Split-Sample Score Tests in Linear Instrumental Variables Regression *

Saraswata Chaudhuri, Thomas Richardson, James Robins and Eric Zivot

Working Paper no. 73
Center for Statistics and the Social Sciences
University of Washington

First Draft: March 30, 2007
This Version: June 26, 2007

*Saraswata Chaudhuri is a graduate student of Economics, University of Washington and the corresponding author of this paper (E-mail: saraswat@u.washington.edu). Thomas Richardson is Associate Professor of Statistics, Department of Statistics, University of Washington. James Robins is Professor of Epidemiology and Biostatistics, Departments of Epidemiology and Biostatistics, Harvard University. Eric Zivot is Associate Professor of Economics, Deparment of Economics, University of Washington. 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). We thank Charles R. Nelson, Richard Parks, Michael D. Perlman, Elaina Rose, Richard Startz, the participants at the North American Econometric Society Summer Meeting (2007) and the seminar participants at the University of Washington (Economics Department) for their helpful comments and suggestions.

Abstract

In this paper we design two split-sample score 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 (USSIV) 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. When the usual regularity conditions are satisfied, the USSIV score test is asymptotically equivalent to the standard score test based on sub-sample one. However, the USSIV score test can be over-sized if the remaining structural coefficients are not identified. This motivates a new score-type test based on a general technique proposed by Robins (2004). The new test is never over-sized and if the remaining structural coefficients are identified, this test is asymptotically equivalent to USSIV score test against $\sqrt{n}$-local alternatives.

KEY WORDS: Instrumental Variables; Sample Splitting.

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 the test 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 inference procedure typically involves three steps – 1) split the sample randomly into two sub-samples, 2) predict the endogenous regressors by multiplying the observed 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) perform the actual inference on the structural coefficients using the dependent variable from sub-sample one and the predicted 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, the 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 (on the boundary) tests like the K-test and the Conditional Likelihood Ratio (CLR) test based on the whole sample [see Kleibergen (2002) and Moreira (2003)]. 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

[1]If we think of each observation on y being included independently in the sample with probability p, then by

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, however, allow for a gain in the asymptotic power under the above situation.

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. Finally, let β be the coefficient of interest.

If the instruments Z are "weak" for β, we cannot estimate β consistently and the standard Wald, likelihood ratio and score tests for testing null hypotheses of the form $H_0 : \beta = \beta_0$ are not valid. [2] Although it is still possible to validly test the null hypothesis using projection tests based on the AR-statistic or the K-statistic or the LR-statistic, the projection tests can be overly 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 for $H_0 : \beta = \beta_0$ which is more powerful than the projection test based on the K-statistic; Zivot et al. (2006) call this the "partial K-test". The partial K-statistic substitutes γ by its limited-information maximum likelihood estimator (constrained by $H_0 : \beta = \beta_0$) in the K-statistic and adjusts the critical value for the test accordingly. This statistic is pivotal when the instruments are strong for γ and boundedly pivotal when the instruments are weak for γ [see Kleibergen (2007)]. The partial K-test can be inverted to obtain confidence regions for β. However, unless the partial K-test is combined with the J-test, not all the points belonging to these confidence regions are necessarily consistent for β even when the instruments are strong for all the structural coefficients [see Kleibergen (2004)]. Moreover, unlike the tests discussed in Dufour and Taamouti (2005), construction of confidence regions based on the partial K-test can be

"randomly missing observations" we mean that p is determined exogenously.

[2] In 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.

numerically costly particularly for larger dimensions of β. [3]

In this paper we first propose a test which rejects the null hypothesis $H_0 : \beta = \beta_0$ for large values of the split-sample score statistic evaluated at $\beta = \beta_0$ and $\gamma = \widehat{\gamma}(\beta_0)$ where $\widehat{\gamma}(\beta_0)$ is the USSIV estimator of γ restricted by the hypothesized value of β [see Angrist and Krueger (1995)]. We call this the USSIV score test. When the instruments are strong for all the structural coefficients the USSIV score test is valid and, against $\sqrt{n}$-local alternatives, it is asymptotically equivalent to the standard score test and the partial K-test (both based on sub-sample one). We show that, unlike the partial K-test, it is possible to analytically obtain confidence regions for β by inverting the USSIV test. However, the USSIV score test is not valid when the instruments are weak for γ.

This motivates a second test which, to our knowledge, is new to the econometrics literature. The new test is based on a general technique proposed by Robins (2004). This test rejects the null hypothesis $H_0 : \beta = \beta_0$ for large values of the infimum of a split-sample version of the efficient score statistic for β where the infimum is taken over all possible values of the nuisance parameter γ belonging to a region which – 1) is asymptotically equivalent to the entire parameter space for γ when the instruments are weak for γ and 2) belongs to the $\sqrt{n}$-neighborhood of the true value of γ with probability one when the instruments are strong for γ. We call this test the "Score-type test" based on Robins method. The present paper makes an important methodological contribution in this respect: the Score-type test is valid irrespective of the strength of the instruments, for either β or γ and, furthermore, when the instruments are strong for γ, the Score-type test is asymptotically equivalent to the USSIV score test against $\sqrt{n}$-local alternatives. It is possible to invert this test to construct confidence regions for β. Such regions are quite generally less conservative than the projection based regions obtained by inverting the split-sample statistic of 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 the score-type test. Section 4 is a Monte-Carlo study investigating the finite sample behaviors of the USSIV score test and the Score-type 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.

[3] Mikusheva (2006) proposes a numerically fast algorithm to construct the K-test-based-confidence-regions for a scalar structural coefficient in a linear IV regression with one endogenous regressor.

2 Linear IV Model with Weak Instruments

2.1 Model:

Consider the following linear structural equations model:

$$y = X\beta + W\gamma + u \tag{2.1}$$

$$X = Z\Pi_X + V_X \tag{2.2}$$

$$W = Z\Pi_W + V_W \tag{2.3}$$

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 there are n observations on y, X, W and Z and we randomly split the sample 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 matrices containing the n_i observations in the ith 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 a set of high level assumptions on the joint asymptotic behavior of the structural errors and the instruments and summarize them under Assumption A.

Assumption A: We assume that the following results hold jointly as $n \to \infty$ 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.

[4] For 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.

3. $\frac{1}{\sqrt{n_i}} Z_i' \left(u_i, V_{Xi}, V_{Wi}\right) \xrightarrow{d} Q^{\frac{1}{2}} \left(\Psi_{Zui}, \Psi_{ZXi}, \Psi_{ZWi}\right)$ where $vec\left(\Psi_{Zui}, \Psi_{ZXi}, \Psi_{ZWi}\right) \sim N(0, \Sigma \otimes I_k)$.

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 relevant random functions based on sub-sample one are asymptotically uncorrelated with those based on sub-sample two. For applications, it is important to check these assumptions on a case by case basis.

2.2 Weak Instrument Framework:

The weak-instrument framework proposed by Staiger and Stock (1997) models Π as local-to-zero. This ties the instrument-strength to the sample size in a way which ensures that as the sample size goes to infinity, the instruments become completely irrelevant at a rate that gives rise to non-trivial asymptotic behavior of the estimators and test statistics for the structural parameters. In Assumption B, we characterize the rank failure of Π using the weak-instrument framework.

Assumption B: $\Pi_X = 1_{[\delta_x=1]}\mathbb{C}_X + 1_{[\delta_x=\frac{1}{2}]}\frac{\mathbb{C}_X}{\sqrt{n}}$ and $\Pi_W = 1_{[\delta_w=1]}\mathbb{C}_W + 1_{[\delta_w=\frac{1}{2}]}\frac{\mathbb{C}_W}{\sqrt{n}}$ where $\mathbb{C}_X$ and $\mathbb{C}_W$ are $k \times m_x$ and $k \times m_w$ matrices of fixed and bounded elements such that $\mathbb{C} = [\mathbb{C}_X, \mathbb{C}_W]$ is full column rank.

The non-random indicator functions involving the δ's correspond to the following four different cases of partial identification of the parameter vector $\theta = (\beta', \gamma')'$ through the instrument strength:

- Case 1: $[\delta_x = \frac{1}{2}$ and $\delta_w = \frac{1}{2}] \Rightarrow \Pi_X = \frac{\mathbb{C}_X}{\sqrt{n}}$ and $\Pi_W = \frac{\mathbb{C}_W}{\sqrt{n}}$ i.e. the instruments are weak for both β and γ and both parameter vectors are asymptotically unidentified.

- Case 2: $[\delta_x = \frac{1}{2}$ and $\delta_w = 1] \Rightarrow \Pi_X = \frac{\mathbb{C}_X}{\sqrt{n}}$ and $\Pi_W = \mathbb{C}_W$ i.e. the instruments are weak for β but strong for γ. Hence γ is identified as long as β is known, but β is asymptotically unidentified even if γ is known.

- Case 3: $[\delta_x = 1$ and $\delta_w = \frac{1}{2}] \Rightarrow \Pi_X = \mathbb{C}_X$ and $\Pi_W = \frac{\mathbb{C}_W}{\sqrt{n}}$ i.e. the instruments are strong for β but weak for γ. Hence β is identified as long as γ is known, but γ is asymptotically unidentified even if β is known.

- Case 4: $[\delta_x = 1$ and $\delta_w = 1] \Rightarrow \Pi_X = \mathbb{C}_X$ and $\Pi_W = \mathbb{C}_W$ i.e. the instruments are strong for both β and γ and both parameter vectors are identified.

The canonical representation of the four cases of partial identification mentioned above is by no means exhaustive, but it is sufficiently rich to highlight the non-degenerate asymptotic results. In these four cases, the instruments being weak for any structural coefficient can be interpreted as the lack of (statistically) "strong" finite-sample association between the instruments and the corresponding endogenous regressor.

For notational convenience we define here the following quantities to be used in the rest of the paper:

$$\begin{aligned}
&\lambda_X = Q^{\frac{1}{2}}\mathbb{C}_X \quad \text{and} \quad \lambda_W = Q^{\frac{1}{2}}\mathbb{C}_W \\
&\nu_{X1} = \lambda_X\sqrt{\zeta} + 1_{\left[\delta_x=\frac{1}{2}\right]}\Psi_{ZX1} \quad \text{and} \quad \nu_{W1} = \lambda_W\sqrt{\zeta} + 1_{\left[\delta_w=\frac{1}{2}\right]}\Psi_{ZW1} \\
&\nu_{X2} = \lambda_X\sqrt{1-\zeta} + 1_{\left[\delta_x=\frac{1}{2}\right]}\Psi_{ZX2} \quad \text{and} \quad \nu_{W2} = \lambda_W\sqrt{1-\zeta} + 1_{\left[\delta_w=\frac{1}{2}\right]}\Psi_{ZW2}.
\end{aligned}$$

3 Testing the Null Hypothesis on Subsets of Parameters

Without loss of generality let β be the parameter of interest and γ the nuisance parameter. [5]

3.1 Why Split the Sample?

We follow the exposition in Wang and Zivot (1998) to motivate sample-splitting (or equivalently using information from both sub-samples). The 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

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

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

$$\nabla_\beta J_{11}(\beta,\gamma) = \widehat{\Pi}'_{X1}Z'_1(y_1 - X_1\beta - W_1\gamma) \tag{3.1}$$

$$\nabla_\gamma J_{11}(\beta,\gamma) = \widehat{\Pi}'_{W1}Z'_1(y_1 - X_1\beta - W_1\gamma) \tag{3.2}$$

where $\widehat{\Pi}_{Xi} = (Z'_iZ_i)^{-1}Z'_iX_i$ and $\widehat{\Pi}_{Wi} = (Z'_iZ_i)^{-1}Z'_iW_i$ 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 (3.1) and (3.2) from (3.5) and (3.6) respectively.

[5] Under Assumption A, the other nuisance parameters Π and Σ can be be replaced by their $\sqrt{n}$-consistent estimators (possibly infeasible from using the true values of β and γ). This cannot be done for γ and hence the rest of the paper only explicitly addresses the different ways of partialling out the nuisance parameter γ from the inference procedure of β.

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

$$\mathcal{WZ}(\beta_0,\gamma_0) = \frac{\nabla J'_{11}(\beta_0,\gamma_0)\left(\widehat{\Pi}'_1 Z'_1 Z_1 \widehat{\Pi}_1\right)^{-1} \nabla J_{11}(\beta_0,\gamma_0)}{\frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\gamma_0)' M_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma_0)} \tag{3.3}$$

and the score test rejects the hypotheses $\beta = \beta_0$ and $\gamma = \gamma_0$ jointly at level α if $\mathcal{WZ}(\beta_0,\gamma_0) > \chi^2_m(1-\alpha)$ where $\chi^2_m(1-\alpha)$ is the $(1-\alpha)$100-th quantile of the central χ^2_m distribution. We denote this statistic by $\mathcal{WZ}$ instead of the standard LM notation to distinguish it from the different versions of score tests described in this paper. Wang and Zivot (1998) point out that under Case 1,

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

which is not a (central) χ^2_m 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 $\widehat{\Pi}_1$, used in the expression of $\nabla J_{11}(\beta,\gamma)$, are correlated (even asymptotically) in the presence of weak instruments and hence the test does not have correct size.

Dufour and Jasiak (2001) replace $\widehat{\Pi}_1$ by $\widehat{\Pi}_2$ in (3.3) and obtain the split-sample score statistic [6]

$$LM(\beta_0,\gamma_0) = \frac{\nabla J'_{21}(\beta_0,\gamma_0)\left(\widehat{\Pi}'_2 Z'_1 Z_1 \widehat{\Pi}_2\right)^{-1} \nabla J_{21}(\beta_0,\gamma_0)}{\frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\gamma_0)' M_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma_0)} \tag{3.4}$$

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

$$\nabla_\beta J_{21}(\beta,\gamma) = \widehat{\Pi}'_{X2} Z'_1(y_1 - X_1\beta - W_1\gamma) = \widehat{X}'_{12}(y_1 - X_1\beta - W_1\gamma) \tag{3.5}$$

$$\nabla_\gamma J_{21}(\beta,\gamma) = \widehat{\Pi}'_{W2} Z'_1(y_1 - X_1\beta - W_1\gamma) = \widehat{W}'_{12}(y_1 - X_1\beta - W_1\gamma) \tag{3.6}$$

and $\widehat{X}_{ij} = Z_i\widehat{\Pi}_{Xj}$ and $\widehat{W}_{ij} = Z_i\widehat{\Pi}_{Wj}$ for $i,j = 1,2$. It is probably appropriate to call (3.5) and (3.6) the "split-sample-gradients" with respect to β and γ respectively. When the instruments are strong, the split-sample-gradients, under suitable normalization, are asymptotically equivalent to the gradients given in (3.1) and (3.2).

When $\beta = \beta_0$ and $\gamma = \gamma_0$, $LM(\beta_0,\gamma_0) \xrightarrow{d} \chi^2_m$. The split-sample score test for jointly testing $\beta = \beta_0$ and $\gamma = \gamma_0$ rejects the hypotheses at level α if $LM(\beta_0,\gamma_0) > \chi^2_m(1-\alpha)$. The test is valid irrespective of the strength of the instruments. See Kleibergen (2002) for a comparison of power of the

[6] This is also referred to as the split-sample Anderson-Rubin statistic by Staiger and Stock (1997), and as the split-sample statistic by Dufour and Taamouti (2005) and Kleibergen (2002).

split-sample score test with the AR and K tests.

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 (3.6) by solving $\nabla_\gamma J_{21}(\beta_0, \gamma) = 0$ and is given by

$$\widehat{\gamma}(\beta_0) = \left(\widehat{W}'_{12} W_1\right)^{-1} \widehat{W}'_{12} \left(y_1 - X_1\beta_0\right) \tag{3.7}$$

Replacing γ_0 by $\widehat{\gamma}(\beta_0)$ in (3.4) we get what we call the USSIV score statistic. The USSIV score statistic is given by

$$\begin{aligned} LM_\beta(\beta_0) &= \frac{\nabla J'_{21}\left(\beta_0, \widehat{\gamma}(\beta_0)\right) \left(\widehat{\Pi}'_2 Z'_1 Z_1 \widehat{\Pi}_2\right)^{-1} \nabla J_{21}\left(\beta_0, \widehat{\gamma}(\beta_0)\right)}{\frac{1}{n_1}\left(y_1 - X_1\beta_0 - W_1\widehat{\gamma}(\beta_0)\right)' M_{Z_1} \left(y_1 - X_1\beta_0 - W_1\widehat{\gamma}(\beta_0)\right)} \\ &= \frac{(y_1 - X_1\beta_0)' \widetilde{M}'_{W_{1,12}} \widehat{X}_{12} \left(\widehat{X}'_{12} M_{\widehat{W}_{12}} \widehat{X}_{12}\right)^{-1} \widehat{X}'_{12} \widetilde{M}_{W_{1,12}} (y_1 - X_1\beta_0)}{\frac{1}{n_1}(y_1 - X_1\beta_0)' \widetilde{M}'_{W_{1,12}} M_{Z_1} \widetilde{M}_{W_{1,12}} (y_1 - X_1\beta_0)} \end{aligned} \tag{3.8}$$

where $\widetilde{M}_{W_{1,12}} := I_{n_1} - W_1(\widehat{W}'_{12} W_1)^{-1} \widehat{W}'_{12}$. We define the USSIV score test as the test which rejects the null hypothesis $H_0 : \beta = \beta_0$ at level α if $LM_\beta(\beta_0) > \chi^2_{m_x}(1 - \alpha)$.

Theorem 3.1 *Let $\beta = \beta_0 + \frac{d_\beta}{\sqrt{n}}$ where $d_\beta = O(1)$. Then under Assumptions A and B, the following results hold as $n \to \infty$:*

1. $n^{\delta_w - \frac{1}{2}}(\widehat{\gamma}(\beta_0) - \gamma) \xrightarrow{d} (\nu'_{w2}\nu_{w1})^{-1}\nu'_{w2} \left[\Psi_{Zu1} + 1_{[\delta_x = 1]}\sqrt{\zeta}\lambda_X d_\beta\right] = \Delta(\beta_0)$ *(say).*
2. $LM_\beta(\beta_0) \xrightarrow{d} \dfrac{\eta'(\beta_0)\left[I_k - \nu_{w2}(\nu'_{w1}\nu_{w2})^{-1}\nu'_{w1}\right] \nu_{x2}(\nu'_{x2} M_{\nu_{w2}} \nu_{x2})^{-1} \nu'_{x2} [I_k - \nu_{w1}(\nu'_{w2}\nu_{w1})^{-1}\nu'_{w2}]\eta(\beta_0)}{\sigma_{uu}(\Delta(\beta_0))}$
   where $\eta(\beta_0) = \Psi_{Zu1} + 1_{[\delta_x = 1]}\nu_{x1} d_\beta$ *and* $\sigma_{uu}(\Delta(\beta_0)) = \sigma_{uu} + 1_{\left[\delta_w = \frac{1}{2}\right]} \Delta'(\beta_0)\left[\sigma_{WW}\Delta(\beta_0) - 2\sigma_{Wu}\right]$.

Remarks: The following observations are worth mentioning here.

1. Under Cases 2 and 4 the USSIV estimator of γ, evaluated at the true β, is $\sqrt{n}$-consistent. Under Case 2 $LM_\beta(\beta_0) \xrightarrow{d} \frac{1}{\sigma_{uu}} \Psi'_{Zu1} P_{[M_{\lambda_W}(\lambda_X\sqrt{1-\zeta} + \Psi_{ZX2})]} \Psi_{Zu1} \sim \chi^2_{m_x}$ and under Case 4 $LM_\beta(\beta_0) \xrightarrow{d} \frac{1}{\sigma_{uu}} \left[\Psi_{Zu1} + \sqrt{\zeta}\lambda_X d_\beta\right]' P_{M_{\lambda_W}\lambda_X} \left[\Psi_{Zu1} + \sqrt{\zeta}\lambda_X d_\beta\right] \sim \chi^2_{m_x}(\sigma^{-1}_{uu}\zeta d'_\beta \lambda'_X M_{\lambda_W} \lambda_X d_\beta)$. Hence in Case

2 the USSIV score test is asymptotically equivalent to the K-test for subsets of parameters (based on sub-sample one) under the null hypothesis, and in Case 4 the asymptotic equivalence holds for any $\sqrt{n}$-local alternative.

2. The USSIV estimator of γ is, however, inconsistent under Cases 1 and 3 when the instruments are weak for γ (the estimator is asymptotically unbiased under the null hypothesis if $\sigma_{uW} = 0$). Not surprisingly, the limiting $\chi^2_{m_x}$ distribution of the USSIV score statistic does not hold under these two cases. Simulation results in the next section show that under Cases 1 and 3, the $\chi^2_{m_x}$ does not stochastically dominate the limiting distribution of the USSIV score statistic under the null hypothesis. Hence, unlike the partial K-test, the USSIV score test is not valid when the instruments are weak for the nuisance parameter γ [see Kleibergen (2004) and Kleibergen (2007)].

Confidence regions for β can be constructed by inverting the USSIV score test as follows:

$$C_{\beta}^{USSIV}(1-\alpha) = \left\{\beta_0 : LM_{\beta}(\beta_0) \leq \chi^2_{m_x}(1-\alpha)\right\} = \{\beta_0 : \beta_0' A_U \beta_0 - 2B_U\beta_0 + C_U \leq 0\} \tag{3.9}$$

where, following Dufour and Taamouti (2005), we define $A_U = X_1' H_U X_1$, $B_U = y_1' H_U X_1$, $C_U = y_1' H_U y_1$ and $H_U = \widetilde{M}'_{W_{1,12}} \left[\widehat{X}_{12}\left(\widehat{X}'_{12} M_{\widehat{W}_{12}} \widehat{X}_{12}\right)^{-1} \widehat{X}'_{12} - \frac{\chi^2_{m_x}(1-\alpha)}{n_1} M_{Z_1}\right] \widetilde{M}_{W_{1,12}}$. As we have already observed from Theorem 3.1, the confidence region in (3.9) has asymptotic coverage probability $1-\alpha$ when the instruments are strong for the nuisance parameter γ. Furthermore, when the instruments are strong for both β and γ, i.e. under Case 4, the asymptotic length of this region is same as that of the regions obtained by inverting the standard score test and the K-test for β (the later two based on sub-sample one). Simulation results in the next section show that the (asymptotic) coverage probability of $C_{\beta}^{USSIV}(1-\alpha)$ is much less than $1-\alpha$ when the instruments are weak for the nuisance parameter, i.e. under Cases 1 and 3.

Now we define a new statistic which helps in exploring the asymptotic relationship between the USSIV score test discussed in this subsection and the Score-type test (based on Robins' method) discussed in the next subsection. Define $LM^*_{\beta}(\beta,\gamma)$ as

$$LM^*_{\beta}(\beta,\gamma) = \frac{(y_1 - X_1\beta - W_1\gamma)' P_{\left[M_{\widehat{W}_{12}}\widehat{X}_{12}\right]} (y_1 - X_1\beta - W_1\gamma)}{\frac{1}{n_1}(y_1 - X_1\beta - W_1\gamma)' M_{Z_1} (y_1 - X_1\beta - W_1\gamma)}. \tag{3.10}$$

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

$$\mathcal{WZ}_\beta(\beta_0) = \frac{(y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))' P_{\left[M_{\widehat{W}_{11}}\widehat{X}_{11}\right]}(y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))}{\widehat{\sigma}_{uu}^{2SLS}(\beta_0)} > \chi^2_{m_x}(1-\alpha) \tag{3.11}$$

where $\widehat{\gamma}_{2SLS}(\beta_0) = (\widehat{W}_{11}'\widehat{W}_{11})^{-1}\widehat{W}_{11}'(y_1 - X_1\beta_0)$ 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 is $\widehat{\sigma}_{uu}^{2SLS}(\beta_0) = \frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))' M_{Z_1}(y_1 - X_1\beta_0 - W_1\widehat{\gamma}_{2SLS}(\beta_0))$. In the context of maximum likelihood (for example, if the structural errors are *iid* Gaussian), we can think of the term $\left(\widehat{\sigma}_{uu}^{2SLS}(\beta_0)\widehat{X}_{11}' M_{\widehat{W}_{11}}\widehat{X}_{11}\right)^{-1}$ 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

$$\mathcal{WZ}^*_\beta(\beta_0) = \frac{(y_1 - X_1\beta_0 - W_1\gamma)' P_{\left[M_{\widehat{W}_{11}}\widehat{X}_{11}\right]}(y_1 - X_1\beta_0 - W_1\gamma)}{\frac{1}{n_1}(y_1 - X_1\beta_0 - W_1\gamma)' M_{Z_1}(y_1 - X_1\beta_0 - W_1\gamma)}. \tag{3.12}$$

The USSIV score statistic given in (3.8) is obtained by replacing $\widehat{X}_{11}$ by $\widehat{X}_{12}$ and $\widehat{W}_{11}$ by $\widehat{W}_{12}$ in the expression of $WZ_\beta(\beta_0)$ in (3.11) and the USSIV estimator of γ, (i.e. $\widehat{\gamma}(\beta_0)$ in (3.7)) 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 (3.10) can be thought of as a split-sample version of the efficient score statistic $WZ^*_\beta(\beta_0, \gamma)$ given in (3.12).

Lemma 3.2 *Let* $\beta = \beta_0 + \frac{d_\beta}{\sqrt{n}}$ *and* $\gamma = \gamma_0 + \frac{d_\gamma}{\sqrt{n}}$ *such that* d_β *and* d_γ *are* $O(1)$. *Then under Assumptions A and B, as* $n \to \infty$:

$$LM^*_\beta(\beta_0, \gamma_0) \xrightarrow{d} \frac{1}{\sigma_{uu}}\eta^{*'}(\beta_0, \gamma_0) P_{M_{\nu_{w2}}\nu_{x2}}\eta^*(\beta_0, \gamma_0) \tag{3.13}$$

where $\eta^*(\beta_0, \gamma_0) = \left[\Psi_{Zu1} + \sqrt{\zeta}\left(1_{[\delta_x=1]}\lambda_X d_\beta + 1_{[\delta_w=1]}\lambda_W d_\gamma\right)\right]$.

Remarks: Several remarks are in order here.

1. At the true β, $LM^*_\beta(\beta, \gamma_0) \xrightarrow{d} \frac{1}{\sigma_{uu}}[\Psi_{Zu1} + 1_{[\delta_w=1]}\lambda_W d_\gamma] P_{M_{\nu_{w2}}\nu_{x2}}[\Psi_{Zu1} + 1_{[\delta_w=1]}\lambda_W d_\gamma] \sim \chi^2_{m_x}$ for any γ_0 in the $\sqrt{n}$-neighborhood of the true γ.

2. When the instruments are strong for the nuisance parameter γ, i.e. under Cases 2 and 4, $LM^*_\beta(\beta_0, \gamma_0) = LM_\beta(\beta_0) + o_p(1)$ for any β_0 and γ_0 in the $\sqrt{n}$-neighborhood of the true β and γ. Hence under Cases 2 and 4 and against $\sqrt{n}$-local alternatives, the USSIV score test is asymptotically equivalent to the infeasible efficient score test which rejects $H_0 : \beta = \beta_0$ at level α if $LM^*_\beta(\beta_0, \gamma) > \chi^2_{m_x}(1-\alpha)$. The later test is infeasible because it uses the unknown true value of γ. It is very important to note that the asymptotic equivalence holds for any γ_0 in the $\sqrt{n}$-neighborhood of the true γ.

Lemma 3.2 holds for any γ_0 in the $\sqrt{n}$-neighborhood of true γ. Hence, if we can obtain a $\sqrt{n}$-consistent estimator of γ, Lemma 3.2 can be used to construct a valid test for the null hypothesis $H_0 : \beta = \beta_0$ regardless of the instrument relevance. The USSIV score test fails under the weak-instrument asymptotics because, even under the null hypothesis, the USSIV estimator $\widehat{\gamma}(\beta_0)$ 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 the 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. [7] Based on the consequences of Lemma 3.2, under Cases 2 and 4 and against $\sqrt{n}$-local alternatives, the USSIV score test can be considered as a benchmark for the tests based only on sub-sample one and the split-sample tests.

3.3 A new Score-type Test based on Robins' method:

In this section we propose a new test for $H_0 : \beta = \beta_0$ based on a general technique proposed by Robins (2004). We call this the Score-type test based on Robins' method. This test is always valid and when the instruments are strong for the nuisance parameter γ, this is asymptotically equivalent to the USSIV score test against $\sqrt{n}$-local alternatives. The Score-type test 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 the split-sample score statistic $LM(\beta_0, \gamma_0)$ rejects the null hypothesis $H_0 : \beta = \beta_0$ at level α when

$$\inf_{\gamma_0 \in \Theta_\gamma} LM(\beta_0, \gamma_0) > \chi^2_m(1-\alpha) \tag{3.14}$$

Assuming that the model is correctly specified, we note that $\inf_{\gamma_0 \in \mathbb{R}^{m_w}} LM(\beta_0, \gamma_0) = \kappa^\dagger_n(\beta_0)$ where

[7] It should be noted that under Cases 1 and 3, the validity of the K-test for β does not depend on the $\sqrt{n}$-consistency of the restricted limited information maximum likelihood estimator of γ [see Kleibergen (2007)].

$\kappa_n^\dagger(\beta_0)$ is the smallest eigen value of the matrix

$$\Xi_P(\beta_0) = \left[\frac{1}{n_1}(y_1 - X_1\beta_0, W_1)' M_{Z_1}(y_1 - X_1\beta_0, W_1)\right]^{-1} \left[(y_1 - X_1\beta_0, W_1)' P_{[\widehat{X}_{12}, \widehat{W}_{12}]}(y_1 - X_1\beta_0, W_1)\right] \tag{3.15}$$

Hence, the projection based test rejects $H_0 : \beta = \beta_0$ at level α if the smallest eigen value of $\Xi_P(\beta_0)$ is greater than $\chi^2_m(1-\alpha)$. Alternatively, the null hypothesis can be rejected if the set $\{\gamma_0 : \gamma_0' A_P \gamma_0 - 2B_P\gamma_0 + C_P < 0\}$ is empty where $A_P = W_1' H_P W_1$, $B_P = (y_1 - X_1\beta_0)' H_P W_1$, $C_P = (y_1 - X_1\beta_0)' H_P (y_1 - X_1\beta_0)$ and $H_P = \left[P_{[\widehat{X}_{12}, \widehat{W}_{12}]} - \frac{\chi^2_m(1-\alpha)}{n_1} M_{Z_1}\right]$. However, the projection test can be very conservative when the instruments are strong for γ and/or β and when the dimension of γ is large relative to that of β.

Now we describe the new Score-type test based on Robins' method. Suppose that, given a specific value β_0, it is possible to construct a $1-\epsilon$ confidence region for γ such that when γ is identified, this region, if non-empty, belongs to the $\sqrt{n}$-neighborhood of γ with probability one. Denote this region by $C_\gamma(1-\epsilon, \beta_0)$. The Score-type test rejects the null hypothesis $H_0 : \beta = \beta_0$ if either $C_\gamma(1-\epsilon, \beta_0)$ is empty or if $\inf_{\gamma_0 \in C_\gamma(1-\epsilon,\beta_0)} LM^*_\beta(\beta_0, \gamma_0) > \chi^2_{m_x}(1-\alpha)$. Theorem 3.3, stated below, helps in understanding the construction and the properties of the Score-type test.

Theorem 3.3 *Let* $LM_\gamma(\beta_0, \gamma_0) := \dfrac{(y_1 - X_1\beta_0 - W_1\gamma_0)' P_{\widehat{W}_{12}}(y_1 - X_1\beta_0 - W_1\gamma_0)}{\frac{1}{n_1}(y_1 - X_1\beta_0 - W\gamma_0)' M_{Z_1}(y_1 - X_1\beta_0 - W\gamma_0)}$ *and define a* $1-\epsilon$ *confidence region for* γ *as* $C_\gamma(1-\epsilon, \beta_0) = \left\{\gamma_0 : LM_\gamma(\beta_0, \gamma_0) \le \chi^2_{m_w}(1-\epsilon)\right\}$. *Then under Assumptions A and B, the following results hold for* $\beta = \beta_0 + \frac{d_\beta}{\sqrt{n}}$ *(where* d_β *is O(1)) as* $n \to \infty$*:*

1. $LM_\gamma(\beta_0, \gamma_0) \xrightarrow{d} \dfrac{[\Psi_{Zu1} + \Psi_{ZW1}(\gamma - \gamma_0) + \sqrt{\zeta}\xi(\beta_0, \gamma_0)]' P_{\nu_{w2}} [\Psi_{Zu1} + \Psi_{ZW1}(\gamma - \gamma_0) + \sqrt{\zeta}\xi(\beta_0, \gamma_0)]}{\sigma_{uu} + 2\sigma_{uW}(\gamma - \gamma_0) + (\gamma - \gamma_0)'\sigma_{WW}(\gamma - \gamma_0)}$
   where $\xi(\beta_0, \gamma_0) = \lim_{n\to\infty} n^{\delta_w - \frac{1}{2}} \lambda_W(\gamma - \gamma_0) + 1_{[\delta_x = 1]}\lambda_X d_\beta$.

2. *under Cases 2 and 4,* $\inf_{\gamma_0 \in C_\gamma(1-\epsilon,\beta_0)} LM^*_\beta(\beta_0, \gamma_0) = LM^*_\beta(\beta_0, \gamma) + o_p(1)$.

Remarks:

1. Recall that, by definition, the USSIV estimator $\widehat{\gamma}(\beta_0)$ satisfies the restriction $\widehat{W}_{12}'(y_1 - X_1\beta_0 - W_1\widehat{\gamma}(\beta_0)) = 0$ and hence $LM_\gamma(\beta_0, \widehat{\gamma}(\beta_0)) = 0$. Therefore, the confidence region $C_\gamma(1-\epsilon, \beta_0)$, defined in the statement of Theorem 3.3 is always non-empty.

2. From Theorem 3.3 it is easy to see that at the true value of β and γ, $LM_\gamma(\beta, \gamma) \xrightarrow{d} \chi^2_{m_w}$ and hence $C_\gamma(1-\epsilon, \beta_0)$ has the correct coverage probability $1-\epsilon$ under the null hypothesis

$H_0 : \beta = \beta_0$. Noting that $\inf_{\gamma_0 \in C_\gamma(1-\epsilon,\beta)} LM^*_\beta(\beta,\gamma_0) \leq LM^*_\beta(\beta,\gamma)$ and using Lemma 3.2 along with standard Bonferroni arguments, it follows that the size of the Score-type test cannot exceed $\alpha + \epsilon$.

3. The asymptotic equivalence in the second statement of Theorem 3.3 depends crucially on the fact that under Cases 2 and 4, $C_\gamma(1-\epsilon,\beta_0)$ belongs to the $\sqrt{n}$-neighborhood of the true γ with probability one. $C_\gamma(1-\epsilon,\beta_0)$ satisfies this criterion because under Cases 2 and 4

$$LM_\gamma(\beta_0,\gamma_0) \xrightarrow{d} \chi^2_{m_w}\left(\lim_{n\to\infty} \zeta \frac{(\sqrt{n}\lambda_W(\gamma-\gamma_0) + 1_{[\delta_x=1]}\lambda_X d_\beta)' P_{\lambda_W}(\sqrt{n}\lambda_W(\gamma-\gamma_0) + 1_{[\delta_x=1]}\lambda_X d_\beta)'}{\sigma_{uu} + 2\sigma_{uW}(\gamma-\gamma_0) + (\gamma-\gamma_0)'\sigma_{WW}(\gamma-\gamma_0)}\right)$$

and the non-centrality parameter of the limiting distribution can be finite only for γ_0 in the $\sqrt{n}$-neighborhood of the true γ. The second statement of the theorem implies that under Cases 2 and 4 and for $\sqrt{n}$-local alternatives the Score-type test is asymptotically equivalent to the infeasible efficient score test rejecting $H_0 : \beta = \beta_0$ at level α if $LM^*_\beta(\beta_0,\gamma) > \chi^2_{m_x}(1-\alpha)$. Lemma 3.2 extends the asymptotic equivalence to the USSIV score test.

The whole idea behind the Score-type test based on Robins' method can be summarized as follows. The unconstrained infimum $\inf_{\gamma_0 \in \mathbb{R}^{m_w}} LM^*_\beta(\beta_0,\gamma_0) = \kappa^\dagger_n(\beta_0)$ where $\kappa^\dagger_n(\beta_0)$ is given by the smallest eigen value of the matrix

$$\Xi(\beta_0) = \left[\frac{1}{n_1}(y_1 - X_1\beta_0, W_1)' M_{Z_1}(y_1 - X_1\beta_0, W_1)\right]^{-1}\left[(y_1 - X_1\beta_0, W_1)' P_{M_{\widehat{W}_{12}}\widehat{X}_{12}}(y_1 - X_1\beta_0, W_1)\right]. \tag{3.16}$$

By definition, $\kappa^\dagger_n(\beta_0) \leq LM^*_\beta(\beta_0,\gamma)$ and hence the test which rejects $H_0 : \beta = \beta_0$ if $\kappa^\dagger_n(\beta_0) > \chi^2_{m_x}(1-\alpha)$ cannot be over-sized. Like the projection test, this particular test can be extremely conservative. The conservativeness of the test is not particularly unattractive when the instruments are weak for the unknown nuisance parameter γ and thus rendering the parameter β asymptotically unidentified. However, there should be scope for improving the power of this test when all the parameters (in particular, γ) are identified. According to Lemma 3.2, this is possible if the value of γ where the infimum of $LM^*_\beta(\beta_0,\gamma)$ is attained can be restricted to the $\sqrt{n}$-neighborhood of the true γ. Denoting the value of γ where the unconstrained infimum is attained by $\gamma^{uc}_{inf}(\beta_0)$, we note that

$$\gamma^{uc}_{inf}(\beta_0) = \left[W_1'\left(P_{M_{\widehat{W}_{12}}\widehat{X}_{12}} - \frac{\kappa^\dagger_n(\beta_0)}{n_1}M_{Z_1}\right)W_1\right]^{-1} W_1'\left(P_{M_{\widehat{W}_{12}}\widehat{X}_{12}} - \frac{\kappa^\dagger_n(\beta_0)}{n_1}M_{Z_1}\right)(y_1 - X_1\beta_0). \tag{3.17}$$

and there is no way to guarantee that $\gamma^{uc}_{inf}(\beta_0)$ is in the $\sqrt{n}$ neighborhood of the true γ. Robins' method allows for an asymptotic trade-off between the size and the power by considering a constrained infimum of the $LM^*_\beta(\beta_0,\gamma)$. Restricting γ to take values in $C_\gamma(1-\epsilon,\beta_0)$ has two important consequences: 1) when the instruments are strong for γ, the constrained infimum of $LM^*_\beta(\beta_0, \gamma)$ can only be attained in the $\sqrt{n}$-neighborhood of the true γ and thus, from Lemma 3.2 and Theorem 3.3, the power of the Score-type test is asymptotically same as the power of the benchmark USSIV score test against $\sqrt{n}$-local alternatives; 2) on the other hand when the instruments are weak for γ, $C_\gamma(1-\epsilon, \beta_0)$ contains the true value of γ with probability $1-\epsilon$ under the null hypothesis and hence the size of the Score-type test is guaranteed not to exceed $\alpha+\epsilon$. To be more conservative about the size, one can choose a small ϵ without affecting the asymptotic power of the Score-type test under Cases 2 and 4 (i.e. when the instruments are strong for the nuisance parameter γ). Simulation results in the next section show that the $\alpha + \epsilon$ is a conservative upper-bound for the size of the Score-type test and asymptotic improvement in power of the test can be made possible by choosing larger values of ϵ. Finally it should be noted that the construction of $C_\gamma(1-\epsilon, \beta_0)$ is not unique, and the AR statistic, the K-statistic and the LR statistic can be used to construct the region. Our particular choice in Theorem 3.3 simply maintains the uniformity of the presentation and we do not make any optimality statement about the choice of $C_\gamma(1-\epsilon, \beta_0)$.

Like other tests described in this paper, it is possible to obtain confidence regions for β by inverting the Score-type test. In particular, the region

$$C_\beta^{Robins}(1-\alpha-\epsilon) = \left\{\beta_0 : C_\gamma(1-\epsilon, \beta_0) \neq \varnothing, inf_{\gamma_0 \in C_\gamma(1-\epsilon,\beta_0)} LM^*_\beta(\beta_0, \gamma_0) \leq \chi^2_{m_x}(1-\alpha)\right\} \quad (3.18)$$

always has asymptotic coverage probability at least $1-\alpha-\epsilon$ and when the instruments are strong for the nuisance parameter γ, this region has the same asymptotic length and coverage probability as the infeasible region $C_\beta^{infeas}(1-\alpha) = \left\{\beta_0 : LM^*_\beta(\beta_0, \gamma) \leq \chi^2_{m_x}(1-\alpha)\right\}$ using the true value of the unknown nuisance parameter γ (and hence same as the region given in (3.9)).

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 (2.1) – (2.3).

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

$$\Sigma = \begin{pmatrix} 1 & \rho_{uX} & \rho_{uW} \\ \rho_{Xu} & 1 & 0 \\ \rho_{Wu} & 0 & 1 \end{pmatrix} \quad (4.1)$$

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 (4.1) 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 (ρ_{uX}, ρ_{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]

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 individual structural coefficients 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 (4.2) 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

$$\mu = \frac{1}{k} \begin{pmatrix} \sigma_{XX} & \sigma_{XW} \\ \sigma_{WX} & \sigma_{WW} \end{pmatrix}^{-\frac{1}{2}'} \begin{pmatrix} n^{\delta_x - \frac{1}{2}} \lambda_X' \\ n^{\delta_w - \frac{1}{2}} \lambda_W' \end{pmatrix} \begin{pmatrix} n^{\delta_x - \frac{1}{2}} \lambda_X' \\ n^{\delta_w - \frac{1}{2}} \lambda_W' \end{pmatrix}' \begin{pmatrix} \sigma_{XX} & \sigma_{XW} \\ \sigma_{WX} & \sigma_{WW} \end{pmatrix}^{-\frac{1}{2}} = \begin{pmatrix} \mu_\beta & 0 \\ 0 & \mu_\gamma \end{pmatrix} \tag{4.2}$$

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 ith element of Π_X is taken as $\sqrt{\frac{\mu_\beta}{n}}$ and the ith element of Π_W is taken as $(-1)^i \sqrt{\frac{\mu_\gamma}{n}}$ for $i = 1, \ldots, k$. [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 $n_1 = [n\zeta]$ observations and the second sub-sample contains $n_2 = n - n_1$ observations. Finally, following the

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

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

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 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 Score-type 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 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} (under our construction 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 Score-type 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 Score-type 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.

[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

[10] In our simulations, we always choose $\epsilon = \alpha$ for the Score-type test.

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 $\zeta = 50\%$). [11] 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 Score-type 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 Score-type 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 the instruments are strong for γ, the Score-type 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 instruments in a single endogenous regressor model, 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 coefficient) 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 Score-type 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

[11] Staiger and Stock (1997) mention that the weak-instrument asymptotics gives reasonably good approximation with 10-20 observations per instrument.

strong for both β and γ, the powers of the Score-type 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 can be greater than that of the projection test based on the AR-statistic.

The simulations show that the instruments for γ have to be very strong for the asymptotic equivalence between the USSIV score test and the Score-type test to hold. However, since $\alpha + \epsilon$ seems to be a too-conservative upper-bound for the size of the Score-type test, it is possible to attain the asymptotic equivalence even with moderately strong instruments by choosing a larger value of ϵ. Note that the choice of ϵ should not matter asymptotically under Cases 2 and 4 (i.e. when the instruments are strong for γ).

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 "unrelated" 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 Score-type 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 Score-type test. In any case, the Score-type 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 Score-type 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, at the same time, are 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, to 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.

A Appendix:

A.1 Proof of Results:

Lemma A.1 *The following results hold jointly under Assumptions A and B as $n \to \infty$:*

1. $n^{\frac{1}{2}-\delta_x}(Z_1'Z_1)^{-\frac{1}{2}}Z_1'X_1 \xrightarrow{d} \lambda_X\sqrt{\zeta} + 1_{[\delta_x=\frac{1}{2}]}\Psi_{ZX1} \equiv \nu_{x1}$.
2. $n^{\frac{1}{2}-\delta_w}(Z_1'Z_1)^{-\frac{1}{2}}Z_1'W_1 \xrightarrow{d} \lambda_W\sqrt{\zeta} + 1_{[\delta_w=\frac{1}{2}]}\Psi_{ZW1} \equiv \nu_{w1}$.
3. $n^{\frac{1}{2}-\delta_x}(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{X2} \xrightarrow{d} \sqrt{\zeta}\left[\lambda_X + 1_{[\delta_x=\frac{1}{2}]}(1-\zeta)^{\delta_x-1}\Psi_{ZX2}\right] \equiv \sqrt{\frac{\zeta}{1-\zeta}}\nu_{x2}$
4. $n^{\frac{1}{2}-\delta_w}(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{W2} \xrightarrow{d} \sqrt{\zeta}\left[\lambda_W + 1_{[\delta_w=\frac{1}{2}]}(1-\zeta)^{\delta_w-1}\Psi_{ZW2}\right] \equiv \sqrt{\frac{\zeta}{1-\zeta}}\nu_{w2}$

Proof of Lemma A.1 Using Assumptions A and B it follows that:

1. $n^{\frac{1}{2}-\delta_x}\left(\frac{Z_1'Z_1}{n_1}\right)^{-\frac{1}{2}}\frac{Z_1'X_1}{\sqrt{n_1}} = \left(\frac{Z_1'Z_1}{n_1}\right)^{\frac{1}{2}'}\mathbb{C}_X\left(\frac{n_1}{n}\right)^{\frac{1}{2}} + \left(\frac{Z_1'Z_1}{n_1}\right)^{-\frac{1}{2}}\frac{Z_1'X_1}{\sqrt{n_1}}n^{\frac{1}{2}-\delta_x} \xrightarrow{d} \nu_{x1}$.
2. $n^{\frac{1}{2}-\delta_w}\left(\frac{Z_1'Z_1}{n_1}\right)^{-\frac{1}{2}}\frac{Z_1'W_1}{\sqrt{n_1}} = \left(\frac{Z_1'Z_1}{n_1}\right)^{\frac{1}{2}'}\mathbb{C}_W\left(\frac{n_1}{n}\right)^{\frac{1}{2}} + \left(\frac{Z_1'Z_1}{n_1}\right)^{-\frac{1}{2}}\frac{Z_1'W_1}{\sqrt{n_1}}n^{\frac{1}{2}-\delta_w} \xrightarrow{d} \nu_{w1}$.
3. $n^{\frac{1}{2}-\delta_x}(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{X2} = \left(\frac{n_1}{n}\right)^{\frac{1}{2}}\left(\frac{Z_1'Z_1}{n_1}\right)^{\frac{1}{2}'}\left[\mathbb{C}_X + \left(\frac{Z_2'Z_2}{n_2}\right)^{-1}\frac{Z_2'V_{X2}}{n^{\delta_x}}\left(\frac{n}{n_2}\right)^{\delta_x-1}\right] \xrightarrow{d} \sqrt{\frac{\zeta}{1-\zeta}}\nu_{x2}$.
4. $n^{\frac{1}{2}-\delta_w}(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{W2} = \left(\frac{n_1}{n}\right)^{\frac{1}{2}}\left(\frac{Z_1'Z_1}{n_1}\right)^{\frac{1}{2}'}\left[\mathbb{C}_W + \left(\frac{Z_2'Z_2}{n_2}\right)^{-1}\frac{Z_2'V_{W2}}{n^{\delta_w}}\left(\frac{n}{n_2}\right)^{\delta_w-1}\right] \xrightarrow{d} \sqrt{\frac{\zeta}{1-\zeta}}\nu_{w2}$. ∎

Proof of Theorem 3.1 Part 1: Using Lemma A.1 it follows that,

$$
\begin{aligned}
n^{\delta_w-\frac{1}{2}}(\widehat{\gamma}(\beta_0)-\gamma) &= \left[\frac{\widehat{\Pi}_{W2}'(Z_1'Z_1)^{\frac{1}{2}}}{n^{\delta_w-\frac{1}{2}}}\frac{(Z_1'Z_1)^{-\frac{1}{2}}Z_1'W_1}{n^{\delta_w-\frac{1}{2}}}\right]^{-1}\frac{\widehat{\Pi}_{W2}'(Z_1'Z_1)^{\frac{1}{2}}}{n^{\delta_w-\frac{1}{2}}} \\
&\quad \times\left[\left(\frac{Z_1'Z_1}{n_1}\right)^{-\frac{1}{2}}\frac{Z_1'}{\sqrt{n_1}}\left(u_1 + V_{X1}\frac{d_\beta}{\sqrt{n}}\right) + \sqrt{\frac{n_1}{n}}\left(\frac{Z_1'Z_1}{n_1}\right)^{\frac{1}{2}'}\Pi_X d_\beta\right] \\
&\xrightarrow{d} (\nu_{w2}'\nu_{w1})^{-1}\nu_{w2}'\left[\Psi_{Zu1} + 1_{[\delta_x=1]}\sqrt{\zeta}\lambda_X d_\beta\right] \equiv \Delta(\beta_0).
\end{aligned}
$$

Part 2: Using Lemma A.1 and noting that

- $n^{1-2\delta_x}\widehat{X}_{12}'\widehat{X}_{12}' = \left[n^{\frac{1}{2}-\delta_x}\widehat{\Pi}_{X2}'(Z_1'Z_1)^{\frac{1}{2}}\right]\left[n^{\frac{1}{2}-\delta_x}(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{X2}'\right] \xrightarrow{d} \frac{\zeta}{1-\zeta}\nu_{x2}'\nu_{x2},$
- $n^{1-\delta_x-\delta_w}\widehat{X}_{12}'\widehat{W}_{12}' = \left[n^{\frac{1}{2}-\delta_x}\widehat{\Pi}_{X2}'(Z_1'Z_1)^{\frac{1}{2}}\right]\left[n^{\frac{1}{2}-\delta_w},(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{W2}'\right] \xrightarrow{d} \frac{\zeta}{1-\zeta}\nu_{x2}'\nu_{w2},$
- $n^{1-2\delta_w}\widehat{W}_{12}'\widehat{W}_{12}' = \left[n^{\frac{1}{2}-\delta_w}\widehat{\Pi}_{W2}'(Z_1'Z_1)^{\frac{1}{2}}\right]\left[n^{\frac{1}{2}-\delta_w}(Z_1'Z_1)^{\frac{1}{2}'}\widehat{\Pi}_{W2}'\right] \xrightarrow{d} \frac{\zeta}{1-\zeta}\nu_{w2}'\nu_{w2},$

it follows that $n^{\frac{1}{2}-\delta_x}\widehat{X}_{12}'\widetilde{M}_{W_{1,12}}(y_1-X_1\beta_0) \xrightarrow{d} \sqrt{\zeta}\nu_{x2}'[I_k-\nu_{w1}(\nu_{w2}'\nu_{w1})^{-1}\nu_{w2}'][\Psi_{Zu1}+1_{[\delta_x=1]}\nu_{x1}d_\beta]$ and hence

$$\begin{aligned} LM_\beta(\beta_0) &= \frac{\left[\frac{\widehat{X}_{12}'\widetilde{M}_{W_{1,12}}(y_1-X_1\beta_0)}{n^{\delta_x-\frac{1}{2}}}\right]'\left[\frac{\widehat{X}_{12}'\widehat{X}_{12}'}{n^{2\delta_x-1}}-\frac{\widehat{X}_{12}'\widehat{W}_{12}'}{n^{\delta_x+\delta_w-1}}\left(\frac{\widehat{W}_{12}'\widehat{W}_{12}'}{n^{2\delta_w-1}}\right)^{-1}\frac{\widehat{W}_{12}'\widehat{X}_{12}'}{n^{\delta_x+\delta_w-1}}\right]^{-1}\left[\frac{\widehat{X}_{12}'\widetilde{M}_{W_{1,12}}(y_1-X_1\beta_0)}{n^{\delta_x-\frac{1}{2}}}\right]}{\frac{1}{n_1}(y_1-X_1\beta_0)'\widetilde{M}_{W_{1,12}}'M_{Z_1}\widetilde{M}_{W_{1,12}}(y_1-X_1\beta_0)} \\ &\xrightarrow{d} \frac{\eta'(\beta_0)\left[I_k-\nu_{w2}(\nu_{w1}'\nu_{w2})^{-1}\nu_{w1}'\right]\nu_{x2}(\nu_{x2}'M_{\nu_{w2}}\nu_{x2})^{-1}\nu_{x2}'[I_k-\nu_{w1}(\nu_{w2}'\nu_{w1})^{-1}\nu_{w2}']\eta(\beta_0)}{\sigma_{uu}(\Delta(\beta_0))} \quad \blacksquare \end{aligned}$$

Proof of Lemma 3.2 Using Lemma A.1 note that:

$$\begin{aligned} \frac{\widehat{X}_{12}'(y_1-X_1\beta_0-W_1\gamma_0)}{n^{\delta_x-\frac{1}{2}}} &= \left[\frac{\widehat{\Pi}_{X2}'(Z_1'Z_1)^{\frac{1}{2}}}{n^{\delta_x-\frac{1}{2}}}\right]\left(\frac{Z_1'Z_1}{n_1}\right)^{-\frac{1}{2}}\left(\frac{Z_1'u_1}{\sqrt{n_1}}+\sqrt{\zeta}\frac{Z_1'V_{X1}}{n_1}d_\beta+\sqrt{\zeta}\frac{Z_1'V_{W1}}{n_1}d_\gamma\right) \\ &\quad+\left[\frac{\widehat{\Pi}_{X2}'(Z_1'Z_1)^{\frac{1}{2}}}{n^{\delta_x-\frac{1}{2}}}\right]\sqrt{\zeta}\left(\frac{Z_1'Z_1}{n_1}\right)^{\frac{1}{2}'}[\mathbb{C}_X d_\beta+\mathbb{C}_W d_\gamma] \\ &\xrightarrow{d} \sqrt{\frac{\zeta}{1-\zeta}}\nu_{x2}'\left[\Psi_{Zu1}+\sqrt{\zeta}\left(1_{[\delta_x=1]}\lambda_X d_\beta+1_{[\delta_w=1]}\lambda_W d_\gamma\right)\right]=\nu_{x2}'\eta^*(\beta_0,\gamma_0) \end{aligned}$$

and similarly

$$\frac{\widehat{W}_{12}'(y_1-X_1\beta_0-W_1\gamma_0)}{n^{\delta_w-\frac{1}{2}}} \xrightarrow{d} \sqrt{\frac{\zeta}{1-\zeta}}\nu_{w2}'\left[\Psi_{Zu1}+\sqrt{\zeta}\left(1_{[\delta_x=1]}\lambda_X d_\beta+1_{[\delta_w=1]}\lambda_W d_\gamma\right)\right]=\nu_{w2}'\eta^*(\beta_0,\gamma_0).$$

Then using the same strategy as in the proof of Theorem 3.1, it follows that $LM_\beta^*(\beta_0,\gamma_0) \xrightarrow{d} \frac{1}{\sigma_{uu}}\eta^{*'}(\beta_0,\gamma_0)P_{M_{\nu_{w2}}\nu_{x2}}\eta^*(\beta_0,\gamma_0)$ since $\frac{1}{n_1}(y_1-X_1\beta_0 0W_1\gamma_0)'M_{Z_1}(y_1-X_1\beta_0-W_1\gamma_0) \xrightarrow{P} \sigma_{uu}$. $\blacksquare$

Proof of Theorem 3.3 Part 1: Using Lemma A.1 as before it follows that:

$$\frac{1}{n_1}(y_1-X_1\beta-W_1\gamma_0)'M_{Z_1}(y_1-X_1\beta-W_1\gamma_0) \xrightarrow{P} \sigma_{uu}+2\sigma_{uW}(\gamma-\gamma_0)+(\gamma-\gamma_0)'\sigma_{WW}(\gamma-\gamma_0),$$

and $\frac{\widehat{W}_{12}'(y_1-X_1\beta_0-W_1\gamma_0)}{n^{\delta_w-\frac{1}{2}}}-\sqrt{\frac{\zeta}{1-\zeta}}\nu_{w2}'\left[\Psi_{Zu1}+\Psi_{ZW1}(\gamma-\gamma_0)+\sqrt{\zeta}\xi(\beta_0,\gamma_0)\right]=o_p(1).$

Therefore,

$$LM_\gamma(\beta_0,\gamma_0) - \frac{[\Psi_{Zu1} + \Psi_{ZW1}(\gamma-\gamma_0) + \sqrt{\zeta}\xi(\beta_0,\gamma_0)]' P_{\nu_{w2}} [\Psi_{Zu1} + \Psi_{ZW1}(\gamma-\gamma_0) + \sqrt{\zeta}\xi(\beta_0,\gamma_0)]}{\sigma_{uu} + 2\sigma_{uW}(\gamma-\gamma_0) + (\gamma-\gamma_0)'\sigma_{WW}(\gamma-\gamma_0)} = o_p(1) \tag{A.1}$$

Part 2: From (A.1) it is easy to see that at the true value of β and γ,

$$LM_\gamma(\beta_0,\gamma_0) \xrightarrow{d} \chi^2_{m_w}\left(\frac{\zeta\xi'(\beta_0,\gamma_0)' P_{\nu_{w2}} \xi(\beta_0,\gamma_0)}{\sigma_{uu} + 2\sigma_{uW}(\gamma-\gamma_0) + (\gamma-\gamma_0)'\sigma_{WW}(\gamma-\gamma_0)}\right)$$

where the non-centrality parameter is finite only for γ_0 such that $\sqrt{n}(\gamma-\gamma_0) = O(1)$. Hence $C_\gamma(1-\epsilon,\beta_0)$ can contain with positive probability only values in the $\sqrt{n}$-neighborhood of the true γ and therefore, $\gamma_{inf}(\beta_0)$, the value where $\inf_{\gamma_0\in C_\gamma(1-\epsilon,\beta_0)} LM^*_\beta(\beta_0,\gamma_0)$ is attained, is also in the $\sqrt{n}$-neighborhood of γ; in particular let $\gamma_{inf}(\beta_0)$ be such that $\gamma = \gamma_{inf}(\beta_0) + \frac{d_\gamma}{\sqrt{n}}$ where $d_\gamma = O_p(1)$. Therefore directly applying Lemma 3.2, we get $\inf_{\gamma_0\in C_\gamma(1-\epsilon,\beta_0)} LM^*_\beta(\beta_0,\gamma_0) = LM^*_\beta(\beta_0,\gamma) + o_p(1)$. ∎

A.2 Rejection Rules for different 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. For the projection-type tests, we follow the algorithm of Dufour and Taamouti (2005).

(I) USSIV Test rejects the null hypothesis $H_0 : \beta = \beta_0$ at level α if $LM_\beta(\beta_0) > \chi^2_{m_x}(1-\alpha)$.

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

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

i.e. if there does not exist any value γ_0 such that $a_{ss}\gamma_0^2 - 2b_{ss}\gamma_0 + c_{ss} \leq 0$ where $a_{ss} = W_1' H_{ss} W_1$, $b_{ss} = W_1' H_{ss}(y_1 - X_1\beta_0)$, $c_{ss} = (y_1 - X_1\beta_0)_1' H_{ss}(y_1 - X_1\beta_0)$ and $H_{ss} = P_{Z_1\widehat{\Pi}_2} - \dfrac{1}{n_1}\chi^2_m(1-\alpha) M_{Z_1}$. Equivalently, reject $H_0 : \beta = \beta_0$ at level α if

$$\left\{b_{ss}^2 - a_{ss}c_{ss} < 0, a_{ss} > 0\right\} \cup \{a_{ss} = b_{ss} = 0, c_{ss} > 0\}$$

III) Score-type Test rejects $H_0 : \beta = \beta_0$ at level at most $\alpha(+\epsilon)$ if

$$\begin{aligned} & \{C_\gamma(1-\epsilon, \beta_0) = \varnothing\} \cup \left\{ \inf_{\gamma \in C_\gamma(1-\epsilon,\beta_0)} LM^*_\beta(\beta_0, \gamma) > \chi^2_m(1-\alpha) \right\} \\ \text{i.e. iff} \quad & \{C_\gamma(1-\epsilon, \beta_0) \cap D_\gamma(1-\alpha, \beta_0)\} = \varnothing \end{aligned}$$

where $C_\gamma(1-\epsilon, \beta_0) = \left\{\gamma_0 | a_1\gamma_0^2 - 2b_1\gamma_0 + c_1 \leq 0\right\}$ and $D_\gamma(1-\alpha, \beta_0) = \left\{\gamma_0 | a_2\gamma_0^2 - 2b_2\gamma_0 + c_2 \leq 0\right\}$, and $a_1 = W_1' H_1 W_1$, $b_1 = W_1' H_1 (y_1 - X_1\beta_0)$, $c_1 = (y_1 - X_1\beta_0)_1' H_1 (y_1 - X_1\beta_0)$, $H_1 = P_{Z_1 \widehat{\Pi}_{W2}} - \frac{1}{n_1}\chi^2_{m_w}(1-\epsilon) M_{Z_1}$, $a_2 = W_1' H_2 W_1$, $b_2 = W_1' H_2 (y_1 - X_1\beta_0)$, $c_2 = (y_1 - X_1\beta_0)_1' H_2 (y_1 - X_1\beta_0)$ and $H_2 = P_{\left[M_{Z_1 \widehat{\Pi}_{W2}} Z_1 \widehat{\Pi}_{X2}\right]} - \frac{1}{n_1}\chi^2_{m_x}(1-\alpha) M_{Z_1}$.

Defining $\Delta_i = b_i^2 - a_i c_i$ for $i = 1, 2$, Score-type test rejects $H_0 : \beta = \beta_0$ at level at most $\alpha + \epsilon$ if any one of the following eleven mutually exclusive conditions are satisfied:

1. $\{\Delta_i < 0, a_i > 0\} \cup \{a_i = b_i = 0, c_i > 0\}$ for $i = 1$ and/or $i = 2$, 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 $i, j = 1, 2$ 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 $i, j = 1, 2$ 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 $i, j = 1, 2$ 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 $i, j = 1, 2$ 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 $i, j = 1, 2$ 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.

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

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

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 $a_0 = W_1' A_0 W_1$, $b_0 = W_1' A_0(y_1 - X_1\beta_0)$, $c_0 = (y_1 - X_1\beta_0)_1' A_0(y_1 - X_1\beta_0)$ and $A_0 = P_{Z_1} - \frac{1}{n_1}\chi^2_k(1-\alpha) M_{Z_1}$. Equivalently, reject $H_0 : \beta = \beta_0$ at level α if

$$\left\{b_0^2 - a_0c_0 < 0, a_0 > 0\right\} \cup \{a_0 = b_0 = 0, c_0 > 0\}$$

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

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

where $\tilde{\gamma}(\beta_0)$ is the limited information maximum likelihood estimator of γ restricted by the null hypothesis, $\tilde{\Pi}(\beta, \gamma) = (Z_1' Z_1)^{-1} Z_1' \left[(X_1, W_1) - (y_1 - X_1\beta - W_1\gamma)\frac{[\tilde{\sigma}_{uX}(\beta,\gamma), \tilde{\sigma}_{uW}(\beta,\gamma)]}{\tilde{\sigma}_{uu}(\beta,\gamma)}\right]$, $\tilde{\sigma}_{uu}(\beta, \gamma) = \frac{1}{n_1}(y_1 - X_1\beta - W_1\gamma)' M_{Z_1}(y_1 - X_1\beta - W_1\gamma)$, $\tilde{\sigma}_{Xu}(\beta, \gamma) = \frac{1}{n_1} X_1' M_{Z_1}(y_1 - X_1\beta - W_1\gamma)$ and $\tilde{\sigma}_{Wu}(\beta, \gamma) = \frac{1}{n_1} W_1' M_{Z_1}(y_1 - X_1\beta - W_1\gamma)$.

B 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 1: $\mu_\beta = 1$ 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
	Score-type 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
	Score-type 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
	Score-type 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 3: $\mu_\beta = 10$ 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
	Score-type 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
	Score-type 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
	Score-type 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. The USSIV Score test, the projection from Split-sample Score test (SS) and the Score-type test combine observations from both sub-samples. The projection from AR test and the Partial-K test are based on sub-sample one containing $n_1 = [n\zeta]$ observations. [$k = 4$, $\rho_{XW} = 0$, $n = 100$].

[a] By Theorem 2, the level of the Score-type test is always lesser than $\alpha + \epsilon$ (= 2ACV 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 2: $\mu_\beta=1$ 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
	Score-type 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
	Score-type 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
	Score-type 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 4: $\mu_\beta=10$ 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
	Score-type 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
	Score-type 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
	Score-type 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. The USSIV Score test, the projection from Split-sample Score test (SS) and the Score-type test combine observations from both sub-samples. The projection from AR test and the Partial-K test are based on sub-sample one containing $n_1=[n\zeta]$ observations. $[k=4,\ \rho_{XW}=0,\ n=100]$.

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

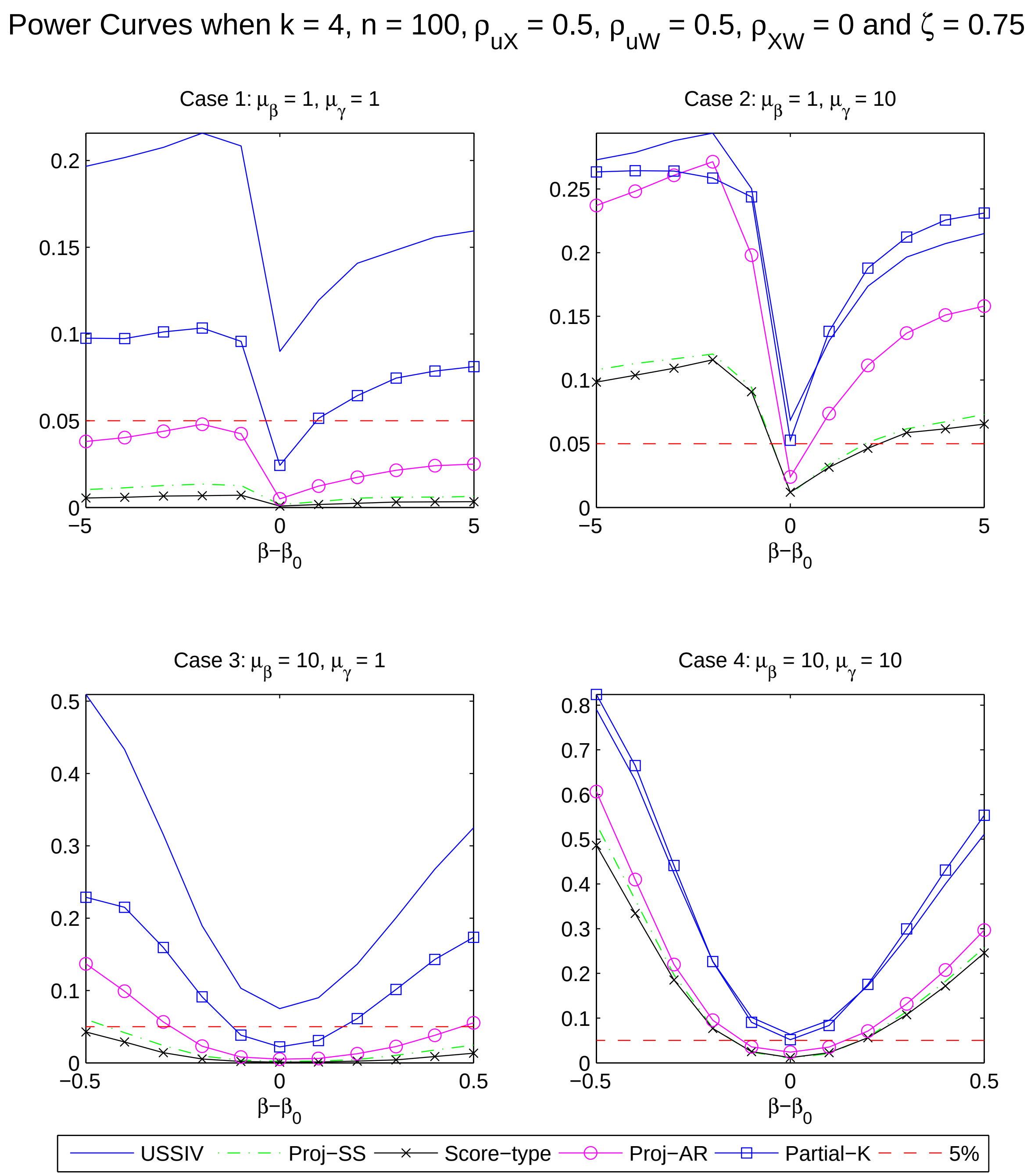

Figure 1: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

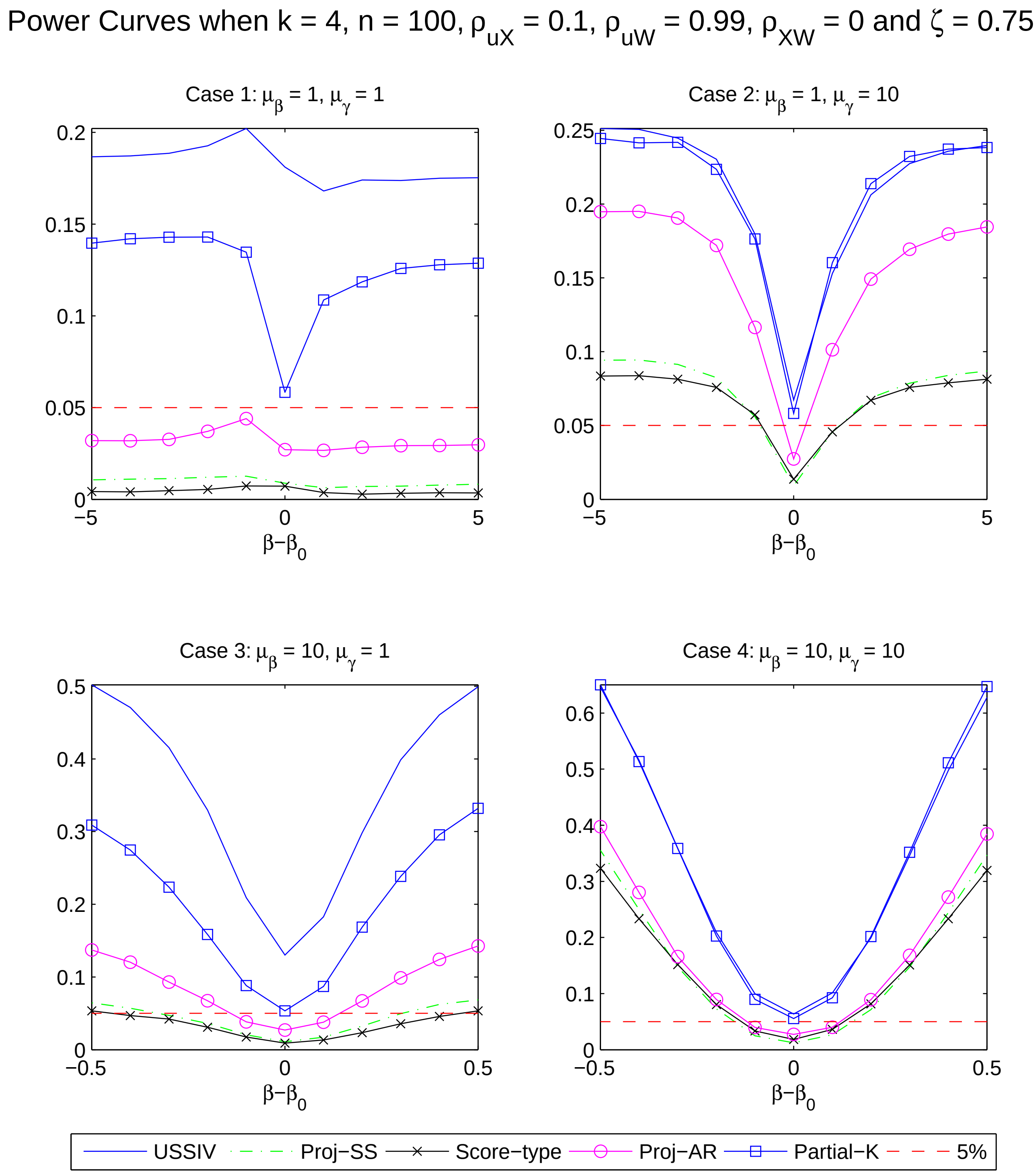

Figure 2: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

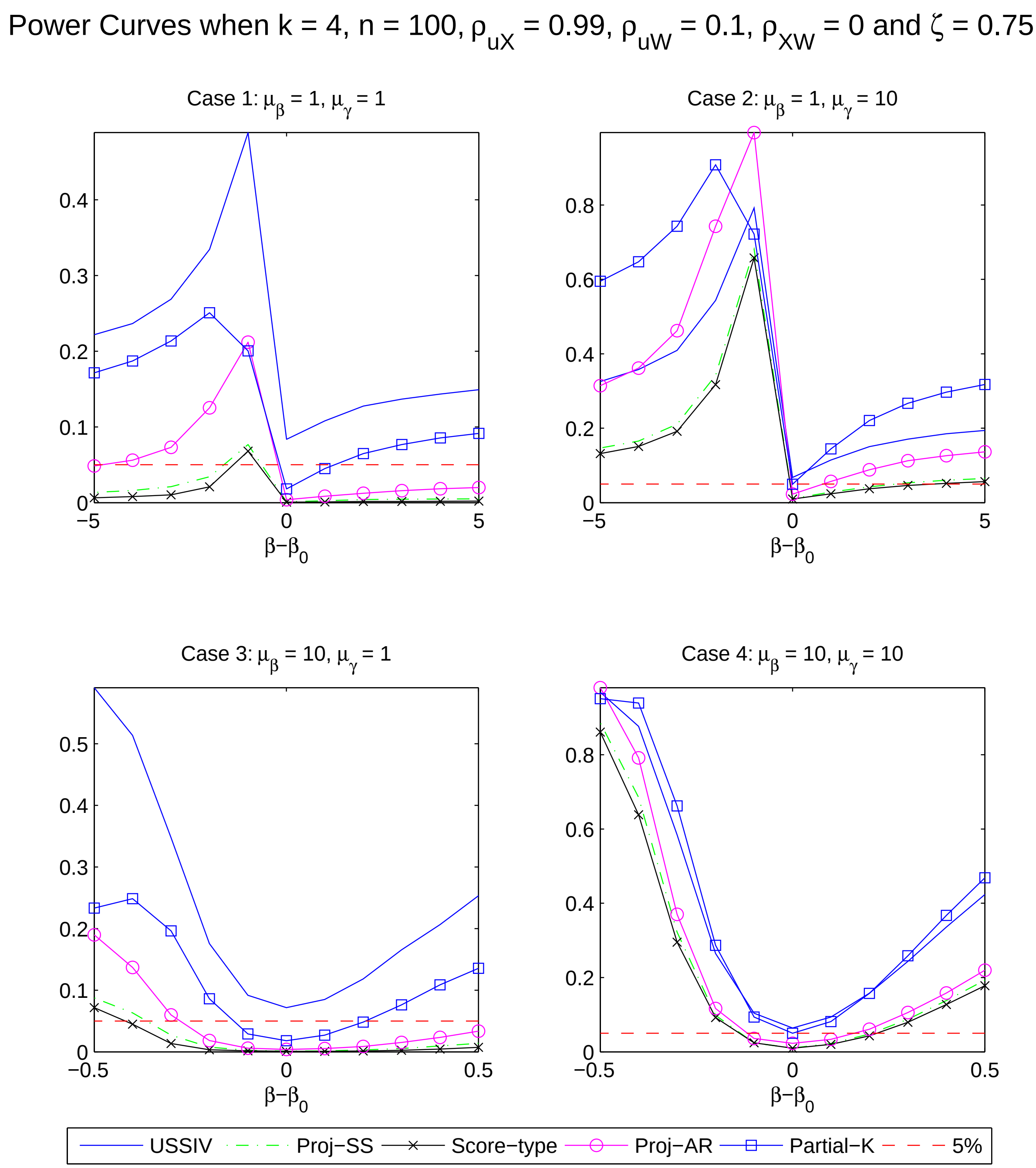

Figure 3: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

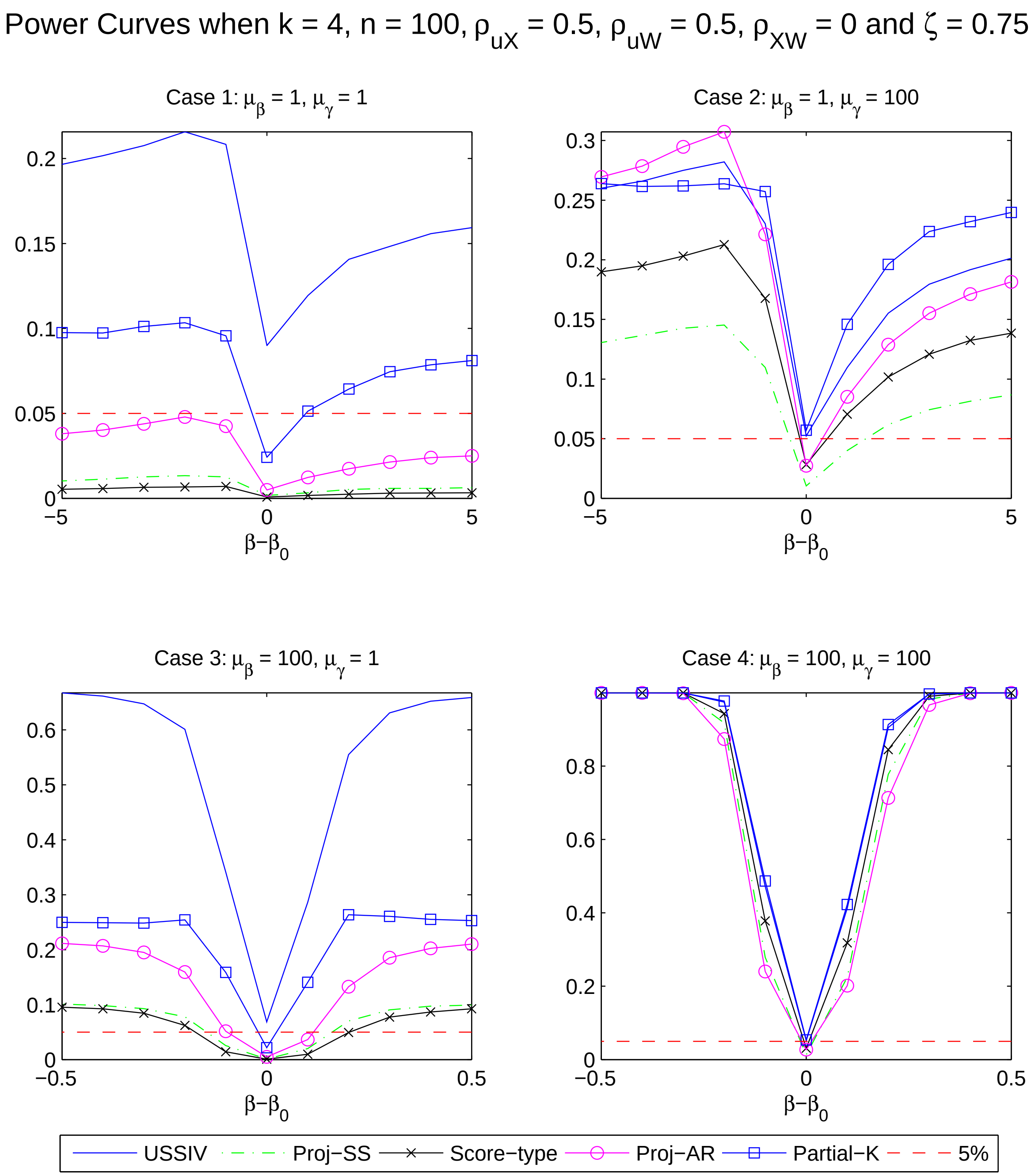

Figure 4: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

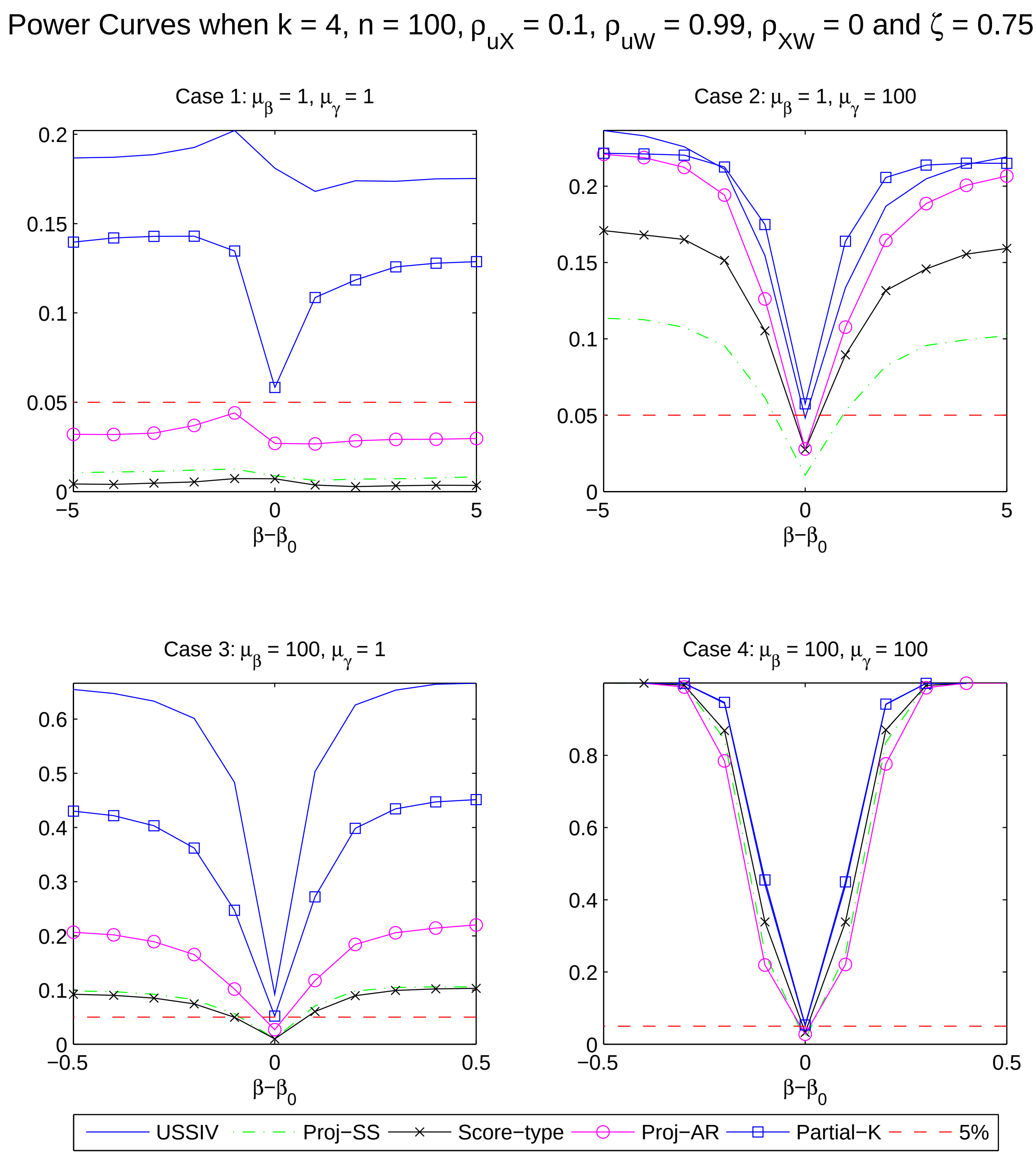

Figure 5: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

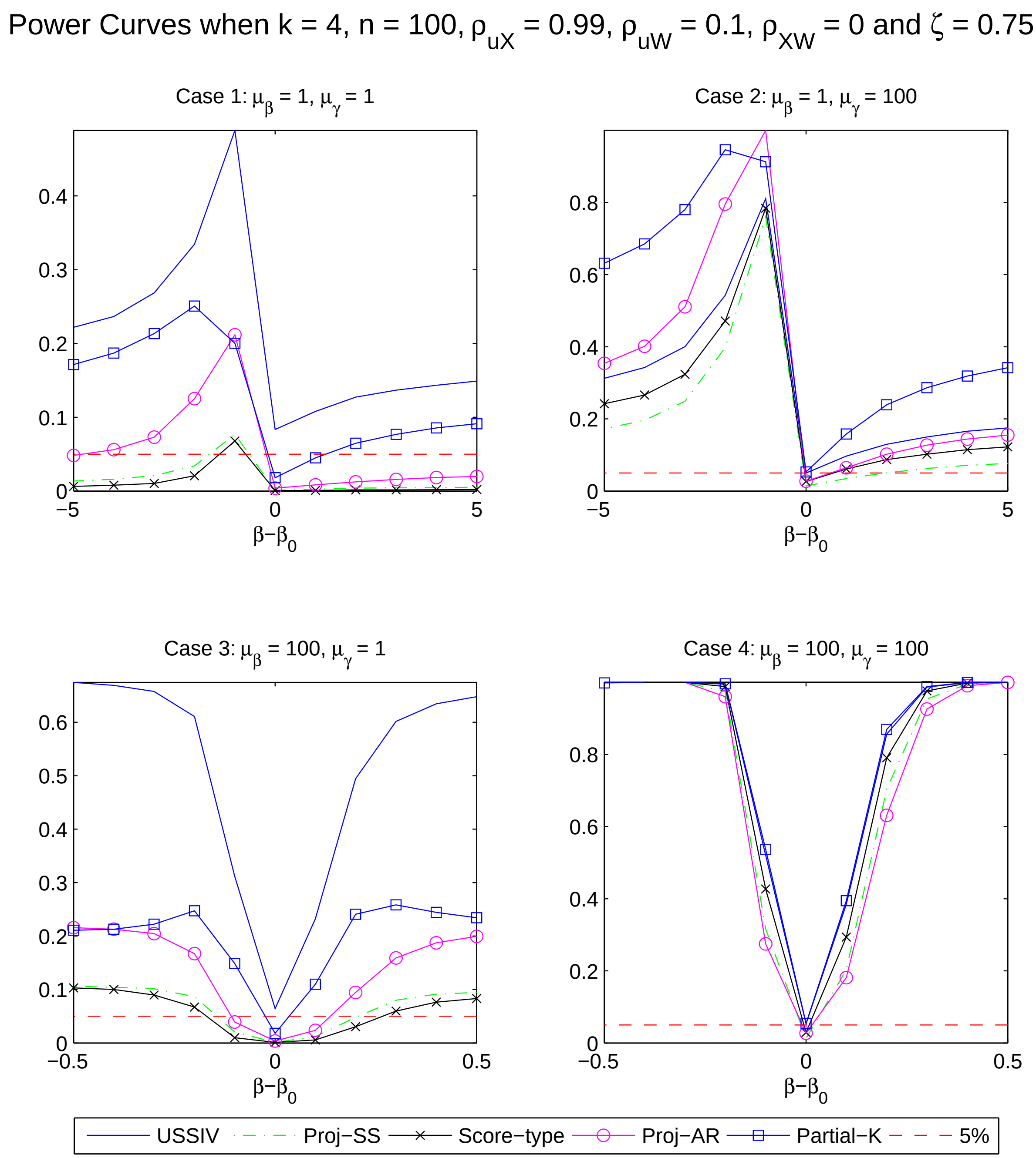

Figure 6: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

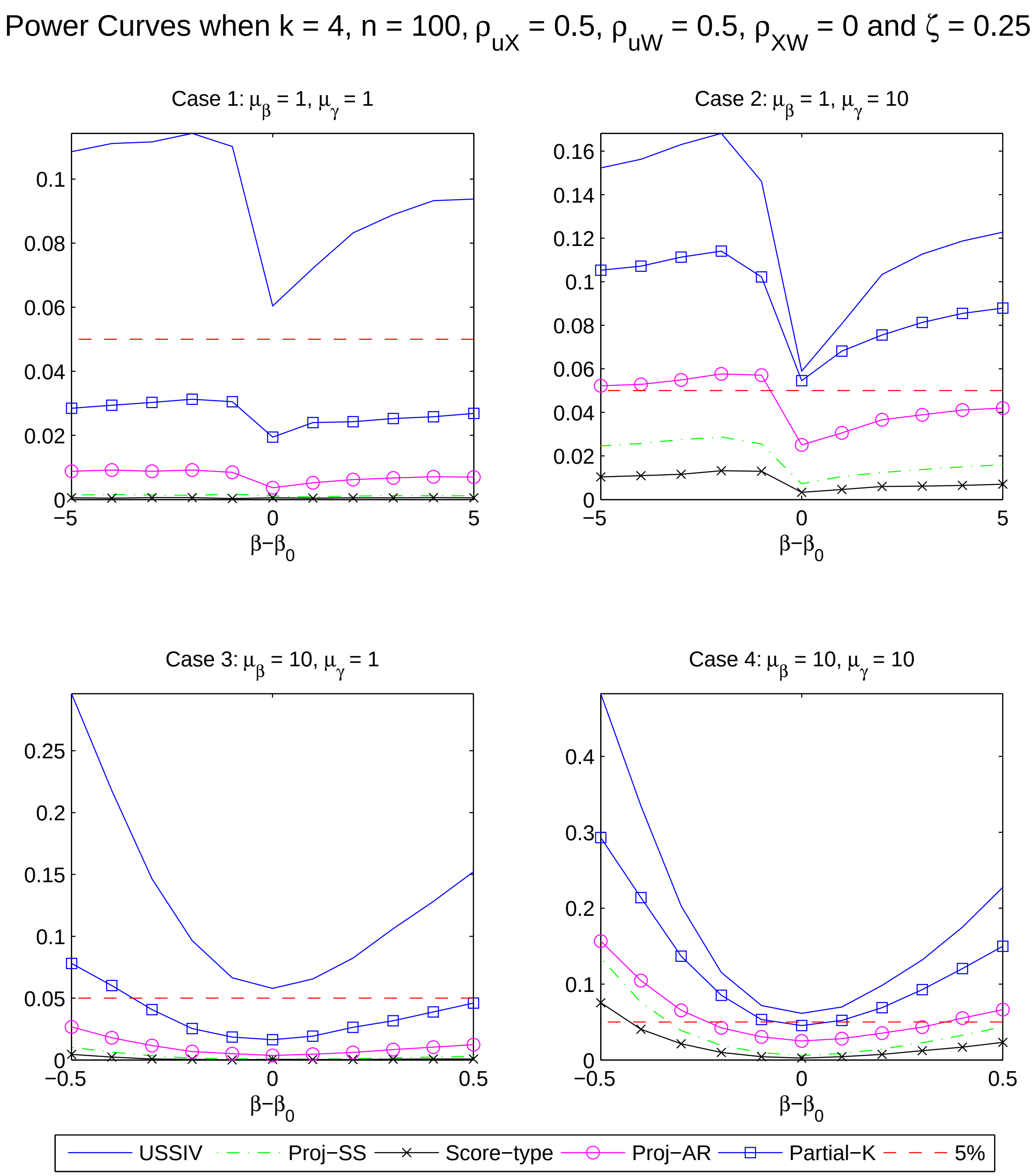

Figure 7: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

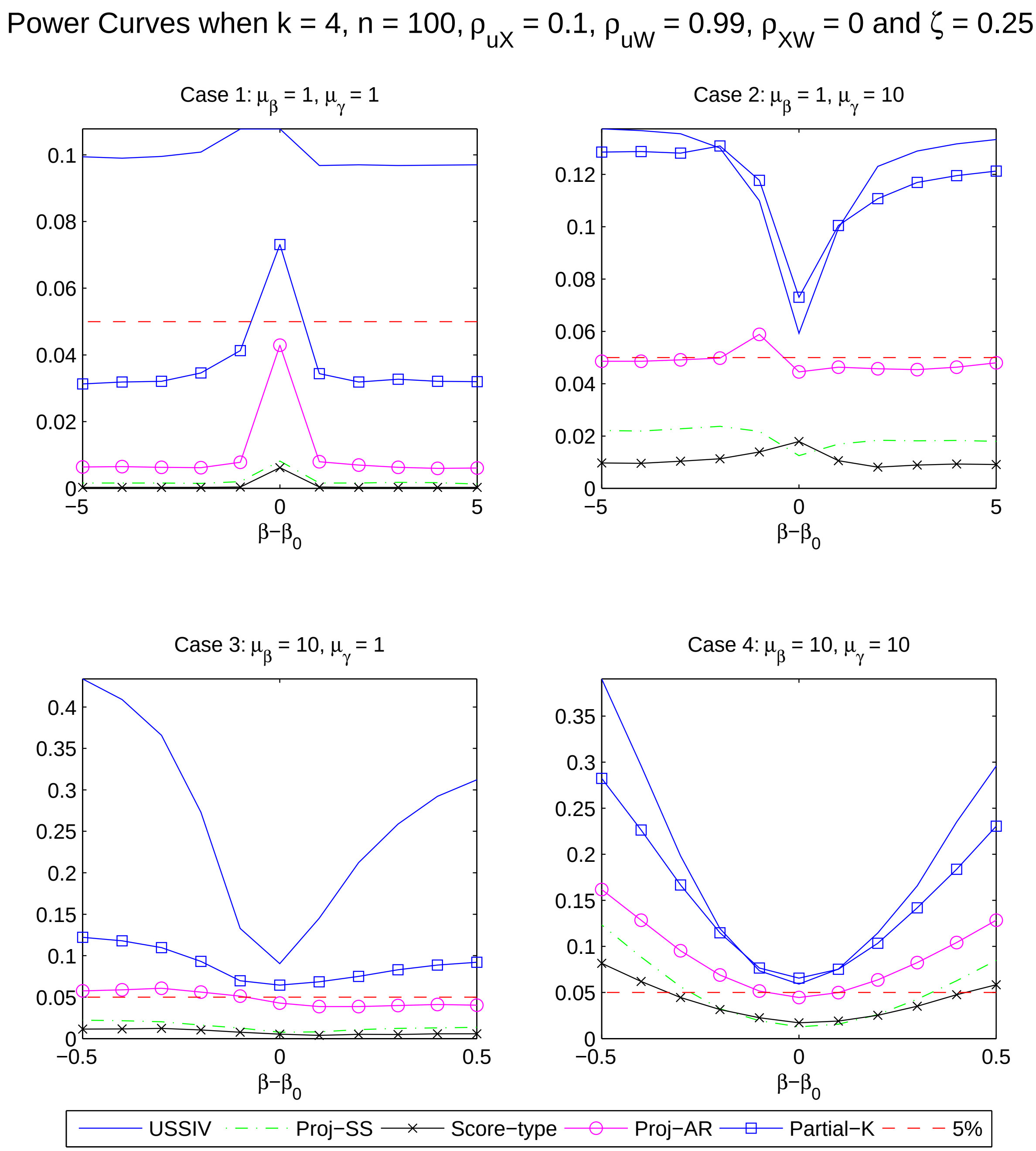

Figure 8: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

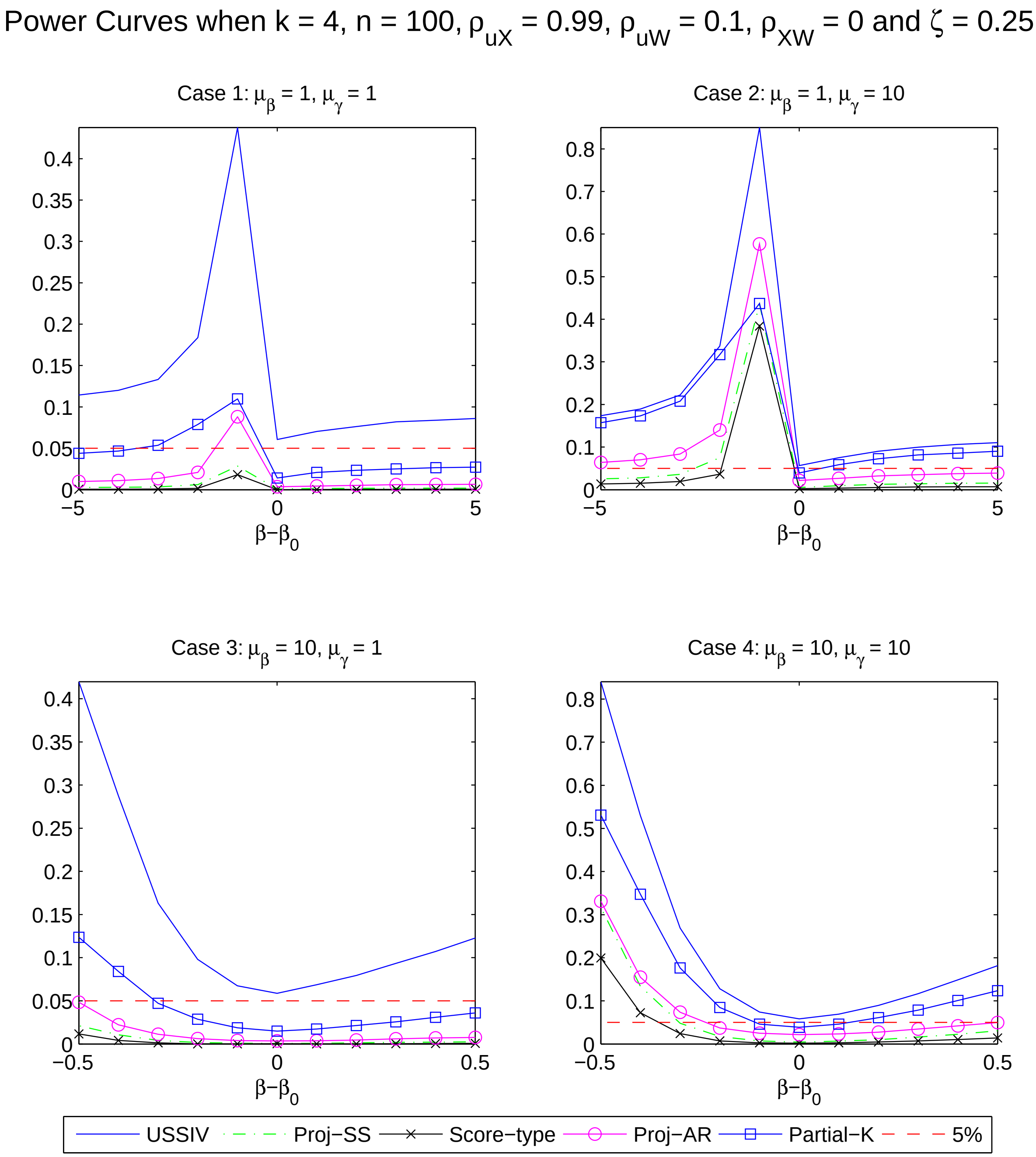

Figure 9: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

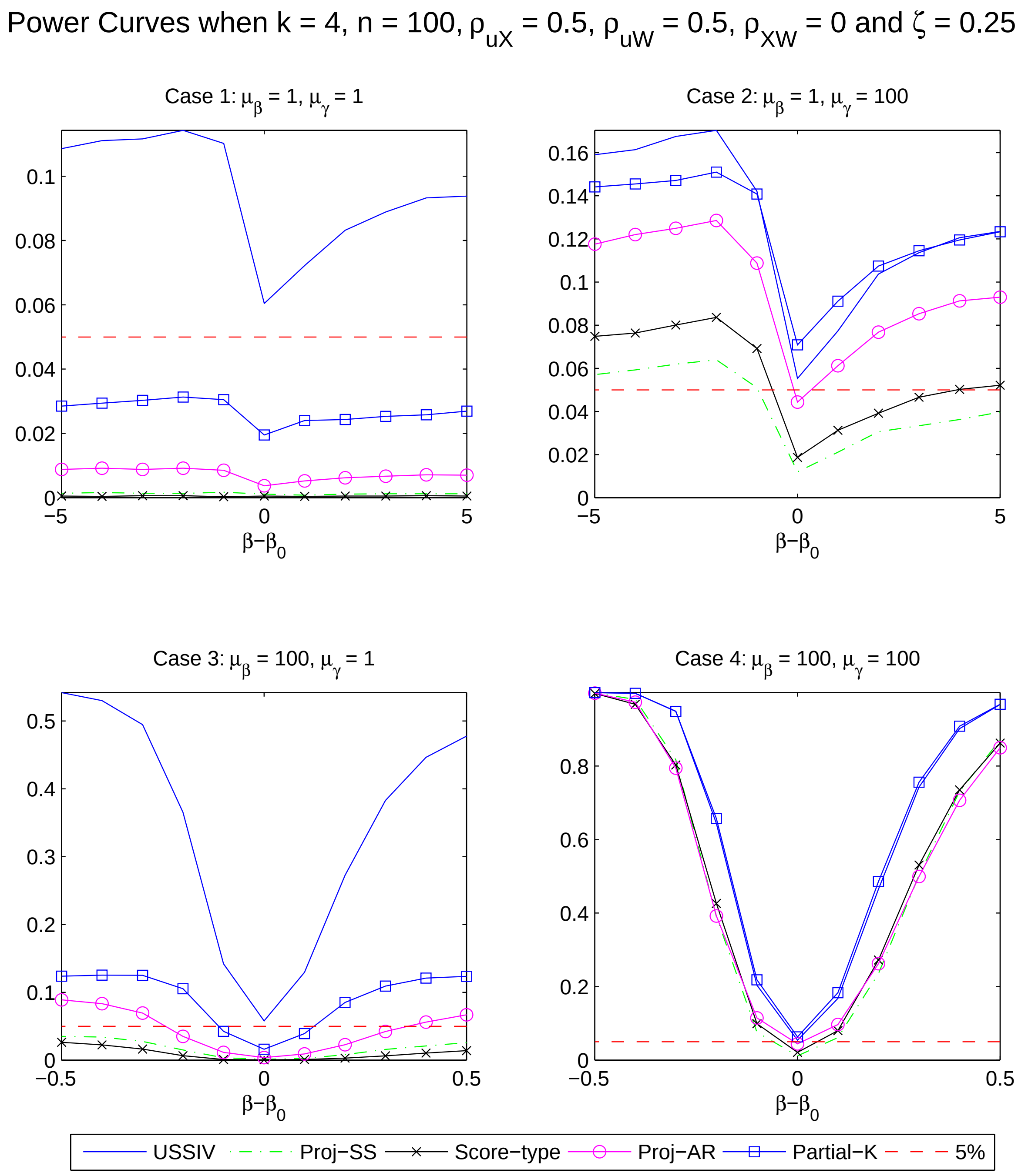

Figure 10: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

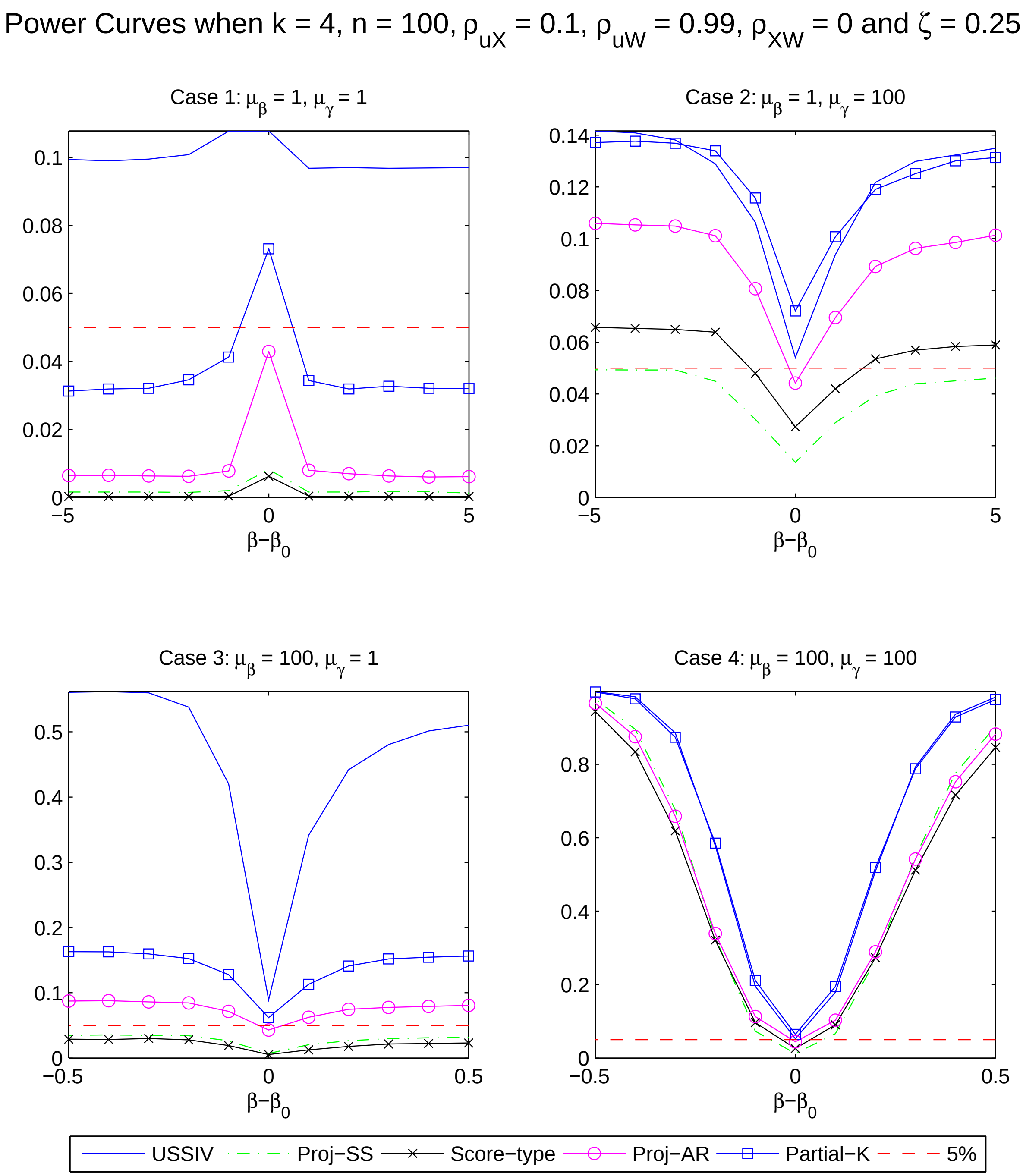

Figure 11: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

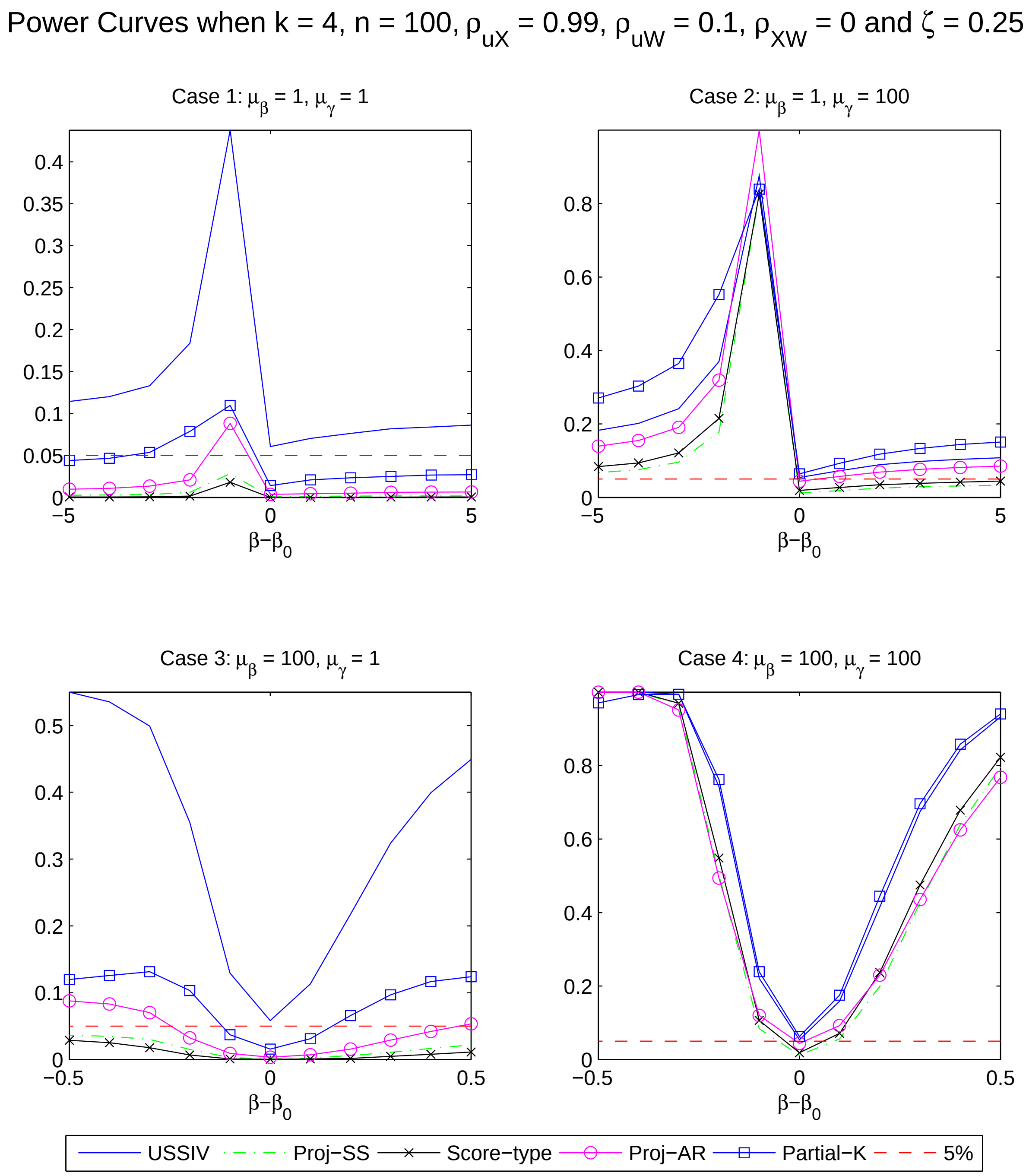

Figure 12: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

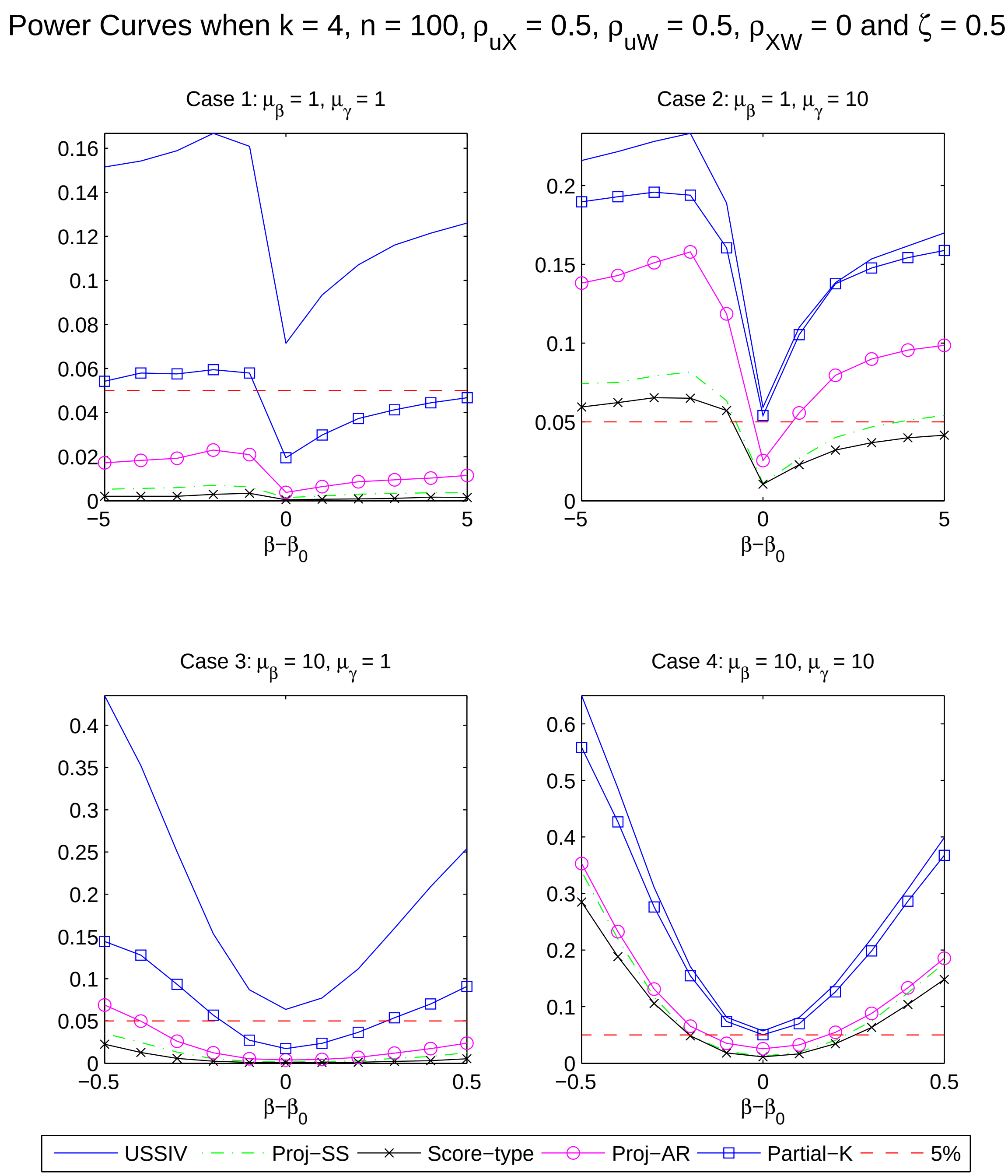

Figure 13: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

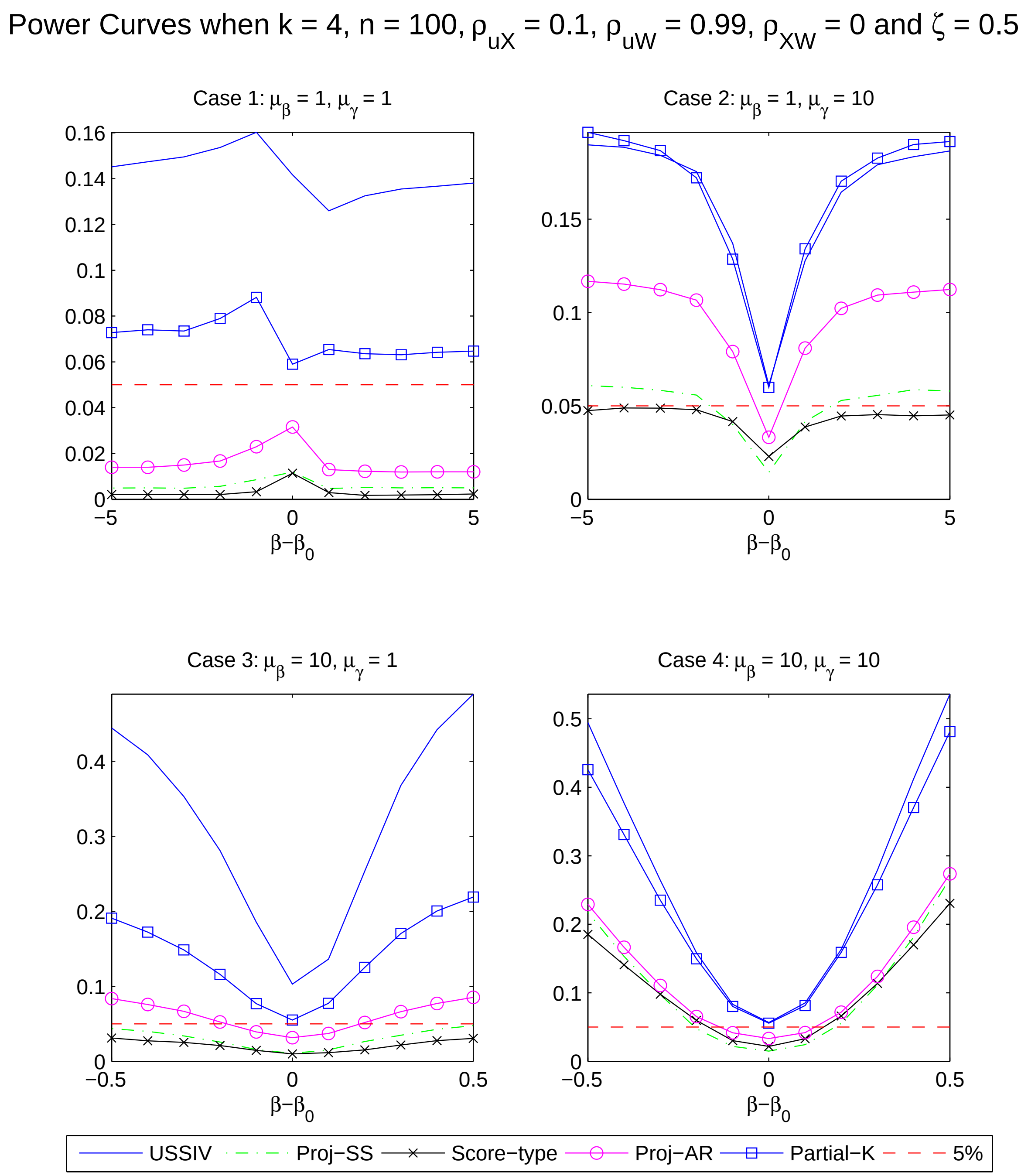

Figure 14: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

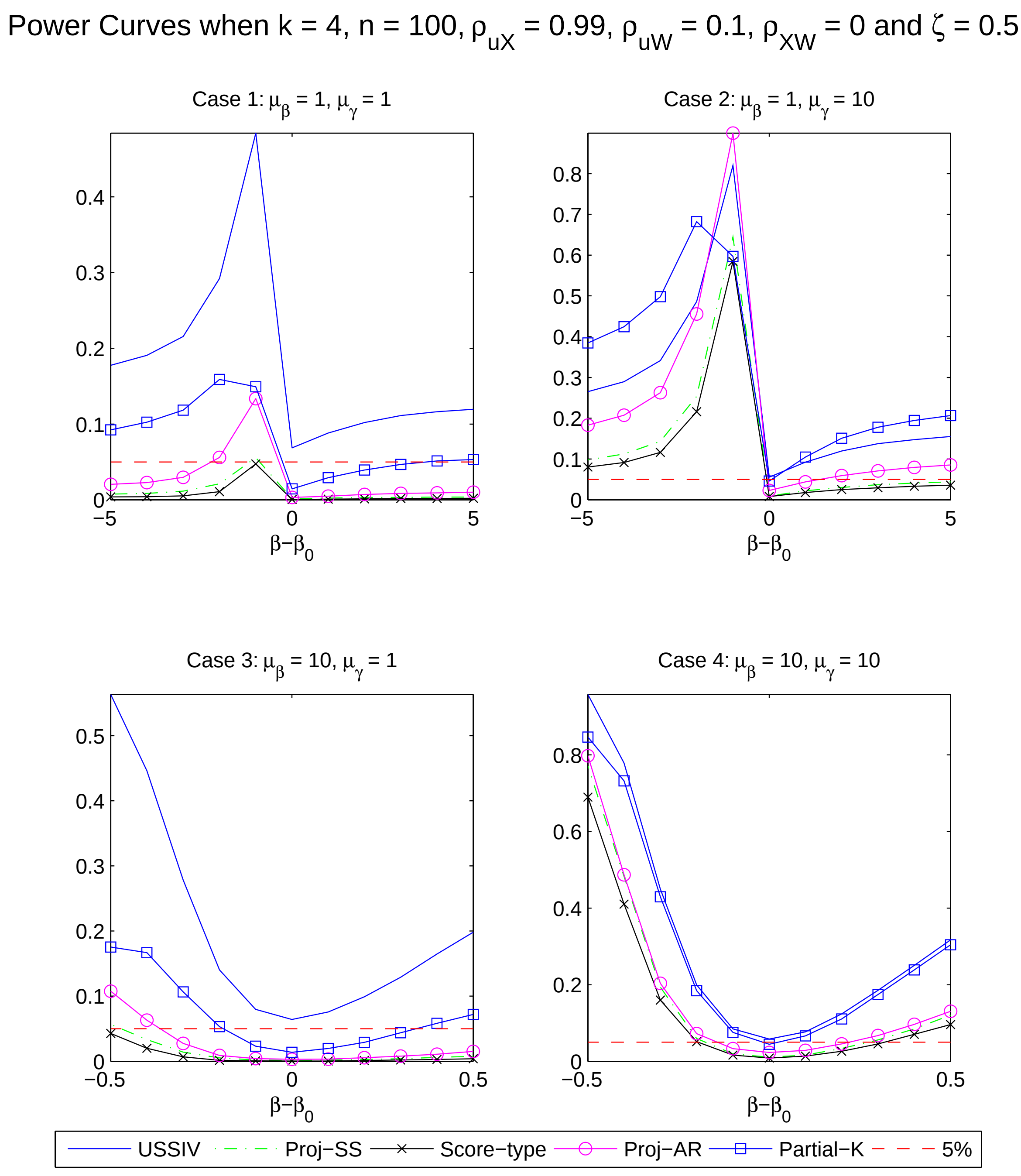

Figure 15: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 10$.

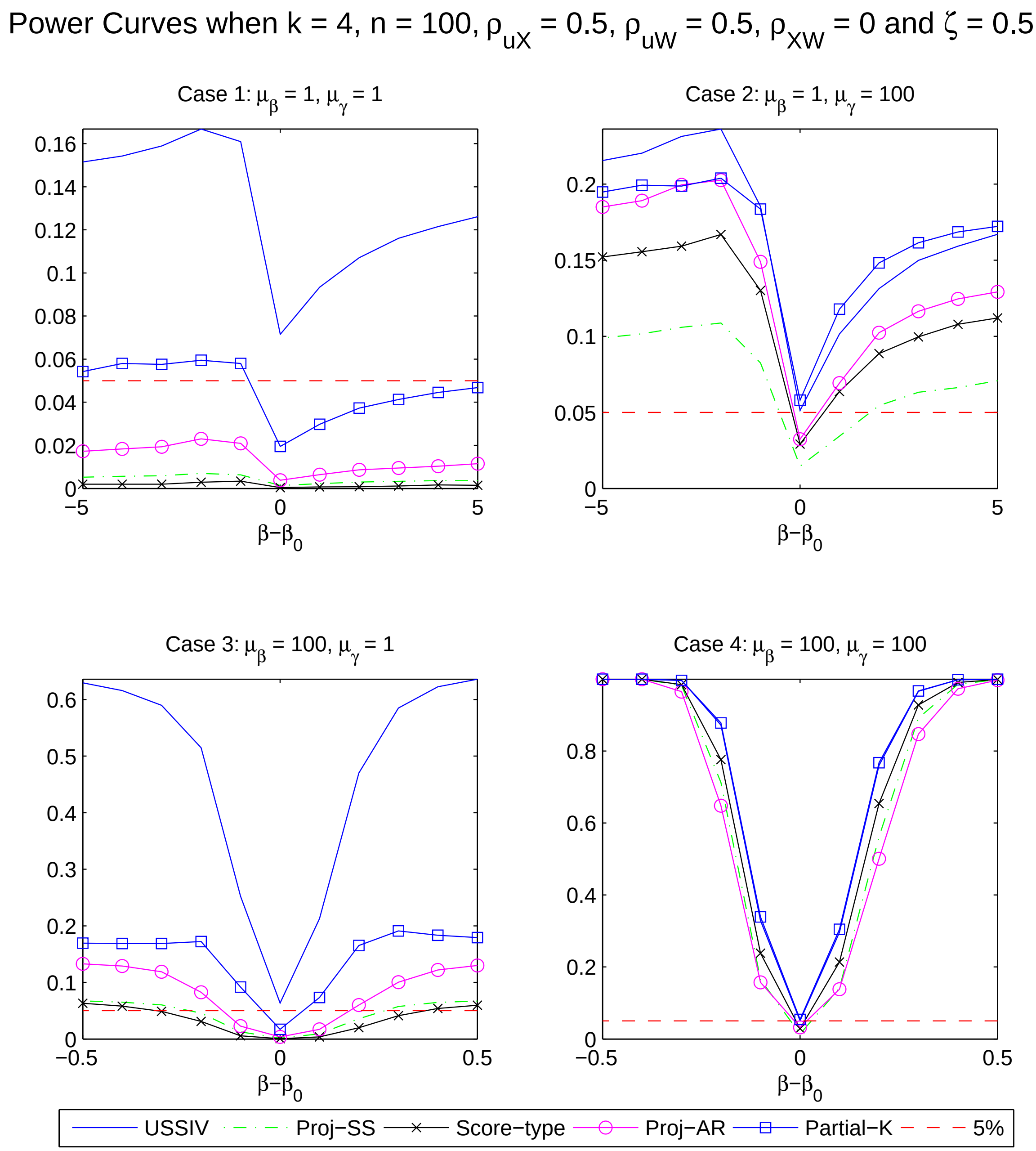

Figure 16: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

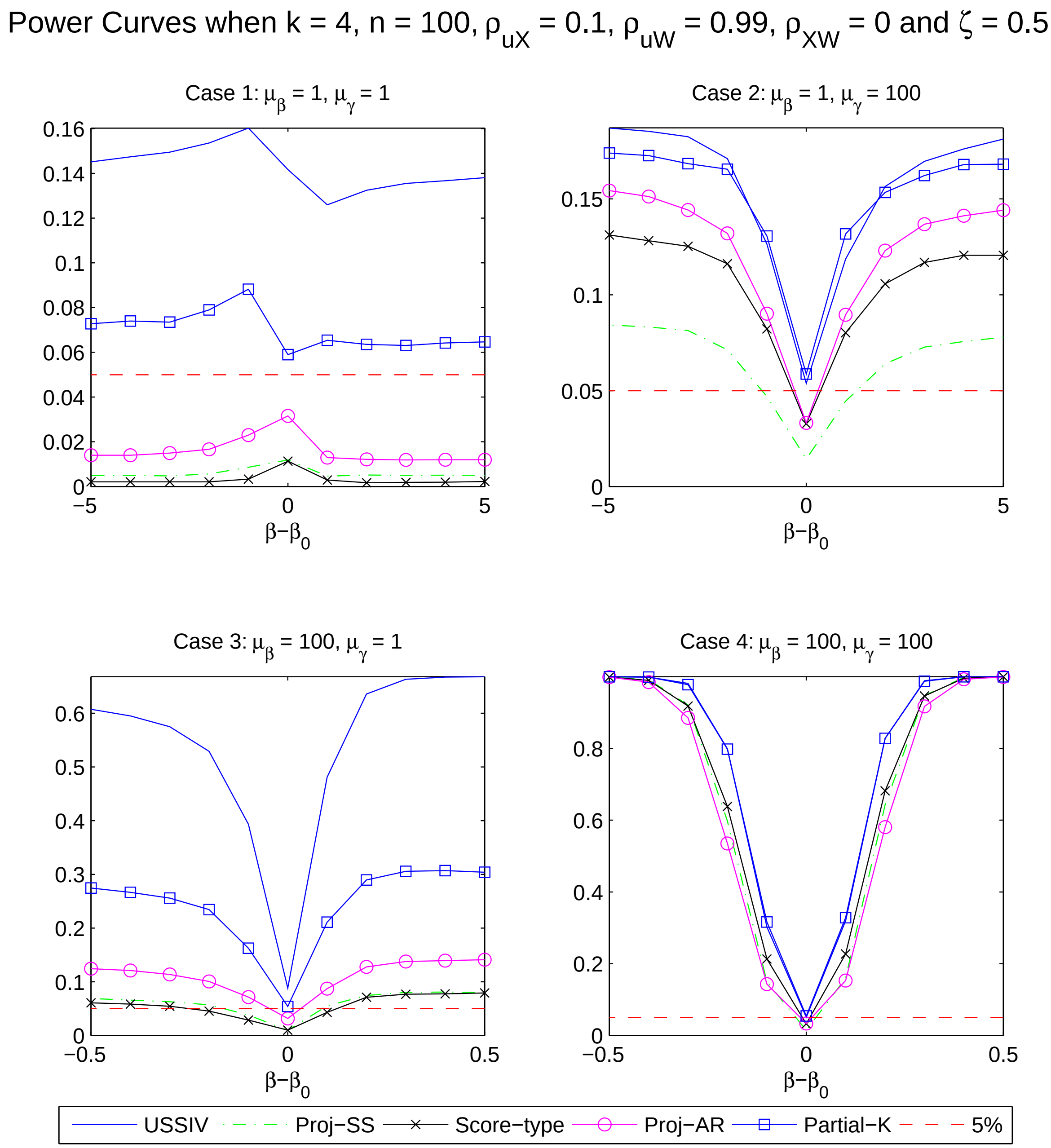

Figure 17: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

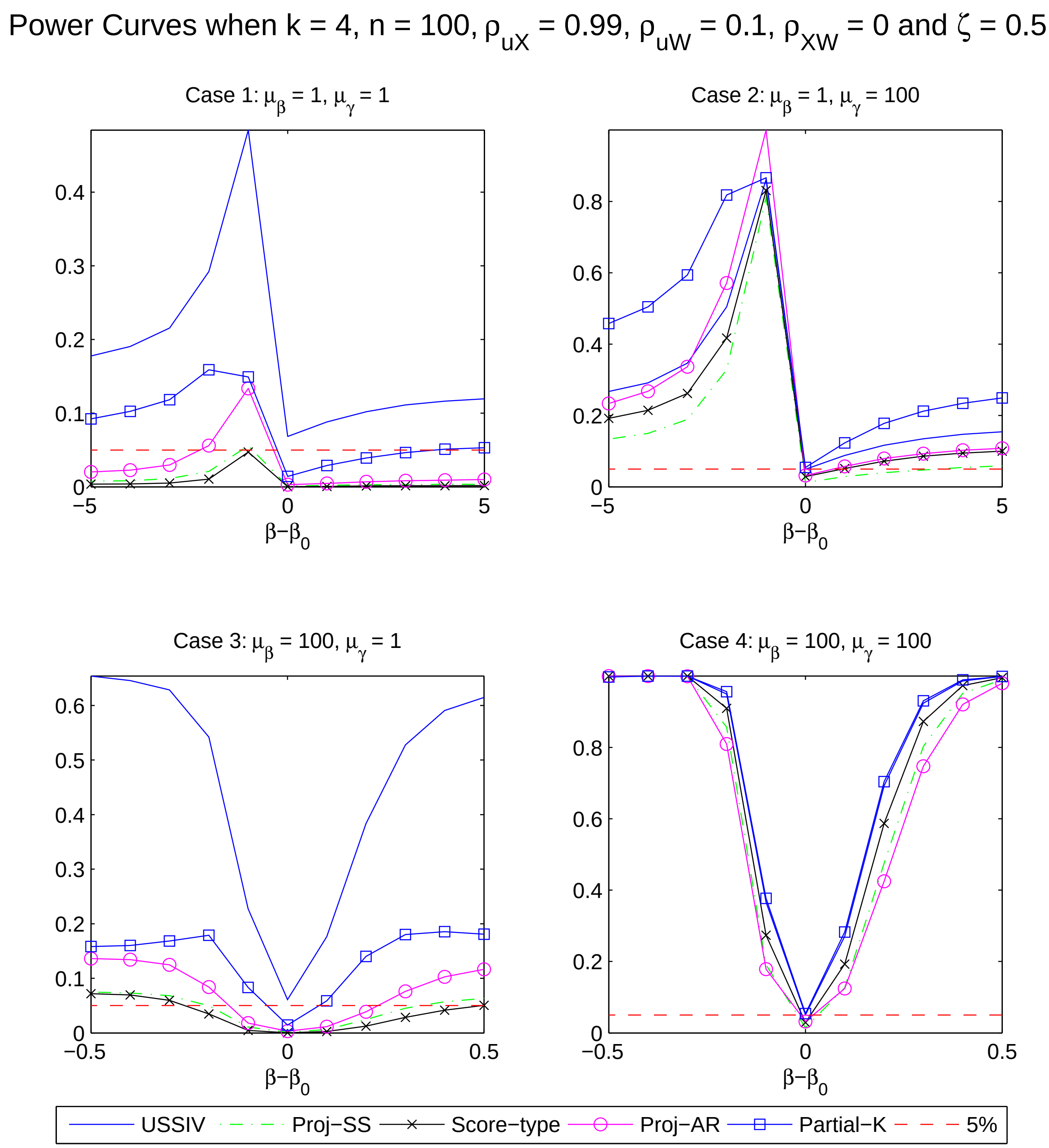

Figure 18: Rejection Rates for tests of $H_0 : \beta = \beta_0$ are computed based on 10,000 Monte-Carlo Trials. Weak instrument characterized by $\mu = 1$ and strong instrument by $\mu = 100$.

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). Spilt-Sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business and Economics Statistics*, 13: 225–235.

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.

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.

Kleibergen, F. (2005). Testing Parameters In GMM Without Assuming That They Are Identified. *Econometrica*, 73: 1103–1123.

Kleibergen, F. (2007). Subset statistic in the linear IV regression model. Technical report, Brown University.

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.

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.

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., 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.