang, Hartwig, and Bowden (2018) mention the possibility of heterogeneity. The Hansen-Sargan test might also reject the Null hypothesis of valid moments when some instruments estimate local effects. In this case, the methods would treat valid IVs with heterogeneous effects as invalid and discard the information contained in them.

Following this comment, I make the assumptions under which the CIM still consistently selects valid IVs explicit: Intuitively, if the largest group of IVs is valid and associated with the same constant effect, the remaining IVs - even when valid - are discarded as invalid.\(^9\) The adjusted

---

\(^9\)I thank Helmut Farbmacher for pointing this out.
shift-share estimate then identifies a weighted combination of a restricted set of industries, as compared to the initial set.

Assume a setting with industry-specific heterogeneous treatment effects, with the index $h \in \{1, \ldots, H\}$. Treatment effects are denoted by $\beta_h$. Invalid instruments yield an estimator which is inconsistent with the constant $c \in \mathbb{R}$, as before. The $J$ estimators in the exactly identified case converge against $Q + 1 \geq H$ values $\phi_q$ with

$$q \in \{0, \ldots, Q\} : \hat{\beta}_j \xrightarrow{P} \phi_q = \beta_h + c.$$  

Each group of IVs now is a set which - if taken on its own - yields an estimator converging against a certain $\phi_q$:

$$G_q = \{j : \hat{\beta}_j \xrightarrow{P} \phi_q = \beta_h + c\}$$

$V \equiv G_0 = \{j : \hat{\beta}_j \xrightarrow{P} \beta_0\}$

$V$ is a group of valid IVs ($c = 0$) which is associated with a certain treatment effect $\beta_0$. Other IVs may be valid or invalid. The new plurality assumption is that instruments with $\beta_0$ (i.e. $c = 0$) form the largest group.

**Assumption 5. LATE Plurality**

$$|V| > \max_q(|G_q|)$$

The new assumption still states that the plurality of instruments has to belong to the same group and have constant treatment effects, but the remaining instruments could be invalid or valid, with effects which differ from those in the plurality group. CIM tries to find the largest group of IVs associated with the same $\beta_h$. If the groups of estimators which converge against a specific $\phi_q$ which is not $\beta_0$ are of smaller cardinality than $V$, then the procedure still identifies the largest group of valid instruments with the same treatment effect, but discards the information of additional valid IVs which estimate other LATEs. Appendix A.2 illustrates the LATE plurality assumption with an example.

If the LATE plurality assumption holds and the adjusted shift-share estimator which uses only IVs selected as valid is used, one concern could be that interpretability is changed, because shares do not sum to one anymore. Does this alter the LATE-like interpretation? If Goldsmith-Pinkham, Sorkin, and Swift’s (2018) Assumption 3 still holds, the estimates can still be interpreted as weighted average. The weights $\frac{z_i \pi_i k}{\sum_{i'=1}^{z_i} \pi_{i' k}}$ still sum to one, even though the shares don’t. What changes is the industry set with which the estimate is computed. In the standard shift-share estimation all industries enter the construction of weights, while for the adjusted methods, invalid shift-share products are not used. Therefore, a weighted combination of a subset of industry-specific instruments is estimated.

4.4. **Power of the Hansen-Sargan test for overidentifying restrictions**

To choose the number of invalid instruments, WFDS follow Belloni, Chen, Chernozhukov, and Hansen (2012) when using the significance level of $0.1/\ln(N)$. It has been shown that the use
of many moment conditions leads to low power when using the Hansen-Sargan test (Bowsher, 2002). Moreover, as noted in Roodman (2009), larger significance levels of the Hansen-Sargan test are more conservative, which is the inverse logic as with conventional tests of coefficient significance. Hence, a more conservative strategy would be to set the significance level of the test used for determining the validity of instruments to a more conventional level, for example to $p = 0.05$. Another practical measure to tackle the problem of too many instruments is to find clusters. One can draw sets randomly or just use naturally occurring groups if heterogeneous effects are also a concern.

The methods that will be applied in the following use the Hansen-Sargan test for overidentifying restrictions to select the number of variables chosen as invalid.

### 4.5. Weak shift-share products

One concern, which is related to the first stage, is that instruments are individually weak. This might be relevant here, because each IV taken on its own is used to predict the endogenous variable. When a certain industry is used to predict variation from a similar industry, the IV can still be expected to be relevant. However, the correlation of shift-share products with variation in the endogenous variable stemming from other industries is probably weak. Individually weak shift-share products are not a problem for the shift-share IV, because the model is exactly identified, and only the composite instrument needs to be relevant.

WFDS are concerned about the consistency of the median estimator. However, in the case with exact identification the weak IV problem is attenuated. Moreover, the oracle properties of adaLasso do not depend on the relative strength of IVs.

If one were still concerned about the strength of IVs, a straightforward solution would be to preselect IVs such that all instruments are strong. However, this could mean to throw out the baby with the bathwater, because some valid but weak instruments, which could contribute to a stronger shift-share instrument would be discarded. A milder alternative to this drastic cure would be to test for relevance of the instruments after selection of IVs, e.g. by looking at the F-statistic of the first stage for the selected model. A weak instrument test could even be incorporated in the stopping rule, alongside the Hansen-Sargan test.

### 4.6. Controlling for invalid shift-share products

One last important question is whether the instruments chosen as invalid are included in the structural equation. If one thinks that there might indeed be a direct effect on the outcome or the invalid product is correlated with both endogenous variable and the error term, then an inclusion makes sense because otherwise omitted variable bias is an issue. However, controlling for them might mechanically increase the variance of estimators because of a loss of degrees of freedom. Therefore, there could be a trade-off between omitting and including the shift-share products from the structural form. A midway would be to sum the endogenous shift-share products to a shift-share control variable so that a single variable is used as control. Effectively, this amounts to restricting the coefficient of all controls to the same value. An alternative would be to include the products when many instruments are chosen as invalid and to omit
them altogether when only very few instruments are chosen as invalid and the selection seems to have happened at random. The cautious researcher can include the shift-share products in a robustness check while reporting the variance of coefficients and first-stage F-statistics.

5. Empirical Applications

In the following, I use the proposed methods in a Monte Carlo exercise to show that the methods outperform the standard shift-share estimator. I then reestimate the effect of Chinese import competition on employment with Autor, Dorn, and Hanson’s (2013) data and of immigration on native labor market outcomes as in Adão, Kolesár, and Morales (2018). I first reproduce the original estimates by using the standard shift-share IV which uses all shift-share products, irrespectively of their validity. I then compare this regression with the result of the adjusted shift-share estimator, using adaLasso and the confidence interval method. I also use subgroups of instruments for which the assumption of constant treatment effects is more likely to hold, estimate the standard shift-share estimator and then the post-adaLasso shift-share estimator with these groups separately.

5.1. Example 1: Monte Carlo simulation

The following Monte Carlo simulations support the view that the shift-share IV estimators adjusted with adaLasso and the confidence interval method performs better in terms of bias as compared to the standard shift-share IV estimator, which uses all products.

The data is created based on the following fixed effects model with unit-root errors, with the structural equation

\[ y_{lt} = x_{lt}\beta + Z_{lt}\alpha + 10C_l + \varepsilon_{lt} \tag{10} \]

and the first stage

\[ x_{lt} = Z_{lt}\gamma + \upsilon_{lt}, \tag{11} \]

where \(\beta\) to 0 and the elements of the vector \(\gamma\) to 0.2. I assume that there are ten shift-share products (\(J = 10\)). To create shares, I draw a matrix with \(J\) columns from a uniform distribution between 0 and 10. I then sum each row up and divide each observation through the sum of the row. In this way, each row sums to one, mimicking the share variable. The share matrix has only \(N/2\) distinct rows, so that the setting is one where \(T = 2\). The shift component is an \(N \times J\) matrix drawn such that \(\text{Shifts} \sim N(1,10)\). The shift-share variable is created by element-wise multiplication of the shift- and share-matrices. \(C\) is drawn from a normal distribution with mean 2 and variance 1, but is fixed over \(t\). It is constructed such that \(\text{Cor}(x_{lt}, C_l) \neq 0\) and \(\text{Cor}(y_{lt}, C_l) \neq 0\).

In this simulation exercise I assume that shifts vary by region and time \(t\). In unreported simulations, when letting the shifts vary only by \(t\), I observed that \(\beta\) is estimated with little precision. I attribute this to perfect correlations between the shifts. These correlations occur mechanically when \(T = 2\). For growing \(T\), this problem disappears. This problem is interesting

\[10^{th}\text{The setup of these simulations follows the setup of the simulations in WFDS.}\]
Figure 1: FD-SSIV, with controls

Note: Performance of shift-share IV adjusted with adaLasso and confidence interval method in Monte Carlo simulations (500 replications). First-differencing done beforehand to eliminate fixed-effects. IVs chosen as invalid are included as controls. Horizontal axis: Number of observations. First row: median absolute deviation (MAD), second row: number of IVs chosen as invalid, third row: relative frequency with which all invalid IVs have been chosen as invalid.
in its own right and deserves to be investigated in greater detail. In this paper, I focus on the selection of valid shift-share products. Therefore, in this example I simply assume that shifts also vary by region.

In a first simulation, the vector indicating endogeneity, $\alpha$ in equation 10, is set to $\alpha = (0.2, 0.2, 0.2, 0, 0, 0, 0, 0)$, so that a majority of shift-share products is still valid. In a second simulation, the vector is set to $\alpha = (0.2, 0.2, 0.5, 0.5, 0.7, 0.7, 0, 0, 0)$, such that only the largest group of IVs is valid.

The error terms are such that

$$
\varepsilon, \nu \sim I(1)
$$

$$
\varepsilon_{lt} = 0.5\nu_{lt} + u_{lt}
$$

$$
u \sim N(0, 1)
$$

$$
t \in \{1, 2\}.
$$

(12)

To eliminate unobserved heterogeneity, I take first differences (FD) of $y_{lt}$, $x_{lt}$ and instruments and then use adaLasso and the confidence interval method. I run Monte Carlo simulations and vary the sample size from 1000 to 30,000 observations, gradually increasing the sample size by 1000. The number of repetitions is 100 for each parameter combination, each time drawing the errors anew.

Results The first baseline estimator with which the new estimators are compared is the standard shift-share, for which all products are assumed to be valid and are used for the construction of the shift-share instrument. The second estimator which serves as comparison is the oracle shift-share estimator, for which only valid IVs are used for the construction of the shift-share IV and invalid ones are used as controls. For both baseline estimators, I first-difference data to get rid of fixed effects. In Figure 1, I first compare these two estimators with the new estimators adjusted by adaLasso and CIM, which also use first-differenced data.11 In a second step, in Figure 2, I go on to compare the baseline estimators with the adjusted ones, but without first-differencing to illustrate the importance of accounting for fixed effects. The main results are that the adjusted estimators with first differencing outperform the standard shift-share estimator for almost all sample sizes and settings and approach the performance of the ideal estimator.

The graphs on the left of Figure 1 depict the setting in which a majority (seven out of ten) of IVs is valid. The graph in the first row on the left, shows the median absolute deviation (MAD) for each sample size. The solid line depicts the performance of the standard shift-share IV. The MAD is at about 0.3 and does not decrease as the sample size gets larger. The oracle shift-share estimator’s median absolute deviation is below 0.1 and gets closer to 0 as N increases. Notably, the MAD of the shift-share estimator adjusted by adaLasso visualized by the dashed-dotted line, equals the MAD of the oracle estimator (the grey, solid line) already for moderate sample sizes as 2000. In the second and third rows, it becomes clear why that is the case: From a sample size

---

11Note, that after transforming the data, in this setting with $T = 2$ one is left with only half the sample size when running the selection algorithms.
of 2000 upwards, only 3 IVs are chosen as invalid on average and in 100% of the cases the chosen IVs are the invalid ones. Therefore, in the setting of this simulation, the adjusted shift-share estimator attains oracle properties from a relatively low sample size on.

The dashed lines represents the performance of the estimator adjusted by the CIM. Again, the MAD quickly approaches the oracle performance. However, now six IVs are chosen as invalid and the relative frequency with which all invalid IVs are selected is only at about 75%. This means that for the specific setting of this simulation the significance level chosen for the stopping rule leads to the selection of some valid IVs as invalid, but of some invalid IVs as valid, even though the performance of the adjusted shift-share estimator still approaches that of the oracle estimator.

The graphs on the right of Figure 1 show the results from the setting in which a plurality of IVs is valid (four out of ten). Here, the estimator adjusted by adaLasso fares only slightly better than the standard shift-share, with a MAD at about 1, which does not decrease with growing sample size. On average, about eight IVs are selected as invalid, but in no MC replication all invalid IVs are correctly selected as invalid. This can be seen from the right graph in row three: the dashed-dotted line and the solid black line coincide. Selection via the CIM yields a performance which is equal to the oracle shift-share from a sample size of about 5000 upwards. The average number of IVs selected as invalid reaches six when $N = 2000$ and the frequency with which all invalid are selected as invalid is close to 100% from $N = 3000$ on. This is in line with the predictions. When the majority rule holds, both methods should work well. AdaLasso is expected to break down when only the plurality rule holds, which it does.

Since in practice one might be concerned that controlling for all invalid shift-share products separately mechanically drives up the variance estimate through the degree of freedom correction, the models selected by adaLasso and CIM are reestimated, once without shift-share products as controls and once with the sum of endogenous shift-share products as control. In row 1 of Figure 6 in the Appendix, I show the MAD of estimations when products chosen as invalid are not used as controls. In row 2, I repeat the estimations with the shift-share control. The results are only marginally worse if anything, with no controls at all and with the shift-share control. However, when shift-share products are correlated among each other, controlling for invalid instruments should make a difference. A look at the correlation structure of IVs can lend guidance on whether controls should be used.

Figure 2 shows that when the selection algorithms are run without taking care of unobserved heterogeneity through first-differencing, their performance can be much worse than that of the standard estimator. The MAD can greatly exceed that of the standard estimator. All measures vary irregularly with sample size. In the case of the plurality assumption holding, for example, CIM almost never chooses all invalid as invalid. Therefore, pre-transforming the data is of central importance when these new methods are applied to a panel data set.

5.2. Example 2: The China Shock

In this section I apply the adjusted shift-share estimator to the estimation of the effect of import exposure on manufacturing employment in the US. I first present the original approach used by
Figure 2: SSIV, with controls

Note: Performance of shift-share IV adjusted with adaLasso and confidence interval method in Monte Carlo simulations (500 replications). No first-differencing beforehand. IVs chosen as invalid are included as controls. Horizontal axis: Number of observations. First row: median absolute deviation (MAD), second row: number of IVs chosen as invalid, third row: relative frequency with which all invalid IVs have been chosen as invalid.
the authors, discuss instrument endogeneity issues and then present my own results. I apply the adaptive Lasso to the entire dataset and then apply the same method to industries from two-digit SIC-codes. I also apply the confidence interval method to the entire dataset.

**Approach**  
Autor, Dorn, and Hanson (2013, ADH) study the impact of Chinese imports on employment in manufacturing in the US. The regression equation is

$$\Delta L_{mt} = \Delta IPW_{uit} \beta_1 + X_{it}' \beta + \gamma_t + \varepsilon_{it},$$  

where the LHS is decadal change in manufacturing employment in commuting zone $i$, $\beta_1$ is the coefficient of interest and $\Delta IPW_{uit}$ is import exposure, defined as $\sum_j z_{ijt} g_{jt}$, where $z_{ijt}$ are the share of workers in commuting zone $i$ employed in industry $j$ at time $t$ and $g_{jt}$ measures the growth of Chinese imports in industry $j$. This regression is estimated in first-differences to get rid of fixed effects and augmented by a time dummy and a set of commuting-zone-level controls. The time period used ranges from 1990 to 2007 and 397 industry shares, indexed by four-digit SIC codes, have been used.

The endogeneity issue that affects this kind of analysis is that both employment and imports might be correlated with unobserved shocks to US demand. To tackle this, a shift-share instrument is used, which replaces the share of workers with the same share ten years earlier, i.e. $t^0 = t - 10$ in Equation 2, and uses import exposure of other high-income countries rather than the US. As the authors use imports to other high income countries as shifts, it comes natural to think of as-good-as-random shifts as motivating the exclusion restriction (Borusyak, Hull, and Jaravel, 2018). Hence, the example is situated in the exogenous shocks setting.

When estimating equation 13 with 2SLS, ADH find a coefficient of -0.596. I report the same coefficient for the original IV estimate in column 1 of table 1.\textsuperscript{12} This is the baseline coefficient to which the results from adjusted shift-share estimation will be compared.

**Instrument endogeneity issues**  
This strategy is only credible to the extent to which demand shocks are not correlated between the US and other high-income countries. The authors discuss the possible invalidity of the computer industry. Demand shocks in this industry might be correlated because the US and other countries are all subject to innovation shocks in information technology. In addition, Goldsmith-Pinkham, Sorkin, and Swift (2018) show that electronic computers display the highest sensitivity-to-misspecification weight, making the validity of this specific shift-share product especially important.

A second industry group for which the exclusion restriction is questioned are construction-related industries. In some of the years in the dataset, high-income countries have undergone housing booms in which construction materials have experienced high demand. This would lead to a correlation of demand shocks across countries and hence invalidate the instruments.

The third industry the authors express concern about is the apparel, footwear and textile industry, because China has been among the main exporters of these products.\textsuperscript{13}

\textsuperscript{12}I use the data provided on Kyrill Borusyak’s Github account.

\textsuperscript{13}The authors use a gravity approach with which they try to isolate supply and trade-cost driven changes in export performance for robustness and show that the results are similar to those of the original IV estimate.
Adaptive Lasso  Using adaLasso, keeping the default significance level of the overidentification test as proposed in WFDS at $0.1/\ln(N) = 0.01375$, the test does not reject the Null hypothesis that all instruments are valid and hence all instruments can be used for the construction of the shift-share instrument. Therefore, the coefficient is identical to the original shift-share estimate (column 2). To account for the problem of too many IVs in the HS-test, I set the significance level to the more conventional $p = 0.05$. After doing this, only motor vehicle parts and accessories (SIC3714) is chosen as invalid. When excluding this industry from the construction of the instrument, the estimate slightly decreases, to -0.554 (column 3).

Table 1: Results of China shock application

|                      | (1) Original | (2) AdaLasso | (3) AdaLasso | (4) AL SIC | (5) CIM | (6) CIM |
|----------------------|-------------|-------------|-------------|-----------|--------|--------|
| Δ Import Exposure    | -0.596***   | -0.596***   | -0.554*     | -0.471    | -0.696* | -0.716* |
|                      | (0.114)     | (0.114)     | (0.116)     | (0.130)   | (0.114)| (0.221)|
| F                    | 47.64       | 47.64       | 48.73       | 39.01     | 47.64  | 26.57  |
| Nr inv               | 0           | 0           | 1           | 26        | 0      | 48     |
| Sign                 | 0.01375     | 0.01375     | 0.05        | 0.01375   | 0.01375| 0.05   |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Adaptive Lasso by SIC-industry  If one argues that treatment effects vary across two-digit SIC industries, the adaLasso can be used for each of these groups separately. The standard shift-share IV estimator is run by two-digit SIC code and the estimates are compared with those adjusted by adaLasso. This also could help tackle the problem of too many IVs which is likely to occur in this example, in which 397 industries are used.

The results are displayed in Figure 3. It is reassuring to see that almost all confidence intervals include the original estimate, offering support for the constant treatment effect assumption. Moreover, in most of the twelve SIC-classes for which some shift-share products are excluded by the algorithm, the point estimates move towards the original estimates and often approach them very closely (the green point estimates often move onto the red line).

Of 397 instruments 26 are chosen as invalid. The exact industries chosen as invalid are listed in table 2.

The 2-digit SIC industry with the most products chosen as invalid is industrial and commercial machinery and computer equipment (SIC35). Five sub-industries have been chosen as invalid in this particular industry. This is especially interesting, because this is exactly the industry that is suspected of invalidity by ADH. Specifically, the industry with the highest sensitivity to misspecification (electronic computers) has also been chosen as invalid. The choice of two sub-industries from SIC36 (electronic, electrical equipments and components, except computer equipment), which is complementary for information technology is consistent with this story. Only using the shift-share products of the computer industry for the construction of the standard instrument leads to a coefficient which is close to zero. Omitting the IVs chosen as invalid from the construction drives the estimate towards the original coefficient found in ADH.

However, the concerns mentioned can not be ruled out completely, as no conclusive evidence on the validity of the exclusion restriction for all industries is available.
Table 2: Industries and countries chosen as invalid

| Analysis | Table, Column | Excluded SIC codes / countries |
|----------|---------------|-------------------------------|
| ADH - China Shock | 2, 3 | 3714 |
| adaLasso | 2, 4 | 2024 2082 2211 2252 2676 2761 2865 3149 3262 3271 3281 3297 3452 3462 3465 3524 3552 3559 3571 3599 3674 3679 3711 3714 3721 3812 |
| adaLasso by SIC | 2, 4 | 2024 2032 2035 2046 2211 2252 2253 2298 2299 2341 2395 2452 2599 2672 2675 2721 2813 2842 2992 3083 3231 3241 3251 3269 3296 3312 3356 3365 3452 3494 3511 3536 3541 3543 3544 3571 3593 3612 3624 3644 3651 3661 3669 3679 3711 3721 3873 3991 |
| CIM | 2, 6 | 2024 2032 2035 2046 2211 2252 2253 2298 2299 2341 2395 2452 2599 2672 2675 2721 2813 2842 2992 3083 3231 3241 3251 3269 3296 3312 3356 3365 3452 3494 3511 3536 3541 3543 3544 3571 3593 3612 3624 3644 3651 3661 3669 3679 |
| Migration | 3, A: 2-4 | Canada, Mexico, Eastern Europe, Philippines, Vietnam, India, South America, England, Belgium, Netherlands, Greece, Italy, Spain, Other USSR and Russia, Thailand, Lebanon, Saudi Arabia, Other |
| adaLasso, employment | 3, B: 2-4 | Canada, Mexico, Eastern Europe, Japan, Korea, Philippines, Scotland, Ireland, Belgium, Greece, Italy, Portugal, Other USSR and Russia, Thailand, Israel/Palestine, Saudi Arabia, Other |
| adaLasso, high-skilled wages | 3, C: 2-4 | Canada, Mexico, Eastern Europe, Japan, Scotland, Italy, Portugal, Other USSR and Russia, Other |
| adaLasso, low-skilled | 3, C: 2-4 | Canada, Mexico, Eastern Europe, Japan, Scotland, Italy, Portugal, Other USSR and Russia, Other |
| CI, employment | 4, A: 2-4 | Canada, Mexico, Scandinavia, Eastern Europe, Japan, Philippines, India, Oceania, Cuba and the West Indies, Oceania, Belgium, Netherlands, Greece, Italy, Other USSR and Russia, Malaysia, Afghanistan, Iran, Maldives, Nepal, Gulf States, Cyprus, Iraq, Israel/Palestine, Jordan, Lebanon, Saudi Arabia, Turkey, Other |
| CI, high-skilled wages | 4, B: 2-4 | Canada, Mexico, Japan, Philippines, Oceania, England, Scotland, Wales, Ireland, Belgium, Greece, Italy, Portugal, Spain, Austria, Other USSR and Russia, Israel/Palestine, Saudi Arabia, Other |
| CI, low-skilled wages | 4, C: 2-4 | Canada, Mexico, Japan, Cuba and West Indies, South America, Scotland, Belgium, Switzerland, Italy, Portugal, Spain, Other USSR and Russia, Afghanistan, Maldives, Iraq, Israel/Palestine, Other |

The class of industries that figures as the one with the second-highest number of invalid IVs is stone, clay, glass and concrete products (SIC32). This is the second of two industries about which ADH express concern. The coefficient from using only SIC32 industries is above zero. Omitting the selected IVs from the construction of the shift-share IV drives the coefficient back to about -0.6.

Malt beverages (such as beer) and ice cream are also chosen as invalid. Correlated demand shocks might be due to heat waves or major sport events which have affected both the US and other high-income countries. Using the entire food and kindred products industry (SIC20) yields a stronger coefficient with a larger confidence interval. Omitting these two sub-industries from the instrument again drives the estimate back to the original coefficient and significantly narrows down the confidence intervals.

Even though the 2-digit SIC industry in which invalid IVs are found could be roughly anticipated, based on the concerns of ADH it was not possible to foresee the exact identity of the invalid IVs. This is only natural, because a typical researcher does not have detailed information about the correlation structure of demand shocks in all 397 4-digit industries.
One might worry, that invalid instruments are detected only when running the analysis using each two-digit cluster at a time. Taking account of this, reconstructing the shift-share instrument and excluding all shift-share products selected as invalid in the two-digit analyses yields a coefficient of -0.471 (column 4), which means a decrease by 21%. Even though the decrease is now somewhat larger, the confidence interval still includes the original estimate.

**Confidence interval method** When applying the confidence interval method, the estimates are still very robust to omitting shift-share products chosen as invalid. Using the value of $\psi$ proposed in Guo, Kang, Cai, and Small (2018) ($\sqrt{20.012\ln(1444)} = 5.4215$), all confidence intervals overlap and as before, the Hansen-Sargan test of overidentifying restrictions is not rejected in the beginning. The post-selection shift-share estimator is identical to the initial one (column 5).

Using a significance level of 0.05 for the test of overidentifying restrictions leads to the classification of 48 shift-share products as invalid. Reconstructing the shift-share IV and estimating the model with invalid products included as controls yields a coefficient of -0.718 (column 6), which is slightly larger but still close to the original estimate.

Information and communication technology related (SIC35 and SIC36, 14 times) and construction-related (SIC32, five times) industries are still the ones which are chosen as invalid most often, in harmony with the explanations proposed in the preceding paragraphs. Now, also five sub-industries from the textile mill products industry (SIC22) are chosen as invalid. This is the third industry ADH have worried about, because of China’s dominant role as an exporter. Interestingly, three of the five industries with the highest sensitivity-to-misspecification weights shown in Goldsmith-Pinkham, Sorkin, and Swift (2018) have been chosen as invalid: telephone apparatus, household audio and video and again electronic computers. The larger number of industries chosen as invalid may also be explained by the potential overselection problem of the confidence interval method present in settings in which the majority rule holds, as seen in the simulation study.

### 5.3. Example 3: Effect of immigration on labor market outcomes

**Approach** The third empirical application is the one used as a motivating example in the introduction: estimating the effect of immigration on local labor market outcomes in the United States. I follow Adão, Kolesár, and Morales (2018) who estimate the linear model

$$\Delta Y_{oit} = \beta \Delta ImmShare_{oit} + Z_{oit}\delta + \varepsilon_{oit},$$

(14)

with 50 occupations $o$, three time periods $t$ (1990, 2000, 2010) and 722 commuting zones $i$. $\Delta Y_{oit}$ is the change in labor market outcome. The three outcomes used are change in log native employment, as well as change in average log weekly wages of high- and low-skilled workers. $\Delta ImmShare_{oit}$ is the change in share of immigrants in total employment. $Z_{oit}$ includes occupation and year fixed effects and $\varepsilon_{oit}$ is an error term. The authors use data from the Census Integrated Public Use Micro Samples and the American Community Survey.\(^{14}\)

\(^{14}\)The data was kindly provided by the authors.
Figure 3: AdaLasso by SIC2-level

Note: Comparison of standard shift-share IV and post-WFDS shift-share IV by SIC class. Standard errors are calculated with the procedure proposed by BHJ and clustered on SIC3 level. Standard shift-share estimate by SIC2-level in blue, adjusted shift-share in green. The red line denotes the original estimate.

Estimating this simple model by OLS does not account for migrant sorting into regions, creating a positive correlation between migrant location and labor-market outcome, which cannot be accounted to the impact of immigration. To tackle this problem, a shift-share instrument which uses origin-specific migrant shares in 1980 and changes in migrant populations is used. The shift-share instrument is hence defined as

$$\Delta X_{itd} \equiv \sum_{j=1}^{57} \text{ImmShare}_{oi1980,j} \frac{\text{Imm}_{t,j} - \text{Imm}_{t-10,j}}{\text{Imm}_{1980,j}},$$

where 57 countries of origin are used.

I reproduce the results of Adão, Kolesár, and Morales (2018) in the first column of Table 3. The effect on the change in log native employment is -0.74, which is statistically significant ($p < 0.01$). The effect on change in average log weekly wage of the high-skilled is at 0.144. The equivalent coefficient of the effect on low-skilled wage is -0.239. Both coefficients are not statistically significant. The original coefficients would suggest negative effects on employment and no effects on wages.

Instrument endogeneity issues One obvious concern would be that migrants in the past and present migrated to certain regions because of their sound economic conditions and not only because of the migrant network. Jaeger, Ruist, and Stuhler (2018) point out that in migration applications the shift-share instrument might fail to meet the exclusion restriction because of a direct effect of $\text{ImmShare}_{oi1980,j}$ used in the instrument on the outcome variable through adjustment processes of capital and native labor. Finally, serial correlation of unobservable shocks might lead to inconsistency when past shocks are correlated with the share used in the
Table 3: Migration application - adaLasso shift-share

|       | Panel A: Change in log native employment | Panel B: Change in avg. log weekly wage, high-skilled | Panel C: Change in avg. log weekly wage, low-skilled |
|-------|------------------------------------------|------------------------------------------------------|-----------------------------------------------------|
| Δ ImShare | -0.739** (-3.28) | -1.176*** (-7.20) | -0.294 (-0.58) | -1.610*** (-3.68) |
| F     | 61.71 | 80.22 | 12.07 | 49.25 |
| Nr inv | 0 | 18 | 18 | 18 |
| Δ ImShare | -0.0418 (-1.24) | -0.0168 (-0.22) | 0.0706 (0.08) |
| F     | 61.71 | 119.7 | 61.78 | 102.8 |
| Nr inv | 0 | 9 | 9 | 9 |

T-statistics in parentheses, significance level for HS-test: 0.00863
+ p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001

Borjas (2003) notes that the spatial approach is likely to be downward biased, due to adjustment processes. The actual effect of migration on employment should in fact be larger and negative. If migrants choose to migrate to regions with persistently low unemployment and native workers choose to migrate in response to the immigration of foreign workers, then the employment-decreasing effect of immigration could well be biased towards zero. For high-skilled wages, however, the story is not so clear. Tech companies might choose to move as well, if they see a surge in low-skilled and a drain of high-skilled labor. Depending on whether employment opportunities vanish more quickly than high-skilled competitors for jobs, the long-term effect of the instrument on wages of the high-skilled could go either way. This might be less of a problem for low-skilled labor if native low-skilled workers are less mobile.

Are there origin countries which are a priori suspect of not fulfilling the exclusion restriction? Card (2009) raises the issue that if the instrument mainly relies on migrants from one country, it could pick up local conditions of locations to which certain groups traditionally migrate. That would be the case especially for Mexican migrants who by far constitute the largest group of immigrants. The second-largest source country is the Philippines.

Interestingly, Mexico is attributed by far the largest Rotemberg-weight in Goldsmith-Pinkham, Sorkin, and Swift (2018), for low-skilled wages, while for high-skilled wages the Philippines receive the highest weight, followed by Mexico. This reflects the fact that Mexican immigrants are poorly educated and migrants from the Philippines are better-educated than natives (Card, 2009).

Jaeger, Ruist, and Stuhler (2018) also note that the measured effect might collapse the effect of longer and shorter lags when shares are serially correlated. The authors propose a multiple instrumentation approach to distinguish the different effects. In my application, the latter concern is not addressed. To also disentangle the time structure of the effect of immigration, the selection procedures would need to allow for multiple endogenous regressors.
Table 4: Migration application - CIM shift-share

|                  | (1)     | (2)     | (3)     | (4)     |
|------------------|---------|---------|---------|---------|
| **Panel A**: Change in log native employment |         |         |         |         |
| Δ ImmShare       | -0.739* | -1.079**| -1.573  | -1.996  |
|                  | (-3.28) | (-5.46) | (-0.55) | (-1.48) |
| F                | 61.71   | 45.25   | 1.076   | 13.61   |
| Nr inv           | 0       | 31      | 31      | 31      |

| **Panel B**: Change in avg. log weekly wage, high-skilled |         |         |         |         |
| Δ ImmShare       | 0.144   | 0.342*  | 0.209   | 0.465*  |
|                  | (1.02)  | (1.87)  | (0.77)  | (2.02)  |
| F                | 61.71   | 104.1   | 27.15   | 82.12   |
| Nr inv           | 0       | 19      | 19      | 19      |

| **Panel C**: Change in avg. log weekly wage, low-skilled |         |         |         |         |
| Δ ImmShare       | -0.239  | -0.134  | -0.130  | 0.0276  |
|                  | (-1.24) | (-0.61) | (-0.26) | (0.07)  |
| F                | 61.71   | 79.25   | 18.63   | 45.27   |
| Nr inv           | 0       | 17      | 17      | 17      |

T-statistics in parentheses, significance level for HS-test: 0.00863
+ p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001

The sensitivity-to-misspecification weights provide a diagnostic of how the invalidity of one origin country affects the consistency of the entire shift-share instrument. However, the weights do not tell us which of the origin countries are invalid and how large the effect is without bias.

**Adaptive Lasso** The result of adaLasso shift-share estimation can be found in table 3. When estimating the effect on employment, 18 countries are chosen as invalid (Panel A). Reconstructing the shift-share instrument with the rest of the products only, without controlling for endogenous shift-share products increases the effect on employment by 60% (column 2). Including the origin-country-specific products as controls makes the coefficient insignificant (column 3). This might be due to an increase in variance, through a loss of degrees of freedom. The problem becomes visible in the sharp decline of the first-stage F-statistic as compared to the case without controls. As a midway, in the fourth column, the shift-share products chosen as invalid are collapsed to a shift-share variable, which controls for the endogenous products. This is equivalent to restricting all shift-share products in the controls to have the same coefficient. This indeed leads to more moderate decreases in the F-statistic. Now, the coefficient has more than doubled and is still highly significant.

When using the wages of high-skilled workers as outcome variable, 17 countries are chosen as invalid. Reconstructing the shift-share IV only with IVs chosen as valid strongly increases the effect by about 140% (Panel B, column 2). The effect is now also marginally significant with a p-value of about 0.07. Including the invalid shift-share products as controls still leads to a strong increase of the coefficient, but statistical significance is lost (column 3). When using the aggregate shift-share control in column 4, the coefficient more than triples (p = 0.052).

For the weekly wages of the low-skilled (Panel C), nine countries are chosen as invalid. The effect was insignificant in the original estimation and stays insignificant with the adjusted method, no matter whether controls are used. Moreover, the coefficients move towards zero. The low
number of invalid IVs is consistent with the idea that low-skilled workers are less mobile and hence adjustment processes of labor are less of a problem for the identification strategy.

Since the number of countries chosen as endogenous is relatively high, I do not run additional robustness checks where the significance level is set even higher. Because the number of shift-share products is more moderate than in the China Shock application, the problem of too many instruments and an underpowered Hansen-Sargan test do not seem to be relevant here.

In unreported regressions, I perform the HS-test with the aggregate shift-share control. This leads to a rejection, and hence these results should be read with a grain of salt.\textsuperscript{16}

**Confidence interval method** Next, I use the confidence interval method to select invalid IVs. The results can be found in Table 4. Now, 31 countries of origin are chosen as invalid. This increases the coefficient to -1.08 when not controlling for invalid products (column 2). Controlling for invalid products leads to an increase of the coefficient of more than 100\% (column 3), but the estimate is not statistically significant anymore. Using the aggregate control leads to a coefficient of -1.996 (\(p = 0.1388\)), in column 4.

When estimating the impact on the wages of high-skilled, 19 countries are chosen as invalid. The results mirror those from the selection with adaLasso. The coefficient using the midway is now significant at the 0.05 level. With regard to low-skilled wages, 17 countries are chosen instead of nine, as when using adaLasso, but qualitatively the result do not change: The coefficients become smaller and are always insignificant.

When selecting with CIM, more products are chosen as invalid. In the case of unemployment, more than 50\% are chosen. Even though it might well be that the overselection problem illustrated in the simulations plays a role here, these results cast severe doubt on the validity of the shift-share strategy in the migration setting. It may well be that of the origin countries commonly used, only a minority is valid.

**5.4. Identity of invalid countries**

The countries chosen as invalid are listed in Table 2. There are a few origin countries which are chosen as invalid very often and catch the eye: Canada, Mexico, Eastern Europe, Philippines, Other USSR and Russia and the category “other countries”. Mexico and the Philippines are the countries which have received the highest sensitivity-to-misspecification weights in Goldsmith-Pinkham, Sorkin, and Swift (2018). The Philippines are chosen as invalid for employment and high-skilled wages, but not for low-skilled wages. This makes sense in light of the fact that immigrants from the Philippines are higher-skilled than native workers.

The choice as invalid of these countries is consistent with economic intuition. Canadian and Mexican migrants settled mostly in border regions. Mexicans settled mostly in Texas and California. California’s economy had and still has a large agricultural sector, and both states are among the wealthiest in the US. Therefore, it is plausible that migrants migrated for exactly

\textsuperscript{16}In principle, one could choose to restrict the coefficient of controls only for a subset of instruments and allow individual coefficients for the remaining ones, but this means to having to check the square of the number of controls models, which becomes computationally infeasible very quickly.
the same reasons in the 1970s as in more recent decades: proximity to the border and economic conditions and not necessarily the previous settlement of compatriots.

During the emigration of Ashkenazi Jews from the Soviet Union in the 1970s and the Post-Soviet countries in the 1990s, hundreds of thousands chose to emigrate to the United States. The new settlers predominantly chose coastal cities. These were cities which had large Jewish communities, but also cities which had experienced lasting prosperity. This makes a violation of the exclusion restriction likely, for Russian and other post-Soviet countries.

In the category “other countries”, all other countries are subsumed. It is not difficult to imagine that any of the above reasons or long-term adjustment effects apply to one of the remaining countries subsumed in this category. Because the algorithm can not distinguish between the countries further, it selects the entire shift-share product as invalid.

For all outcome variables, there is a large overlap of origin countries chosen as invalid by both methods. Many of these, such as Belgium, Netherlands, Italy, Spain or Saudi Arabia were not suspected of violating the exclusion restriction a priori. This illustrates that the proposed methods can enhance shift-share estimation beyond robustness checks which rely on intuition.

5.5. Summary of results

Overall, the application of the adjusted shift-share IV to the estimation of employment effects of Chinese import competition and immigration suggests that the proposed methods indeed select groups which were suspected of invalidity. The findings are relatively robust to the omission of invalid industries in the example of ADH.

In the migration setting, a large number of origin countries is chosen as invalid and coefficients change qualitatively. The insignificant results in columns 3 for employment and high-skilled wages might be due to the fact that indeed no effect is present, or that the variance estimate increases because of a loss of degrees of freedom. The strong increase of coefficients when using the shift-share control in columns 4 provides some evidence that the coefficient is in fact higher than suggested by the standard estimate. This would mean that migration has more extreme effects: stronger adverse effects on employment and stronger beneficial effects on wages of the high-skilled. The stronger negative effects are consistent with what Borjas (2003) expects. The presence of shift-share products which violate the exclusion restriction could be an explanation for the only moderate negative estimates in standard shift-share estimation.

Notably, not only products from the industries which were suspected of invalidity are chosen as invalid and not the entire suspected industries are chosen. Therefore, the methods yield tangible benefits, since without these methods, exclusion of products is subject to the discretion of researchers.

Why do the results suggest endogeneity of many shift-share products in the migration setting, while the problem is less acute in the application of ADH? Apart from the reasons discussed above, a further reason could be that the shifts used in ADH are imports to other high-income countries, while migration to the US is measured both in the endogenous variable and in the instrument. The migration literature relies strongly on the exogeneity of shares and no attempt to exogenize the shifts is undertaken. One way to attenuate this problem could be to motivate
the exclusion restriction through quasi-randomness of shifts and to use country-of-origin specific push factors related to war, civil liberties or natural disasters. This kind of approach has been used by Llull (2017) among others.

6. Conclusion

In this paper I presented a method that addresses the problem of endogenous shift-share instruments. I proposed an adjusted estimator which uses the adaptive Lasso (Windmeijer, Farbmacher, Davies, and Smith, 2018) and the confidence interval method (Windmeijer, Liang, Hartwig, and Bowden, 2019) to select instruments as invalid and exclude them from the construction of the adjusted shift-share IV. In the China shock example, the methods choose a few industries as invalid, but do not yield large changes in practical significance, suggesting a low inconsistency of the unadjusted estimator. In the migration setting, however, many products are chosen as invalid. The adjusted estimator yields stronger effects for employment and wages of the high skilled. Hence, the migration setting doesn’t seem well suited for the standard approach and adjusting the estimator via selection of invalid shift-share products makes a qualitative difference.

Recent methodological literature discusses ways in which the exclusion restriction in shift-share estimation might be fulfilled (Borusyak, Hull, and Jaravel, 2018; Goldsmith-Pinkham, Sorkin, and Swift, 2018), ways in which it might be violated (Jaeger, Ruist, and Stuhler, 2018), diagnostics to analyze which instruments are prone to lead to a large bias of the estimator (Goldsmith-Pinkham, Sorkin, and Swift, 2018) and improved inference (Adão, Kolesár, and Morales, 2018). I contribute to this literature by applying newly developed methods to substantially relax the exclusion restriction, requiring that only a majority or a plurality of products needs to be valid. The presented methods might be regarded as complementary to the ones developed in the recent literature. Before using these methods, it is important to carefully think about the nature of the exclusion restriction and which source of exogeneity is most feasible, as suggested by Goldsmith-Pinkham, Sorkin, and Swift (2018). I suggest using the adjusted method whenever some shares or shifts are suspected to be directly correlated with the outcome variable. The ado-files in the supplementary material offer a simple way to apply the proposed methods.

The methods applied here also allow for IVs associated with local average treatment effects but discard the information contained in them. Moreover, the methods cannot accommodate multiple endogenous regressors. These could be promising avenues for future research.
Appendices

A. Methodological appendix

A.1. Adaptive LASSO Details

In the following I provide some more details on the adaptive Lasso.

The $L_2$-norm is denoted by $||.||_2$. The projection matrices are $P_Z \equiv Z'Z^{-1}Z'$ and $M_Z \equiv I - P_Z$. The moment conditions $E[Z'\varepsilon] = 0$ can be rewritten to

$$\Gamma = \alpha + \gamma\beta,$$

where $\Gamma = E[Z'Z]^{-1}E[Z'y]$ and $\gamma = E[Z'Z]^{-1}E[Z'd]$.

From the moment conditions in equation (16), WFDS define a vector $\pi$ with elements $\pi_j \equiv \frac{\Gamma_j}{\gamma_j}$, where subscript $j$ denotes the $j$-th element of the vector. The sample equivalent of this is $\hat{\pi}_j = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j}$. From the median of the vector $\hat{\pi}$, $\beta_m \equiv med(\hat{\pi})$ one can yield a consistent estimator of $\hat{\alpha}_m = \hat{\Gamma} - \hat{\gamma}\hat{\beta}_m$. The adaLasso as proposed in Zou (2006) can then be used, where the initial consistent estimate is given by $\hat{\alpha}_m$.

The adaLasso minimization problem is

$$\hat{\alpha}^{\lambda}_{ad} = \arg\min_\alpha \frac{1}{2}||y - \tilde{Z}\alpha||_2^2 + \lambda_n \sum_{j=1}^J |\alpha_j|/\hat{\pi}_j,$$

where $\tilde{Z} = M_D Z$ and $\hat{d}$ is the linear projection of $d$ on $Z$. The adaptive Lasso estimator of $\beta$ is then retrieved by

$$\hat{\beta}^{\lambda}_{ad} = \frac{\hat{d}(y - Z\hat{\alpha}^{\lambda}_{ad})}{\hat{d}'\hat{d}}.$$

In summary, the estimation procedure works as follows:

1. Compile the vector $\hat{\pi}$
2. Take its median $\hat{\beta}_m$
3. Calculate $\hat{\alpha}_m$
4. Estimate $\alpha^{\lambda}_{ad}$ by adaptive Lasso
5. Calculate $\hat{\beta}^{\lambda}_{ad}$
6. Post-adaLasso 2SLS: 2SLS regression with the instruments chosen as invalid included as controls in the structural equation and those chosen as valid used as IVs. In this application, the post-adaLasso estimation is the just-identified shift-share IV.
A.2. LATE Plurality example

An example should help clarify the setting from section 4.3. Assume there are $H = 3$ treatment effects and 20 IVs. The treatment effects are $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = -2$. There are $Q + 1 = 7$ groups $G_q$ (including $\mathcal{V}$).

The groups are structured as follows:

| Group | # of elements | $\phi_q$       |
|-------|---------------|----------------|
| $\mathcal{V}$ | 6             | $\beta_0 = 0$ |
| $G_1$  | 2             | $\phi_1 = \beta_0 + 2 = 2$ |
| $G_2$  | 1             | $\phi_2 = \beta_0 + 3 = 3$ |
| $G_3$  | 3             | $\phi_3 = \beta_1 = 1$ |
| $G_4$  | 4             | $\phi_4 = \beta_1 + 5 = 6$ |
| $G_5$  | 2             | $\phi_5 = \beta_2 = -2$ |
| $G_6$  | 2             | $\phi_6 = \beta_2 - 3 = -5$ |

The distribution of LATEs is visualized below. The IVs which are selected as valid by the CIM asymptotically are those associated with $\beta_0$. Here, IVs associated with $\beta_0$ make up the largest group of IVs.
A.3. Documentation for ado-files

The following subsection provides the documentation for the ssada- and sscim- programs in Stata.

**Preliminaries:** Save ssada and sscim to your personal ado-directory.

A.3.1. AdaLasso shift-share

The Stata implementation of adaLasso shift-share is called ssada. The code is a variation of sivreg (Farbmacher, 2017) and shares its syntax. The differences are that in ssada analytical weights are allowed, the adjusted shift-share instrument is created and the post-estimation regression is a shift-share regression instead of 2SLS. Moreover, the standard errors to be reported in the post-adaLasso regression can be chosen and locals containing valid and invalid IVs are returned. As sivreg, ssada also requires moremata.

**Syntax**

```
ssada depvar indepvars [if] [in] [aw], ///
endog(varlist) exog(varlist) id(string) [options]
```

**Options**

**Required:**

- `endog` Endogenous variable
- `exog` Exogenous controls as well as potentially endogenous single products used for construction of the shift-share IV. The shift-share products should have the following naming: e.g. stub1, stub2, stub3, ...
- `id` String denoting variables by which observations are identified

**Optional:**

- `aw` Only analytical weights (aweight) are allowed
- `vce` Specifies the type of standard error reported. Same as in standard vce-option. Default is robust
- `c` real specifying the significance level as \( c/\ln(n) \) for the Andrews-Hansen stopping rule. Default is 0.1

**Stored results** ssada stores the results of the last post-adaLasso ivregress-command in `e()`. Moreover, the following macros are returned:

- `e(wv)` A local containing the varnames of variables chosen as valid by the adaLasso algorithm
- `e(wi)` A local containing the varnames of variables chosen as invalid by the adaLasso algorithm
A.3.2. Confidence Interval Method shift-share

The setup of sscim is adopted from sivreg (Farbmacher, 2017, a large part of lines 1-150).

Syntax

sscim depvar indepvars [if] [in] [aw], ///
endog(varlist) exog(varlist) ssstub(string) [options]

Options

Required:
endog Endogenous variable
exog Exogenous controls as well as potentially endogenous single products used for construction of the shift-share IV
ssstub Stub of shift-share products. The shift-share products should have the following naming: e.g. stub1, stub2, stub3, ...

Optional:
aw Only analytical weights (aweight) are allowed
vce Specifies the type of standard error reported. Same as in standard vce-option. Default is robust
c real specifying the significance level as $c/\ln(n)$ for the Andrews-Hansen stopping rule. Default is 0.1
step Specifies the fraction by which the critical value is shrunk at each step in the confidence interval method. Default is 0.99
psif Specifies initial critical value with which confidence intervals are calculated, according to $\psi = \text{psif} \times \sqrt{2.01^2 \times \ln(N)}$. Set this larger than one if in the beginning already more than one IV is chosen as invalid. Default is 1.

Stored results sscim stores the results of the last post-CIM ivregress-command in e() and the following macros:

- e(vc) A local containing the varnames of variables chosen as valid by the confidence interval method
- e(ic) A local containing the varnames of variables chosen as invalid by the confidence interval method

Post-estimation

For both ssada and sscim, the same post-estimation results as for ivregress apply.

For calculation of corrected standard errors as in Adão, Kolesár, and Morales (2018), Borusyak, Hull, and Jaravel’s (2018) ssaggregate command can be used.
Example

drop _all
set seed 123

* CREATE DATA *
set obs 500
gen i = _n

* Structural equation error:
gen eps = rnormal(0,1)

* First-stage error:
gen ups = rnormal(0,1)
gen yps = 0.5*eps + ups

* Single products:
forv i = 1/10{
gen ssp'i' = rnormal(0,2)
}
gen D = ssp1 * 0.2 + ssp2 * 0.2 + ssp3 * 0.2 + ssp4 * 0.2 + ///
ssp5 * 0.2 + ssp6 * 0.2 + ssp7 * 0.2 + ssp8 * 0.2 + ///
ssp9 * 0.2 + ssp10 * 0.2 + ups

//Majority assumption fulfilled
gen Y1 = D*0 + 0.2 * ssp1 + 0.2 * ssp2 + 0.2 * ssp3 + 0 * ssp4 + 0 * ssp5 + ///
0 * ssp6 + 0 * ssp7 + 0 * ssp8 + ///
0 * ssp9 + 0 * ssp10 + eps
//Plurality assumption fulfilled
gen Y2 = D*0 + 0 * ssp1 + 0 * ssp2 + 0 * ssp3 + 0.2 * ssp4 + 0.4 * ssp5 + ///
0.6 * ssp6 + 0.8 * ssp7 + 1 * ssp8 + ///
1.2 * ssp9 + 1.5 * ssp10 + eps

* RUN PROGRAMS *
//ssp1, ssp2 and ssp3 should be chosen as invalid
ssada Y1 D, endog(D) exog(ssp*) id(i)
sscim Y1 D, endog(D) exog(ssp*) ssstub(ssp) psif(10) step(0.999)

//All except ssp1, ssp2 and ssp3 should be chosen as invalid
ssada Y2 D, endog(D) exog(ssp*) id(i)
sscim Y2 D, endog(D) exog(ssp*) ssstub(ssp) psif(10) step(0.999)
B. Figures

Figure 4: Comparison of standard shift-share-IV and post-WFDS shift-share-IV by SIC class. Standard errors are calculated with the procedure proposed by BHJ and clustered on SIC3 level. Significance level of HS-test has been set to 0.05.

Figure 5: Comparison of standard shift-share-IV estimates with and without Adão, Kolesár, and Morales (2018) standard error correction.
Figure 6: MAD of adjusted shift-share IV. Row 1: invalid IVs not chosen as controls. Row 2: invalid IVs collapsed to shift-share control.
References

Adão, R., M. Kolesár, and E. Morales (2018). Shift-Share Designs: Theory and Inference. *arXiv preprint arXiv:1806.07928*.

Andrews, D. W. (1999). Consistent Moment Selection Procedures for Generalized Method of Moments Estimation. *Econometrica 67*(3), 543–563.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review 103*(6), 2121–68.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain. *Econometrica 80*(6), 2369–2429.

Borjas, G. J. (1999). The Economic Analysis of Immigration. In *Handbook of Labor Economics*, Volume 3, pp. 1697–1760. Elsevier.

Borjas, G. J. (2003). The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *The Quarterly Journal of Economics 118*(4), 1335–1374.

Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-Experimental Shift-Share Research Designs. *arXiv preprint arXiv:1806.01221*.

Bowsher, C. G. (2002). On Testing Overidentifying Restrictions in Dynamic Panel Data Models. *Economics Letters 77*(2), 211–220.

Caner, M., X. Han, and Y. Lee (2018). Adaptive Elastic Net GMM Estimation with Many Invalid Moment Conditions: Simultaneous Model and Moment Selection. *Journal of Business & Economic Statistics 36*(1), 24–46.

Card, D. (2009). Immigration and inequality. *American Economic Review 99*(2), 1–21.

Farbmacher, H. (2017). Sivreg: Stata Module to Perform Adaptive Lasso with Some Invalid Instruments.

Gautier, E., A. Tsybakov, and C. Rose (2011). High-Dimensional Instrumental Variables Regression and Confidence Sets. *arXiv preprint arXiv:1105.2454*.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018). Bartik Instruments: What, When, Why, and How. Technical report, National Bureau of Economic Research.

Guo, Z., H. Kang, T. T. Cai, and D. S. Small (2018). Confidence Intervals for Causal Effects with Invalid Instruments by Using Two-Stage Hard Thresholding with Voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Han, C. (2008). Detecting Invalid Instruments Using L1-GMM. *Economics Letters 101*(3), 285–287.
Jaeger, D. A., J. Ruist, and J. Stuhler (2018). Shift-Share Instruments and the Impact of Immigration. Technical report, National Bureau of Economic Research.

Kang, H., A. Zhang, T. T. Cai, and D. S. Small (2016). Instrumental Variables Estimation with Some Invalid Instruments and Its Application to Mendelian Randomization. *Journal of the American Statistical Association* 111(513), 132–144.

Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2015). Identification and Inference with Many Invalid Instruments. *Journal of Business & Economic Statistics* 33(4), 474–484.

Llull, J. (2017). The Effect of Immigration on Wages: Exploiting Exogenous Variation at the National Level. *Journal of Human Resources*, 0315–7032R2.

Mullainathan, S. and J. Spiess (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives* 31(2), 87–106.

Roodman, D. (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and statistics* 71(1), 135–158.

Windmeijer, F., H. Farbmacher, N. Davies, and G. D. Smith (2018). On the Use of the LASSO for Instrumental Variables Estimation with Some Invalid Instruments. *Journal of the American Statistical Association* (just-accepted), 1–32.

Windmeijer, F., X. Liang, F. Hartwig, and J. Bowden (2018). The Confidence Interval Method for Selecting Valid Instruments. Technical report, University of Bristol.

Windmeijer, F., X. Liang, F. P. Hartwig, and J. Bowden (2019). The confidence interval method for selecting valid instrumental variables. Technical report, Department of Economics, University of Bristol, UK.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101(476), 1418–1429.