

Split-Sample Score Tests in Linear Instrumental Variables Regression

Saraswata Chaudhuri^{*}, Thomas Richardson[†], James Robins[‡], Eric Zivot[§]

March 28, 2007

Abstract

In this paper we design two split-sample tests for subsets of structural coefficients in a linear Instrumental Variables (IV) regression. Sample splitting serves two purposes – 1) validity of the resultant tests does not depend on the identifiability of the coefficients being tested and 2) it combines information from two unrelated samples one of which need not contain information on the dependent variable. The tests are performed on sub-sample one using the regression coefficients obtained from running the so-called first stage regression on sub-sample two (sample not containing information on the dependent variable). The first test uses the unbiased split-sample IV estimator of the remaining structural coefficients constrained by the hypothesized value of the structural coefficients of interest [see Angrist and Krueger (1995)]. We call this the USSIV score test. The USSIV score test is asymptotically equivalent to the standard score test based on sub-sample one when the standard regularity conditions are satisfied. However, the USSIV score test can be over-sized if the remaining structural coefficients are not identified. This motivates another test based on Robins (2004), which we call the Robins-test. The Robins-test is never oversized and if the remaining structural coefficients are identified, the Robins-test is asymptotically equivalent to USSIV score test against \sqrt{n} -local alternatives.

^{*}Department of Economics, University of Washington

[†]Department of Statistics, University of Washington

[‡]Departments of Epidemiology and Biostatistics, Harvard University

[§]Department of Economics, University of Washington

Contents

1	Introduction	3
2	Linear IV Model with Weak Instruments	7
2.1	Model:	7
2.2	Weak Instrument Framework:	8
3	Testing the Null Hypothesis on Subsets of Parameters	9
3.1	Why Split the Sample?	10
3.2	USSIV Score Test:	12
3.3	The Robins-Test:	15
4	Finite Sample Behavior of Split-Sample Score Tests	19
4.1	Monte-Carlo Design and Parameter Specifications:	19
4.2	Rejection Rates when the Null Hypothesis is True:	21
4.3	Rejection Rates when the Null Hypothesis is False:	22
5	Conclusion	23
6	Appendix	26
6.1	Proofs of Results:	26
6.2	K-Test, Partial K-Test and CLR-Test:	28
6.3	Rejection Rules for Projection Type Tests:	29
7	Tables and Figures	32

1 Introduction

In this paper we propose a new split-sample score test for subsets of structural coefficients in linear Instrumental Variables (IV) models and show that it is valid irrespective of the relevance of the instruments. Our test is quite generally less conservative than the projection test based on the split-sample (score) statistic proposed by Dufour and Jasiak (2001).

Split-sample methods of estimating structural coefficients in linear IV models were proposed by Angrist and Krueger (1995) to avoid biased estimation in the presence of irrelevant instruments. The estimation procedure typically involves three steps – 1) split the sample randomly into two sub-samples, 2) compute the predicted values of the endogenous regressors by multiplying the values of the instruments in sub-sample one with the coefficients obtained by running the so-called first-stage regression on sub-sample two and 3) use sub-sample one to estimate the structural coefficients by running a regression of the dependent variable on the predicted values of the endogenous regressors.

Dufour and Jasiak (2001) show that a split-sample version of the score test [see Wang and Zivot (1998)] can be used to jointly test all the structural coefficients in a linear IV regression. Simulation results from Kleibergen (2002) show that when the degree of over-identification is large, split-sample score test can be more powerful than the Anderson-Rubin (AR) test based on the whole sample [see Dufour (1997) and Staiger and Stock (1997)].

A common argument against sample-splitting is the wastage of information which results in loss of power (efficiency). For example, the split-sample score test for jointly testing all the structural coefficients is less powerful than other similar tests like the K-test [see Kleibergen (2002)] and the Conditional Likelihood Ratio (CLR) test [see Moreira (2003)] based on the whole sample. But there is another way to look at it – with randomly missing data, the split-sample score test can actually use more information than the K-test or the CLR test

mentioned above. For example, consider the familiar IV regression model:

$$\begin{aligned} y &= Y\theta + u \\ Y &= Z\Pi + V \end{aligned}$$

where u and V are the unobserved correlated structural errors, Y is the set of endogenous regressors and Z is the set of instrumental variables. Let θ be the parameter of interest and suppose that all the regularity conditions are satisfied. Also, suppose that we have n independent and identically distributed observations on Y and Z , but only $n_1 < n$ observations on y where $n_2 = n - n_1$ observations of y are randomly missing.¹ In this situation the K-test and the CLR-test, in their present form, can use only n_1 of the observations on all the variables and the remaining n_2 observations on Y and Z are wasted, whereas the split-sample score test actually uses all the information available. See Angrist and Krueger (1992) for another example where sample splitting (in other words, using two different samples) is a reasonable option. The structure of the split-sample score test does not allow for a gain in power under the above situation. However, Monte-Carlo experiments (not reported in this paper) show that if a majority of the observations on y are missing, the match between the nominal and the asymptotic size is possibly better for the split-sample score test than it is for the AR-test, the K-test or the CLR-test based on sub-sample one alone (i.e. the complete observations).

The present paper extends the split-sample score test to subsets of parameters. Suppose that in the IV model described above, $\theta = (\beta', \gamma')'$ and $Y = (X, W)$, i.e. β is the coefficient associated with the endogenous regressor X and γ is the coefficient associated with the other endogenous regressor W . For example, if we are interested in the returns to schooling, X can be considered as the years of schooling, W as the years of experience and y as logarithm of the wage earned by the individual. Both X and W presumably depend on the unobserved ability of the individual which is probably correlated with the individual's earning, and hence X and W can be argued to be endogenous. For simplicity, we do not mention the exogenous control variables included in the regression. As mentioned above, β is the coefficient of interest. If

¹If we think of each observation on y being included independently in the sample with probability p , then by "randomly missing observations" we mean that p is determined exogenously.

the instruments Z are “weak” for β , we cannot estimate β consistently but it is still possible to validly test null hypotheses of the form $H_0 : \beta = \beta_0$ using projection test based on the AR-statistic or the K-statistic or the LR-statistic [see Staiger and Stock (1997), Kleibergen (2002) and Moreira (2003)].² However, projection tests of $H_0 : \beta = \beta_0$ can be very conservative when the instruments are “strong” for either γ or β .

Kleibergen (2004) proposes an alternative way of using the K-statistic to obtain a valid test of $H_0 : \beta = \beta_0$ when the instruments are strong for γ ; Zivot et al. (2006) call this the “partial” K-test. The partial K-statistic replaces γ by its limited-information maximum likelihood (LIML) estimator (constrained by $H_0 : \beta = \beta_0$) in the K-statistic and adjusts the degree of freedom of the asymptotic distribution accordingly. The partial K-test is more powerful than the projection test based on the K-statistic and the partial K-statistic is pivotal when the instruments are strong for γ . Given that strong instruments are not easy to find, it is not advisable to simply assume that the instruments are strong for γ . Hence a reliable use of the partial K-test must be accompanied by a pretest for the strength of instruments for γ and possibly with proper size-correction thereafter [see Hall et al. (1996)].³ However, similar to Guggenberger and Smith (2005) and Zivot et al. (2006), we do not find any evidence of major upward-size distortion of the partial K-test but we do observe that the size of the partial K-test increases with the level of endogeneity of W (i.e. the regressor associated with the structural coefficient γ).

In the present paper we propose a new test, the Robins-test, which is valid irrespective of the strength of the instruments, for either β or γ , under weak assumptions about the underlying data generating process. When all the standard regularity conditions are satisfied, the power of this test is asymptotically equivalent to the power of the standard score test (based on

²In the rest of the paper “valid tests” is synonymous to “tests that are not over-sized”. If the instruments are “weak” for any structural coefficient in the sense of Staiger and Stock (1997), then the coefficient is asymptotically unidentified. However, strictly speaking, “strong” instrument for any structural coefficient in the sense of Staiger and Stock (1997) does not necessarily mean that the parameter is identified [see WI Case 3 in Zivot et al. (2006)]. In the present paper we do not consider this case and use the term “strong” instrument loosely to mean that the corresponding structural coefficient is identified.

³However, to our knowledge, there does not exist any systematic way of pretesting the strength of instruments in the presence of multiple endogenous regressors. A widely used (abused) measure of instrumental strength is the partial R^2 statistic proposed by Shea (1997). Zivot and Chaudhuri (2007) point out that the partial R^2 statistic for γ , in this case, can be totally misleading if the instruments are weak for β .

sub-sample one). When the rank condition is not satisfied, the standard score test is invalid, whereas, our method can still be reliably used for inference on subsets of structural coefficients. We also show that our method is computationally easy to implement and we recommend its use for testing subsets of structural coefficients in linear IV models. As shown by Zivot et al. (1998), we can also form a confidence interval (of any desired level and probably conservative) for β as the collection of values β_0 which cannot be rejected by our test (at the desired level). Again, if all the standard regularity conditions are satisfied, any point belonging to such a confidence interval (of non-trivial level) will belong to the \sqrt{n} -neighborhood of the true value of the structural coefficients with probability one. So like the standard tests, our method can also be used for consistent interval-estimation of subsets of structural coefficients when the standard regularity conditions are satisfied.

The main contributions of this paper are summarized below:

- First we propose a split-sample score test for the null hypothesis $H_0 : \beta = \beta_0$ which is valid when the instruments are strong for γ . Let $\hat{\gamma}(\beta_0)$ be the Unbiased-Split-Sample-Instrumental-Variables (USSIV) estimator of γ constrained by the null hypothesis $H_0 : \beta = \beta_0$ [see Angrist and Krueger (1995)]. We define a test which rejects the null hypothesis at level α if the split-sample score statistic evaluated at $\beta = \beta_0$ and $\gamma = \hat{\gamma}(\beta_0)$ exceeds the $(1 - \alpha)100$ -th quantile of the central χ^2 distribution with degrees of freedom equal to the dimension of β [see Dufour and Jasiak (2001)]. We call this the USSIV score test. When the instruments are strong for γ , the USSIV score test for β is valid and asymptotically equivalent to the partial K-test (based on sub-sample one) against \sqrt{n} -local alternatives. When the instruments are weak for γ , we do not have control over the size of the USSIV score test. The USSIV score test serves the purpose of a benchmark for split-sample score tests (or tests based on sub-sample one) of $H_0 : \beta = \beta_0$ when the instruments are strong for γ .
- Finally, we design a new testing procedure for the null hypothesis $H_0 : \beta = \beta_0$ which is always valid. When the instruments are strong for γ , it is asymptotically equivalent to the USSIV score test. To our knowledge this kind of test was first proposed by Robins (2004) and hence we call it the Robins-test. The Robins-test is quite generally more

powerful than the projection test based on the split-sample score statistic proposed by Dufour and Jasiak (2001).

The rest of the paper is organized as follows: Section 2 discusses the model and the weak instrument framework, Section 3 describes testing of hypotheses on subsets of parameters using the USSIV score test and Robins-test, Section 4 is a Monte-Carlo study investigating the finite sample behaviors of the USSIV score test and Robins-test, we conclude in Section 5.

We use the following notations throughout. Consider any $n \times m$ matrix A . If A has full column rank then $P_A = A(A'A)^{-1}A'$ and $M_A = I_n - P_A$. If A is a symmetric, positive semidefinite matrix then $A = A^{\frac{1}{2}}A^{\frac{1}{2}'} where $A^{\frac{1}{2}}$ is the lower-triangular Cholesky factor of A .$

2 Linear IV Model with Weak Instruments

2.1 Model:

Suppose that we have the following structural equation model:

$$y = X\beta + W\gamma + u \quad (1)$$

$$X = Z\Pi_X + V_X \quad (2)$$

$$W = Z\Pi_W + V_W \quad (3)$$

where y is the dependent variable, X and W are the endogenous regressors, Z is the instrument and u , V_X and V_W are the unobserved correlated structural errors.⁴ Let the dimensions of β , γ , Π_X and Π_W be respectively $m_x \times 1$, $m_w \times 1$, $k \times m_x$ and $k \times m_w$. Let $m = m_x + m_w$ and m, m_x, m_w and k be fixed and finite numbers. We assume that the order condition $k \geq m$ is satisfied. We do not, however, impose the restriction of full column rank on $\Pi = [\Pi_X, \Pi_W]$.

Suppose that we have n observations on y , X , W and Z and we randomly split the sample

⁴For simplicity, we leave out the included exogenous variables. Adding them in the model does not entail any fundamental change in our results because it is possible to find \sqrt{n} -consistent estimators for the corresponding coefficients when the true values of β and γ are known.

in two parts – the first part containing $n_1 = [n\zeta]$ observations and the second part containing $n_2 = n - n_1$ observations where ζ is a fixed number in the interval $(0, 1)$. Let y_i , X_i , W_i and Z_i represent the n_i observations in the i th sub-sample ($i = 1, 2$) where the observations are stacked in rows. To motivate alternatively, we can also assume that y_2 is missing and hence the AR-test, the K-test and the CLR test, in their present form, can only be performed based on the first sub-sample. The asymptotic results discussed in this paper do not depend on whichever motivation is used – sample splitting or missing y_2 . However, for the purpose of the simulations we use the missing y_2 scenario.

We make the following high level assumptions summarized under Assumption A.

Assumption A: We assume that the following results hold jointly as $n \rightarrow \infty$ for $i = 1, 2$:

1. $\frac{1}{n_i} (u_i, V_{Xi}, V_{Wi})' (u_i, V_{Xi}, V_{Wi}) \xrightarrow{P} \Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{uX} & \sigma_{uW} \\ \sigma_{Xu} & \sigma_{XX} & \sigma_{XW} \\ \sigma_{Wu} & \sigma_{WX} & \sigma_{WW} \end{pmatrix}$ where Σ is a symmetric, positive definite matrix.
2. $\frac{1}{n_i} Z_i' Z_i \xrightarrow{P} Q$ where Q is a symmetric, positive definite matrix.
3. $\frac{1}{\sqrt{n_i}} Z_i' (u_i, V_{Xi}, V_{Wi}) \xrightarrow{d} Q^{\frac{1}{2}} (\Psi_{Zui}, \Psi_{ZXi}, \Psi_{ZW_i})$ where $vec(\Psi_{Zui}, \Psi_{ZXi}, \Psi_{ZW_i}) \sim N(0, \Sigma \otimes I_k)$.
4. Finally, we assume that $\Psi_{Zu1}, \Psi_{ZX1}, \Psi_{ZW1}$ are uncorrelated with Ψ_{ZX2} and Ψ_{ZW2} .

See Staiger and Stock (1997) and Kleibergen (2002) for discussions on the first three assumptions. The fourth assumption ensures that the random functions computed from sub-sample one are asymptotically uncorrelated with those computed from sub-sample two.

2.2 Weak Instrument Framework:

The weak-instrument framework proposed by Staiger and Stock (1997) models Π as $O(n^\delta)$ where $\delta = 0$ means that the instruments are strong and $\delta = -\frac{1}{2}$ means that they are weak for the corresponding structural coefficients. The framework ties the instrument-strength to the sample size in a way which ensures that as the sample size goes to infinity the instru-

ments become completely irrelevant at a rate that gives non-trivial asymptotic properties. We use the weak-instrument framework to characterize four cases of partial identification under Assumption A [see Zivot et al. (2006), Choi and Phillips (1992) and Phillips (1989)] as given below:

- Case 1: $\Pi_X = \frac{C_X}{\sqrt{n}}$ and $\Pi_W = \frac{C_W}{\sqrt{n}}$ i.e. both β and γ are asymptotically unidentified.
- Case 2: $\Pi_X = \frac{C_X}{\sqrt{n}}$ and $\Pi_W = C_W$ i.e. β is asymptotically unidentified but γ is identified.
- Case 3: $\Pi_X = C_X$ and $\Pi_W = \frac{C_W}{\sqrt{n}}$ i.e. β is identified but γ is asymptotically unidentified.
- Case 4: $\Pi_X = C_X$ and $\Pi_W = C_W$ i.e. both β and γ are identified.

where $C = [C_X, C_W]$ is a full-column-rank, non-random matrix with fixed and bounded elements. This is by no means an exhaustive list of all possible cases, but is sufficiently rich to highlight the interesting asymptotic results.

We also define $\Lambda = (\lambda_X, \lambda_W) = Q^{\frac{1}{2}}C$ and note that under the special case $\lambda'_X \lambda_W = 0$ and $\sigma_{XW} = 0$, Zivot et al. (2006) define the so-called concentration parameter (measuring the instrument strength) for β and γ respectively as

$$\mu_\beta = \frac{n^{(1+2\delta_x)}}{k} \sigma_{XX}^{-\frac{1}{2}'} \lambda'_X \lambda_X \sigma_{XX}^{-\frac{1}{2}} \quad \text{and} \quad \mu_\gamma = \frac{n^{(1+2\delta_w)}}{k} \sigma_{WW}^{-\frac{1}{2}'} \lambda'_W \lambda_W \sigma_{WW}^{-\frac{1}{2}}$$

where $\delta_x = -\frac{1}{2}$ under Cases 1 and 2 and $\delta_x = 0$ under Cases 3 and 4. Similarly $\delta_w = -\frac{1}{2}$ under Cases 1 and 3 and $\delta_w = 0$ under Cases 2 and 4. However, in the rest of the paper we do not impose the restrictions $\lambda'_X \lambda_W = 0$ and $\sigma_{XW} = 0$ (except in Section 5: simulation study).

3 Testing the Null Hypothesis on Subsets of Parameters

Without loss of generality we treat β as the parameters of interest and γ as the nuisance parameters.

3.1 Why Split the Sample?

We follow the exposition by Wang and Zivot (1998) to motivate sample-splitting (or equivalently using information from both sub-samples). The Wang and Zivot score statistic for jointly testing $\beta = \beta_0$ and $\gamma = \gamma_0$, based on sub-sample one, uses the gradients of the following objective function with respect to β and γ

$$\max_{\beta, \gamma} J(\beta, \gamma) = -\frac{1}{2}(y_1 - X_1\beta - W_1\gamma)'P_{Z_1}(y_1 - X_1\beta - W_1\gamma).$$

The gradients with respect to β and γ are, respectively, given by

$$\nabla_{\beta} J_{11}(\beta, \gamma) = \hat{\Pi}'_{X1} Z'_1 (y_1 - X_1\beta - W_1\gamma) \quad (4)$$

$$\nabla_{\gamma} J_{11}(\beta, \gamma) = \hat{\Pi}'_{W1} Z'_1 (y_1 - X_1\beta - W_1\gamma) \quad (5)$$

where $\hat{\Pi}_{Xi} = (Z'_i Z_i)^{-1} Z'_i X_i$ and $\hat{\Pi}_{Wi} = (Z'_i Z_i)^{-1} Z'_i W_i$ are the first-stage estimators of Π_X and Π_W based on sub-sample i ($= 1, 2$). The subscript on ∇J is used to distinguish (4) and (5) from (8) and (9) respectively.

Let $\nabla J_{11}(\beta, \gamma) = [\nabla_{\beta} J'_{11}(\beta, \gamma), \nabla_{\gamma} J'_{11}(\beta, \gamma)]'$ and $\hat{\Pi}_i = [\hat{\Pi}_{Xi}, \hat{\Pi}_{Wi}]$ for $i = 1, 2$. The Wang and Zivot score statistic is defined as

$$\mathcal{WZ}(\beta_0, \gamma_0) = \frac{\nabla J'_{11}(\beta_0, \gamma_0) \left(\hat{\Pi}'_1 Z'_1 Z_1 \hat{\Pi}_1 \right)^{-1} \nabla J_{11}(\beta_0, \gamma_0)}{\frac{1}{n_1} (y_1 - X_1\beta_0 - W_1\gamma_0)' (y_1 - X_1\beta_0 - W_1\gamma_0)} \quad (6)$$

and the score test rejects the hypotheses $\beta = \beta_0$ and $\gamma = \gamma_0$ jointly at level α if $\mathcal{WZ}(\beta_0, \gamma_0) > \chi_m^2(1 - \alpha)$ where $\chi_m^2(1 - \alpha)$ is the $(1 - \alpha)100$ -th quantile of the central χ_m^2 distribution. Wang and Zivot (1998) point out that under Case I, if β_0 and γ_0 are the true values, then

$$\mathcal{WZ}(\beta_0, \gamma_0) \xrightarrow{d} \frac{1}{\sigma_{uu}} \Psi'_{Zu1} P_{[\lambda_X + \Psi_{ZX1}, \lambda_W + \Psi_{ZW1}]} \Psi_{Zu1}$$

which is not a (central) χ_m^2 distribution because Ψ_{Zu1} is correlated with Ψ_{ZX1} and Ψ_{ZW1} . The problem arises because $\frac{1}{\sqrt{n_1}} Z'_1 (y_1 - X_1\beta - W_1\gamma)$ and $\hat{\Pi}_1$, used in the expression of $\nabla J_{11}(\beta, \gamma)$, are asymptotically correlated in the presence of weak instruments.

Dufour and Jasiak (2001) replace $\widehat{\Pi}_1$ by $\widehat{\Pi}_2$ in (6) and obtain the split-sample score statistic ⁵

$$LM(\beta_0, \gamma_0) = \frac{\nabla J'_{21}(\beta_0, \gamma_0) \left(\widehat{\Pi}'_2 Z'_1 Z_1 \widehat{\Pi}_2 \right)^{-1} \nabla J_{21}(\beta_0, \gamma_0)}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma_0)' (y_1 - X_1 \beta_0 - W_1 \gamma_0)} \quad (7)$$

where $\nabla J_{21}(\beta, \gamma) = [\nabla_{\beta} J'_{21}(\beta, \gamma), \nabla_{\gamma} J'_{21}(\beta, \gamma)]'$ denotes the new gradient vector using information from both the sub-samples and

$$\nabla_{\beta} J_{21}(\beta, \gamma) = \widehat{\Pi}'_{X2} Z'_1 (y_1 - X_1 \beta - W_1 \gamma) \quad (8)$$

$$\nabla_{\gamma} J_{21}(\beta, \gamma) = \widehat{\Pi}'_{W2} Z'_1 (y_1 - X_1 \beta - W_1 \gamma) \quad (9)$$

It is probably more appropriate to call (8) and (9) the “pseudo-gradients” with respect to β and γ respectively. When the instruments are strong, the pseudo-gradients are asymptotically equivalent to the gradients given in (4) and (5).

When $\beta = \beta_0$ and $\gamma = \gamma_0$, $LM(\beta_0, \gamma_0) \xrightarrow{d} \chi_m^2$. The split-sample score test for jointly testing $\beta = \beta_0$ and $\gamma = \gamma_0$ rejects the hypotheses at level α if $LM(\beta_0, \gamma_0) > \chi_m^2(1 - \alpha)$. The test is valid irrespective of the strength of the instruments.

In a separate Monte-Carlo study (not reported in this paper) we compare the finite sample behavior of the (joint) split-sample score test, the AR-test, the K-test and the CLR-test (based on sub-sample one) for different levels of instrument relevance and endogeneity. When the number of observations in sub-sample one is small, the nominal sizes of the AR-test, the K-test and the CLR-test exceed the asymptotic size. On the other hand, the split-sample score test has correct nominal size even when the number of observations in sub-sample one is small. This is probably not surprising given that the two sub-samples are actually drawn independently under our Monte-Carlo design. However, the fact that the split-sample score test uses more information than the AR, the K and the CLR tests, based on sub-sample one only, does

⁵This is also referred to as the split-sample Anderson-Rubin statistic by Staiger and Stock (1997), as the split-sample statistic by Dufour and Taamouti (2005) and Kleibergen (2002). However, we use the expression of the denominator of the statistic from Wang and Zivot (1998). Since we will restrict ourselves to \sqrt{n} -local alternatives, the choice of denominator is not going to matter asymptotically.

not imply that it is more powerful in finite samples. Asymptotically these tests (except the AR test in the over-identified case) have the same power against \sqrt{n} -local alternatives when the instruments are strong.

In the following sub-sections we exploit the (asymptotic) absence of correlation between the two sub-samples to design split-sample score tests for subsets of structural coefficients i.e. for the null hypothesis $H_0 : \beta = \beta_0$.

3.2 USSIV Score Test:

Given any value β_0 , the USSIV estimator of γ , as defined in Angrist and Krueger (1995), is obtained from (9) by solving $\nabla_\gamma J_{21}(\beta_0, \gamma) = 0$ and is given by

$$\hat{\gamma}(\beta_0) = \left(\hat{\Pi}'_{W2} Z'_1 W_1 \right)^{-1} \hat{\Pi}'_{W2} Z'_1 (y_1 - X_1 \beta_0) \quad (10)$$

Replacing γ_0 by $\hat{\gamma}(\beta_0)$ in (7) we get what we call the USSIV-score statistic. The USSIV score statistic is given by

$$\begin{aligned} LM_\beta(\beta_0) &= \frac{\nabla J'_{21}(\beta_0, \hat{\gamma}(\beta_0)) \left(\hat{\Pi}'_2 Z'_1 Z_1 \hat{\Pi}_2 \right)^{-1} \nabla J_{21}(\beta_0, \hat{\gamma}(\beta_0))}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))' (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))} \\ &= \frac{\nabla_\beta J'_{21}(\beta_0, \hat{\gamma}(\beta_0)) \left(\hat{\Pi}'_{X2} Z'_1 M_{Z_1 \hat{\Pi}_{W2}} Z_1 \hat{\Pi}_{X2} \right)^{-1} \nabla_\beta J_{21}(\beta_0, \hat{\gamma}(\beta_0))}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))' (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))} \end{aligned} \quad (11)$$

where (11) is obtained by using the definition of the USSIV estimator of γ given β_0 . The USSIV score test rejects the null hypothesis $H_0 : \beta = \beta_0$ at level α if $LM_\beta(\beta_0) > \chi^2_{m_x}(1 - \alpha)$.

Theorem 1: When $n \rightarrow \infty$ and $\beta = \beta_0 + \frac{d_\beta}{\sqrt{n}}$, the asymptotic distribution of the USSIV score statistic under Cases 2 and 4 is given by

- In Case 2: $LM_\beta(\beta_0) \xrightarrow{d} \chi^2_{m_x}$.
- In Case 4: $LM_\beta(\beta_0) \xrightarrow{d} \chi^2_{m_x} \left(\frac{\zeta d'_\beta \lambda'_X P_{M_{\lambda_W} \lambda_X} \lambda_X d_\beta}{\sigma_{uu}} \right)$.

We prove Theorem 1 in the Appendix. Theorem 1 says that in Case 2, where the instruments

are weak for β but strong for γ , the USSIV score test cannot distinguish the true β from the \sqrt{n} -local alternatives. In Case 4, where the instruments are strong for both β and γ , the USSIV score test behaves like the standard score test for a subset of structural coefficients in a linear IV model using the whole sample – the only difference being the non-centrality parameter of the asymptotic distribution which is a fraction ζ of the non-centrality parameter of the asymptotic distribution of the standard score statistic based on the whole sample. Theorem 1 does not tell us anything about Cases 1 and 3 where the instruments are weak for γ . Noting that under Cases 1 and 3,

$$\widehat{\gamma}(\beta_0) - \gamma \xrightarrow{d} \left[\left(\lambda_W \sqrt{1 - \zeta} + \Psi_{ZW2} \right)' \left(\lambda_W \sqrt{\zeta} + \Psi_{ZW1} \right) \right]^{-1} \left(\lambda_W \sqrt{1 - \zeta} + \Psi_{ZW2} \right)' \Psi_{Zu1} > o_p(1)$$

when $H_0 : \beta = \beta_0$ is true, it can be shown that the asymptotic distribution of the USSIV score statistic under $H_0 : \beta = \beta_0$ does not follow $\chi_{m_x}^2$ unless the regressor W in (1) is (weakly) exogenous. Similarity between the asymptotic behaviors of the USSIV test and the partial K-test in Cases 2 and 4 follow from Theorem 1.

Now we define a new statistic which helps to explore the asymptotic relationship between the USSIV score test discussed in this section and Robins-test discussed in the next section. Consider any γ_* in the \sqrt{n} -neighborhood of the true γ i.e. let $\gamma = \gamma_* + \frac{d_\gamma}{\sqrt{n}}$. Define $LM_\beta^*(\beta_0, \gamma_*)$ as

$$LM_\beta^*(\beta_0, \gamma_*) = \frac{(y_1 - X_1\beta_0 - W_1\gamma_*)' P_{[M_{Z_1\hat{\Pi}_{W2}} Z_1\hat{\Pi}_{X2}]} (y_1 - X_1\beta_0 - W_1\gamma_*)}{\frac{1}{n_1} (y_1 - X_1\beta_0 - W_1\gamma_*)' (y_1 - X_1\beta_0 - W_1\gamma_*)}. \quad (12)$$

It is easier to motivate the two statistics $LM_\beta(\beta_0)$ and $LM_\beta^*(\beta_0, \gamma_*)$ if we think of the standard IV regression without any weak instrument i.e. when sample splitting is not necessary. In this setting the standard score test based on sub-sample one rejects $H_0 : \beta = \beta_0$ at level α if

$$\mathcal{WZ}_\beta(\beta_0) = \frac{(y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))' P_{[M_{Z_1\hat{\Pi}_{W1}} Z_1\hat{\Pi}_{X1}]} (y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))}{\widehat{\sigma}_{uu}^{2SLS}} > \chi_{m_x}^2(1 - \alpha) \quad (13)$$

where $\widehat{\gamma}_{2SLS}(\beta_0) = \left(\widehat{\Pi}_1' Z_1' Z_1 \widehat{\Pi}_1 \right)^{-1} \widehat{\Pi}_1' Z_1' (y_1 - X_1\beta_0)$ is the constrained two-stage least squares (2SLS) estimator of γ satisfying $\nabla_\gamma J_{11}(\beta_0, \widehat{\gamma}_{2SLS}(\beta_0)) = 0$ and the estimator of the residual variance under the null hypothesis is $\widehat{\sigma}_{uu}^{2SLS} = \frac{(y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))' (y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))}{n_1}$. In

the context of maximum likelihood (for example, if the structural errors jointly follow Gaussian distribution), we can think of the term $\left(\hat{\sigma}_{uu}^{2SLS} \hat{\Pi}'_{X1} Z'_1 M_{Z_1 \hat{\Pi}_{W1}} Z_1 \hat{\Pi}_{X1}\right)^{-1}$ as the top-left $m_x \times m_x$ block in the inverse of the (Hessian-based) estimator of the Information Matrix. i.e. as the estimator of the information bound for estimating β . Equivalently, it is the inverse of the variance estimator of the efficient score function for β evaluated at the $(\sqrt{n}$ -neighborhood of the) true values of β and γ . If we define the efficient score statistic for β as the quadratic form of the sample efficient score function with respect to the inverse of its estimated variance, then the efficient score statistic, based on sub-sample one, and evaluated at β_0 and γ_* is given by

$$\mathcal{WZ}_\beta^*(\beta_0) = \frac{(y_1 - X_1\beta_0 - W_1\gamma_*)' P_{[M_{Z_1 \hat{\Pi}_{W1}} Z_1 \hat{\Pi}_{X1}]} (y_1 - X_1\beta_0 - W_1\gamma_*)}{\frac{1}{n_1} (y_1 - X_1\beta_0 - W_1\gamma_*)' (y_1 - X_1\beta_0 - W_1\gamma_*)}. \quad (14)$$

The USSIV score statistic given in (11) is obtained by replacing $\hat{\Pi}_1$ by $\hat{\Pi}_2$ in the expression of $WZ_\beta(\beta_0)$ in (13) and the USSIV estimator of γ , (i.e. $\hat{\gamma}(\beta_0)$ in (10)) can be thought of as the split-sample version of the 2SLS estimator of γ . Similarly the statistic $LM_\beta^*(\beta_0, \gamma_*)$ given in (12) can be thought of as a split-sample version of the efficient score statistic $WZ_\beta^*(\beta_0, \gamma_*)$ given in (14). Lemma 1 discusses the asymptotic properties of $LM_\beta^*(\beta_0, \gamma_*)$ which are used to prove Theorem 1. It also helps to motivate the new split-sample score test introduced in the next section.

Lemma 1: Let $1_{[\delta_x=0]}$ be a dummy variable taking value one if $\delta_x = 0$ and taking value zero if $\delta_x = -\frac{1}{2}$ where δ_x is such that $\Pi_X = O(n^{\delta_x})$. Similarly define the dummy variable $1_{[\delta_w=0]}$ where δ_w is such that $\Pi_W = O(n^{\delta_w})$. Suppose that $\beta = \beta_0 + \frac{d_\beta}{\sqrt{n}}$ and $\gamma = \gamma_* + \frac{d_\gamma}{\sqrt{n}}$. Also, let $\mathbb{B} = M_{[\lambda_W \sqrt{1-\zeta} + (1-1_{[\delta_w=0]})\Psi_{ZW2}]} \left(\lambda_X \sqrt{1-\zeta} + (1-1_{[\delta_x=0]})\Psi_{ZX2} \right)$ and $\mathbb{D} = [\Psi_{Zu1} + \sqrt{\zeta} (1_{[\delta_x=0]}\lambda_X d_\beta + 1_{[\delta_w=0]}\lambda_W d_\gamma)]$. When $n \rightarrow \infty$ the asymptotic distribution of $LM_\beta^*(\beta_0, \gamma_*)$ is given by

$$LM_\beta^*(\beta_0, \gamma_*) \xrightarrow{d} \frac{\mathbb{D}' P_{\mathbb{B}} \mathbb{D}}{\sigma_{uu}} \quad (15)$$

We prove Lemma 1 in the Appendix. Several points are worth mentioning here.

1. The asymptotic distribution of $LM_\beta^*(\beta, \gamma_*)$ is central $\chi_{m_x}^2$, i.e. at the true value of β the asymptotic distribution does not depend on the strength of the instruments.
2. Under Case 2, the asymptotic distribution of $LM_\beta^*(\beta_0, \gamma_*)$ is again central $\chi_{m_x}^2$.

3. Under Case 4, the asymptotic distribution of $LM_{\beta}^*(\beta_0, \gamma_*)$ is non-central $\chi_{m_x}^2$ with non-centrality parameter given by $\frac{\zeta d'_{\beta} \lambda'_X P_{M_{\lambda_W} \lambda_X} \lambda_X d_{\beta}}{\sigma_{uu}}$.

The above observations suggest asymptotic equivalence between $LM_{\beta}(\beta_0)$ and $LM_{\beta}^*(\beta_0, \gamma_*)$ when instruments are strong for γ . See the proof of Theorem 1 for a formal statement of asymptotic equivalence between $LM_{\beta}(\beta_0)$ and $LM_{\beta}^*(\beta_0, \gamma_*)$ under cases 2 and 4.

Lemma 1 holds for any γ_* in the \sqrt{n} -neighborhood of true γ . Hence, if we can obtain a \sqrt{n} -consistent estimator of γ , Lemma 1 can be used to construct a valid test of the null hypothesis $H_0 : \beta = \beta_0$ regardless of the instrumental relevance. The USSIV score test fails under the weak-instrument asymptotics because even under the null hypothesis the estimator $\hat{\gamma}(\beta_0)$ is not consistent for γ in Cases 1 and 3. It should not, however, be possible to find a \sqrt{n} -consistent estimator of γ when instruments are weak for γ rendering it asymptotically unidentified. Hence it is not possible to find a valid test for $H_0 : \beta = \beta_0$ (unless V_W is uncorrelated with u and V_X) using tests like the USSIV score test or for that matter, any “standard” test. However, as discussed earlier, under Cases 2 and 4 and against \sqrt{n} -local alternatives, the USSIV score test can be used as a benchmark for the tests based only on sub-sample one and the split-sample tests.

3.3 The Robins-Test:

In this section we propose a new test for $H_0 : \beta = \beta_0$ which is always valid and when the instruments are strong it is asymptotically equivalent to the USSIV score test against \sqrt{n} -local alternatives. Our test is based on the general testing procedure proposed by Robins (2004), and we call this the Robins-test. When the instruments are strong for γ , the Robins-test based on the split-sample score function is quite generally less conservative than the projection test based on the split-sample score statistic.

If γ belongs to the parameter space $\Theta_{\gamma} \subseteq \mathbb{R}^{m_w}$, the projection test based on $LM(\beta_0, \gamma_0)$ rejects

the null hypothesis $H_0 : \beta = \beta_0$ at level α if

$$\inf_{\gamma_0 \in \Theta_\gamma} LM(\beta_0, \gamma_0) > \chi_m^2(1 - \alpha) \quad (16)$$

The projection test can be very conservative when the instruments are strong for γ (and β).

Now we describe the Robins-test. Suppose that, given a specific value β_0 , it is possible to construct a $1 - \epsilon$ confidence region for γ and let us denote it by $C_\gamma(1 - \epsilon, \beta_0)$. The Robins-test rejects the null hypothesis $H_0 : \beta = \beta_0$ if either $C_\gamma(1 - \epsilon, \beta_0)$ is empty or if $\inf_{\gamma_0 \in C_\gamma(1 - \epsilon, \beta_0)} LM_\beta^*(\beta_0, \gamma_0) > \chi_{m_x}^2(1 - \alpha)$. Theorem 2 stated below shows that the size of the Robins-test cannot exceed $\alpha + \epsilon$ and when the instruments are strong for γ the test is asymptotically equivalent to the size- α USSIV score test against \sqrt{n} -local alternatives.

Before stating Theorem 2, we note that there exist different methods to construct confidence regions like $C_\gamma(1 - \epsilon, \beta_0)$. A general method of constructing the confidence region is as follows. Let $T(\gamma_0|\beta_0)$ denote a random function (test statistic) evaluated at $\beta = \beta_0$ and $\gamma = \gamma_0$ and suppose that it converges to ξ (which does not depend on any unknown parameters) when β_0 and γ_0 are the true values of β and γ . Then a $1 - \epsilon$ confidence region for γ is given by $C_\gamma(1 - \epsilon, \beta_0) = \{\gamma_0 | T(\gamma_0|\beta_0) \leq \xi(1 - \epsilon)\}$ where $\xi(1 - \epsilon)$ is the $(1 - \epsilon)100$ -th quantile of the distribution of ξ . By definition $C_\gamma(1 - \epsilon, \beta_0)$ contains the true value of γ with probability $1 - \epsilon$ when the null hypothesis $H_0 : \beta = \beta_0$ is true. If y_2 is not missing, the above confidence region can be constructed based on the whole sample otherwise we need to construct it based on sub-sample one. This method can be applied to the AR statistic, the K statistic or the LR statistic which have pivotal asymptotic distributions.⁶ When the instruments are strong for γ , $C_\gamma(1 - \epsilon, \beta_0)$, obtained by inverting the AR-test, the K-test or the CLR-test, can only contain values γ_0 such that $\|\gamma - \gamma_0\| = O\left(n^{-\frac{1}{2}}\right)$ [see Chaudhuri (2007)].

However, we consider a different choice of $T(\gamma_0|\beta_0)$ to construct $C_\gamma(1 - \epsilon, \beta_0)$ in this paper.

⁶For the LR statistic, the asymptotic distribution conditional on an ancillary statistic is pivotal [see Moreira (2003)].

We define

$$C_\gamma(1-\epsilon, \beta_0) = \left\{ \gamma_0 \left| T(\gamma_0|\beta_0) = \frac{\nabla_\gamma J'_{21}(\beta_0, \gamma_0) \left(\hat{\Pi}'_{W2} Z'_1 Z_1 \hat{\Pi}_{W2} \right)^{-1} \nabla_\gamma J_{21}(\beta_0, \gamma_0)}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma_0)' (y_1 - X_1 \beta_0 - W_1 \gamma_0)} \leq \chi_{m_w}^2 (1-\epsilon) \right. \right\}. \quad (17)$$

The only reason behind this choice is to maintain the uniformity of the presentation and we do not make any optimality statement regarding the choice of $C_\gamma(1-\epsilon, \beta_0)$. However, given that the (joint) tests for all structural coefficients (against \sqrt{n} -local alternatives) using the K statistic or the LR statistic (based on sub-sample one) and the split-sample score test have the same asymptotic power under strong instruments, the choice of $C_\gamma(1-\epsilon, \beta_0)$ should not matter at least when the instruments are strong for γ . Lemma 2 summarizes the relevant asymptotic properties of $C_\gamma(1-\epsilon, \beta_0)$.

Lemma 2: Let $1_{[\delta_x=0]}$ be a dummy variable taking value one if $\delta_x = 0$ and zero if $\delta_x = -\frac{1}{2}$ where δ_x is such that $\Pi_X = O(n^{\delta_x})$. Similarly define the dummy variable $1_{[\delta_w=0]}$ where δ_w is such that $\Pi_W = O(n^{\delta_w})$. Suppose that $\beta = \beta_0 + \frac{d_\beta}{\sqrt{n}}$ and $\gamma = \gamma_* + \frac{d_\gamma}{\sqrt{n}}$. Also, for notational simplicity let us define $\mathbb{G} = \lambda_W \sqrt{1-\zeta} + (1 - 1_{[\delta_w=0]}) \Psi_{ZW2}$ and $\mathbb{H} = \lim_{n \rightarrow \infty} \left[\Psi_{Zu} + \sqrt{\zeta} 1_{[\delta_x=0]} \lambda_X d_\beta + \left([(1 - 1_{[\delta_w=0]}) + 1_{[\delta_w=0]} \sqrt{n}] \sqrt{\zeta} \lambda_W + \Psi_{ZW1} \right) (\gamma - \gamma_0) \right]$. When $n \rightarrow \infty$,

$$\begin{aligned} T(\gamma_0|\beta_0) &= \frac{\nabla_\gamma J'_{21}(\beta_0, \gamma_0) \left(\hat{\Pi}'_{W2} Z'_1 Z_1 \hat{\Pi}_{W2} \right)^{-1} \nabla_\gamma J_{21}(\beta_0, \gamma_0)}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma_0)' (y_1 - X_1 \beta_0 - W_1 \gamma_0)} \\ &\xrightarrow{d} \frac{\mathbb{H}' P_{\mathbb{G}} \mathbb{H}}{\sigma_{uu} + (\gamma - \gamma_0)' [1_{[\delta_w=0]} \lambda'_W \lambda_W + \sigma_{WW}] (\gamma - \gamma_0) + 2\sigma_{uW}(\gamma - \gamma_0)} \end{aligned} \quad (18)$$

We prove Lemma 2 in the Appendix. We list below those properties of $C_\gamma(1-\epsilon, \beta_0)$ which are useful for the construction of the Robins-test.

1. It is clear from (18) that $T(\gamma_0|\beta_0) \xrightarrow{d} \chi_{m_w}^2$ when β_0 and γ_0 are the true values of β and γ . Hence under the null hypothesis $H_0 : \beta = \beta_0$, the confidence region $C_\gamma(1-\epsilon, \beta_0)$ contains the true value of γ with probability $1-\epsilon$.
2. When the instruments are strong for γ , the asymptotic distribution in (18) is that of a

non-central $\chi_{m_w}^2$ whose non-centrality parameter is finite iff $\|\gamma - \gamma_0\| = O(n^{-\frac{1}{2}})$. Hence for $\beta_0 = \beta - \frac{d_\beta}{\sqrt{n}}$, the confidence region $C_\gamma(1 - \epsilon, \beta_0)$ can contain, with positive probability, only those values that are in a \sqrt{n} -neighborhood of the true γ .

Now we state the main result of this paper, Theorem 2.

Theorem 2: The Robins-test which rejects $H_0 : \beta = \beta_0$ if either $C_\gamma(1 - \epsilon, \beta_0)$ given in (17) is empty or if $\inf_{\gamma_0 \in C_\gamma(1 - \epsilon, \beta_0)} LM_\beta^*(\beta_0, \gamma_0) > \chi_{m_x}^2(1 - \alpha)$ has the following properties. As $n \rightarrow \infty$:

1. The size of the Robins-test never exceeds $\alpha + \epsilon$.
2. When γ is identified, the Robins-test is at least as powerful as the projection test described in (16) and it is also asymptotically equivalent to the USSIV score test against \sqrt{n} -local alternatives.

We prove Theorem 2 in the Appendix. When the instruments are strong for γ , Theorem 2 states that the Robins-test is asymptotically equivalent to the USSIV score test, our benchmark, against \sqrt{n} -local alternatives. However, unlike the USSIV score test, we have control over the size of Robins-test when the instruments are weak for γ .

Theorem 2 also implies that the confidence region for β obtained by inverting the Robins-test (i.e. the collection of all β_0 's which cannot be rejected) has at least $\alpha + \epsilon$ coverage probability asymptotically and when the instruments are strong for γ , this confidence region is asymptotically less conservative than the projection confidence region based on the joint split-sample score test of β and γ . Since, under weak instruments, any confidence region for β is unbounded with positive probability making the expected length of the confidence region infinite, it is not possible to analytically compare the expected length of the confidence regions obtained from the projection test and Robins-test.⁷

Following Dufour and Taamouti (2005), we also show that (a possibly) infinite grid search is not required to perform the Robins-test. The rejection rule for the Robins-test boils down

⁷See Mikusheva (2006) for analytical comparison of the expected arc-length of the confidence regions (expressed in spherical co-ordinates) for scalar β in a related context.

to testing quadratic inequalities in terms of γ [see Appendix]. Similarly we can construct the confidence region for β based on the Robins-test by analytically solving quadratic inequalities in terms of β and γ . However, we note that the existence of analytical methods for solving the inequalities exist because the statistics $LM_{\beta}^*(\beta_0, \gamma_0)$ and $T(\gamma_0|\beta_0)$ are ratios of quadratics in terms of β_0 and γ_0 and it may not be possible to avoid grid search when the Robins-test is applied to other statistics.

4 Finite Sample Behavior of Split-Sample Score Tests

In this section we perform Monte-Carlo experiments to study the finite sample behavior of the different score tests under different levels of instrument relevance and endogeneity. Our Monte-Carlo design closely follows Zivot et al. (2006). We describe below the Monte-Carlo Design and the data generating process for the model described in (1) – (3).

4.1 Monte-Carlo Design and Parameter Specifications:

The structural errors $[u, V_X, V_W]$ are generated by drawing n independent random samples from $N_3(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho_{uX} & \rho_{uW} \\ \rho_{Xu} & 1 & 0 \\ \rho_{Wu} & 0 & 1 \end{pmatrix} \quad (19)$$

If V_X and V_W are correlated, the level of endogeneity of the regressor X depends on the correlations between V_X and u , V_X and V_W and V_W and u . Our choice of Σ in (19) simplifies the set-up by ensuring that the level endogeneity of X depends only on the correlation between V_X and u and similarly the level of endogeneity of W depends only on the correlation between V_W and u . Because $\rho_{XW} = 0$, the overall endogeneity of the model can be measured by the quantity $\rho_{uX}^2 + \rho_{uW}^2$. We make three different choices for the pair $(\rho_{uX}, \rho_{uW}) = (0.5, 0.5)$, $(0.1, 0.99)$ and $(0.99, 0.1)$. X and W are moderately (and equally) endogenous in the first case, X is highly endogenous and W is mildly endogenous in the second case, X is mildly endogenous and W is highly endogenous in the third case.⁸

⁸For positive-definiteness of Σ we need $\rho_{uX}^2 + \rho_{uW}^2 < 1$.

The instruments Z are generated by drawing n independent random samples from $N_k(0, Q)$ independently of the structural errors. For simplicity we choose $Q = I_k$ where $k = 4$ is chosen arbitrarily.

To our knowledge, there does not exist a universally accepted measure of instrumental relevance for a particular structural coefficient in a linear IV model with more than one endogenous regressor. However, for a model with a single endogenous regressor, the instruments are considered weak for the structural coefficient if the concentration parameter is less than 10 [see Staiger and Stock (1997)]. We follow Zivot et al. (2006) and impose the restriction(s) $\lambda'_X \lambda_W = 0$ (and $\sigma_{XW} = 0$) such that the concentration matrix given in (20) is diagonal where the first diagonal element corresponds to the concentration parameter for β and the second one to the concentration parameter for γ . Defining $\delta_x = -\frac{1}{2}$ for Cases 1 and 2 and $\delta_x = 0$ for Cases 3 and 4 and similarly $\delta_w = -\frac{1}{2}$ for Cases 1 and 3 and $\delta_w = 0$ for Cases 2 and 4, the concentration matrix is given by

$$\mu = \frac{1}{k} \begin{pmatrix} \sigma_{XX} & \sigma_{XW} \\ \sigma_{WX} & \sigma_{WW} \end{pmatrix}^{\frac{1}{2}'} \begin{pmatrix} n^{\frac{1+2\delta_x}{2}} \lambda'_X \\ n^{\frac{1+2\delta_w}{2}} \lambda'_W \end{pmatrix} \begin{pmatrix} n^{\frac{1+2\delta_x}{2}} \lambda'_X \\ n^{\frac{1+2\delta_w}{2}} \lambda'_W \end{pmatrix}' \begin{pmatrix} \sigma_{XX} & \sigma_{XW} \\ \sigma_{WX} & \sigma_{WW} \end{pmatrix}^{\frac{1}{2}} = \begin{pmatrix} \mu_\beta & 0 \\ 0 & \mu_\gamma \end{pmatrix} \quad (20)$$

We choose $\Pi = [\Pi_X, \Pi_W]$ such that $\Pi'_X \Pi_W = 0$ and such that $\mu_\beta = 1$ when the instruments are weak for β and $\mu_\beta = 10$ when the instruments are strong for β . Similarly $\mu_\gamma = 1$ when the instruments are weak for γ and $\mu_\gamma = 10$ when the instruments are strong for γ . In particular, the i th element of Π_X is taken as $\sqrt{\frac{\mu_\beta}{n}}$ and the i th element of Π_W is taken as $(-1)^i \sqrt{\frac{\mu_\gamma}{n}}$ for $i = 1, \dots, k$.⁹

We choose the structural coefficients $\beta = 1$ and $\gamma = 10$. We take the sample size $n = 100$ and randomly split the sample into two sub-samples where the first sub-sample contains $n_1 = \lceil n\zeta \rceil$ observations and the second sub-sample contains $n_2 = n - n_1$ observations. Finally, following the missing y_2 motivation, we assume that y_2 is not observable (delete y_2 from sub-sample two). The results reported below are based on 10,000 Monte-Carlo trials. The instrument

⁹In Section 4.3, we also consider strong instrument for β (γ) characterized by $\mu_\beta = 100$ ($\mu_\gamma = 100$).

matrix Z is kept fixed over the 10,000 trials.

The null hypothesis of interest is $H_0 : \beta = \beta_0$ and we compare the finite sample behaviors of the USSIV score test, the Robins-test and the projection test based on the split-sample score statistic under the null and alternative hypotheses.¹⁰ We also compare the finite sample behaviors of these tests with two other tests considered by Zivot et al. (2006) – the partial K-test and the projection test based on the AR statistic [see Appendix]. The partial K-test and the projection test based on the AR statistic are performed based on sub-sample one alone, so that power of all the tests (except the AR-test) are asymptotically equal when the instruments are strong for both β and γ .

4.2 Rejection Rates when the Null Hypothesis is True:

Tables 1 and 2 summarize the nominal size of the above five tests for different levels of endogeneity, instrument relevance, critical values and proportion of observations in sub-sample one. As discussed before, none of the tests are over-sized when the instruments are strong for γ (i.e. under Cases 2 and 4). Under Cases 1 and 3, the USSIV score test over-rejects the null hypothesis when it is true. The rate of over-rejection of the USSIV score test also increases with the level of endogeneity of W . This should not be surprising because the asymptotic-bias of the constrained USSIV estimator of γ increases with ρ_{uW} (the asymptotic bias does not depend on the level of endogeneity of X and under the null hypothesis it does not depend on the correlation between V_X and V_W). The projection tests based on the AR and the split-sample score statistics are conservative. Theorem 2 gives an upper bound of $\alpha + \epsilon$ for the level of the Robins-test when the instruments are weak for γ . However, simulation results indicate that the upper bound is overly conservative and the nominal size of the Robins-test does not exceed α . Unlike the USSIV score test, the partial K-test is not over-sized even when the instruments are weak for γ . Similar to the findings of Guggenberger and Smith (2005) and Zivot et al. (2006), our simulation results show that the partial K-test is, in fact, size-distorted downward when the instruments are weak for γ . [Chaudhuri (2007) shows how the downward size-distortion of the partial K-test affects its power.]

¹⁰For the Robins-test we always choose $\epsilon = \alpha$.

[INSERT TABLE-1 AND TABLE-2 HERE.]

4.3 Rejection Rates when the Null Hypothesis is False:

Figures 1 - 3 plot the (nominal) power curves of the different tests using the 5% critical values under different specifications of the error covariance. In Figure 1, both X and W are moderately endogenous, in Figure 2, X is mildly endogenous whereas W is highly endogenous and in Figure 3, X is highly endogenous whereas W is mildly endogenous. We choose $\zeta = 75\%$ without loss of generality. Other (non-extreme and reasonable) choices of ζ do not change the results (Figures 7 - 12 use $\zeta = 25\%$ and Figures 13 - 18 use $\zeta = 50\%$). To highlight the fact that when the instruments are weak for γ , the high power of the USSIV score test comes at the cost of its upward size-distortion and the validity of the Robins-test and projection test based on split-sample score statistic comes at the cost of their low power, we choose not to plot the size-adjusted powers.

[INSERT FIGURES 1 – 3 HERE.]

The projection test based on the split-sample score statistic and the Robins-test are extremely conservative when the instruments are weak for γ . The USSIV score test has high power in all cases, but because of its upward size-distortion under Cases 1 and 3, it cannot be reliably used in practice. The partial K-test and the projection test based on the AR statistic are not over-sized and at the same time are more powerful than the other tests. However, the power of neither of these two tests dominate each other uniformly. Theorem 2 shows that when instruments are strong for γ , the Robins-test is asymptotically equivalent to the USSIV score test and hence more powerful than the projection test based on the split-sample score statistic. This result cannot be verified by simulations when “strong-instruments” for γ is characterized by $\mu_\gamma = 10$. However, when we consider strong instruments by taking the corresponding concentration parameter to be 100, the claims of Theorem 2 are verified. We note that the characterization of strong instrument, by taking the concentration parameter at least 10, is proposed by Staiger and Stock (1997) and Stock and Yogo (2005) to ensure that

the relative bias of the 2SLS estimator with respect to that of the OLS estimator (of the corresponding structural parameter) does not exceed 10%. This characterization need not be appropriate under our framework. In Figures 4 - 6, we plot the same power curves taking the concentration parameter to be 100 when the corresponding instruments are strong.

[INSERT FIGURES 4 – 6 HERE.]

In Figures 4 – 6, the power of the Robins-test dominates the power of the projection test based on the split-sample score statistic when the instruments are strong for γ . When the instruments are weak for γ , powers of both these tests are close to zero and neither test can distinguish the true value of β from the false ones. However, these tests are valid unlike the USSIV score test which is over-sized when the instruments are weak for γ . We also note that when the instruments are strong for both β and γ , the powers of the Robins-test and the projection test based on the split-sample score statistic are very close to that of the partial-K test and the USSIV score test, and are greater than that of the projection test based on the AR-statistic.

The simulations as a whole encourage the use of the partial K-test and the projection test based on the AR statistic. This is similar to the conclusions of Zivot et al. (2006). We also note that the instruments for γ have to be very strong for the asymptotic equivalence between the USSIV score test and the Robins-test to hold. However, the Robins-test and the projection test based on the split-sample score statistic could possibly be used reliably without over-rejecting the true parameter value of β under all the cases discussed in this paper.

5 Conclusion

In the present paper we show how to construct valid tests for subsets of structural coefficients by splitting the sample in two parts or, in other words, by combining information from two “un-related” samples one of which need not contain information on the dependent variable. The USSIV score test for subsets of structural coefficients ($H_0 : \beta = \beta_0$) against \sqrt{n} -local alternatives is as powerful as the partial K-test (based on sub-sample one) when the instruments are strong for the remaining structural coefficients (γ), but it is severely over-sized

otherwise. On the other hand, the Robins-test is never over-sized and at the same time it is asymptotically as powerful as the USSIV score test when the instruments are strong for the remaining structural coefficients. However, moderate strength of instruments (for example, when the corresponding concentration parameter takes the value 10) may not be enough to ensure the asymptotic equivalence between USSIV score test and the Robins-test. In any case, the Robins-test can be reliably used for testing subsets of structural coefficients in a linear Instrumental Variables model. The projection test based on the (joint) split-sample score statistic is also never over-sized, but can be extremely conservative. The power of Robins-test is more than that of the projection test based on the (joint) split-sample score statistic when the instruments are strong for the remaining structural coefficients. Similar to the finding of Zivot et al. (2006), our simulation results also indicate that, for the sample sizes considered here, the projection test based on the AR statistic and the partial K-test (both based on sub-sample one) are never over-sized and not less powerful than the split-sample tests described in the present paper.

In this paper, we introduced Robins' method for testing hypotheses on subsets of parameters and subsequently constructing valid confidence regions under partial identification. It is a projection-based method that can be substantially less conservative than projections from pivotal statistics for testing the significance of all parameters jointly. The application of Robins' method requires two statistics: a valid test of the parameters of interest when the nuisance parameters are known; a valid confidence set for the nuisance parameters when the parameters of interest are known. In principle, Robin's method can be applied to the linear IV model in a non-split sample context using, for example, the results of Kleibergen (2004), and to nonlinear models estimated by the generalized method of moments using the results of Kleibergen (2005). These extensions are the subject of our future research.

Acknowledgements

Research in this paper was supported by: the Gary Waterman Distinguished Scholar Fund; a seed grant from the Center for Statistics and the Social Sciences at the University of Washington (PI:Zivot); NSF grant DMS 0505865 (PI: Richardson); NIH grant R01 AI032475

(PI:Robins). Saraswata Chaudhuri thanks Professors Charles R Nelson, Richard Startz and the seminar participants at the University of Washington (Economics Department) for their helpful comments and suggestions.

6 Appendix

6.1 Proofs of Results:

It helps to prove Lemma 1 before proving Theorem 1.

Proof of Lemma 1: Using Assumption A, it is easy to see that

$n^{-\delta_x} \hat{\Pi}_{X2} \xrightarrow{d} C_X + (1 - \zeta)^{-\frac{1}{2}} (1 - 1_{[\delta_x=0]}) Q^{-\frac{1}{2}} \Psi_{ZX2}$ and $n^{-\delta_w} \hat{\Pi}_{W2} \xrightarrow{d} C_W + (1 - \zeta)^{-\frac{1}{2}} (1 - 1_{[\delta_w=0]}) Q^{-\frac{1}{2}} \Psi_{ZW2}$. Then it follows directly from Assumption A that

$$\frac{\left[\frac{1}{n_1} \frac{\hat{\Pi}'_{X2}}{n^{\delta_x}} Z'_1 M_{Z_1 \hat{\Pi}_{W2}} Z_1 \frac{\hat{\Pi}_{X2}}{n^{\delta_x}} \right]^{-\frac{1}{2}} \frac{\hat{\Pi}'_{X2}}{n^{\delta_x}} Z'_1 M_{Z_1 \frac{\hat{\Pi}_{W2}}{n^{\delta_w}}} \frac{y_1 - X_1 \beta_0 - W_1 \gamma_*}{\sqrt{n_1}}}{\sqrt{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma_*)' (y_1 - X_1 \beta_0 - W_1 \gamma_*)}} \xrightarrow{d} \frac{(\mathbb{B}' \mathbb{B})^{-\frac{1}{2}'} \mathbb{B}' \mathbb{D}}{\sqrt{\sigma_{uu}}}$$

and hence $LM^*_{\beta}(\beta_0, \gamma_*) \xrightarrow{d} \frac{\mathbb{D}' P_{\mathbb{B}} \mathbb{D}}{\sigma_{uu}}$. [Q.E.D.]

Proof of Theorem 1: We restrict attention to Cases 2 and 4. Using (10) and the proof of Lemma 1, we note that

1. $y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0) = \left[I_{n_1} - W_1 \left(\hat{\Pi}'_{W2} Z'_1 W_1 \right)^{-1} \hat{\Pi}'_{W2} Z'_1 \right] \left[u_1 + Z_1 \Pi_X \frac{d_{\beta}}{\sqrt{n}} + V_{X1} \frac{d_{\beta}}{\sqrt{n}} \right]$
2. $\hat{\Pi}_{W2} \xrightarrow{P} C_W$.

Hence, Assumption A gives that $\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))' (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0)) \xrightarrow{P} \sigma_{uu}$

and

$$\frac{\left[\frac{1}{n_1} \frac{\hat{\Pi}'_{X2}}{n^{\delta_x}} Z'_1 M_{Z_1 \hat{\Pi}_{W2}} Z_1 \frac{\hat{\Pi}_{X2}}{n^{\delta_x}} \right]^{-\frac{1}{2}} \frac{\hat{\Pi}'_{X2}}{n^{\delta_x}} \left[I_k - \frac{Z'_1 W_1}{n_1} \left(\hat{\Pi}'_{W2} \frac{Z'_1 W_1}{n_1} \right)^{-1} \hat{\Pi}'_{W2} \right] \frac{Z'_1}{\sqrt{n_1}} \left[u_1 + Z_1 \Pi_X \frac{d_{\beta}}{\sqrt{n}} + V_{X1} \frac{d_{\beta}}{\sqrt{n}} \right]}{\sqrt{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))' (y_1 - X_1 \beta_0 - W_1 \hat{\gamma}(\beta_0))}} \xrightarrow{d} \frac{(\mathbb{E}' \mathbb{E})^{-\frac{1}{2}'} \mathbb{E}' \mathbb{F}}{\sqrt{\sigma_{uu}}}$$

where $\mathbb{E} = M_{\lambda_W \sqrt{1-\zeta}} \left(\lambda_X \sqrt{1-\zeta} + (1 - 1_{\delta_x=0}) \Psi_{ZX2} \right)$ and $\mathbb{F} = [\Psi_{Zu1} + \sqrt{\zeta} 1_{[\delta_x=0]} \lambda_X d_{\beta}]$. Un-

der Cases 2 and 4 it is easy to see that by definition $\mathbb{B} = \mathbb{E}$ and $\mathbb{D} = \mathbb{F}$ and hence

$$LM_\beta(\beta_0) = LM_\beta^*(\beta_0, \gamma_*) + o_p(1) \quad (21)$$

Direct application of Lemma 1 proves Theorem 1. [Q.E.D.]

Proof of Lemma 2: Using Lemma 1, we get

$$\begin{aligned} & \frac{\left(\frac{\hat{\Pi}'_{W2}}{n^{\delta_w}} \frac{Z'_1 Z_1}{n_1} \frac{\hat{\Pi}_{W2}}{n^{\delta_w}} \right)^{-\frac{1}{2}} \frac{\hat{\Pi}'_{W2}}{n^{\delta_w}} \frac{Z'_1}{\sqrt{n_1}} \left[u_1 + Z_1 \Pi_X \frac{d_\beta}{\sqrt{n}} + V_{X1} \frac{d_\beta}{\sqrt{n}} + Z_1 \Pi_W (\gamma - \gamma_0) + V_{W1} (\gamma - \gamma_0) \right]}{\sqrt{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma_0)' (y_1 - X_1 \beta_0 - W_1 \gamma_0)}} \\ & \xrightarrow{d} \frac{(\mathbb{G}' \mathbb{G})^{-\frac{1}{2}} \mathbb{G}' \mathbb{H}}{\sqrt{\sigma_{uu} + (\gamma - \gamma_0)' [1_{[\delta_w=0]} \lambda'_W \lambda_W + \sigma_{WW}] (\gamma - \gamma_0) + 2\sigma_{uW} (\gamma - \gamma_0)}} \end{aligned}$$

$$\text{Hence } T(\gamma_0 | \beta_0) \xrightarrow{d} \frac{\mathbb{H}' P_{\mathbb{G}} \mathbb{H}}{\sigma_{uu} + (\gamma - \gamma_0)' [1_{[\delta_w=0]} \lambda'_W \lambda_W + \sigma_{WW}] (\gamma - \gamma_0) + 2\sigma_{uW} (\gamma - \gamma_0)}. \quad [\text{Q.E.D.}]$$

Proof of Theorem 2:

Part 1: Using Lemma 2, it is straightforward to see

$$\begin{aligned} & Pr_{\beta_0} \left[\{C_\gamma(1 - \epsilon, \beta_0) = \emptyset\} \cup \left\{ \{C_\gamma(1 - \epsilon, \beta_0) \neq \emptyset\} \cap \left\{ \inf_{\gamma_0 \in C_\gamma(1 - \epsilon, \beta_0)} LM_\beta^*(\beta_0, \gamma_0) > \chi_{m_x}^2(1 - \alpha) \right\} \right\} \right] \\ & \leq Pr_{\beta_0} \left[\{\gamma \notin C_\gamma(1 - \epsilon, \beta_0)\} \cup \left\{ \inf_{\gamma_0 \in C_\gamma(1 - \epsilon, \beta_0) \neq \emptyset} LM_\beta^*(\beta_0, \gamma_0) > \chi_{m_x}^2(1 - \alpha) \right\} \right] \\ & \leq Pr_{\beta_0} [\{\gamma \notin C_\gamma(1 - \epsilon, \beta_0)\} \cup \{LM_\beta^*(\beta_0, \gamma) > \chi_{m_x}^2(1 - \alpha)\}] \\ & \leq Pr_{\beta_0} [\gamma \notin C_\gamma(1 - \epsilon, \beta_0)] + Pr_{\beta_0} [LM_\beta^*(\beta_0, \gamma) > \chi_{m_x}^2(1 - \alpha)] \\ & = \epsilon + \alpha \end{aligned}$$

Part 2: When γ is asymptotically identified, Lemma 2 gives that only values from the \sqrt{n} -neighborhood of the true γ can be contained in $C_\gamma(1 - \epsilon, \beta_0)$ with positive probability. Hence for any β_0 such that $\beta - \beta_0 = \frac{d_\beta}{\sqrt{n}}$, the value $\hat{\gamma}_{\inf}(\beta_0)$ where the $\inf_{\gamma_0 \in C_\gamma(1 - \epsilon, \beta_0)} LM_\beta^*(\beta_0, \gamma_0)$ is attained should be in the \sqrt{n} -neighborhood of the true γ . (21) gives

$$\inf_{\gamma_0 \in C_\gamma(1 - \epsilon, \beta_0)} LM_\beta^*(\beta_0, \gamma_0) = LM_\beta^*(\beta_0, \hat{\gamma}_{\inf}(\beta_0)) = LM_\beta(\beta_0) + o_p(1). \quad (22)$$

The following steps show that the Robins-test is more asymptotically powerful than the projection test when γ is identified.

$$\begin{aligned}
& Pr_{\beta} \left[\inf_{\gamma_0 \in \Theta_{\gamma}} LM(\beta_0, \gamma_0) \leq \chi_m^2(1 - \alpha) \right] \\
& \geq Pr_{\beta} [LM(\beta_0, \hat{\gamma}(\beta_0)) \leq \chi_m^2(1 - \alpha)] \\
& = Pr_{\beta} [LM_{\beta}(\beta_0) \leq \chi_m^2(1 - \alpha)] \\
& \geq Pr_{\beta} [LM_{\beta}(\beta_0) \leq \chi_{m_x}^2(1 - \alpha)] .
\end{aligned}$$

Hence the USSIV score test is more powerful than projection test. Using (22) we can see that the Robins-test is asymptotically more powerful than projection test whenever γ is identified. [Q.E.D.]

6.2 K-Test, Partial K-Test and CLR-Test:

I) K-Test [see Kleibergen (2002)] rejects the null hypothesis $H : \beta = \beta_0, \gamma = \gamma_0$ at level α if

$$K(\beta_0, \gamma_0) = \frac{(y_1 - X_1\beta_0 - W_1\gamma_0)' P_{Z_1\tilde{\Pi}(\beta_0, \gamma_0)}(y_1 - X_1\beta_0 - W_1\gamma_0)}{\frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\gamma_0)' P_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma_0)} > \chi_m^2(1 - \alpha)$$

where $\tilde{\Pi}(\beta_0, \gamma_0) = \begin{bmatrix} \tilde{\Pi}_X & \tilde{\Pi}_W \end{bmatrix}_{(\beta_0, \gamma_0)}$ where $\tilde{\Pi}_X(\beta_0, \gamma_0) = (Z_1'Z_1)^{-1} Z_1' \left[X_1 - (y_1 - X_1\beta_0 - W_1\gamma_0) \frac{\sigma_{Xu}(\beta_0, \gamma_0)}{\sigma_{uu}(\beta_0, \gamma_0)} \right]$, $\tilde{\Pi}_W(\beta_0, \gamma_0) = (Z_1'Z_1)^{-1} Z_1' \left[W_1 - (y_1 - X_1\beta_0 - W_1\gamma_0) \frac{\sigma_{Wu}(\beta_0, \gamma_0)}{\sigma_{uu}(\beta_0, \gamma_0)} \right]$, $\sigma_U(\beta_0, \gamma_0) = \frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\gamma_0)' M_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma_0)$, $\sigma_{Xu}(\beta_0, \gamma_0) = \frac{1}{n_1} X_1' M_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma_0)$ and $\sigma_{Wu}(\beta_0, \gamma_0) = \frac{1}{n_1} W_1' M_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma_0)$.

II) Partial K-Test [see Kleibergen (2004)] rejects the null hypothesis $H_0 : \beta = \beta_0$ at level α if

$$K(\beta_0, \tilde{\gamma}(\beta_0)) = \frac{(y_1 - X_1\beta_0 - W_1\tilde{\gamma}(\beta_0))' P_{Z_1\tilde{\Pi}(\beta_0, \tilde{\gamma}(\beta_0))}(y_1 - X_1\beta_0 - W_1\tilde{\gamma}(\beta_0))}{\frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\tilde{\gamma}(\beta_0))' M_{Z_1}(y_1 - X_1\beta_0 - W_1\tilde{\gamma}(\beta_0))} > \chi_{m_x}^2(1 - \alpha)$$

where $\tilde{\gamma}(\beta_0)$ solve $\tilde{\Pi}'_W(\beta_0, \tilde{\gamma}(\beta_0)) Z_1'(y_1 - X_1\beta_0 - W_1\tilde{\gamma}(\beta_0)) = 0$.

III) CLR-Test [see Moreira (2003)] rejects the null hypothesis $H : \beta = \beta_0, \gamma = \gamma_0$ at level α

if

$$L(\beta_0, \gamma_0) = \frac{1}{2} \left[S'S - T'T + \sqrt{[S'S + T'T]^2 - 4[S'ST'T - (S'T)^2]} \right] > \xi_{LR}(1 - \alpha, T'T = \tau, k)$$

where $S = \frac{1}{\sqrt{\sigma_{uu}(\beta_0, \gamma_0)}} (Z_1' Z_1)^{-\frac{1}{2}} Z_1' (y_1 - X_1 \beta_0 - W_1 \gamma_0)$, $T = (Z_1' Z_1)^{-\frac{1}{2}} Z_1' [y_1, X_1, W_1] \hat{\Omega}^{-1} A (A' \hat{\Omega}^{-1} A)^{-\frac{1}{2}}$, $\hat{\Omega} = \frac{1}{n_1} [y_1, X_1, W_1]' M_{Z_1} [y_1, X_1, W_1]$ and $A = [(\beta_0', \gamma_0')', I_m]$. Finally, $\xi(1 - \alpha, \tau, k)$ is the $(1 - \alpha)100$ -th quantile of the null-distribution of LR_0 given $T'T = \tau$ and the number of instruments k .

6.3 Rejection Rules for Projection Type Tests:

We discuss the rejection rules for different tests testing the null hypothesis $H_0 : \beta = \beta_0$. Following our Monte-Carlo setting, we do it for the special case where γ is scalar. However, it is not hard to extend the results to vector valued γ . Our discussion is based on Dufour and Taamouti (2005).

I) Projection Test based on $AR(\beta_0, \gamma)$ based on sub-sample one rejects $H_0 : \beta = \beta_0$ at level α if

$$\inf_{\gamma \in \Theta_\gamma} AR(\beta_0, \gamma) = \inf_{\gamma \in \Theta_\gamma} \frac{(y_1 - X_1 \beta_0 - W_1 \gamma)' P_{Z_1} (y_1 - X_1 \beta_0 - W_1 \gamma)}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma)' M_{Z_1} (y_1 - X_1 \beta_0 - W_1 \gamma)} > \chi_k^2(1 - \alpha)$$

i.e. if there does not exist any value of γ such that $a_0 \gamma^2 - 2b_0 \gamma + c_0 \leq 0$ where $a_0 = W_1' A_0 W_1$, $b_0 = W_1' A_0 (y_1 - X_1 \beta_0)$, $c_0 = (y_1 - X_1 \beta_0)' A_0 (y_1 - X_1 \beta_0)$ and $A_0 = P_{Z_1} - \frac{1}{n_1} \chi_k^2(1 - \alpha) M_{Z_1}$. Equivalently, reject $H_0 : \beta = \beta_0$ at level α if

$$\{b_0^2 - a_0 c_0 < 0, a_0 > 0\} \cup \{a_0 = b_0 = 0, c_0 > 0\}$$

II) Projection Test based on $LM(\beta_0, \gamma)$ rejects $H_0 : \beta = \beta_0$ at level α if

$$\inf_{\gamma \in \Theta_\gamma} LM(\beta_0, \gamma) = \inf_{\gamma \in \Theta_\gamma} \frac{\nabla J_{21}'(\beta_0, \gamma) \left(\hat{\Pi}_2' Z_1' Z_1 \hat{\Pi}_2 \right)^{-1} \nabla J_{21}(\beta_0, \gamma)}{\frac{1}{n_1} (y_1 - X_1 \beta_0 - W_1 \gamma)' (y_1 - X_1 \beta_0 - W_1 \gamma)} > \chi_m^2(1 - \alpha)$$

i.e. if there does not exist any value of γ such that $a_1\gamma^2 - 2b_1\gamma + c_1 \leq 0$ where $a_1 = W_1' A_1 W_1$, $b_1 = W_1' A_1 (y_1 - X_1 \beta_0)$, $c_1 = (y_1 - X_1 \beta_0)'_1 A_1 (y_1 - X_1 \beta_0)$ and $A_1 = P_{Z_1 \hat{\Pi}_2} - \frac{1}{n_1} \chi_m^2 (1 - \alpha) I_{n_1}$. Equivalently, reject $H_0 : \beta = \beta_0$ at level α if

$$\{b_1^2 - a_1 c_1 < 0, a_1 > 0\} \cup \{a_1 = b_1 = 0, c_1 > 0\}$$

III) Robins-Test rejects $H_0 : \beta = \beta_0$ at level atmost $\alpha(+\epsilon)$ if

$$\{C_\gamma(1 - \epsilon, \beta_0) = \emptyset\} \cup \left\{ \inf_{\gamma \in C_\gamma(1 - \epsilon, \beta_0)} LM_\beta^*(\beta_0, \gamma) > \chi_m^2(1 - \alpha) \right\}$$

i.e. iff $\{C_\gamma(1 - \epsilon, \beta_0) \cap D_\gamma(1 - \alpha, \beta_0)\} = \emptyset$

where $C_\gamma(1 - \epsilon, \beta_0) = \{\gamma_0 | a_2 \gamma_0^2 - 2b_2 \gamma_0 + c_2 \leq 0\}$ and $D_\gamma(1 - \alpha, \beta_0) = \{\gamma_0 | a_3 \gamma_0^2 - 2b_3 \gamma_0 + c_3 \leq 0\}$ and

$$\begin{aligned} a_2 &= W_1' A_2 W_1, & a_3 &= W_1' A_3 W_1 \\ b_2 &= W_1' A_2 (y_1 - X_1 \beta_0), & b_3 &= W_1' A_3 (y_1 - X_1 \beta_0) \\ c_2 &= (y_1 - X_1 \beta_0)'_1 A_2 (y_1 - X_1 \beta_0), & c_3 &= (y_1 - X_1 \beta_0)'_1 A_3 (y_1 - X_1 \beta_0) \\ A_2 &= P_{Z_1 \hat{\Pi}_{W_2}} - \frac{1}{n_1} \chi_{m_w}^2 (1 - \epsilon) I_{n_1}, & A_3 &= P_{[M_{Z_1 \hat{\Pi}_{W_2}} Z_1 \hat{\Pi}_{X_2}]} - \frac{1}{n_1} \chi_{m_x}^2 (1 - \alpha) I_{n_1} \end{aligned}$$

Defining $\Delta_i = b_i^2 - a_i c_i$ for $i = 2, 3$, Robins-test rejects $H_0 : \beta = \beta_0$ at level at most $\alpha + \epsilon$ if any one of the following six mutually exclusive conditions are satisfied:

1. $\{\Delta_i < 0, a_i > 0\} \cup \{a_i = b_i = 0, c_i > 0\}$ for $i = 2$ and/or $i = 3$, i.e. if at least one of the intervals $C_\gamma(1 - \epsilon, \beta_0)$ and $D_\gamma(1 - \alpha, \beta_0)$ is empty.
2. $\left\{ a_i = 0, b_i > 0, a_j = 0, b_j < 0, \frac{c_j}{2b_j} < \frac{c_i}{2b_i} \right\}$, for $i, j = 2, 3$ and $i \neq j$ i.e. if the intervals are of the form $\left[\frac{c_i}{2b_i}, +\infty \right]$ and $\left[-\infty, \frac{c_j}{2b_j} \right]$ where $\frac{c_j}{2b_j} < \frac{c_i}{2b_i}$.
3. $\left\{ a_i = 0, b_i > 0, a_j > 0, \Delta_j \geq 0, \frac{b_j + \sqrt{\Delta_j}}{a_j} < \frac{c_i}{2b_i} \right\}$ for $i, j = 2, 3$ and $i \neq j$ i.e. if the intervals are of the form $\left[\frac{c_i}{2b_i}, +\infty \right]$ and $\left[\frac{b_j - \sqrt{\Delta_j}}{a_j}, \frac{b_j + \sqrt{\Delta_j}}{a_j} \right]$ where $\frac{b_j + \sqrt{\Delta_j}}{a_j} < \frac{c_i}{2b_i}$.
4. $\left\{ a_i = 0, b_i < 0, a_j > 0, \Delta_j \geq 0, \frac{b_j - \sqrt{\Delta_j}}{a_j} > \frac{c_i}{2b_i} \right\}$ for $i, j = 2, 3$ and $i \neq j$ i.e. if the

- intervals are of the form $\left[-\infty, \frac{c_i}{2b_i}\right]$ and $\left[\frac{b_j - \sqrt{\Delta_j}}{a_j}, \frac{b_j + \sqrt{\Delta_j}}{a_j}\right]$ where $\frac{b_j - \sqrt{\Delta_j}}{a_j} > \frac{c_i}{2b_i}$.
5. $\left\{ a_i > 0, \Delta_i \geq 0, a_j < 0, \Delta_j \geq 0, \frac{b_i - \sqrt{\Delta_i}}{a_i} > \frac{b_j + \sqrt{\Delta_j}}{a_j}, \frac{b_i + \sqrt{\Delta_i}}{a_i} < \frac{b_j - \sqrt{\Delta_j}}{a_j} \right\}$ for $i, j = 2, 3$ and $i \neq j$ i.e. if the intervals are of the form $\left[\frac{b_i - \sqrt{\Delta_i}}{a_i}, \frac{b_i + \sqrt{\Delta_i}}{a_i}\right]$ and $\left[-\infty, \frac{b_j + \sqrt{\Delta_j}}{a_j}\right] \cup \left[\frac{b_j - \sqrt{\Delta_j}}{a_j}, +\infty\right]$ where $\frac{b_i - \sqrt{\Delta_i}}{a_i} > \frac{b_j + \sqrt{\Delta_j}}{a_j}$ and $\frac{b_i + \sqrt{\Delta_i}}{a_i} < \frac{b_j - \sqrt{\Delta_j}}{a_j}$.
6. $\left\{ a_i > 0, \Delta_i \geq 0, a_j > 0, \Delta_j \geq 0, \frac{b_i - \sqrt{\Delta_i}}{a_i} > \frac{b_j + \sqrt{\Delta_j}}{a_j} \right\}$ for $i, j = 2, 3$ and $i \neq j$ i.e. if the intervals are of the form $\left[\frac{b_i - \sqrt{\Delta_i}}{a_i}, \frac{b_i + \sqrt{\Delta_i}}{a_i}\right]$ and $\left[\frac{b_j - \sqrt{\Delta_j}}{a_j}, \frac{b_j + \sqrt{\Delta_j}}{a_j}\right]$ where $\frac{b_i - \sqrt{\Delta_i}}{a_i} > \frac{b_j + \sqrt{\Delta_j}}{a_j}$.

The above set of conditions are very useful and it reduces a grid search over (possibly) the whole real line to testing just eleven mutually exclusive conditions which are easy to verify.

7 Tables and Figures

Error Correlations		$\rho_{uX} = 0.5, \rho_{uW} = 0.5$			$\rho_{uX} = 0.1, \rho_{uW} = 0.99$			$\rho_{uX} = 0.99, \rho_{uW} = 0.1$		
ACV: α (in %)		1	5	10	1	5	10	1	5	10
Case I: $\mu_\beta = 1$ and $\mu_\gamma = 1$										
$\zeta = 25\%$	USSIV-Test	1.2	6	11.8	3.3	10.8	17.3	1.1	6.1	11.9
	Proj SS-Test	0	0.1	0.3	0.1	0.8	2.5	0	0.1	0.3
	Robins-Test ^a	0	0.1	0.2	0	0.6	2.7	0	0	0.2
	Proj AR-Test	0.1	1.1	2.5	2.5	7	11.4	0.1	1	2.1
	Partial K-Test	0.1	1.7	5.7	1.2	6.2	11.7	0.1	1.2	4
$\zeta = 50\%$	USSIV-Test	1.5	7.2	13.2	6.7	14.2	20.5	1.6	6.9	12.9
	Proj SS-Test	0	0.1	0.5	0.2	1.2	2.9	0	0.1	0.4
	Robins-Test	0	0	0.4	0	1.1	3.2	0	0	0.3
	Proj AR-Test	0.1	0.6	1.7	1.2	4.2	7.7	0.1	0.5	1.2
	Partial K-Test	0.1	1.8	5	1	5.5	10.8	0.1	1.3	4.5
$\zeta = 75\%$	USSIV-Test	2.7	9	15.7	10.1	18.1	24.9	2.2	8.4	14.7
	Proj SS-Test	0	0.2	0.6	0.1	0.9	2.6	0	0.1	0.5
	Robins-Test	0	0.1	0.5	0	0.1	2.9	0	0.1	0.5
	Proj AR-Test	0.1	0.7	1.9	0.7	3.2	6.5	0	0.5	1.5
	Partial K-Test	0.2	2.3	6	1	5.5	11.1	0.2	1.7	4.7
Case III: $\mu_\beta = 10$ and $\mu_\gamma = 1$										
$\zeta = 25\%$	USSIV-Test	1.2	5.8	11.8	2.3	9	15.4	1	5.9	12.3
	Proj SS-Test	0	0.1	0.4	0.1	0.8	2.4	0	0.1	0.3
	Robins-Test	0	0.1	0.3	0	0.5	2.4	0	0	0.2
	Proj AR-Test	0.1	1.1	2.5	2.5	7	11.4	0.1	1	2.1
	Partial K-Test	0.1	1.4	4.3	1	5.3	10.6	0.1	1.3	4.1
$\zeta = 50\%$	USSIV-Test	1.5	6.4	12.6	3.9	10.3	16.7	1.4	6.4	12.4
	Proj SS-Test	0	0.1	0.5	0.1	1.1	2.5	0	0.1	0.4
	Robins-Test	0	0.1	0.4	0	1	3	0	0	0.3
	Proj AR-Test	0.1	0.6	1.7	1.2	4.2	7.7	0.1	0.5	1.2
	Partial K-Test	0.1	1.7	4.9	0.9	5	10.4	0.1	1.3	4.2
$\zeta = 75\%$	USSIV-Test	1.8	7.5	13.6	5.8	13	19.8	1.7	7.2	13.2
	Proj SS-Test	0	0.2	0.6	0.2	1.1	2.7	0	0.1	0.4
	Robins-Test	0	0.1	0.5	0.1	0.9	2.9	0	0.1	0.3
	Proj AR-Test	0.1	0.7	1.9	0.7	3.2	6.5	0	0.5	1.5
	Partial K-Test	0.2	2.1	5.2	0.8	5	10.1	0.2	1.7	4.6

Table 1: Empirical Size of different tests for $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo trials. USSIV Score test, Projection from Split-sample Score test (SS) and Robins test combine observations from both sub-samples. Projection from AR test and partial-K tests are based on sub-sample one containing $n_1 = \lceil n\zeta \rceil$ observations. [$k = 4, \rho_{XW} = 0, n = 100$]

^aBy Theorem 2, the level of the Robins-test is always lesser than $\alpha + \epsilon$ ($= 2\text{ACV}$ by our specification), but is asymptotically equal to α if the instruments are strong for γ .

Error Correlations		$\rho_{uX} = 0.5, \rho_{uW} = 0.5$			$\rho_{uX} = 0.1, \rho_{uW} = 0.99$			$\rho_{uX} = 0.99, \rho_{uW} = 0.1$		
ACV: α (in %)		1	5	10	1	5	10	1	5	10
Case II: $\mu_\beta = 1$ and $\mu_\gamma = 10$										
$\zeta = 25\%$	USSIV-Test	1.1	5.9	11.6	1	5.9	11.8	1	5.7	11.4
	Proj SS-Test	0.1	0.7	2.1	0.1	1.3	3.4	0	0.5	2
	Robins-Test ^a	0	0.3	1.9	0.1	1.8	5.7	0	0.2	1.6
	Proj AR-Test	1.3	4.7	8.4	2.7	7.5	12.2	1.1	4.1	7.2
	Partial K-Test	0.7	4.8	9.6	1.2	6.2	11.7	0.3	3.3	7.7
$\zeta = 50\%$	USSIV-Test	1.1	5.9	11.3	1.2	6.1	11.7	1.1	5.6	11.1
	Proj SS-Test	0.1	1.1	2.9	0.2	1.4	3.1	0.1	0.9	2.5
	Robins-Test	0.1	1.1	3.9	0.2	2.3	5.9	0	0.8	3.1
	Proj AR-Test	0.9	3.6	6.7	1.2	4.3	7.9	0.8	3.1	6.6
	Partial K-Test	0.8	5	10.2	1	5.6	11	0.7	4.2	9.3
$\zeta = 75\%$	USSIV-Test	1.7	6.9	12.8	1.4	6.7	12.8	1.7	6.7	12.5
	Proj SS-Test	0.1	1.1	2.7	0.1	0.9	2.9	0.2	1.1	2.6
	Robins-Test	0.1	1.2	3.9	0.1	1.4	4.5	0.1	1	3.3
	Proj AR-Test	0.7	3.1	5.9	0.8	3.3	6.7	0.6	2.9	5.8
	Partial K-Test	0.9	4.9	10.6	0.9	5.4	11	0.8	4.7	9.7
Case IV: $\mu_\beta = 10$ and $\mu_\gamma = 10$										
$\zeta = 25\%$	USSIV-Test	0.9	6.1	11.6	1	5.9	11.7	0.8	5.8	12
	Proj SS-Test	0	0.6	2.1	0.1	1.3	3.5	0	0.4	1.7
	Robins-Test	0	0.3	2.1	0	1.7	5.6	0	0.2	1.6
	Proj AR-Test	1.3	4.7	8.4	2.7	7.5	12.2	1.1	4.1	7.2
	Partial K-Test	0.5	3.9	8.6	0.9	5.4	10.7	0.3	3.2	7.6
$\zeta = 50\%$	USSIV-Test	1.2	5.7	11	1.2	5.7	11.5	1.3	5.8	11.2
	Proj SS-Test	0.2	1.2	2.8	0.2	1.5	3.1	0.1	1.1	2.6
	Robins-Test	0	1.1	3.9	0.2	2.2	5.8	0	0.9	3.4
	Proj AR-Test	0.9	3.6	6.7	1.2	4.3	7.9	0.8	3.1	6.6
	Partial K-Test	0.8	4.6	9.8	1	5.2	10.3	0.7	4.1	9.3
$\zeta = 75\%$	USSIV-Test	1.3	6.4	12.6	1.4	6.4	12.4	1.5	6.4	12.3
	Proj SS-Test	0.2	1	2.8	0.2	1.2	3.1	0.2	1	2.7
	Robins-Test	0.1	1.2	4.1	0.2	1.8	5.2	0.1	1	3.7
	Proj AR-Test	0.7	3.1	5.9	0.8	3.3	6.7	0.6	2.9	5.8
	Partial K-Test	0.8	4.8	9.8	1	5.1	10.3	0.8	4.7	9.6

Table 2: Empirical Size of different tests for $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo trials. USSIV Score test, Projection from Split-sample Score test (SS) and Robins test combine observations from both sub-samples. Projection from AR test and partial-K tests are based on sub-sample one containing $n_1 = [n\zeta]$ observations. [$k = 4, \rho_{XW} = 0, n = 100$]

^aBy Theorem 2, the level of the Robins-test is always lesser than $\alpha + \epsilon$ ($= 2ACV$ by our specification), but is asymptotically equal to α if the instruments are strong for γ .

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.5$, $\rho_{uW} = 0.5$, $\rho_{XW} = 0$ and $\zeta = 0.75$

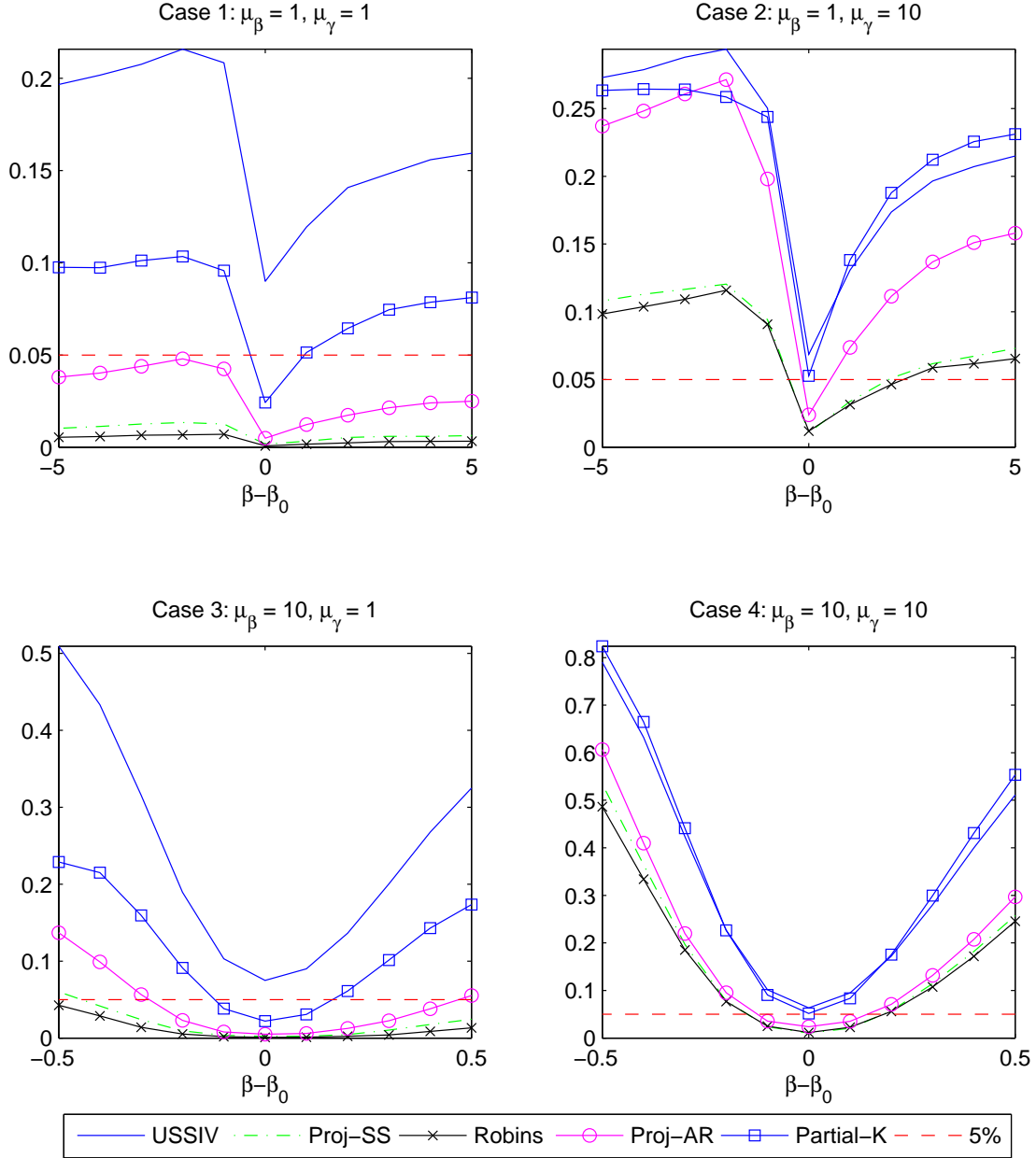


Figure 1: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.1$, $\rho_{uW} = 0.99$, $\rho_{XW} = 0$ and $\zeta = 0.75$

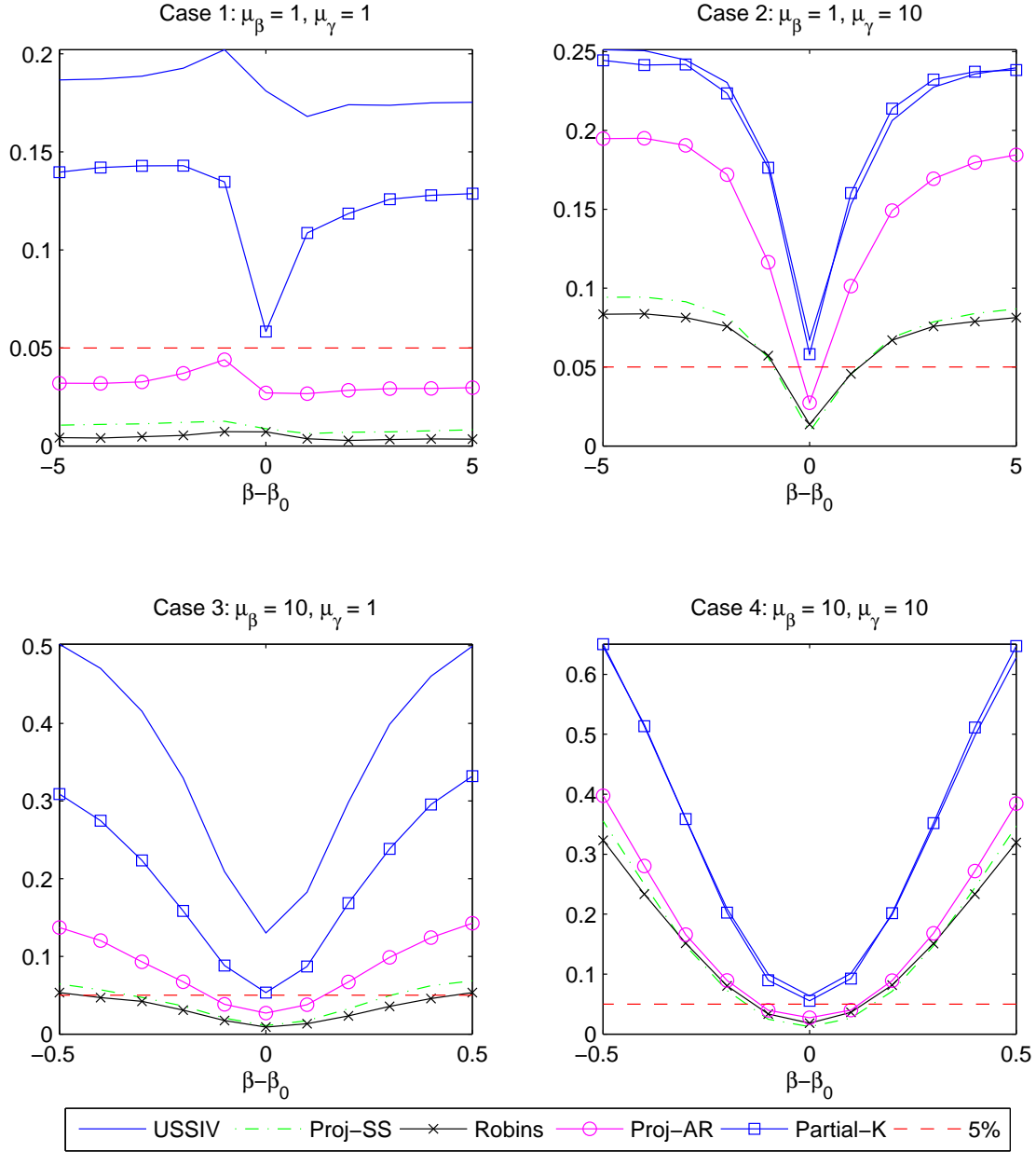


Figure 2: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.99$, $\rho_{uW} = 0.1$, $\rho_{XW} = 0$ and $\zeta = 0.75$

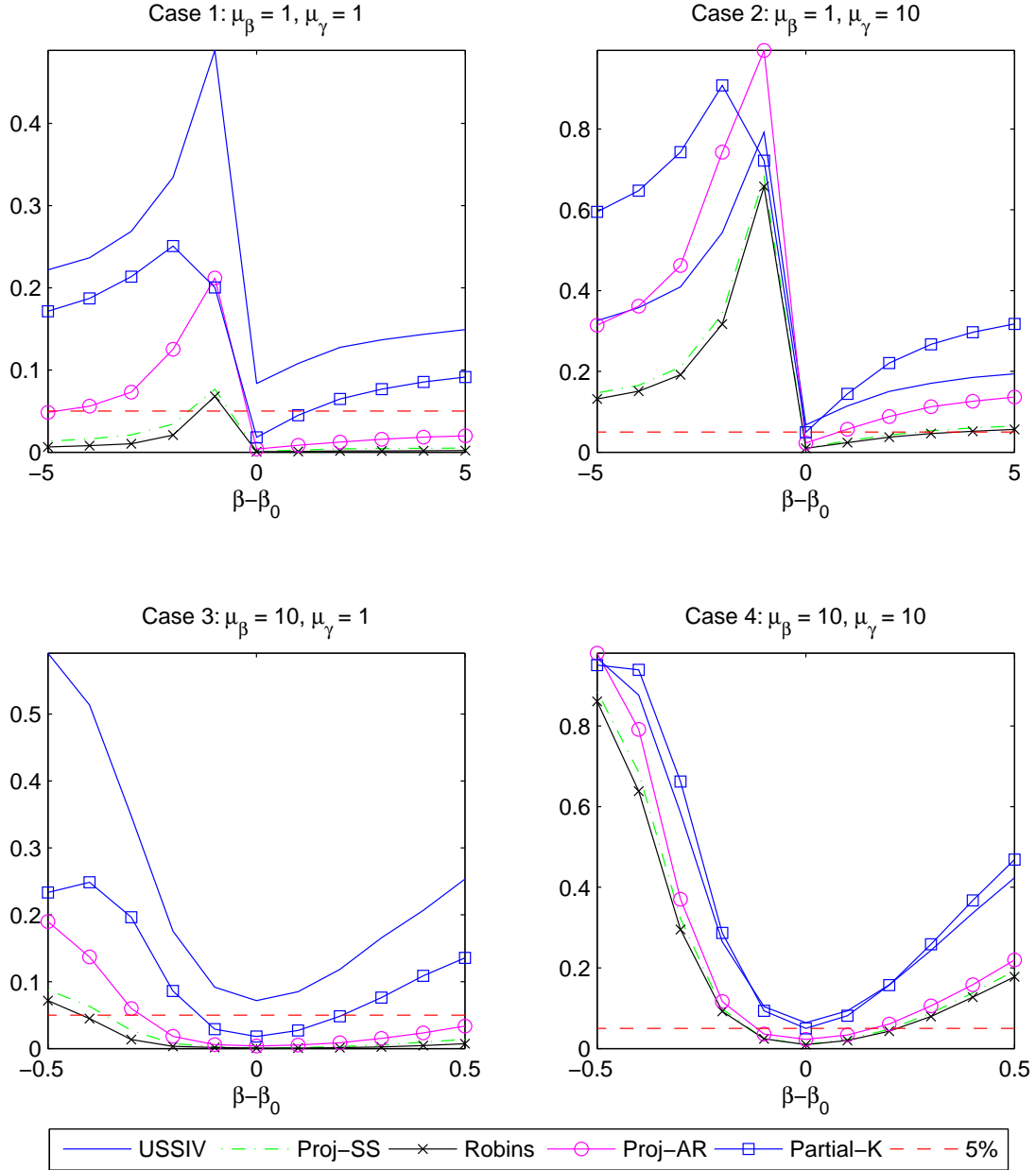


Figure 3: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.5$, $\rho_{uW} = 0.5$, $\rho_{XW} = 0$ and $\zeta = 0.75$

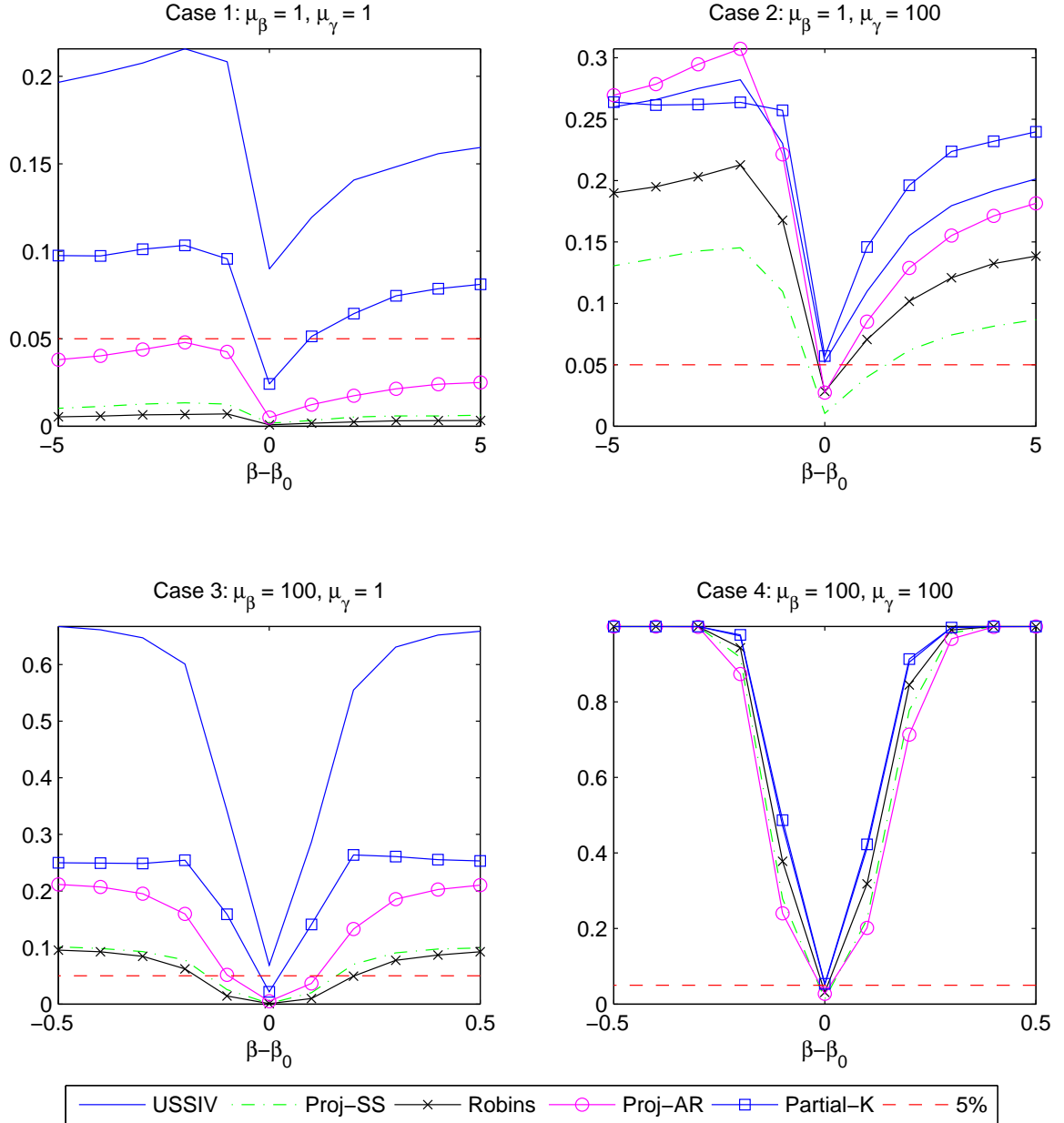


Figure 4: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.1$, $\rho_{uW} = 0.99$, $\rho_{XW} = 0$ and $\zeta = 0.75$

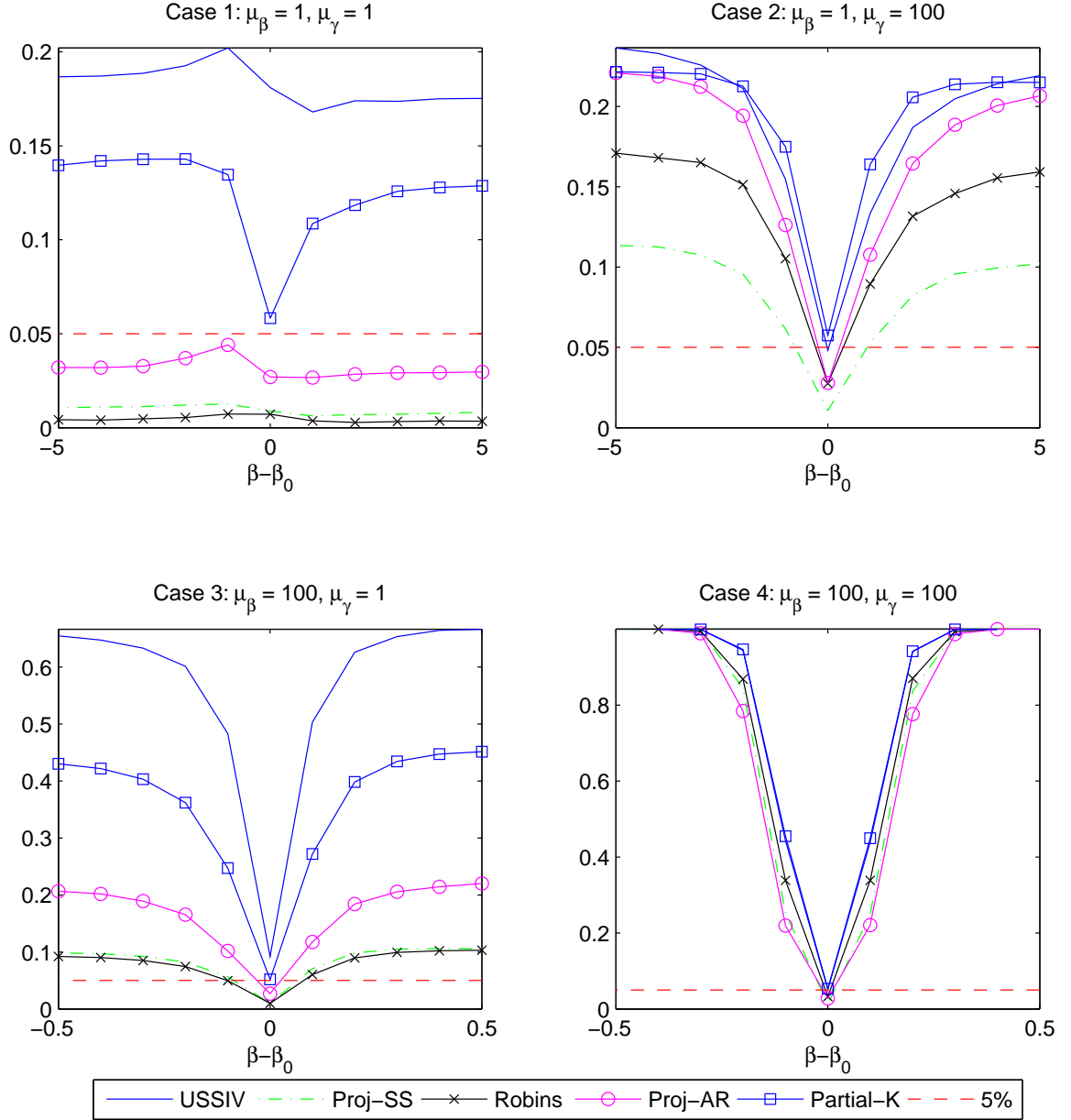


Figure 5: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.99$, $\rho_{uW} = 0.1$, $\rho_{XW} = 0$ and $\zeta = 0.75$

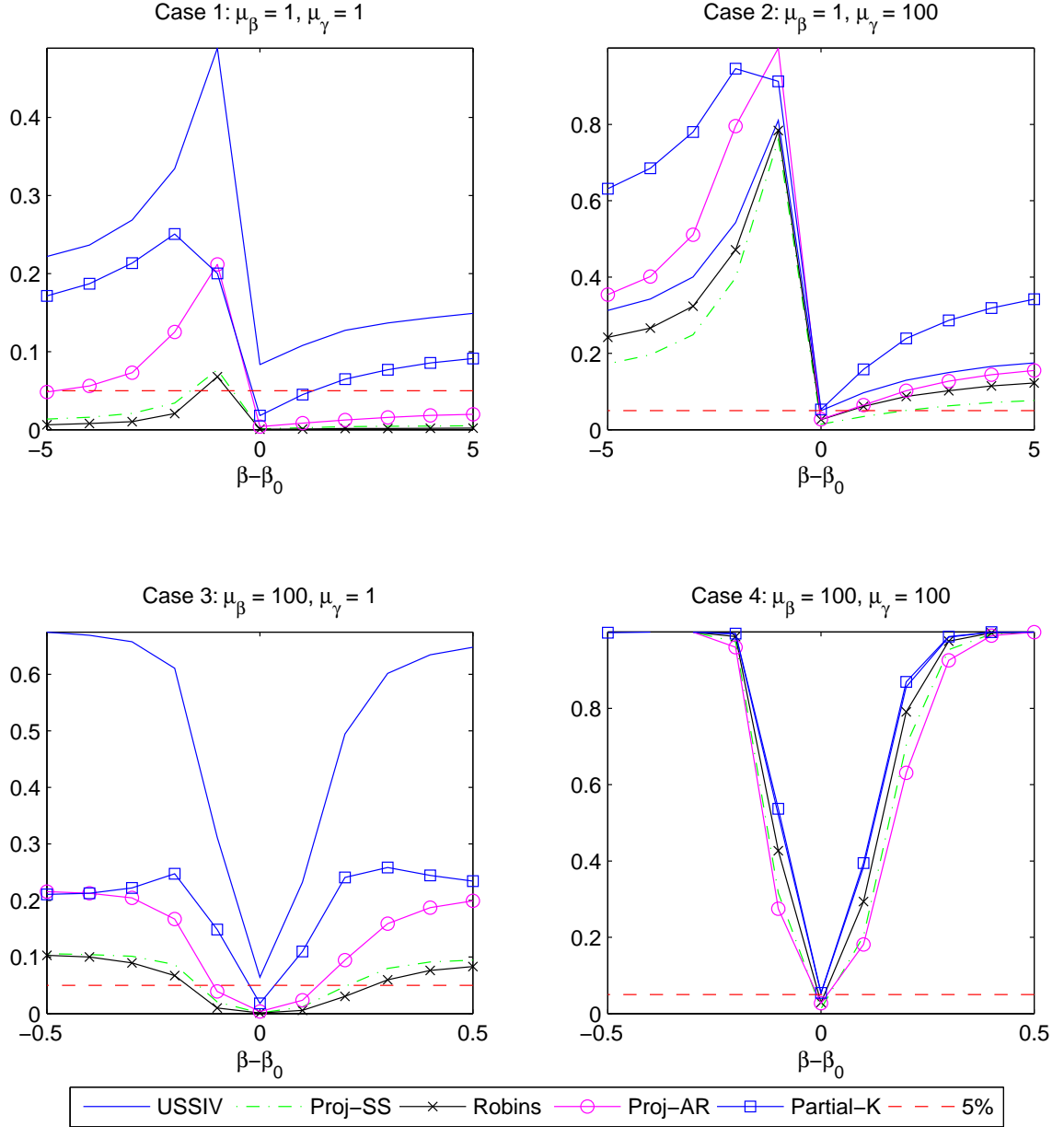


Figure 6: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.5$, $\rho_{uW} = 0.5$, $\rho_{XW} = 0$ and $\zeta = 0.25$

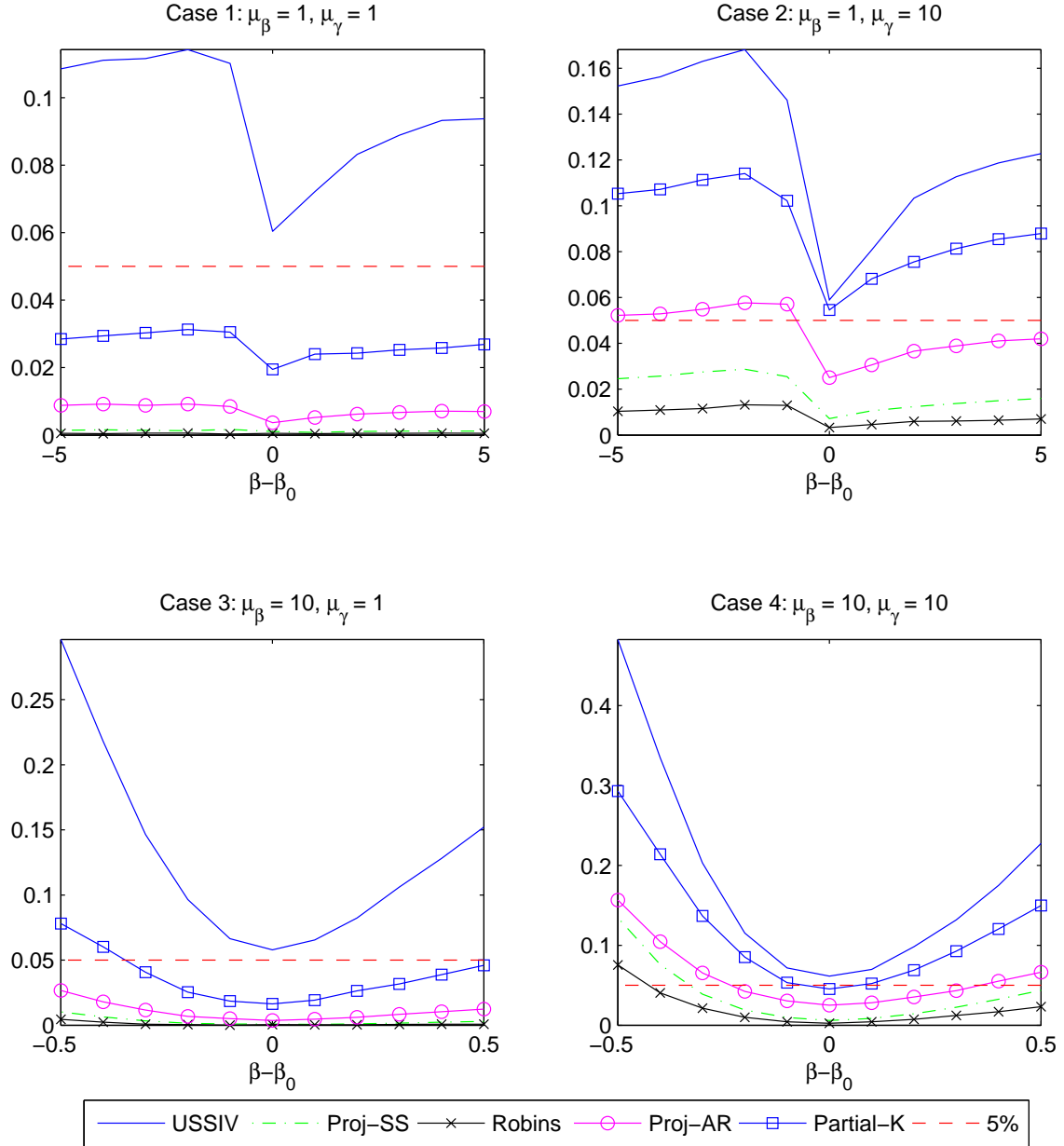


Figure 7: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.1$, $\rho_{uW} = 0.99$, $\rho_{XW} = 0$ and $\zeta = 0.25$

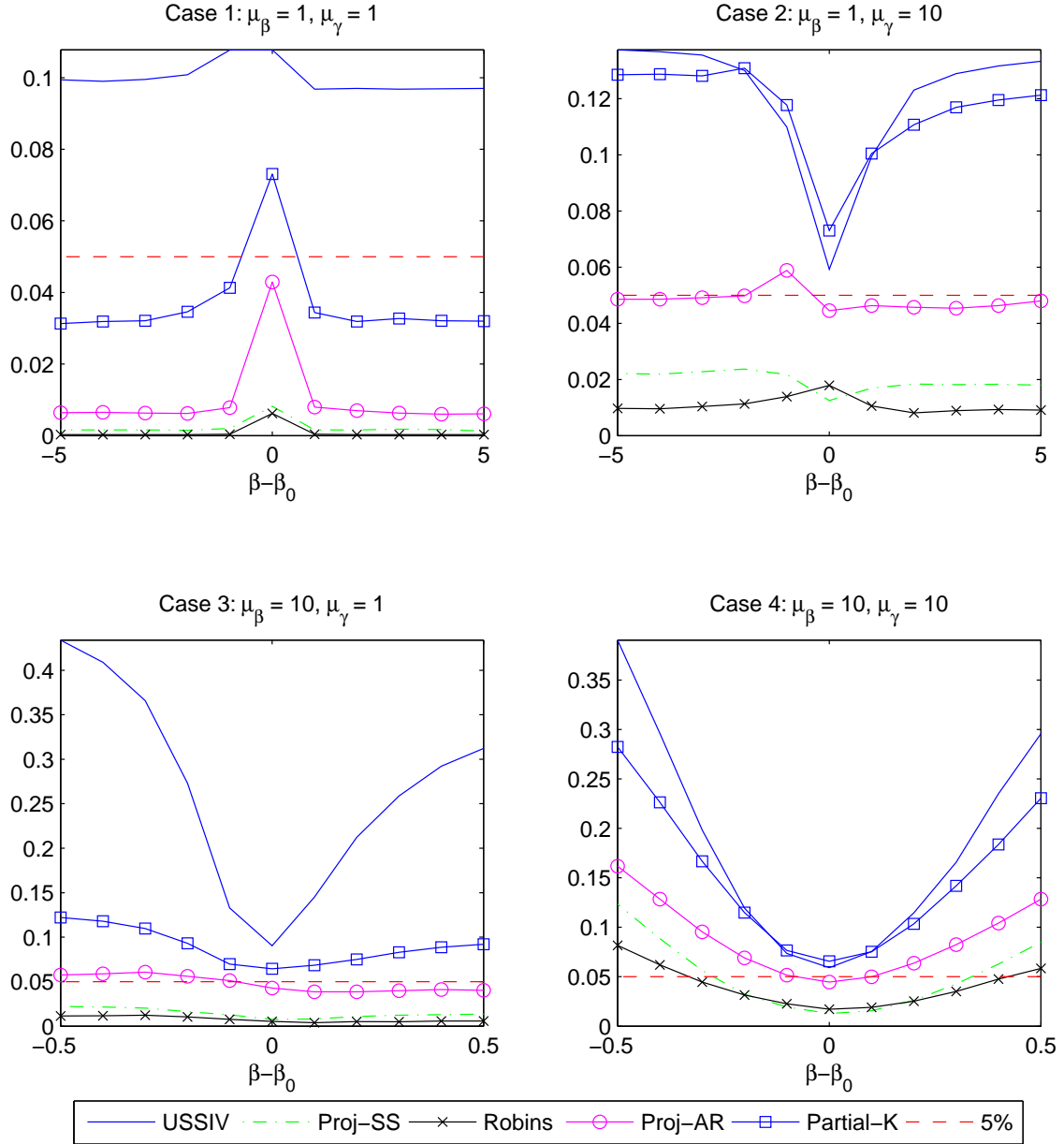


Figure 8: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.99$, $\rho_{uW} = 0.1$, $\rho_{XW} = 0$ and $\zeta = 0.25$

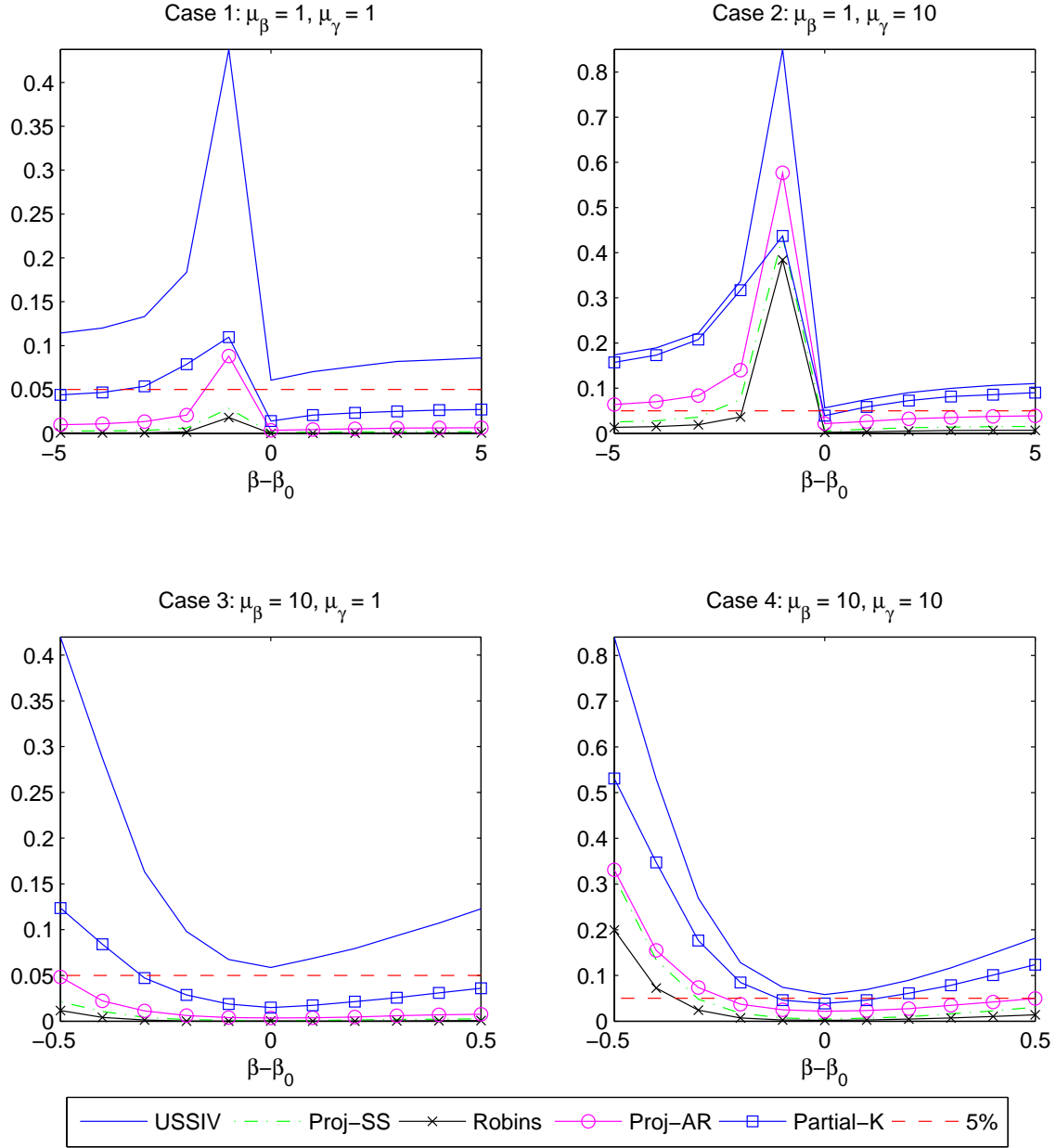


Figure 9: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.5$, $\rho_{uW} = 0.5$, $\rho_{XW} = 0$ and $\zeta = 0.25$

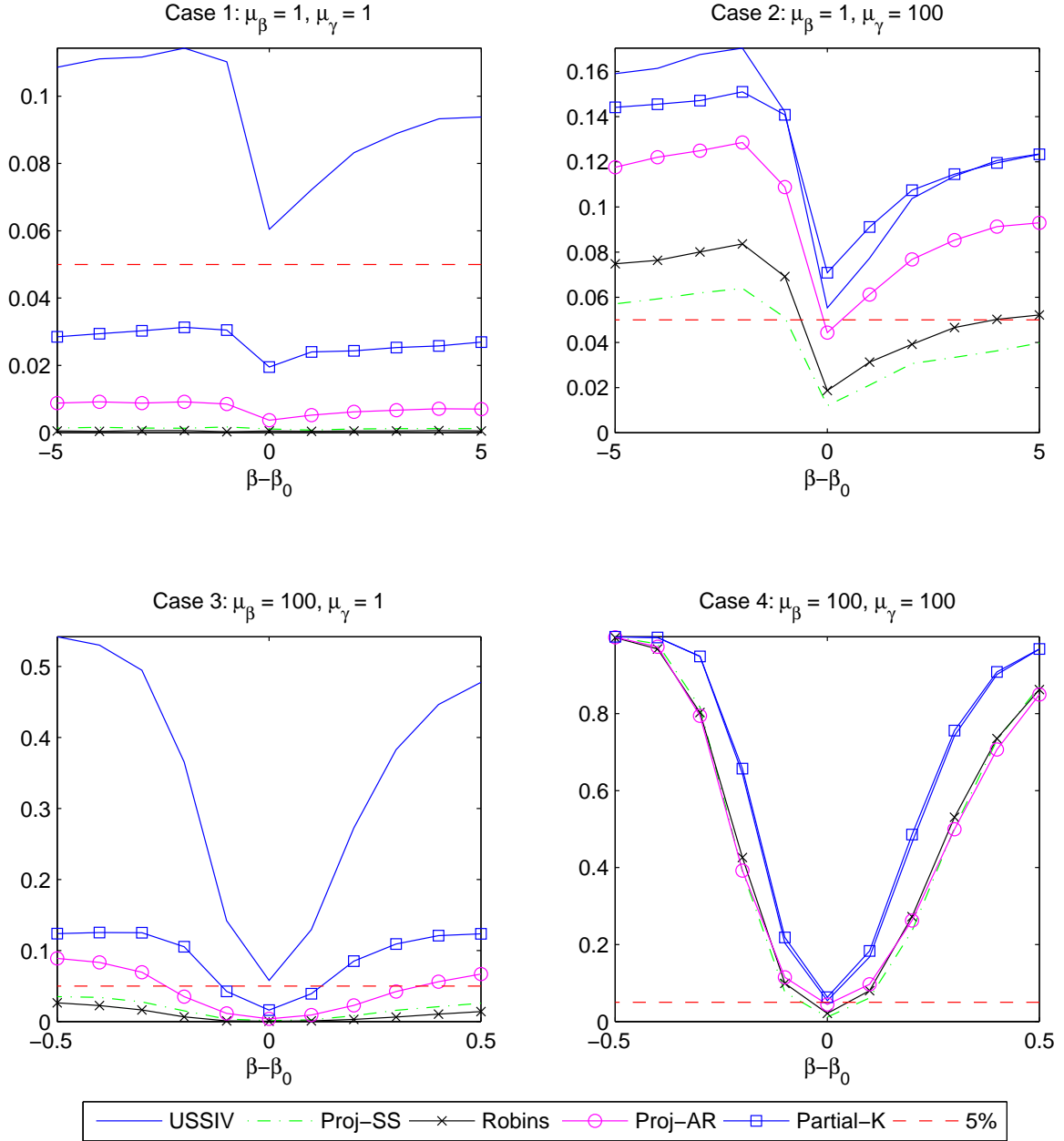


Figure 10: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.1$, $\rho_{uW} = 0.99$, $\rho_{XW} = 0$ and $\zeta = 0.25$

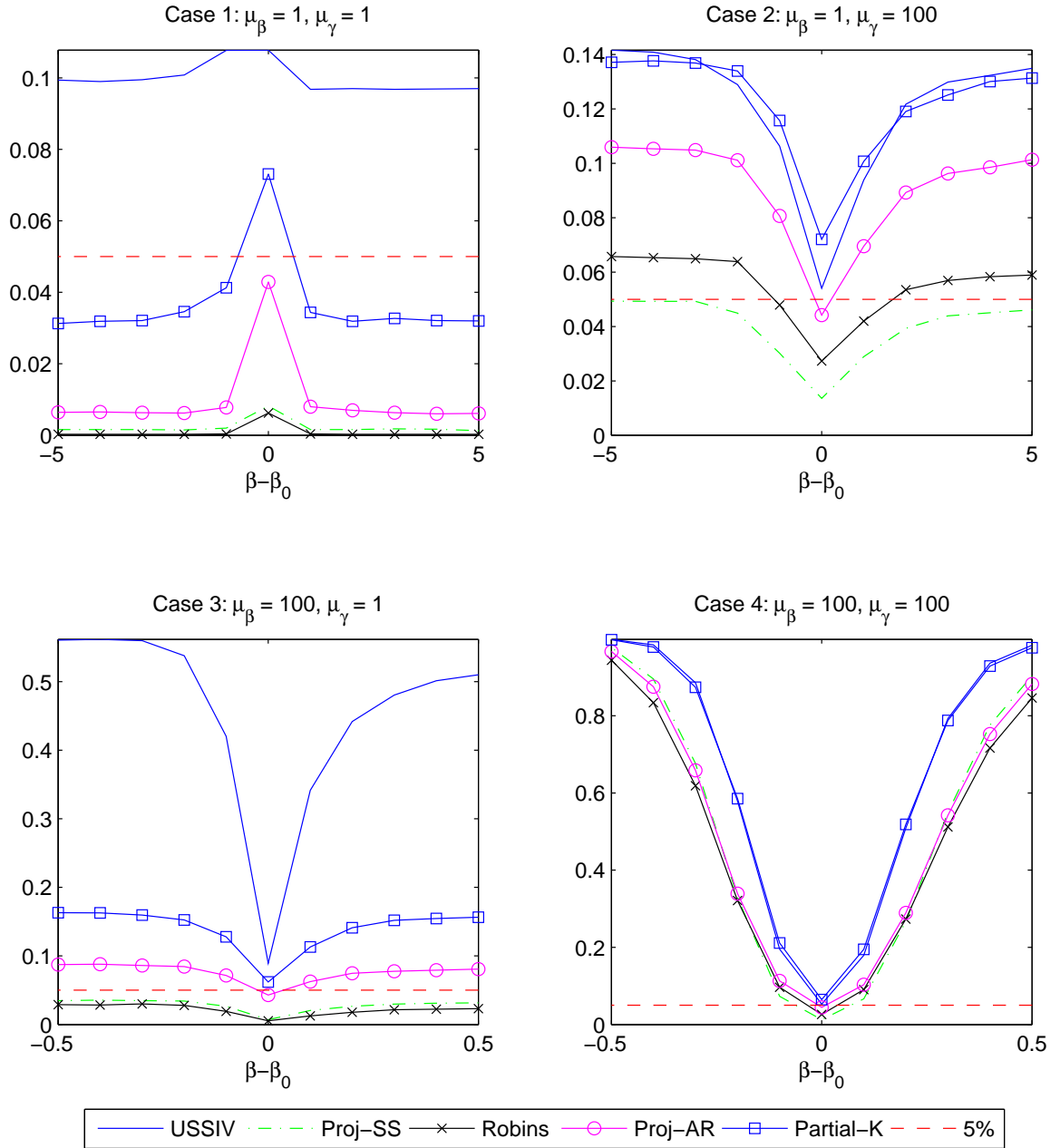


Figure 11: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.99$, $\rho_{uW} = 0.1$, $\rho_{XW} = 0$ and $\zeta = 0.25$

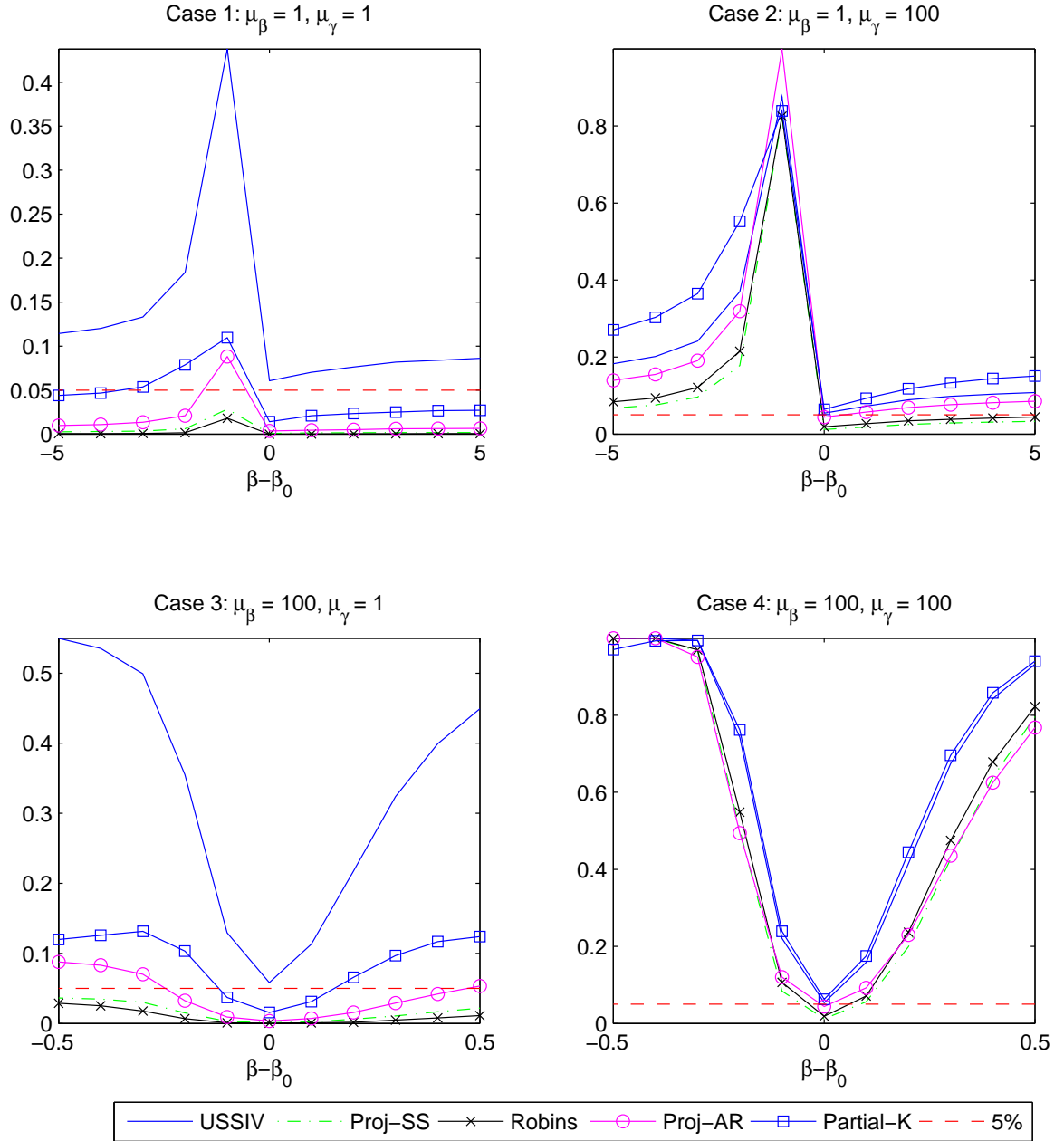


Figure 12: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.5$, $\rho_{uW} = 0.5$, $\rho_{XW} = 0$ and $\zeta = 0.5$

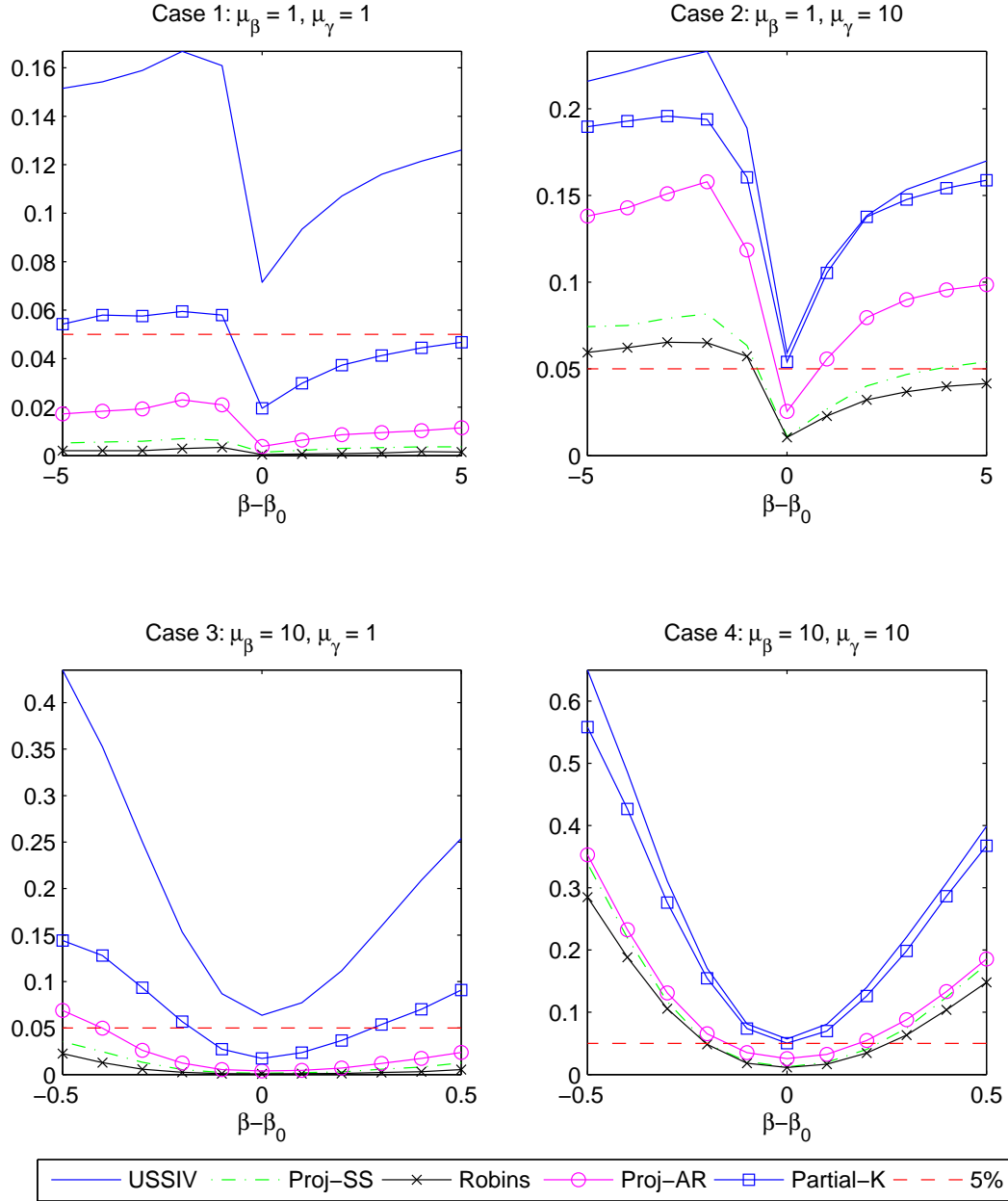


Figure 13: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.1$, $\rho_{uW} = 0.99$, $\rho_{XW} = 0$ and $\zeta = 0.5$

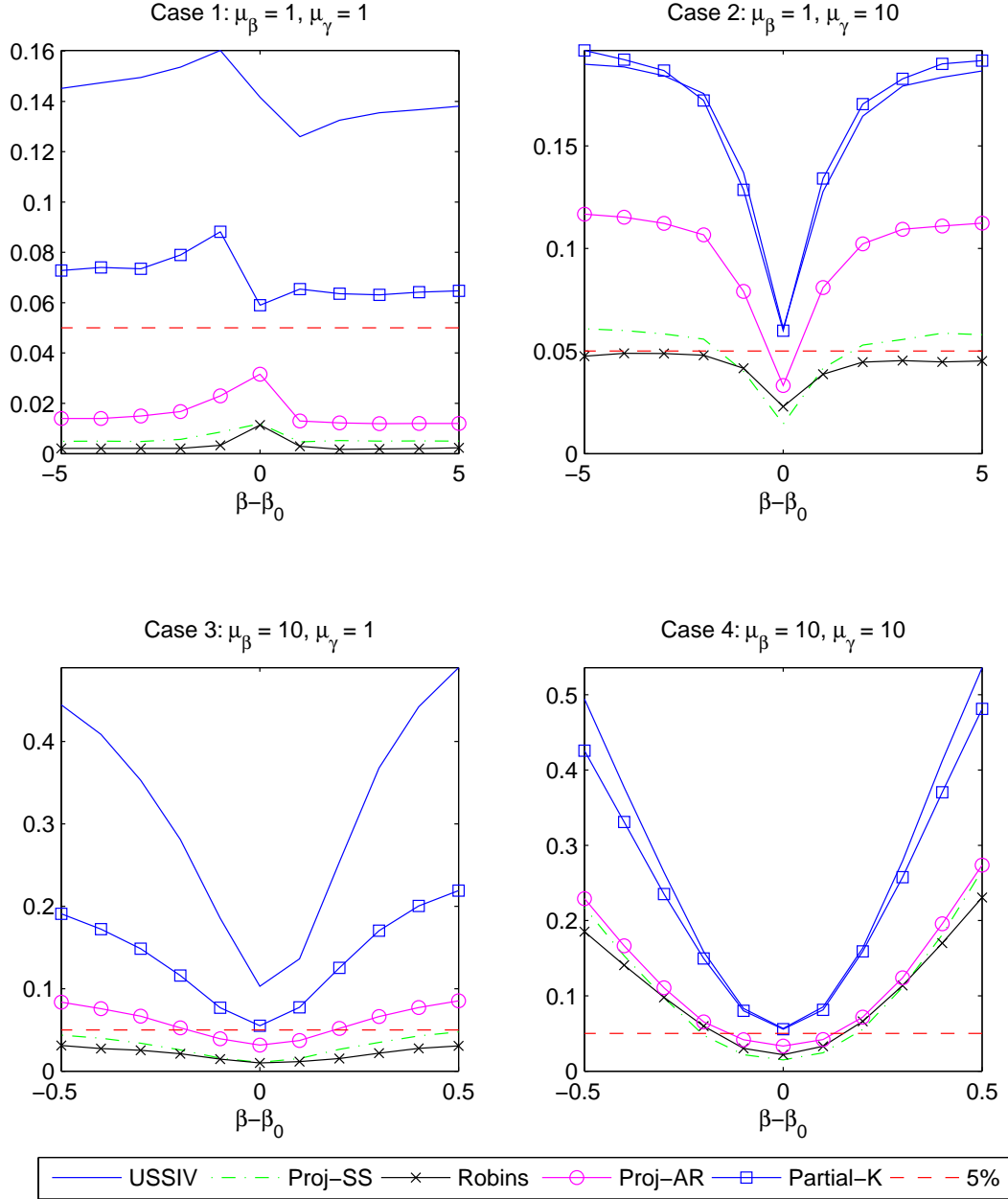


Figure 14: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.99$, $\rho_{uW} = 0.1$, $\rho_{XW} = 0$ and $\zeta = 0.5$

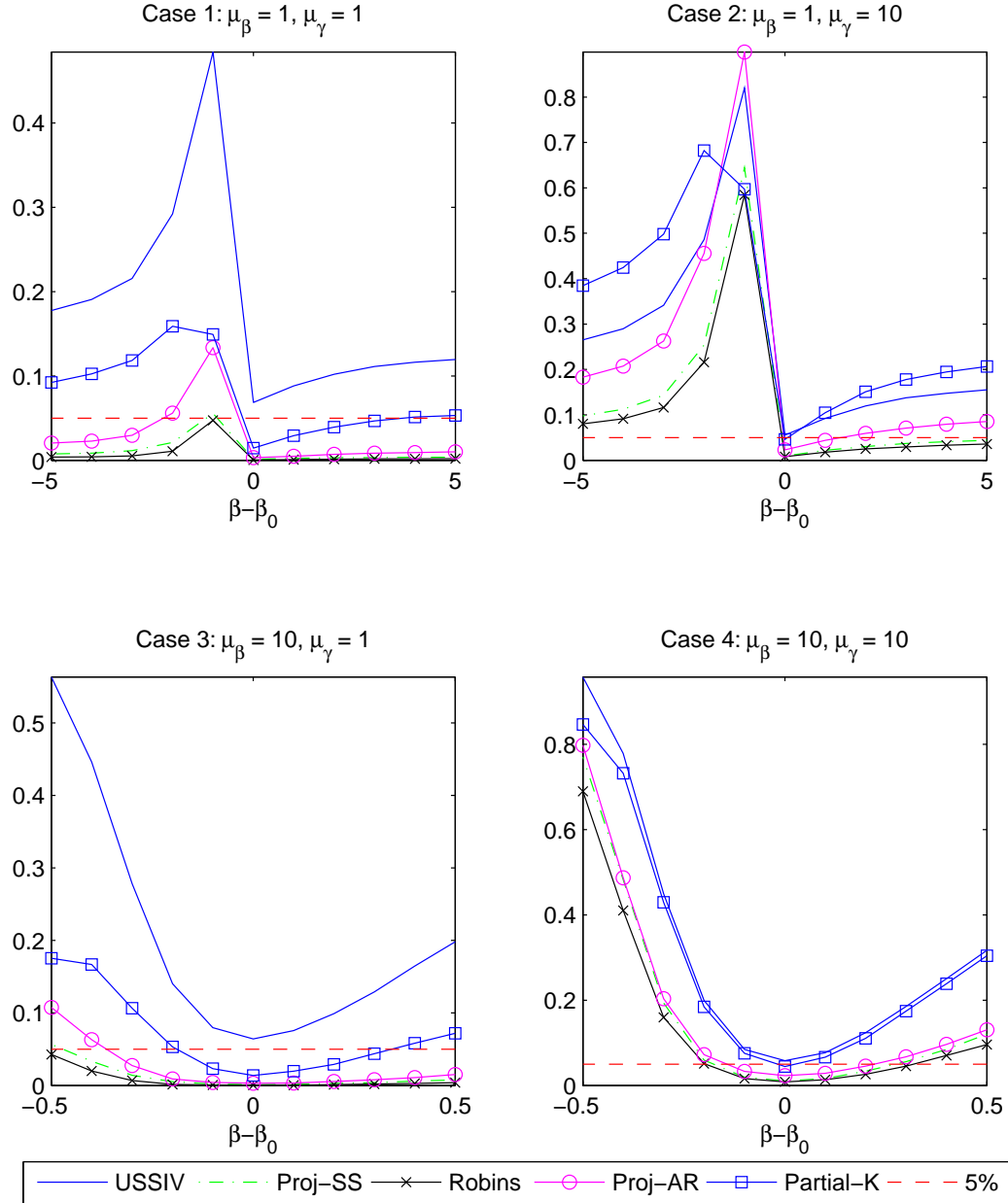


Figure 15: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.5$, $\rho_{uW} = 0.5$, $\rho_{XW} = 0$ and $\zeta = 0.5$

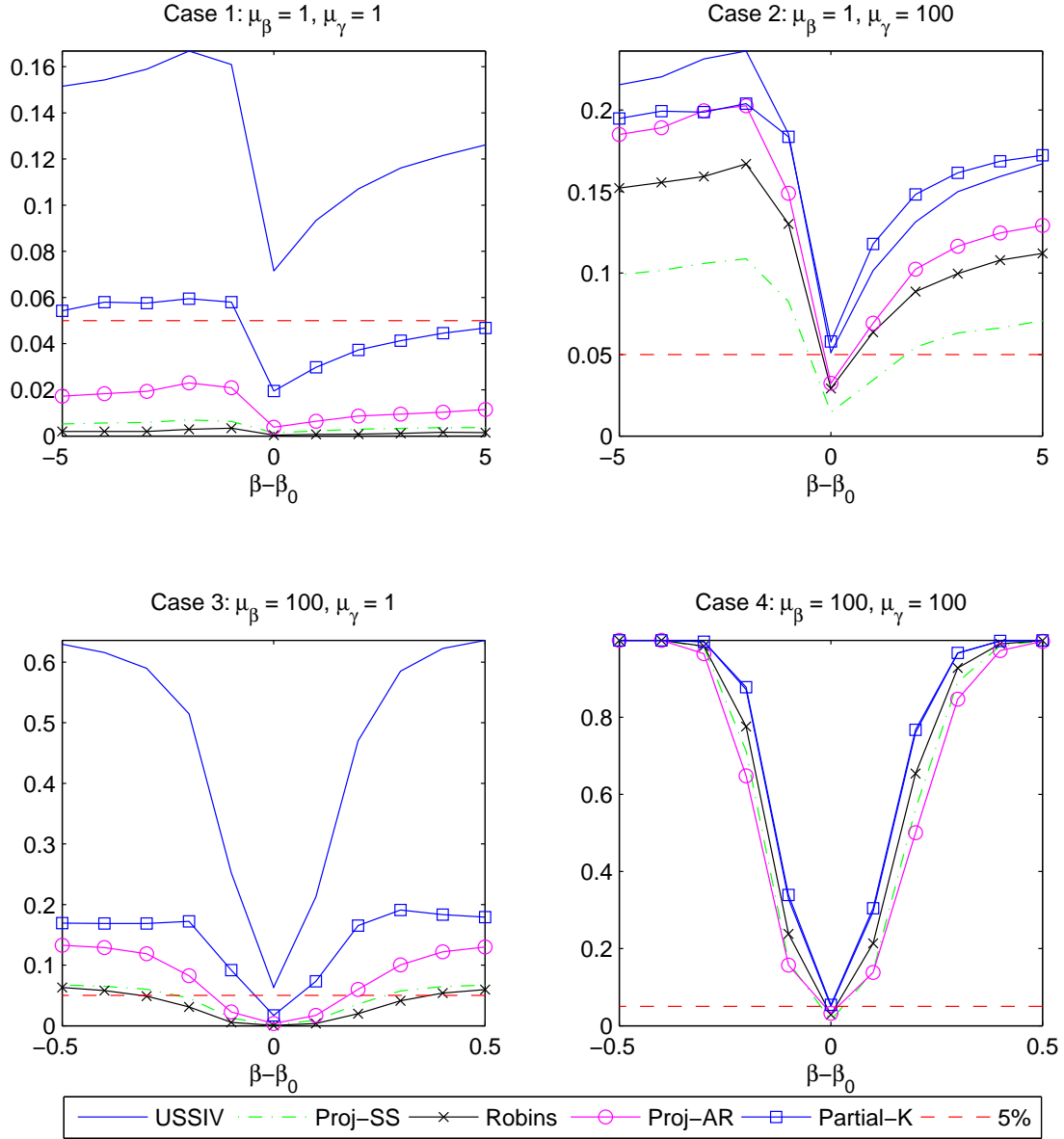


Figure 16: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.1$, $\rho_{uW} = 0.99$, $\rho_{XW} = 0$ and $\zeta = 0.5$

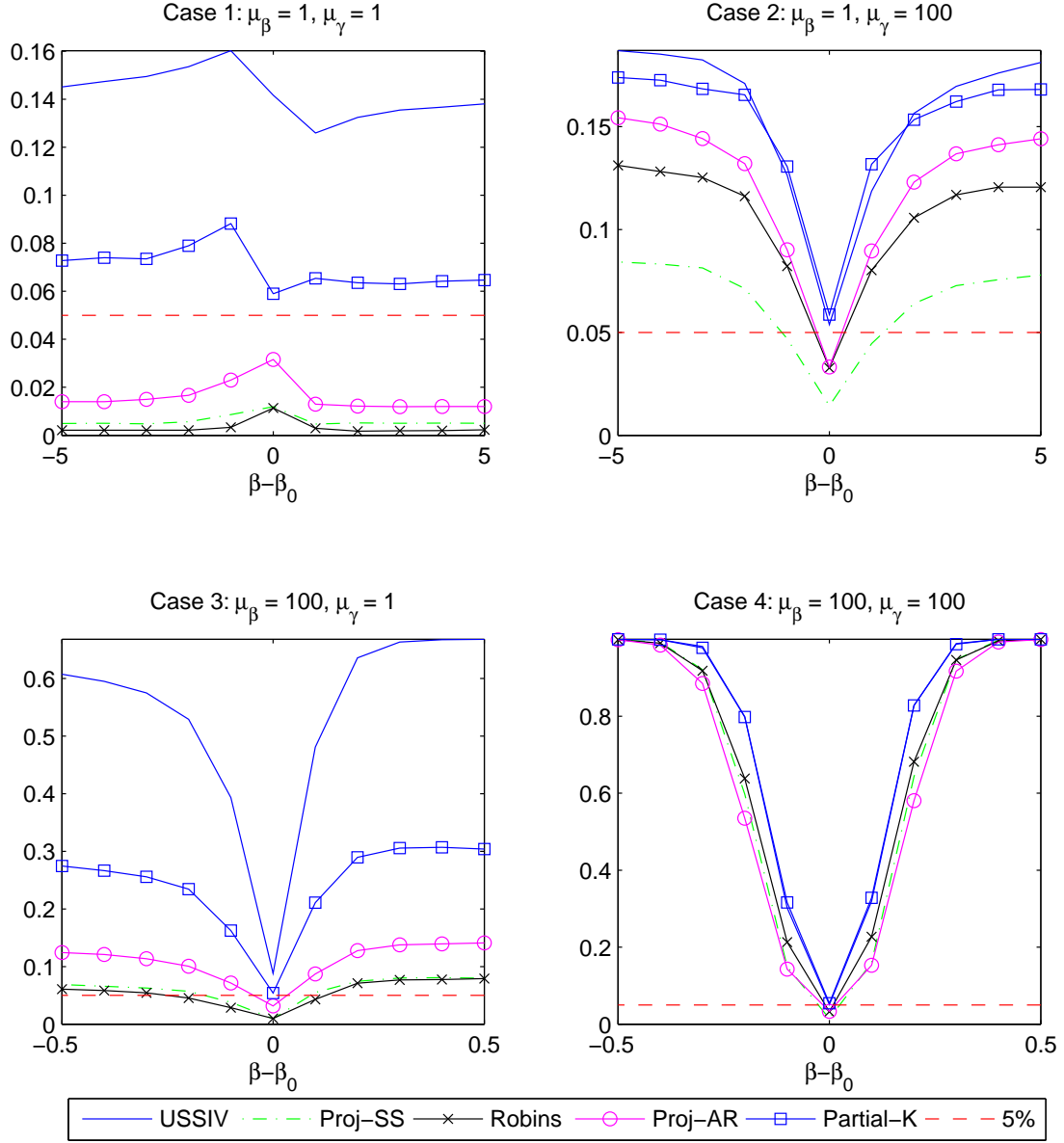


Figure 17: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

Power Curves when $k = 4$, $n = 100$, $\rho_{uX} = 0.99$, $\rho_{uW} = 0.1$, $\rho_{XW} = 0$ and $\zeta = 0.5$

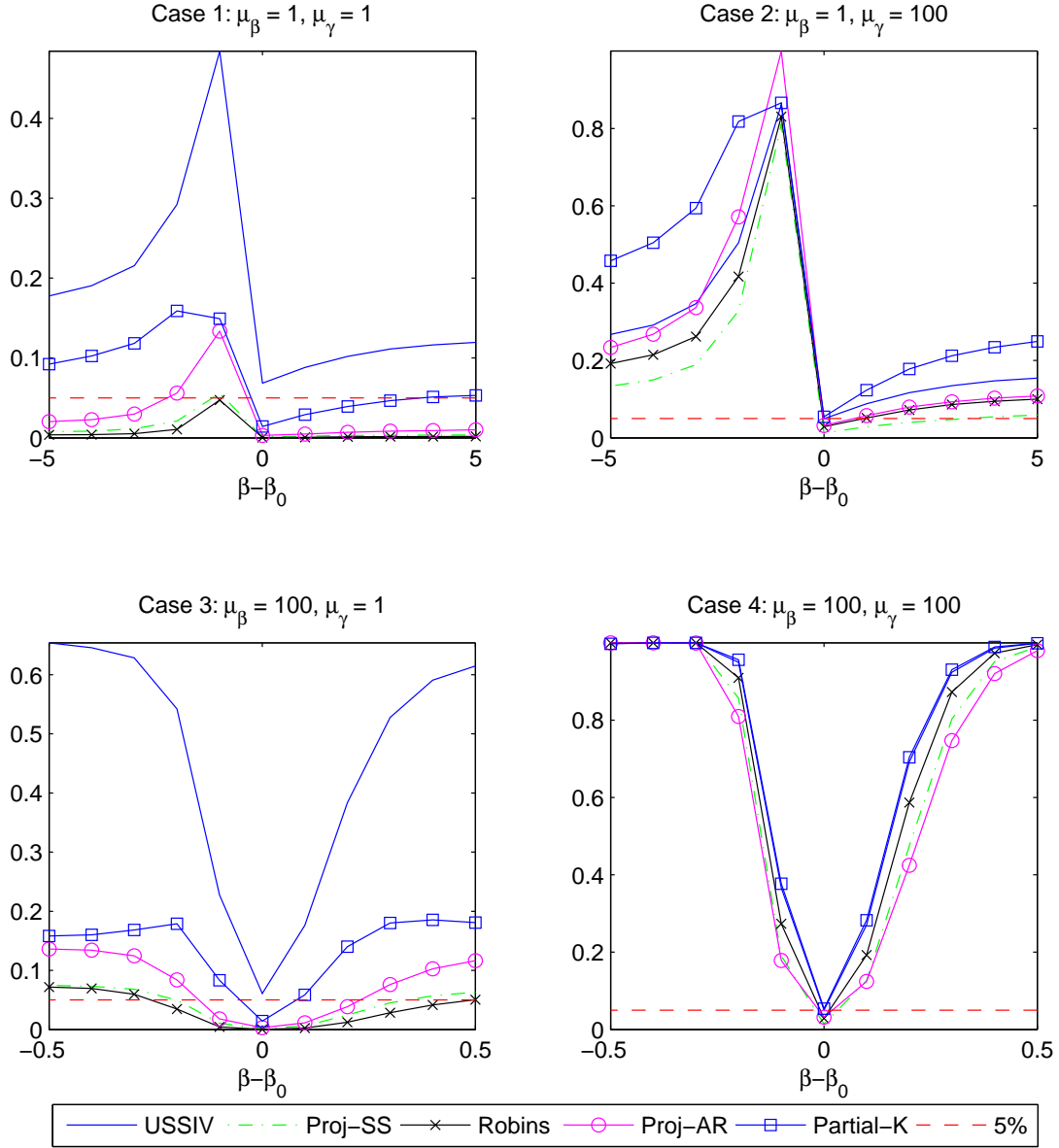


Figure 18: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

References

- Angrist, J. and Krueger, A. B. (1992). The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of American Statistical Association*, 87: 328–336.
- Angrist, J. and Krueger, A. B. (1995). Split-Sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business and Economics Statistics*, 13: 225–235.
- Chaudhuri, S. (2007). Testing of Hypotheses in Partially Identified Linear Instrumental Variables Model. Job-Market Paper.
- Choi, I. and Phillips, P. C. B. (1992). Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations. *Journal of Econometrics*, 51: 113–150.
- Dufour, J. M. (1997). Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, 65: 1365–1388.
- Dufour, J. M. and Jasiak, J. (2001). Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review*, 42: 815–843.
- Dufour, J. M. and Taamouti, M. (2005). Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments. *Econometrica*, 73: 1351–1365.
- Guggenberger, P. and Smith, R. (2005). Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification. *Econometric Theory*, 21: 667–709.
- Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996). Judging Instrumental Relevance in Instrumental Variables Estimation. *International Economic Review*, 37: 283–298.
- Kleibergen, F. (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70: 1781–1803.
- Kleibergen, F. (2004). Testing Subsets of Parameters In The Instrumental Variables Regression Model. *The Review of Economics and Statistics*, 86: 418–423.

- Mikusheva, A. (2006). Robust Confidence Sets in the Presence of Weak Instruments. Department of Economics, Harvard University.
- Moreira, M. J. (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71: 1027–1048.
- Phillips, P. C. B. (1989). Partially Identified Econometric Models. *Econometric Theory*, 5: 181–240.
- Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D. Y. and Heagerty, P., editors, *In Proceedings of the Second Seattle Symposium on Biostatistics*. New York: Springer.
- Shea, J. (1997). Instrument Relevance in Multivariate linear models: A simple measure. *The Review of Economics and Statistics*, LXXIX: 348–352.
- Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65: 557–586.
- Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press.
- Wang, J. and Zivot, E. (1998). Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments. *Econometrica*, 66: 1389–1404.
- Zivot, E. and Chaudhuri, S. (2007). The Distribution of Shea’s Partial R^2 Statistic in the Presence of Weak Instruments. Unpublished.
- Zivot, E., Startz, R., and Nelson, C. (1998). Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, 39: 1119–1144.
- Zivot, E., Startz, R., and Nelson, C. (2006). Inference in Weakly Identified Instrumental Variables Regression. In Corbae, D., Durlauf, S. N., and Hansen, B. E., editors, *Frontiers in Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pages 125–166. Cambridge University Press.