ESSAYS ON PRICE DISCOVERY AND MODEL SELECTION IN PRESENCE OF WEAK INSTRUMENTS

MICHAEL NELSON AGUESSY

A THESIS
IN
THE DEPARTMENT
OF
ECONOMICS

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
CONCORDIA UNIVERSITY
MONTRÉAL, QUÉBEC, CANADA

NOVEMBER 2020

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Mr. Michael Nelson Aguessy**

Entitled: **Essays on Price Discovery and Model Selection in Presence of Weak Instruments**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Economics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

______________________ Chair
Dr. Martin French

______________________ External Examiner
Dr. Firmin Doko Tchatoka

______________________ External to Program
Dr. Sen Arusharka

______________________ Examiner
Dr. Han Xintong

______________________ Examiner
Dr. Bryan Campbell

______________________ Thesis Supervisor
Dr. Prosper Dovonon

Approved by ______________________
Dr. Christian Sigouin, Graduate Program Director

September 29, 2020 ______________________
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Essays on Price Discovery and Model Selection in Presence of Weak Instruments

Michael Nelson Aguessy, Ph.D.
Concordia University, 2020

This thesis organized in three chapters, essentially covers two main fields: finance and econometric theory with an application to macroeconomics. The first chapter proposes a methodology to uniquely measure price discovery, the mechanism by which the price of a security or an asset cross-listed on multiple markets is determined. The second chapter develops an information criterion that remain robust in presence of weaker instruments. Finally, the third chapter illustrates the benefits of optimal instruments' selection in assessing the impact of an example of monetary policy.

Being a process that allows market participants to uncover the real worth of an asset in a timely manner, the price discovery may lead to arbitrage opportunities. As such, the Information Share (IS) commonly used to measure it, needs to be as accurate as possible to help mitigate related market inefficiencies. In the first chapter of this thesis, we investigate the identification issues encountered by the IS due to its sensitivity to price ordering. This translates to price innovations vectors leading to a serious lack of robustness of the IS metric. Exploiting some statistical features of price innovations, we propose to use Independent Component Analysis (ICA) in order to decompose the residuals into independent signals. Compared to leading measures in the literature, our approach is shown to perform well in the standard two-market data framework. We also obtain consistent results while extending our simulations to larger number of markets framework, notably the three-market set-up. We finally confirm our findings by studying the mechanism of price discovery in two analogous empirical applications. The first analyzes futures and spot prices in the European Union Allowances (EUAs) market for CO_2 emissions, and the second concentrates on three Exchange Traded Funds (ETFs) tracking the performance of the Russell 2000 index. Our evidence suggests that futures prices and the IWM (ETF issued by iShares), respectively dominate their companions in contribution to price discovery.

The second chapter is motivated by the fact that the usual exogeneity assumption is essential to the least squares estimator as it guarantees its consistency. However, when this condition fails, the explanatory variable is said to be endogenous and Instrumental Variable (IV) regressions is one of the

methods available to the researcher to obtain consistent estimates. In response to the importance of the instruments selection step in the construction of a good IV estimator, we propose the alternative Relevant Moment Selection Criterion (aRMSC). This information criterion improves model selection when instruments are only weakly correlated with the endogenous variable. Through Monte Carlo simulations, we first illustrate that existing information criteria are not robust to these types of issues; naively selecting the larger models. We benefit from recent development on the importance of the strength of identification in achieving efficient estimation, and leverage it to evaluate how this may affect instruments selection when the candidate instruments available to the researcher are equally weakened or a pool of instruments with various strengths. Our evidence suggests that despite their weakness some instruments still contribute to improving the estimator's efficiency, in such a way that the selection of the most parsimonious model is possible.

In the final chapter of this thesis, we first illustrate the performance of our proposed information criterion in a macroeconomic application. Moreover, we study empirically the relationship between news from forward guidance and monetary policy. We account for interactions both between various macroeconomic variables while considering their own lagged values using a structural vector autoregressive (VAR) model including interest rates, consumer price index, industrial production and excess bond premium. Our analysis relies on Gertler and Karadi's (2015) high frequency identification (HFI) approach for monetary policy shocks to extend monetary policy indicators to the 2 year government bond rate even though authors initially considered it as facing weak instruments issues. The aRMSC allows us to identify relevant instruments in the VAR model with the 2 year government bond rate and compare our results to those predicted with the 1 year rate, a stronger instrument. We also consider the limited information maximum likelihood estimator (LIML) to improve the instruments' selection. All together, our results highlight that the model based on the optimal set of instruments in comparison to the model with the naive inclusion of all instruments from the candidate set, produces more accurate impulse responses for economic and financial variables regardless of the estimator used to obtain the alternative information criterion.

Acknowledgments

First, I would like to express my deep gratitude to my supervisor, Dr. Prosper Dovonon, for all the support, encouragement, and motivation. Without his guidance and assistance, this Ph.D. would not have been achievable.

I would like to offer my sincere thanks to the member of my thesis committee, Dr. Martin French, Dr. Firmin Doko Tchatoka, Dr. Sen Arusharka, Dr. Han Xintong and Dr. Bryan Campbell, for their invaluable and critical comments, which enabled me to address the weaknesses of the essays and aim to improve them.

My sincere recognition also goes to the former and current Graduate Program Directors, Dr. Effrosyni Diamantoudi, Dr. Szilvia Pápai, Dr. Damba Lkhagvasuren, and Dr. Christian Sigouin (current), and the former and current Chairs, Dr. Greg LeBlanc, and Dr. Jorgen Hansen (current).

I will always think of my dear friends and my Montreal family being present and supportive during this long journey. My special thanks go to Shadi, Nickesha, Liang, Ehsan, Kokouvi and Youssouf.

I wish to thank the staff members of the department of economics, Ms. Lucy Gilson, Ms. Elise Melançon, Ms. Lise Gosselin, Ms. Bonnie Janicki. Their help, patience, and diligence were very much appreciated.

Finally, above all, I would like to thank my beloved parents, Félicienne and Christophe who constantly encouraged and inspired me with their words and prayers, my beloved sisters, Tilda, Gloria, Tasha and Nelly for their unconditional support, and last but not least, my sweet wife, Cordie who blessed me with our two children, Enanmy and Hadriel, and her full dedication and love in every single step of this adventure.

Contents

List of Figures

List of Tables

Chapter 1

Measuring the Price Discovery through the Independent Component Analysis

1.1 Introduction

In Economics, the equilibrium price is defined as being the price level at which market demand equals market supply. As long as demand and supply are well approximated, it should be fairly trivial to derive an equilibrium price. When the good or service of concern is a financial market product, it is not straightforward to identify its demand or its supply. The mechanism of recovering the corresponding equilibrium price, or the real worth of the asset, is called the price discovery process. It relies on interactions between market participants in order to disclose the true asset valuation. In this framework, if the markets were able to function efficiently, the assets would trade at the right price, eliminating anomalies due to bubbles or crises. Instead, over the last four decades, there has been a rich literature that acknowledges financial market inefficiencies (see, e.g.,Banz (1981); De BONDT and Thaler (1985) and Fama (1998)).

Among others, two main measures of price discovery emerged from the studies of Hasbrouck (1995) and Gonzalo and Granger (1995). The first method is the so-called Information share (IS) due to Hasbrouck (1995). IS attempts to capture the variance of innovations in the long run price of an asset. It relies on a decomposition of those innovations. The second method, the Gonzalo-Granger permanent/transitory (PT) decomposition of cointegrated time series, exploits the fact that microstructural noise does not affect prices indefinitely. It expresses the common efficient price, also called common factor, in terms of a weighting vector distributed over contemporaneous prices.

Lehmann (2002) in his analysis of both methods obtained that they correctly measure price discovery when it takes place in one market. While Hasbrouck (2002) explains, after extensive simulations and comparison of the measures, that the problem with Gonzalo-Granger's method is that it is suboptimal at predicting relatively close future prices, it does not suffer the IS identification issues. Indeed, when the innovations to transaction prices are correlated, it is less obvious for the Hasbrouck IS to uniquely determine price discovery as it relies on a Cholesky decomposition.

As a result, the Hasbrouck IS is often presented in an interval which may diverge in response to the extent of the correlation. Baillie et al. (2002) suggest to average information shares to determine a mid-point of the lower and upper bounds to obtain a unique IS. Lien and Shrestha (2009) underline that this solution is arbitrary and propose the Modified Information Share (MIS) based on the square roots of the eigenvalues of the innovations' covariance matrix. Even though this approach has stronger foundations than a simple average and may resolve the issue, it also arbitrarily excludes negative roots of the obtained eigenvalues. Later, Jong and Schotman (2010) expand the Hasbrouck (1993) methodology to a multivariate setting in fragmented markets and obtain a new measure. Finally, Sultan and Zivot (2015) exploiting the homogeneity property of the innovations' standard deviations, construct a new alternative to the IS, the Price Discovery Share (PDS). As we will illustrate in our extensive simulations, this approach turns out to be insensitive to correlation matters but may report negative shares, making their interpretations very difficult.

The main goal of our study is to propose a methodology leading to a unique IS while ensuring that shares are positive and intuitive. In other words, contributions to price discovery are measured by proportions that sum-up to hundred percent. The closest to our research is the study of Grammig and Peter (2013) who obtain a unique IS when price changes exhibit tail dependence. Authors assume that the idiosyncratic innovations in market prices have a particular distribution. Indeed, they represent the innovations as a mixture of two serially uncorrelated Gaussian random vectors. The first part of the mixture is of unit variance and represents the tranquil "no news" regime. The second part, with diagonal covariance matrix represents the volatile "news" regime.

Although, many studies find attractive the use of finite normal mixture distributions to capture features of financial market prices, similar to Grammig and Peter (2013), they operate in a parametric framework and are restrictive in some sense. A non-parametric or data driven method may be of interest. Furthermore, the tail dependence is not obvious to observe in real data. Grammig and Peter (2013), revisiting (Uhrig-Homburg and Wagner, 2009) application to the European carbon emission prices report hardly interpretable results. They conclude that this is due to lack of tail dependence in the data. Our proposed methodology which utilizes a statistical technique used in blind source separation for the extraction of independent components from mixed signals, proves to offer very attractive properties in the quest of a unique IS. In addition to that, it offers less restrictive constraints regarding the distribution of price innovations.

Our next section will attempt, to thoroughly present the price structure documented in the literature when studying price discovery before introducing the reader to ICA. Then, the following natural step will be to explain how the ICA can improve the Hasbrouck IS. In the third section, this improvement will be illustrated simultaneously using Monte Carlo simulations on existing simulated market data and compared to IS and PDS measures. We recall that while the standard simulation framework used in the relevant literature, the two-market case (for the "Roll" model and the private and public information models) was introduced in Hasbrouck (2002), the ICA approach not only overcomes the correlation problem in the innovations but also offers the flexibility, at no cost, to extend the simulations to a higher number of markets. Therefore, we also present the simulations in a three-market set-up. This allows us to better illustrate the multiple prices context observed in practice and mostly to determine if the IS, the PDS and the ICA based methods consistently measure price discovery process.

Finally, in the remainder of the paper, we empirically illustrate our findings with two applications. The first analyzes (Uhrig-Homburg and Wagner, 2009) application to the European carbon emissions market. The Second evaluates the performance of our approach at measuring the individual contribution of three Exchange-Traded Funds to the price discovery mechanism in the Russell 2000 index. We conclude that the ICA information share correctly identifies the futures market (regardless of maturities) and the IWM (issued by IShares) as price discovery leaders in the respective applications.

1.2 Methodology

1.2.1 Hasbrouck IS and price discovery

Economic theory suggests that when a good is exchanged, the equilibrium price at which it is offered depends on competition between the sellers. Indeed, sellers want to attract as many buyers as they can, while maximizing their profits. In that framework, the equilibrium price will fluctuate around a common or minimum price at which each seller is willing to offer its good on the market.

When it comes to one good offered on many markets, the principle remains the same, even though the venues of exchange are different. This principle is a key characteristic of market microstructure models in which all markets prices exhibit a common component called the "efficient price" and any new information is impounded into the prices through specific market microstructure effects.

More rigorously, consider n market transaction prices which diverge by their specific microstructure effect. As illustrated by Lehmann (2002), at each time period t, prices can be expressed as follows

$$p_t = \ell_n m_t + s_t \tag{1.1}$$

where $p_t = (p_{1,t}, \cdots, p_{n,t})'$ is a vector of the market prices, ℓ_n is an n -dimensional vector of ones and the efficient price $m_t = m_{t-1} + \upsilon_t$ is a random walk as is standard in the literature and υ_t is the homoscedastic efficient price innovation and is uncorrelated with future s_t . Similarly to Hasbrouck (2002), the market microstructure effects $s_t = (s_{1,t}, \cdots, s_{n,t})'$ may represent bid-ask bounces, discreteness or inventory effects. The process s_t is a zero mean and covariance stationary process with a Wold representation

$$s_t = \nu_t + \sum_{j=1}^{\infty} \Upsilon_j \nu_{t-j} \text{ with } \sum_{j=0}^{\infty} \Upsilon_j < \infty \, ,$$

where $\nu_t = (\nu_{1,t}, \cdots, \nu_{n,t})'$, $E(\nu_t) = 0$, $E(\nu_t'\nu_s) = 0$ for all $s \neq t$, $E(\nu_t \nu_t') = \Sigma_\nu$ and Υ_0 is equal to an identity matrix of order n.

In this multivariate framework, the price dynamics can be represented as a Vector Autoregressive (VAR) of order p . Therefore, there exists a linear combination of prices which is stationary and admits a Vector Error Correction Model (VECM) representation of order $(p-1)$

$$\Delta p_t = \alpha \beta' p_{t-1} + \Gamma_1 \Delta p_{t-1} + \cdots + \Gamma_{p-1} \Delta p_{t-p+1} + u_t$$

in which α is an $n \times (n-1)$ loading matrix of same dimension as β the cointegration matrix as defined by Lütkepohl (2007). $\beta' p_t$ is stationary and each of its element represents a cointegrating relation, with $\beta' = [\ell_{n-1} \; -I_{n-1}]$ a matrix of linearly independent rows, as is often used in the literature.

Exploiting the stationarity of Δp_t , Lehmann (2002) obtained the following Wold representation:

$$\begin{aligned} \Delta p_t &= \Xi(L)u_t = \ell_n \upsilon_t + \Delta s_t \\ &= \ell_n \upsilon_t + (1-L)\Upsilon(L)\upsilon_t \end{aligned} \tag{1.2}$$

with $\Xi(L) = \sum_{j=0}^{\infty} \Xi_j^* L^j$ and $\Xi_0 = I_n$, the usual $n \times n$ identity matrix.

Now using the Beveridge-Nelson decomposition, p_t and Δp_t may be expressed as follows:

$$\begin{aligned} p_t &= p_0 + \Xi(1)\sum_{s=0}^{t} u_s + \Xi^*(L)u_t, \\ \Delta p_t &= \Xi(1)u_t + (1-L)\Xi^*(L)u_t \end{aligned} \tag{1.3}$$

where $\Xi(1) = \beta_\perp \left[\alpha'_\perp (I_n - \sum_{i=1}^{p-1} \Gamma_i)\beta_\perp\right]^{-1} \alpha'_\perp$, $\Xi^*(L)u_t = \sum_{j=0}^{\infty} \Xi_j^* u_{t-j}$ is an $I(0)$ process, p_0 contains initial values and $a_\perp$ is the orthogonal complement of a . In our multivariate framework, the n-dimensional vector $\Xi(1)$ has identical rows $\xi' = \sum_{j=0}^{\infty} \Xi_j^*$ and equation (1.2) yields $\Delta p_t = \Xi(1)u_t = \ell_n \upsilon_t$.

Therefore, we can express υ_t as following

$$\upsilon_t = \xi' u_t \tag{1.4}$$

where ξ' evaluates the permanent impact, on the efficient price, of a price innovation u_t .

This cointegration result in microstructure models is an introduction to market information shares. Indeed, in order to account for the correlation in price innovations from one market to another, Hasbrouck (1995) defined them as having a factor structure,

$$u_t = F\varepsilon_t \tag{1.5}$$

where ε_t is an $(n \times 1)$ vector of random variable with $E(\varepsilon_t) = 0$ and $Var\,(\varepsilon_t) = I_n$. F is the lower triangular factor of the Cholesky decomposition of $Var\,(u) = \Sigma_u$.

To compute market j 's information share, the contribution of market j to the variance in price innovations of an asset, Hasbrouck (1995) needed to estimate the variance of υ_t :

$$\begin{aligned} Var(v_t) &= Var(\xi' u_t) \qquad (1.6) \\ &= \xi' F F' \xi \qquad (1.7) \end{aligned}$$

Then, this quantity will be decomposed into n components, each one quantifying contribution of an individual market. as this methodology is closely related to the treatment of the prediction error variance decomposition of Hamilton, he decided to benefit from the triangularization of the covariance matrix. In this way, he would easily identify market contributions recursively. The resulting Hasbrouck information shares will depend on the Cholesky factor F and will be expressed as following,

$$IS = \frac{[\xi' F]^{(2)}}{\xi' F F' \xi} \qquad (1.8)$$

where $^{(2)}$ denotes element-wise squaring.

However, the Cholesky decomposition is very sensitive to column ordering and permutation. Indeed, the matrix F differs depending on the initial order given to market prices.

Consider for example, $n = 2$ markets, the price vector would be $p_t^1 = (p_{1,t}, p_{2,t})'$ and its permutation $p_t^2 = (p_{2,t}, p_{1,t})'$. Their respective innovations $u_t^1 = (u_{1,t}, u_{2,t})'$ and $u_t^2 = (u_{2,t}, u_{1,t})'$ will have the same off diagonal covariance elements, while the individual Cholesky decompositions of those covariances will be different. Actually, the Cholesky decomposition in the first case puts more weight on market 1's contribution to the price discovery while the second ranks the contribution of market 2 higher than market 1's.

In order to address this permutation problem, Hasbrouck (1995) suggested to report the Information Shares as an interval, specifying their lower and upper bounds. This may give rise to severe accuracy problems when the range of this interval is large. The section below will review existing measures of price discovery and compare them to the IS.

1.2.2 Existing Measures of Price Discovery

As mentioned earlier in the introduction, if we were to categorize the existing measures of price discovery, we would account for two main groups. The first category would account for papers modifying the IS or exploiting its statistical features to provide a unique IS. The second category would be formed by studies suggesting a different measure of the mechanism. We range our research in the first category as proposing a new measure of price discovery is beyond the scope of this paper.

In this category of interest, we enumerate below few of those variants of the IS and explain briefly how they exploit the original measure and compare them to our approach.

Variant of IS

Baillie et al. (2002) establish a clear difference between the IS and the Gonzalo-Granger PT models. They state that both methods use different definitions of price discovery. In particular the IS measure is closely related to Stock and Watson (1988) trends in economic time series. It assumes that the prices contain two parts, a permanent component (the common trend) and a transitory component. The Gonzalo-Granger PT method instead expresses the common trend as a combination of the prices. It measures perturbations in the prices as solely depending on a weighting coefficient and the common factor. Their price discovery measure directly exploits the changes in the prices through error correction model. Baillie et al. (2002) find the methods to converge to the same results when the VECM residuals are uncorrelated. They suggest that the midpoint of the lower and upper bounds would represent the IS. One issue with this approach is when the residuals are correlated. This in fact is what causes the original IS to diverge.

Lien and Shrestha (2009) exploiting the factor structure in equation (1.5), use matrix diagonalization to reformulate the covariance matrix of innovations. As there exist a well-known relationship between the variance and the correlation, this approach is closely related to the use of brute force to decompose the variance into a product of its positive square roots. In their framework, $\Sigma_u = \hat{F}\hat{F}'$ with $\hat{F} = (G\Lambda^{-1/2}G'V^{-1})^{-1}$ where Λ is a diagonal matrix containing the eigenvalues of the covariance matrix, G is the matrix formed by the respective eigenvectors and V is a diagonal matrix containing the innovations' standard deviations. We may notice that although this approach solves the order problem, it destroys the correlation structure, initially targeted by Hasbrouck imposing the representation in equation (1.5). Indeed the goal of this factor structure is to report the direction and the size of shocks from each specific markets to the efficient market price.

Grammig and Peter (2013) also rewrite the factor structure as $u_t = We_t$ where W is a non singular matrix and e_t is a vector of serially uncorrelated innovations. In which case e_t is represented by a mixture of Gaussian random variables as follows:

$$\begin{cases} e_{1,t} \sim N(0, I_n) & \text{with probability} \quad \gamma \\ e_{2,t} \sim N(0, \Psi) & \text{with probability} \quad 1-\gamma \end{cases}$$

where Ψ is a diagonal matrix with positive elements $\psi_1, \cdots, \psi_n$. $(1-\gamma)$ is the probability that we are in a volatile regime and γ when applicable. This approach as opposed to the previous one, preserves the initial factor structure but requires some constraints for the identification of the weight matrix W . Indeed, it imposes that the components of W are all positive and the elements on the diagonal are all larger than elements of the same row. This implies that u_t varies in the same direction as e_t and shocks initiated in market 1 through e_t have a stronger impact on u_t than shocks originating from market 2. Our approach will also keep the factor structure while suggesting a non-parametric approximation of the Cholesky factor.

Other Measures of Price Discovery

As we mentioned earlier, the most known alternative measure of price discovery is the Gonzalo-Granger PT method. More recently, Sultan and Zivot (2015) among others, apply Euler's theorem to decompose the volatility of the permanent shocks. Recall equation (1.4) may be rewritten $v_t = \xi' u_t = \sum_{k=1}^{n} \xi_k u_{k,t}$ and the standard deviation of the permanent shock $\sigma_v = (\xi' \Sigma_u \xi)^{1/2}$ is an homogeneous function linear in ξ' . Therefore by Euler's theorem σ_v is expressed as a weighted sum of marginal contributions from each market:

$$\sigma_v = \sum_{i=1}^{n} \frac{\partial \sigma_v}{\partial \sigma_i} = \xi_1 \frac{\partial \sigma_v}{\partial \sigma_1} + \cdots + \xi_n \frac{\partial \sigma_v}{\partial \sigma_n}.$$

Sultan and Zivot (2015) define the ratio of the i - th term of this sum and the innovations' volatility as the Price Discovery Share of market i. As this approach is fully symmetric, the authors underline its invariance to price ordering. However, they also raise the fact that it may report negative values, which compares to the natural risk reducer in risk management. Although, in their opinion, those negative shares may offer an additional understanding of information transmission mechanism in financial markets, they remain counter intuitive. The next section will introduce the ICA approach and explain how it leads to a unique positive information share.

1.2.3 Independent Component Analysis (ICA)

General Framework

The Independent Component Analysis is a statistical method introduced by Jutten and Herault (1991) and Comon (1994). It has been used in many fields as a tool that can separate linearly mixed independent components. Popular in the signal processing literature, it can decompose a vector of mixture signals X in terms of vectors of its independent components S , and a mixing matrix A

$$X(t) = AS(t) \tag{1.9}$$

where $S(t)$ is unobservable.

As an illustration, consider the well known "cocktail-party problem" with two speakers recorded simultaneously in the same room by two different microphones. The signals recorded at each time period t respectively named $x_1(t)$ and $x_2(t)$, appear to be mixtures of both speeches (speakers voices) $s_1(t)$ and $s_2(t)$ such that

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \end{aligned} \tag{1.10}$$

where the coefficient a_{ij} for $i, j = 1, 2$ can be approximated by the distance between each microphone as suggested by Hyvärinen and Oja (2000). The uniqueness of the mixing matrix $\left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$ in this approach relies on the vectors s_1 and s_2 being drawn from non-Gaussian distributions.

In particular, it has been often raised in the blind source separation (BSS) literature that in the case of normality of both independent components, the mixing matrix is not identifiable. Indeed, the joint density of both independent components is perfectly symmetric and we can only estimate the ICA model up to an orthogonal transformation. Consequently, it is sufficient that one of the independent components has a non-Gaussian distribution, to find $s_1(t)$ and $s_2(t)$ statistically independent and consistently estimate the mixing matrix.

It is important to highlight this fact since the ICA method, in determining the source signals, assumes that they are independent. Therefore, this approach becomes relevant when it is not too restrictive, similarly to the cocktail party example, to assume that the microphones record the voices independently of each other.

Some measures to detect non-Gaussian distribution, are the Kurtosis or fourth order cumulant and the negentropy. As the Kurtosis has been found to be very sensitive to outliers (Huber (1985)), the negentropy is a more appealing measure but difficult to estimate. Later, Hyvärinen et al. (2004) proposed to approximate it using a non-quadratic function. They also introduced the FastICA algorithm, which is an iterative fixed point method maximizing the approximation (see Appendix 1.6.1 for more details).

Permutation Ambiguity

Factor methods, often exhibit some identification problems. The ICA does not bypass this inconvenience as it suffers some ordering problems. Consider, the signals obtained from separating the mixed signals in equation (1.10). x_1 and x_2 may not be individually associated with each source signal s_1 and s_2 as expected. All we know from this system is that each recorded signal is obtained from a combination of the independent components. This issue has been of minor concern to neural network literature, because the order of components does not affect the results.

The testing-and-acceptance approach proposed by Cheung and Xu (2001) using the Relative Hamming Distance or the Mean Squared Error criterion are the ordering methods frequently employed in ICA empirical studies to reconstruct mixed signals. Even though, the Relative Hamming Distance has been used by Lu et al. (2009) in time series applications and offered good results, a typical criterion for evaluating dependency in data is the correlation of variances or correlation of squares, often referred to as correlation of energies in physics. Hyvärinen (2013) suggests that it highlights some underlying process that determines the level of activity of the components. As opposed to the standard measures of correlation, which are proportional to the coefficients a_{ij} for $i, j = 1, 2$, this criterion will measure the nonlinear correlation between the signals. Applying it in our two signals case, this criterion implies the computation of the coefficient of correlation between the squares of source signals and the squares of the respective ICs. When the coefficient between the source signal i and the IC $_i$ is close to one, this implies that the IC may be attributed to that source signal. Also, it suggests that if the ICA properly separates the information incorporated in the mixed signals, the correlation coefficient between the source signal i and the IC $_j$ will be closer to zero.

Alternatively, it may happen in practice that both components are strongly correlated with the same source signal. In that situation, we will consider the decomposition to be uninformative about the data features. In other words, the data are not rich enough for the ICA to, consistently, separate the mixed signals or the linear ICA may not be suitable to capture the complexity of the information carried by the original signals.

The next section will provide some intuition in the use of ICA in order to construct Information Shares.

1.2.4 A Unique Information Share Through Independent Component Analysis

In this section, we consider the same factor structure given earlier $u_t = F\varepsilon_t$ where ε_t and F are unknown. For simplicity, let $n = 2$ markets, this implies that the equation (1.4) will write,

$$\upsilon_t = \xi' \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \xi' \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are idiosyncratic shocks from markets 1 and 2 . This means that F_{ij}, the component of row i and column j of F , measures the impact and the strength of the shocks $\varepsilon_{i,t}$ originated from market i at time t . Intuitively, this factor structure is to model the composite innovations u_t as a linear combination of specific shocks propagating from markets i to j .

It clearly appears that if it is possible to measure ε_t accurately, the contribution of each market to the innovations u_t , and the direction and size of the contribution quantified by F , we would marginally contribute to the rich literature on price discovery mechanism.

As mentioned earlier, the ICA approach can extract independent components, as long as it is realistic to assume that their mixture signals are not gaussian and their separated source signals are independent. In our example, this implies nongaussianity of the composite innovations ($u_{1,t}$ and $u_{2,t}$) and statistical independence between the idiosyncratic innovations ($\varepsilon_{1,t}$ and $\varepsilon_{2,t}$) incorporated in the prices. Indeed, because they measure shocks specific to each market, the $\varepsilon_t s$ are considered independent. This is very plausible since this independence does not forbid $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ from being prompted by the same economic situation. In fact, this is similar to the economic intuition of Grammig and Peter (2013), who assume that for market i , F_{ii} will be larger than F_{ij} , as the impact of a shock should be stronger on its originating market. Therefore, F preserves the fact that a linear combination of 'local' effects from each market give rise to *global* $u_{1,t}$ and $u_{2,t}$.

Consequently under these conditions, we rewrite u_t as the decomposition of two independent components e_t through a mixing matrix A

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}$$

Under this specification, we eliminate the ranking problem caused by the Cholesky decomposition. The above equation leads to a new expression for the variance of υ_t :

$$Var(\upsilon_t) = \xi' \Sigma_u \xi = \xi' A \Sigma_e A' \xi,$$

where $\Sigma_e = I_n$ from the preprocessing for the ICA method, in which the original signals are centered and whitened, with zero means and unit variances (see Hyvärinen and Oja (2000) for more details)

By analogy with the equation (1.8) the information shares can now be calculated as following,

$$IS_{ICA} = \frac{[\xi' A]^{(2)}}{\xi' A A' \xi}$$

This procedure only requires the composite shocks u_t in equation (1.10) to be non-Gaussian and the choice of non linearity function $g(\cdot)$ in the fixed point algorithm (see Appendix 1.6.1) to be suitable to separate independent components embodied in u_t . The IS_{ICA} will then return the same values either when we reorder the series of prices or permute them. Using simulated market data in the next section, we attempt to evaluate the performance of the IS_{ICA} and compare them to some existing measures of price discovery.

1.3 Simulated Market Data

In this section, using Monte Carlo simulations, we simulate various market data models $1,000$ times. In each round, we compute the ICA Information Shares and compare our results to the Hasbrouck IS and the Price Discovery Share (PDS) of Sultan and Zivot (2015). We report for the Hasbrouck IS its Lower and Upper bounds (IS_L , IS_U), and Standard Deviations (Std. Dev) and the 95% Confidence Interval (CI) for all. The prices Vector Error Correction Model will be estimated at order $q = 20$ as this is the benchmark in the literature.

1.3.1 Two Markets Models

The "Roll" Model

In this section, we simulate the two-markets "Roll" model using a similar set-up as in Hasbrouck (2002), we consider transaction prices p_{it} for market $i = 1, 2$ such that the prices deviate from an efficient price m_t by the bid-ask bounce or trade direction and the cost of trading c . Each of them are defined as follows:

- the efficient price: $m_t = m_{t-1} + v_t$ where $v_t \sim N(0, \sigma_v^2)$
- the trade direction: $q_{it} = \begin{Bmatrix} -1 & \text{with probability} & 1/2 \\ +1 & \text{with probability} & 1/2 \end{Bmatrix}$ for $i = 1, 2$
- the transaction price: $p_{it} = m_t + cq_{it}$ for $i = 1, 2$, $c = 1$ and $\sigma_v = 1$ for $T = 100,000$ which is the number of observations.

The trading cost c is the same for traders on each market as each has the same information about the efficient price at the same time. For $q_{it} = -1$, the trader on market i sells the asset and $q_{it} = +1$ the trader buys it. In this cost framework and the probability distribution of the trade direction, the price discovery on each market is expected to be 50% .

Table 1 summarizes the average information shares obtained after $1,000$ simulations of the "Roll" model. The Hasbrouck original information shares report that the contribution of market 1 to price discovery are located between 21% and 79% . This is similar to the results obtained in the literature for this model. The PDS of Sultan and Zivot (2015) returns 50% either when the prices are kept in one order (p_1, p_2) or when the order is changed to (p_2, p_1) . This is due to the perfect symmetry allowed by the homogeneity property of the standard deviations in the PDS measure. The ICA also happens to separate the composite innovations properly and reports an information share very close to 50% for market 1.

The standard deviations in Table 1 , in the cases of PDS and ICA are larger than in the IS case. This comes from the fact that in different rounds of simulation, due to the random draws, some of the information shares diverge slightly from the expected 50% . Nevertheless, the 95% CI remains pretty sharp, especially in the ICA case.

Private Information Model

In the private information model from Hasbrouck (2002) we have the same framework as previously, except that

- the efficient price is $m_t = m_{t-1} + \lambda q_{1t}$ with $\lambda = 1$ as a liquidity parameter,
- the transaction prices $p_{1t} = m_t + cq_{1t}$ and $p_{2t} = m_{t-1} + cq_{2t}$ where q_{it} for market $i = 1, 2$, still takes the values $+1$ and -1 each with probability $\frac{1}{2}$.

Notice that the price on market 1 , p_{1t} , incorporates the information about the efficient price at time t , which in turn depends on the trade direction on market 1 . This new set-up implies that market 1 is expected to contribute for 100% to price discovery.

Table 2 summarizes the shares reported by the three type of measure. These are very close to what the structural model suggests, 100% . Again the 95% CI reports a much sharper bound than in the first model, even in the case of Hasbrouck IS.

Private and Public Information Model

In the third model, also from Hasbrouck (2002), we have the same framework as in the first example except that

- the efficient price: $m_t = m_{t-1} + \lambda q_{1t} + \upsilon_t$ where $\upsilon_t \sim N(0, \sigma_\upsilon^2)$ and $\sigma_\upsilon = 1$ and $\lambda = 1$
- the transaction price: $p_{1t} = m_t + c_1 q_{1t}$ and $p_{2t} = m_{t-1} + c_2 q_{2t}$ where $c_1 = 1$, $c_2 = 0$ and q_{it} for market $i = 1, 2$ remains the same as in the previous two examples.

This model suggests that market 1 still contributes at 100% to price discovery as its trade direction at time t determines the efficient price m_t . However, the trading costs are different such that $c_1 > c_2$, the cost of market-making is higher for traders of market 1 . Hasbrouck (2002) uses for this model the extreme value $c_2 = 0$, meaning that it is costless for market 2 , to trade at lagged prices.

Table 3 shows that the PDS measure is the closest to the expected 100% share. While none of them returned 100% , the PDS of Sultan and Zivot (2015) is close to the IS upper bound IS_U, and ICA captures the IS lower bound IS_L . Also in this design, with $c_2 = 0$, p_{2t} is exactly equal to the lagged value of the efficient price m_{t-1} , therefore Δp_t used in the VECM depends essentially on λ, q_{1t} and

v_t . This implies that the VECM residuals will merely be a mixture of Gaussian distributions with various means. This may explain why the ICA approach did not happen to accurately separate the composite innovations as the uniqueness of its decomposition is sensitive to Gaussian source signals.

In order to avoid this aspect to lead to inconclusive results, we allow the model to have a fatter tail distribution for the innovations, through a GARCH(1,1) model such as the one used to generate mean-adjusted DAX returns in Lütkepohl (2007). We obtain a significant improvement for all shares with values close to 100% . As a result, for the simulations on the Private and Public Information Model in the three-market models, GARCH(1,1) innovations are used.

The Modified "Roll" Model

This fourth and final example is taken from Sultan and Zivot (2015). This is a modified two-markets "Roll" model with one market mainly leading in price discovery contribution. This structural model will differ from the previous one by offering a shared contribution to price discovery as opposed to the 50% - 50% or all-or-nothing context of the last three examples.

Transaction prices are defined as:

$$p_{1t} = Dm_t + (1 - D)m_{t-1} + cq_{1t}$$

$$p_{2t} = (1 - D)m_t + Dm_{t-1} + cq_{2t}$$

where D is a binary variable such that $D = 1$ with probability 0.7 and $D = 0$ with probability 0.3 .

This structural model suggests that market 1 has a higher contribution to price discovery 70% and market 2 , 30% .

Table 4 reports again for Hasbrouck's IS a wide range, 51% to 81% , while for a predicted contribution to price discovery of 70% PDS and ICA report respectively 71% and 72% in the structural model for market 1. Also the 95% CI shows that any divergence in the shares throughout the simulations is very small.

1.3.2 Three Market Models

From the two-market simulations, it appears that PDS and ICA outperform the original Hasbrouck IS. However, it is important to highlight that all those price discovery measures depend on $\alpha_\perp$ the orthogonal to the vector of error correction coefficient in equation (1.3) through ξ' . Indeed, in a typical two-market framework, the elements of $\alpha_\perp$ have the same sign, this results in the PDS often being between 0% and 100% . When the number of markets is increased in our simulations, for example to three markets, we may notice that the PDS method can report values outside of the interval 0% and 100% . Indeed, as the approach is based on Euler's homogeneity theorem and exploits the fact that the sum of partial derivatives equals to the total derivative, it only requires that the sum of all price discovery shares is equal to 1 . We explore previous market data models in a context of a higher number of prices and compare them for the three measures.

The "Roll" Model

In this example of "Roll", we consider transaction prices p_{it} for market $i = 1, 2, 3$. The only modification to the initial model will be the number of markets involved in the transactions such that the contribution to price discovery is shared between all. We define the trade direction for the third market as $q_{3t} = \pm 1$, each value distributed with probability $\frac{1}{2}$ .The Information Share for each market should be a third of 100% .

Table 5 reports that the original Hasbrouck's IS is between 9.5% and 70.5% while PDS and ICA find respectively 33.4% with a sharper 95% CI for ICA.

Private Information Model

In this three-market set up, market 1 leads in price discovery with 100% of information share. The efficient price and the parameters continue to have the same expressions as previously but the transaction price for market 3 is $p_{3t} = m_{t-1} + cq_{3t}$ where $q_{3t} = \pm 1$ with probability $\frac{1}{2}$ and $c = 1$.

Table 6 reports again a wide range for Hasbrouck's IS and 92% of contribution to the price discovery for both PDS and ICA, which is not very close to the expected share from the structural model, 100% . This design may need to be improved in the three-price case. We recall that we did not wish to change the designs commonly used in the literature. However, it is important to point out that PDS and ICA perform similarly in this model.

Private and Public Information Model

Again the only modification to this model from the two-market case will be in the transaction prices. We add a third price expressed as $p_{3t} = m_{t-1} + c_3 q_{3t}$ where the trade direction takes the values 1 or -1 half of the time. The trading costs will be set to $c_1 = 1$, $c_2 = 0.01$, $c_3 = 0.01$. They satisfy $c_1 > c_2 = c_3$, showing that the cost is higher for informed traders and cheaper for traders on other markets. The costs of trading for markets 2 and 3 is the same since we assume that their traders have the same information. Also $c_2 = c_3 \neq 0$, because the VECM is not invertible when both equal 0 . In this set-up we expect market 1 to have a 100% information share.

In Table 7 , all the shares are close to the expected 100% share. However, the PDS measure reports on average, that shares are higher than 100% . Investigating the shares for markets 2 and 3 in that context, we notice that instead of being 0% as expected, those PDS shares were negative and pretty high compared to ICA shares. Yet, as mentioned earlier, we observed that the PDS shares sum up to 100%. This is insured by the Euler's homogeneity theorem exploited by the PDS approach.

The Modified "Roll " Model

This final three-market set up will illustrate how the three measures perform when we have more diversity in the contributions to price discovery. This model will offer a shared contribution to price discovery similar to its analogous model in the two-market case. The transaction prices will be:

$p_{1t} = D_1 m_t + (1 - D_1) m_{t-1} + c q_{1t}$

$p_{2t} = D_2 m_t + (1 - D_2) m_{t-1} + c q_{2t}$

$p_{3t} = D_3 m_t + (1 - D_3) m_{t-1} + c q_{3t}$ where $q_{it} = \pm 1$,with probability $1/2$ for $i = 1, 2, 3$;

D_1, D_2 and D_3 are binary variables such that

$D_1 = 1$, $D_2 = 0$ and $D_3 = 0$ with probability 0.5

$D_1 = 0$, $D_2 = 1$ and $D_3 = 0$ with probability 0.2

$D_1 = 0$, $D_2 = 0$ and $D_3 = 1$ with probability 0.3 .

This structural model suggests that market 1 has a higher contribution to price discovery 50% , market 2 , 20% and market 3 , 30% .

In Table 8, the results show that PDS and ICA, consistently return a contribution to price discovery of about 52% , this is close to the expected 50% from the structural model. The standard deviations and the 95% CI suggest that in models close to real market structure with multiple markets contributing to price discovery the PDS and the ICA perform similarly with the ICA still offering a sharper bound. Nevertheless, it will be convenient to evaluate its performances in strong empirical application with real market data.

1.4 Empirical Applications

1.4.1 Existing Applications

Beyond the high and low frequencies data used in most of the empirical studies on price discovery, the majority of existing applications illustrates the presence of price discovery in two schemes. The first scheme is between prices of products cross listed on different markets. In this case, the price discovery process push prices to converge to the same common price in order to avoid arbitrage opportunities. The second scheme considers products whose prices affect each other. In the latter, the assets do not exhibit the exact same prices but both depend on the same common trend.

Higher Frequency Data

Hasbrouck (1995) in his pioneer work illustrates the mechanism in the first scheme: one security and many markets. Determining the contributions to price discovery of thirty Dow Jones individual securities between the New York Stock Exchange (NYSE) and its competitors use high frequency quotes data. He suggests that by going to a higher frequency, the researcher reduces the divergence of the IS interval caused by the time aggregation. Jong and Schotman (2010) in there study of price discovery in fragmented markets, exploit two-minute interval high frequency quotes for 123 trading days. Similarly, with an emphasis on the variety of the stocks, Boehmer and Wu (2013) study a large panel of NYSE-listed stocks, and Brogaard et al. (2014), 120 stocks traded both on NASDAQ and NYSE grouping them by market capitalization. As reducing the time aggregation does not obviously solve the IS problem, Sultan and Zivot (2015) to illustrate their PDS, report their measure of price discovery at various frequencies. They notice that while the price leader remains the same, the reported value of price discovery differ depending on the time frequency but also on the level volatility.

Lower Frequency Data

As opposed to high frequency prices, the low frequency data are often less affected by microstructure effects. Moreover, they are more easily accessible, notably through Datastream or Bloomberg among others. Blanco et al. (2005) exploit the data of the Bank of England to study price discovery in daily Credit Derivatives Swaps (CDS) and Credit Spreads for 33 US and European entities. Later, Grammig and Peter (2013) revisiting their study confirm that five year CDSs contribute more to price discovery. Lien and Shrestha (2014) also use daily data covering the second half of 2009 to study price discovery in the CDS and corporate bonds markets, before observing the mechanism in the commodities energy market. In a similar context, Figuerola-Ferretti and Gonzalo (2010) also report daily futures to dominate daily spot prices in non-ferrous metals. Finally, in a more recent type of market, the market for the CO2 emissions, Uhrig-Homburg and Wagner (2009) also conclude that futures lead spot prices in price discovery regardless of the 2005 - 2007 experimental phase.

We consider both schemes of applications. Our first application will revisit Uhrig-Homburg and Wagner (2009) study of the prices of CO2 emissions allowances (two markets, low frequency application), while the second will study Russell 2000 Exchange Traded Funds (three-price, high frequency application).

1.4.2 Price Discovery in European CO2 Emissions

The European Union Allowance (EUA) is a permit authorizing the states of the European Union (EU) to emit a certain amount of CO2 under the EU Emission Trading Scheme (EU ETS). Each country determines the quantity of EUAs it allocates to its companies. The EU ETS therefore gives the opportunity to companies to acquire the right to pollute additionally in order to reach a higher level of output.

The installation of these emissions market has taken place in multiple phases. The first trading period ran from 2005 to 2007 , the second from 2008 to 2012 and the third phase started in 2013 . In the application of Uhrig-Homburg and Wagner (2009) that we will revisit in this section, they emphasize the first phase. They study the evidence of price discovery in these newly developed markets between the observed futures and the theoretical futures prices.

We attempt to reconstruct as closely as possible their data by extracting daily futures data from Bloomberg from June 24, 2005 to November 15, 2006. The theoretical futures are constructed using the following formula,

$$TF_t = e^{r(\tau - t)} S_t \tag{1.11}$$

where S_t denotes the spot price at time t , TF_t is the associated theoretical futures, r is the corresponding risk free rate and τ is the time to maturity. Similar to the authors, we proxy the risk free rate with the Euribor midquotes for maturities less than one year, and the Euroswap midquotes for the rest. The EUA futures considered are those expiring in December of each year. Indeed, they are the most liquid. In particular, we will compute the IS for futures maturing in December 2006 and 2007 to evaluate if, similarly to previous findings, futures dominate spot prices in price discovery.

In order to obtain the estimated innovations in (1.5), we respectively estimate the VECM of order 4 for the couple of futures and theoretical futures of 2006 (F06/TF06) and the VECM of order 3 for 2007 (F07/TF07). The lag orders have been obtained from the analysis of Akaike (AIC) and Schwartz (BIC) information criteria. Regarding the series for 2006 , the BIC selects the VAR(1) as best fit for the prices and the AIC the VAR(5). Even though the AIC often tends to select more lags than the BIC, the lag order 4 offer a richer feature for the VECM representation. In the case of the year 2007 , we choose the BIC which correctly selects the lag order 4 while the AIC directs us to include 5 lags. For brevity, we do not expose the detailed lag selection results here. Although, the impact of the order of the VECM used in calculating the IS is an interesting question, it relates to the impact of model misspecification on measures of price discovery and is left for future research.

Table 9 reports the price discovery measures obtained following our estimations. In the left panel, for the extensive data, the PDS and the ICA conclude that the futures contribute more to disclose the EUAs prices, whether their maturity is for December 2006 or 2007 . The PDS finds that F06 at 61% for the closest maturity and 75% for F07, while the ICA approach returns 79% for F06 and 74% for F07. At the same time, Hasbrouck's IS, returns very wide intervals. This divergence makes it impossible to conclude which market leads in price discovery. Indeed, it suggests a strong correlation in the residuals of spot and futures markets, regardless of the maturities. This may explain why we obtained various lag suggestions for our VECM estimation earlier. This ambiguity motivated Uhrig-Homburg and Wagner (2009) to investigate the VECM lag selection in a short-listed data, from December 2005 to November 2006 .Similarly to the work reported here, they find for both maturities that the VAR(2) fits better to the short-listed data, therefore the VECM(1) representation is more adequate to capture this cointegration relationship.

The price discovery measures for the short-listed data are presented in the right panel of our table.

We notice a small improvement of the IS lower and upper bounds with a range of 70% compared to 90% in the first case. The PDS and the ICA approaches both strongly indicate that the futures lead in price discovery regardless of the maturity, with values above 95% . Also, the PDS performs slightly better than the ICA in this short-listed data context. Nevertheless, the ICA information share seems to be more robust when the appropriate VECM order is not straight forward (left panel of Table 9). Indeed, we clearly see that all three measures remain consistent, reporting similar values for both maturities. This confirms our previous comment about the issue on the effect of misspecification in particular poor lag selection on the measures. In the next section, we will look at a three market prices application and evaluate the performance of the measures.

1.4.3 Price Discovery in the Russell 2000 Exchange Traded Funds (ETFs)

General ETFs Framework

Created in January 1993 , the first ETF was the S&P500's SPY (issued by SPRD). Subsequently, ETFs became more popular and are now omnipresent in financial markets, with more than $1,000$ ETFs traded by the end of 2011 . As of December 2014 , the total net assets of ETFs was almost two trillion dollars, that is 13% of total net assets managed by long-term mutual funds (source: www.etfdb.com). Since they share similar characteristics, ETFs are similar to mutual funds and regulated in the same manner. However, through their market structure, they offer lower costs and are more tax efficient. They are accessible to various types of investors as they do not require sophisticated investment strategies. Also, they trade immediately as opposed to standard mutual funds whose transactions are completed at the market closure.

Even though, ETFs accommodate better from individual investors to hedge funds managers, they give rise to arbitrage opportunities. In particular, Ben-David et al. (2011) discuss arbitrage opportunities when the ETF and its Net Asset Valuation (NAV) diverge considerably.

To understand further these arbitrage activities, we need briefly to describe the process for ETFs creation and redemption. ETFs are created by Authorized Participants (APs) in such a way that they replicate closely an index. This replication is motivated by the increasing demand for particular stocks on the market. Once created, APs provide an ETF company with the considered basket of securities. The latter will then issue shares amounting to the created ETF value and sold on the exchange market by APs. When the demand for the ETF in question rises, APs can sponsor the creation of more ETFs by the same process. Alternatively, APs may return ETFs shares to the ETF company and redeem them at a fair value. They obtain back the basket of securities which are more in demand on market. We may notice that when ETFs are not injected on the exchange market, they cannot be bought or sold.

The Russell 2000 ICA Information Shares

The Russell 2000 index groups the 2000 bottom companies of the Russell 3000 . An ETF for the Russell 2000 index is a fund that tracks the value of these companies. It trades like any common stock and is substantially liquid. Usually, several ETFs track the same index.Whether it is the case or not, the acquisition of one or more will depend on each investor's market strategy. For example, the performance of the S&P500 is tracked by SPY, IVV (issued by Ishares) and VOO (issued by Vanguard). As a result, market makers or traders participating in the ETFs market desire to benefit from any informational advantage accessible to them, notably the ETF leader. Sultan and Zivot (2015) found SPY to contribute more to price discovery than IVV.

We revisit this research question in a small capitalisation context. Our motivation comes first from the fact that price discovery is not uniquely important in large capitalization markets. The Russell 2000 represents the mutual funds indexed as "small cap" companies. It is to the The Russell 2000, the analogous of the S&P 600 for the S&P 500 index. Second, our approach allows it to go beyond the usual two most popular ETFs for tracking an index's performance. Indeed, even if the ETF contribution to price discovery is weak in comparison to its competitors, it may still exist and our methodology reports it.

The Russell 2000 performance is tracked by the TWOK (issued by SPRD), the IWM and VTWO (issued by Vanguard). Similar to other authors, we select bid-ask midpoints to reduce microstructure noise in transaction prices. The source for these minute by minute mid-quotes data is Bloomberg. They usually exhibit some non-trading periods which generate mistakes in bid and ask values. Barndorff-Nielsen et al. (2008) suggested to correct trades and quotes data to reduce the impact of those mistakes. As they recommend, we collect the data within the daily trading period, from 9:30 AM to 4:00 PM. However, as we only have access to a quite short time series, we will exclude those mistakes following one pratical option available in Bloomberg: we report, for non-trading minutes,the previous trading price. This price is the immediate price observed by traders.

Also, in order to picture usual trading situations, we attempt to quantify the price discovery in low and high volatility periods. In order to measure the volatility, we use the Russell 2000 volatility index

(RVX) of the Chicago Board Options Exchange. We notice in low volatility days that the index value is around 18 , while it doubles in high volatility days. In particular, we select the flash crash day of August 24^{th} 2015 (390 observations) where the RVX reached 46 . Regarding the low volatility days, our analysis will focus on two periods of 1950 observations each, the business days of November 30th to December 4^{th} of 2015 and January 25^{th} to 29^{th} of 2016 for which the respective average volatilities were 18.4 and 26 .

The Russell 2000 for US equity is stored under the ticker "RTY" in Bloomberg. We concentrate our study on the New York Stock Exchange (NYSE) market as transaction prices for this index on the Nasdaq or Bats markets are no different.

A descriptive analysis of the three ETFs selected shows that they clearly fluctuate together. Even though they do not exhibit the same levels of prices, most of the companies in their individual top 10 holdings are the same (source: www.etfdb.com). More rigorously, Figures 1 - 3 show that these time series are possibly cointegrated. The only exception is on August 24^{th}, where they seem to fluctuate differently with the TWOK that fell sharply from 68.72 to 65.36 from 9:47 AM to 9:56 AM before rising back to 67.07 . These changes may look negligible because we consider midquotes. When looking carefully at the bid and ask prices, we notice that this variation results from a decrease in the ask prices from 72.9 to 66.7. Indeed, that day was compared to May 6^{th} 2010 , the flash crash day. A study from Goldman Sachs from December 2015 on global investment reported that the series of rules and procedures enhanced after 2010 were insufficient to prevent the trading halts on August 24^{th} . As a result, several market and limit orders were executed at prices impacted by the volatility and multiple investors incurred losses. From the opening to 10:15 AM, the Nasdaq market suffered 1102 trading halts for the ETFs passing their Limit Up Limit Down (LULD) halts compared to their NAV.

In Table 10, we present the descriptive statistics of the midquotes under the scope of our analysis. In the first panel of the table, we observe that the TWOK is the cheapest ETF among those selected with an average price of 67 dollars per share while the most expensive ETF, the IWM is worth 112 dollars. Throughout the day, the range of the prices is six dollars for the first two ETFs more affected by the flash crash compared to the TWOK which exhibit only a range of about four dollars. This is confirmed by the standard deviations above one dollar for IWM and VTWO and around 75 cents for the third ETF.

In the second panel representing normal trading periods, the midquotes are higher averaging approximately 119 dollars for IWM, 95 dollars for VTWO and 71 dollars for TWOK. For the range within this longer time period, it was respectively about four dollars, three dollars and two dollars. The ranking is the same in regard to the standard deviations, expressing more dispersion in the iShares ETF prices.

In the last panel, again in a five days time span but with mid-low volatility, the average midquote prices are much smaller. Their ranges are of similar size to those in the second panel, while their standard deviations are all smaller than 75 cents.

As suggested in the literature and the methodology used in this paper, the price innovations are the residuals obtained from the estimation of the adequate VECM model. For brevity, we do not expose the steps of the estimation. The optimal VECM lags has been determined by Akaike and Schwartz information criteria and the existence of cointegration relationship has been tested with the Johansen cointegration test.

Although Hasbrouck (1995) suggests higher frequency as solution for reducing contemporaneous residual correlation, many studies still obtained a wide range for the IS at no more than 1-sec interval (see Huang (2002) and Grammig et al. (2005)). At the same time, increasing the time aggregation would downsize the effect of microstructure noise on the measures. Since our available data are at 1min frequency, it will be interesting to look at lower frequencies whether the measures would return different values.

Therefore, Table 11 reports the Information shares obtained for three frequencies: one minute (1min), five minutes (5min), ten minutes (10min). We do not show the Hasbrouck IS as it would imply computing the IS for all the six possible permutations of the prices and determine the lower and upper bounds. We focus attention uniquely on the PDS and the ICA as they outperform the IS in our previous examples.

In general, we observe that the ICA selects the iShares as leading ETF in contribution to the disclosure of the real value of the Russell 2000. The contributions found differ for the various frequencies considered but remain always larger than 60% for the IWM (iShares). On the contrary, the price discovery share of Sultan and Zivot (2015) finds negative contributions to the price discovery in some cases.In particular, for the VTWO on a high volatility day and the TWOK in normal and low volatility days (all at 1min intervals). It also returns negative values for the TWOK at 10min and 5min frequencies respectively on August 24^{th} and November 30^{th} to December 4^{th}. The interpretation of negative values for shares is not very intuitive. When they are close to zero, they may be interpreted as zeroes. They are hardly interpretable when they are negative with large absolute value like the -52%

contribution to price discovery for the TWOK for January 25^{th} to January 29^{th}. Additionally, since the shares must sum up to one, the lack of rationale behind negative shares also implies a lack of intuition on shares with positive and larger contribution than 100% (see for example 147% contribution for the VTWO in our case of study).

We also notice on the high volatility day that both approaches conclude that IWM is the price leader at 1min and 10min frequencies. At 5min frequency, both ICA and PDS choose the VTWO as price leader, however the ICA interestingly returns zero as contribution for the IWM. This remark casts some doubts on shares obtained for the VTWO and the TWOK. This urges us to investigate further how the ICA approach performs at decomposing the residuals into independent components on these time series.

ICA Permutation Ambiguity in the Russell 2000 Information Share

In this section, we seek to evaluate the robustness of our results by verifying whether they suffer the ICA ambiguity problem. This will be realized by investigating the strength of the correlation between the independent components obtained from the ICA and the composite innovations collected from the VECM estimation. Indeed, recall the ICA approach attempts to separate mixed signals into independent source signals. That is, the resulting components are independent but individually contribute to the formation of each original signal. For example, the first composite innovation in our three-price empirical application is expressed as $\widehat{u}_1 = A_{11}IC_1 + A_{12}IC_2 + A_{13}IC_3$, where $\widehat{u}_i$ is the residual associated to the price i , A_{ij} is the element at the row i and column j of the mixing matrix A associated with each ICj for $j = 1, 2, 3$.

As pointed out earlier, we may use the correlation of energies as a criterion to measure dependence between the independent innovations and the original composite innovations. Table 12 summarizes the coefficients of correlation between the squares of each original innovation and the three independent components. It reports that on the high volatility day at 1min frequency, the ICA approach efficiently identifies the price innovations and their associated independent components. Indeed, the analysis of the correlation coefficients clearly associates $\widehat{u}_1$ with IC_2 , $\widehat{u}_2$ with IC_1 and $\widehat{u}_3$ with IC_3 (with all coefficients closed to one). As the FastICA algorithm (see Appendix 1.6.1) does not maintain the order of the independent components, we recall that this step is very important to permute adequately the columns of the mixing matrix. In our application, this permutation prevents us from confusing the information shares of prices 2 and 1 .

For the other frequencies and the lower volatility days, the coefficients of correlation show that among all the independent components obtained, only one is strongly tied to all three price innovations. This matches with one of the exceptions highlighted above, in the section on the ICA permutation ambiguity.

This analysis is confirmed graphically, when we plot the innovations and the ICs in the case of the high volatility day (see Figures 4,5 and 6 in Appendix 1.6.2). We clearly observe at 1min frequency that each independent component fluctuates similarly to a unique residual. While for the other cases, the composite innovations graphs are very similar. In particular at 5min and 10min, respectively, the estimated innovations of the VTWO and TWOK and those of the IWM and the VTWO are almost identical. This explains why the ICA struggles to divide them in three different components, they all appear like homothetic transformations of the same original signals. In other words, the ICA approach does not properly separate the original signals. We observe the same for the other time periods studied. This suggests that the ICA information shares obtained above, except for the case of august 24th at 1min frequency, should be interpreted with caution.

1.5 Conclusion

This paper suggests the use of the ICA approach in finding a unique IS. Indeed, it is often pointed out the severe identification issues that the Hasbrouck IS suffers. When the correlation among the transaction prices innovations is strong, the IS exhibits a large gap between its lower and upper bounds. This divergence makes the identification of the market leading in the price discovery process difficult while this information may be very useful for policy makers and other market participants. Various studies propose an improvement of the measure in order to resolve the uniqueness problem but are qualified as arbitrary since they are often based on intuition.

Although, many of those studies construct new measures of the price discovery, others like Grammig and Peter (2013) remain in line with Hasbrouck (1995) IS. They attempt to overcome its price ordering difficulty by assuming that the innovations may follow a mixture of Gaussian but require the existence of tail dependence in the data. Our approach instead is semi-parametric and therefore less restrictive. Indeed, with its very appealing applications in the neural network literature, the ICA approach offers a unique framework for separating mixed signals in independent components. In this context, it is sufficient for the composite innovations in transaction prices to satisfy some statistical properties, to be partitioned in components specific to each market.

We illustrate our approach by applying it to the popular two-market simulated models. Our results are consistent with those predicted by the structural models. We recall that the designs of those simulations are standard in the price discovery literature and are the necessary first step of validation of price discovery measures. Keeping the same features, we further simulate three-price cases and obtain robust results while some of the leading measures of price discovery return negative shares. In the last section, we conduct two empirical applications illustrating both simulations set-up , one to the European CO2 emissions market and the other to the Russell 2000 ETFs market.

Our results for the EUA market comply with the PDS results, consistently selecting the futures market as price leader. Nevertheless, in this context of non unique lag order selection for the VECM estimation, the ICA approach returns closer contributions to price discovery for different maturities. The impact of models misspecification such as issues linked to lag selection, length or frequency of time series on the various measures are left for future research.

On the other side, our application to the Russell 2000 confirms the existence of a mechanism of price discovery in the ETFs market. In particular, exploring various frequencies (1min, 5min, 10min) and volatility days, we observe at 1min frequency that the ICA clearly identifies the IWM as the price leader. It is important to highlight that our choice of data is slightly unorthodox, therefore we may not quantify its effect on the measures. We mean that, first the ETFs are from the same market (NYSE) and second, they are not built with the exact same companies. We assumed that tracking the same index and being largely formed by the same companies is enough for reliable data construction. In future work, it is compelling to look at cross-listed ETFs while considering a nonlinear ICA approach in order to allow for a richer decomposition of the price innovations.

1.6 APPENDIX

1.6.1 Appendix A: Independent Component Analysis

The ICA method is based on capturing the diverse features in the data formed by the observed signals. From the equation (1.9), if we knew A we could recover $S(t)$ by $S(t) = WX(t)$ where $W = A^{-1}$ with A invertible. The whole ICA framework will be derived from this principle. It will attempt to estimate W maximizing the independence between the source signals S .

There exists few ways of exploiting the information carried by the observed signals in order to estimate W. Cardoso (1997) , notably, employed the Maximum Likelihood estimation to separate source. Nevertheless, this method requires an accurate approximation of the density functions of the independent components. Alternatively, one could use the Kullback-Leibler divergence, the minimization of mutual information as implemented by Bell and Sejnowski (1995) or the maximization of the nongaussianity strongly advocated by Hyvärinen et al. (2004) in their studies. In this paper, we utilize the ICA method based on the maximisation of the nongaussianity.

The nongaussianity has often been used in the literature as a criterion for independence. In fact, it provides information about how different the observed signals are from Gaussian signals. Two ways of measuring nongaussianity are the Kurtosis or fourth order cumulant and the negentropy. Because the Kurtosis has been found to be very sensitive to outliers (Huber (1985)), the negentropy is a more appealing measure of nongaussianity.

Before giving the definition of the negentropy, we need to recall that the entropy of a random variable evaluates the degree of information that the observation of the variable gives. The more random, i.e unpredictable and unstructured the variable is, the larger its entropy. Cover and Thomas (2012) and Papoulis (1991) defined the (differential) entropy H of a random variable y with density $f(y)$ as:

$$H(y) = -\int f(y)\log f(y)dy$$

The latter proved that *a gaussian variable has the largest entropy among all random variables of equal variance*. As a result, using the difference between the entropy of gaussian signals and the entropy of observed signals, the negentropy will allow us to measure the nongaussianity.

The negentropy J is usually defined as

$$J(y) = H(y_{gauss}) - H(y)$$

where y_{gauss} is a gaussian random variable with same variance as y .

However, the negentropy is difficult to estimate. Jones and Sibson (1987) proposed an approximation using higher order moments including the Kurtosis. Since the kurtosis is not robust to outliers, Hyvarinen (1998) proposed an approximation using a non quadratic function $G(\cdot)$

$$J(y) \propto \left[E\left\{G(y)\right\} - E\left\{G(v)\right\}\right]^2$$

where v is a Gaussian variable of zero mean and unit variance. $G(\cdot)$ is chosen carefully for approximating the negentropy. Usually an adequate choice of G is $G(u) = \log(\cosh(u))$ or $G(u) = -\exp(-u^2/2)$.

Later, Hyvärinen and Oja (2000) introduced the FastICA algorithm, which is an iterative fixed point method maximizing the approximation. For one computational unit, they found that the optimal mixing matrix would be obtained by choosing a separating vector

$$w^+ = E\{\tilde{x}g(w^T\tilde{x})\} - E\{g'(w^T\tilde{x})\}w$$

where w is an initial weight vector that will be updated at each iteration with $w = \frac{w^+}{\|w^+\|}$ until the dot-product of the new and old values of w is close enough to 1. $\tilde{x}$ is obtained by centering (making x zero mean variables) and whitening ($E(\tilde{x}\tilde{x}^T) = I$) the observed signals. The function $g(\cdot)$ used here is the derivative of a non quadratic function $G(\cdot)$.

This iterative procedure for one unit algorithm can be extended to the computation of n units, meaning n weight vectors $w_1, \cdots, w_n$, where each represents one row of W . However, this extension requires to decorrelate the outputs $w_1^T x, \cdots, w_n^T x$ after each iteration to avoid that they converge to the same maxima.

The classical method of decorrelation is the deflation scheme based on Gram-Schmidt-like . Indeed, estimating the q vectors $w_1, \cdots, w_q$ for q independent components, we run the one-unit fixed point algorithm for w_{q+1} and after each iteration step subtract from w_{q+1} the projections of previously estimated q vectors and then re-normalize w_{q+1} :

$$w_{q+1} = w_{q+1} - \sum_{j=1}^{q} w_{q+1}^T w_j w_j \text{with } w_{q+1} = \frac{w_{q+1}}{\parallel w_{q+1} \parallel}$$

This decorrelation method as pointed out by Karhunen et al. (1997) may prioritize some vectors over others. The symmetric decorrelation instead estimates all the independent component at the same time. It utilizes W , the matrix $(w_1, \cdots, w_n)'$ of the vectors, and has it converging through the following iterative algorithm:

1. Let $W = W/\sqrt{||WW'||}$

   Repeat 2. until convergence:

2. Let $W = \frac{3}{2}W - \frac{1}{2}WW'W$

where the norm in step 1 can be almost any ordinary matrix norm (see Hyvärinen et al. (2004) for more details).

1.6.2 Appendix B: Figures

Figure 1: Bid-Ask Midprices, August 24th 2015

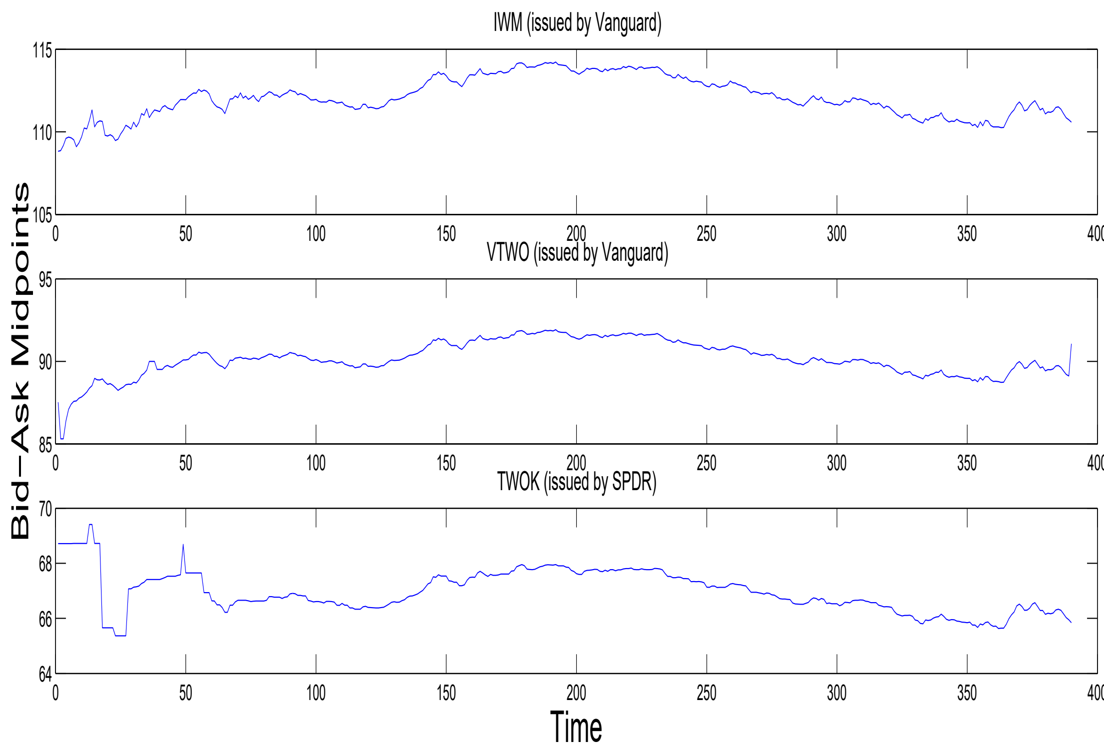


Figure 2: Bid-Ask Midprices, November 30th 2015 to December 4th 2015

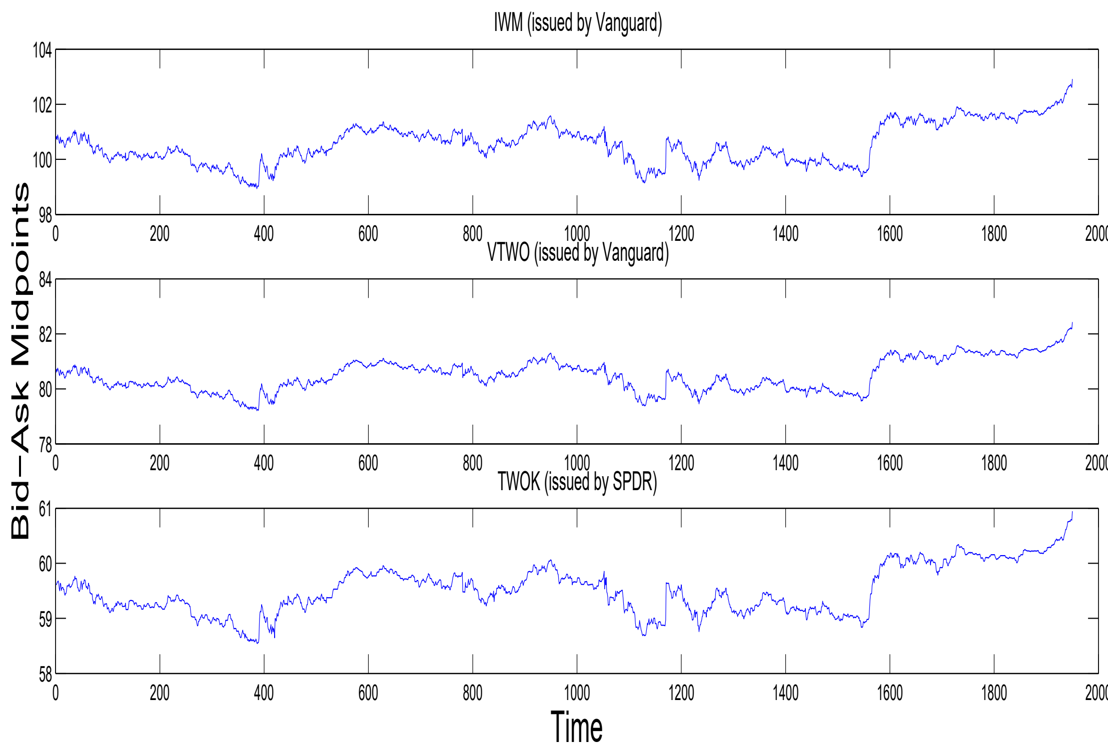


Figure 3: Bid-Ask midprices, January 25th 2016 to January 29th 2016

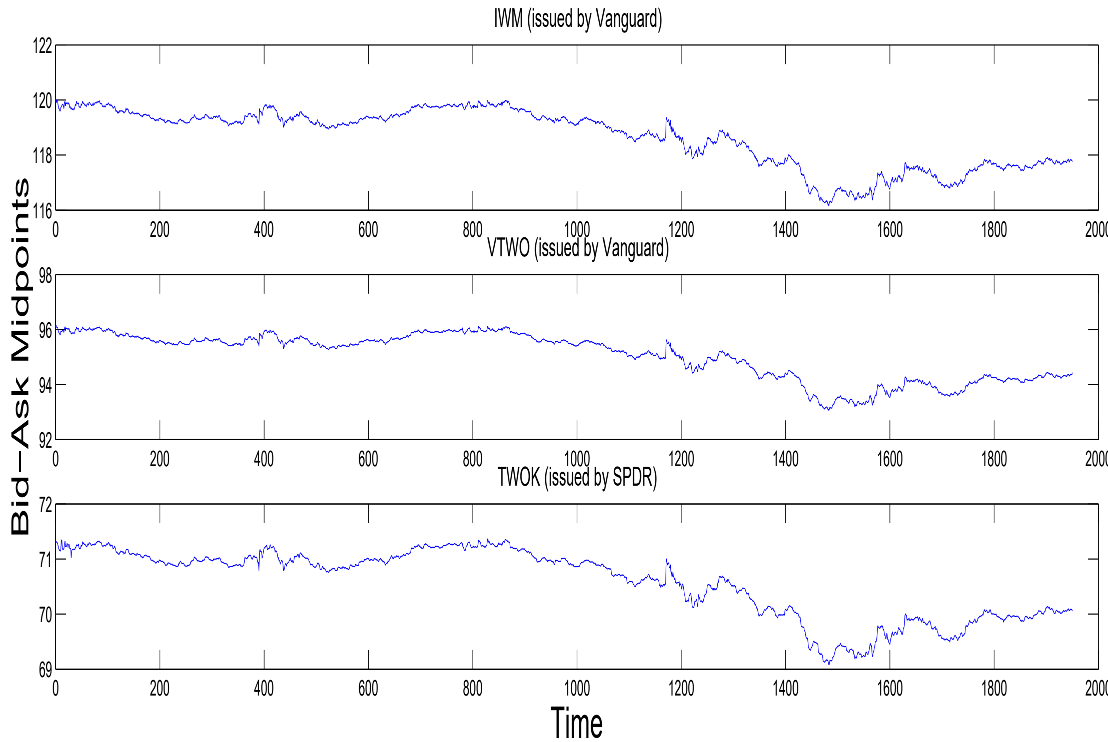


Figure 4: Correlation of energies at 1min frequency, August 24th 2015

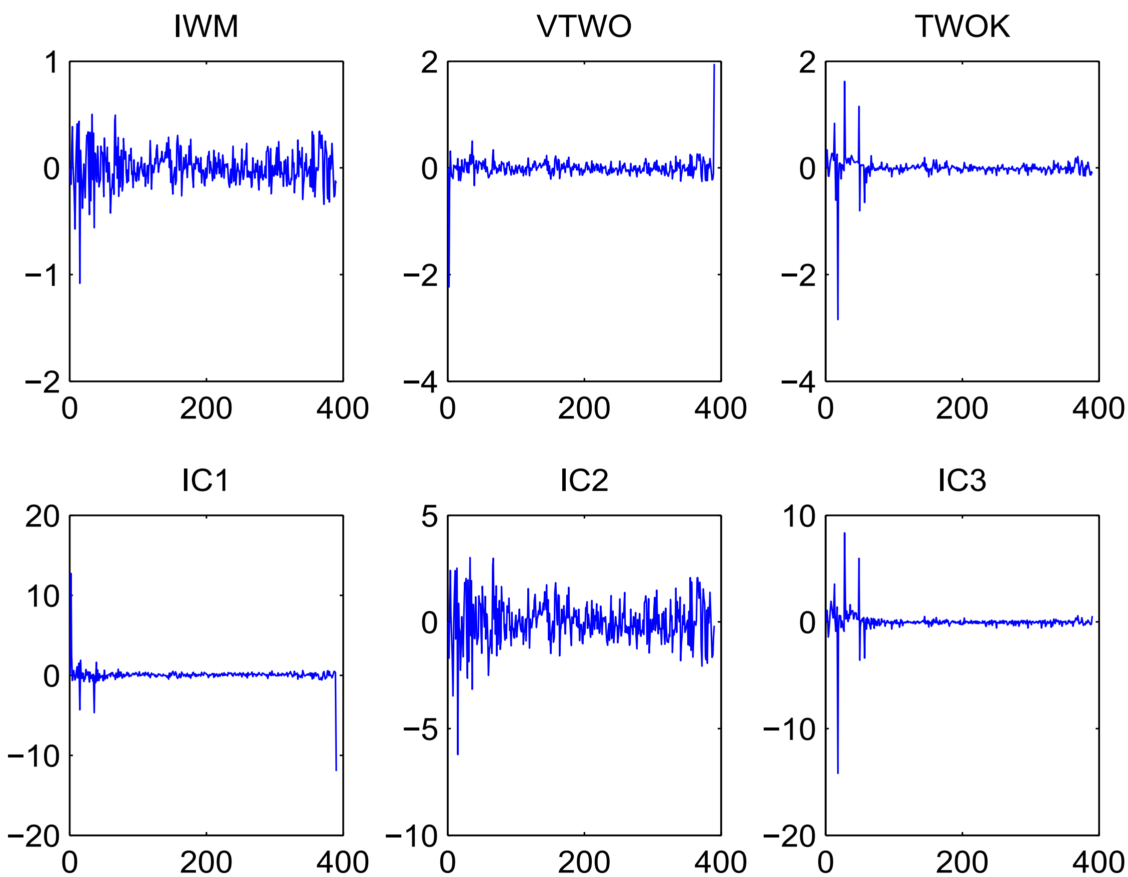


Figure 5: Correlation of energies at 5min frequency, August 24th 2015

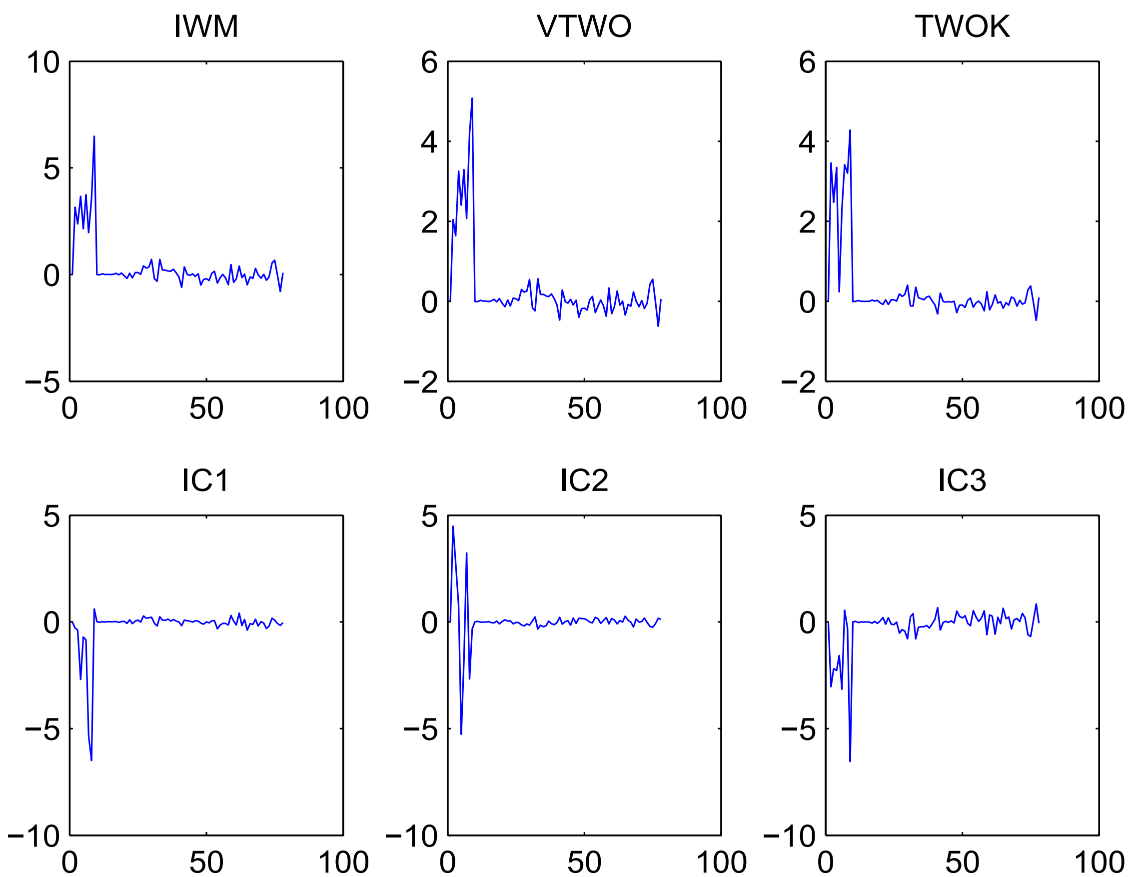
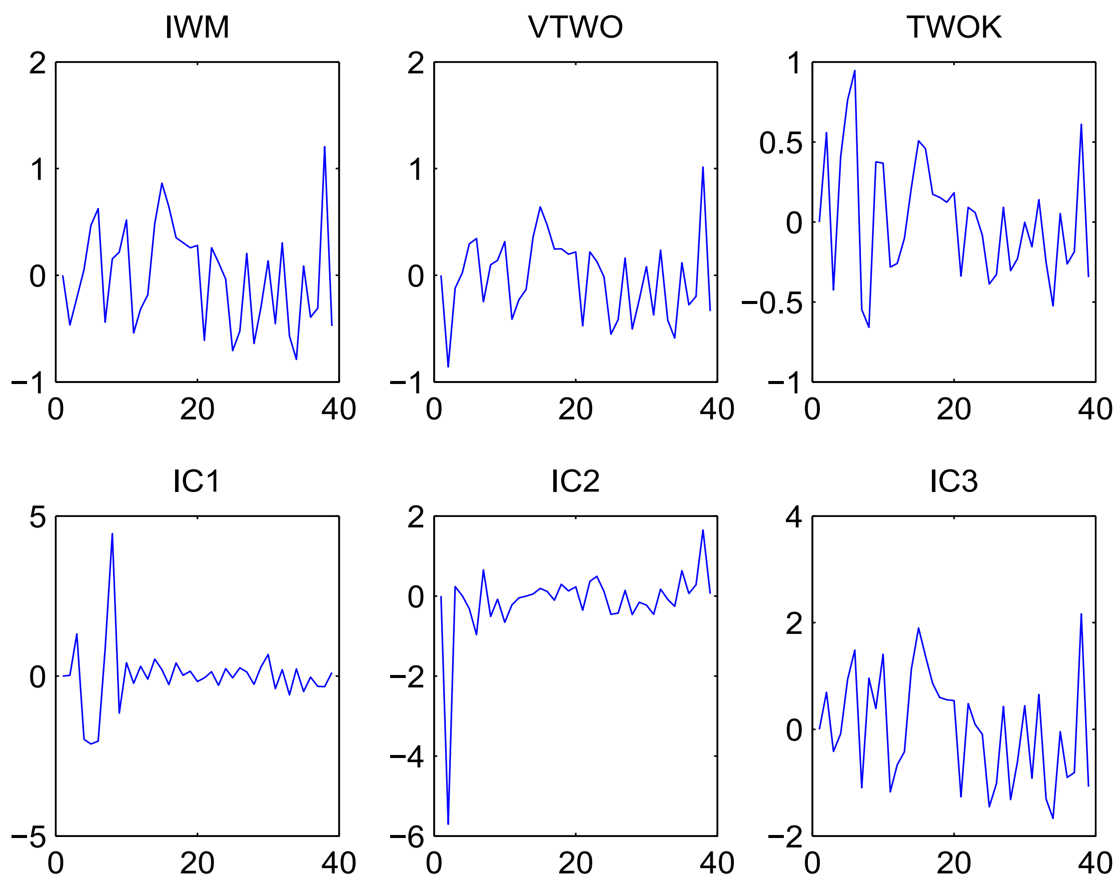


Figure 6: Correlation of energies at 10min frequency, August 24th 2015



1.6.3 Appendix C: Tables

Table 1: Two markets "Roll" model

Model 1 is simulated $1,000$ times.

Efficient Price: $m_t = m_{t-1} + \upsilon_t$, $\upsilon_t \sim N(0, \sigma_\upsilon^2)$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$; Transaction price: $p_{it} = m_t + cq_{it}$ for $i = 1, 2$; $c = 1$ and $\sigma_\upsilon = 1$ for $T = 100,000$ observations; Price discovery in the structural model for market 1 is 50%.

	Mean	Std. Dev	95% CI
Hasbrouck	0.211	0.012	[0.188, 0.234]
	0.788	0.012	[0.765, 0.811]
Sultan et al	0.500	0.022	[0.455, 0.544]
ICA	0.499	0.019	[0.461, 0.537]

Table 2: Two markets private information model

Model 2 is simulated $1,000$ times.

Efficient Price: $m_t = m_{t-1} + \lambda q_{1t}$ with $\lambda = 1$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$; Transaction price: $p_{1t} = m_t + cq_{1t}$ and $p_{2t} = m_{t-1} + cq_{2t}$ $c = 1$ and $T = 100,000$ observations; Price discovery in the structural model for market 1 is 100%.

	Mean	Std. Dev	95% CI
Hasbrouck	0.999	0.032	[0.936, 1.0]
	0.999	0.032	[0.936, 1.0]
Sultan et al	0.999	0.032	[0.936, 1.0]
ICA	0.999	0.032	[0.936, 1.0]

Table 3: Two markets private and public information model

Model 3 is simulated $1,000$ times.

Efficient Price: $m_t = m_{t-1} + \lambda q_{1t} + \upsilon_t$, $\upsilon_t \sim N(0, \sigma_\upsilon^2)$, $\sigma_\upsilon = 1$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$; Transaction price: $p_{1t} = m_t + c_1 q_{1t}$ and $p_{2t} = m_{t-1} + c_2 q_{2t}$; $c_1 = 1$, $c_2 = 0$ and $T = 100,000$ observations; Price discovery in the structural model for market 1 is 100%.

	Mean	Std. Dev	95% CI
Hasbrouck	0.900	0.008	[0.883, 0.917]
	0.984	0.003	[0.978, 0.991]
Sultan et al	0.975	0.005	[0.966, 0.984]
ICA	0.900	0.010	[0.880, 0.919]

Table 4: Two markets modified "Roll" model

Model 4 is simulated $1,000$ times.
Efficient Price: $m_t = m_{t-1} + \upsilon_t$, $\upsilon_t \sim N(0, \sigma_\upsilon^2)$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$; Transaction price: $p_{1t} = Dm_t + (1-D)m_{t-1} + cq_{1t}$ and $p_{2t} = (1-D)m_t + Dm_{t-1} + cq_{2t}$ $D = 1$ with probability 0.7 and $D = 0$ with probability 0.3 , $c = 1$ and $\sigma_\upsilon = 1$ for $T = 100,000$ observations; Price discovery in the structural model for market 1 is 70%.

	Mean	Std. Dev	95% CI
Hasbrouck	0.509	0.015	[0.480, 0.538]
	0.807	0.011	[0.784, 0.829]
Sultan et al	0.706	0.016	[0.674, 0.738]
ICA	0.718	0.017	[0.684, 0.751]

Table 5: Three-market "Roll" model

Model 1 for three markets is simulated $1,000$ times.
Efficient Price: $m_t = m_{t-1} + \upsilon_t$, $\upsilon_t \sim N(0, \sigma_\upsilon^2)$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$; Transaction price: $p_{it} = m_t + cq_{it}$ for $i = 1, 2, 3$; $c = 1$ and $\sigma_\upsilon = 1$ for $T = 100,000$ observations; Price discovery in the structural model for market 1 is 33.33% .

	Mean	Std. Dev	95% CI
Hasbrouck	0.095	0.009	[0.077, 0.112]
	0.706	0.014	[0.678, 0.733]
Sultan et al	0.334	0.024	[0.286, 0.382]
ICA	0.334	0.016	[0.302, 0.366]

Table 6: Three-market private information model

Model 2 for three markets is simulated $1,000$ times.
Efficient Price: $m_t = m_{t-1} + \lambda q_{1t}$ with $\lambda = 1$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$ for $i = 1, 2, 3$; Transaction price: $p_{1t} = m_t + cq_{1t}$, $p_{2t} = m_{t-1} + cq_{2t}$ and $p_{3t} = m_{t-1} + cq_{3t}$ $c = 1$ and $T = 100,000$ observations; Price discovery in the structural model for market 1 is 100% .

	Mean	Std. Dev	95% CI
Hasbrouck	0.705	0.013	[0.679, 0.730]
	0.949	0.006	[0.936, 0.961]
Sultan et al	0.919	0.008	[0.902, 0.935]
ICA	0.924	0.008	[0.908, 0.940]

Table 7: Three-market private and public information model

Model 3 for three markets with $GARCH(1,1)$ errors is simulated $1,000$ times.
Efficient Price: $m_t = m_{t-1} + \lambda q_{1t} + \upsilon_t$, $\upsilon_t = \sigma^2_{t/t-1}\epsilon_t$, $\epsilon_t \sim N(0,1)$, $\sigma^2_{t/t-1} = 0.0003 + 0.12\upsilon^2_{t-1} + 0.771\sigma^2_{t-1/t-2}$; Trade direction: $q_{it} = \pm 1$, each with prob. $1/2$; Transaction price: $p_{1t} = m_t + c_1 q_{1t}$, $p_{2t} = m_{t-1} + c_2 q_{2t}$ and $p_{3t} = m_{t-1} + c_3 q_{3t}$ $c_1 = 1$, $c_2 = 0.01$, $c_3 = 0.01$ and $T = 100,000$ observations; Price discovery in the structural model for market 1 is 100%.

	Mean	Std. Dev	95% CI
Hasbrouck	0.9994	0.0003	$[0.999, 1.0]$
	0.9997	0.0005	$[0.999, 1.0]$
Sultan et al	1.0184	0.0305	$[0.957, 1.079]$
ICA	0.9994	0.0004	$[0.999, 1.0]$

Table 8: Three-market modified "Roll" model

Model 4 for three markets is simulated $1,000$ times.
Efficient Price: $m_t = m_{t-1} + \upsilon_t$, $\upsilon_t \sim N(0, \sigma^2_\upsilon)$; Trade direction: $q_{it} = \pm 1$, each with pr. $1/2$ for $i = 1,2,3$; Transaction price: $p_{1t} = D_1 m_t + (1 - D_1) m_{t-1} + c q_{1t}$where $D_1 = 1$, $D_2 = 0$ and $D_3 = 0$ with probability 0.5; $p_{2t} = D_2 m_t + (1 - D_2) m_{t-1} + c q_{2t}$ where $D_1 = 0$, $D_2 = 1$ and $D_3 = 0$ with probability 0.2; $p_{3t} = D_3 m_t + (1 - D_3) m_{t-1} + c q_{3t}$, $D_1 = 0$, $D_2 = 0$ and $D_3 = 1$ with probability 0.3 $c = 1$ and $\sigma_\upsilon = 1$ for $T = 100,000$ observations; Price discovery in the structural model for market 1 is 50%.

	Mean	Std. Dev	95% CI
Hasbrouck	0.270	0.014	$[0.244, 0.297]$
	0.695	0.013	$[0.668, 0.723]$
Sultan et al	0.517	0.020	$[0.477, 0.557]$
ICA	0.518	0.018	$[0.483, 0.554]$

Table 9: Price Discovery in the CO2 Emissions Markets

Information Shares	**06/2005 - 11/2006**				**12/2005 - 11/2006**			
	F06	**TF06**	**F07**	**TF07**	**F06**	**TF06**	**F07**	**TF07**
Hasbrouck	0.97	0.03	0.99	0.01	0.998	0.002	0.998	0.002
	0.07	0.93	0.11	0.89	0.77	0.23	0.77	0.23
Sultan et al	0.61	0.39	0.75	0.25	0.99	0.01	0.99	0.01
ICA	0.79	0.21	0.74	0.26	0.98	0.02	0.98	0.02

Table 10: Russell 2000 ETFs descriptive statistics

Stats	**Aug 24, 2015**			**Nov 30 - Dec 4, 2015**			**Jan 25 - Jan 29, 2016**		
	IWM	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**
Min	108.82	85.30	65.36	116.15	93.06	69.08	98.93	79.21	58.54
Max	114.22	91.92	69.41	120.02	96.15	71.37	102.92	82.43	60.94
Mean	112.08	90.16	66.96	118.66	95.05	70.57	100.56	80.49	59.51
Std. Dev	1.21	1.06	0.76	1.03	0.83	0.61	0.72	0.58	0.43

Table 11: Price Discovery in the Russell 2000 ETFs

Aug 24, 2015

Information Shares	**1min**			**5min**			**10min**		
	IWM	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**
Sultan et al.	0.59	−0.10	0.51	0.32	0.65	0.03	0.79	0.26	−0.05
ICA	0.66	0.05	0.29	0.00	0.62	0.38	0.66	0.19	0.15

Nov 30 - Dec 4, 2015

Information Shares	**1min**			**5min**			**10min**		
	IWM	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**
Sultan et al.	0.31	0.87	−0.18	0.20	0.88	−0.08	0.20	0.72	0.08
ICA	0.76	0.16	0.08	0.81	0.17	0.02	0.67	0.31	0.02

Jan 25 - Jan 29, 2016

Information Shares	**1min**			**5min**			**10min**		
	IWM	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**	**IWM**	**VTWO**	**TWOK**
Sultan et al.	0.05	1.47	−0.52	0.39	0.41	0.20	0.16	0.64	0.20
ICA	0.66	0.24	0.10	0.88	0.03	0.09	0.87	0.01	0.12

Table 12: Correlation between innovations and ICs

Aug 24, 2015

Innovations	**1min**			**5min**			**10min**		
	IC$_1$	**IC**$_2$	**IC**$_3$	**IC**$_1$	**IC**$_2$	**IC**$_3$	**IC**$_1$	**IC**$_2$	**IC**$_3$
$\widehat{u}_1$	0.09	1	0.01	0.27	0.26	0.94	0.13	0.06	0.95
$\widehat{u}_2$	0.99	0.05	0.00	0.50	0.31	0.81	0.15	0.54	0.74
$\widehat{u}_3$	0.01	0.03	1	0.53	0.41	0.74	0.48	0.19	0.48

Nov 30 - Dec 4, 2015

Innovations	**1min**			**5min**			**10min**		
	IC$_1$	**IC**$_2$	**IC**$_3$	**IC**$_1$	**IC**$_2$	**IC**$_3$	**IC**$_1$	**IC**$_2$	**IC**$_3$
$\widehat{u}_1$	0.03	0.27	0.97	0.06	0.93	0.09	0.95	0.06	0.06
$\widehat{u}_2$	0.03	0.29	0.99	0.06	0.96	0.04	0.97	0.06	0.01
$\widehat{u}_3$	0.16	0.27	1	0.1	0.99	0.04	0.98	0.02	0.02

Jan 25 - Jan 29, 2016

Innovations	**1min**			**5min**			**10min**		
	IC$_1$	**IC**$_2$	**IC**$_3$	**IC**$_1$	**IC**$_2$	**IC**$_3$	**IC**$_1$	**IC**$_2$	**IC**$_3$
$\widehat{u}_1$	0.12	0.95	0.11	0.13	0.95	0.17	0.15	0.01	0.93
$\widehat{u}_2$	0.11	0.97	0.05	0.09	0.97	0.15	0.16	0.00	0.93
$\widehat{u}_3$	0.08	1	0.07	0.02	0.99	0.05	0.11	0.01	0.99

Chapter 2

Model Selection with Possibly Weak Instruments

2.1 Introduction

Ordinary least squares is the most intuitive method for modelling the relationship between two or more variables. However, it relies on strong assumptions among which, one requires explanatory variables to be uncorrelated with the disturbance term. In case of failure of this condition, instrumental variable regressions are the answer to eliminate the inconsistency generated by this limitation. Examples of cases in which independent variables are correlated with the error term are very frequent in economic studies. They range from situations in which explanatory variables are measured with error or correlated with an omitted variable to simultaneous equations models. They may also be encountered in time series regression models when the lagged dependent variable is correlated with the error term that is itself often autoregressive. This makes instrumental variables methods extremely useful in numerous applications in economics where instruments must be both valid and relevant.

The validity or exogeneity condition (often termed orthogonality condition) simply implies that instruments are not correlated with the error term in the original regression, while the relevance condition suggests that the set of candidate instruments must contain only those making a significant contribution to the explanatory power of the endogenous regressors. While studies show that the failure of both conditions lead to biased and inconsistent estimators, it might be tempting to rely on only one of them when the second is difficult to achieve. Unfortunately, as pointed out by Murray (2006), strong and almost valid instruments tend to bias little the two stage least squares estimator as opposed to almost valid weak instruments. This suggests that the weak instruments problem might require a more particular focus than the other condition. In this paper we are interested in instruments' selection in the case of mildly weak instruments, when they are still able to improve estimation. We will explore instrument selection when instruments' strength deteriorates and propose a suitable information criterion.

Our study is in line with a great amount of work in the literature on weak instruments and weak identification. A well known example of weak identification issues is illustrated in Angrist and Krueger's (1991) Cross-sectional Instrumental Variable (IV) model where the authors estimate returns to education using quarter of birth and its interactions as instruments for educational attainment. As opposed to common beliefs at the time, suggesting that the inclusion of a wide number of instruments in IV regressions is sufficient to improve the IV estimator, Angrist and Krueger (1991) had at their disposal more than 178 interaction terms and covariates as instruments for schooling but obtained estimators with poor asymptotic inferences. Bound et al. (1995) demonstrate that despite Angrist and Krueger (1991) large sample size data (329,509 observations), their estimates are biased and inconsistent as a result of the weak correlation between the endogenous variable and its instruments. The weakness of instruments cannot simply be categorized as a small sample problem. At least in terms of model selection, the increase of the sample size is not a solution to the issues generated by weak identification.

Another illustration of weak identification problems occurs in Campbell (2003) estimation of the elasticity of inter-temporal substitution between consumption growth (ΔC_{t+1}) and the return on some asset i ($r_{i,t+1}$) from t to $t+1$ given a set of instruments. Under the homoskedasticity condition, it should be possible to recover the two stage least square (2SLS) estimate of the regression of $r_{i,t+1}$ on ΔC_{t+1} using the reverse regression (regression of ΔC_{t+1} on $r_{i,t+1}$). In its comparison of the results of both regressions, Campbell (2003) obtained very different confidence intervals for their respective estimates. Earlier, Hahn and Hausman (2002) who propose a statistic to test the null of strong instruments by comparing both regressions (forward and reverse) in the Consumption Capital Asset Pricing Model (CCAPM). In the same order, several procedures on assessing and detecting instruments weakness, build on these characteristics of regressions in presence of weak correlation.

Stock et al. (2002) for example, in their study, rely on the first stage $F-$statistic to test the null hypothesis that the ratio of concentration parameter to the number of candidate instruments belongs to a certain threshold. They recommend that $F-$statistic exceeds 10, to have reliable 2SLS inference. The evaluation of the R^2 and the partial R^2 of Shea (1997) goes in the same direction.

In response to those inferential issues, some papers suggest robust tests in presence of weak instruments. Among fully robust tests, the first to point out is the Anderson-Rubin statistic (see Dufour and Taamouti (2005)) which is minimized to derive the Limited Information Maximum Likelihood (LIML) estimator. Similar to the Anderson-Rubin statistic, there are also the Kleibergen (2002) and Moreira (2009) statistics which are part of the Gaussian tests family.

As an alternative to those fully robust test which are sometimes difficult to implement when there are more than one endogenous regressor, there exists partially robust methods in the sense that they improve on the 2SLS estimator. They range in methods based on k-class estimators: the LIML estimator, the Fuller k-estimator and more recently the Bias adjusted Two Stage Least Squares (B2SLS) of Donald and Newey (2001) estimator. In general, for this class of estimators the normalized bias tend to be smaller as the sample size is larger than the squared degree of overidentification, in particular when the number of instruments becomes larger (see Bekker (1994) for further details on the comparison between the LIML and the 2SLS estimators).

In recent years, researchers have been more attracted by the impact on estimation of instruments which are not completely weak and how they contribute to consistent estimation. The goal being to investigate how much information we discard by assuming that instruments are weak and consequently eliminate them. Hahn and Kuersteiner (2002) establish in the linear IV framework following Staiger and Stock (1997) that it is possible to obtain consistency when instruments are nearly weak. While, Caner (2009) extend this analysis to the Generalized Method of Moments (GMM) framework following Stock and Wright (2000).

In the same context, the goal of this paper is to propose a model selection technique that will allow the researcher to be able to recognize instruments that may contribute to improving linear models and include them when required. Andrews (1999) is the first to introduce the GMM information criterion $(GMMIC)$ which evaluates the selection vector c representing the set of candidate instruments. His criterion depends on the $J-$ test statistic of over-identifying restrictions based on the moment conditions selected by c with a penalty term function of the same c, the sample size and the number of explanatory variables. As it is raised by Hall and Peixe (2003), because the $GMMIC$ is constructed using the J-test statistic of over-identifying restrictions, it relies solely on the orthogonality property of the instruments without consideration for their relevance. In other words, it tends to select instruments based uniquely on their validity.

In response to this moment selection violation, Hall et al. (2007) propose the Relevant Criterion for Moment Selection (RMSC), a criterion built using the long run canonical correlations and the entropy of the limiting distribution of the GMM estimator. While their findings only briefly discuss how this criterion may be interesting to analyze few scenarios of weak identification in linear IV models, our research will build on their methodology to construct an information criterion robust to weak instruments issues. This chapter has also been motivated by the recent development on the different forms of weak identification and the possibility of efficient estimation and inference in their presence, notably in Antoine and Renault (2009, 2012) and Andrews and Cheng (2012).

The rest of this chapter is organized in five sections where we present the general framework of the study in the first. The second section covers the contents of the relevant literature on existing information criteria. The third section proposes the robust information criterion and the section four confirms our findings in simulations before the concluding remarks in the last section.

2.2 Model Set-up and General Framework

Consider the linear IV model:

$$\begin{cases} Y = X\theta + U \\ X = Z\Pi + V \end{cases} \tag{2.1}$$

With Y the T-vector of realizations of the dependent variable, X the (T, p)-matrix of p explanatory variables, some of which may be endogenous, Z the (T, k)-matrix of instrumental variables; U and V, T-vector and (T, p)-matrix of errors, respectively; θ and Π, p-vector and (k, p)-matrix of parameters, respectively.

In order to introduce weak identification in our model, we follow the standards in the existing literature which consider that the level of weakness does not affect directly the parameter θ but Instead, appears at the instruments level.

To allow for variability in strength of the instruments, we rewrite

$$\Pi = \Lambda_T^{-1} C \equiv \begin{pmatrix} T^{-\delta_1} C_1 \\ T^{-\delta_2} C_2 \end{pmatrix}, \quad \text{with} \quad \Lambda_T = \begin{pmatrix} T^{\delta_1} I_{k_1} & 0 \\ 0 & T^{\delta_2} I_{k_2} \end{pmatrix}$$

for some $0 \leq \delta_1 \leq \delta_2 < 1/2$ and C_j, (k_j, p)-matrix for $j = 1, 2$, $k_1 + k_2 = k$. $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$ partitioned according to Π, ie, Z_j, (T, k_j)-matrix for $j = 1, 2$. Thus we can write the system (2.2) as:

$$\begin{cases} Y = X\theta + U \\ X = Z_1 \frac{C_1}{T^{\delta_1}} + Z_2 \frac{C_2}{T^{\delta_2}} + V \end{cases}$$

When $\delta_1 = \delta_2$, the instruments in Z_1 and Z_2 have equal strength while those in Z_1 are stronger than those in Z_2 if $\delta_1 < \delta_2$.

The estimation of the parameter θ gives rise to some issues when it cannot be uniquely determined. Our paper is concerned by the nearly weak identification case introduced by Hahn and Kuersteiner (2002) in linear models and later Caner (2009) in a non linear context. In particular, we aim to construct a model selection criterion that will be robust to this kind of nearly weakness.

More specifically, our setup completely excludes the *'near non-identified'* case of Hahn and Kuersteiner (2002) (where δ_1 and δ_2 are greater than $1/2$) and the $\sqrt{T}$ weakly identified case in the seminal paper of Staiger and Stock (1997). In other words for our two sets of instruments, the nearly weakness is such that a consistent estimate of θ will converge at a slower rate $\sqrt{T}$. Our framework is also related to the "*'semi-strong'* case defined in Andrews and Cheng (2012) where $\delta_1 = \delta_2 = 0$ coincides with the standard global strong identification case.

In comparison to Hahn and Kuersteiner (2002) who investigate nearly weakness when all instruments possess similar strength, our paper relaxes this condition and accounts for the possibility of nearly weak instruments with different level of weakness.

Let us consider the Two stage least squares estimator $\hat{\theta}$ associated to the model in (2.2):

$$\hat{\theta} = \left(X' P_Z X\right)^{-1} X' P_Z Y, \tag{2.2}$$

with $P_Z = Z(Z'Z)^{-1}Z'$ the usual orthogonal projection matrix.

For consistent estimation of the parameter vector $\hat{\theta}$, we need the following regularity conditions

Assumption 2.2.1. (i) *$\left\{\omega_t \equiv (Y_t, X_t, Z_t) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^k : t = 1, \ldots, T\right\}$ is a sample of identically and independently distributed random vectors with finite second moments.*

(ii) *C is full column rank and*

$$\Delta \equiv \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} = \begin{pmatrix} E\left(Z_{1t} Z'_{1t}\right) & E\left(Z_{1t} Z'_{2t}\right) \\ E\left(Z_{2t} Z'_{1t}\right) & E\left(Z_{2t} Z'_{2t}\right) \end{pmatrix}$$

is nonsingular.

(iii) $E(Z_tU_t) = 0$ *and* $E(Z_tV_t) = 0$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t'U_t \xrightarrow{d} \mathcal{N}(0, \sigma_u^2\Delta), \quad \textit{and} \quad \frac{Z'V}{\sqrt{T}} = O_p(1)$$

where $\sigma_u^2 = E(U_t^2)$.

2.2.1 (i) restricting the sample considered to be independently and identically distributed is standard and merely imposed for simplicity. 2.2.1 (ii) guaranties that instruments are non redundant and the usual rank condition on $E(Z_tX_t')$. Finally, 2.2.1 (iii) assumes exogeneity of Z_t and homoscedasticity of U_t. The latter is for exposition purpose since in the heteroskedastic case, the second moments necessary for the construction of information criterion are heavier to carry out in proofs. The limit properties in 2.2.1 (iii) of Z_t are useful in establishing the asymptotic distribution of the estimator of interest in this paper.

From the expression in (2.2), we write

$$\begin{aligned}\hat{\theta} - \theta &= (X'P_ZX)^{-1}X'P_ZU \\ &= \left[(X'Z)(Z'Z)^{-1}(Z'X)\right]^{-1}\left[(X'Z)(Z'Z)^{-1}(Z'U)\right]\end{aligned}$$

To establish consistency of the 2SLS estimator we need to determine the order of magnitude of the three components that enter the above expression. Notice that,

$$\begin{aligned}T^{-1+\delta_1}X'Z &= T^{-1+\delta_1}\Pi'Z'Z + T^{-1+\delta_1}V'Z \\ &= \begin{pmatrix} C_1' & T^{-\delta_2+\delta_1}C_2' \end{pmatrix}\left(\frac{Z'Z}{T}\right) + T^{-\frac{1}{2}+\delta_1}\left(\frac{V'Z}{\sqrt{T}}\right)\end{aligned} \tag{2.3}$$

where $0 \leq \delta_1 < \delta_2 < 1/2$ implies that the quantity with C_2 vanishes as T grows to infinity, while for $\delta_1 = \delta_2 < 1/2$ we find the same result as in the *'nearly weak'* instrument case of Hahn and Kuersteiner (2002) and C remains full.

From assumptions 2.2.1(ii) and (iii), we conclude that

Lemma 2.2.2. $T^{-1+\delta_1}X'Z \sim C_1'\begin{pmatrix}\Delta_{11} & \Delta_{12}\end{pmatrix}$ *for* $0 \leq \delta_1 < \delta_2 < 1/2$ *and* $T^{-1+\delta_1}X'Z \sim C'\begin{pmatrix}\Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22}\end{pmatrix}$ *for* $\delta_1 = \delta_2 < 1/2$.

Using this lemma we rewrite

$$T^{\frac{1}{2}-\delta_1}(\hat{\theta}-\theta) = \left[\left(T^{-1+\delta_1}X'Z\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(T^{-1+\delta_1}Z'X\right)\right]^{-1}\left[\left(T^{-1+\delta_1}X'Z\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'U}{\sqrt{T}}\right)\right] \equiv O_p(1)$$

This guarantees the consistent estimation of the 2SLS estimator. Hence we write

Proposition 2.2.3. *Under the assumptions 2.2.1*

$$T^{\frac{1}{2}-\delta_1}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \sigma_u^2\left(A\Delta^{-1}A'\right)^{-1}\right)$$

where $A = p\lim\left(T^{-1+\delta_1}X'Z\right) = C_1'\begin{pmatrix} \Delta_{11} & \Delta_{12} \end{pmatrix}$ *and* $\Delta = p\lim\left(\frac{Z'Z}{T}\right)$.

In practice δ_1 is not observed, our goal is to demonstrate that model selection is still possible when we have instruments of various strengths such as in this scenario.

2.3 Performance of Existing Criteria

2.3.1 Existing Selection Methods

Ng (2013) study methods used for the selection of relevant predictors to forecast time series. She identified two main categories of model selection criteria:

- Models with a small number, N, of predictors relative to the sample size T

- Models with a large number, N, of predictors relative to the sample size T.

This categorization is far from exhaustive but necessary to define the specific application of the information criterion proposed in this paper.

Criterion Based Methods

We consider criterion based methods, which usually imply multiple combinations of the predictors with the target to select the candidate model that optimizes the criterion's metric. Mallows criterion,

$\mathbf{C}_p$ is one of the first criteria based on the minimisation of the mean squared error. Under regularity conditions, this criterion is expressed as follows:

$$C_p = \frac{SSR_p}{\hat{\sigma}^2} - T + 2p$$

where SSR_p is the residual sum of squares associated with a subset of p elements of the candidate set, and $\hat{\sigma}^2$ an accurate estimate of the residual mean square after the regression on the complete candidate set. Mallows raised that researchers must be cautious using this criterion as it may not distinguish close competitor models.

Ng (2013) defined this criterion as belonging to a large family of criteria that minimizes an estimator of the mean square error combined with a penalty term:

$$\arg \min_{p=1,\cdots,p_{\max}} \left(\log(\hat{\sigma}_p^2) + p\frac{C_T}{T} \right)$$

where $\hat{\sigma}_p^2$ measures the model's fit when p regressors are considered. p measures the model's complexity and $\frac{C_T}{T}$ is the term used to punish complexity in favor of parsimony.

She also mentioned that the determination of C_T is crucial in the construction of a strong criterion. It is easy to choose a penalty term that goes to fast to infinity preventing it to penalize appropriately the information related part of the criterion. In the literature, the values of C_T mostly encountered are $log(T)$ and 2, which are respectively proportional to BIC criterion type and AIC criterion type. The proposed information criterion in this paper can be classified in the above described family of criteria. As a result, after the construction of our criterion, we investigate in section 2.3.2, the performance in simulations of the AIC, BIC and HQIC to choose the best family of penalty functions.

Ng (2013) also considered other sequential testing procedures that are not covered in this paper. We refer the reader to the original paper for further references and details on the question.

Regularization and Dimension Reduction Methods

The main disadvantage inherent to criterion based methods for selection is the number of permutation possible (2^N) when the candidate set of models is quite large. In particular when the candidate set is as large as the sample size, the potential good fit obtained may only be due to noise. Regularization and dimension reduction techniques are solution to these issues.

In a standard framework, the instrument set Z is of full column rank, but when some instruments are weakly correlated with the regressand (eigenvalues of Z are nearly 0) and the parameter estimate becomes very sensitive. Regularization techniques alleviates the issue by assigning smaller weights to weaker instruments. They are treated as less important , this is termed as shrinkage method.

A general shrinkage method is the bridge regression which resolves the following problem,

$$\hat{\Pi} = \arg\min_{\theta \in \Theta} \|X - Z\Pi\|_2^2 + \gamma \sum_{j=1}^{M+N} |\Pi_j|^{\eta}, \eta > 0$$

γ is the shrinkage parameter. It is fairly easy to notice that when $\eta = 0$ this is equivalent to the least squares regressions context. As an illustration, note that information criterion and Sequential testing procedures are based on L_0 regularization.

When $\eta = 2$, we obtain the ridge estimator with an L_2 penalty, also known as the Tikhonov regularization. As discussed in the relevant literature, L_1 penalisation ($\eta = 1$) not only shrinks coefficients toward zero as L_2 penalization, but set them to zero as well discarding some instruments from the candidate set. This method, commonly termed LASSO is often used for instruments selection. Zou (2006) suggests an Adaptive LASSO procedure by adjusting the weight on the penalty function; this will provide LASSO oracle properties. In continuation of that, Belloni et al. (2012) improves the estimator in terms of rate of convergence and bias by applying OLS post LASSO estimation. While, it is also possible to combine L_1 and L_2 penalization generating an Elastic Net estimator which applies strictly convex penalization, we refer the reader to the related literature for a complete review on LASSO techniques. He may also consider the literature on Forward Stagewise and least absolute regression studies for further details.

In regards to dimension reduction techniques, Principal Component Analysis, factor augmented and later reduced rank and partial least squares regressions can be used for the purpose of model selection. However, they usually imply latent variables or estimated components that generate a dilemma between what to target or not to target: bias or efficiency. As recalled in Ng (2013) the AIC - BIC criteria dilemma comes from the fact that targeting both model consistency and optimal prediction is not straight forward, model selection procedures cannot be both consistent and minimax rate optimal.

However, one could use asymptotic properties of estimators to derive a robust information criterion. This is the main goal of this paper which considers cases when most of the candidate instruments are weaker or all them are weak with mixed strength. Selection using regularization or dimension reduction techniques with mixed identification strength is also of interest for future research.

2.3.2 Classical Information Criteria

To investigate the performance of existing criteria, and illustrate the goal of our research we consider the following data generating process

$$\begin{aligned} Y &= X\theta + U \\ X &= z_1\pi_{1T} + z_2\pi_{2T} + V, \text{ where } \pi_{jT} = \frac{c_j}{T^{\delta_j}};\ j = 1, 2. \end{aligned}$$

This framework is to that of Hall et al. (2007) with the difference that we introduce weakness in the instruments through the weak correlation between the instruments and the endogenous variable. We also allow in our simulation the possibility to consider multiple endogenous regressors with various design of weak correlation.

For simplicity, in the rest of the paper we only consider the case of one endogenous setting $\theta_0 = 0.1$, $c_1 = 1.48$ and $c_2 = 1.48$ and the case of two endogenous regressors with $\theta_0 = (0.1; 0.1)'$, $c_1 = (1.48, 0)'$ and $c_2 = (0; 1.48)'$. This means that only z_1 and z_2 are relevant instruments among all candidate instruments.

We include four extra instruments z_3, z_4, z_5, z_6 orthogonal to each other and z_1, z_2, U and V drawn from the same common distribution $\mathcal{N}(0, I_T)$ with $Cov(U_t, V_t) = \rho$. We then conduct Monte Carlo simulations of all possible combinations of the instruments, 63 (57) for the case of one endogenous variable (two endogenous variables). The sample sizes considered are 100 and 500 for $10,000$ replications.

In our first simulation results presented in the table 13, we report various level of weakness, we consider the cases where $\delta_1 = \delta_2 = \{0, 0.1, 0.2, 0.3, 0.4\}$ to account for situations in which instruments are stronger with $0 \leq \delta < 1/4$ or weaker with $1/4 \leq \delta < 1/2$. Our objective is to evaluate how standard criteria perform when we weaken the instruments by increasing δ. This will give us some evidence on the robustness of those criteria and an opportunity to identify the criteria that perform best among them.

The information criteria that were computed are, the classical information criteria, Akaike (AIC), Schwartz (BIC), Hannan Quinn (HQIC), Andrews (1999) GMMBIC moment selection criterion and the RMSC. The RMSC criterion is a penalized entropy measure that is minimized over candidate models to obtain the most relevant model. Below is the summary of the expressions of those criteria used in our simulations:

1. Akaike (AIC), Schwartz (BIC), Hannan Quinn (HQIC)

$$AIC = \frac{2\ln|\hat{V}|}{T} + \frac{2p}{T}$$

$$BIC = \frac{2\ln|\hat{V}|}{T} + \frac{2p\ln(T)}{T}$$

$$HQIC == \frac{2\ln|\hat{V}|}{T} + \frac{2\ln(\ln(T))}{T}$$

2. Andrews (1999):GMMIC

$$GMMIC = J_T + p\ln(T)$$

3. Donald and Newey (2001): MSE_{DN}

$$MSE_{DN} = H^{-1}\left[\hat{\sigma}_{uv}\hat{\sigma}'_{uv}\frac{k^2}{T} + \hat{\sigma}^2_u\frac{f'\left(I - P^k\right)f}{T}\right]H^{-1}$$

   where $f = P_Z X$, $H = \frac{f'f}{T}$, $\hat{\sigma}_{uv} = \frac{\hat{U}'\hat{V}}{T}$ and $\hat{\sigma}^2_u = \frac{\hat{U}'\hat{U}}{T}$. P^k is the projection matrix with the generalized inverse of $(Z'Z)$. We use in our simulations the standard projection matrix as $(Z'Z)$ is easily invertible in our framework.

4. Hall et al. (2007) : RMSC

$$RMSC = \ln|\hat{V}| + (k-p)\frac{\ln(\sqrt{T})}{\sqrt{T}}$$

In all the formula, $|\cdot|$ represents the determinant function, $\hat{V}$ is the standard estimator of the variance of $\hat{\theta}$ (under homoscedasticity, valid and relevant instruments), p is its dimension, k is the number of instruments in the candidate set, T is the sample size and J_T is the J-test statistic of over identifying restrictions.

Simulations conducted respectively for sample sizes $T = 100$ and $T = 500$, demonstrate that the RMSC outperforms the GMMBIC, the mean square error criterion (MSE-DN) of Donald and Newey (2001) and the standard classical criteria.

We summarize the results of the various criteria for selection of the correct model including exactly z_1 and z_2 in Figure 7. The standard criteria, AIC, BIC and HQIC are represented by the red curve joining red $\times$ signs, the GMMBIC criterion of AndrewsConsistentMomentSelection1999 is represented by the dashed blue curve joining blue square signs,the MSE criterion proposed by Donald and Newey (2001) is represented by the blue curve joining blue diamond signs and the RMSC criterion is represented by the dotted and dashed blue curve joining blue star signs. We observe at $T = 100$, the standard criteria and the GMMBIC fail to select the correct model over the $10,000$ simulation runs while the MSE-DN performs slightly better than them but shrinks toward 0 as δ increases toward 0.4. The RMSC performs strongly at selecting the correct model but start decreasing sharply from $\delta \geq 0.1$. At $T = 500$ the sharp decrease of the RMSC only starts at $\delta \geq 0.2$.

Remark that all the criteria improve when the sample size increases from $T = 100$ to $T = 500$, demonstrating their asymptotic properties.

The detailed simulation results are presented in Tables 13 and 14 (in Appendix 2.7.3). We notice that the AIC, BIC and HQIC returns exactly the same selection frequencies and are never capable of selecting the correct model (the model with exactly z_1 and z_2). When they attempt to identify the optimal number of relevant instruments, those standard criteria retain models containing both relevant instruments and irrelevant instruments. Regardless of when we are evaluating a model including only the strong instrument ($\delta = 0$) or when we add weaker instruments with $0.1 \leq \delta \leq 0.4$, the selection frequencies remains very low for the standard criteria and the increase of the sample size to 500 does not very much improve the results.

Remark that the AIC , BIC and HQIC all have the same information part in their formulae, meaning that their differences in penalty terms are insufficient to identify the most parsimonious model.

Regarding the GMMBIC, the results show that it is ineffective at identifying the best models even when the instruments are all strong. This would be because the J-test statistic is minimal for valid instruments and as the candidate set contains only orthogonal instruments, the criterion constructed using this statistic recommends the inclusion of the largest possible set of candidate orthogonal instruments as they are all valid (non correlated with the error term U). They GMMBIC returns 100%

of the time the full candidate set.

In the case of the MSE criterion proposed by Donald and Newey (2001), the results recommend at $T = 100$, $\delta \leq 0.2$ the selection of the models including solely one of the relevant instruments with similar probability or the model including only the irrelevant instruments when $\delta = 0.4$. Even when we increase the sample size to $T = 500$, the proportions are slightly increased for the models with only one relevant instrument and slightly decreased for the only irrelevant case.

Finally, in the case of the RMSC which performs better than all other criteria discussed above, for smaller sample size at $T = 100$ and larger sample $T = 500$, we notice that it is sensitive to the inclusion of strong instruments in the candidate set. For example the RMSC reports 99.63% (100%) of the time the model including exactly z_1 and z_2, when $\delta = 0$ (instruments are strong) for $T = 100$ ($T = 500$). This performance changes as we weaken the instruments through a decrease of the strength δ of the instruments. In particular, for $\delta = 0.1$, $RMSC$ looses 6% of its accuracy at $T = 100$ and when $\delta \geq 0.2$, the RMSC identifies the correct model,only 67.42% and then 28.54% of the time versus 97.14% and then 47.24% at $T = 500$.

This supports the results in the weak instruments literature which suggest that the closer are the instruments' weakness to the $\sqrt{T}$ level (δ close to 0.5) the worse are the estimators properties. As a matter of fact when the instruments are weak, all the criteria hardly distinguish models with irrelevant instruments versus those with relevant instruments. This suggests that they make no difference between weak and irrelevant instruments at this point.

These simulations indicate that the existing criteria would improve if it is possible to treat instruments' weakness conveniently. They underline the concerns raised in the literature about the importance of the level of weakness in the selection of most parsimonious models. In the next section, we will investigate what factors reduce the selection performance of the RMSC in the instrument selection context.

2.4 A Robust Model Selection Procedure

2.4.1 The Selection Criterion

In this section we discuss the construction of the information criteria for model selection in the standard framework in order to propose an information criterion more robust to weak identification as defined earlier. To better understand how the weakness of the correlation between instruments and the endogenous regressors affect model selection, we reviewed in the previous section the performance of classical information criteria. As we noticed that the RMSC performs best, we consider it as starting point for the construction of a robust criterion

Recall that model selection is alternatively a function of the asymptotic variance of the regression's estimator or of an approximation of the mean squared error of regressions. In this context, we consider the asymptotic distribution of our estimator and we need to extract the asymptotic variance of $\hat{\theta}$ obtained in theorem (2.2).

As it is done in the standard case, the researcher may naively consider the following estimator of the asymptotic variance of $\hat{\theta}$

$$\hat{V}_T = \hat{\sigma}_u^2 \left(\frac{X'P_Z X}{T}\right)^{-1} \quad \text{where} \quad \hat{\sigma}_u^2 = (Y - X\hat{\theta})'(Y - X\hat{\theta})/T. \tag{2.4}$$

The only issue is that this would not converge towards the true variance since in our weak instruments context, $X'Z$ is found to be of $O_p(T^{\delta_1 - 1})$.As a result a consistent estimate of the asymptotic variance of $\hat{\theta}$ will write

$$\begin{aligned} \tilde{V}_T &= \hat{\sigma}_u^2 \left[\left(T^{-1+\delta_1} X'Z\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(T^{-1+\delta_1} Z'X\right)\right]^{-1} \\ &= T^{-2\delta_1}\hat{V}_T \end{aligned} \tag{2.5}$$

Now, following Hall et al. (2007) who introduced the RMSC using the entropy of the limiting distribution of $\hat{\theta}$ to measure the information carried by the estimator, we exploit the result of Ahmed and Gokhale (1989) to derive the entropy associated to the approximated asymptotic distribution in equation (2.5)

$$\begin{aligned} ent_\theta &= \frac{1}{2}p(1 + \ln(2\pi)) + \frac{1}{2}\ln|\tilde{V}_T| \\ &= \frac{1}{2}p(1 + \ln(2\pi)) - p\delta_1 \ln T + \frac{1}{2}\ln|\hat{V}_T|. \end{aligned} \tag{2.6}$$

The above entropy is slightly different from the one obtained by Hall et al. (2007) Its second term is not constant and may change from model to model depending on the set of instruments selected. Therefore a naive use of $\hat{V}_T$ as estimate of the asymptotic variance in the construction of a model selection criterion is misleading. In the situation when $\delta_1 = \delta_2 = 0$, meaning that all the instruments are strong, the information captured by the entropy above will be the same as that considered in the RMSC. It would also be the case if there is a large number of strong instruments with $\delta_1 = 0$, as the weaker instruments will be discarded, the information content in the entropy would not change.

On the contrary when the candidate set contains, in large majority, weak instruments with $0 < \delta_1 < \delta_2 < 1/2$, the second term in the entropy fails to vanish and would contribute to the fluctuation in the information measure.

Dividing the entropy expression in equation (2.6) by $\ln(T)$, we obtain

$$\frac{ent_\theta}{\ln(T)} = \frac{p(1 + \ln(2\pi))}{\ln(T)} - 2p\delta_1 + \frac{\ln|\hat{V}_T|}{\ln(T)}.$$

The information-related part of the entropy above can therefore effectively be considered as:

$$\frac{\ln|\hat{V}_T|}{\ln(T)} - p\delta_1 \tag{2.7}$$

In practice δ_1 is not observed and it has not been proposed in the literature any approach to estimate its value. Therefore, it remains an area for future research. In the meantime, we retain the information content of the entropy in equation (2.6) and propose a penalty term that will adequately penalize the information criterion to rank optimally models in presence of weak instruments.

We omit $p\delta_1$, and adjust the level of the penalty function of the criterion. The resulting family of information criteria for model selection that we label Adjusted Relevant Moment Selection Criterion is given by

$$aRMSC = \ln\left|\hat{V}_T\right| + \kappa_T \ln T \tag{2.8}$$

where κ_T is the usual penalty term and $\hat{V}_T$ is the standard estimate of the variance of the 2SLS estimator $\hat{\theta}$.

The main difference between the proposed criterion and that of Hall et al. (2007) is that the RMSC accounts only for cases where all directions of the parameter space are estimated at the standard $\sqrt{T}$ rate. In other words, when all instruments are strong. RMSC type information criteria would be less efficient at identifying weaker instruments even when they are pertinent to the model. Indeed, in that case $\delta_1 \neq 0$ and the RMSC's information content consisted of the entropy in equation (2.6) would asymptotically tend to infinity pushing the criterion to diverge. This is well illustrated in the Section

2.2 as we weakened the relevant instruments, the RMSC which offered the best performance among its peers diverged progressively.

Our family of information criterion accounts for the fact that consistent estimation is possible at a slower rate than the standard rate $\sqrt{T}$ by correcting the RMSC's information related term, by $p\delta_1 \ln(T)$ to bring it to a level where, combined with the usual penalty term, it will make the criterion consistent. The consistency of the information criterion and the choice of the penalty term $\kappa_T \ln(T)$ will be discussed in the next section.

2.4.2 Convergence of the criterion

Following the notation of Andrews (1999) and Hall et al. (2007), we consider a selection vector c of dimension $k_{\max} \times 1$ ($k_{\max}$ being the maximal number of instruments in the candidate set) for which the elements take the values of 0 or 1. At row i , $c_i = 1$ means that the instrument z_i is selected and $c_i = 0$ implies that the instrument is not included in the model. We regroup all the possible combinations of instruments in candidate set $\mathcal{C}$. Remark that each vector c in $\mathcal{C}$ always has the same dimension and determines the number of instruments included in the model with $|c| = c'c \geq p$.

For exposition purpose, the statistics of interest are now indexed by c and so $\hat{\theta}(c)$ denotes the IV estimator resulting from the selection vector $c; V_\theta(c)$ is the asymptotic variance of its limiting distribution.

To avoid missing *relevant* instruments in the current framework, the researcher may consider including all the $k_{\max}$ instruments of the candidate set to reach asymptotic efficiency. However, it has been reported Hall and Peixe (2003) in their simulations results that redundant instruments have negative effect on inference. Similar negative effect on finite sample bias has been established by Newey and Smith (2004). Consequently, Hall et al. (2007) formally defined relevant instruments to be equivalent to the minimal set of instruments necessary to achieve the same asymptotic efficiency as all the $k_{\max}$ instruments. We maintain below the same definition as Hall et al. (2007)

Definition 2.4.1. *c_r is the selection vector associated with the relevant instruments if the following properties hold:*

(i) *$c_r \in \mathcal{C}$;*

(ii) *$V_\theta(\iota_{k_{\max}}) = V_\theta(c_r)$ where $\iota_{k_{\max}}$ is a $k_{\max} \times 1$ vector of ones;*

(iii) *$V_\theta(c_{r,1}) - V_\theta(c_r)$ is positive semi definite for $c_r = c_{r,1} + c_{r,2}$ and $c_{r,1}, c_{r,2} \in \mathcal{C}$.*

Part (ii) of this definition states that the model accounting for the maximum number of instruments attain the asymptotic efficiency as the model including only the relevant set of instruments. In other words, additional instruments to those selected by the vector c_r, are redundant and do not reduce the estimator's asymptotic variance. While part (iii) implies that no subset of the relevant instruments can achieve smaller variance than the model based on c_r. Indeed, if $c_{r,1}$ was able to generate a smaller asymptotic variance of $\hat{\theta}$, we would conclude with probability equal to one that c_r contains redundant instruments and cannot be considered as a set of relevant instruments.

To determine c_r, we rely on the information criterion introduced in (2.8) with a penalization term κ_T , a function of sample size and the size of the estimating function. Note that parsimony is sought relatively to the number of moment restriction and not the number of parameter estimates which is always p. Specifically, we write

$$aRMSC(c) = \ln\left|\hat{V}_T(\theta(c))\right| + \kappa(|c|, T). \ln T \tag{2.9}$$

Therefore, to estimate c_r, the researcher will minimize the information criterion over $\mathcal{C}$:

$$\hat{c}_T = \arg\min_{c\in\mathcal{C}} aRMSC(c)$$

Similar to Hall et al. (2007), let

$$\mathsf{C}_{eff} = \{c; c \in \mathcal{C} : V_\theta(c) = V_\theta(\iota_{k_{\max}})\}$$

and

$$\mathsf{C}_{\min} = \left\{c; c \in \mathsf{C}_{eff} : |c| \leq |\bar{c}| \text{ for all} \bar{c} \in \mathsf{C}_{eff}\right\}$$

We impose the following assumptions to characterize the set of selection vectors.

Assumption 2.4.2. **(i)** *c_r satisfies definition 2.4.1 and $\mathsf{C}_{\min} = \{c_r\}$;*

(ii) *$\tilde{V}_T\left(\theta(c)\right) = V\left(\theta(c)\right) + O_p(\tau_T^{-1})$ where $\tau_T \to \infty$ as $T \to \infty$*

(iii) *For any $\{\bar{c}, \tilde{c}\} \in \mathcal{C}$, such that $|\bar{c}| > |\tilde{c}|$, $\tau_T. \ln(T)\left(\kappa(|\bar{c}|, T) - \kappa(|\tilde{c}|, T)\right) \to +\infty$ as $T \to +\infty$, and $\kappa(|c|, T) = o(1)$ for every $c \in \mathcal{C}$.*

This is similar to Assumption 4 of Hall et al. (2007) Part (i) guaranties the identification of c_r. Part (ii) follows from their choice of τ_T , the rate of convergence of the estimate of the asymptotic variance toward the true variance $V(\theta(c))$. This assumption will turn out to be crucial in helping to choose an appropriate level of penalization to constrain the information part of the criterion. Finally, part (iii) defines the strictly increasing monotonicity of $\kappa(|\cdot|, T)$ in c, required for the convergence of the criterion.

For the range of models considered in this paper, let c_i be the vector that selects at least one redundant or irrelevant instrument and no relevant instrument. From definition 2.4.1, any addition of instruments in $\mathcal{C}$ to c_r cannot reduce further the estimated asymptotic variance. This means that at most

$$\tilde{V}_T(\theta(c_r + c_i)) - \tilde{V}_T(\theta(c_r)) = M \tag{2.10}$$

where M is a positive semi definite matrix.

This implication turns out to be essential in establishing the consistency of our criterion. Since the criterion is a function of the asymptotic variance, its convergence depends on its ability to recognize the most parsimonious model, the model constructed using the set of instruments selected by c_r.

Theorem 2.4.3. *Under assumptions (2.4.2) (i)-(iii), $\hat{c}_T$ converges in probability to c_r as $T \to \infty$ and*

$$\lim_{T \to +\infty} \mathbb{P}\left\{[aRMSC(c_r + c_i) - aRMSC(c_r)] \geq 0\right\} = 1$$

2.4.3 Choice of Penalty Function

Hall et al. (2007) consider the Schwartz (BIC type) and Hannan Quinn (HQIC type) penalty functions given below

$$\kappa_{BIC}(|c|, T) = (|c| - p)\frac{\ln(\ell_T)}{\ell_T} \quad \text{and} \quad \kappa_{HQIC}(|c|, T) = (|c| - p)\frac{Q\ln(\ln(\ell_T))}{\ell_T} \tag{2.11}$$

with $Q > 2$ and $\ell_T : \ell_T(\hat{V}_T - V) = O_P(1)$. In other words, ℓ_T is the rate of convergence of their estimate of the asymptotic variance $\hat{V}_T$.

This implies that the natural analogous to ℓ_T in our framework would be τ_T the rate of convergence of the consistent estimator $\tilde{V}_T(\theta(c))$ of $V(\theta(c))$. As a result our first step in choosing a penalty term will be to determine the rate of convergence of the asymptotic variance used in such a way that

$\kappa(|\cdot|, T)$ satisfies Assumptions 2.4.2 (iii).

Recall the consistent estimator of the asymptotic variance in equation (2.5)

$$\tilde{V}_T\left(\theta(c)\right) = T^{-2\delta_1}\hat{\sigma}_u^2\left(\frac{X'P_ZX}{T}\right)^{-1}$$

We now formalize its rate of convergence in the following proposition.

Proposition 2.4.4. *Assume assumptions 2.2.1 hold, we determine the rate of convergence of the estimator of the consistent variance of θ given a candidate set c of instruments included in the model*

$$\tilde{V}_T\left(\theta(c)\right) - V\left(\theta(c)\right) = O_P(T^{\delta_1 - \frac{1}{2}})$$

Proof. Proof in Appendix 2.7.1 . □

This result allows us to confirm that, in our framework, the rate of convergence of the consistent estimator of the asymptotic variance depends on the strength of instruments included in the model under consideration. In particular in the case of mixed strength, this rate of convergence depends on the level of weakness (measured by δ_1) of strongest set of instruments regardless of the weakness of the other instruments. In practice, it suffices that the set of stronger instruments contains enough element to allow full identification of the estimator of θ.

The main implication of this result is that it is not obvious to choose an optimal penalty function for our information criterion without estimating δ_1. While the estimation of δ_1 is left for future research, our study is definitely a first step in demonstrating that model selection is possible under mixed identification strength. We confirm these findings by Monte Carlo simulations in the next section and revisit those of Andrews (1999) and Hall and Peixe (2003) who report that the BIC-type penalty function performs better than the AIC and HQIC type penalty functions.

As a result from the simulations in Section 2.5, we consider the BIC type penalty as recommended by Hall et al. (2007) but investigate the performance of the AIC and HQIC type penalties. Our penalty function, using the BIC type penalty writes

$$\ln T \times \kappa(|c|, T) = \ln T \times \left(1 - \frac{p}{|c|}\right)\frac{\ln(\sqrt{T})}{\sqrt{T}}$$

where our choice of positive strictly increasing function of $|c|$ for all values of p is $\left(1 - \frac{p}{|c|}\right)$ differs from that of Hall et al. (2007) to prevent the penalty from diverging to infinity too fast.

The resulting alternative criterion would be

$$aRMSC(c) = \ln \left| \hat{\sigma}_u^2 \left(\frac{X' P_Z X}{T} \right)^{-1} \right| + \ln T \times \left(1 - \frac{p}{|c|} \right) \times \frac{\ln(\sqrt{T})}{\sqrt{T}} \tag{2.12}$$

where $\hat{\sigma}_u^2 = (Y - X\hat{\theta})'(Y - X\hat{\theta})/T$.

Remark that choosing the optimal penalty term is a non-trivial question and is left for future research. We refer the reader to a companion paper by Dovonon et al. (2020) for an approach of selection within a family of robust penalty functions independent of the instruments strength.

2.5 Simulations

In this section, we carry out Monte Carlo simulations to study the finite sample properties of the proposed criterion (aRMSC). In this framework, we consider the 2SLS version of the criterion as well as the Limited Information Maximum Likelihood (LIML) version of the criterion. Our goal being to explore the proposed criterion's performance when the researcher switch to k-class estimators as they are well known to be more robust in presence of weak identification.

To give some background on the k-class estimators, we follow the set-up of Wang and Doko Tchatoka (2018) in order to define the k-class estimator associated to the linear regression in the equation (2.2)

$$\hat{\theta}(\kappa) = \left[X' \left(P_Z - \frac{\kappa}{T - |c|} M_Z \right) X \right]^{-1} \left[X' \left(P_Z - \frac{\kappa}{T - |c|} M_Z \right) Y \right]$$

where for $\kappa = 0$ we obtain the 2SLS estimator. For the LIML estimator $\kappa = \kappa_{LIML}$ where κ_{LIML} is the smallest root of the determinant $\left| \frac{\kappa}{T-|c|} \widetilde{Y}' M_Z \widetilde{Y} - \widetilde{Y}' P_Z \widetilde{Y} \right| = 0$ with $\widetilde{Y} = [Y \vdots X]$ and $M_Z = I_T - P_Z$ and $P_Z = Z(Z'Z)^{-1}Z'$.

Again, we keep the same set-up as in Section 2.3.2, with the cases of one or two endogenous regressors combine with six candidate instruments. For the purpose of comparison results are shown for both aRMSC and RMSC. Notice that, by aRMSC we mean the aRMSC with BIC penalty function as a result from the comparison with the AIC and HQIC type penalty functions. The performances of the three types of penalty term will be discussed later in this section.

Figures 8 - 11 highlight the simulation results. We report for the 2SLS estimator, Figures 8 and 9 respectively for the case of one endogenous regressor $(p = 1)$ and two endogenous regressors $(p = 2)$, for various level of instruments strength measured by δ_1 and δ_2. In particular, we distinguish two cases; situations where the instruments' weakness varies while keeping them at the same strength with $'\delta_1 = \delta_2\,'$ and situations in which one set of instruments is weaker than the other with $'\delta_1 < \delta_2 = 0.4'$. The same simulations are done for the LIML estimator and reported in Figures 10 and 11 respectively.

Notice that for more details on the empirical selection probabilities and comparisons between the BIC type penalty function and the AIC and HQIC type penalty functions, we report them in Tables 15-26 in Appendix 2.7.3. They are the tables combined to obtain the hit rates plotted in Figures 8 - 11. The tables present the selection probabilities for the 2SLS and LIML estimators side by side for easy comparison, grouping together cases where $'\delta_1 = \delta_2'$ and cases where $'\delta_1 < \delta_2 = 0.4'$. Each table therefore reports the results for one sample size.

All the figures highlight the performance of the RMSC and the aRMSC from small to large samples, specifically for the sample sizes $T = \{100, 500, 1000, 10000, 50000, 100000\}$. In particular, each figure contains 8 sub figures of plots of the growth of empirical selection frequencies (hit rate) by sample size for the RMSC (represented by the dotted and dashed blue curve joining blue circle signs) and the aRMSC (represented by the red curve joining blue pentagram signs).

In the Figure 8 of both criteria based on 2SLS estimator and one endogenous regressor, for cases when $'\delta_1 = \delta_2 = 0'$ and $'\delta_1 = \delta_2 = 0.1'$, we observe that the RMSC and the aRMSC perform exactly similarly with hit rates reaching almost 100% as early as $T = 500$. These observations are confirmed in Table 16 at $T = 500$, where we respectively report in the column $'z_1 + z_2\,'$ for both rows of $'\delta_1 = \delta_2 = 0'$ and $'\delta_1 = \delta_2 = 0.1'$,99.75% and 92.21% for the aRMSC and 100% for the RMSC; while at $T = 100$ in Table 15 we reported 55.44% and 39.47% versus 99.63% and 93.78%. This clearly demonstrates an improvement of the both criteria as the sample size increases in these situations. The same result is observed in larger samples.

Instead, when $'\delta_1 = \delta_2 = 0.2'$, the RMSC outperforms the aRMSC in small sample sizes with the blue curve dominating the red curve until $T = 10,000$, where both criteria's curves converge towards 100% of selection of the correct model including exactly z_1 and z_2 (see Table 18 for further details). This convergence is confirmed at $T = 50,000$ and $T = 100,000$ (Tables 19 and 20).

When $'\delta_1 = \delta_2 = 0.3'$, we notice again the domination of the blue curve over the red curve until $T = 10,000$ where the trend is reversed with the aRMSC dominating the RMSC with an increasing gap as T increases. Finally, the case of $'\delta_1 = \delta_2 = 0.4'$ is not presented in the sub figures because both criteria deteriorates with an advantage to the aRMSC which exhibit higher frequencies than the RMSC when the model includes at least one relevant instrument. Even in large samples with $T = 100,000$, the aRMSC selects the correct model only 4.44% of the time versus 2.1% for the RMSC (see details in Table 20 in Appendix 2.7.3). We understand that the instruments are too weak in this context to carry enough informative content for the model. Yet, the aRMSC outperforms the RMSC in this case.

When we consider the LIML estimator in the same conditions, in Figure 10, we observe similar behaviour for $'\delta_1 = \delta_2 = 0, 0.1, 0.2'$. However, for $'\delta_1 = \delta_2 = 0.3'$ the RMSC is boosted and is not any more outperformed by the aRMSC and both criteria now converge in large samples. In Table 20, for the LIML estimator when $'\delta_1 = \delta_2 = 0.3'$, the RMSC selects the correct model 87.04% of the time versus 87.17% for the aRMSC. When $'\delta_1 = \delta_2 = 0.4'$, the selection proportions are now 8.65% for the RMSC against 9.4%. Interestingly the aRMSC improves on the RMSC as the instruments are weakened.

In Figure 9, the case of two endogenous regressors for the 2SLS estimator, when $'\delta_1 = \delta_2 = 0$, $0.1'$ the RMSC and the aRMSC select the correct model 100% of the time even in small samples. While when $'\delta_1 = \delta_2 = 0.3'$ the aRMSC clearly dominates the RMSC and reaches almost 100% of selection of the model with z_1 and z_2 from $T = 10,000$. Indeed, considering the results in Table 24, when $'\delta_1 = \delta_2 = 0.3'$, the aRMSC selects the correct model 97% time while de RMSC only reports 51%. When $'\delta_1 = \delta_2 = 0.4'$, we notice that both criteria perform poorly indicating again that the instruments are too weak. Yet, the red curve representing the aRMSC still dominates the blue curve representing the RMSC and their gap shrinks as T increases to $T = 100,000$. In Figure 11, we notice that the LIML improves again the results of both criteria.

In the case of mixed instruments' strengths, our simulations in the case of one endogenous variable (Figures 8 and 10), when $'\delta_1 = 0, 0.1 < \delta_2 = 0.4'$, the RMSC and the aRMSC perform similarly selecting the correct model with probability one from quite small samples. When $'\delta_1 = 0.2 < \delta_2 = 0.4'$, the RMSC does not perfectly select the correct model (i.e. the columns $'z_1\,'$, the relevant model) even in large samples as opposed to the aRMSC which reaches 99% selection frequency from $T = 1000$. When $'\delta_1 = 0.3 < \delta_2 = 0.4'$, the aRMSC curve remains dominant but performs below 50%. Indeed, at $T = 1000$ in Table 23 the aRMSC reports 35.42% versus 8.2% for the RMSC.

Considering the case when the criteria are based on the LIML estimator, in the same mixed instruments' strength case, the results are similar to the 2SLS case in terms of aRMSC dominance. While it only slightly improves both criteria in the case of one endogenous regressor (see Figures 8 and 10), its impact in the case of two endogenous regressors (Figures 9 and 11) is marginally larger.

The LIML therefore improves the RMSC bridging the gap between it and the aRMSC in the 2SLS case.

On the other side, it is important to raise that in some cases of Figures 8 - 11, in particular in mixed instruments' strength cases with $'\delta_1 < \delta_2'$, the hit rates do not grow with the sample sizes. This indicates that although we outperform the RMSC in these cases, we struggle to consistently converge toward 100%. This results from our choice of penalization function which is not optimal yet. As discussed in the section 2.4.3, this is not a trivial problem and is left for future research.

Finally, regarding the choice of the penalty functions, we notice in our simulations results of Tables 15 - 26 that in the case of one endogenous regressor, the aRMSCs with AIC and HQIC type penalty functions dominate the one with BIC type penalty function in smaller samples. For example, at $T = 100, 500, 1000$ for $'\delta_1 = \delta_2 = 0, 0.1'$ the aRMSC(HQIC) beats the aRMSC(BIC) for the 2SLS estimator. For the LIML estimator, we observe the same behaviour until $'\delta_1 = \delta_2 = 0, 0.1, 0.2'$ highlighting the improvement when passing from the 2SLS to LIML. As $T = 10,000$ to $T = 100,000$ either the aRMSC(BIC) performs similarly or outperforms the aRMSC(HQIC). Also when $'\delta_1 = \delta_2'$ increases until 0.4 or $'\delta_1 < \delta_2 = 0.4'$ we observe the same behaviour at all sample sizes. The later illustrate that the aRMSC(BIC) consistently selects best the correct model as the sample size increases or the instruments are weaker with $\delta_i (i = 1, 2)$ increasing towards 0.4. All these observations are confirmed in the case of the LIML estimator and unanimously in the case of two endogenous regressors.

Overall, our simulation exercise illustrates that the Adjusted Relevant Moment Selection Criterion performs well in small to large values of δ_i $(i = 1, 2)$, while the RMSC fails to handle these cases, as per its decreasing hit rate as the sample size increases for higher values δ_i. Moreover, our result tables show that in comparison to the aRMSC with AIC and HQIC type penalty functions, the aRMSC with the BIC type penalty term performs best, confirming the observations of Andrews (1999) and Hall et al. (2007).

2.6 Conclusion

In this paper, we investigate the problem of model selection in presence of weak instruments. We consider cases where within the set of candidate instruments, there are instruments of the similar strength and cases in which the instruments with different strengths are included in the model. In this framework, we propose an information criterion,the aRMSC that proves to be robust to weak identification situations. In deriving this information criterion, we first evaluate the performance of the existing criteria and realize that the RMSC of Hall et al. (2007) is best at selecting relevant instruments. However, the RMSC does not encompass the type of weak identification presented in this paper. Indeed, in a linear regression instrumental variable model, the RMSC turns out to be very sensitive to the increasing instruments' weakness.

Using the entropy of the 2SLS estimator, we obtain that its limiting distribution depends on the strength of the IVs and the number of endogenous regressors. We then constructed, our robust criterion to this type of weakness and tested our results in Monte Carlo simulations. Overall, 5 results stand out from our study:

- Among groups of instruments with similar strength, the aRMSC is more robust at selecting models with weaker instruments than the RMSC with selection probabilities consistently converging towards one as sample sizes increase;
- When we consider instruments with mixed strength, the aRMSC beats the RMSC again but its value does not pass the 50% selection frequency threshold even in large samples. This highlights that the criterion could be improved to converge toward selection probability equal to one;
- Added to the fact that the criterion is flexible and may benefit from new development in terms of efficient estimators, we show that the aRMSC with the BIC type penalty term exhibit better consistency properties than the criteria with the AIC and HQIC type penalty functions;
- The criteria performances often decrease after $\delta_i = 0.3$ $(i = 1, 2)$,questioning the hard threshold of $\delta_i = \frac{1}{4}$ used to define nearly strong identification versus nearly weak identification in the existing literature Andrews and Cheng (2012)
- LIML which is usually attractive in presence of a large number of instruments, still improves on the 2SLS estimator in presence of a finite number of instruments even if marginally.

2.7 Appendix

2.7.1 Appendix A: Proofs

Proof of Theorem 2.4.3: (Convergence of the criterion)

We define $\Delta_T(c, c_r) = aRMSC(c) - aRMSC(c_r)$.Assuming that assumptions 2.4.2 and the definition 2.4.1 of relevant instruments are satisfied, there are only two possibilities for consistent estimation:

(i) the candidate model includes all the relevant instruments in such a way that the asymptotic variance reaches its minimum regarless of the inclusion of additional irrelevant variables with

$$V(\theta(c_r + c_i)) = V(\theta(c_r))$$

In this case, we write

$$\begin{aligned}\Delta_T(c_r + c_i, c_r) = &\left[\ln\left|\tilde{V}_T(\theta(c_r + c_i))\right| - \ln|V(\theta(c_r + c_i))|\right] - \left[\ln\left|\tilde{V}_T(\theta(c_r))\right| - \ln|V(\theta(c_r))|\right] \\ &+ \ln T \times [\kappa(|c_r + c_i|, T) - \kappa(|c_r|, T)]\end{aligned}$$

Therefore, by assumption 2.4.2 (ii),

$$\tau_T \Delta_T(c_r + c_i, c_r) = O_p(1) + \tau_T \ln T \times [\kappa(|c_r + c_i|, T) - \kappa(|c_r|, T)] \qquad (2.13)$$

and by assumption 2.4.2 (iii) as $\kappa(\cdot, T)$ be a strictly increasing function in c, we have

$$\lim_{T\to\infty} \tau_T \ln T \times [\kappa(|c_r + c_i|, T) - \kappa(|c_r|, T)] = +\infty \qquad (2.14)$$

and the expression in (2.13) is positive with probability one as $T \to \infty$.

(ii) the candidate model includes only a subset of the relevant instruments and consequently does not reach the minimum variance with

$$V(\theta(c)) - V(\theta(c_r)) = M$$

where M is positive semi definite and different from 0.

This implies that the difference between the two variance-covariance matrices is positive semi definite and by Magnus and Neudecker (1999)Theorem 22, , we have in this case,

$$|V(\theta(c))| > |V(\theta(c_r))|$$

Applying the logarithm neperian function to the expression we obtain

$$\ln|V(\theta(c))| - \ln|V(\theta(c_r))| \geq 0$$

and $\tilde{V}_T(\theta(c)) \xrightarrow{p} V(\theta(c))$ and $\tilde{V}_T(\theta(c_r)) \xrightarrow{p} V(\theta(c_r))$ with both finite limits, we have

$$\begin{aligned}\Delta_T(c, c_r) &= \ln\left|\tilde{V}_T(\theta(c))\right| - \ln\left|\tilde{V}_T(\theta(c_r))\right| + \ln T \times [\kappa(|c|, T) - \kappa(|c_r|, T)] \\ &= \ln|V(\theta(c))| - \ln|V(\theta(c_r))| + o_p(1)\end{aligned}$$

As a result, in this case $\Delta_T(c, c_r)$ is positive with probability one as $T \to \infty$.

Taken together case (i) and (ii) lead to $\hat{c}_T \xrightarrow{p} c_r$ as $T \to \infty$. $\square$

Proof of Theorem 2.4.4: (Order of magnitude of $\tilde{V}_T$)

$$\tilde{V}_T(\theta(c)) = T^{-2\delta_1}\hat{\sigma}_u^2\left(\frac{X'P_ZX}{T}\right)^{-1} \tag{2.15}$$

To obtain its order of magnitude we need to find the rate of convergence of $\hat{\sigma}_u^2$.

$$\begin{aligned}\hat{\sigma}_u^2 &= \frac{(Y - X\hat{\theta})'(Y - X\hat{\theta})}{T} \\ &= \frac{U'U}{T} + 2\left(\hat{\theta} - \theta\right)\frac{X'U}{T} + \left(\hat{\theta} - \theta\right)'\left(\frac{X'X}{T}\right)\left(\hat{\theta} - \theta\right)\end{aligned}$$

We use as key input,

$$\frac{X'U}{T} \equiv T^{-1/2}C'\begin{pmatrix} T^{-\delta_1}I_{k_1} & 0 \\ 0 & T^{-\delta_2}I_{k_2}\end{pmatrix}(T^{-1/2}Z'U) + T^{-1}V'U$$

where by assumption 2.2.1 (iii) $T^{-1/2}Z'U = O_P(1)$ as well as $T^{-1}V'U = O_P(1)$.

Hence,

$$\frac{X'U}{T} = o_p(1) \times O_P(1) + O_P(1) = O_P(1)$$

Using the fact that $\sigma^2 = \frac{U'U}{T}$ and that the rate of convergence of $\hat{\theta}$ towards θ is $(\frac{1}{2} - \delta_1)$, we deduce that

$$\hat{\sigma}_u^2 - \sigma^2 = O_P(T^{\delta_1 - 1/2})$$

Now the order of magnitude of the last term $\left(\frac{X'P_Z X}{T}\right)^{-1}$ can be determined by noting that

$$\begin{aligned}
T^{-1+\delta_1} Z'X &= T^{-1+\delta_1} \Pi' Z'Z + T^{-1+\delta_1} V'Z \\
&= \begin{pmatrix} C_1' & T^{-\delta_2+\delta_1} C_2' \end{pmatrix} \left(\frac{Z'Z}{T}\right) + T^{-\frac{1}{2}+\delta_1} \left(\frac{V'Z}{\sqrt{T}}\right) = O_P(1)
\end{aligned}$$

because $\delta_1 \leq \delta_2 < \frac{1}{2}$ the first term of the equation above is $O_P(1)$ and $\left(\frac{V'Z}{\sqrt{T}}\right)$ as well is $O_P(1)$ by by assumption 2.2.1 (iii).

Combining the above results we obtain,

$$\begin{aligned}
\left(\frac{X'P_Z X}{T}\right)^{-1} &= \left[\left(T^{-1} X'Z\right) \left(\frac{Z'Z}{T}\right)^{-1} \left(T^{-1} Z'X\right)\right]^{-1} \\
&= \left[O_P(T^{-\delta_1}) \times O_P(1) \times O_P(T^{-\delta_1})\right]^{-1} \\
&= O_P(T^{2\delta_1})
\end{aligned}$$

We can now conclude that the rate of convergence of our consistent estimator of the variance, $\tilde{V}_T(c)$, is as following

$$\tilde{V}_T(c) - V(c) = O_P(T^{\delta_1 - 1/2})$$

□

2.7.2 Appendix B: Figures

Figure 7: Proportion of best model selection (Hit rate) by AIC, BIC, HQIC, GMMBIC, MSE-DN and RMSC for models with one endogenous variable. Sample size $T = 100$; 500. Number of replications: 10,000.

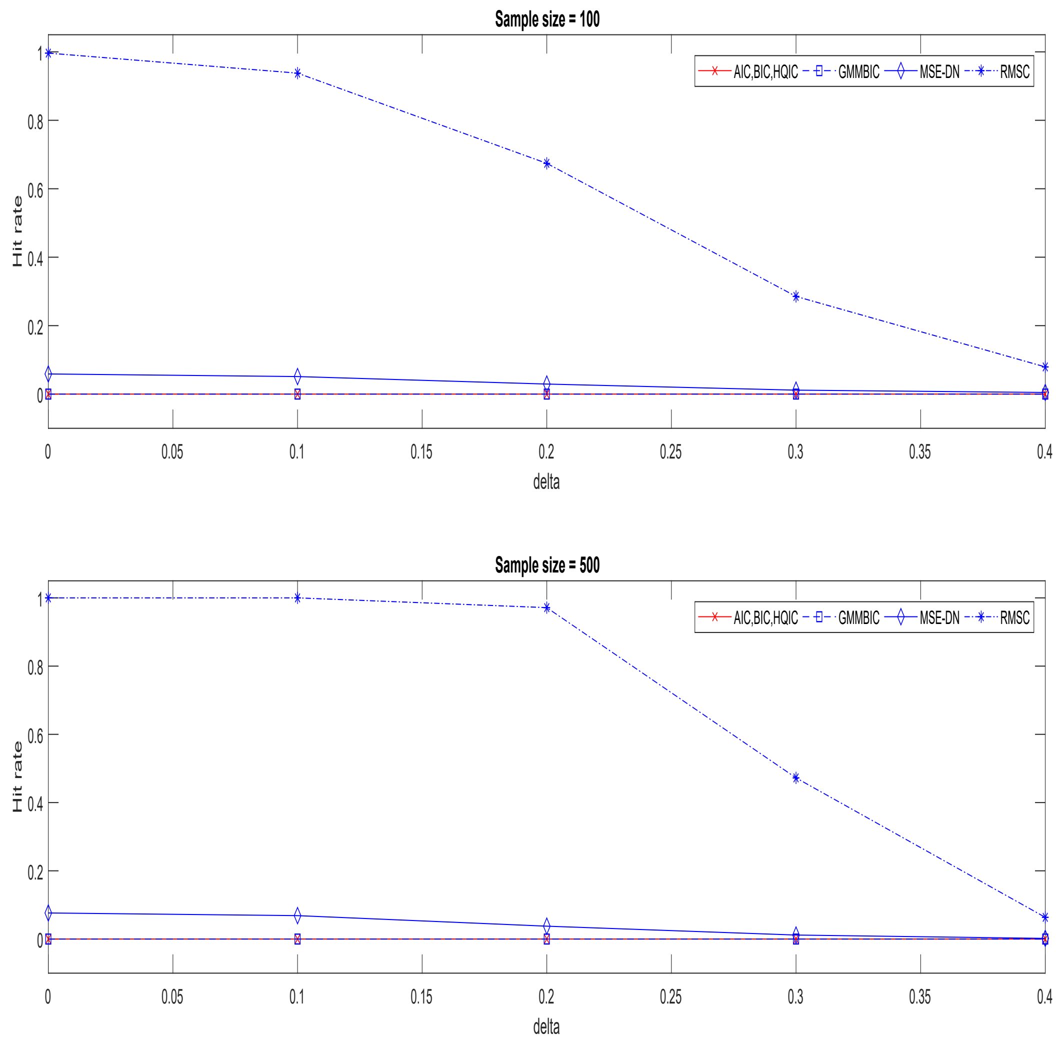


Figure 8: Hit rate of aRMSC and RMSC with 2SLS: model with one endogenous variable ($p = 1$). Sample size $T = 100$; 500; 1,000; 10,000; 50,000; 100,000. Number of replications: 10,000.

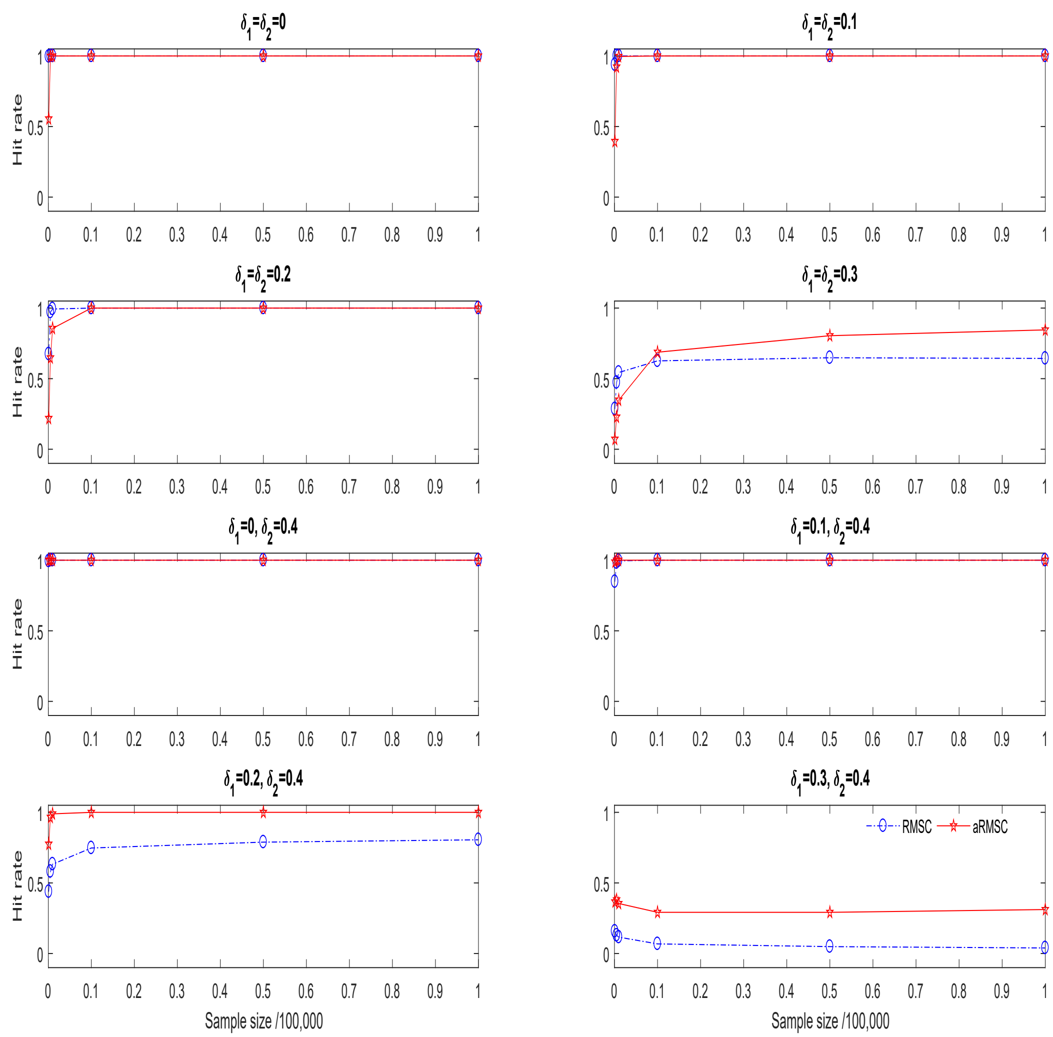


Figure 9: Hit rate of aRMSC and RMSC with 2SLS: model with two endogenous variable ($p = 2$). Sample size $T = 100$; 500; 1,000; 10,000; 50,000; 100,000. Number of replications: 10,000.

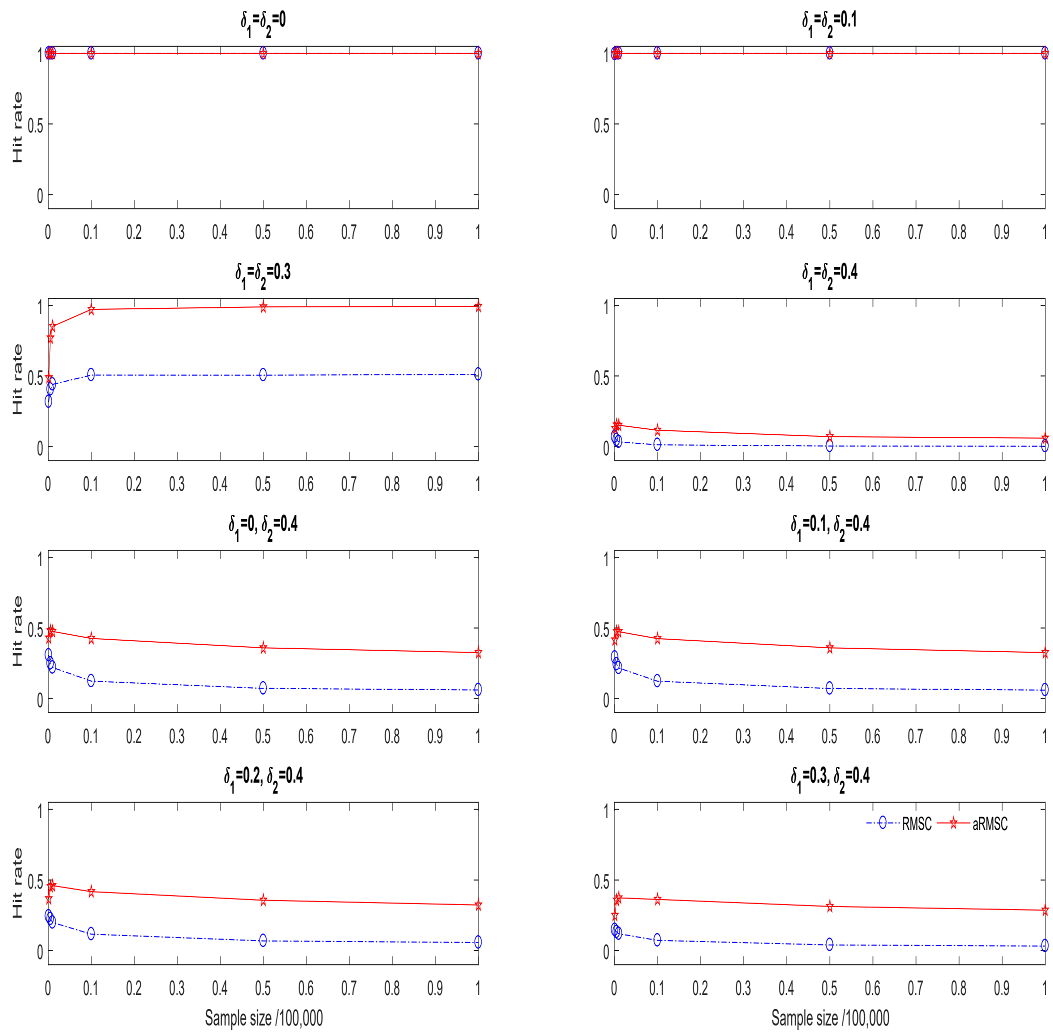


Figure 10: Hit rate of aRMSC and RMSC with LIML: model with one endogenous variable ($p = 1$). Sample size $T = 100$; 500; 1,000; 10,000; 50,000; 100,000. Number of replications: 10,000.

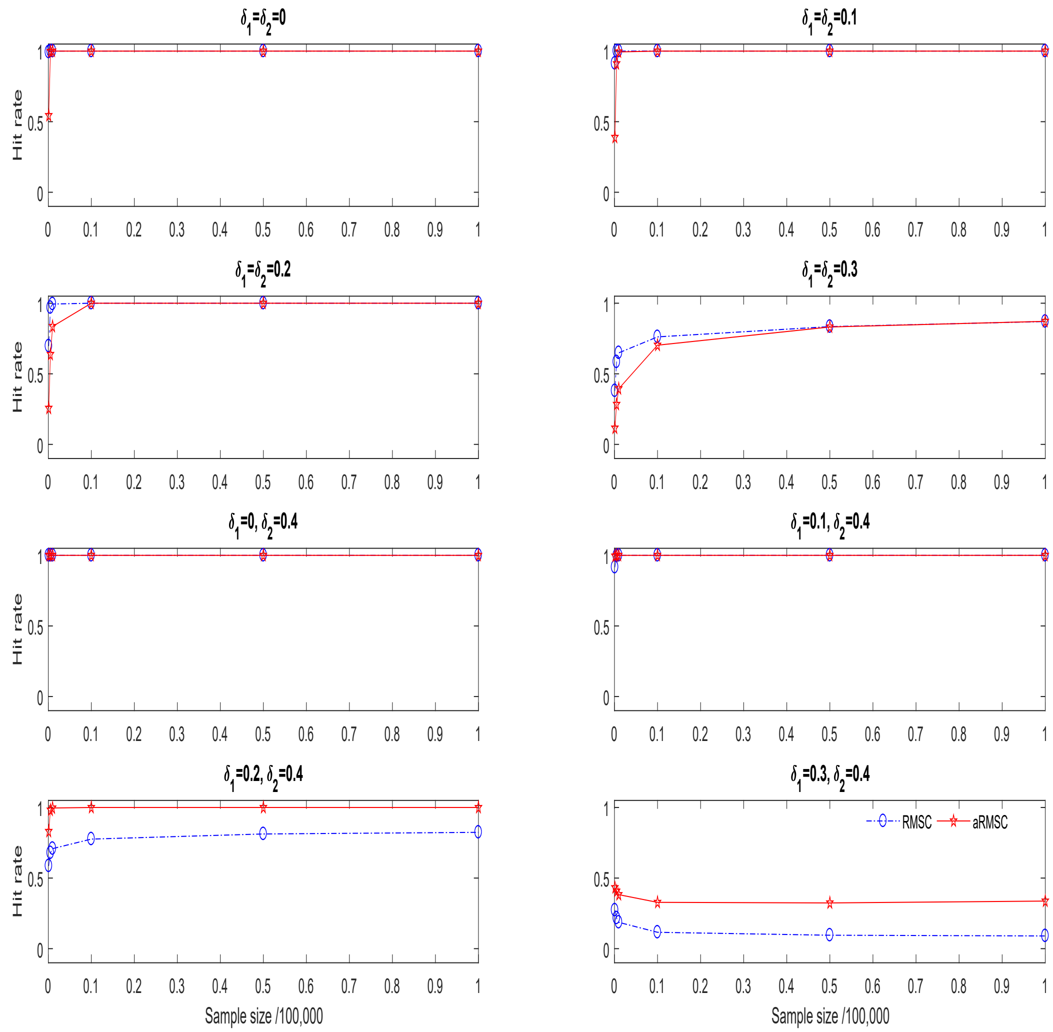
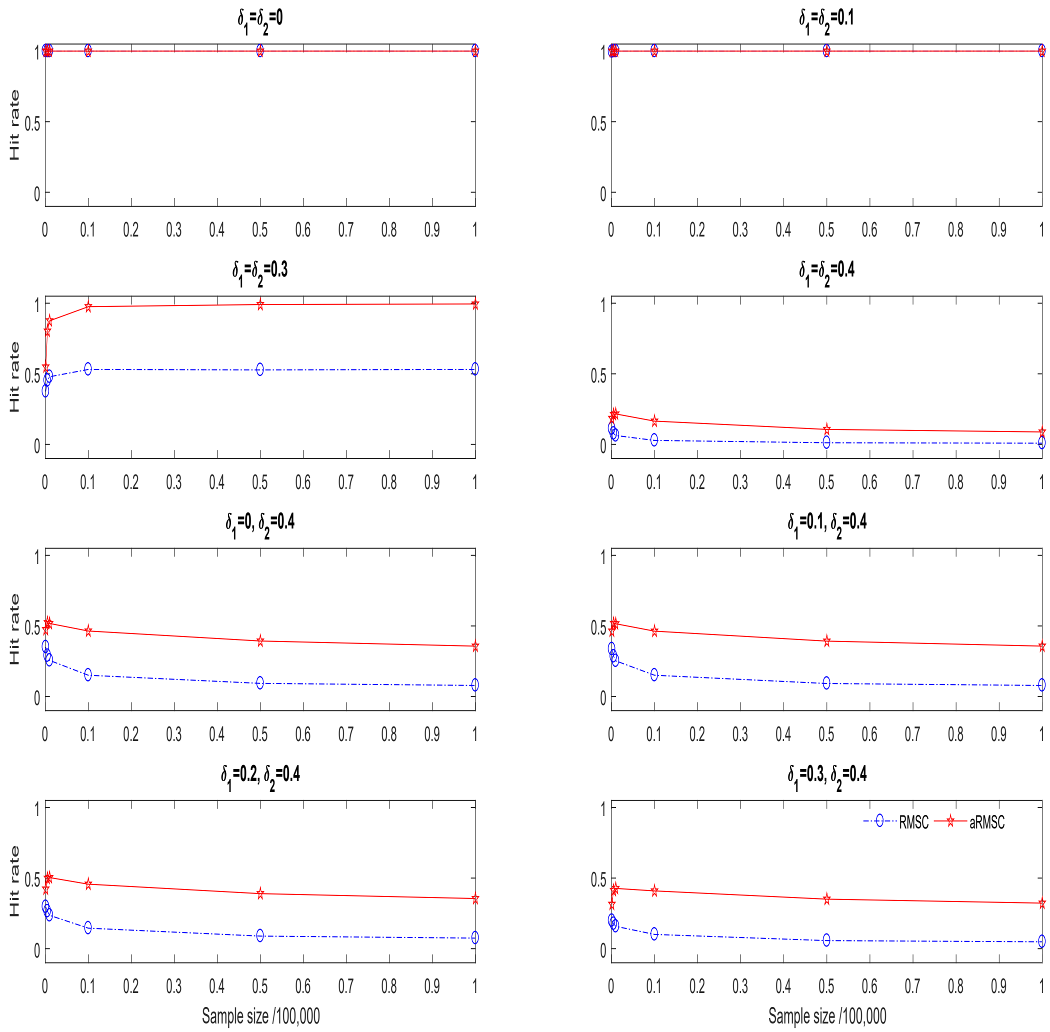


Figure 11: Hit rate of aRMSC and RMSC with LIML: model with two endogenous variable ($p = 2$). Sample size $T = 100$; 500; 1,000; 10,000; 50,000; 100,000. Number of replications: 10,000.



2.7.3 Appendix C: Tables

Table 13: Empirical Selection Probabilities of Some Existing Criteria, T=100.

		T=100									
	δ	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.91
	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.91
AIC,BIC,HQIC	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.85
	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.41	0.53
	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.46	0.19
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
GMMBIC	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0	0.40	0.40	0.06	0.00	0.00	0.00	0.00	0.13	0.00	0.00
	0.1	0.36	0.37	0.05	0.00	0.00	0.00	0.00	0.22	0.00	0.00
MSE_DN	0.2	0.30	0.30	0.03	0.00	0.00	0.00	0.00	0.37	0.00	0.00
	0.3	0.20	0.20	0.01	0.01	0.00	0.00	0.00	0.57	0.00	0.00
	0.4	0.11	0.11	0.00	0.01	0.01	0.00	0.00	0.74	0.02	0.00
	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.03	0.03	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RMSC	0.2	0.11	0.11	0.67	0.03	0.03	0.04	0.00	0.00	0.01	0.00
	0.3	0.11	0.11	0.29	0.13	0.13	0.06	0.00	0.12	0.07	0.00
	0.4	0.06	0.06	0.08	0.10	0.11	0.02	0.00	0.47	0.10	0.00

Table 14: Empirical Selection Probabilities of Some Existing Criteria, T=500.

		T=500									
	δ	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
AIC,BIC,HQIC	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.92
	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.92
	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.91
	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.20	0.67
	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.14	0.10
GMMBIC	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
MSE_DN	0	0.44	0.44	0.08	0.00	0.00	0.00	0.00	0.04	0.00	0.00
	0.1	0.42	0.42	0.07	0.00	0.00	0.00	0.00	0.09	0.00	0.00
	0.2	0.36	0.36	0.04	0.00	0.00	0.00	0.00	0.24	0.00	0.00
	0.3	0.23	0.23	0.01	0.00	0.00	0.00	0.00	0.53	0.00	0.00
	0.4	0.09	0.09	0.00	0.00	0.00	0.00	0.00	0.81	0.00	0.00
RMSC	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.01	0.01	0.97	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.05	0.05	0.47	0.05	0.05	0.12	0.01	0.17	0.02	0.00
	0.4	0.02	0.03	0.06	0.03	0.03	0.02	0.00	0.78	0.02	0.00

Table 15: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of one endogenous regressor (p=1) and T=100.

			T=100																			
			2SLS										LIML									
	δ_1	δ_2	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																					
aRMSC(AIC)	0	0.4	0.99	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.85	0.00	0.11	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.44	0.02	0.23	0.17	0.04	0.03	0.00	0.03	0.04	0.00	0.59	0.04	0.27	0.01	0.04	0.00	0.00	0.05	0.01	0.00
	0.3	0.4	0.16	0.05	0.16	0.15	0.12	0.03	0.00	0.23	0.10	0.00	0.27	0.10	0.23	0.02	0.05	0.00	0.00	0.33	0.00	0.00
aRMSC(BIC)	0	0.4	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.81	0.00	0.06	0.01	0.00	0.03	0.03	0.00	0.03	0.01	0.91	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.36	0.02	0.08	0.04	0.01	0.08	0.10	0.01	0.24	0.07	0.57	0.03	0.21	0.00	0.02	0.04	0.02	0.05	0.06	0.00
	0.3	0.4	0.12	0.04	0.04	0.03	0.03	0.05	0.08	0.18	0.39	0.06	0.26	0.09	0.14	0.01	0.03	0.05	0.03	0.33	0.07	0.01
aRMSC(HQIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.99	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.78	0.07	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.00	0.83	0.09	0.01	0.00	0.01	0.00	0.00	0.05	0.01	0.00
	0.3	0.4	0.37	0.14	0.03	0.02	0.03	0.02	0.02	0.25	0.11	0.01	0.43	0.19	0.05	0.00	0.00	0.00	0.00	0.32	0.00	0.00
RMSC	0	0.4	0.98	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.71	0.00	0.08	0.02	0.00	0.05	0.05	0.00	0.06	0.02	0.86	0.00	0.12	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.2	0.4	0.28	0.01	0.07	0.04	0.01	0.08	0.12	0.01	0.29	0.09	0.50	0.02	0.22	0.01	0.02	0.06	0.04	0.05	0.08	0.01
	0.3	0.4	0.09	0.03	0.03	0.03	0.02	0.05	0.08	0.17	0.43	0.08	0.22	0.07	0.13	0.02	0.03	0.05	0.04	0.32	0.11	0.01
	$\delta_1 = \delta_2$																					
aRMSC(AIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.03	0.03	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.11	0.11	0.67	0.03	0.03	0.04	0.00	0.00	0.01	0.00	0.14	0.14	0.70	0.01	0.01	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.11	0.11	0.29	0.13	0.13	0.06	0.00	0.12	0.07	0.00	0.19	0.18	0.38	0.03	0.03	0.00	0.00	0.18	0.00	0.00
	0.4	0.4	0.06	0.06	0.08	0.10	0.11	0.02	0.00	0.47	0.10	0.00	0.12	0.13	0.13	0.03	0.03	0.00	0.00	0.57	0.00	0.00
aRMSC(BIC)	0	0	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.03	0.03	0.81	0.00	0.00	0.08	0.03	0.00	0.02	0.00	0.04	0.04	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.09	0.09	0.33	0.01	0.01	0.15	0.14	0.00	0.13	0.06	0.14	0.14	0.63	0.00	0.01	0.04	0.01	0.00	0.02	0.00
	0.3	0.3	0.09	0.08	0.08	0.03	0.03	0.08	0.11	0.08	0.35	0.08	0.17	0.17	0.26	0.02	0.02	0.07	0.04	0.18	0.07	0.01
	0.4	0.4	0.04	0.04	0.01	0.03	0.03	0.03	0.04	0.40	0.34	0.04	0.11	0.12	0.07	0.01	0.01	0.03	0.02	0.56	0.05	0.01
aRMSC(HQIC)	0	0	0.22	0.22	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.23	0.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.30	0.31	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.31	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.35	0.35	0.22	0.00	0.00	0.04	0.02	0.00	0.02	0.00	0.37	0.37	0.26	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	0.3	0.3	0.29	0.29	0.07	0.02	0.02	0.03	0.03	0.14	0.09	0.01	0.34	0.35	0.12	0.00	0.00	0.00	0.00	0.19	0.00	0.00
	0.4	0.4	0.16	0.15	0.02	0.02	0.02	0.01	0.01	0.49	0.12	0.01	0.20	0.20	0.03	0.00	0.00	0.00	0.00	0.56	0.00	0.00
RMSC	0	0	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.02	0.02	0.76	0.00	0.00	0.11	0.05	0.00	0.03	0.01	0.03	0.03	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.06	0.06	0.29	0.01	0.01	0.16	0.17	0.00	0.17	0.08	0.11	0.11	0.63	0.01	0.01	0.07	0.03	0.00	0.03	0.00
	0.3	0.3	0.06	0.06	0.06	0.03	0.03	0.08	0.12	0.07	0.39	0.10	0.14	0.14	0.24	0.02	0.02	0.08	0.06	0.17	0.11	0.01
	0.4	0.4	0.03	0.03	0.01	0.02	0.02	0.03	0.04	0.39	0.38	0.05	0.10	0.10	0.06	0.02	0.02	0.04	0.03	0.56	0.08	0.01

Table 16: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of one endogenous regressor (p=1) and T=500

			T=500																			
			2SLS										LIML									
	δ_1	δ_2	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																					
aRMSC(AIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.98	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.58	0.00	0.28	0.09	0.00	0.03	0.00	0.01	0.01	0.00	0.68	0.00	0.31	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	0.3	0.4	0.13	0.02	0.21	0.09	0.04	0.06	0.00	0.43	0.04	0.00	0.22	0.05	0.28	0.00	0.01	0.00	0.00	0.44	0.00	0.00
aRMSC(BIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.98	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.50	0.00	0.12	0.02	0.00	0.09	0.10	0.00	0.13	0.05	0.67	0.00	0.26	0.00	0.00	0.03	0.01	0.01	0.01	0.00
	0.3	0.4	0.09	0.01	0.05	0.02	0.01	0.06	0.08	0.40	0.22	0.06	0.21	0.04	0.19	0.00	0.01	0.05	0.03	0.44	0.03	0.00
aRMSC(HQIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.96	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	0.3	0.4	0.38	0.08	0.07	0.00	0.01	0.01	0.01	0.42	0.02	0.00	0.41	0.10	0.08	0.00	0.00	0.00	0.00	0.41	0.00	0.00
RMSC	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	0.99	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.56	0.00	0.12	0.02	0.00	0.08	0.08	0.00	0.10	0.04	0.72	0.00	0.23	0.00	0.00	0.02	0.01	0.01	0.01	0.00
	0.3	0.4	0.12	0.01	0.06	0.02	0.01	0.06	0.08	0.40	0.19	0.05	0.22	0.05	0.20	0.00	0.01	0.04	0.02	0.44	0.02	0.00
	$\delta_1 = \delta_2$																					
aRMSC(AIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.01	0.01	0.97	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.05	0.05	0.47	0.05	0.05	0.12	0.01	0.17	0.02	0.00	0.10	0.11	0.58	0.01	0.01	0.00	0.00	0.19	0.00	0.00
	0.4	0.4	0.02	0.03	0.06	0.03	0.03	0.02	0.00	0.78	0.02	0.00	0.06	0.07	0.10	0.00	0.00	0.00	0.00	0.77	0.00	0.00
aRMSC(BIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.01	0.01	0.64	0.00	0.00	0.16	0.10	0.00	0.07	0.03	0.02	0.02	0.95	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.3	0.04	0.04	0.12	0.01	0.01	0.12	0.17	0.15	0.24	0.11	0.10	0.10	0.42	0.00	0.00	0.09	0.04	0.20	0.03	0.01
	0.4	0.4	0.02	0.02	0.01	0.01	0.01	0.02	0.02	0.77	0.10	0.02	0.06	0.06	0.06	0.00	0.00	0.02	0.01	0.77	0.01	0.00
aRMSC(HQIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.04	0.04	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.17	0.18	0.65	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.18	0.19	0.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.24	0.25	0.23	0.00	0.00	0.04	0.02	0.19	0.02	0.00	0.26	0.26	0.28	0.00	0.00	0.00	0.00	0.20	0.00	0.00
	0.4	0.4	0.08	0.09	0.03	0.00	0.00	0.01	0.00	0.78	0.01	0.00	0.10	0.11	0.04	0.00	0.00	0.00	0.00	0.76	0.00	0.00
RMSC	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.01	0.01	0.70	0.00	0.00	0.14	0.08	0.00	0.05	0.02	0.02	0.02	0.95	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.3	0.04	0.05	0.14	0.01	0.01	0.13	0.16	0.15	0.21	0.09	0.11	0.11	0.44	0.00	0.00	0.07	0.03	0.20	0.02	0.00
	0.4	0.4	0.02	0.03	0.02	0.01	0.01	0.02	0.02	0.77	0.09	0.01	0.06	0.07	0.06	0.00	0.00	0.01	0.01	0.77	0.01	0.00

Table 17: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of one endogenous regressor (p=1) and T=1000

			T=1000																			
			2SLS										LIML									
	δ_1	δ_2	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																					
aRMSC(AIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.63	0.00	0.28	0.06	0.00	0.02	0.00	0.00	0.00	0.00	0.71	0.00	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.12	0.01	0.23	0.06	0.02	0.06	0.00	0.50	0.02	0.00	0.19	0.03	0.29	0.00	0.01	0.00	0.00	0.49	0.00	0.00
aRMSC(BIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.55	0.00	0.13	0.02	0.00	0.08	0.08	0.00	0.10	0.04	0.70	0.00	0.26	0.00	0.00	0.02	0.01	0.00	0.00	0.00
	0.3	0.4	0.08	0.01	0.05	0.02	0.00	0.06	0.08	0.49	0.16	0.06	0.18	0.02	0.21	0.00	0.00	0.05	0.02	0.49	0.02	0.01
aRMSC(HQIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.99	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.35	0.05	0.09	0.00	0.00	0.01	0.01	0.48	0.01	0.00	0.38	0.06	0.10	0.00	0.00	0.00	0.00	0.46	0.00	0.00
RMSC	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.68	0.00	0.12	0.01	0.00	0.06	0.05	0.00	0.06	0.02	0.79	0.00	0.19	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.4	0.12	0.01	0.07	0.01	0.01	0.06	0.07	0.49	0.12	0.04	0.20	0.03	0.22	0.00	0.00	0.03	0.01	0.49	0.01	0.00
	$\delta_1 = \delta_2$																					
aRMSC(AIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.03	0.03	0.54	0.03	0.03	0.13	0.01	0.21	0.01	0.00	0.07	0.07	0.65	0.00	0.00	0.00	0.00	0.21	0.00	0.00
	0.4	0.4	0.02	0.02	0.05	0.02	0.02	0.02	0.00	0.85	0.01	0.00	0.04	0.05	0.08	0.00	0.00	0.00	0.00	0.82	0.00	0.00
aRMSC(BIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.73	0.00	0.00	0.13	0.08	0.00	0.04	0.02	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.3	0.02	0.02	0.14	0.01	0.01	0.13	0.17	0.18	0.22	0.11	0.07	0.06	0.48	0.00	0.00	0.10	0.04	0.22	0.02	0.01
	0.4	0.4	0.01	0.02	0.01	0.01	0.00	0.01	0.02	0.85	0.06	0.01	0.04	0.04	0.05	0.00	0.00	0.02	0.01	0.83	0.01	0.00
aRMSC(HQIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.07	0.07	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.08	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.18	0.18	0.35	0.00	0.00	0.04	0.02	0.21	0.01	0.00	0.20	0.20	0.39	0.00	0.00	0.00	0.00	0.21	0.00	0.00
	0.4	0.4	0.06	0.06	0.03	0.00	0.00	0.00	0.00	0.84	0.00	0.00	0.07	0.07	0.04	0.00	0.00	0.00	0.00	0.82	0.00	0.00
RMSC	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.82	0.00	0.00	0.10	0.05	0.00	0.02	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.03	0.03	0.19	0.01	0.01	0.14	0.15	0.19	0.17	0.08	0.08	0.07	0.54	0.00	0.00	0.06	0.02	0.22	0.01	0.00
	0.4	0.4	0.02	0.02	0.02	0.00	0.01	0.02	0.02	0.85	0.04	0.01	0.05	0.05	0.06	0.00	0.00	0.01	0.00	0.83	0.00	0.00

Table 18: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of one endogenous regressor (p=1) and T=10000

			T=10000																			
			2SLS										LIML									
	δ_1	δ_2	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																					
aRMSC(AIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.75	0.00	0.24	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.78	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.07	0.00	0.26	0.02	0.00	0.05	0.00	0.59	0.00	0.00	0.12	0.00	0.34	0.00	0.00	0.00	0.00	0.54	0.00	0.00
aRMSC(BIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.70	0.00	0.16	0.00	0.00	0.05	0.04	0.00	0.03	0.01	0.78	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.05	0.00	0.06	0.01	0.00	0.07	0.09	0.59	0.10	0.05	0.11	0.00	0.25	0.00	0.00	0.05	0.02	0.55	0.01	0.00
aRMSC(HQIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.29	0.00	0.14	0.00	0.00	0.00	0.00	0.56	0.00	0.00	0.33	0.01	0.15	0.00	0.00	0.00	0.00	0.52	0.00	0.00
RMSC	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.90	0.00	0.08	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.94	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.10	0.00	0.10	0.00	0.00	0.07	0.07	0.59	0.05	0.02	0.14	0.00	0.30	0.00	0.00	0.01	0.00	0.54	0.00	0.00
	$\delta_1 = \delta_2$																					
aRMSC(AIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.63	0.00	0.00	0.12	0.00	0.24	0.00	0.00	0.01	0.01	0.76	0.00	0.00	0.00	0.00	0.23	0.00	0.00
	0.4	0.4	0.00	0.00	0.04	0.00	0.01	0.02	0.00	0.93	0.00	0.00	0.02	0.02	0.07	0.00	0.00	0.00	0.00	0.89	0.00	0.00
aRMSC(BIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.92	0.00	0.00	0.05	0.02	0.00	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.17	0.00	0.00	0.15	0.18	0.24	0.17	0.10	0.01	0.01	0.60	0.00	0.00	0.10	0.04	0.23	0.02	0.01
	0.4	0.4	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.92	0.03	0.01	0.02	0.02	0.04	0.00	0.00	0.02	0.01	0.89	0.01	0.00
aRMSC(HQIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.02	0.02	0.69	0.00	0.00	0.01	0.00	0.25	0.00	0.00	0.03	0.03	0.70	0.00	0.00	0.00	0.00	0.23	0.00	0.00
	0.4	0.4	0.02	0.02	0.03	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.03	0.03	0.05	0.00	0.00	0.00	0.00	0.89	0.00	0.00
RMSC	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.30	0.00	0.00	0.17	0.14	0.24	0.11	0.05	0.01	0.01	0.72	0.00	0.00	0.03	0.00	0.23	0.00	0.00
	0.4	0.4	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.92	0.02	0.01	0.02	0.02	0.06	0.00	0.00	0.01	0.00	0.89	0.00	0.00

Table 19: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of one endogenous regressor (p=1) and T=50000

			T=50000																			
			2SLS										LIML									
	δ_1	δ_2	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																					
aRMSC(AIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.79	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.05	0.00	0.30	0.02	0.00	0.10	0.00	0.52	0.00	0.00	0.10	0.00	0.43	0.00	0.00	0.00	0.00	0.47	0.00	0.00
aRMSC(BIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.77	0.00	0.17	0.00	0.00	0.03	0.02	0.00	0.01	0.00	0.81	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.03	0.00	0.05	0.01	0.00	0.07	0.11	0.52	0.14	0.08	0.09	0.00	0.27	0.00	0.00	0.09	0.05	0.47	0.03	0.01
aRMSC(HQIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.29	0.00	0.20	0.00	0.00	0.01	0.00	0.51	0.00	0.00	0.32	0.00	0.22	0.00	0.00	0.00	0.00	0.46	0.00	0.00
RMSC	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.96	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.08	0.00	0.12	0.00	0.00	0.09	0.08	0.52	0.07	0.03	0.12	0.00	0.37	0.00	0.00	0.03	0.01	0.47	0.00	0.00
	$\delta_1 = \delta_2$																					
aRMSC(AIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.65	0.00	0.00	0.17	0.00	0.18	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.17	0.00	0.00
	0.4	0.4	0.00	0.00	0.03	0.00	0.00	0.03	0.01	0.92	0.01	0.00	0.01	0.01	0.08	0.00	0.00	0.01	0.00	0.88	0.00	0.00
aRMSC(BIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.98	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.16	0.00	0.00	0.15	0.19	0.18	0.20	0.12	0.00	0.00	0.60	0.00	0.00	0.13	0.06	0.17	0.03	0.01
	0.4	0.4	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.92	0.04	0.02	0.01	0.01	0.03	0.00	0.00	0.02	0.02	0.88	0.03	0.01
aRMSC(HQIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.80	0.00	0.00	0.01	0.00	0.18	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.16	0.00	0.00
	0.4	0.4	0.01	0.01	0.04	0.00	0.00	0.01	0.00	0.92	0.00	0.00	0.02	0.02	0.08	0.00	0.00	0.00	0.00	0.88	0.00	0.00
RMSC	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.34	0.00	0.00	0.18	0.14	0.18	0.11	0.05	0.00	0.00	0.80	0.00	0.00	0.03	0.00	0.17	0.00	0.00
	0.4	0.4	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.92	0.03	0.01	0.01	0.02	0.05	0.00	0.00	0.02	0.01	0.88	0.01	0.00

Table 20: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of one endogenous regressor (p=1) and T=100000

		T=100000																				
		2SLS											LIML									
	δ_1	δ_2	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1	z2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																					
aRMSC(AIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.81	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.82	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.04	0.00	0.31	0.03	0.00	0.16	0.01	0.45	0.00	0.00	0.09	0.00	0.50	0.00	0.00	0.00	0.00	0.41	0.00	0.00
aRMSC(BIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.79	0.00	0.17	0.00	0.00	0.02	0.01	0.00	0.01	0.00	0.82	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.03	0.00	0.05	0.01	0.00	0.08	0.13	0.45	0.16	0.10	0.09	0.00	0.27	0.00	0.00	0.11	0.07	0.41	0.04	0.02
aRMSC(HQIC)	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.31	0.00	0.24	0.00	0.00	0.01	0.00	0.44	0.00	0.00	0.34	0.00	0.26	0.00	0.00	0.00	0.00	0.40	0.00	0.00
RMSC	0	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.4	0.98	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.4	0.07	0.00	0.13	0.00	0.00	0.10	0.11	0.45	0.09	0.04	0.12	0.00	0.42	0.00	0.00	0.03	0.01	0.41	0.00	0.00
	$\delta_1 = \delta_2$																					
aRMSC(AIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.64	0.00	0.00	0.21	0.01	0.14	0.00	0.00	0.00	0.00	0.87	0.00	0.00	0.00	0.00	0.13	0.00	0.00
	0.4	0.4	0.00	0.00	0.02	0.00	0.00	0.04	0.01	0.92	0.01	0.00	0.01	0.01	0.09	0.00	0.00	0.02	0.00	0.87	0.00	0.00
aRMSC(BIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.16	0.00	0.00	0.16	0.20	0.14	0.22	0.13	0.00	0.00	0.60	0.00	0.00	0.15	0.07	0.13	0.04	0.01
	0.4	0.4	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.91	0.03	0.02	0.01	0.01	0.02	0.00	0.00	0.02	0.03	0.87	0.04	0.01
aRMSC(HQIC)	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.85	0.00	0.00	0.02	0.00	0.14	0.00	0.00	0.00	0.00	0.87	0.00	0.00	0.00	0.00	0.13	0.00	0.00
	0.4	0.4	0.01	0.01	0.04	0.00	0.00	0.01	0.01	0.92	0.00	0.00	0.02	0.02	0.09	0.00	0.00	0.00	0.00	0.87	0.00	0.00
RMSC	0	0	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.00	0.00	0.34	0.00	0.00	0.19	0.15	0.14	0.12	0.06	0.00	0.00	0.83	0.00	0.00	0.04	0.00	0.13	0.00	0.00
	0.4	0.4	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.92	0.03	0.01	0.01	0.01	0.05	0.00	0.00	0.02	0.02	0.87	0.02	0.00

Table 21: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of Two endogenous regressors (p=2) and T=100

			T=100															
			2SLS								LIML							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
aRMSC(AIC)	0	0.4	0.31	0.03	0.00	0.43	0.15	0.00	0.09	0.00	0.35	0.05	0.00	0.41	0.11	0.00	0.09	0.00
	0.1	0.4	0.29	0.03	0.00	0.43	0.16	0.00	0.09	0.00	0.34	0.05	0.00	0.41	0.11	0.00	0.09	0.00
	0.2	0.4	0.24	0.02	0.00	0.43	0.20	0.00	0.11	0.00	0.30	0.04	0.00	0.42	0.14	0.00	0.10	0.00
	0.3	0.4	0.15	0.01	0.00	0.39	0.29	0.00	0.16	0.00	0.20	0.03	0.00	0.41	0.21	0.00	0.14	0.00
aRMSC(BIC)	0	0.4	0.38	0.04	0.00	0.23	0.17	0.00	0.15	0.03	0.43	0.06	0.00	0.23	0.14	0.00	0.12	0.02
	0.1	0.4	0.36	0.04	0.00	0.22	0.17	0.00	0.16	0.05	0.41	0.06	0.00	0.23	0.15	0.00	0.13	0.02
	0.2	0.4	0.31	0.03	0.00	0.20	0.18	0.00	0.19	0.09	0.37	0.06	0.00	0.21	0.16	0.00	0.16	0.04
	0.3	0.4	0.20	0.02	0.00	0.16	0.19	0.00	0.27	0.16	0.26	0.04	0.01	0.20	0.19	0.00	0.22	0.07
aRMSC(HQIC)	0	0.4	0.43	0.05	0.00	0.23	0.14	0.00	0.13	0.02	0.47	0.08	0.00	0.22	0.12	0.00	0.10	0.01
	0.1	0.4	0.42	0.05	0.00	0.22	0.15	0.00	0.13	0.03	0.46	0.07	0.00	0.22	0.12	0.00	0.11	0.01
	0.2	0.4	0.37	0.04	0.00	0.20	0.16	0.00	0.16	0.06	0.42	0.07	0.00	0.21	0.13	0.00	0.13	0.03
	0.3	0.4	0.25	0.03	0.00	0.17	0.18	0.00	0.23	0.13	0.31	0.06	0.01	0.20	0.17	0.00	0.19	0.06
RMSC	0	0.4	0.14	0.01	0.00	0.17	0.24	0.00	0.29	0.15	0.18	0.03	0.00	0.22	0.24	0.00	0.25	0.07
	0.1	0.4	0.12	0.01	0.00	0.15	0.22	0.00	0.31	0.20	0.17	0.02	0.00	0.20	0.24	0.00	0.27	0.10
	0.2	0.4	0.08	0.00	0.00	0.11	0.19	0.00	0.34	0.28	0.12	0.02	0.00	0.17	0.24	0.00	0.31	0.14
	0.3	0.4	0.04	0.00	0.00	0.06	0.15	0.00	0.37	0.38	0.07	0.01	0.00	0.13	0.24	0.00	0.36	0.18
	$\delta_1 = \delta_2$																	
aRMSC(AIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.82	0.00	0.00	0.17	0.01	0.00	0.00	0.00	0.84	0.00	0.00	0.15	0.01	0.00	0.00	0.00
	0.3	0.3	0.32	0.00	0.00	0.44	0.19	0.00	0.05	0.00	0.38	0.00	0.00	0.43	0.14	0.00	0.05	0.00
	0.4	0.4	0.07	0.01	0.01	0.30	0.34	0.00	0.27	0.01	0.11	0.02	0.03	0.34	0.25	0.01	0.23	0.00
aRMSC(BIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.90	0.00	0.00	0.06	0.02	0.00	0.01	0.01	0.92	0.00	0.00	0.05	0.02	0.00	0.01	0.00
	0.3	0.3	0.41	0.00	0.00	0.18	0.15	0.00	0.15	0.10	0.48	0.01	0.01	0.19	0.14	0.00	0.13	0.05
	0.4	0.4	0.10	0.01	0.01	0.11	0.19	0.00	0.35	0.22	0.15	0.03	0.04	0.16	0.21	0.01	0.31	0.09
aRMSC(HQIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.93	0.00	0.00	0.04	0.01	0.00	0.01	0.00	0.94	0.00	0.00	0.04	0.01	0.00	0.01	0.00
	0.3	0.3	0.49	0.00	0.00	0.18	0.13	0.00	0.12	0.07	0.55	0.01	0.01	0.18	0.12	0.00	0.10	0.03
	0.4	0.4	0.13	0.02	0.02	0.13	0.19	0.00	0.33	0.19	0.18	0.04	0.05	0.17	0.19	0.01	0.28	0.07
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	0.94	0.00	0.00	0.04	0.01	0.00	0.01	0.01	0.95	0.00	0.00	0.03	0.01	0.00	0.01	0.00
	0.2	0.2	0.45	0.00	0.00	0.16	0.13	0.00	0.15	0.11	0.50	0.00	0.00	0.16	0.13	0.00	0.13	0.07
	0.3	0.3	0.09	0.00	0.00	0.10	0.18	0.00	0.32	0.32	0.13	0.00	0.00	0.16	0.24	0.00	0.30	0.18
	0.4	0.4	0.02	0.00	0.00	0.04	0.13	0.00	0.39	0.42	0.04	0.01	0.01	0.11	0.22	0.01	0.42	0.17

Table 22: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of Two endogenous regressors (p=2) and T=500

			T=500															
			2SLS								LIML							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
aRMSC(AIC)	0	0.4	0.25	0.01	0.00	0.44	0.23	0.00	0.07	0.00	0.29	0.01	0.00	0.44	0.19	0.00	0.07	0.00
	0.1	0.4	0.24	0.01	0.00	0.44	0.24	0.00	0.07	0.00	0.28	0.01	0.00	0.44	0.19	0.00	0.07	0.00
	0.2	0.4	0.22	0.01	0.00	0.42	0.26	0.00	0.08	0.00	0.27	0.01	0.00	0.43	0.21	0.00	0.07	0.00
	0.3	0.4	0.13	0.00	0.00	0.37	0.34	0.00	0.14	0.01	0.18	0.01	0.00	0.41	0.28	0.00	0.11	0.01
aRMSC(BIC)	0	0.4	0.31	0.01	0.00	0.23	0.22	0.00	0.17	0.06	0.36	0.02	0.00	0.25	0.20	0.00	0.14	0.03
	0.1	0.4	0.31	0.01	0.00	0.23	0.22	0.00	0.17	0.07	0.35	0.02	0.00	0.25	0.20	0.00	0.14	0.04
	0.2	0.4	0.29	0.01	0.00	0.20	0.21	0.00	0.20	0.09	0.33	0.02	0.00	0.23	0.20	0.00	0.16	0.05
	0.3	0.4	0.18	0.01	0.00	0.14	0.20	0.00	0.28	0.19	0.24	0.01	0.00	0.19	0.22	0.00	0.23	0.10
aRMSC(HQIC)	0	0.4	0.48	0.02	0.00	0.23	0.15	0.00	0.10	0.02	0.52	0.03	0.00	0.22	0.13	0.00	0.08	0.01
	0.1	0.4	0.47	0.02	0.00	0.23	0.15	0.00	0.10	0.03	0.52	0.03	0.00	0.23	0.13	0.00	0.08	0.01
	0.2	0.4	0.45	0.02	0.00	0.22	0.15	0.00	0.12	0.04	0.50	0.03	0.00	0.22	0.13	0.00	0.09	0.02
	0.3	0.4	0.36	0.01	0.00	0.18	0.18	0.00	0.18	0.09	0.41	0.03	0.00	0.21	0.17	0.00	0.14	0.05
RMSC	0	0.4	0.16	0.00	0.00	0.20	0.25	0.00	0.26	0.12	0.20	0.01	0.00	0.24	0.27	0.00	0.22	0.07
	0.1	0.4	0.15	0.00	0.00	0.19	0.25	0.00	0.27	0.14	0.19	0.01	0.00	0.23	0.26	0.00	0.23	0.08
	0.2	0.4	0.13	0.00	0.00	0.15	0.21	0.00	0.30	0.21	0.16	0.01	0.00	0.20	0.25	0.00	0.26	0.11
	0.3	0.4	0.06	0.00	0.00	0.08	0.17	0.00	0.34	0.34	0.09	0.01	0.00	0.14	0.24	0.00	0.32	0.19
	$\delta_1 = \delta_2$																	
aRMSC(AIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.98	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.02	0.00	0.00	0.00	0.00
	0.3	0.3	0.40	0.00	0.00	0.42	0.15	0.00	0.02	0.00	0.45	0.00	0.00	0.41	0.12	0.00	0.02	0.00
	0.4	0.4	0.05	0.00	0.00	0.22	0.39	0.00	0.29	0.04	0.08	0.01	0.01	0.30	0.35	0.00	0.23	0.02
aRMSC(BIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.52	0.00	0.00	0.15	0.12	0.00	0.13	0.08	0.57	0.00	0.00	0.16	0.12	0.00	0.10	0.05
	0.4	0.4	0.06	0.00	0.00	0.08	0.18	0.00	0.36	0.31	0.11	0.01	0.01	0.16	0.24	0.00	0.33	0.15
aRMSC(HQIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.77	0.00	0.00	0.11	0.06	0.00	0.04	0.02	0.80	0.00	0.00	0.10	0.05	0.00	0.03	0.01
	0.4	0.4	0.15	0.01	0.00	0.13	0.20	0.00	0.30	0.20	0.21	0.02	0.02	0.19	0.21	0.00	0.25	0.09
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.92	0.00	0.00	0.04	0.02	0.00	0.01	0.01	0.93	0.00	0.00	0.04	0.02	0.00	0.01	0.01
	0.3	0.3	0.21	0.00	0.00	0.13	0.18	0.00	0.26	0.23	0.26	0.00	0.00	0.16	0.20	0.00	0.23	0.15
	0.4	0.4	0.02	0.00	0.00	0.04	0.13	0.00	0.37	0.44	0.04	0.01	0.01	0.11	0.23	0.00	0.39	0.21

Table 23: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of Two endogenous regressors (p=2) and T=1000

			T=1000															
			2SLS								LIML							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
aRMSC(AIC)	0	0.4	0.22	0.00	0.00	0.43	0.27	0.00	0.08	0.00	0.26	0.01	0.00	0.44	0.23	0.00	0.07	0.00
	0.1	0.4	0.22	0.00	0.00	0.42	0.27	0.00	0.08	0.00	0.25	0.01	0.00	0.44	0.23	0.00	0.07	0.00
	0.2	0.4	0.20	0.00	0.00	0.42	0.29	0.00	0.09	0.01	0.24	0.01	0.00	0.43	0.25	0.00	0.08	0.00
	0.3	0.4	0.12	0.00	0.00	0.36	0.35	0.00	0.15	0.02	0.16	0.01	0.00	0.40	0.31	0.00	0.12	0.01
aRMSC(BIC)	0	0.4	0.28	0.00	0.00	0.23	0.23	0.00	0.19	0.07	0.32	0.01	0.00	0.25	0.22	0.00	0.16	0.04
	0.1	0.4	0.28	0.00	0.00	0.22	0.23	0.00	0.19	0.08	0.32	0.01	0.00	0.25	0.22	0.00	0.16	0.04
	0.2	0.4	0.26	0.00	0.00	0.21	0.22	0.00	0.21	0.10	0.30	0.01	0.00	0.24	0.22	0.00	0.18	0.06
	0.3	0.4	0.17	0.00	0.00	0.14	0.19	0.00	0.28	0.22	0.21	0.01	0.00	0.19	0.23	0.00	0.24	0.12
aRMSC(HQIC)	0	0.4	0.48	0.01	0.00	0.23	0.15	0.00	0.10	0.03	0.52	0.02	0.00	0.23	0.14	0.00	0.08	0.01
	0.1	0.4	0.47	0.01	0.00	0.23	0.15	0.00	0.10	0.03	0.52	0.02	0.00	0.23	0.14	0.00	0.08	0.02
	0.2	0.4	0.46	0.01	0.00	0.22	0.16	0.00	0.11	0.04	0.50	0.02	0.00	0.23	0.14	0.00	0.09	0.02
	0.3	0.4	0.37	0.01	0.00	0.19	0.17	0.00	0.17	0.09	0.43	0.02	0.00	0.21	0.16	0.00	0.13	0.05
RMSC	0	0.4	0.16	0.00	0.00	0.19	0.26	0.00	0.26	0.13	0.19	0.01	0.00	0.23	0.27	0.00	0.23	0.07
	0.1	0.4	0.15	0.00	0.00	0.19	0.25	0.00	0.26	0.14	0.19	0.01	0.00	0.23	0.26	0.00	0.23	0.08
	0.2	0.4	0.13	0.00	0.00	0.16	0.23	0.00	0.29	0.19	0.17	0.01	0.00	0.21	0.25	0.00	0.25	0.11
	0.3	0.4	0.06	0.00	0.00	0.09	0.18	0.00	0.33	0.34	0.10	0.01	0.00	0.14	0.24	0.00	0.32	0.19
	$\delta_1 = \delta_2$																	
aRMSC(AIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.3	0.44	0.00	0.00	0.41	0.13	0.00	0.02	0.00	0.48	0.00	0.00	0.40	0.11	0.00	0.01	0.00
	0.4	0.4	0.03	0.00	0.00	0.20	0.38	0.00	0.31	0.07	0.06	0.01	0.00	0.28	0.38	0.00	0.24	0.03
aRMSC(BIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.55	0.00	0.00	0.15	0.11	0.00	0.11	0.07	0.60	0.00	0.00	0.15	0.11	0.00	0.09	0.05
	0.4	0.4	0.05	0.00	0.00	0.07	0.17	0.00	0.35	0.36	0.09	0.01	0.01	0.14	0.25	0.00	0.34	0.17
aRMSC(HQIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.85	0.00	0.00	0.08	0.03	0.00	0.03	0.01	0.88	0.00	0.00	0.07	0.03	0.00	0.02	0.01
	0.4	0.4	0.15	0.00	0.00	0.14	0.20	0.00	0.29	0.21	0.22	0.01	0.01	0.19	0.23	0.00	0.25	0.10
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	0.98	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.3	0.3	0.27	0.00	0.00	0.14	0.17	0.00	0.23	0.19	0.32	0.00	0.00	0.16	0.18	0.00	0.21	0.14
	0.4	0.4	0.02	0.00	0.00	0.04	0.13	0.00	0.35	0.46	0.04	0.00	0.00	0.11	0.24	0.00	0.38	0.22

Table 24: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of Two endogenous regressors (p=2) and T=10000

			T=10000															
			2SLS								LIML							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
aRMSC(AIC)	0	0.4	0.12	0.00	0.00	0.34	0.36	0.00	0.15	0.02	0.15	0.00	0.00	0.38	0.33	0.00	0.12	0.01
	0.1	0.4	0.12	0.00	0.00	0.34	0.36	0.00	0.15	0.02	0.15	0.00	0.00	0.38	0.33	0.00	0.13	0.01
	0.2	0.4	0.12	0.00	0.00	0.34	0.36	0.00	0.16	0.02	0.15	0.00	0.00	0.37	0.33	0.00	0.13	0.01
	0.3	0.4	0.07	0.00	0.00	0.27	0.38	0.00	0.23	0.04	0.10	0.00	0.00	0.32	0.37	0.00	0.19	0.03
aRMSC(BIC)	0	0.4	0.16	0.00	0.00	0.18	0.25	0.00	0.26	0.14	0.20	0.00	0.00	0.22	0.27	0.00	0.23	0.09
	0.1	0.4	0.16	0.00	0.00	0.18	0.25	0.00	0.26	0.14	0.20	0.00	0.00	0.22	0.27	0.00	0.23	0.09
	0.2	0.4	0.15	0.00	0.00	0.17	0.25	0.00	0.27	0.16	0.19	0.00	0.00	0.21	0.27	0.00	0.23	0.10
	0.3	0.4	0.10	0.00	0.00	0.11	0.20	0.00	0.30	0.29	0.13	0.00	0.00	0.15	0.24	0.00	0.28	0.19
aRMSC(HQIC)	0	0.4	0.43	0.00	0.00	0.22	0.18	0.00	0.13	0.05	0.46	0.00	0.00	0.23	0.17	0.00	0.11	0.03
	0.1	0.4	0.42	0.00	0.00	0.22	0.18	0.00	0.13	0.05	0.46	0.00	0.00	0.23	0.17	0.00	0.11	0.03
	0.2	0.4	0.42	0.00	0.00	0.22	0.18	0.00	0.13	0.05	0.46	0.00	0.00	0.23	0.17	0.00	0.11	0.03
	0.3	0.4	0.36	0.00	0.00	0.20	0.18	0.00	0.17	0.09	0.41	0.00	0.00	0.21	0.18	0.00	0.14	0.06
RMSC	0	0.4	0.11	0.00	0.00	0.16	0.26	0.00	0.29	0.17	0.14	0.00	0.00	0.20	0.28	0.00	0.26	0.12
	0.1	0.4	0.11	0.00	0.00	0.16	0.26	0.00	0.30	0.18	0.14	0.00	0.00	0.20	0.28	0.00	0.26	0.12
	0.2	0.4	0.10	0.00	0.00	0.15	0.24	0.00	0.30	0.20	0.13	0.00	0.00	0.19	0.27	0.00	0.27	0.13
	0.3	0.4	0.06	0.00	0.00	0.08	0.18	0.00	0.32	0.36	0.09	0.00	0.00	0.13	0.23	0.00	0.32	0.23
	$\delta_1 = \delta_2$																	
aRMSC(AIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.51	0.00	0.00	0.37	0.11	0.00	0.01	0.00	0.53	0.00	0.00	0.36	0.09	0.00	0.01	0.00
	0.4	0.4	0.01	0.00	0.00	0.10	0.30	0.00	0.41	0.18	0.03	0.00	0.00	0.17	0.36	0.00	0.34	0.10
aRMSC(BIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.62	0.00	0.00	0.12	0.09	0.00	0.10	0.07	0.65	0.00	0.00	0.12	0.09	0.00	0.09	0.06
	0.4	0.4	0.02	0.00	0.00	0.04	0.13	0.00	0.33	0.49	0.04	0.00	0.00	0.10	0.22	0.00	0.37	0.27
aRMSC(HQIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.97	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.98	0.00	0.00	0.02	0.00	0.00	0.00	0.00
	0.4	0.4	0.12	0.00	0.00	0.11	0.18	0.00	0.30	0.29	0.16	0.00	0.00	0.16	0.23	0.00	0.28	0.16
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.44	0.00	0.00	0.14	0.13	0.00	0.16	0.13	0.47	0.00	0.00	0.14	0.13	0.00	0.15	0.11
	0.4	0.4	0.01	0.00	0.00	0.03	0.11	0.00	0.32	0.53	0.03	0.00	0.00	0.08	0.21	0.00	0.38	0.30

Table 25: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of Two endogenous regressors (p=2) and T=50000

			T=50000															
			2SLS								LIML							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
aRMSC(AIC)	0	0.4	0.07	0.00	0.00	0.28	0.38	0.00	0.22	0.05	0.09	0.00	0.00	0.32	0.36	0.00	0.19	0.03
	0.1	0.4	0.07	0.00	0.00	0.28	0.38	0.00	0.22	0.05	0.09	0.00	0.00	0.32	0.36	0.00	0.19	0.03
	0.2	0.4	0.07	0.00	0.00	0.28	0.38	0.00	0.23	0.05	0.09	0.00	0.00	0.31	0.37	0.00	0.20	0.04
	0.3	0.4	0.04	0.00	0.00	0.21	0.37	0.00	0.29	0.08	0.06	0.00	0.00	0.26	0.37	0.00	0.25	0.06
aRMSC(BIC)	0	0.4	0.10	0.00	0.00	0.15	0.25	0.00	0.30	0.19	0.12	0.00	0.00	0.19	0.28	0.00	0.27	0.14
	0.1	0.4	0.10	0.00	0.00	0.15	0.25	0.00	0.30	0.20	0.12	0.00	0.00	0.19	0.28	0.00	0.27	0.14
	0.2	0.4	0.09	0.00	0.00	0.15	0.24	0.00	0.31	0.21	0.12	0.00	0.00	0.18	0.27	0.00	0.28	0.15
	0.3	0.4	0.05	0.00	0.00	0.09	0.17	0.00	0.32	0.35	0.08	0.00	0.00	0.13	0.22	0.00	0.32	0.25
aRMSC(HQIC)	0	0.4	0.36	0.00	0.00	0.22	0.20	0.00	0.16	0.07	0.39	0.00	0.00	0.23	0.19	0.00	0.14	0.05
	0.1	0.4	0.36	0.00	0.00	0.22	0.20	0.00	0.16	0.07	0.39	0.00	0.00	0.23	0.19	0.00	0.14	0.05
	0.2	0.4	0.36	0.00	0.00	0.21	0.20	0.00	0.16	0.07	0.39	0.00	0.00	0.23	0.19	0.00	0.14	0.05
	0.3	0.4	0.31	0.00	0.00	0.19	0.19	0.00	0.19	0.12	0.35	0.00	0.00	0.21	0.19	0.00	0.17	0.08
RMSC	0	0.4	0.08	0.00	0.00	0.14	0.25	0.00	0.32	0.22	0.10	0.00	0.00	0.18	0.28	0.00	0.30	0.15
	0.1	0.4	0.07	0.00	0.00	0.14	0.25	0.00	0.32	0.22	0.10	0.00	0.00	0.18	0.28	0.00	0.30	0.15
	0.2	0.4	0.07	0.00	0.00	0.13	0.24	0.00	0.33	0.24	0.09	0.00	0.00	0.17	0.27	0.00	0.30	0.16
	0.3	0.4	0.04	0.00	0.00	0.08	0.17	0.00	0.33	0.39	0.06	0.00	0.00	0.12	0.22	0.00	0.33	0.27
	$\delta_1 = \delta_2$																	
aRMSC(AIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.51	0.00	0.00	0.38	0.10	0.00	0.01	0.00	0.53	0.00	0.00	0.37	0.09	0.00	0.01	0.00
	0.4	0.4	0.00	0.00	0.00	0.05	0.23	0.00	0.43	0.29	0.01	0.00	0.00	0.10	0.32	0.00	0.39	0.18
aRMSC(BIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.63	0.00	0.00	0.12	0.09	0.00	0.09	0.07	0.65	0.00	0.00	0.12	0.09	0.00	0.08	0.06
	0.4	0.4	0.00	0.00	0.00	0.02	0.09	0.00	0.31	0.57	0.01	0.00	0.00	0.06	0.19	0.00	0.37	0.36
aRMSC(HQIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.4	0.4	0.07	0.00	0.00	0.08	0.17	0.00	0.32	0.36	0.11	0.00	0.00	0.13	0.23	0.00	0.31	0.23
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.51	0.00	0.00	0.13	0.12	0.00	0.13	0.11	0.53	0.00	0.00	0.13	0.12	0.00	0.12	0.09
	0.4	0.4	0.00	0.00	0.00	0.02	0.08	0.00	0.31	0.59	0.01	0.00	0.00	0.05	0.19	0.00	0.37	0.37

Table 26: Empirical Selection Probabilities of aRMSC (AIC,BIC,HQIC) in comparison to RMSC in the case of Two endogenous regressors (p=2) and T=100000

			T=100000															
			2SLS								LIML							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
aRMSC(AIC)	0	0.4	0.06	0.00	0.00	0.26	0.37	0.00	0.25	0.06	0.08	0.00	0.00	0.29	0.37	0.00	0.21	0.05
	0.1	0.4	0.06	0.00	0.00	0.26	0.37	0.00	0.25	0.06	0.08	0.00	0.00	0.29	0.37	0.00	0.21	0.05
	0.2	0.4	0.06	0.00	0.00	0.25	0.37	0.00	0.25	0.06	0.07	0.00	0.00	0.29	0.38	0.00	0.22	0.05
	0.3	0.4	0.03	0.00	0.00	0.19	0.36	0.00	0.32	0.10	0.05	0.00	0.00	0.24	0.38	0.00	0.27	0.07
aRMSC(BIC)	0	0.4	0.08	0.00	0.00	0.14	0.24	0.00	0.32	0.22	0.10	0.00	0.00	0.18	0.27	0.00	0.29	0.16
	0.1	0.4	0.08	0.00	0.00	0.14	0.24	0.00	0.32	0.22	0.10	0.00	0.00	0.18	0.27	0.00	0.29	0.16
	0.2	0.4	0.08	0.00	0.00	0.14	0.23	0.00	0.32	0.24	0.10	0.00	0.00	0.17	0.27	0.00	0.29	0.17
	0.3	0.4	0.04	0.00	0.00	0.08	0.17	0.00	0.33	0.38	0.06	0.00	0.00	0.12	0.22	0.00	0.32	0.27
aRMSC(HQIC)	0	0.4	0.33	0.00	0.00	0.22	0.21	0.00	0.17	0.08	0.36	0.00	0.00	0.23	0.21	0.00	0.14	0.06
	0.1	0.4	0.33	0.00	0.00	0.22	0.21	0.00	0.17	0.08	0.36	0.00	0.00	0.23	0.21	0.00	0.14	0.06
	0.2	0.4	0.32	0.00	0.00	0.22	0.21	0.00	0.17	0.08	0.36	0.00	0.00	0.23	0.20	0.00	0.15	0.06
	0.3	0.4	0.29	0.00	0.00	0.19	0.19	0.00	0.20	0.13	0.32	0.00	0.00	0.21	0.20	0.00	0.17	0.09
RMSC	0	0.4	0.07	0.00	0.00	0.13	0.24	0.00	0.33	0.24	0.09	0.00	0.00	0.17	0.27	0.00	0.30	0.17
	0.1	0.4	0.07	0.00	0.00	0.13	0.24	0.00	0.33	0.24	0.09	0.00	0.00	0.17	0.27	0.00	0.30	0.17
	0.2	0.4	0.06	0.00	0.00	0.12	0.23	0.00	0.33	0.25	0.08	0.00	0.00	0.16	0.27	0.00	0.31	0.18
	0.3	0.4	0.03	0.00	0.00	0.07	0.16	0.00	0.33	0.41	0.05	0.00	0.00	0.11	0.21	0.00	0.33	0.30
	$\delta_1 = \delta_2$																	
aRMSC(AIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.51	0.00	0.00	0.37	0.10	0.00	0.01	0.00	0.53	0.00	0.00	0.36	0.09	0.00	0.01	0.00
	0.4	0.4	0.00	0.00	0.00	0.04	0.20	0.00	0.43	0.34	0.01	0.00	0.00	0.08	0.28	0.00	0.41	0.22
aRMSC(BIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.63	0.00	0.00	0.12	0.09	0.00	0.09	0.07	0.65	0.00	0.00	0.12	0.09	0.00	0.09	0.06
	0.4	0.4	0.00	0.00	0.00	0.02	0.08	0.00	0.30	0.60	0.01	0.00	0.00	0.05	0.17	0.00	0.37	0.40
aRMSC(HQIC)	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.4	0.4	0.06	0.00	0.00	0.07	0.16	0.00	0.32	0.39	0.09	0.00	0.00	0.12	0.22	0.00	0.32	0.26
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.53	0.00	0.00	0.13	0.11	0.00	0.12	0.11	0.56	0.00	0.00	0.13	0.11	0.00	0.12	0.09
	0.4	0.4	0.00	0.00	0.00	0.01	0.07	0.00	0.29	0.62	0.01	0.00	0.00	0.05	0.16	0.00	0.37	0.41

Chapter 3

Instrument Selection in Monetary Policy Surprises

3.1 Introduction

In studies of macroeconomic and financial variables or combinations of both, researchers aim at formalizing the dynamic structure inherent to their interactions. Often, they need to account for recent past values of the aggregates as a policy shock may impact them over multiple future periods. A natural approach that offers economists the opportunity to introduce in their framework, simultaneously, impacts from lagged values and exogenous explanatory variables is the vector autoregressive (VAR) model. In general, from the estimated VAR models, residuals of the variable of interest are stimulated by one standard deviation or unity to generate a reaction from the different variables in the system. These reaction functions are the so-called impulse responses. They indicate to the policy maker the impact of shocks in terms of size, direction and length of propagation to other variables. In that context, a Cholesky decomposition is often used to disentangle shocks' impacts, however it requires identification of shocks that have contemporaneous effects on the others. Many papers like Eichenbaum and Evans (1995) achieve identification using this approach and assuming that US monetary policy shocks have no immediate effect on foreign interest rates. However, number of them acknowledge that various global rates contemporaneously react to US monetary policy shocks. In response, (Gertler and Karadi, 2015, hereafter GK), following Stock and Watson (2012) and Mertens and Ravn (2013) propose an identification approach of structural monetary policy shocks in a VAR model. Their method bypasses the recursive ordering issue relative to Cholesky decomposition, allowing for more flexibility.

Rogers et al. (2018) rely on that flexibility to study Unconventional Monetary Policy and International Risk using external instruments as identification method. They raise that this method is attractive because it does not impose the usual implausible short run restrictions required by the standard approach. Later, Husted et al. (2019) in their *event study* construct Monetary Policy Uncertainty (MPU) indexes, new measures of uncertainty about federal policy actions and their consequences. They show under a variety of VAR identification schemes including GK, that positive shocks to uncertainty about monetary policy robustly raise credit spreads and reduce output.

Our study revisits GK's findings, with a focus on the two year government bond rate (2YR) as policy indicator. Our choice of the 2YR results from arguments in the paper, in favor of rates with longer than one year horizon. GK thought that, as policy indicator, the 2YR could be of interest but it exhibits a first stage F statistic lower than 5.2 for a combination of all instruments while a value above 10 is the recommended by the rule of thumb. Several consequences of weak correlation between the endogenous variable and its instruments have been discussed, including bias and inconsistency of estimators, fallacious t-type tests and Wald-type tests which are then badly approximated by standard asymptotics.

Few well known empirical examples of weak identification issues are first, Angrist and Krueger (1991) study of return to education using quarter of birth and its interactions as instruments for educational attainment. They obtained estimators with poor asymptotic inferences. Their findings are confirmed by Bound et al. (1995) who demonstrate that despite Angrist and Krueger (1991) large sample size data (329,509 observations), their estimates were biased and inconsistent as a result of the weak correlation between the endogenous variable and its instruments.

A second example pertains to the study of Galí and Gertler (1999) on modeling inflation dynamics using the new Keynesian Phillips curve and US post-war data. Their results suggest that while real marginal costs are statistically significant, inflation dynamics are largely forward-looking. Kleibergen et al. (2009) argue that those results are unreliable because marginal costs' coefficients are close to zero. This leads the new Keynesian Phillips curve to be flat as a result of limited exogenous variation in inflation forecasts. Canova and Sala (2009) and Nason and Smith (2008) support this evidence.

As a final illustration, we consider Campbell (2003) who estimates the elasticity of inter temporal substitution between consumption growth and the return on some assets given a set of instruments. He noticed that confidence intervals are sensitive to reverse regression even when errors are homoskedastic.

In this chapter, while we illustrate the empirical performance of an information criterion robust to weak instruments and the importance of optimal selection in instrumental variables regression in presence of weak instruments, we study high frequency identification and monetary policy surprises on credit costs. The plan for the remainder of this paper is as follows. In the next section we describe the VAR methodology and the high frequency identification approach. In section 3, we present the data we use in the empirical analysis. In section 4, we report our empirical results and conclude in Section 5.

3.2 VAR Methodology and Identification of Monetary Policy Surprises

To study the mechanism of monetary policy transmission, we need to estimate the effects of monetary policy shocks on a mixture of economic and financial variables. We estimate a vector autoregressive model as this is conventionally done in the literature. In that purpose, we follow GK who identified monetary surpises using external instruments in their VAR model, namely, High Frequency Identification of policy shocks.

Their methodology which is a variation of Stock and Watson (2012) and Mertens and Ravn (2013) consider Y_t to be the vector of economic and financial variables, A nonsingular and $C_j \ \forall \ j \geq 1$ conformable coefficient matrices, and ε_t a vector of structural shocks. The general structural form of the subsequent VAR is given by

$$AY_t = \sum_{j=1}^{p} C_j Y_{t-j} + \varepsilon_t \tag{3.1}$$

Multiplying each side of the equation by A^{-1} yields the following reduced form representation

$$Y_t = \sum_{j=1}^{p} B_j Y_{t-j} + u_t \tag{3.2}$$

where $u_t = S\varepsilon_t$ is the reduced form shock, with $B_j = A^{-1}C_j, S = A^{-1}$ and p is the number of lags considered for the VAR model.

The expression of u_t implies that the reduced form errors can be related to a set of underlying structural shocks ε_t. Let s denote the column in matrix S corresponding to the impact on each element of the vector of reduced form residuals u_t of the structural policy shock ε_t. Accordingly, we need to estimate the following equation to compute various impulse responses to monetary shocks

$$Y_t = \sum_{j=1}^{p} B_j Y_{t-j} + S\varepsilon_t \tag{3.3}$$

As we extensively discussed in our first chapter, the ordering of variables is crucial when a Cholesky decomposition is used for identification, it requires to impose short-run restrictions. Reversely, the ordering is irrelevant when the adopted identification approach involves external instruments. Indeed, because GK were not interested in computing a variance decomposition or the impulse responses to other shocks, they conclude that it is unnecessary to identify all the coefficients of the matrix S, but rather only the elements of the column s.

Let Z_t be a vector of instrumental variables and consider a partition of ε_t as $(\varepsilon_t^p, \varepsilon_t^{q\prime})'$ where ε_t^p is the monetary policy shock and ε_t^q is the vector of all other shocks. As it is required in the instrumental variable literature, it is fairly realistic to assume that Z_t satisfy respectively the following validity and relevance conditions:

$$\begin{cases} E\left(Z_t \varepsilon_t^{q\prime}\right) = 0 \\ E\left(Z_t \varepsilon_t^{p\prime}\right) = \phi \end{cases} \tag{3.4}$$

That is Z_t is correlated with the monetary policy shock and orthogonal to all other structural shocks.

To obtain the estimate of s we follow GK:

1. We first estimate the VAR model (3.2) and obtain the vector $\widehat{u_t}$ of reduced form residuals.

2. Then we let $\widehat{u_t^p}$ be the reduced form residual from the equation for the policy indicator, $\widehat{u_t^q}$ the reduced form residuals from all the other equations excluding the policy indicator and s^q be the associated elements of s corresponding to a unit increase in the policy shock ε_t^p. It is now possible to estimate the ratio s^q/s^p from the two stage least squares regression of $\widehat{u_t^q}$ on $\widehat{u_t^p}$, using the instrument set Z_t.

Regarding the model selection part, recall that in our chapter 2 we constructed the aRMSC in a linear regression case, as a result it easily applies to the high frequency identification framework. We summarize the two stage least squares system of equations below:

$$\begin{cases} \widehat{u_t^q} = \frac{s^q}{s^p}\widehat{u_t^p} + \eta_t \\ \widehat{u_t^p} = \Pi Z_t + e_t \end{cases} \tag{3.5}$$

where η_t and e_t are the usual error terms.

In the first stage of the two stage least squares estimation, the authors regress $\widehat{u_t^p}$ on Z_t, the obtained fitted value, $\widehat{\widehat{u_t^p}}$, intuitively captures the impact of news about monetary policy that is carried by the high frequency instruments Z_t. Given that the variation in $\widehat{u_t^p}$ are only generated by ε_t^p (see equation 3.3), then regressing u_t^q on the fitted value of $\widehat{u_t^p}$ at the second stage yields a consistent estimate of the ratio s^q/s^p.

The aRMSC in this context writes:

$$aRMSC(c) = \ln\left|\widehat{V}(c)\right| + \frac{1}{2} \times \left(1 - \frac{p}{|c|}\right) \times \frac{(\ln T)^2}{\sqrt{T}} \tag{3.6}$$

where T is the sample size, c is a selection vector of dimension $k_{\max} \times 1$ ($k_{\max}$ being the maximal number of instruments in the candidate set) for which the elements take the values of 0 or 1. At row i, $c_i = 1$ means that the instrument z_i is selected and $c_i = 0$ implies that the instrument is not included in the model. $|c|$ is the number of instruments of the candidate set included in the model and is always greater or equal to p the number of endogenous variables. In the current application, $\widehat{V}(c)$ would correspond to an estimator of the asymptotic variance of the 2SLS estimator and $p = 1$ as it is either the one year rate or the two year rate.

Finally following GK, s^p is derived by estimating the variance-covariance matrix of the reduced form residuals of the VAR model (3.2) and exploiting the HFI approach to recursively determine its value. This process leads to the identification of exogenous monetary policy surprises (see GK for further details).

3.3 Data

Following GK, we examine the response of various market interest rates to surprises in various policy indicators, using interest rate futures surprises on FOMC dates as instruments. Specifically, the surprise is measured by the change in expectation of future rates within a short term window bracketing the time of the monetary policy announcement. To ensure that this change of expectation results only from news about the FOMC decision, the surprises in futures rates are usually measured within a thirty minutes window of the announcement.

For the policy indicators, we consider mainly the one year government bond rate and the two year government bond rate. We also consider the same instruments as GK: the surprise in the current Federal Funds futures rate (FF1), the surprise in the three month ahead futures rate (FF4); and the full (Gürkaynak et al., 2005, hereafter GSS) instruments set, the six month (ED2), nine month (ED3) and one year (ED4) ahead futures on three month Eurodollar deposits.

We use a wide range of economic and financial variables over the period of July 1979 to June 2012. In particular, we estimate a monthly VAR model with 12 lags as in GK. Unfortunately, the instruments are only available in the study from January 1991 through June 2012. Even though this is a shorter sample period than the one available for the full VAR model, we assume that instruments selected over this period would remain the same when more data is available and did not constrained the VAR to the instruments shorter length length.

Similar to Rogers et al. (2018) we measure U.S. monetary policy shocks at the zero lower bound including the 2008 crisis period. GK highlight concerns that over zero lower bound period, the central bank might have limited capability in leveraging over the one year government bond rate as instrument. Swanson and Williams (2014) make the case that this constraint would be less efficient when it comes to the the two year rate. Indeed, GK argue, using a safe interest rate with a longer maturity than the Fed Funds rate would allow us to account for shocks to forward guidance in the overall measure of monetary policy shocks. Bernanke et al. (2004), GSS, Hanson and Stein (2015) and others also found evidence that supports the argument that the Federal Reserve forward guidance strategy operates with a roughly two year horizon.

As a result, while GK's conceptually preferred indicator was the two year government bond rate, they found following Stock et al. (2002) rule of thumb recommending a threshold value of ten for the first stage F statistic, that a weak instrument problem might be present. Reversely, the one year government bond rate exhibits F statistic values safely above the threshold. Even when using the GSS instruments set that had strong explanatory power in the high frequency data, none of the instruments' combinations meet the threshold for the two year rate. Therefore, GK safely moved towards the one year rate to avoid weak identification issues.

Essentially, our study revisits results of GK with an emphasis on the instruments selection process for the one year and two year government bond rates. Indeed, we expect that using our alternative Relevant Moment Selection Criterion (aRMSC) that is more robust to weaker instruments, we would improve the instruments' selection and by the same opportunity the impulse responses. Following GK, we estimate a monthly VAR model that includes two economic variables, log of the industrial production and the log of the consumer price index, the two year government bond rate (the policy indicator), and a credit spread,specifically the excess bond premium of Gilchrist and Zakrajšek (2012).

3.4 Empirical Results

3.4.1 Optimal Instruments Selection

We consider all five possible instruments: FF1, FF4, ED2, ED3, ED4 in the cases were the policy indicator is either the one year government bond rate or the two year government bond rate. This implies that they will represent full set of candidate instruments included in Z_t in equation (3.5). We explore all possible combinations for them and select the one that minimizes our information criterion, the aRMSC. More specifically, we conduct instrument selection using the reduced form VAR residuals of each of non-policy indicator variables: the logarithm of the consumer price index (CPI), the logarithm of the industrial production (IP) and the excess bond premium (EBP). As a result, we determine the best instrument combination for each of them and compared the optimal sets.

The results in the case of the one year government bond rate recommend the FF4 as best instruments to capture monetary policy surprises (see Table 27). This confirms the choice made by GK to use FF4 as instrument when the policy indicator is the one year rate. Also, as the F statistic of the first stage regression in this case was safely above the threshold of 10, the model FF4 would be exempt of weak instrument problems.

In the case of the two year government bond rate as policy indicator, the authors naively consider the full set of GSS instruments to compare its impulse responses with that of the one year rate with FF4 as instrument. It is common in empirical studies to assume that it is always better to include the larger number of instruments even when they are redundant and do not significantly improve the model. The perspective being that it is not very costly to use large number of instruments when there is enough degree of freedom, when the population size is large enough.

Indeed, when repeating the same selection process as in the case of the one year rate, we did not obtain the full set of GSS as best set of instruments. Instead, the aRMSC recommends the combination of FF4 and ED3 as best set of instruments for the two year rate policy indicator.

In its construction, the first stage F statistic utilizes the mean squared errors information as input to determine a best set of instruments. However, it is only reliable for inference when the statistic is greater than or equal to 10. When this condition is not satisfied, authors recommend to be cautious. For example, Kleibergen et al. (2009) show that in presence of weak instruments, the rule of Stock and Yogo (2005) is not verified and the test has less power than the Anderson and Rubin statistic which is more identification robust. As a result, standard inferential results are fallacious. These suggest that the obtained explanatory power for the individual instruments of the GSS set by GK might not be coherent with the reality.

The weak instrument issues are confirmed in our selection results, when we attempt selection using classical information criteria, Akaike, Schwartz and Hanan-Quinn. We notice that the results are not unanimous, the criteria seem confused between recommending the full candidate set as best instruments set and the full candidate set excluding FF1. In particular, using the 2SLS estimator they suggest for EBP that the best instrument set is the full candidate set but using the LIML estimator, the FF1 is excluded from the selected set. This divergence of choice might be explained by the weak correlation of instruments with the endogenous variables. Nevertheless, the fact that classical criteria are suggesting all instruments imply that they might all be valid but not guaranteed to all be relevant. Indeed, in empirical studies the relevance condition is often omitted or overlooked. In our following analysis of monetary policy shocks we will focus on the importance of choosing optimal instruments set.

3.4.2 Monetary Policy Shocks

Impact of Policy Indicator Choice

In GK, the results of the VAR estimated with the one year government bond rate are discussed in details. We do not find necessary to present all the advantages of the external instruments approach in comparison to the Cholesky identification method as this is covered in GK. We jump to the comparison of the impulse responses generated with the VAR model based on the two year government bond rate and its optimal set of instruments with the model with the one year rate and FF4 as instrument.

Figure 13 displays on the same graphs the impulse responses of our replication of GK's results (with the one year rate as policy indicator and the three months ahead Funds rate future surprise) in solid line and that of our two year rate policy indicator in dashed line. Reporting the impulse responses for both the economic and financial on the same graphs allow us to compare at the same time level and dynamics. The top sub figure shows that the one year rate and the two year rate exhibit very similar values. Indeed, it shows that a one standard deviation surprise monetary tightening in either one of both policy indicators generates around 25 basis points increase. The curves remain very close until the 24th month where the 2YR starts to slightly dominate the 1YR.

In the second panel, consistent with conventional theory, we observe a small decline in the consumer price index roughly by 4 basis points for the 1YR versus 6 basis points for the 2YR case. This gap between the two impulse responses grow bigger and reaches roughly 6 basis points after two years and half. Relative to the impact of the shock this is a pretty significant difference although both policy indicators present strongly identical dynamic. Similarly consistent with standard theory, we notice in the third sub figure, an important decline in industrial production with a minimum around a year and half like GK but with 5 basis points gap between the 1YR and the 2YR. Remark that the curves are almost identical for the first quarter, then they diverge before re converging around the forty fifth month horizon.

Finally, in the excess bond premium reaction function, there is an increase, on the impact of roughly 11 basis points for the 1YR versus 16 basis points for the 2YR. In this last sub figure, the curves diverge over the first year before converging after the month number 15. In any case, the two year rate reports a larger impact on the various economic and financial variables than the one year rate. It is hard to conclude which impulse response is more accurate but the findings in terms of directions of the shocks, that are very similar to the combination of the one year rate and the three months ahead Funds rate future strong instrument are very encouraging. Also, an interesting question is to measure monetary policy surprises resulting from the combination of multiple policy indicators, since the aRMSC offers the flexibility to select instruments in presence of multiple endogenous variables, we leave it for future research.

Impact of Optimal Instruments Choice

We seize the opportunity in the current framework to contemplate the impact of the selection of optimal instruments versus the naive inclusion of all instruments. Some authors recommend to include all instruments in the model when it is not very costly to do so. Indeed, in our application the candidate set only contains five potential instruments, one might be tempted to rely on all of them to construct the impulse responses whether it is appropriate to proceed in such a manner or not. In particular, in presence weak instruments issues, the extensive selection results of the Monte Carlo simulations, presented in the chapter 2 suggest caution.

We present in Figures 14 and 15, monetary policy shocks for the CPI, IP and EBP when the 2 year rate is used as policy indicator. We consider both the 2SLS and the LIML estimators, when the optimal instruments are included in the model as opposed to the case when all instruments are naively included.

Two main observations are noticeable. In Figure 14, we notice using the 2SLS estimator with all instruments that the impact of shocks are largely under estimated in comparison to the optimal set of instruments. This difference in sizes of the impacts generates accuracy concerns. Specifically, in the case of IP both shocks start with opposite signs. While the expected sign should initially be positive before decreasing as observed in the GK's 1YR policy shock impact on the industrial production (see Figure 12).

The second observation displayed in Figure 15, reports that when the selected instruments are optimal the impulse responses based on the 2SLS and the LIML estimators exhibit a relatively small gap over the 48 impulse responses periods. Meanwhile, when using all instruments in the candidate set they only converge from time to time. Additionally, here the sample size for instrument selection is 258 observations which is relatively large for macroeconomics applications. While, it is expected under mild conditions that the LIML and the 2SLS converge in larger samples, in our application, it is definitely not the case when using a non optimal set of instruments.

3.5 Conclusion

In this chapter, we review GK's study of monetary policy surprises using the external instrument approach. While the authors mentioned their preference for the 2YR as policy indicator over the 1YR, they were forced to adopt the latter because the 2YR was weakly correlated with the set of instruments available to them. Indeed, the FF1 was found to be the best instrument for the 1YR as its associated first stage F statistic was safely above 10 as recommended by Stock et al. (2002). We leverage our weak instruments robust information criterion , the alternative Relevant Moment Selection Criterion, to analyze the case of the 2YR as policy indicator. we also compare our results to those of GK in terms of monetary shocks and the impact of better instrument selection process.

Overall our findings can be summarized in three categories. First, using the aRMSC, we confirm that the best instrument for the 1YR is effectively the three months ahead Funds rate future. This implies that is efficient for selection when instruments are strong. It keeps the same properties as the RMSC but performs better in presence of weak instruments. For the 2YR, as opposed to GK who use the full GSS set of instruments, the aRMSC recommends FF4 combined with ED3 as best instruments set. Additionally, tracing out the VAR dynamics resulting from a shock on one of both policy indicator, we notice that all the impulse responses produced with the 2YR show similar variations as the 1YR (in terms of shape). However, in terms of size, the impact of shocks from the 2YR are always larger than those from the one year rate. We do not claim more accuracy in the 2YR than in the 1YR. Nevertheless, the lower F statistic will probably be less considered a hard threshold in the future, as consistent instrument selection is possible when exploring monetary policy shocks.

Secondly, the naive inclusion of all instruments in the candidate set, may not always improve estimation in presence of weak instruments. Indeed, the impulse responses obtained when including optimal instruments in the model versus the full candidate set are quite different. The driven wedge between those impulse responses is not reduced even when we consider the LIML estimator, member of the k-class estimators known to be more robust to weak instruments. Finally, the 2SLS and LIML estimators approaches generate much closer impulse responses when the optimal number of instruments is used. Although, this is a totally empirical finding it would be of interest to reflect on the question and this is left for future research.

3.6 Appendix

3.6.1 Appendix A: Figures

Figure 12: 1YR and 2YR shocks using HFI compared with Cholesky restrictions

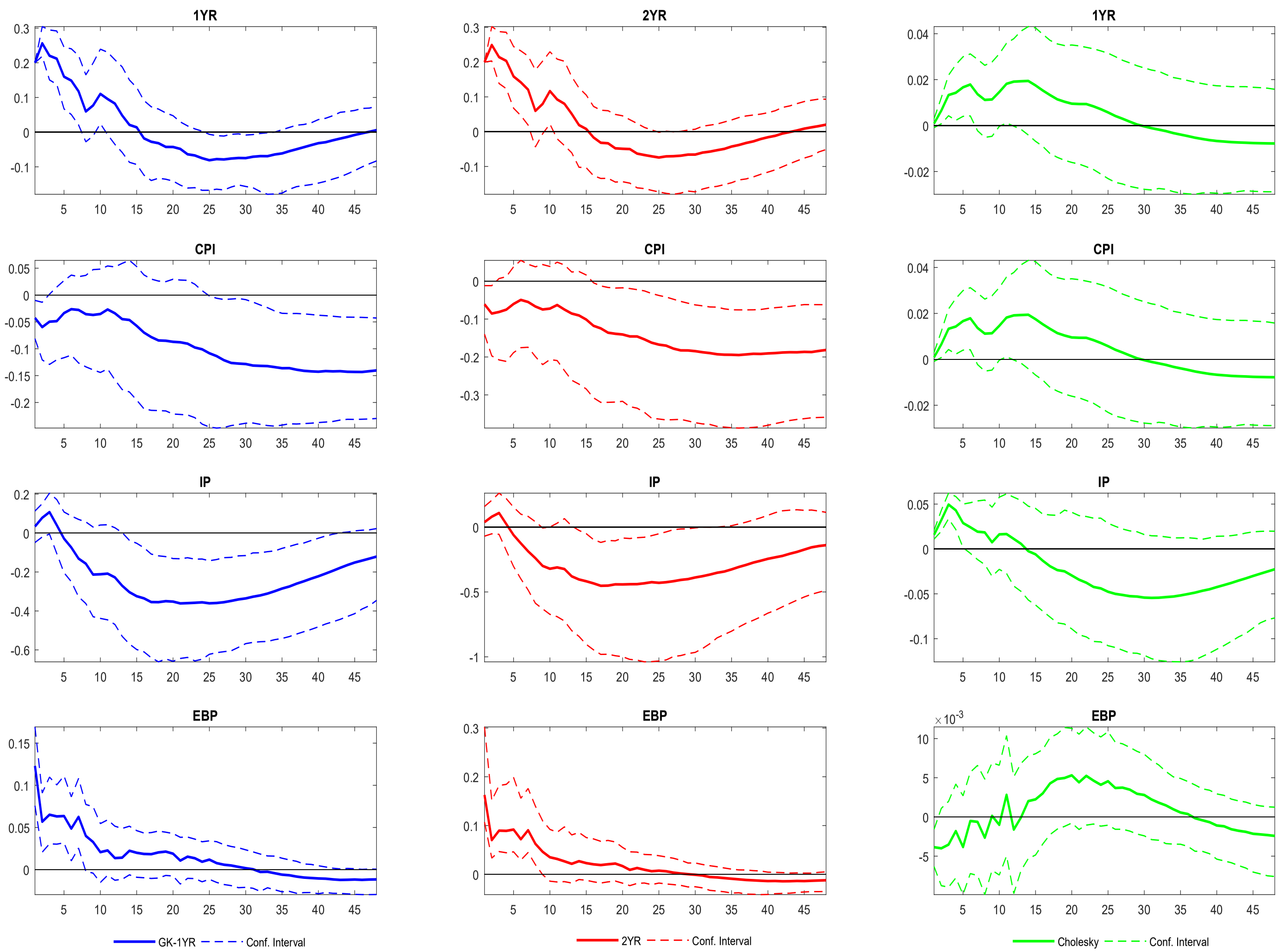


Figure 13: 2 year rate shock with consumer price index, industrial production and excess bond premium.

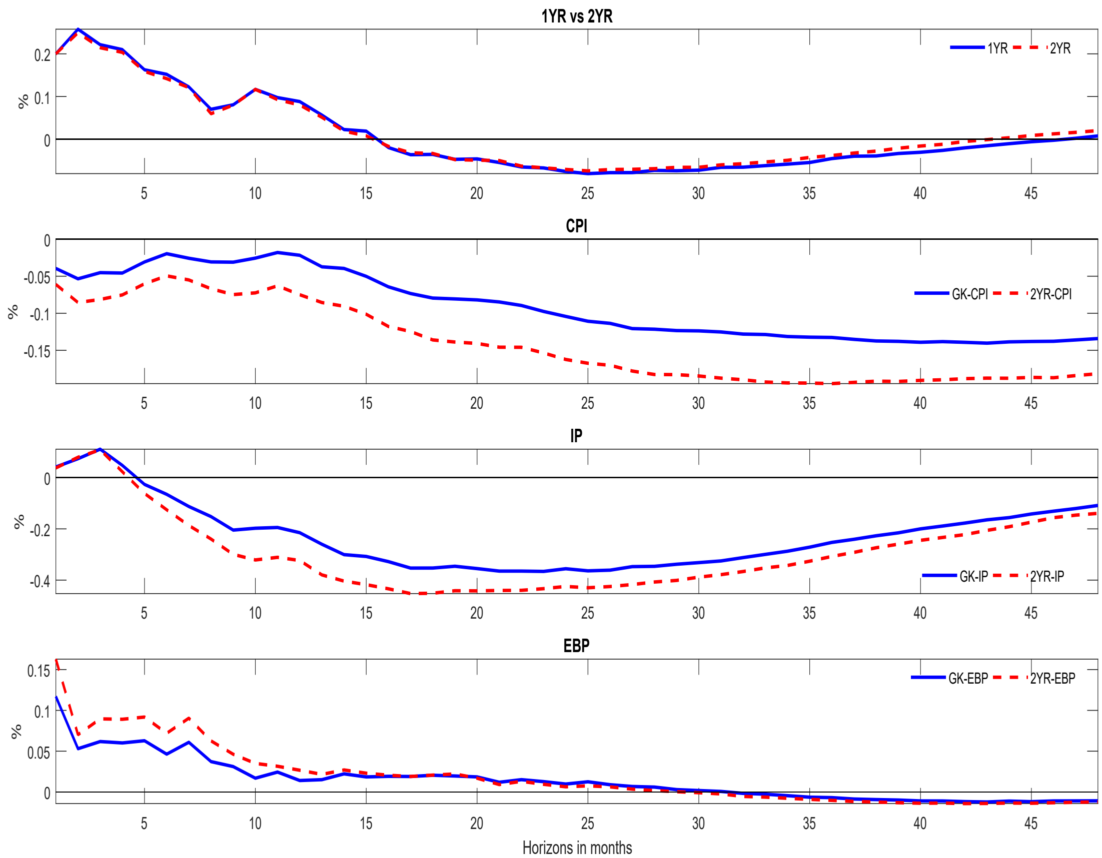


Figure 14: 2 year rate shock with 2SLS estimator and Optimal versus All instruments.

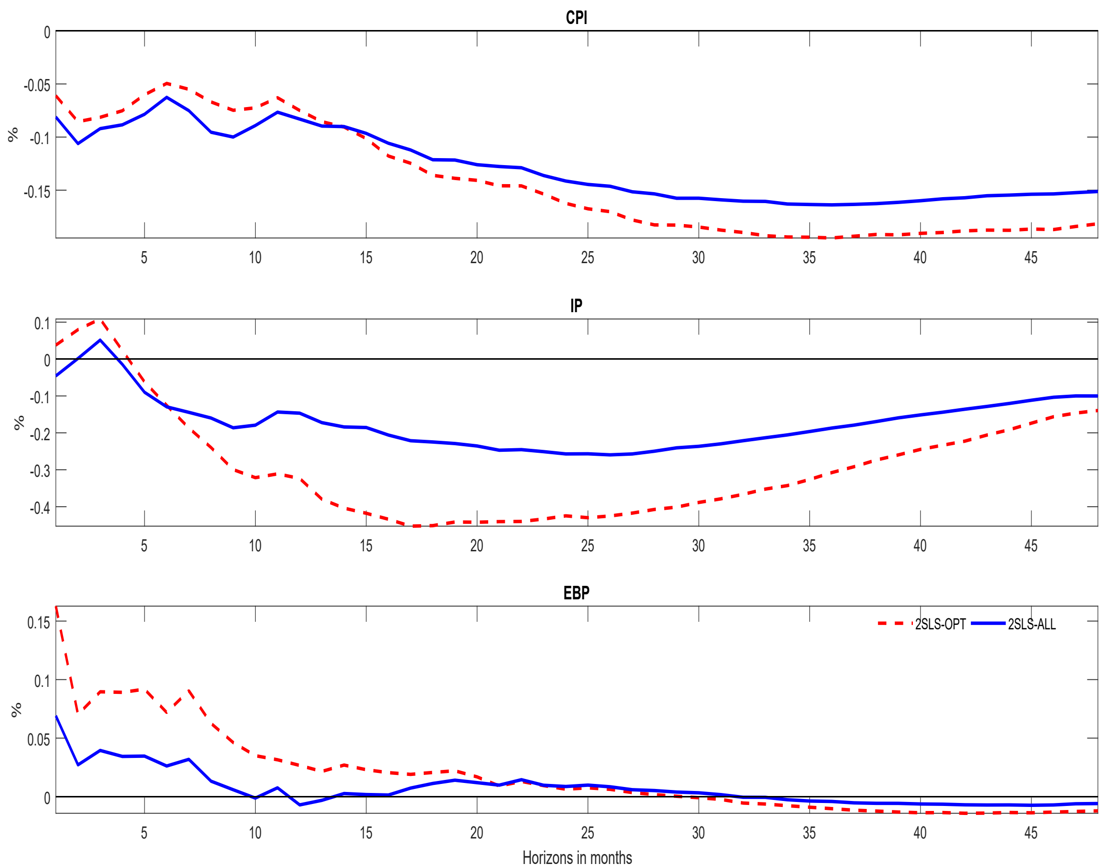
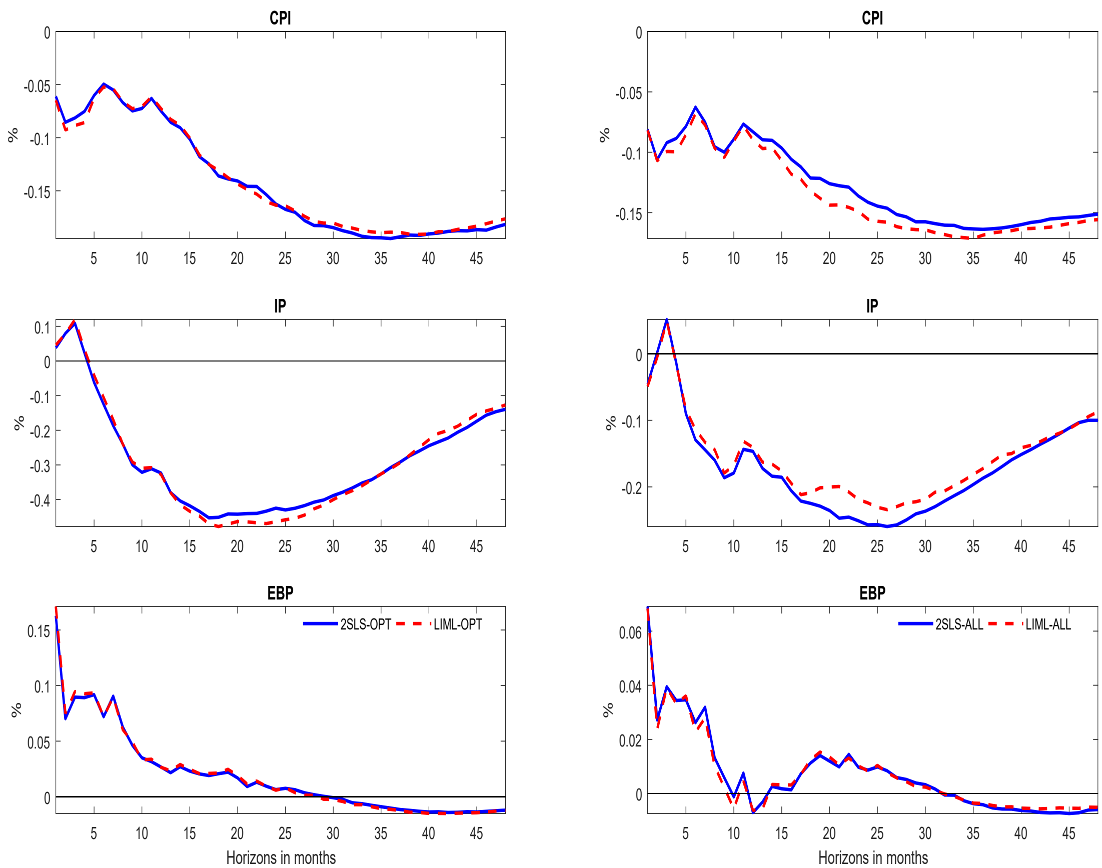


Figure 15: 2 year rate shock with 2SLS versus LIML estimator for Optimal and All instruments.



3.6.2 Appendix B: Tables

Table 27: Instruments Selection with AIC, BIC, HQIC and aRMSC

Information Criteria	1YR			2YR (2SLS)			2YR (LIML)		
	CPI	IP	EBP	CPI	IP	EBP	CPI	IP	EBP
Classical AIC (Akaike)	All	All	All	All	Except FF1	All	All	Except FF1	Except FF1
Classical BIC (Schwartz)	All	All	All	All	Except FF1	All	All	Except FF1	Except FF1
Classical HQIC (Hannan-Quinn)	All	All	All	All	Except FF1	All	All	Except FF1	Except FF1
aRMSC	FF4	FF4	FF4	FF4+ED3	FF4+ED3	FF4+ED3	FF4+ED3	FF4+ED3	FF4+ED3

Bibliography

N. A. Ahmed and D. V. Gokhale. Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory*, 35(3):688–692, May 1989. ISSN 0018-9448. doi: 10.1109/18.30996.

Donald W. K. Andrews. Consistent Moment Selection Procedures for Generalized Method of Moments Estimation. *Econometrica*, 67(3):543–563, May 1999. ISSN 1468-0262. doi: 10.1111/1468-0262.00036.

Donald W. K. Andrews and Xu Cheng. Estimation and Inference with Weak, Semi-Strong, and Strong Identification. *Econometrica*, 80(5):2153–2211, September 2012. ISSN 0012-9682.

Joshua D. Angrist and Alan B. Krueger. Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991. ISSN 0033-5533. doi: 10.2307/2937954.

Bertille Antoine and Eric Renault. Efficient GMM with nearly-weak instruments. *Econometrics Journal*, 12:S135–S171, January 2009. ISSN 1368-423X. doi: 10.1111/j.1368-423X.2009.00279.x.

Bertille Antoine and Eric Renault. Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics*, 170(2):350–367, 2012. ISSN 0304-4076.

Richard T. Baillie, G. Geoffrey Booth, Yiuman Tse, and Tatyana Zabotina. Price discovery and common factor models. *Journal of Financial Markets*, 5(3):309–321, July 2002. ISSN 1386-4181. doi: 10.1016/S1386-4181(02)00027-7.

Rolf W. Banz. The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18, March 1981. ISSN 0304-405X. doi: 10.1016/0304-405X(81)90018-0.

Ole E. Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. *Econometrica*, 76(6):1481–1536, November 2008. ISSN 1468-0262. doi: 10.3982/ECTA6495.

Paul A. Bekker. Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, 62(3):657–681, 1994. ISSN 0012-9682. doi: 10.2307/2951662.

Anthony J. Bell and Terrence J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159, November 1995. ISSN 0899-7667. doi: 10.1162/neco.1995.7.6.1129.

A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica*, 80(6):2369–2429, 2012. ISSN 1468-0262. doi: 10.3982/ECTA9626.

Itzhak Ben-David, Francesco Franzoni, Rabih Moussawi, et al. *ETFs, Arbitrage, and Contagion*. University of Geneva, 2011.

Ben S. Bernanke, Vincent R. Reinhart, and Brian P. Sack. Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment. SSRN Scholarly Paper ID 632381, Social Science Research Network, Rochester, NY, September 2004.

Roberto Blanco, Simon Brennan, and Ian W. Marsh. An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps. *The Journal of Finance*, 60(5):2255–2281, October 2005. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2005.00798.x.

Ekkehart Boehmer and Juan (Julie) Wu. Short Selling and the Price Discovery Process. *The Review of Financial Studies*, 26(2):287–322, February 2013. ISSN 0893-9454. doi: 10.1093/rfs/hhs097.

John Bound, David A. Jaeger, and Regina M. Baker. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995. ISSN 0162-1459. doi: 10.2307/2291055.

Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High-Frequency Trading and Price Discovery. *The Review of Financial Studies*, 27(8):2267–2306, August 2014. ISSN 0893-9454. doi: 10.1093/rfs/hhu032.

John Y. Campbell. Chapter 13 Consumption-based asset pricing. In *Handbook of the Economics of Finance*, volume 1 of *Financial Markets and Asset Pricing*, pages 803–887. Elsevier, January 2003. doi: 10.1016/S1574-0102(03)01022-7.

Mehmet Caner. Testing, Estimation in GMM and CUE with Nearly-Weak Identification. *Econometric Reviews*, 29(3):330–363, December 2009. ISSN 0747-4938. doi: 10.1080/07474930903451599.

Fabio Canova and Luca Sala. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449, 2009. ISSN 0304-3932.

J. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, April 1997. ISSN 1558-2361. doi: 10.1109/97.566704.

Yiu-ming Cheung and Lei Xu. Independent component ordering in ICA time series analysis. *Neurocomputing*, 41(1–4):145–152, October 2001. ISSN 0925-2312. doi: 10.1016/S0925-2312(00)00358-1.

Pierre Comon. Higher Order StatisticsIndependent component analysis, A new concept? *Signal Processing*, 36(3):287–314, April 1994. ISSN 0165-1684. doi: 10.1016/0165-1684(94)90029-9.

Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley \& Sons, 2012.

Werner F. M. De BONDT and Richard Thaler. Does the Stock Market Overreact? *The Journal of Finance*, 40(3):793–805, July 1985. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1985.tb05004.x.

Stephen G. Donald and Whitney K. Newey. Choosing the Number of Instruments. *Econometrica*, 69(5):1161–1191, 2001. ISSN 0012-9682.

Prosper Dovonon, Firmin Doko Tchatoka, and Michael Aguessy. Relevant Moment Selection under mixed Identification Strength, April 2020.

Jean-Marie Dufour and Mohamed Taamouti. Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments. *Econometrica*, 73(4):1351–1365, 2005. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2005.00618.x.

Martin Eichenbaum and Charles Evans. Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates. *The Quarterly Journal of Economics*, 110(4):975–1009, 1995.

Eugene F. Fama. Market efficiency, long-term returns, and behavioral finance1. *Journal of Financial Economics*, 49(3):283–306, September 1998. ISSN 0304-405X. doi: 10.1016/S0304-405X(98)00026-9.

Isabel Figuerola-Ferretti and Jesús Gonzalo. Modelling and measuring price discovery in commodity markets. *Journal of Econometrics*, 158(1):95–107, September 2010. ISSN 0304-4076. doi: 10.1016/j.jeconom.2010.03.013.

Jordi Galí and Mark Gertler. Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222, October 1999. ISSN 0304-3932. doi: 10.1016/S0304-3932(99)00023-9.

Mark Gertler and Peter Karadi. Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76, January 2015. ISSN 1945-7707. doi: 10.1257/mac.20130329.

Simon Gilchrist and Egon Zakrajšek. Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, 102(4):1692–1720, June 2012. ISSN 0002-8282. doi: 10.1257/aer.102.4.1692.

Jesus Gonzalo and Clive Granger. Estimation of Common Long-Memory Components in Cointegrated Systems. *Journal of Business & Economic Statistics*, 13(1):27–35, January 1995. ISSN 0735-0015. doi: 10.2307/1392518.

Joachim Grammig and Franziska J. Peter. Telltale Tails: A New Approach to Estimating Unique Market Information Shares. *Journal of Financial and Quantitative Analysis*, 48(02):459–488, April 2013. ISSN 1756-6916. doi: 10.1017/S0022109013000215.

Joachim Grammig, Michael Melvin, and Christian Schlag. Internationally cross-listed stock prices during overlapping trading hours: Price discovery and exchange rate effects. *Journal of Empirical Finance*, 12(1):139–164, 2005. ISSN 0927-5398.

Refet S. Gürkaynak, Brian Sack, and Eric Swanson. The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models. *American Economic Review*, 95(1):425–436, March 2005. ISSN 0002-8282. doi: 10.1257/0002828053828446.

Jinyong Hahn and Jerry Hausman. A New Specification Test for the Validity of Instrumental Variables. *Econometrica*, 70(1):163–189, 2002. ISSN 0012-9682.

Jinyong Hahn and Guido Kuersteiner. Discontinuities of weak instrument limiting distributions. *Economics Letters*, 75(3):325–331, May 2002. ISSN 0165-1765. doi: 10.1016/S0165-1765(01)00622-X.

Alastair R. Hall and Fernanda P. M. Peixe. A Consistent Method for the Selection of Relevant Instruments. *Econometric Reviews*, 22(3):269–287, January 2003. ISSN 0747-4938. doi: 10.1081/ETC-120024752.

Alastair R. Hall, Atsushi Inoue, Kalidas Jana, and Changmock Shin. Information in generalized method of moments estimation and entropy-based moment selection. *Journal of Econometrics*, 138(2):488–512, June 2007. ISSN 0304-4076. doi: 10.1016/j.jeconom.2006.05.006.

Samuel Hanson and Jeremy Stein. Monetary Policy and Long-Term Real Rates. *Journal of Financial Economics*, 115(3):429–448, 2015.

Joel Hasbrouck. Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement. *Review of Financial Studies*, 6(1):191–212, January 1993. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/6.1.191.

Joel Hasbrouck. One Security, Many Markets: Determining the Contributions to Price Discovery. *The Journal of Finance*, 50(4):1175–1199, September 1995. ISSN 0022-1082. doi: 10.2307/2329348.

Joel Hasbrouck. Stalking the "efficient price" in market microstructure specifications: An overview. *Journal of Financial Markets*, 5(3):329–339, July 2002. ISSN 1386-4181. doi: 10.1016/S1386-4181(02)00029-0.

Roger D. Huang. The Quality of ECN and Nasdaq Market Maker Quotes. *The Journal of Finance*, 57(3):1285–1319, June 2002. ISSN 1540-6261. doi: 10.1111/1540-6261.00461.

Peter J. Huber. Projection Pursuit. *The Annals of Statistics*, 13(2):435–475, 1985. ISSN 0090-5364.

Lucas Husted, John Rogers, and Bo Sun. Monetary policy uncertainty. *Journal of Monetary Economics*, July 2019. ISSN 0304-3932. doi: 10.1016/j.jmoneco.2019.07.009.

A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, June 2000. ISSN 0893-6080. doi: 10.1016/S0893-6080(00)00026-5.

Aapo Hyvarinen. New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing Systems*, pages 273–279, 1998.

Aapo Hyvärinen. Independent component analysis: Recent advances. *Phil. Trans. R. Soc. A*, 371 (1984):20110534, February 2013. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2011.0534.

Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, April 2004. ISBN 978-0-471-46419-8.

M. C. Jones and Robin Sibson. What is Projection Pursuit? *Journal of the Royal Statistical Society. Series A (General)*, 150(1):1–37, 1987. ISSN 0035-9238. doi: 10.2307/2981662.

Frank De Jong and Peter C. Schotman. Price Discovery in Fragmented Markets. *Journal of Financial Econometrics*, 8(1):1–28, January 2010. ISSN 1479-8409, 1479-8417. doi: 10.1093/jjfinec/nbp015.

Christian Jutten and Jeanny Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991. ISSN 0165-1684. doi: 10.1016/0165-1684(91)90079-X.

J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo. A class of neural networks for independent component analysis. *IEEE Transactions on Neural Networks*, 8(3):486–504, May 1997. ISSN 1045-9227. doi: 10.1109/72.572090.

Frank Kleibergen. Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70(5):1781–1803, 2002. ISSN 0012-9682.

Frank Kleibergen, Sophocles Mavroeidis, Fabio Canova, John C. Chao, Norman R. Swanson, Jean-Marie Dufour, Anna Mikusheva, Jonathan H. Wright, Motohiro Yogo, Eric Zivot, and Saraswata Chaudhuri. Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve [with Comments and Rejoinder]. *Journal of Business & Economic Statistics*, 27(3):293–339, 2009. ISSN 0735-0015.

Bruce N. Lehmann. Some desiderata for the measurement of price discovery across markets. *Journal of Financial Markets*, 5(3):259–276, July 2002. ISSN 1386-4181. doi: 10.1016/S1386-4181(02)00025-3.

Donald Lien and Keshab Shrestha. A new information share measure. *Journal of Futures Markets*, 29(4):377–395, April 2009. ISSN 1096-9934. doi: 10.1002/fut.20356.

Donald Lien and Keshab Shrestha. Price Discovery in Interrelated Markets. *Journal of Futures Markets*, 34(3):203–219, March 2014. ISSN 1096-9934. doi: 10.1002/fut.21593.

Chi-Jie Lu, Tian-Shyug Lee, and Chih-Chou Chiu. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 47(2):115–125, May 2009. ISSN 0167-9236. doi: 10.1016/j.dss.2009.02.001.

Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media, July 2007. ISBN 978-3-540-26239-8.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics (Revised Edition)*. John Wiley & Sons Ltd, 1999.

Karel Mertens and Morten O. Ravn. The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, 103(4):1212–1247, June 2013. ISSN 0002-8282. doi: 10.1257/aer.103.4.1212.

Marcelo Moreira. Tests with Correct Size When Instruments Can Be Arbitrarily Weak. *Journal of Econometrics*, 152:131–140, October 2009. doi: 10.1016/j.jeconom.2009.01.012.

Michael P. Murray. Avoiding Invalid Instruments and Coping with Weak Instruments. *Journal of Economic Perspectives*, 20(4):111–132, December 2006. ISSN 0895-3309. doi: 10.1257/jep.20.4.111.

James Nason and Gregor Smith. The New Keynesian Phillips curve: Lessons from single-equation econometric estimation. *Economic Quarterly*, 94(Fall):361–395, 2008.

Whitney K. Newey and Richard J. Smith. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, 72(1):219–255, 2004. ISSN 0012-9682.

Serena Ng. Variable Selection in Predictive Regressions. Handbook of Economic Forecasting, Elsevier, 2013.

Athanasios Papoulis. *Stochastic Process and Probability, Random Variables*. McGraw-Hill, 1991.

John H. Rogers, Chiara Scotti, and Jonathan H. Wright. Unconventional Monetary Policy and International Risk Premia. *Journal of Money, Credit and Banking*, 50(8):1827–1850, 2018. ISSN 1538-4616. doi: 10.1111/jmcb.12511.

John Shea. Instrument Relevance in Multivariate Linear Models: A Simple Measure. *The Review of Economics and Statistics*, 79(2):348–352, 1997.

Douglas Staiger and James H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586, May 1997. ISSN 0012-9682. doi: 10.2307/2171753.

James H Stock and Mark W Watson. Variable Trends in Economic Time Series. *Journal of Economic Perspectives*, 2(3):147–174, August 1988. ISSN 0895-3309. doi: 10.1257/jep.2.3.147.

James H Stock and Mark W Watson. Disentangling the Channels of the 2007-2009 Recession. Working Paper 18094, National Bureau of Economic Research, May 2012.

James H. Stock and Jonathan H. Wright. GMM with Weak Identification. *Econometrica*, 68(5): 1055–1096, September 2000. ISSN 1468-0262. doi: 10.1111/1468-0262.00151.

James H. Stock and Motohiro Yogo. Testing for Weak Instruments in Linear IV Regression. In Donald W. K. Andrews and James H. Stock, editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, Cambridge, 2005. ISBN 978-0-521-84441-3. doi: 10.1017/CBO9780511614491.006.

James H. Stock, Jonathan H. Wright, and Motohiro Yogo. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, 20 (4):518–529, October 2002. ISSN 0735-0015.

Syed Galib Sultan and Eric Zivot. *Price Discovery Share-An Order Invariant Measure of Price Discovery with Application to Exchange-Traded Funds*. PhD thesis, University of Washington, 2015.

Eric T. Swanson and John C. Williams. Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates. *American Economic Review*, 104(10):3154–3185, October 2014. ISSN 0002-8282. doi: 10.1257/aer.104.10.3154.

Marliese Uhrig-Homburg and Michael Wagner. Futures Price Dynamics of CO_2 Emission Allowances: *An Empirical Analysis of the Trial Period*. *The Journal of Derivatives*, 17(2):73–88, November 2009. ISSN 1074-1240. doi: 10.3905/JOD.2009.17.2.073.

Wenjie Wang and Firmin Doko Tchatoka. On Bootstrap inconsistency and Bonferroni-based size-correction for the subset Anderson–Rubin test under conditional homoskedasticity. *Journal of Econometrics*, 207(1):188–211, November 2018. ISSN 0304-4076. doi: 10.1016/j.jeconom.2018.07.003.

Hui Zou. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429, December 2006. ISSN 0162-1459. doi: 10.1198/016214506000000735.