UNIVERSITÀ DI SIENA 1240

QUADERNI DEL DIPARTIMENTO

DI ECONOMIA POLITICA E STATISTICA

Federico Crudu

Errors-in-Variables Models with Many Proxies

n. 774 – Gennaio 2018

Errors-in-Variables Models with Many Proxies

Federico Crudu*†

University of Siena and CRENoS

January 2018

Abstract

This paper introduces a novel method to estimate linear models when explanatory variables are observed with error and many proxies are available. The empirical Euclidean likelihood principle is used to combine the information that comes from the various mismeasured variables. We show that the proposed estimator is consistent and asymptotically normal. In a Monte Carlo study we show that our method is able to efficiently use the information in the available proxies, both in terms of precision of the estimator and in terms of statistical power. An application to the effect of police on crime suggests that measurement errors in the police variable induce substantial attenuation bias. Our approach, on the other hand, yields large estimates in absolute value with high precision, in accordance with the results put forward by the recent literature.

Key words: data combination, empirical Euclidean likelihood, errors-in-variables, instrumental variables.
JEL classification: C13, C26, C30, C36.

*Deparment of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, federico.crudu@unisi.it.

†I am grateful to Emilio Porcu, Tiziano Razzolini, Giovanni Mellace, Zsolt Sándor and Alessia Paccagnini for helpful comments and suggestions.

1 Introduction

It is well known that the ordinary least squares (OLS) estimator is biased and inconsistent when one or more regressors are measured with error. A textbook solution for the so called measurement error problem is the use of instrumental variables (IV). When two mismeasured variables or proxies are available, it is possible to use the second proxy as an instrument for the first. Relevant papers in this context are Ashenfelter & Krueger (1994), Borjas (1995), Barron et al. (1997). Ashenfelter & Krueger (1994) measure the effect of schooling on wages on a sample of twins. The variable that approximates schooling is measured as the difference between the self-reported schooling levels. The second proxy variable is the difference of the sibling's reported schooling differences (see Krueger & Lindahl, 2001, for a survey on the effects of measurement errors on education data). Borjas (1995) studies the link between ethnic externalities and ethnic neighborhoods. In this context, self-reported parental skills induce a considerable downward bias in the correlation between earnings of fathers and earning of sons. The presence of measurement errors is taken into account by using as an instrument the average skills of the father as reported by other siblings. Barron et al. (1997) study the effect of on-the-job training on the determination of wages. In this case, establishment-reported training is a natural instrument for worker-reported training.

Clearly, if two proxies are available, it is possible to build two consistent IV estimators. Interestingly, the literature has recently put forward methods that efficiently use the information coming from two or multiple proxies, see for example Lubotsky & Wittenberg (2006), Andersson & Møen (2016) and Chalfin & McCrary (2017). Andersson & Møen (2016) propose a clever way to optimally combine IV estimators when two proxy variables are available. Chalfin & McCrary (2017), in a conceptually similar framework, solve the measurement error problem by means of a generalized method of moments (GMM) approach. The key difference between the two approaches is that the latter defines a set of moment conditions where the parameters of interest are constrained to be the same and are simultaneously estimated. The former method, on the other hand, estimates two different models. The estimated parameters are then linearly combined.

Often, the unobserved explanatory variable can be proxied by several mismeasured variables. In this context, Lubotsky & Wittenberg (2006) propose a method to combine the available proxies in a single indicator. Other related papers are, e.g., Bernal et al. (2016) and Black & Smith (2006), who study the effect of school quality when multiple proxies for quality are available. Fryer et al. (2013) build an index for crack cocaine based on a number of indirect measures such as cocaine arrests, cocaine-related emergency room visits, cocaine induced drug deaths, crack mentions in newspapers and drug busts. In an interesting paper on the role of social capital on financial development in Italy, Guiso et al. (2004) use participation to referendums and blood donation as approximations to social capital.

As we have seen in the above examples, proxies are widely used in applied economics because they allow us to quantify variables that would be otherwise latent. However, if not carefully treated, proxies may deliver severely biased estimates. This paper proposes a novel method to estimate linear models in the presence of error-ridden covariates when a possibly large but fixed number of proxies is available. Intuitively, using all proxies should not only produce consistent estimates but also increase precision. We consider the classical errors-in-variables (EIV) framework with independent measurement errors. The main idea is to generate various sets of moment conditions where for each set of moments we build an objective function according to the empirical Euclidean likelihood (EEL) principle (Owen, 2001; Newey & Smith, 2004; Antoine et al., 2007). The estimator, called CombEEL, is obtained by optimizing the sum of such objective functions. This idea is similar to what Tsao & Wu (2006) propose to test for a common mean in the presence of many variables and heteroskedasticity (see also Tsao & Wu, 2015). Using standard asymptotic results it is possible to show that the proposed estimator is consistent and asymptotically normal. The numerical results indicate that the CombEEL estimator works very well, in terms of bias and mean square error, when the number of proxies is relatively large, even when the quality of the proxies deteriorates. We also find that the power of the t test increases proportionally with the number of proxies, suggesting that the estimator efficiently exploits the information contained in the mismeasured variables.

The outline of the paper is as follows. Section 2 introduces the classical EIV model, Section 3 describes the EEL estimation method and the corresponding asymptotic results, while Section 4 includes the Monte Carlo experiments and an application to the effect of police on crime. To provide further insight on the workings of the method, the two proxy case is illustrated throughout the paper. Finally, Section 5 concludes. Proofs and auxiliary results are deferred to the Appendix.

2 The Classical Errors-in-Variables Model

In what follows we describe the classical EIV model for our specific problem. Let us first introduce some notation: bold upper case letters indicate matrices, bold lower case letters indicate vectors, vectors are meant to be column vectors unless differently stated. Let us consider the model

$$\boldsymbol{y} = \boldsymbol{x}_1^*\delta + \boldsymbol{Z}_2\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{1}$$

where the n-vector $\boldsymbol{x}_1^*$ is a latent exogenous variable and can only be observed with error, δ is a scalar coefficient, $\boldsymbol{Z}_2$ is a $n \times K_2$ matrix of exogenous covariates, $\boldsymbol{\gamma}$ is its associated K_2-vector of parameters, while $\boldsymbol{\varepsilon}$ is zero mean and potentially heteroskedastic.[1] Let us now suppose that we observe a bunch of proxy variables $\boldsymbol{x}_1^{(\eta)} = \boldsymbol{x}_1^* + \boldsymbol{e}^{(\eta)}$ where $\eta = 1, \ldots, J$.[2] The measurement errors $\boldsymbol{e}^{(\eta)}$ are independent among themselves, they have zero mean but do not necessarily share the same variance. They are also independent of the structural error $\boldsymbol{\varepsilon}$ and the unobservable explanatory variable $\boldsymbol{x}_1^*$. So, if we use $\boldsymbol{x}^{(\eta)}$ as a regressor, we

[1] The more general scenario where the latent variable is, say, a $n \times K_1$ matrix is rather similar.

[2] Lubotsky & Wittenberg (2006) consider a more general definition of measurement error. In our notation, $\boldsymbol{x}_1^{(\eta)} = \rho\boldsymbol{x}_1^* + \boldsymbol{e}^{(\eta)}$. Our approach does not allow for this type of measurement error yet. This issue is currently under investigation.

can define

$$\boldsymbol{y} = \boldsymbol{x}_1^{(\eta)}\delta + \boldsymbol{Z}_2\boldsymbol{\gamma} + \boldsymbol{u}^{(\eta)} = \boldsymbol{X}^{(\eta)}\boldsymbol{\beta} + \boldsymbol{u}^{(\eta)}$$

where $\boldsymbol{u}^{(\eta)} = \boldsymbol{\varepsilon} - \boldsymbol{e}^{(\eta)}\delta$, $\boldsymbol{X}^{(\eta)} = \left(\boldsymbol{x}_1^{(\eta)}, \boldsymbol{Z}_2\right)$ is a $n \times (1 + K_2)$ matrix and $\boldsymbol{\beta} = (\delta, \boldsymbol{\gamma}^\top)^\top$. As suggested, among others, by Andersson & Møen (2016), if the measurement errors $\boldsymbol{e}^{(\eta)}$ are independent we may use the remaining proxies as instruments.

2.1 Moment Conditions

In this section we describe how to generate moment conditions when we have a potentially large number of proxies. Such moment conditions are the basis of the estimation method described in Section 3. Let us assume that $J \geq 2$ and define

$$\boldsymbol{Z}^{(\zeta)} = \left(\boldsymbol{x}_1^{(\zeta)}, \boldsymbol{Z}_2\right),$$

where $\zeta \in \{1, \dots, J\}\backslash\{\eta\}$ and $\eta \in \{1, \dots, J\}$ and denote $\boldsymbol{z}_i^{(\zeta)^\top}$ as the i-th row of $\boldsymbol{Z}^{(\zeta)}$. Accordingly, we define $\boldsymbol{x}_i^{(\eta)^\top}$ as the i-th row of $\boldsymbol{X}^{(\eta)}$ and $\boldsymbol{z}_{2i}^{(\zeta)^\top}$ as the i-th row of $\boldsymbol{Z}_2^{(\zeta)}$. Then, we can construct the following set of just-identified moment conditions

$$\mathbb{E}\left[\boldsymbol{z}_i^{(\zeta)}\left(y_i - x_{1i}^{(\eta)}\delta_0 - \boldsymbol{z}_{2i}^\top\boldsymbol{\gamma}_0\right)\right] = \mathbb{E}\left[\boldsymbol{z}_i^{(\zeta)}\left(y_i - \boldsymbol{x}_i^{(\eta)^\top}\boldsymbol{\beta}_0\right)\right] = \boldsymbol{0} \tag{2}$$

where $\boldsymbol{\beta}_0 = (\delta_0, \boldsymbol{\gamma}_0^\top)^\top$ is the true parameter and it is unique. The pair (η, ζ) takes values in the index set

$$\mathcal{A}(J) := \{(\eta, \zeta)|\eta \in \{1, \dots, J\}, \zeta \in \{1, \dots, J\}\backslash\{\eta\}\}. \tag{3}$$

Since the matrix $\boldsymbol{Z}^{(\zeta)}$ has $1 + K_2$ columns and the cardinality of the index set is $|\mathcal{A}(J)| = (J-1)J$, we get $K = (1 + K_2)(J-1)J$ moment conditions.[3] The moment conditions in Equation (2) are the same as the IV method of Andersson & Møen (2016).

2.2 The Two Proxy Case: Moment Conditions

Let us assume, as in Andersson & Møen (2016), that $J = 2$. This is, we have two proxy variables $\boldsymbol{x}^{(\eta)} = \boldsymbol{x}_1^* + \boldsymbol{e}^{(\eta)}$, $\eta = 1, 2$ and two models

$$\boldsymbol{y} = \boldsymbol{X}^{(\eta)}\boldsymbol{\beta} + \boldsymbol{u}^{(\eta)}, \quad \eta = 1, 2.$$

Estimating either model by OLS would produce biased and inconsistent estimates. However, we can easily see that

$$\boldsymbol{y} = \boldsymbol{x}_1^*\delta + \boldsymbol{Z}_2\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = (\boldsymbol{x}^{(\eta)} - \boldsymbol{e}^{(\eta)})\delta + \boldsymbol{Z}_2\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \boldsymbol{x}^{(\eta)}\delta + \boldsymbol{Z}_2\boldsymbol{\gamma} + \boldsymbol{u}^{(\eta)}, \quad \eta = 1, 2.$$

Clearly, $\boldsymbol{u}^{(1)} = \boldsymbol{\varepsilon} - \boldsymbol{e}^{(1)}\delta$ and $\boldsymbol{u}^{(2)} = \boldsymbol{\varepsilon} - \boldsymbol{e}^{(2)}\delta$. This implies that $\boldsymbol{x}^{(2)}$ is not correlated with $\boldsymbol{u}^{(1)}$ and, by the same line of reasoning, $\boldsymbol{x}^{(1)}$ is not correlated with $\boldsymbol{u}^{(2)}$. Furthermore, $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are correlated. Hence, $\boldsymbol{x}^{(2)}$ could serve as an instrument when $\boldsymbol{x}^{(1)}$ is used as a regressor, while $\boldsymbol{x}^{(1)}$ would be the instrument when $\boldsymbol{x}^{(2)}$ is used as a regressor. In the two proxy case, this approach would produce two estimators for δ, say $\widehat{\delta}^{(\eta)}$, $\eta = 1, 2$. Andersson & Møen (2016) suggest obtaining an estimator that uses the information from both proxies

[3] We could generate moment conditions in various ways. For example, let $\boldsymbol{z}_i^{(\eta)\top}$ be the i-th row of $\boldsymbol{Z}^{(\eta)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\eta-1)}, \boldsymbol{x}^{(\eta+1)}, \ldots, \boldsymbol{x}^{(J)}, \boldsymbol{Z}_2)$ and $\boldsymbol{z}_{2i}^\top$ be the i-th row of $\boldsymbol{Z}_2$. Then,

$$\mathbb{E}\left[\boldsymbol{z}_i^{(\eta)}\left(y_i - x_{1i}^{(\eta)}\delta_0 - \boldsymbol{z}_{2i}^\top\boldsymbol{\gamma}_0\right)\right] = \boldsymbol{0},$$

a set of overidentified moment conditions. We refrain from using this approach as a potentially high degree of overidentification may introduce substantial bias in the estimator. It would be though interesting to see what happens when the number of proxies grows proportionally with the sample size. In this case, one could invoke certain results in Newey & Windmeijer (2009).

through a convex combination of the type

$$\widehat{\delta}_c = c\widehat{\delta}^{(1)} + (1-c)\widehat{\delta}^{(2)}, \quad c \in (0,1).$$

The estimator is optimal in the sense that minimizes the variance of $\widehat{\delta}_c$ with respect to c.

3 The Empirical Euclidean Likelihood Approach

The objective function produced by the CombEEL is simple to define and it is very similar to the objective function that characterizes the GMM estimator. Nonetheless, there are some important differences, in particular when the number of proxies is large. The framework considered in this paper implies that a potentially large number of moments is included in the estimation problem. This presents two issues. The first issue is that the GMM estimator tends to perform poorly when the degree of overidentification is large. This problem could be easily fixed by using the empirical likelihood estimator or some other member of the generalized empirical likelihood family of estimators (Newey & Smith, 2004). The second issue is that there may be a high level of collinearity among moments when $J > 2$. An immediate consequence is that the (efficient) variance covariance matrix may be singular. This problem does not arise in the case of the CombEEL estimator, as shown in this Section.

For the sake of simplicity and to generalize the scope of our problem, let us define the following moment condition model

$$\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{w}_i^{(\eta,\zeta)}, \boldsymbol{\beta}_0\right)\right] = \mathbb{E}\left[\boldsymbol{g}_i^{(\eta,\zeta)}\left(\boldsymbol{\beta}_0\right)\right] = \mathbf{0} \tag{4}$$

where $\boldsymbol{w}_i^{(\eta,\zeta)}$ is a L-vector of data, $\boldsymbol{g} : \mathbb{R}^L \times \mathcal{B} \to \mathbb{R}^{1+K_2}$ and $\boldsymbol{\beta}_0 \in \mathcal{B} \subset \mathbb{R}^{1+K_2}$. The linear case described in Section 2 corresponds to $\boldsymbol{w}_i^{(\eta,\zeta)} = \left(y_i, x_{1i}^{(\eta)}, x_{1i}^{(\zeta)}, \boldsymbol{z}_{2i}^\top\right)^\top$ and $\boldsymbol{g}_i^{(\eta,\zeta)}\left(\boldsymbol{\beta}\right) = \boldsymbol{z}_i^{(\zeta)}\left(y_i - \boldsymbol{x}_i^{(\eta)^\top}\boldsymbol{\beta}\right)$. Following, for example, Newey & Smith (2004), we can define the EEL as a sum of quadratic functions (see also Antoine et al., 2007). Thus, for the generic pair

$(\eta, \zeta) \in \mathcal{A}(J)$, we find

$$R_n^{(\eta,\zeta)}(\boldsymbol{\beta}, \boldsymbol{\lambda}^{(\eta,\zeta)}) = \frac{1}{2} \sum_{i=1}^{n} \left(1 + \boldsymbol{\lambda}^{(\eta,\zeta)\top} \boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})\right)^2$$

where the auxiliary parameter vector $\boldsymbol{\lambda}^{(\eta,\zeta)}$ comes from the fact that the EEL can be derived from a Lagrangian optimization problem. Thus, the objective function that characterizes the CombEEL estimator is given by the sum of all the individual objective functions:

$$\begin{aligned} R_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_n^{(\eta,\zeta)}(\boldsymbol{\beta}, \boldsymbol{\lambda}^{(\eta,\zeta)}) = \sum_{(\eta,\zeta)\in\mathcal{A}(J)} \left(\frac{1}{2} \sum_{i=1}^{n} \left(1 + \boldsymbol{\lambda}^{(\eta,\zeta)\top} \boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})\right)^2 \right) \\ &= \frac{n}{2} \sum_{(\eta,\zeta)\in\mathcal{A}(J)} \left(1 + 2\boldsymbol{\lambda}^{(\eta,\zeta)\top} \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta}) + \boldsymbol{\lambda}^{(\eta,\zeta)\top} \widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta}) \boldsymbol{\lambda}^{(\eta,\zeta)}\right) \end{aligned} \tag{5}$$

where $\widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})$ and

$$\widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta}) \boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})^\top . \tag{6}$$

By taking the first order conditions of Equation (5) with respect to $\boldsymbol{\lambda}^{(\eta,\zeta)}$ we find

$$\widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)}(\boldsymbol{\beta}) = -\widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta})^{-1} \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta}),$$

which implies

$$\begin{aligned} R_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{n}{2} \sum_{(\eta,\zeta)\in\mathcal{A}(J)} \left(1 - \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta})^\top \widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta})^{-1} \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta})\right) \\ &= \frac{n}{2} \left(|\mathcal{A}(J)| - \widehat{\boldsymbol{g}}(\boldsymbol{\beta})^\top \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1} \widehat{\boldsymbol{g}}(\boldsymbol{\beta})\right) \\ &= \frac{n}{2} \left(|\mathcal{A}(J)| - Q_n(\boldsymbol{\beta})\right) \end{aligned}$$

where

$$Q_n(\boldsymbol{\beta}) = \widehat{\boldsymbol{g}}(\boldsymbol{\beta})^\top \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1} \widehat{\boldsymbol{g}}(\boldsymbol{\beta}) . \tag{7}$$

The K-vector of moment conditions $\widehat{\boldsymbol{g}}(\boldsymbol{\beta})$ is obtained by stacking $\widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta})$, $(\eta,\zeta) \in \mathcal{A}(J)$ and $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})$ is a $K \times K$ block diagonal matrix where its generic diagonal element is $\widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta})$. We can estimate the parameter vector $\boldsymbol{\beta}$ by minimizing the quadratic form $Q_n(\boldsymbol{\beta})$. Hence,

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} Q_n(\boldsymbol{\beta}) \tag{8}$$

is the CombEEL estimator for $\boldsymbol{\beta}$.

3.1 The Two Proxy Case: Objective Function

To clarify the mechanics of the CombEEL estimator, let us build the objective function for the simple two proxy case. The index set is $\mathcal{A}(2) = \{(1,2),(2,1)\}$ and clearly $|\mathcal{A}(2)| = 2$. The objective function is then defined as

$$\begin{aligned} R_n(\boldsymbol{\beta},\boldsymbol{\lambda}) &= \frac{n}{2}\left(|\mathcal{A}(2)| - \widehat{\boldsymbol{g}}^{(1,2)}(\boldsymbol{\beta})^\top \widehat{\boldsymbol{\Omega}}^{(1,2)}(\boldsymbol{\beta})^{-1}\widehat{\boldsymbol{g}}^{(1,2)}(\boldsymbol{\beta}) - \widehat{\boldsymbol{g}}^{(2,1)}(\boldsymbol{\beta})^\top \widehat{\boldsymbol{\Omega}}^{(2,1)}(\boldsymbol{\beta})^{-1}\widehat{\boldsymbol{g}}^{(2,1)}(\boldsymbol{\beta})\right) \\ &= \frac{n}{2}\left(2 - \widehat{\boldsymbol{g}}(\boldsymbol{\beta})^\top \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\widehat{\boldsymbol{g}}(\boldsymbol{\beta})\right) \end{aligned}$$

where $\widehat{\boldsymbol{g}}(\boldsymbol{\beta}) = \left(\widehat{\boldsymbol{g}}^{(1,2)}(\boldsymbol{\beta})^\top, \widehat{\boldsymbol{g}}^{(2,1)}(\boldsymbol{\beta})^\top\right)^\top$ is the composite vector of moment conditions, while

$$\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) = \begin{pmatrix} \widehat{\boldsymbol{\Omega}}^{(1,2)}(\boldsymbol{\beta}) & \boldsymbol{O} \\ \boldsymbol{O} & \widehat{\boldsymbol{\Omega}}^{(2,1)}(\boldsymbol{\beta}) \end{pmatrix}$$

is the block diagonal matrix that includes in the main diagonal the variance covariance matrices of the partial moment conditions where $\boldsymbol{O}$ is a $(1+K_2)\times(1+K_2)$ matrix of zeros and

$$\widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_{1i}^{(\eta)}\delta - \boldsymbol{z}_{2i}^\top\boldsymbol{\gamma})^2 \boldsymbol{z}_i^{(\zeta)}\boldsymbol{z}_i^{(\zeta)\top}, \quad (\eta,\zeta) \in \mathcal{A}(2) = \{(1,2),(2,1)\}.$$

3.2 Asymptotic Results

We now provide the main asymptotic results. To derive consistency and asymptotic normality we use the same assumptions as in Newey & Smith (2004). The treatment of the asymptotic results is general and the linear model presented in Section 2 can be seen as a special case.

Theorem 1 (Consistency)**.** *Let us assume that*

1. $\boldsymbol{\beta}_0 \in \mathcal{B}$ *is the unique solution to* $\mathbb{E}[\boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})] = \mathbf{0}$ *and* $\mathcal{B}$ *is compact,*
2. $\boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})$ *is continuous at each* $\boldsymbol{\beta} \in \mathcal{B}$ *with probability one,*
3. $\mathbb{E}[\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})\|^{\alpha}] < \infty$ *for some* $\alpha > 2$,
4. $\boldsymbol{\Omega}^{(\eta,\zeta)}(\boldsymbol{\beta})$ *is non singular,*

then $\widehat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$ *and* $\widehat{\boldsymbol{\lambda}} \to_p \mathbf{0}$.

Let us define

$$\boldsymbol{V} = \left(\boldsymbol{G}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{G}\right)^{-1} \boldsymbol{G}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{S} \boldsymbol{\Omega}^{-1} \boldsymbol{G} \left(\boldsymbol{G}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{G}\right)^{-1} = \boldsymbol{H}^\top \boldsymbol{S} \boldsymbol{H}$$

where $\boldsymbol{H}$ is implicitly defined, $\boldsymbol{S} = \lim_{n\to\infty} n \operatorname{Var}[\widehat{\boldsymbol{g}}(\boldsymbol{\beta})]$ is a potentially singular matrix, while $\boldsymbol{\Omega}$ is a block diagonal matrix whose generic diagonal element is $\boldsymbol{\Omega}^{(\eta,\zeta)} = \lim_{n\to\infty} n \operatorname{Var}[\widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta})]$. The $K \times (1+K_2)$ matrix $\boldsymbol{G}$ is obtained by stacking the $(J-1)J$ population gradient matrices $\boldsymbol{G}^{(\eta,\zeta)} = \mathbb{E}[\boldsymbol{G}_i^{(\eta,\zeta)}(\boldsymbol{\beta})]$, where $\boldsymbol{G}_i^{(\eta,\zeta)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})$.

Theorem 2 (Asymptotic normality)**.** *If the assumptions in Theorem 1 hold and*

1. $\boldsymbol{\beta}_0$ *is an interior point of* $\mathcal{B}$,
2. $\boldsymbol{g}_i^{(\eta,\zeta)}(\boldsymbol{\beta})$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* $\boldsymbol{\beta}_0$ *and* $\mathbb{E}[\sup_{\boldsymbol{\beta}\in\mathcal{N}} \|\boldsymbol{G}_i^{(\eta,\zeta)}(\boldsymbol{\beta})\|] < \infty$,
3. $\boldsymbol{G}^{(\eta,\zeta)}$ *is full rank,*

then $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\to_d N\left(\mathbf{0},\boldsymbol{V}\right)$.

An estimator of $\boldsymbol{V}$, say $\widehat{\boldsymbol{V}}$, requires an estimator for each of its components. $\boldsymbol{\Omega}$ can be estimated as suggested in Equation (6). In order to estimate $\boldsymbol{G}$ and $\boldsymbol{S}$ we use

$$\widehat{\boldsymbol{G}}^{(\eta,\zeta)}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{G}_i^{(\eta,\zeta)}(\boldsymbol{\beta}),\qquad \widehat{\boldsymbol{S}}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(\boldsymbol{\beta})\boldsymbol{g}_i(\boldsymbol{\beta})^\top.$$

3.3 The Two Proxy Case: Variance Estimation

As in the general case, the estimation of $\boldsymbol{V}$ reduces to substituting the matrices that compose $\boldsymbol{V}$, i.e. $\boldsymbol{\Omega}$, $\boldsymbol{G}$ and $\boldsymbol{S}$ with their sample counterparts $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})$, $\widehat{\boldsymbol{G}}(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{S}}(\boldsymbol{\beta})$. Hence,

$$\widehat{\boldsymbol{G}}^{(1,2)}(\boldsymbol{\beta})=\frac{1}{n}\boldsymbol{X}^{(1)^\top}\boldsymbol{Z}^{(2)},\quad \widehat{\boldsymbol{G}}^{(2,1)}(\boldsymbol{\beta})=\frac{1}{n}\boldsymbol{X}^{(2)^\top}\boldsymbol{Z}^{(1)}$$

$$\widehat{\boldsymbol{S}}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{z}_i^{(1)}(y_i-x_{1i}^{(2)}\delta-\boldsymbol{z}_{2i}^\top\boldsymbol{\gamma})\\ \boldsymbol{z}_i^{(2)}(y_i-x_{1i}^{(1)}\delta-\boldsymbol{z}_{2i}^\top\boldsymbol{\gamma})\end{pmatrix}\begin{pmatrix}\boldsymbol{z}_i^{(1)}(y_i-x_{1i}^{(2)}\delta-\boldsymbol{z}_{2i}^\top\boldsymbol{\gamma})\\ \boldsymbol{z}_i^{(2)}(y_i-x_{1i}^{(1)}\delta-\boldsymbol{z}_{2i}^\top\boldsymbol{\gamma})\end{pmatrix}^\top.$$

The estimator for $\boldsymbol{\Omega}$ is given in Section 3.1.

4 Numerical Results

In this section we study the finite sample properties of the CombEEL estimator by means of some Monte Carlo experiments and an application to real data. In particular, we study the performance of the estimator in terms of bias and mean squared error (MSE) and we analyze the size and power properties of the t test associated to the parameter of interest δ. The data example considers the problem of Chalfin & McCrary (2017) on the estimation of the elasticity of crime with respect to the presence of police. The numerical calculations are carried out in `R` (R Core Team, 2017).

4.1 Monte Carlo Experiments

We consider $J = 2, \ldots, 10$ throughout. The $J = 2$ case is compared to the optimal estimator of Andersson & Møen (2016) denoted as OptIV and the estimator of Chalfin & McCrary (2017) denoted as GMM. Let us consider the model

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{\iota}\gamma + \boldsymbol{x}_1^*\delta + \boldsymbol{\varepsilon}, \\
\boldsymbol{x}_1^{(\eta)} &= \boldsymbol{x}_1^* + \boldsymbol{e}^{(\eta)}
\end{aligned}
$$

where $\boldsymbol{\iota}$ is a vector of ones. The true parameters are $\gamma = \delta = 1$ and $\boldsymbol{\varepsilon}$ is sampled from a normal distribution with zero mean and variance $\mathrm{Var}[\varepsilon_i] = \sigma_i^2$ where $\sigma_i = \sqrt{(x_{1i}^*)^2}$. The latent variable $\boldsymbol{x}_1^*$ is sampled from a standard normal distribution. The measurement errors are also normally distributed, their variances though follow the three scenarios below:

1. $\mathrm{Var}[e_i^{(\eta)}] = 1 - \frac{\eta-1}{10}$, $\eta = 1, \ldots, 10$,
2. $\mathrm{Var}[e_i^{(\eta)}] = 1$, $\eta = 1, \ldots, 10$,
3. $\mathrm{Var}[e_i^{(\eta)}] = \eta$, $\eta = 1, \ldots, 10$.

The sample size is $n = 500$, while the number of replications is 5000. The first scenario corresponds to a situation where the quality of the proxies improves as J increases. In the second scenario the quality of the proxies remains constant, while in the third scenario it deteriorates with J. The performance of the CombEEL estimator is measured in terms of bias, mean square error (MSE) and size of the test for the null hypothesis $H_0 : \delta = 1$. Furthermore, the power of the t test is analyzed as J increases.

Bias and MSE

The results for the bias are collected in Figures 1, 2 and 3. We notice that the bias of the OptIV estimator is larger, in absolute value, than the bias of the other estimators (about -0.008). The GMM estimator, on the other hand, has a bias that moves approximately between -0.003 and -0.004. The bias of the CombEEL estimator varies between 0.002 and -0.002 as J increases. The MSE of the CombEEL estimator shows a clear tendency to decrease with the number of proxies. For $J = 2$, the OptIV has the largest MSE, while the MSE of the CombEEL estimator is the smallest.

Figure 1 about here.

Figure 2 about here.

Figure 3 about here.

Figure 4 about here.

Figure 5 about here.

Figure 6 about here.

Size and Power

The size of the CombEEL estimator is generally very close to the nominal size for the first two scenarios (see Figures 7 and 8). When the variance of the measurement error increases we can observe some size distortion as the number of proxies increases (Figure 9). In general the size properties of the CombEEL estimator outperform those of the competing estimators. With respect to power, we observe that it increases as the number of proxies increases (see Figures 10, 11 and 12). This phenomenon is observable independently of the quality of the proxies. It is interesting to notice that in the third scenario (increasing variance) the t test has very little power for alternatives close to the true value and when only two proxies are used. Increasing the number of proxies improves the power properties of the test (see Figure 12). Intuitively, this would suggest that the CombEEL estimator is able to efficiently extract information from the available variables even when the quality of the proxies is low.

Figure 7 about here.

Figure 8 about here.

Figure 9 about here.

Figure 10 about here.

Figure 11 about here.

Figure 12 about here.

4.2 The Effect of Police on Crime

Let us consider the model proposed by Chalfin & McCrary (2017) on the effect of police on crime.[4] Crime growth rates y_i are defined as a function of police growth rate x_i^* and other exogenous covariates $\boldsymbol{z}_i$:

$$y_i = \delta x_i^* + \boldsymbol{z}_i^\top \boldsymbol{\gamma} + \varepsilon_i.$$

Typically, police growth rates are based on the the Uniform Crime Reports (UCR) provided by the Federal Bureau of Investigation (FBI). Interestingly, Chalfin & McCrary (2017) find clear evidence of measurement errors in the UCR data with respect to the number of sworn police officers. Another source of information is the U.S. Census Bureau's Annual Survey of Government (ASG). The ASG survey provides information on all city government officials. This data set can be used to build an alternative measure for the number of sworn police officers. The final data set includes 10589 data points for 49 years, from 1962 to 2010, 45

[4]The data set can be downloaded from prof. Justin McCrary's webpage: https://eml.berkeley.edu/ jmccrary/.

U.S. states and 242 cities. The model considers as the dependent variable the growth rate of crime y_i. We consider various types of property crimes, violent crimes and aggregate measures. The main explanatory variable is the growth rate of police measured via the UCR data and the ASG survey. The model also includes the growth rate of population and city-year controls. So, the two proxies related to the latent variable are

$$x_i^{(\eta)} = x_i^* + u_i^{(\eta)}, \quad \eta = UCR, ASG.$$

The moment conditions for the estimator of Chalfin & McCrary (2017) are then defined as

$$\boldsymbol{g}_i(\boldsymbol{\beta}) = w_i \begin{pmatrix} x_i^{(ASG)}(y_i - \delta_1 x_i^{(UCR)} - \boldsymbol{z}_i^\top \boldsymbol{\gamma}_1) \\ \boldsymbol{z}_i(y_i - \delta_1 x_i^{(UCR)} - \boldsymbol{z}_i^\top \boldsymbol{\gamma}_1) \\ x_i^{(UCR)}(y_i - \delta_2 x_i^{(ASG)} - \boldsymbol{z}_i^\top \boldsymbol{\gamma}_2) \\ \boldsymbol{z}_i(y_i - \delta_2 x_i^{(ASG)} - \boldsymbol{z}_i^\top \boldsymbol{\gamma}_2) \end{pmatrix}$$

where w_i is 2010 population in levels. Here, the obvious restriction is $\delta_1 = \delta_2 = \delta$, while $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are allowed to differ.

Table 1 reports the estimates for the elasticity of crime for each crime considered. The second and third columns show the estimate of the elasticity of crime computed via the CombEEL estimator and OptIV estimator respectively. The fourth column replicates the results of column 9 in Table 3 of Chalfin & McCrary (2017).[5]

Chalfin & McCrary (2017) stress the fact that the attenuation bias associated to measurement errors in the police variable substantially distorts the estimate of the elasticity of crime and, in particular, violent crime such as murder. Their biased-corrected estimates suggest that previous studies have underestimated the magnitude of such elasticity. Our estimates confirm their results. More specifically, the elasticity of murder is larger than the one produced by both the GMM estimator and the OptIV estimator. With respect to the former, the CombEEL is also characterized by a smaller standard error. It is interesting to

[5]Our GMM estimates are slightly different from those obtained by Chalfin & McCrary (2017). This is probably due to differences in the optimization algorithm.

y_i	CombEEL	OptIV	GMM
Murder	-0.779	-0.653	-0.672
	(0.221)	(0.202)	(0.237)
Rape	-0.286	-0.242	-0.270
	(0.202)	(0.145)	(0.235)
Robbery	-0.538	-0.558	-0.560
	(0.108)	(0.087)	(0.108)
Aggravated assault	-0.087	-0.098	-0.098
	(0.118)	(0.104)	(0.126)
Burglary	-0.278	-0.216	-0.240
	(0.083)	(0.061)	(0.089)
Larceny	-0.099	-0.081	-0.084
	(0.062)	(0.051)	(0.063)
Motor vehicle theft	-0.320	-0.347	-0.341
	(0.093)	(0.072)	(0.100)
Sum of violent crimes	-0.352	-0.343	-0.344
	(0.089)	(0.072)	(0.092)
Sum of property crimes	-0.181	-0.173	-0.175
	(0.057)	(0.044)	(0.060)
Cost-weighted sum of violent crimes	-0.601	-0.492	-0.511
	(0.191)	(0.176)	(0.209)
Cost-weighted sum of property crimes	-0.271	-0.277	-0.275
	(0.066)	(0.050)	(0.072)
Cost-weighted sum of all crimes	-0.535	-0.457	-0.477
	(0.159)	(0.140)	(0.174)

Note: standard errors are in parentheses.

Table 1: Elasticity of crime with respect to police.

notice that the empirical results are in line with the results in the Monte Carlo simulations. In the two proxy case we observe in fact that the CombEEL tends to have smaller bias (see Figure 1 and Figure 2).

5 Conclusions

This paper introduces a novel method to estimate a linear model when explanatory variables are observed with error and multiple proxies are available. It is based on the EEL principle and, by standard methods, it is possible to show that the estimator is consistent and asymptotically normal. The Monte Carlo simulations show its competitive properties in finite samples. In particular, the numerical results suggest that the CombEEL estimator works very well when the number of proxies is relatively large, even when the quality of the proxies deteriorates. We also find that the power of the t test increases proportionally with the number of proxies. Intuitively, this suggests that the estimator is able to efficiently extract information from the available proxies at least under the scenarios taken into consideration. An application on the estimation of the elasticity of crime confirms the conjecture of Chalfin & McCrary (2017) that previous results are unreliable due to the distorting effect of the attenuation bias. Furthermore, while we focus on the EIV case, the same idea is potentially applicable to a variety of situations that involve the combination of models, as long as at least a subset of the parameter vector of interest is shared by the aforementioned models.

Appendix: Proofs

This section collects the proofs for Theorem 1 and Theorem 2. Let us introduce some notation. The first derivative of $R_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are $R_{n,\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and $R_{n,\boldsymbol{\lambda}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ respectively. Second and cross derivatives are denoted as $R_{n,\boldsymbol{\beta\beta}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, $R_{n,\boldsymbol{\lambda\lambda}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and $R_{n,\boldsymbol{\beta\lambda}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$. The norm of a generic vector, say $\boldsymbol{a}$, is defined as $\|\boldsymbol{a}\| := \sqrt{\boldsymbol{a}^\top \boldsymbol{a}}$. Similarly, the norm of a generic matrix $\boldsymbol{A}$ is defined as $\|\boldsymbol{A}\| := \sqrt{\mathrm{trace}\,(\boldsymbol{A}^\top \boldsymbol{A})}$. Finally, in the context of the mean value theorem, $\dot{\boldsymbol{\beta}}$ and $\dot{\boldsymbol{\lambda}}$ denote mean values.

Proof of Theorem 1. Let us define $Q(\boldsymbol{\beta}) = \boldsymbol{g}(\boldsymbol{\beta})^\top \boldsymbol{\Omega}(\boldsymbol{\beta})^{-1}\boldsymbol{g}(\boldsymbol{\beta})$, the population version of $Q_n(\boldsymbol{\beta})$. Hence, by taking the norm and by triangle inequality

$$
\begin{aligned}
|Q_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| \leq{}& \|\widehat{\boldsymbol{g}}\,(\boldsymbol{\beta}) - \boldsymbol{g}\,(\boldsymbol{\beta})\|^2 \left\|\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\right\| + 2\,\|\widehat{\boldsymbol{g}}\,(\boldsymbol{\beta}) - \boldsymbol{g}\,(\boldsymbol{\beta})\| \left\|\widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\right\| \|\boldsymbol{g}\,(\boldsymbol{\beta})\| \\
&- \|\boldsymbol{g}\,(\boldsymbol{\beta})\|^2 \left\|\boldsymbol{\Omega}(\boldsymbol{\beta})^{-1} - \widehat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\right\|.
\end{aligned}
$$

So, by the uniform weak law of large numbers and the continuous mapping theorem, $\sup_{\boldsymbol{\beta}\in\mathcal{B}} |Q_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| \to_p 0$, which implies $\widehat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$. Moreover, since $\widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)} = \widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\widehat{\boldsymbol{\beta}})^{-1}\widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\widehat{\boldsymbol{\beta}})$, by the following mean value argument

$$
\widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta}_0) + \widehat{\boldsymbol{G}}^{(\eta,\zeta)}(\dot{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),
$$

the uniform weak law of large numbers and the continuous mapping theorem, consistency of $\widehat{\boldsymbol{\beta}}$, we get $\widehat{\boldsymbol{\lambda}} \to_p \boldsymbol{0}$. □

Proof of Theorem 2. By computing the first order conditions for the parameters of interest $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ we get

$$
\begin{aligned}
\boldsymbol{0} &= \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)}) = -\frac{n}{2} \sum_{(\eta,\zeta)\mathcal{A}(J)} \left(\frac{\boldsymbol{\lambda}^{(\eta,\zeta)\top}}{n} \frac{\partial \widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) \frac{\boldsymbol{\lambda}^{(\eta,\zeta)}}{n} - 2\frac{\boldsymbol{\lambda}^{(\eta,\zeta)\top}}{n} \frac{\partial \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right), \\
\boldsymbol{0} &= R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)}) = -\left(\widehat{\boldsymbol{\Omega}}^{(\eta,\zeta)}(\boldsymbol{\beta})\boldsymbol{\lambda}^{(\eta,\zeta)} - \widehat{\boldsymbol{g}}^{(\eta,\zeta)}(\boldsymbol{\beta}) \right).
\end{aligned}
$$

Let us now take a mean value expansion of the first order conditions about the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$:

$$
\begin{aligned}
\mathbf{0} = \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)}) &= \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \mathbf{0}) + \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}\boldsymbol{\lambda}^{(\eta,\zeta)}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})\widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)} \\
&+ \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}\boldsymbol{\lambda}^{(\eta,\zeta)}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})\sqrt{n}\widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)} \\
&+ \sum_{(\eta,\zeta)\in\mathcal{A}(J)} R_{n,\boldsymbol{\beta}\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{0} = R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)}) &= R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}}(\boldsymbol{\beta}_0, \mathbf{0}) + R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}\boldsymbol{\lambda}^{(\eta,\zeta)}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})\widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)} \\
&+ R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \sqrt{n}R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}}(\boldsymbol{\beta}_0, \mathbf{0}) + R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}\boldsymbol{\lambda}^{(\eta,\zeta)}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})\sqrt{n}\widehat{\boldsymbol{\lambda}}^{(\eta,\zeta)} \\
&+ R_{n,\boldsymbol{\lambda}^{(\eta,\zeta)}\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)})\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).
\end{aligned}
$$

The equations above can be rewritten as

$$
\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \sqrt{n}\widehat{\boldsymbol{g}}(\boldsymbol{\beta}_0) \end{pmatrix} + \begin{pmatrix} \sum_{(\eta,\zeta)\in\mathcal{A}} \widehat{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)}) & \dot{\boldsymbol{G}}^\top \\ \dot{\boldsymbol{G}} & \dot{\boldsymbol{\Lambda}} \end{pmatrix} \begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ \sqrt{n}\widehat{\boldsymbol{\lambda}} \end{pmatrix}.
$$

Hence, invoking again the uniform weak law of large numbers in concert with the continuous mapping theorem, we have

$$
\begin{pmatrix} \sum_{(\eta,\zeta)\in\mathcal{A}} \widehat{R}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\dot{\boldsymbol{\beta}}, \dot{\boldsymbol{\lambda}}^{(\eta,\zeta)}) & \dot{\boldsymbol{G}}^\top \\ \dot{\boldsymbol{G}} & \dot{\boldsymbol{\Lambda}} \end{pmatrix}^{-1} = - \begin{pmatrix} \mathbf{0} & \boldsymbol{G}^\top \\ \boldsymbol{G} & \boldsymbol{\Lambda} \end{pmatrix}^{-1} + o_p(1)
$$

Thus,

$$\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ \sqrt{n}\widehat{\boldsymbol{\lambda}} \end{pmatrix} = -\begin{pmatrix} \mathbf{0} & \boldsymbol{G}^\top \\ \boldsymbol{G} & \boldsymbol{\Lambda} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0} \\ \sqrt{n}\widehat{\boldsymbol{g}}(\boldsymbol{\beta}_0) \end{pmatrix} + o_p(1) = -\begin{pmatrix} -\boldsymbol{\Sigma} & \boldsymbol{H}^\top \\ \boldsymbol{H} & \boldsymbol{\Xi} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \sqrt{n}\widehat{\boldsymbol{g}}(\boldsymbol{\beta}_0) \end{pmatrix} + o_p(1)$$

$$= -\begin{pmatrix} \boldsymbol{H}^\top \sqrt{n}\widehat{\boldsymbol{g}}(\boldsymbol{\beta}_0) \\ \boldsymbol{\Xi}\sqrt{n}\widehat{\boldsymbol{g}}(\boldsymbol{\beta}_0) \end{pmatrix} + o_p(1)$$

where

$$\boldsymbol{\Sigma} = \left(\boldsymbol{G}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{G}\right)^{-1}, \ \boldsymbol{H} = \boldsymbol{\Omega}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}, \ \boldsymbol{\Xi} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^\top\boldsymbol{\Omega}^{-1}.$$

To conclude the proof, standard application of the central limit theorem provides

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to_d N(\mathbf{0}, \boldsymbol{V})$$

where $\boldsymbol{V} = \boldsymbol{H}^\top \boldsymbol{S} \boldsymbol{H}$. □

References

Andersson, J. & Møen, J. (2016). A simple improvement of the IV-estimator for the classical errors-in-variables problem. *Oxford Bulletin of Economics and Statistics*, *78*, 113–125.

Antoine, B., Bonnal, H. & Renault, E. (2007). On the efficient use of the informational content of estimating equations: implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics*, *138*, 461–487.

Ashenfelter, L. & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review*, *84*, 1157–1173.

Barron, J. M., Berger, M. C. & Black, D. A. (1997). How well do we measure training? *Journal of Labor Economics*, *15*, 507–528.

Bernal, P., Mittag, N. & Qureshi, J. A. (2016). Estimating effects of school quality using multiple proxies. *Labor Economics*, *39*, 1–10.

Black, D. A. & Smith, J. A. (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics*, *24*, 701–728.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review*, *85*, 365–390.

Chalfin, A. & McCrary, J. (2017). Are U.S. cities underpoliced? Theory and evidence. *Review of Economics and Statistics*.

Chen, X., Hong, H. & Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, *49*, 901–937.

Fryer, R. G., Heaton, P. S., Levitt, S. D. & Murphy, K. M. (2013). Measuring crack cocaine and its impact. *Economic Inquiry*, *51*, 1651–1681.

Fuller, W. (1987). *Measurement Error Models.* Wiley.

Guiso, L., Sapienza, P. & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review*, *94*, 526-556.

Krueger, A. & Lindahl, M. (2001). Education for growth: why and for whom? *Journal of Economic Literature*, *39*, 1101–1136.

Lubotsky, D. & Wittenberg, M. (2006). Interpretation of regressions with multiple proxies. *Review of Economics and Statistics*, *88*, 549–562.

Newey, W. & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, *72*, 219–255.

Newey, W. & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, *77*, 687–719.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Tsao, M. & Wu, C. (2006). Empirical likelihood inference for a common mean in the presence of heteroscedasticity. *The Canadian Journal of Statistics*, *34*, 45–59.

Tsao, M. & Wu, F. (2015). Two-sample extended empirical likelihood for estimating equations. *Journal of Multivariate Analysis*, *142*, 1–15.

Wansbeek, T. & Meijer, E. (2000). *Measurement Error and Latent Variables in Econometrics*. North-Holland, Amsterdam.

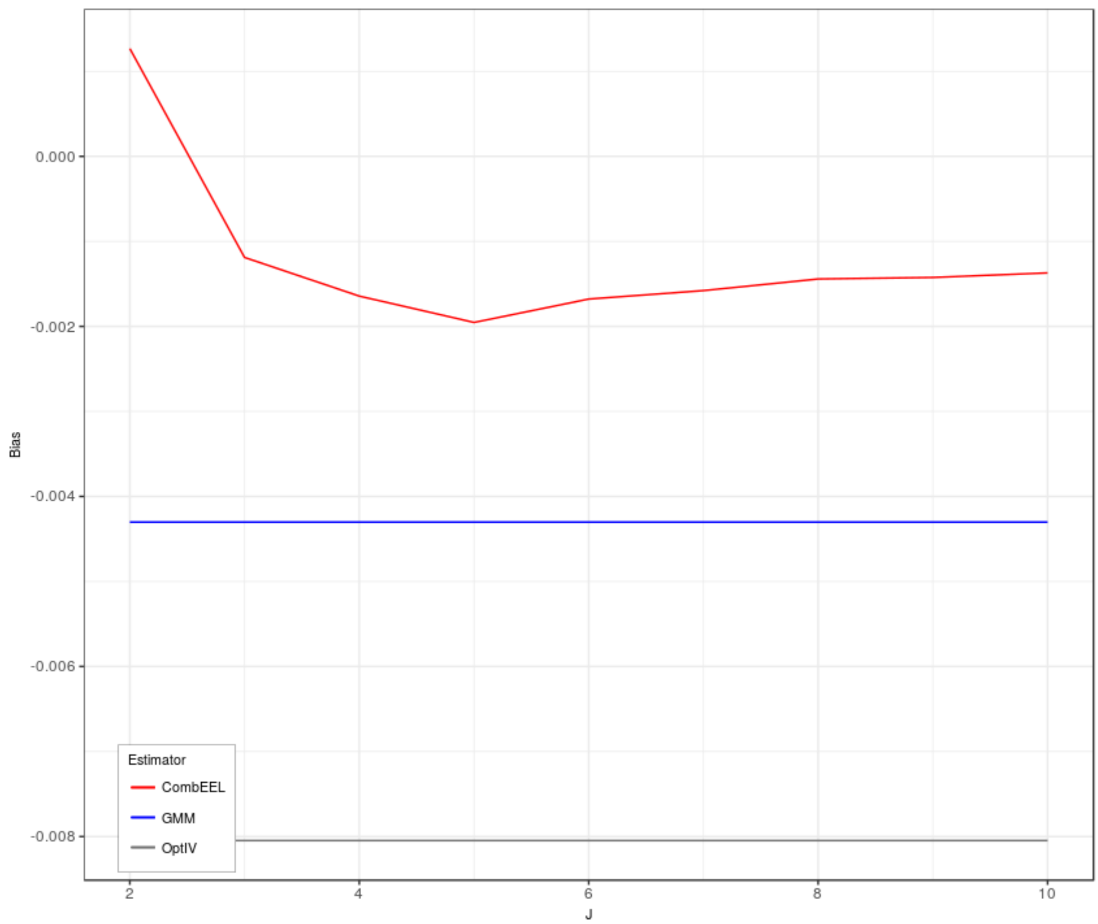

Figure 1: Bias of the CombEEL estimator for $\eta = 2, \ldots, 10$ under scenario 1 (decreasing variance). The plot also features the size of the GMM estimator and the OptIV estimator for $J = 2$.

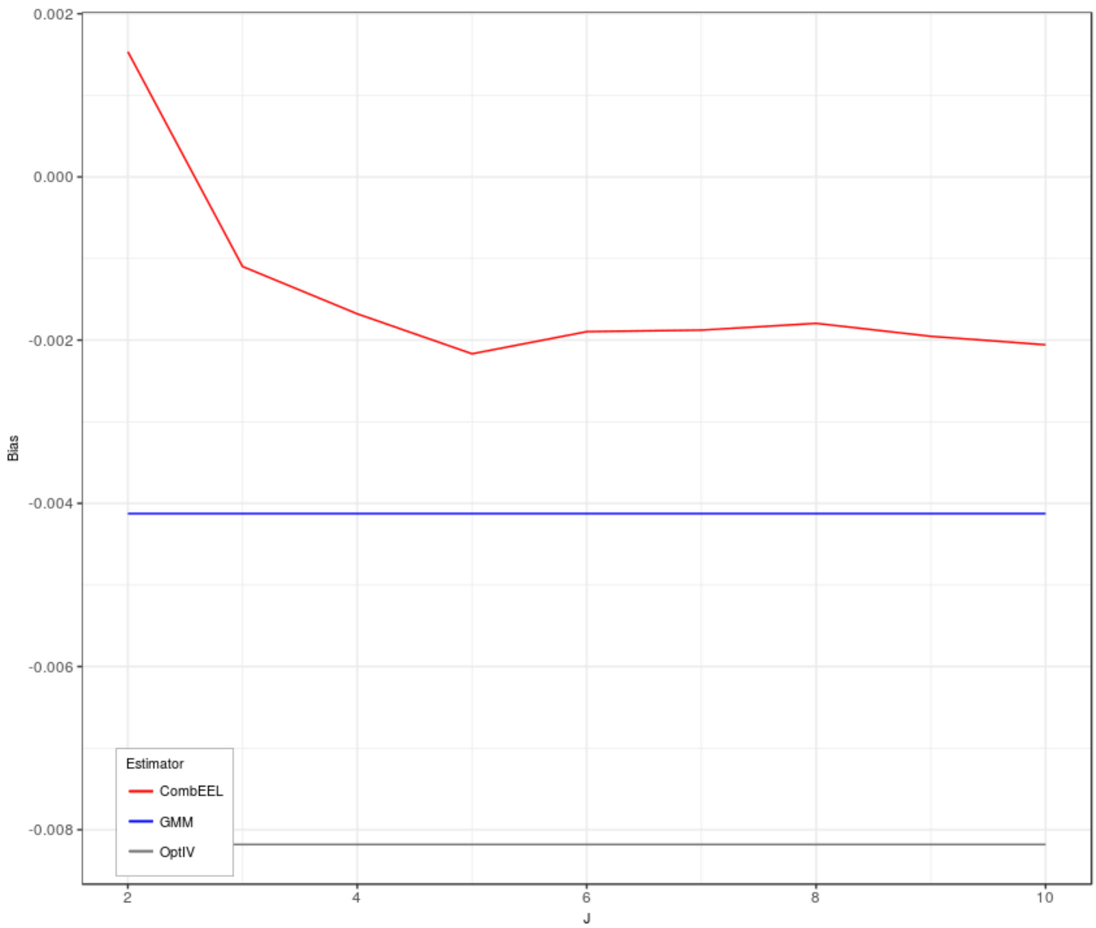

Figure 2: Bias of the CombEEL estimator for $\eta = 2, \ldots, 10$ under scenario 2 (constant variance). The plot also features the bias of the GMM estimator and the OptIV estimator for $J = 2$.

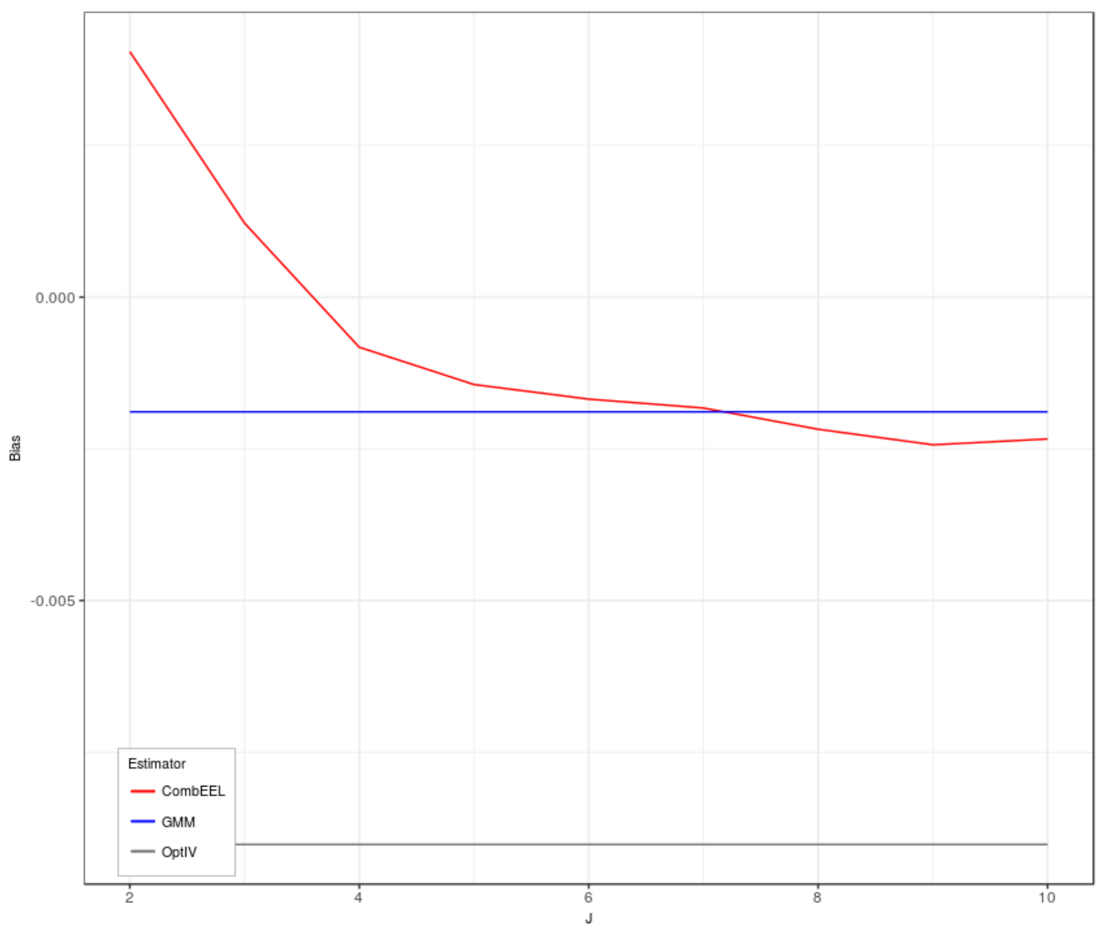

Figure 3: Bias of the CombEEL estimator for $\eta = 2, \ldots, 10$ under scenario 3 (increasing variance). The plot also features the bias of the GMM estimator and the OptIV estimator for $J = 2$.

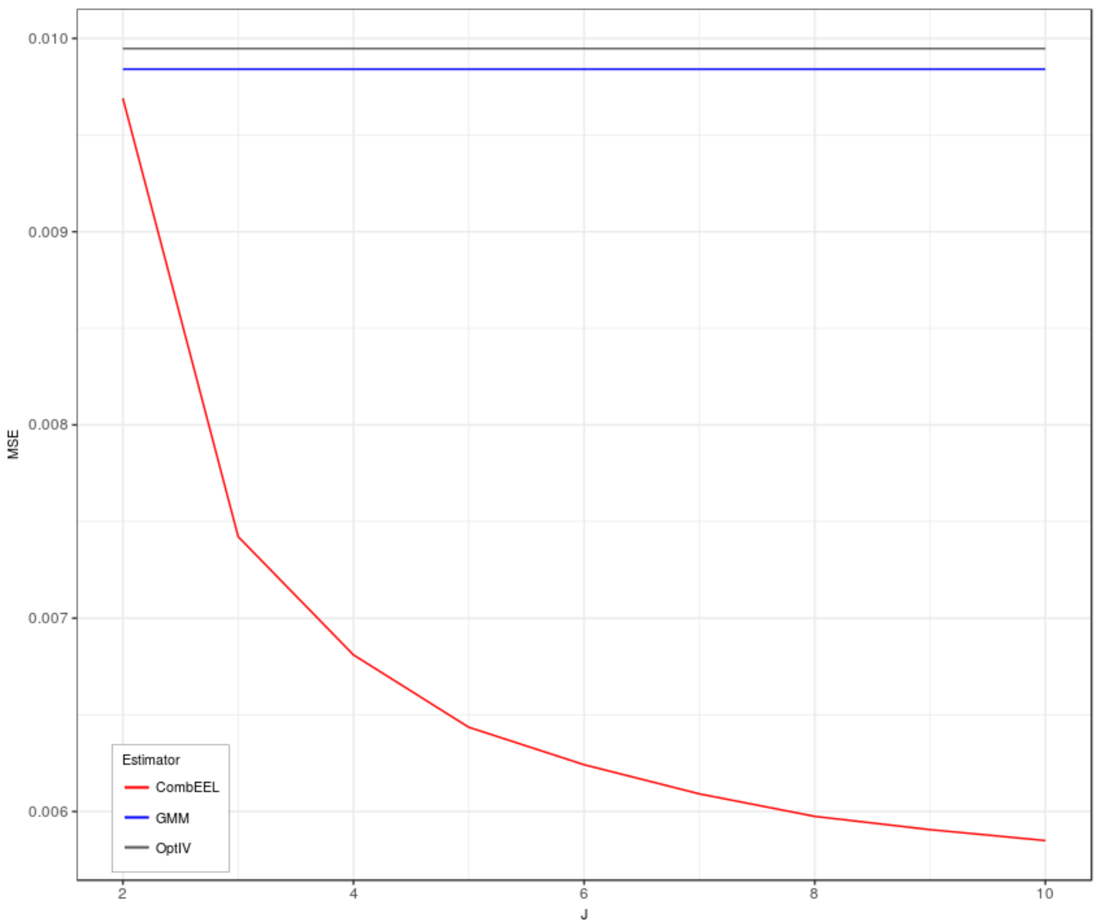

Figure 4: MSE of the CombEEL estimator for $\eta = 2, \ldots, 10$ under scenario 1 (decreasing variance). The plot also features the MSE of the GMM estimator and the OptIV estimator for $J = 2$.

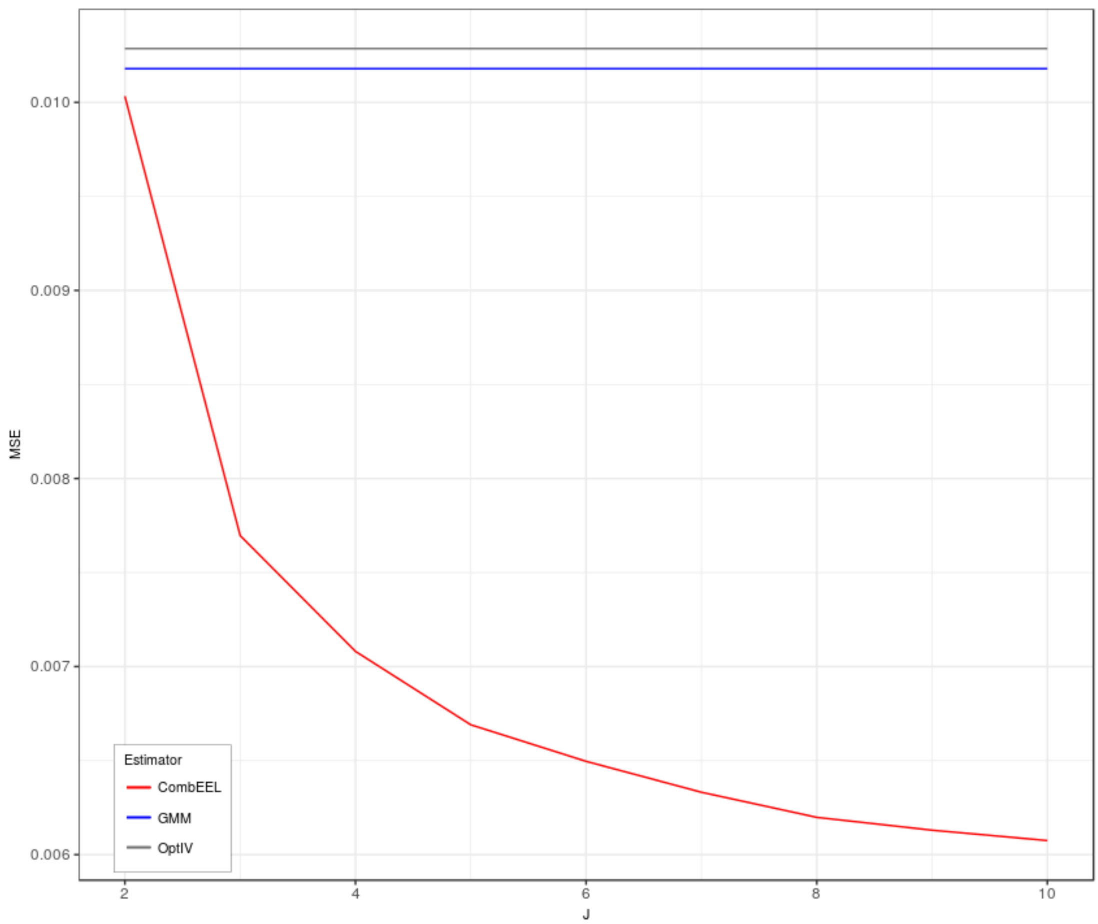

Figure 5: MSE of the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 2 (constant variance). The plot features the size for the GMM estimator and the OptIV estimator for $J = 2$.

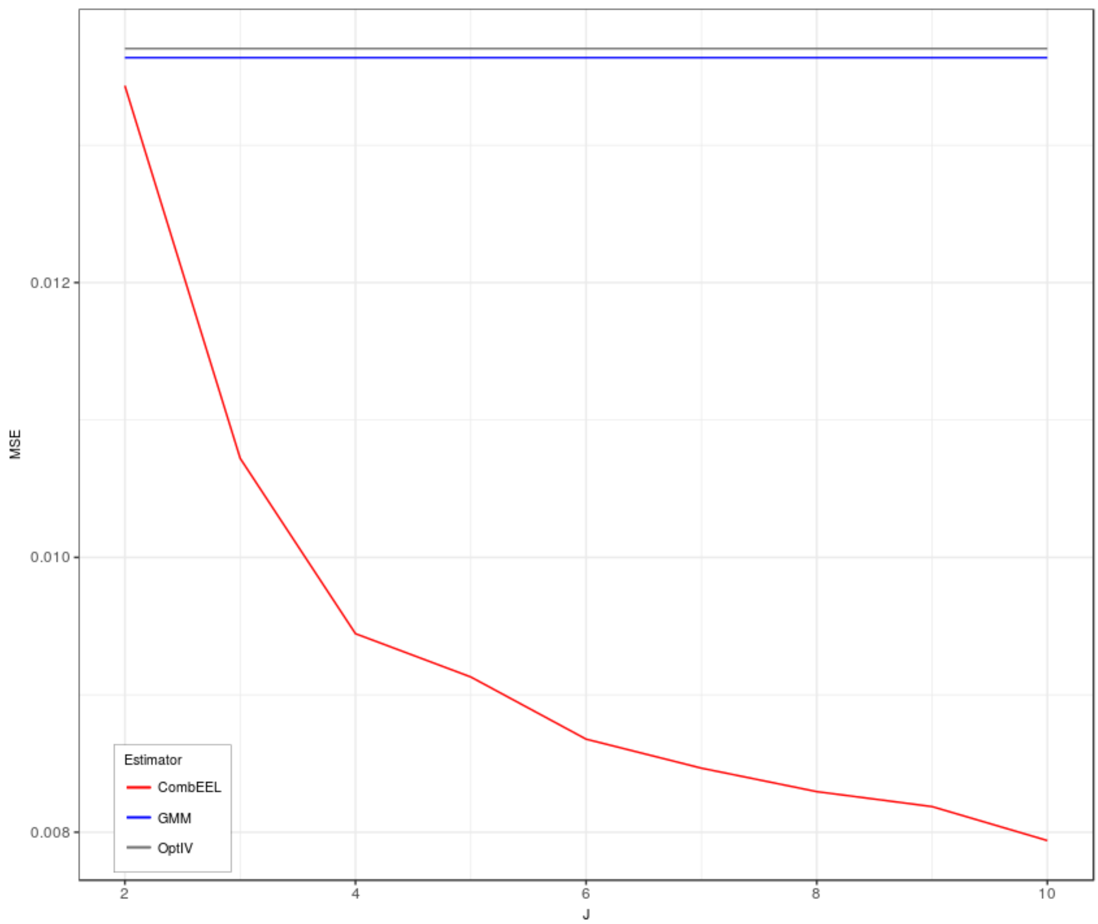

Figure 6: MSE of the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 2 (increasing variance). The plot features the size for the GMM estimator and the OptIV estimator for $J = 2$.

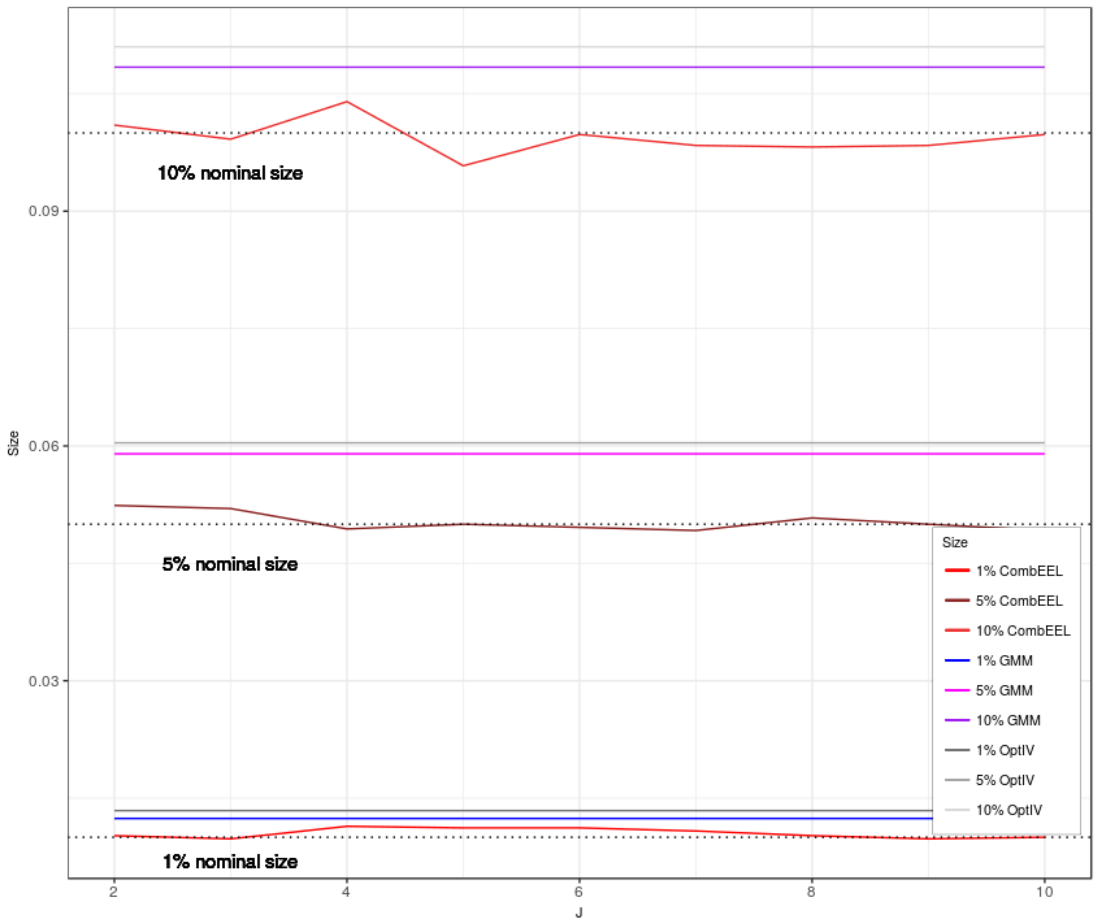

Figure 7: Size of the t test associated to the CombEEL estimator for $\eta = 2, \dots, 10$ and $H_0 : \delta = 1$ under scenario 1 (decreasing variance). The plot also features the size for the GMM estimator and the OptIV estimator for $J = 2$. The dotted lines indicate the corresponding nominal levels.

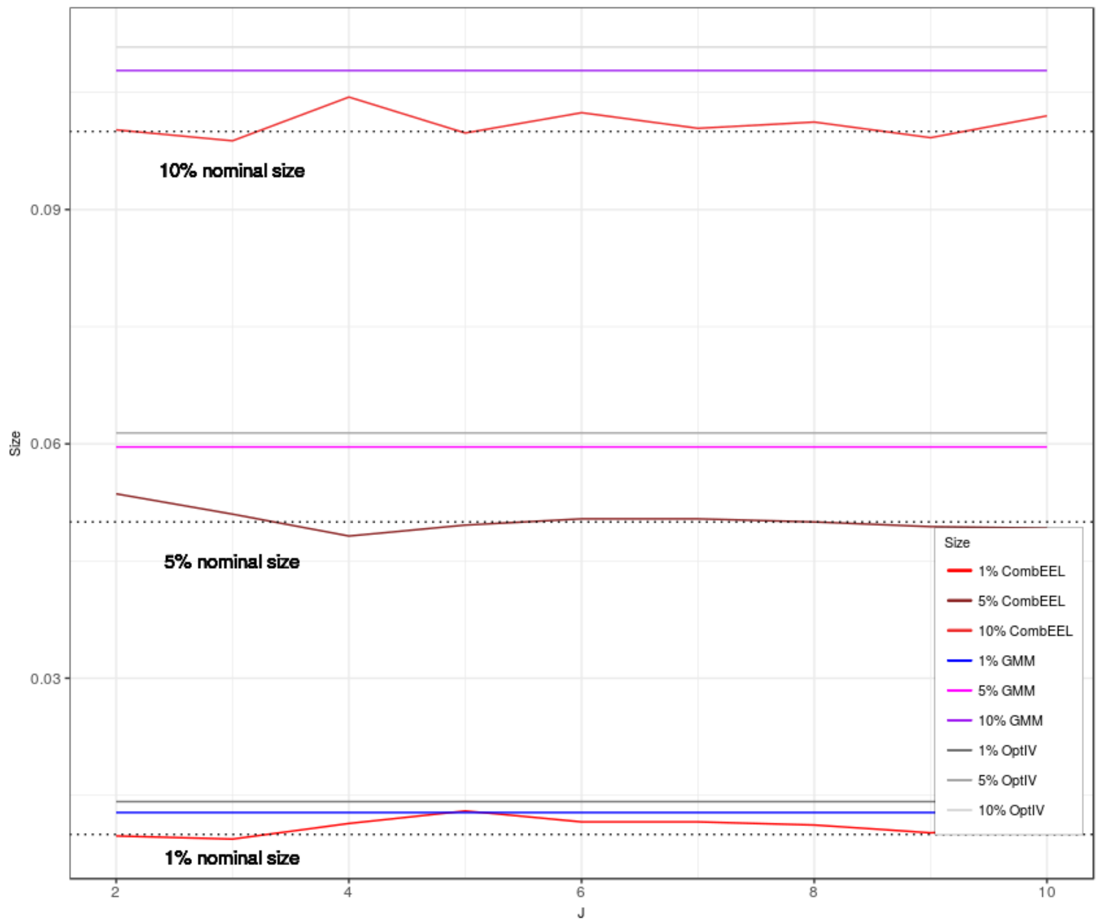

Figure 8: Size of the t test associated to the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 2 (constant variance). The plot also features the size for the GMM estimator and the OptIV estimator for $J = 2$. The dotted lines indicate the corresponding nominal levels.

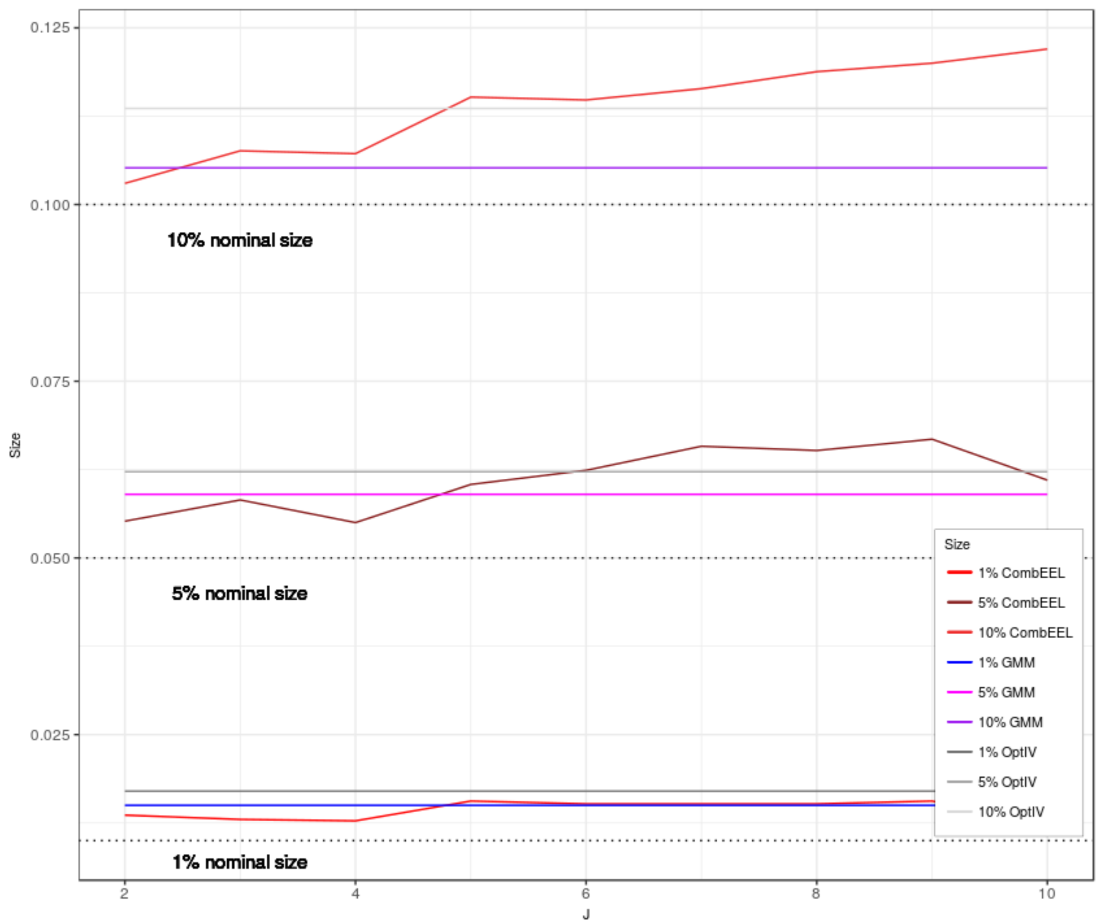

Figure 9: Size of the t test associated to the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 3 (increasing variance). The plot also features the size for the GMM estimator and the OptIV estimator for $J = 2$. The dotted lines indicate the corresponding nominal levels.

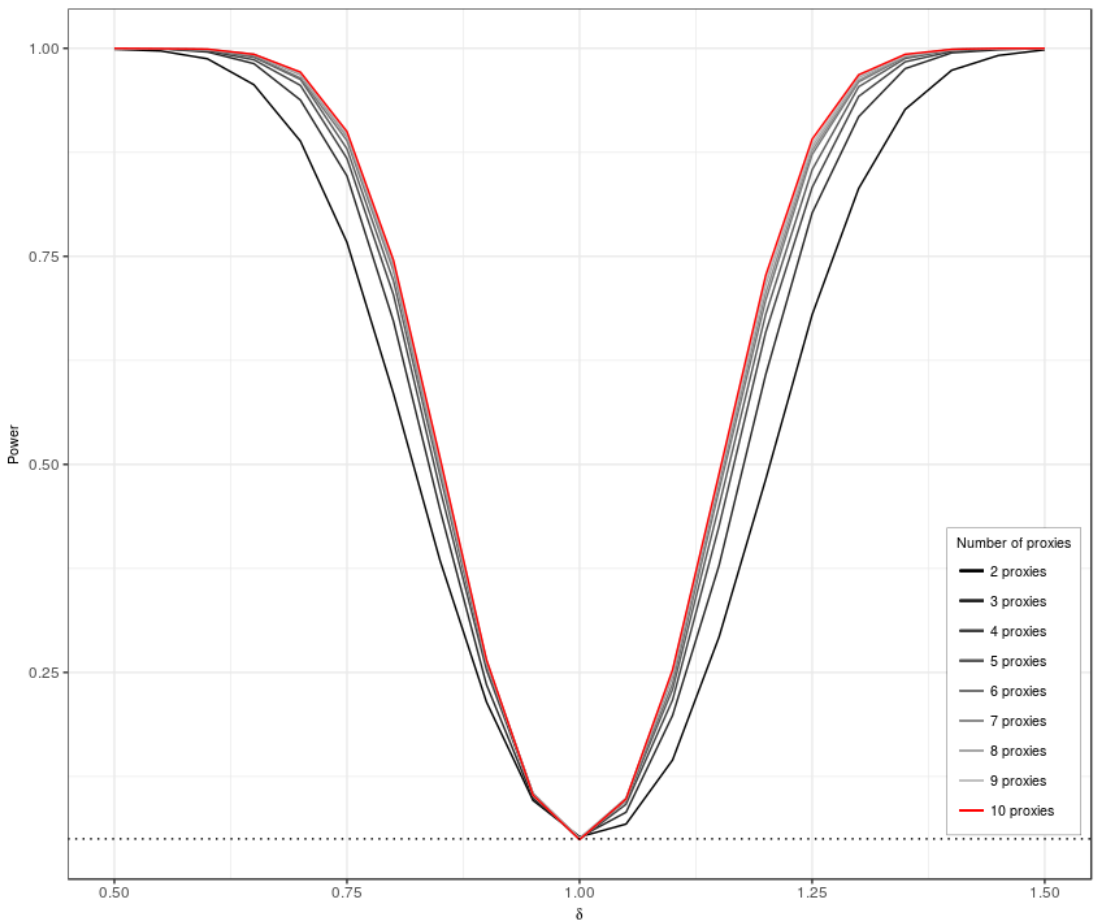

Figure 10: Power of the t test associated to the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 1 (decreasing variance). The dotted horizontal line indicates the 5% nominal level.

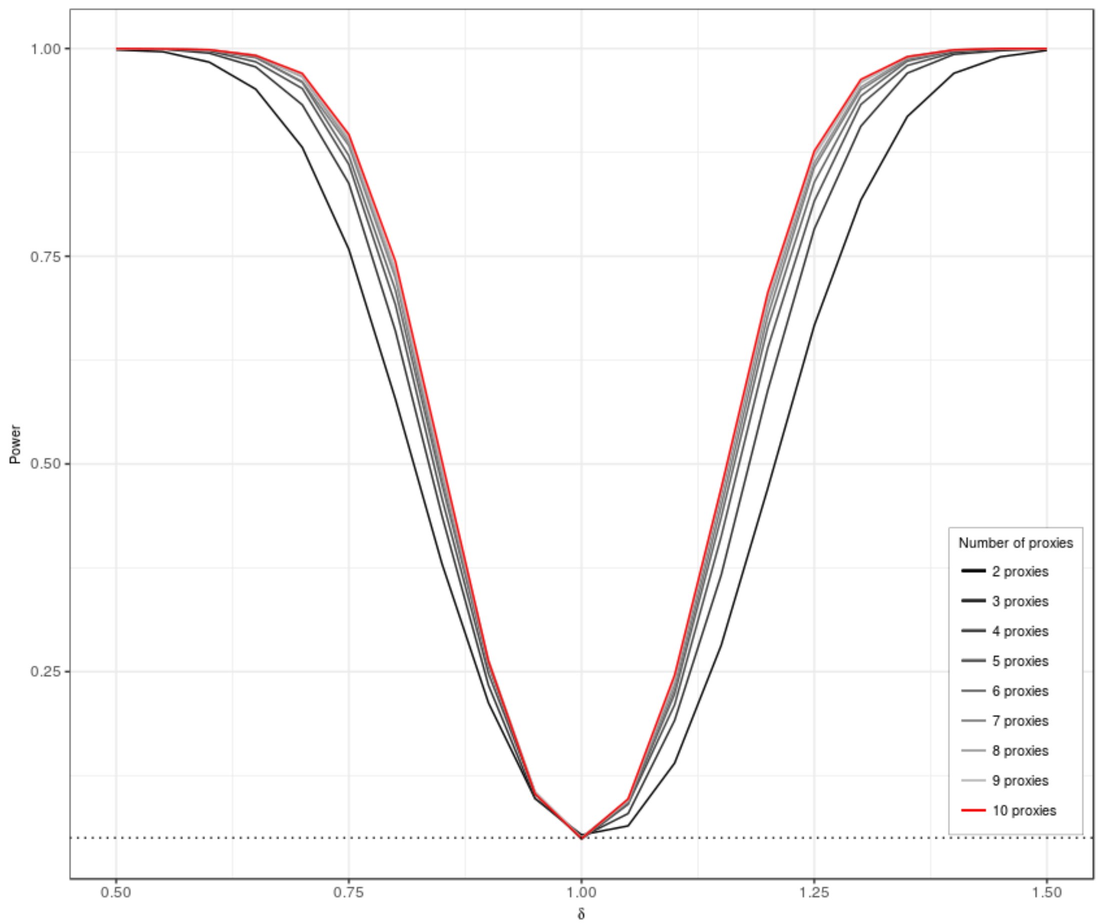

Figure 11: Power of the t test associated to the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 2 (constant variance). The dotted horizontal line indicates the 5% nominal level.

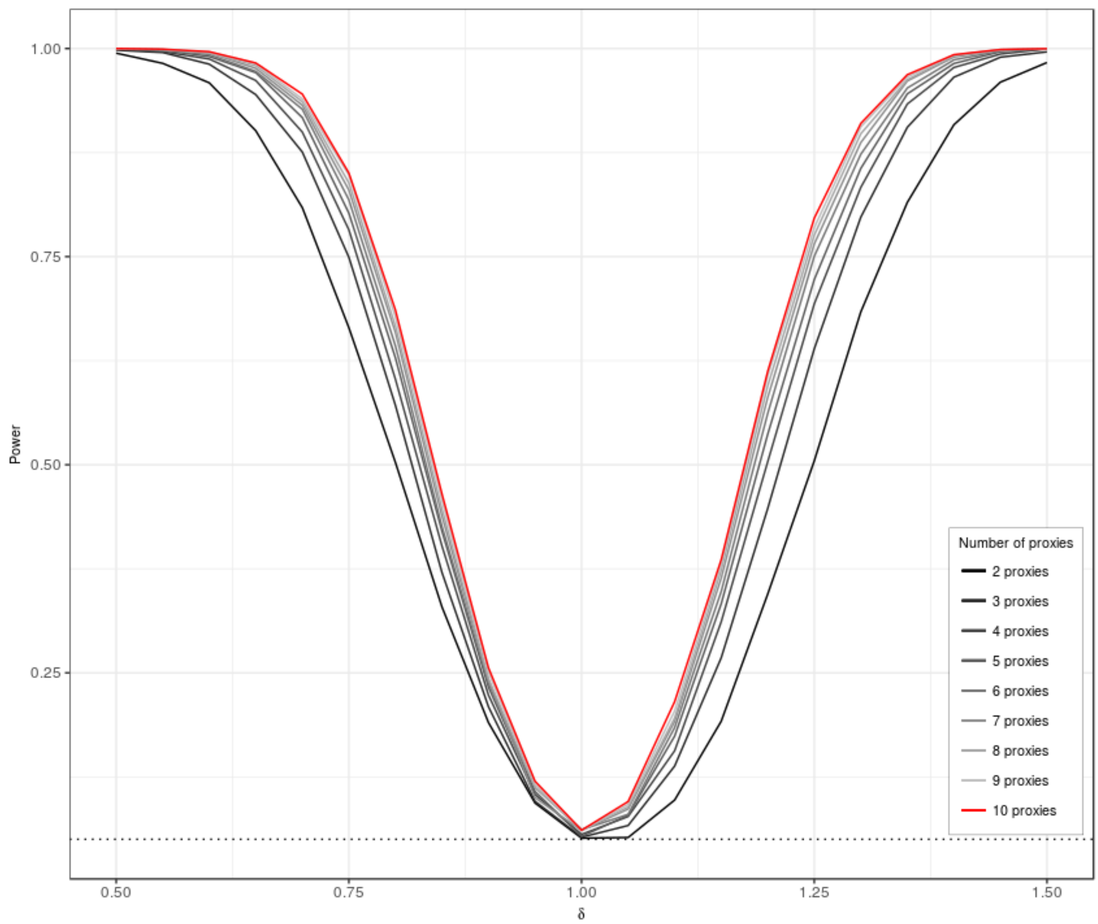

Figure 12: Power of the t test associated to the CombEEL estimator for $\eta = 2, \ldots, 10$ and $H_0 : \delta = 1$ under scenario 3 (increasing variance). The dotted horizontal line indicates the 5% nominal level.