DISCUSSION PAPER SERIES

No. 10390

FACTOR BASED IDENTIFICATION-ROBUST INFERENCE IN IV REGRESSIONS

George Kapetanios, Lynda Khalaf
and Massimiliano Marcellino

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

FACTOR BASED IDENTIFICATION-ROBUST INFERENCE IN IV REGRESSIONS

George Kapetanios, Lynda Khalaf and Massimiliano Marcellino

Discussion Paper No. 10390
February 2015
Submitted 22 January 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

FACTOR BASED IDENTIFICATION-ROBUST INFERENCE IN IV REGRESSIONS[†]

Abstract

Robust methods for IV inference have received considerable attention recently. Their analysis has raised a variety of problematic issues such as size/power trade-offs resulting from weak or many instruments. We show that information-reduction methods provide a useful and practical solution to this and related problems. Formally, we propose factor-based modifications to three popular weak-instrument-robust statistics, and illustrate their validity asymptotically and in finite samples. Results are derived using asymptotic settings that are commonly used in both the factor and weak instrument literatures. For the Anderson-Rubin statistic, we also provide analytical finite sample results that do not require any underlying factor structure. An illustrative Monte Carlo study reveals the following. Factor based tests control size regardless of instruments and factor quality. All factor based tests are systematically more powerful than standard counterparts. With informative instruments and in contrast with standard tests: (i) power of factor-based tests is not affected by k even when large, and (ii) weak factor structure does not cost power. An empirical study on a New Keynesian macroeconomic model suggests that our factor-based methods can bridge a number of gaps between structural and statistical modeling.

JEL Classification:
Keywords: factor model, identification-robust inference, IV regression, new Keynesian model, principle components and weak instruments

George Kapetanios g.kapetanios@qmul.ac.uk
Queen Mary, University of London

Lynda Khalaf lynda.khalaf@carleton.ca
Carleton University

Massimiliano Marcellino massimiliano.marcellino@unibocconi.it
Bocconi University, IGIER and CEPR

[†] We would like to thank the Editor Herman van Dijk, two anonymous Referees, Bertille Antoine and Jim Stock, for useful comments, and Jean-Marie Dufour and Mohamed Taamouti for useful suggestions and for sharing their code and data. We are also grateful to Zhenjiang Lin for excellent research assistance. The usual disclaimers apply.

1 Introduction

Identification concerns have revolutionized the practice of instrumental variable (IV) based econometrics (see Stock (2010)). Identification problems, which refer to the possibility of inferring characteristics of interest from observable data, are particularly enduring in IV regressions. In such contexts, identifying structural parameters depends on the quality of available instruments, and it is often quite difficult in economics to find informative instruments from observed data. Such difficulties have been at the heart of IV-based econometrics since the early 1990s and various inference methods are now available that are considered identification-robust [IdR], that is, that correct for the possibility of weak instruments.[1]

The field has been profoundly affected by the introduction of such methods. Nevertheless, important questions are still open and concern, among others, problems arising from the number of considered instruments. Since commonly used economic models rarely provide guidance for instrument choice, the number of instruments used in empirical studies is often much larger than the number of instrumented variables and sometimes quite large relative to the sample size. This practice uses up degrees of freedom which is likely to cause size distortions and/or power losses. In this paper, we consider this problem from an IdR perspective.

To set focus, assume that interest centers on testing the null hypothesis $\beta = \beta^0$, with β^0 known, in the IV regression $Y = \mathbf{Y}_1\beta + U$, where Y is a T-dimensional vector and $\mathbf{Y}_1$ is a $T \times G$ matrix of endogenous regressors. In this case, one of the first proposed IdR methods that may be traced back to Anderson and Rubin (1949) [thereafter AR] requires testing, in the context of an artificial regression of $Y - \mathbf{Y}_1\beta^0$ on k instruments, the exclusion of these k instruments. Mapping the test into a regression on instruments allows one to use standard techniques whose size is not affected by the quality of instruments. However, in over-identified applications, testing G restrictions [on β] would require assessing k constraints [the number of instruments tested out] where k may be much larger than G. This difficulty has guided related research since Dufour (1997) and Staiger and Stock (1997). Prominent alternatives include the LM-type test of Kleibergen (2002) and the conditional LR test by Moreira (2003), both motivated by a desire to minimize the effects of over-identification.

When k is large relative to G, then the question of whether k is large relative to T is central. Research aiming to tackle associated problems has prompted important changes in

[1] See for example Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Stock and Wright (2000), Stock, Wright, and Yogo (2002), Kleibergen (2002), Dufour (2003), Moreira (2003), Dufour and Taamouti (2005), Kleibergen (2005), Andrews, Moreira, and Stock (2006), Dufour and Taamouti (2007), Andrews and Stock (2007), and Chaudhuri, Richardson, Robins, and Zivot (2010).

asymptotic analysis. These include frameworks in which both the sample size, T, and the number of instruments, k, can grow infinitely.[2] In this context, Andrews and Stock (2007) and Newey and Windmeijer (2009) show that the Anderson-Rubin statistic and the statistics of Kleibergen (2002) and Moreira (2003) remain valid under specific conditions. Nevertheless, reported finite sample studies suggest that many available procedures - although valid in principle - are highly sensitive to the number of instruments. For example, Andrews and Stock (2007) emphasize that k should not be too large relative to T. Hansen, Hausman, and Newey (2008) argue that some commonly held assumptions on the magnitude of k relative to T are likely to prove more useful with large cross-section datasets than in time series settings with small samples. Mikusheva (2009) raises similar concerns with regards to methods proposed by Kleibergen (2005). In other words, although work on large k asymptotics is ongoing, many and possibly weak instruments still raise difficult problems in practice. The present paper reflects such a perspective, via a focus on information-reduction.

Information reduction methods including principle components and factor analysis are popular nowadays to analyze a wide spectrum of economic models. Their usefulness for improving standard IV methods including the 2SLS estimator has recently been demonstrated; see Bai and Ng (2010), Kapetanios and Marcellino (2010) and the references therein. To the best of our knowledge, available factor-IV results still focus on standard criteria including for example standard IV-based t-tests and more generally Wald criteria. The above cited works on IdR methods seem to suggest that standard statistics cannot be improved so conceptually different IdR criteria have been suggested instead. This raises concerns as to whether factor-IV salvages standard criteria. Furthermore, whether factor reduction works in conjunction with the alternative IdR class of statistics remains an open question. This paper addresses both problems. We introduce factor-IV in the context of three leading and representative IdR statistics namely the AR statistic and the statistics of Kleibergen (2002) and Moreira (2003). We motivate the use of factor based tests on the grounds of both controlling size and improving power.

Alternative statistics are available that rely on different estimation objectives. For example, Guggenberger and Smith (2005) and Guggenberger and Smith (2008) focus on empirical likelihood. Guggenberger, Kleibergen, Mavroeidis, and Chen (2012) propose partialled-out statistics for inference on subsets of parameters. Our results suggest that factor-analysis is worth considering in these contexts as well. We nevertheless focus on the AR test and tests

[2]See *e.g.* Chao and Swanson (2005), Stock and Yogo (2005), Newey and Windmeijer (2009) and Antoine and Lavergne (2014).

from Kleibergen (2002) and Moreira (2003) since these fundamental methods have set standards for subsequent work in this literature and have been extensively studied. In conjunction with these statistics, we show that factor analysis methods are appealing with time series and small samples, as occurs for example in macro-economics. At the same time, when factor approaches are combined with IV-regressions, identification requires both strong instruments as well as strong factor structures.[3] Since these effects are hard to disentangle, we show that our proposed factor modifications are valid whether instruments are weak or strong and whether hypothesized factor structures are weak or strong.

Overall, this paper has four main contributions. First, we extend the AR method as well as the K-test from Kleibergen (2002) and the LR test of Moreira (2003) to cover the case where instruments are selected via principle components; we refer to resulting statistics as their factor-based counterparts. Formally, we consider the case where endogenous regressors ($\mathbf{Y}_1$ in the above example) depend, weakly or strongly, on a - possibly large - number of - possibly - unobservable factors, and where the set of available instruments depends, weakly or strongly, on these factors. In this context, we show that all factor-based criteria achieve size control asymptotically, given commonly used regularity assumptions.[4] In particular, we assume that the (true) factors are valid instruments but the instruments individually might not be. In other words, our asymptotic results do not rule out the possibility that instruments are contemporaneously correlated with the structural error.[5]

Second, we show that the factor-based AR statistic remains exactly pivotal given commonly used finite sample assumptions except the $k < T$ one which we relax.[6] In particular, valid instruments are required but the instruments and the underlying factor structure can be weak or strong. This result does not call for any assumptions on the relative size of T and k, as long as the number of retained factors is smaller than T. Our factor-AR test is thus useful even if k is larger than T, in which case the original AR test would be infeasible.

Third, we analyze size and power properties of the considered statistics via a Monte Carlo study with emphasis on over-instrumentation, specification, weak instruments and weak fac-

[3]On weak structures: in standard factor analysis, see *e.g.* Boivin and Ng (1981) and Onatski (2012), and in factor IV, see Bai and Ng (2010) and Kapetanios and Marcellino (2010).

[4]These include assumptions by Staiger and Stock (1997) [on weak-IV], Andrews and Stock (2007) and Newey and Windmeijer (2009) [on many and weak IV], and Bai and Ng (2010) and Kapetanios and Marcellino (2010) [on weak factors].

[5]On the effects of invalid instruments, see Ashley (2009), Berkowitz, Caner, and Fang (2008), Doko-Tchatoka and Dufour (2008), Hahn and Hausman (2005), Chetty, Friedman, Glaeser, Imbens, and Kolesar (2011). Chetty, Friedman, Glaeser, Imbens, and Kolesar (2011) also provide a comprehensive survey of related on-going works.

[6]These include assumptions by Dufour (1997), Dufour and Taamouti (2005) and Andrews, Moreira, and Stock (2006), as well as Dufour and Taamouti (2007) who allow for a mis-specified reduced form equation.

tors. We also study the standard Wald test in our framework. Results can be summarized as follows. Our factor based tests control size regardless of instruments and factor structure weakness. In contrast to all factor IdR criteria, the standard factor-IV Wald test has both size and power problems in our context. Given strong instruments, all factor based tests are systematically more powerful than their standard counterparts. More importantly, and in contrast with standard tests, power is not affected by k even when large. In the latter context, we discuss an empirically relevant case building on Dufour and Taamouti (2007) (DT) where standard tests have practically no power while the power of their factor counterparts is close to one. Interestingly, even when the factor structure is not strong, in general power does not suffer as long as instruments are not weak.[7] Taken collectively, power results suggest to set the number of factors at or close to the number of right-hand-side endogenous variables, whether the latter corresponds to the number of true factors or not.

Finally, we apply our proposed inference methods to structural New Keynesian models. We consider a system of three equations with US data, and various different specifications based on Clarida, Galí, and Gertler (1999), Galí and Gertler (1999), Galí, Gertler, and Lopez-Salido (2001), Linde (2005), Benati (2008), Mavroeidis (2010), Kapetanios and Marcellino (2010) and Dufour, Khalaf, and Kichian (2013). Our focus is motivated by the recent surveys of Schorfheide (2013) and Mavroeidis, Plagborg-Moller, and Stock (2014), as well as the Cochrane (2011) critique. To address some of the empirical challenges that plague Dynamic Stochastic General Equilibrium (DSGE) models, Schorfheide (2013) proposes, among other strategies: (i) to use identification-robust methods, and (ii) to "connect" the structure - in some way - to "richer data-sets".[8] Our exercise takes both suggestions in consideration. We combine model-based instruments with factors based on Stock and Watson (2005). Results suggest that factor-based instruments can bridge a number of important gaps between structural and statistical macroeconomic models.

The paper is organized as follows. In Section 2 we review the IdR inferential procedures. In Section 3 we introduce and analyze the proposed factor based statistics. In Section 4, we

[7]On the other hand, if $\mathbf{Y_1}$ depends on only few variables in the large instrument set, and the set of instruments has little collinearity (weak factor structure), then factor analysis would be suboptimal with respect to the use of the proper subset of relevant instruments.

[8]IdR inference has gained popularity in macroeconometrics. See Dufour, Khalaf, and Kichian (2006), Nason and Smith (2008), Kleibergen and Mavroeidis (2009) and companion discussions and references therein, Magnusson and Mavroeidis (2010), Dufour, Khalaf, and Kichian (2010) and Dufour, Khalaf, and Kichian (2013). On the identification of DSGE models, see also Canova and Sala (2009), Iskrev (2010), Komunjer and Ng (2011), Andrews and Mikusheva (2014), Guerron-Quintana, Inoue, and Kilian (2013) and the references therein. On exploiting data-rich approaches in DSGEs including factor analysis, see also Consolo, Favero, and Paccagnini (2009) and the references therein.

analyze instrument and factor weakness from an asymptotic and finite sample perspective. In Section 5, we consider several extensions of the basic framework, including non i.i.d. errors and parameter non-linearity; we also study selection of the number of factors. In Section 6 we assess the finite sample size and power properties of the proposed procedures. Our empirical analysis is reported in section 7. Section 8 summarizes the main results and concludes.

2 Robust IV: an overview

This section reviews the considered robust IV methods within their initially proposed Data Generation Process (DGP), which aims to motivate the factor-based setups we define subsequently and facilitate presentation, extensions and proofs.

Adopting the notation [maintained throughout the paper] that vectors are identified by capital letters, e.g. $Z = (z_1, ..., z_T)'$ is $T \times 1$, and matrices by bold capital letters, e.g. $\mathbf{Z} = (Z_1, ..., Z_k)$ is $T \times k$, where T denotes the sample size, we consider the system

$$Y = \mathbf{Y}_1 \beta + U, \tag{1}$$

$$\mathbf{Y}_1 = \mathbf{X\Pi} + \mathbf{V}, \tag{2}$$

where $\mathbf{Y}_1$ is a $T \times G$ matrix of endogenous regressors with typical element denoted by $y_{1,it}$, $\mathbf{X}$ is a $T \times k$ matrix of valid instruments (exogenous variables), with $k \geq G$, β and $\mathbf{\Pi}$ are, respectively, a $G \times 1$ vector and $k \times G$ matrix of unknown parameters, and $[U : \mathbf{V}]$ is a $T \times (G + 1)$ matrix of mean zero errors terms.

The AR test statistic associated with testing

$$H_0 : \beta = \beta^0, \tag{3}$$

using $\mathbf{X}$ as instruments, takes the form

$$AR(\beta^0|\mathbf{X}) = \frac{T-k}{k} \frac{(Y - \mathbf{Y}_1\beta^0)' [I - M(\mathbf{X})] (Y - \mathbf{Y}_1\beta^0)}{(Y - \mathbf{Y}_1\beta^0)' [M(\mathbf{X})] (Y - \mathbf{Y}_1\beta^0)} \tag{4}$$

where for any full-column rank matrix $\mathbf{A}$, $M(\mathbf{A}) = I - N(\mathbf{A})$ and $N(\mathbf{A}) = \mathbf{A} (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}'$. If we can condition on X for statistical analysis and the error terms are $i.i.d.$ across rows, and if U is Gaussian, then the AR statistic follows in finite samples an $F(k, T - k)$ null distribution. Dufour and Taamouti (2005) provide analytical solutions for inverting this test leading to confidence sets for the individual components of β. The Gaussian hypothesis on U and strong exogeneity of X can be relaxed so that, under the usual regularity conditions of classical regression [see for example Dufour and Jasiak (2001) and Andrews, Moreira, and

Stock (2006)], $k \times AR(\beta^0|\mathbf{X}) \overset{asy}{\sim} \chi^2(k)$. We will refer to this version of AR as ARS. Staiger and Stock (1997) reconsider the limiting properties of this statistic using a specific weak-IV asymptotic framework that has become, since then, a basic standard in this literature. Stock and Wright (2000) generalize the latter to non-linear and non-*i.i.d.* settings via GMM.

Dufour (2003) argues that a point-optimal although infeasible instrument corresponds to

$$\bar{Z} = \mathbf{X\Pi}. \tag{5}$$

See also Dufour, Khalaf, and Kichian (2006). The LM procedure of Wang and Zivot (1998) may be obtained as an AR test based on $\bar{Z}$ setting $\mathbf{\Pi}$ to

$$\hat{\mathbf{\Pi}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1. \tag{6}$$

The same holds for the K-test from Kleibergen (2002) for which $\mathbf{\Pi}$ is replaced by its constrained LIML estimator:

$$\hat{\mathbf{\Pi}}^0 = \hat{\mathbf{\Pi}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[Y - \mathbf{Y}_1\beta^0\right]\frac{[Y - \mathbf{Y}_1\beta^0]' M(\mathbf{X})\mathbf{Y}_1}{[Y - \mathbf{Y}_1\beta^0]' M(\mathbf{X})\,[Y - \mathbf{Y}_1\beta^0]}. \tag{7}$$

Under the weak-IV regularity conditions in Kleibergen (2002), $k \times AR(\beta^0|\mathbf{X}\hat{\mathbf{\Pi}}^0) \overset{asy}{\sim} \chi^2(G)$. Kleibergen (2005) extends the latter to non-*i.i.d.* and non-linear setting of GMM.

The LR statistic of Moreira (2003) is based on a Gaussian model with a known covariance matrix $\mathbf{\Omega}$ for $[U : \mathbf{V}]$ and fixed regressors:

$$LR(\beta^0|\mathbf{X}) = \frac{b_0'\mathbf{Y}'N(\mathbf{X})\mathbf{Y}b_0}{b_0'\mathbf{\Omega}b_0} - \min_b \frac{b'\mathbf{Y}'N(\mathbf{X})\mathbf{Y}b}{b'\mathbf{\Omega}b} \tag{8}$$

where $\mathbf{Y}=(Y, \mathbf{Y}_1)$, $b = (1, -\beta')'$ and $b_0 = (1, -\beta^{0\prime})'$. The null distribution of LR *conditional* on $\mathbf{T}_0 = \mathbf{X}'\mathbf{Y}\mathbf{\Omega}^{-1}(\beta^0, 1)'$, that is, given that $\mathbf{T}_0$ takes the value $\mathbf{t}_0$, does not depend on $\mathbf{\Pi}$; the latter distribution can be computed (as a function of $\mathbf{t}_0$) and used to construct a test robust to $\mathbf{\Pi}$. Furthermore, Moreira suggests

$$LR_1(\beta^0|\mathbf{X}) = \frac{b_0'\mathbf{Y}'N(\mathbf{X})\mathbf{Y}b_0}{b_0'\hat{\mathbf{\Omega}}b_0} - \hat{l}, \quad \hat{l} = \min_\beta \frac{b'\mathbf{Y}'N(\mathbf{X})\mathbf{Y}b}{b'\hat{\mathbf{\Omega}}b}, \quad \hat{\Omega} = \frac{\mathbf{Y}'M(\mathbf{X})\mathbf{Y}}{T-k} \tag{9}$$

where $\hat{l}$ corresponds to the smallest eigenvalue of $\hat{\mathbf{\Omega}}^{-1/2}\mathbf{Y}'N(\mathbf{X})\mathbf{Y}\hat{\mathbf{\Omega}}^{-1/2}$, and another variant based on the LR statistic with an unknown $\mathbf{\Omega}$

$$LR_2(\beta^0|\mathbf{X}) = \frac{T}{2}\ln\left(1 + \frac{b_0'\mathbf{Y}'N(\mathbf{X})\mathbf{Y}b_0}{b_0'\mathbf{Y}'M(\mathbf{X})\mathbf{Y}_0}\right) - \frac{T}{2}\ln\left(1 + \frac{\hat{l}}{T-k}\right).$$

$LR_1(\beta^0|\mathbf{X})$ and $LR_2(\beta^0|\mathbf{X})$ can be used with the same critical values as with $LR(\beta^0|\mathbf{X})$, using for $\mathbf{t}_0$ its plug-in counterpart, obtained by replacing $\mathbf{\Omega}$ by $\hat{\mathbf{\Omega}}$ in $\mathbf{t}_0$.

Andrews, Moreira, and Stock (2006) study these statistics using the weak-IV asymptotics of Staiger and Stock (1997). On validity with many instruments, see Andrews and Stock (2007) and Newey and Windmeijer (2009). It is worth noting that minimizing the AR statistic leads to the LIML estimator; see Carrasco and Tchuente (2012) on LIML with many instruments. Our proposed factor-based modification and associated framework are presented next.

3 The factor based robust inferential procedures

Maintaining the above notation, let us now consider the DGP

$$Y = \mathbf{Y}_1\beta + U, \tag{10}$$

$$\mathbf{Y}_1 = \mathbf{F\Pi} + \mathbf{V}, \tag{11}$$

$$\mathbf{X} = \mathbf{F\Lambda_k} + \mathbf{E}. \tag{12}$$

The main differences with respect to the DGP (1) in Section 2 are that now the endogenous variables $\mathbf{Y}_1$ depend on r unobservable and independent factors $f_t = (f_{1,t}, ..., f_{r,t})'$, with $r < k$ and $r < T$, grouped in the $T \times r$ matrix $\mathbf{F} = (f_1, ..., f_T)'$. In turn, each element of $\mathbf{X}$ depends on the common factors $\mathbf{F}$ via the $r \times k$ loadings $\mathbf{\Lambda_k} = [\lambda_{ij}]$, and on an idiosyncratic component, *i.e.*, an element of the $T \times k$ matrix $\mathbf{E}$. We add a subscript k to the loadings matrix to indicate clearly that its dimension depends on k, which will be assumed to tend to infinity, as is standard in this literature. Conditions on the errors $\mathbf{E} = (e_1, ..., e_T)'$ are specified below.

Assumption 1

a. $E||f_t||^4 \leq M < \infty$, $T^{-1}\sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \mathbf{\Sigma}_f$ for some $r \times r$ positive definite matrix $\mathbf{\Sigma}_f$. $\mathbf{\Lambda_k}$ has bounded elements. Further $||\mathbf{\Lambda_k}\mathbf{\Lambda_k}'/k - \mathbf{D}|| \to 0$, as $k \to \infty$, where $\mathbf{D}$ is a positive definite matrix.

b. $E(e_{i,t}) = 0$, $E|e_{i,t}|^8 \leq M$ where $e_t = (e_{1,t}, ..., e_{k,t})'$. The variance of e_t is denoted by $\mathbf{\Sigma}_e$. f_s and e_t are independent for all s, t.

c. For $\tau_{i,j,t,s} \equiv E(e_{i,t}e_{j,s})$ the following hold

Assumption 2

a. $k^{-1}\sum_{i=1}^{k}\sum_{j=1}^{k}|\tau_{i,j,s,s}| \leq M$ for all s

b. $(kT)^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{j=1}^{k}|\tau_{i,j,t,s}| \leq M$

c. For every (t,s), $E|(k)^{-1/2}\sum_{i=1}^{k}(e_{i,s}e_{i,t}-\tau_{i,i,s,t})|^4 \leq M$

Assumption 3 *$(u_t, v_{1,t}, ..., v_{G,t})'$, $t = 1, ..., T$, is a martingale difference sequence with covariance matrix Σ_{uv} of finite norm. u_t and $v_{j,t}$ have finite fourth moments.*

Assumption 4 *$E(f_{i,t}u_t) = 0$, $i = 1, ..., r$. $E(f_{i,t}v_{j,t}) = 0$, $i = 1, ..., r$; $j = 1, ..., G$. $E(x_t x_t')$ is nonsingular for all k and t. $E(x_t x_t')$ has full column rank k. x_t have finite fourth moments.*

Assumption 1 is standard in the factor literature. In particular, it is used in Stock and Watson (2002b), Stock and Watson (2002a), Bai and Ng (2002) and Bai (2003) to prove consistency and asymptotic normality (at certain rates) of the principal component based estimator of the factors, and by Bai and Ng (2006) to show consistency of the parameter estimators in factor augmented regressions. Assumption 3 will be relaxed below to allow for more general dependence structures. Via assumptions 3 and 4, we contribute to the literature studying properties of IV methods allowing for many, possibly weak and possibly invalid instruments. Indeed, standard IV typically requires that instruments are "valid", that is $E(x_{i,t}u_t) = 0$, $i = 1, ..., k$. This is not guaranteed under Assumption 4, which requires that the unobservable components, that is, the true factors, are valid. We show that using factor methods to reduce a large number of possibly invalid instruments prior to using the considered statistics will eventually evacuate the "invalid component" of instruments asymptotically. These assumptions also cover the case where the number of instruments k exceeds the number of observations T, as long as r is small.

Assumptions 1 and 4 are assumed to hold throughout the paper, unless stated otherwise. For further reference, we note that (10)-(12) imply

$$Y - \mathbf{Y}_1\beta^0 = \mathbf{Y}_1\left(\beta - \beta^0\right) + U = \mathbf{F\Pi}\left(\beta - \beta^0\right) + U + \mathbf{V}\left(\beta - \beta^0\right). \tag{13}$$

The statistic (4) relies on $\mathbf{X}$ as the instrument, in view of (2). Relaxing the latter, a GMM-type perspective for inference on (1) given any $T \times r$ full-column rank matrix of instruments $\mathbf{Z}$ implies testing the exclusion of $\mathbf{Z}$, in the artificial regression of $Y - \mathbf{Y}_1\beta^0$ on $\mathbf{Z}$, using the F-statistic

$$AR(\beta^0|\mathbf{Z}) = \frac{T-r}{r}\frac{(Y - \mathbf{Y}_1\beta^0)'\left[I - M(\mathbf{Z})\right](Y - \mathbf{Y}_1\beta^0)}{(Y - \mathbf{Y}_1\beta^0)'\left[M(\mathbf{Z})\right](Y - \mathbf{Y}_1\beta^0)}. \tag{14}$$

For further reference [mainly in section 4], we introduce the pivotal statistic given any $\mathbf{Z}$:

$$\overline{AR}(\mathbf{Z}) = \frac{T-r}{T}\frac{U'\left[I - M(\mathbf{Z})\right]U}{U'\left[M(\mathbf{Z})\right]U}. \tag{15}$$

Let $\hat{\mathbf{F}}$ refer to the $T \times r$ matrix of the principal components of $\mathbf{X}$ that are commonly used [in line with Stock and Watson (2002b) and Stock and Watson (2002a)] as estimators for $\mathbf{F}$ in factor models akin to (12). Replacing $\mathbf{Z}$ in (14) by $\mathbf{F}$ and $\hat{\mathbf{F}}$ leads to two statistics of interest, $AR(\beta^0|\mathbf{F})$ and $AR(\beta^0|\hat{\mathbf{F}})$. We next show that, under mild conditions, $AR(\beta^0|\mathbf{F})$, and $AR(\beta^0|\hat{\mathbf{F}})$ are asymptotically equivalent, namely, the unobservable true factors can be replaced by their estimates.[9]

Theorem 1 *In the context of (10)-(12) under assumptions 1-4* $(AR(\beta^0|\mathbf{F}) - AR(\beta^0|\hat{\mathbf{F}}))$ *is* $o_p(1)$ *when* $T^{0.5}/k = o(1)$, *where* $AR(\beta^0|.)$ *is defined by (14).*

The proofs for this and subsequent theorems are provided in the appendix. A similar result applies for the Factor-ARS statistic, whose distribution is therefore the same as that of the ARS test with r instruments.

We introduce the Factor-LM statistics based on Wang and Zivot (1998), as $AR(\beta^0|\mathbf{F}\hat{\mathbf{\Pi}}_{\mathbf{F}})$, $AR(\beta^0|\hat{\mathbf{F}}\hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}})$, replacing $\mathbf{Z}$ in (14) by $\mathbf{F}\hat{\mathbf{\Pi}}_{\mathbf{F}}$ and $\hat{\mathbf{F}}\hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}}$ respectively, with

$$\hat{\mathbf{\Pi}}_{\mathbf{F}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Y}_1, \qquad \hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}} = (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\mathbf{Y}_1.$$

Theorem 2 *In the context of (10)-(12) under assumptions 1-4* $(AR(\beta^0|\mathbf{F}\hat{\mathbf{\Pi}}_{\mathbf{F}}) - AR(\beta^0|\hat{\mathbf{F}}\hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}}))$ *is* $o_p(1)$ *when* $T^{0.5}/k$ *is* $o(1)$, *where* $AR(\beta^0|.)$ *is defined by (14).*

The proof is similar to that for the Factor-AR test and the result follows from the fact that $\hat{\mathbf{F}} - \mathbf{F}\mathbf{H}$ is $o_p(1)$, where $\mathbf{H}$ is some non singular rotation matrix, since $\mathbf{H}$ does not enter in the computation of the statistic. This ensures that $AR(\beta^0|\mathbf{F}\hat{\mathbf{\Pi}}_{\mathbf{F}})$ and $AR(\beta^0|\hat{\mathbf{F}}\hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}})$ are asymptotically equivalent.

We next define the factor version of Kleibergen's (2002)'s K-tests [Factor-K] as $AR(\beta^0|\mathbf{F}\hat{\mathbf{\Pi}}^0_{\mathbf{F}})$, $AR(\beta^0|\hat{\mathbf{F}}\hat{\mathbf{\Pi}}^0_{\hat{\mathbf{F}}})$, replacing $\mathbf{Z}$ in (14) by $\mathbf{F}\hat{\mathbf{\Pi}}^0_{F}$ and $\hat{\mathbf{F}}\hat{\mathbf{\Pi}}^0_{\hat{\mathbf{F}}}$ respectively where, with reference to (7):

$$\hat{\mathbf{\Pi}}^0_{\mathbf{F}} = \hat{\mathbf{\Pi}}_{\mathbf{F}} - (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\left[Y - \mathbf{Y}_1\beta^0\right]\frac{\left[Y - \mathbf{Y}_1\beta^0\right]' M(\mathbf{F})\mathbf{Y}_1}{\left[Y - \mathbf{Y}_1\beta^0\right]' M(\mathbf{F})\left[Y - \mathbf{Y}_1\beta^0\right]}, \tag{16}$$

$$\hat{\mathbf{\Pi}}^0_{\hat{\mathbf{F}}} = \hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}} - (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\left[Y - \mathbf{Y}_1\beta^0\right]\frac{\left[Y - \mathbf{Y}_1\beta^0\right]' M(\hat{\mathbf{F}})\mathbf{Y}_1}{\left[Y - \mathbf{Y}_1\beta^0\right]' M(\hat{\mathbf{F}})\left[Y - \mathbf{Y}_1\beta^0\right]}. \tag{17}$$

The asymptotic distribution for the Factor-K test remains

$$\left(r \times AR(\beta^0|\hat{\mathbf{F}}\hat{\mathbf{\Pi}}^0_{\hat{\mathbf{F}}})\right) \overset{asy}{\sim} \chi^2(r),$$

[9] In section 4, we also analyze $\overline{AR}(\hat{\mathbf{F}})$.

under the weak-IV regularity conditions in Kleibergen (2002), since once again the statistic is invariant to the use of $\mathbf{F}$ or $\mathbf{FH}$ and the principal components are consistent for the space spanned by the true factors.

Finally, we introduce the Factor version of Moreira's statistics:

$$LR(\beta^0|\hat{\mathbf{F}}) = \frac{b_0'\mathbf{Y}'N(\hat{\mathbf{F}})\mathbf{Y}b_0}{b_0'\mathbf{\Omega}b_0} - \hat{\lambda}, \qquad \hat{\lambda} = \min_b \frac{b'\mathbf{Y}'N(\hat{\mathbf{F}})\mathbf{Y}b}{b'\mathbf{\Omega}b}$$
$$LR(\beta^0|\mathbf{F}) = \frac{b_0'\mathbf{Y}'N(\mathbf{F})\mathbf{Y}b_0}{b_0'\mathbf{\Omega}b_0} - \lambda, \qquad \lambda = \min_b \frac{b'\mathbf{Y}'N(\mathbf{F})\mathbf{Y}b}{b'\mathbf{\Omega}b}$$

where $\mathbf{Y}=(Y, \mathbf{Y}_1)$, $b = (1, -\beta')'$, $b_0 = (1, -\beta^{0\prime})'$ and $LR(\beta^0|.)$ is given by (8). If the factors are known, the Factor-LR statistic has the same limiting distribution as the standard LR statistic. Hence, we only need to study $(LR(\beta^0|\mathbf{F}) - LR(\beta^0|\hat{\mathbf{F}}))$.

Theorem 3 *In the context of (10)-(12) under assumptions 1-4 $LR(\beta^0|\mathbf{F}) - LR(\beta^0|\hat{\mathbf{F}})$ is $o_p(1)$ when $T^{0.5}/k = o(1)$.*

The Factor versions of Moreira's LR1 and LR2 statistics are constructed along the same lines, and a similar result applies for their asymptotic distribution.

4 Conditioning information weakness

In this section we discuss the implications of weakness in either instruments or factors for the performance of the proposed tests. The effect of instrument weakness is that the tests have lower power. The main effect of factor structure weakness is that factor estimates approximate the true factors sufficiently well to enable Theorems 1-3 to hold, only under stricter conditions. We also study the $\Lambda_k = 0$ extreme case, where we use principal components (factor estimates) as instruments in the absence of a factor structure, and introduce an alternative finite sample framework that justifies $AR(\beta^0|\hat{\mathbf{F}})$ regardless of any underlying factor structure.

4.1 Asymptotic analysis

To discuss the issue of weak instruments we first replace (11) by

$$\mathbf{Y}_1 = \mathbf{F}\mathbf{\Pi}_T + \mathbf{V}$$

where now $\mathbf{\Pi}_T$ is some non-stochastic sequence of matrices over T. Weakness in this context is equivalent to the idea that $\mathbf{\Pi}_T'\mathbf{F}'\mathbf{F}\mathbf{\Pi}_T = o_p(T)$. In particular, it is reasonable to model weakness by specifying that

$$\|\mathbf{\Pi}_T\| = O(T^{-\vartheta}), 0 \leq \vartheta \leq 1/2. \tag{18}$$

The following Theorem focuses on the AR test although similar results can be shown for the other tests we proposed.

Theorem 4 *Let (18) hold. Under assumptions 1-4,*

$$AR(\beta^0|\hat{\mathbf{F}}) = O_p\left(T^{1-2\vartheta}\right), \tag{19}$$

when the true value of the coefficient is $\beta^1 \neq \beta^0$.

Next, we examine the question of weak factor models. In this case (12) is replaced by

$$\mathbf{X} = \mathbf{F}\tilde{\mathbf{\Lambda}}_{\mathbf{k}} + \mathbf{E}.$$

where $\tilde{\mathbf{\Lambda}}_k = [\tilde{\lambda}_{ij,k}]$ and $\tilde{\lambda}_{ij,k} = \lambda_{ij}/k^{\alpha}, 0 < \alpha \leq 1/2$.While such a model of factor loadings may appear restrictive, discussions in Kapetanios and Marcellino (2010) suggest that it can accommodate a variety of different weak factor loading structures. The idea for this model is that the factor structure is weak in the sense that, as the cross-sectional dimension of the dataset increases, the factors explain a diminishing proportion of the variance of the data whose limit as $k \to \infty$, is zero. Another way to characterise weakness comes from noting that in a standard factor model the r largest eigenvalues of $\Lambda_k'\Lambda_k$ tend to infinity at rate k, while the rest of the eigenvalues are zero, since $\Lambda_k'\Lambda_k$ has rank equal to r. Further, under regularity conditions, the r largest eigenvalues of the covariance matrix of x_t tend to infinity at rate k, as well, while the remaining eigenvalues are bounded. Then, weakness can be characterised by these two sets of largest eigenvalues increasing with k, but at a rate lower than k. For example, recalling that assumption 1(a) specifies that λ_{ij} are bounded and that $\mathbf{\Lambda_k}\mathbf{\Lambda_k}'/k$ has a nonsingular limit, it is easy to show that $\mu_1(\tilde{\Lambda}_k'\tilde{\Lambda}_k) = O(1)$, in the extreme case where $\alpha = 1/2$, where $\mu_1(A)$ denotes the maximum eigenvalue of A. This implies that the covariance matrix of x_t has bounded eigenvalues, in contrast to a standard factor model where that covariance has a maximum eigenvalue that rises with k, at rate k. Kapetanios and Marcellino (2010) discuss this model in detail. It is worth noting that this local to zero model for $\tilde{\lambda}_{ij,k}$ need not be the only way to characterise a weak factor model. However, it is the only one, whose properties have, to the best of our knowledge, been theoretically considered. For our purposes it is important to note what the implications of this weakness are for the inferential tools we propose. The following Theorem does that.

Theorem 5 *Let $\lambda_{ij,k} = \lambda_{ij}/k^{\alpha}$, $0 < \alpha < 1/4$. Further let $k = o(T^{1/4\alpha})$ and $C_{kT}^{-1}T^{1/2} = o(1)$ where*

$$C_{kT} \equiv \min\left(k^{1/2-3\alpha}T^{1/2}, k^{1-3\alpha}, k^{-2\alpha}\min(k,T)\right). \tag{20}$$

Then, under assumptions 1-4, the following hold: (i) $(AR(\beta^0|\mathbf{F}) - AR(\beta^0|\hat{\mathbf{F}})) = o_p(1)$*, where* $AR(\beta^0|.)$ *is defined by (14), (ii)* $(AR(\beta^0|\mathbf{F}\hat{\mathbf{\Pi}}_{\mathbf{F}}) - AR(\beta^0|\hat{\mathbf{F}}\hat{\mathbf{\Pi}}_{\hat{\mathbf{F}}})) = o_p(1)$*, and (iii)* $LR(\beta^0|\mathbf{F}) - LR(\beta^0|\hat{\mathbf{F}}) = o_p(1)$*.*

This Theorem follows from Lemma 1 and Theorem 5 of Kapetanios and Marcellino (2010). One way to provide intuition for this condition is to note that as $\alpha \to 0$, namely, the factor structure becomes strong, the condition $C_{kT}^{-1} T^{1/2} = o(1)$ becomes equivalent to $\sqrt{T}/k = o(1)$, which is the standard condition for the asymptotic equivalence of using true and estimated factors.

4.2 Finite sample analysis

Theorem 6 *In the context of (10)-(12) where both* $\mathbf{E}$ *and* $\mathbf{F}$ *are independent of* U*,* $AR(\beta^0|\hat{\mathbf{F}})$ *is distributed, under the null hypothesis (3), like* $\overline{AR}(\hat{\mathbf{F}})$*, where* $\overline{AR}(.)$ *is defined by (15), with no further assumptions on* $\mathbf{\Pi}$ *and* $\mathbf{\Lambda_k}$*, so the null distribution of* $AR(\beta^0|\hat{\mathbf{F}})$ *is completely determined by the distribution of* U *given* $\hat{\mathbf{F}}$*.*

If U is further assumed *i.i.d.* Gaussian, then the $F(r, T-r)$ distribution will provide size-correct critical points for $AR(\beta^0|\hat{\mathbf{F}})$, regardless of the quality of instruments and factors and for finite T and any k, that is, without any assumptions on the relative size of T and k, as long as T exceeds r. Since the error terms $\mathbf{E}$ and $\mathbf{V}$ are also evacuated from $\overline{AR}(\hat{\mathbf{F}})$, the above shows that the null distribution of $AR(\beta^0|\hat{\mathbf{F}})$ does not depend on the specification of the idiosyncratic errors in the factor model for $\mathbf{X}$ as long as $\mathbf{X}$ is independent from U.

Our proof of Theorem 6 relies on (13). This same decomposition can also serve to assess the key determinants of power for $AR(\beta^0|\hat{\mathbf{F}})$. In particular, for a given difference $(\beta - \beta^0)$, the power will decrease in the case of weak instruments ("small" $\mathbf{\Pi}$). When the $\mathbf{X}$ are used as instruments, the strength of the factor structure ($\mathbf{\Lambda_k}$) also matters.

Theorem 6 suggests an alternative interpretation of the factor AR test, beyond the scope of model (10)-(12). One specification of interest consists in reconsidering (1)-(2) where (2) is high-dimensional, or more concretely, where k is large relative to T and may be even larger than T, and/or $rank(\mathbf{X}) < k$. The following thus disregards (10)-(12) and companion assumptions.

Theorem 7 *In the context of (1)-(2) with* $\mathbf{X}$ *fixed or independent from* U *and* $rank(\mathbf{X}) \leq k$*, let* $\hat{\mathbf{F}}$ *refer to the matrix of (the first)* r *principle components of* $\mathbf{X}$*. Define the factor AR statistic* $AR(\beta^0|\hat{\mathbf{F}})$ *using* $\hat{\mathbf{F}}$ *for* $\mathbf{Z}$ *in (14). Then, under the null hypothesis (3),* $AR(\beta^0|\hat{\mathbf{F}})$ *is* *distributed like $\overline{AR}(\hat{\mathbf{F}})$, where $\overline{AR}(.)$ is defined by (15), with no further assumptions on the rank of $\mathbf{\Pi}$ nor on k relative to T, so the null distribution of $AR(\beta^0|\hat{\mathbf{F}})$ is completely determined by the distribution of U given $\hat{\mathbf{F}}$.*

Using (15), simulation-based Monte Carlo or parametric bootstrap p-values can be obtained by simulation [see e.g. Dufour and Khalaf (2002), Beaulieu, Dufour, and Khalaf (2007), Beaulieu, Dufour, and Khalaf (2010) or Beaulieu, Dufour, and Khalaf (2013)] if the distribution of U is given up to an unknown scalar. If U is *i.i.d.* Gaussian then $AR(\beta^0|\hat{\mathbf{F}})$ follows an $F(r, T-r)$ null distribution.

Theorem 7 does not treat $\hat{\mathbf{F}}$ as "estimates" of some underlying factors. So from a finite sample perspective, this proves the validity of $AR(\beta^0|\hat{\mathbf{F}})$ without assuming any underlying factor structure, of any form. The definition of a principle component implies that for any r

$$\mathbf{Y}_1 = \mathbf{X\Pi} + \mathbf{V} \Rightarrow \mathbf{Y}_1 = \hat{\mathbf{F}}\mathbf{\Pi}_* + \mathbf{V}_* \tag{21}$$

where $\mathbf{\Pi}_*$ may or may not have full rank. We are thus led back to the test's original framework, which ensures size control whether $\mathbf{\Pi}_*$ is zero, close to zero or rank deficient, with no further assumptions.

Finally, note that in theorems 6 and 7 we could use any linear combinations of the instruments rather than specifically principal components, making the results of more general applicability.

5 Extensions

5.1 Heteroskedasticity and Serial Correlation

We now relax assumption 3 and allow for correlation and heteroskedasticity in the errors U of equation (10) when considering the Factor AR test. We formalize this with the following assumption, which substitutes for assumption 3:

Assumption 5 *u_t is a zero mean process with finite variance. The process $f_t u_t$ satisfies the conditions for the application of some central limit theorem for weakly dependent processes, with a zero mean asymptotic normal limit.*

$$S_{fu} = \lim_{T\to\infty} T \left[E\left(\left[T^{-1} \sum_{t=1}^{T} u_t f_t \right] \left[T^{-1} \sum_{t=1}^{T} u_t f_t \right]' \right) \right]$$

exists and is nonsingular. u_t have finite eighth moments.

The latter is a high level assumption, given in this form for generality. More primitive conditions on the errors, e.g. mixing with polynomially declining mixing coefficients or near epoque dependence (see Davidson (1994)) are sufficient for Assumption 5 to hold. The assumption that u_t has finite eighth moment is needed to analyse terms of the form $T^{-1}\sum_{t=j+1}^{T} u_t^2\left(Hf_t - \widehat{f}_t\right)$ through Lemma A.1 of Bai and Ng (2006).

In principle, the AR method [see for example Stock and Wright (2000) or Beaulieu, Dufour, and Khalaf (2010)] can be adapted to accommodate such deviations from the *i.i.d.* case. To do so, revisit the artificial regression of $Y - \mathbf{Y}_1\beta^0$ on $\mathbf{Z}$ from section 3. In this context, the usual Newey-West Wald-HAC statistic may be used instead of the F-statistic:

$$AR\text{-HAC}(\beta^0|\hat{\mathbf{F}}) = T^{-1}\left(Y - \mathbf{Y}_1\beta^0\right)'\hat{\mathbf{F}}\left(\hat{S}_{\hat{f}u}\right)^{-1}\hat{\mathbf{F}}'\left(Y - \mathbf{Y}_1\beta^0\right) \tag{22}$$

where $\hat{S}_{\hat{f}u}$ is an estimate of S_{fu} that can be obtained by a HAC procedure, such as that developed in Newey and West (1987). For example, using a Bartlett kernel, we have

$$\widehat{S}_{\hat{f}u,h} = \widehat{\Phi}_0 + \sum_{j=1}^{h}\left(1 - \frac{j}{h+1}\right)(\widehat{\Phi}_j + \widehat{\Phi}_j'), \widehat{\Phi}_j = T^{-1}\sum_{T=j+1}^{T}\widehat{u}_t^{\hat{f}}\widehat{u}_{t-j}^{\hat{f}}\hat{f}_t\hat{f}_{t-j}',$$

where h is the length of the window (bandwidth), $\hat{U}^{\hat{f}} = (\widehat{u}_1^{\hat{f}}, ..., \widehat{u}_T^{\hat{f}})'$, $\hat{U}^{f} = (Y - \mathbf{Y}_1\beta^0) - \hat{\mathbf{F}}\hat{\vartheta}^{\hat{f}}$, $\hat{\vartheta}^{\hat{f}} = (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'(Y - \mathbf{Y}_1\beta^0)$. Then, assuming the use of the Bartlett kernel, we have the following Theorem that mirrors Theorem 1.

Theorem 8 *Under assumptions 1-4 and 5, under the null hypothesis (3) assuming that $\widehat{S}_{\hat{f}u,h}$ is used as an estimate of S_{fu} with $h = o\left(T^{1/2}\right)$, and when $T^{0.5}/k = o_p(1)$, (AR-HAC$(\beta^0|\hat{\mathbf{F}})$ − AR-HAC$(\beta^0|\mathbf{F})$) is $o_p(1)$.*

If $\mathbf{F}$ and $\mathbf{U}$ satisfy the usual exogeneity, stationarity and ergodicity assumptions that validate the usual Newey-West statistic, then the AR-HAC$(\beta^0|\mathbf{F})$ statistic will follow a limiting $\chi^2(r)$ under the null hypothesis, and so will AR-HAC$(\beta^0|\hat{\mathbf{F}})$ under the condition of Theorem 8. Because the residuals are a function of β^0, inverting this statistic cannot be done analytically since the method of Dufour and Taamouti (2005) will no longer work. Numerical methods will be required, as applied for example by Dufour, Khalaf, and Kichian (2010).

5.2 Nonlinearity

We next consider relaxing the assumption of linearity in parameters. Formally, we generalize equation (10) into

$$Y = \mathbf{Y}_1\varphi(\beta) + U,$$

where $\varphi(.)$ is a possibly non-linear function of β and the latter remains the structural parameter of interest. The reduced forms (11) and (12) are unchanged. The above defined AR and factor AR statistics remain valid in finite samples to test $\beta = \beta^0$, regardless of whether the function $\varphi(.)$ is well behaved or presents discontinuities. Combining non-linearity with possibly non-i.i.d. errors leads to the framework of Stock and Wright (2000), in which case the Wald-HAC criterion would correspond (numerically) or will at least be asymptotically equivalent to a GMM objective function.

5.3 Selecting the number of factors

In a conventional factor analysis, the number of factors is estimated (possibly allowing for zero as an outcome) using information criteria (IC) as proposed *e.g.* by Bai and Ng (2002) or Onatski (2009). Penalties with such IC are however too strict when factor structures are weak. The problem (specifically that of estimating the number of factors) within Factor-IV, calls for a different although related treatment. Broadly, three issues deserve notice.

First, results of the previous section suggest reliance on IC that account for weakness, as proposed *e.g.* by Kapetanios and Marcellino (2010). Second, the exact distributional results in section 3.2 are unaffected, as long as the IC are applied to the instruments set only (in a way that does not use information on the model's endogenous regressors). Finally, recall that our framework - unlike standard factor IV - does not require $r \geq G$. In particular, size control is achieved, at least asymptotically, even when Λ_k is zero (in which case $r = 0$). From the power perspective: (i) consistency does not require using *all* relevant factors, and (ii) adding more factors even when relevant does not necessarily translate into effective power gains.

On balance, our Monte Carlo results reported below support the following simple strategy. If more than zero factors are selected via IC, we recommend disregarding the number of factors suggested by IC, and using G factors instead.

6 Monte Carlo evaluation

We conduct a simulation study based on the design and notation of DT, using the following Data Generation Process (DGP):

$$Y = Y_1\beta_1 + Y_2\beta_2 + U, \quad (Y_1, Y_2) = (F_1, F_2)\mathbf{\Pi}_2 + X_3\delta + (V_1, V_2), \quad \mathbf{X}_2 = (F_1, F_2)\mathbf{\Lambda_k} + \mathbf{E}, \tag{23}$$

$$(u_t, v_{1t}, v_{2t})' \overset{i.i.d.}{\sim} N(0, \Sigma),\ \Sigma = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .3 \\ .8 & .3 & 1 \end{pmatrix},\ t = 1, ..., T.$$

Here $\mathbf{X}_2$ is a $T \times k_2$ matrix of valid instruments and X_3 is a $T \times 1$ omitted instrument vector which is not taken into account when computing the different statistics. We consider X_3 in line with DT. In contrast with DT who generate each of the elements of $\mathbf{X}_2$ as $i.i.d.$ $N(0,1)$, we however assume that $\mathbf{X}_2$ is generated by a factor model, where F_1 and F_2 are $T \times 1$ unobservable independent factors whose elements are generated respectively as $i.i.d.$ $N(0, 0.4)$ and $i.i.d.$ $N(0, 0.2)$, $\mathbf{\Lambda_k}$ is a $2 \times k_2$ matrix of loadings whose elements are all i.i.d. draws from $N(1,1)$, while the error terms collected in the $T \times k_2$ matrix $\mathbf{E}$ are $i.i.d.$ $N(0, 0.4)$ and independent of the factors and all the other errors in the model. Hence, each element of $\mathbf{X}_2$ is $i.i.d.$ $N(0,1)$ as in DT, but it is correlated with each other element of $\mathbf{X}_2$ because of the common factors F_1 and F_2 that drive $\mathbf{X}_2$. The elements of X_3 are instead generated as $i.i.d.$ $N(0,1)$, independent of $\mathbf{X}_2$. The statistics under analysis are the original and factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003).

The correlation of each element of U with the corresponding elements of V_1 and V_2 is equal to 0.8, so that the variables Y_1 and Y_2 are endogenous and the instrumental variables $\mathbf{X}_2$ are necessary. We follow DT and set $\beta_1 = 1/2$, $\beta_2 = 1$, $\delta = \lambda(1,1)'$, where λ can take the values 0 or 1. The larger the value of λ, the more relevant the omitted instrument X_3, in the following sense: with respect to our theoretical framework, the error term of the regression of (Y_1, Y_2) on (F_1, F_2) includes the extra additive component $X_3\delta$ which affects the correlation between disturbances and thereby the extent of endogeneity. The matrix $\mathbf{\Pi}_2$ controls the strength of the instruments and it is defined as $\mathbf{\Pi}_2 = \rho\mathbf{\Pi}$, where ρ is either 0.01 (weak instruments) or 1 (strong instruments), and as in DT $\mathbf{\Pi}$ is obtained from the identity matrix by keeping the first k_2 lines and the first 2 columns. The number of instruments k_2 varies from 2 to 50, the sample size is $T = 100$, the number of replications is 1000, and the conditional LR critical values are computed using the same number of replications.

For each statistic, we consider first a version based on the $\mathbf{X}_2$ instruments, and then a factor based version where the unobservable factors F_1 and F_2 are substituted by the first two principal components of $\mathbf{X}_2$. We assess the size of the tests by computing the empirical rejection probability of the null hypothesis $H_0 : \beta = (1/2, 1)'$, when the nominal level of the tests is 5%. To evaluate the (size-adjusted) power, we test the same null hypothesis $H_0 : \beta = (1/2, 1)'$ when $\beta = (1/2 + x, 1 + x)'$ and x is either 0.1 or 0.5.

Size and power are evaluated for four parameters combinations. It can be either $\delta = 0$ [no omitted instrument] or $\delta = 1$ [X_3 is an omitted instrument]. For each value of δ, we then consider designs with either weak identification ($\rho = 0.01$) or strong identification ($\rho = 1$). The size and power results for instrument based statistics are presented in Tables 1 and 2 for, respectively, $\delta = 0$ and $\delta = 1$. Tables 3 and 4 present the corresponding results for the (estimated) factor based statistics. In each Table we report results for different values of k_2, ranging from 2 to 50.

Size results in Tables 1-2 suggest that except for the AR statistic, when $\mathbf{X}_2$ are considered as instruments, size distortions are more likely and more severe the more instruments are considered, even with strong instruments. Many and weak instruments create more sizeable distortions, and omission of X_3 when relevant exacerbates distortions. On balance and with weak instruments, the LR1 and LR2 tests are better behaved than the K test but remain oversized with large k_2. In contrast, and conforming with our finite sample analysis, the AR test has the correct size for all parameter configurations considered. Its chi-square counterpart (ARS) performs slightly worse, yet as argued by DT, this is a design in which the F critical value is appropriate and thus preferable in finite samples.

For all the considered tests (except AR which is correctly sized), size distortions increase progressively with k_2, which suggests that our factor based solution holds promise. Indeed, Tables 3-4 confirm that when the two estimated factors are used instead of $\mathbf{X}_2$, size distortions of all tests based on these factors basically disappear. All the empirical rejection frequencies are in the range 4%-6% for any value of the δ parameter, ρ and k_2. So for all considered parameter configurations, our factor based tests achieve size control. This result is remarkable given the small sample considered here ($T = 100$).

Table 3 provides equally remarkable results on power, given strong instruments. With $x = 0.1$, all factor based tests, including the AR test, are systematically more powerful than their standard counterpart. In addition, while the power of the latter decreases with k_2, the factor based statistics are not affected as k_2 increases (power seems stable over k_2). When $x = 0.5$, all statistics have power close to one. When $\delta = 1$, and the omitted instrument becomes relevant, Tables 2 and 4 underscore the power gains resulting from factors. In particular, while the standard tests have practically no power with large k_2, the power of their factor counterparts remains close to one. Finally, as expected, a comparison of Tables 1 and 3 suggests that when the instruments are weak all statistics have no power.

Another issue we explore in this Monte Carlo study is the question of how the standard Wald test works in our framework. It is well known that the Wald test suffers in the presence of a fixed number of weak instruments. It is therefore instructive to consider its behaviour in this study. Preliminary investigation shows that if an increasing number of instruments is used, the test overrejects very badly irrespective of whether the instruments are weak or strong. Therefore, we focus on the case where the factors are used as instruments. We use the setup underlying Table 3 (i.e. no omitted instruments). Similar results are obtained in the case where there is an omitted instrument. When the instruments (factors) are strong the Wald test is correctly sized but has very low power (similar to the rejection probabilities under the null) in the case where $x = 0.1$. The power is high when $x = 0.5$. In the more interesting case of weak instruments the test is undersized when two factors are used. We then consider the case where there are two true factors but we extract more than two factors from the large set of instruments. While the AR test remains well behaved, in all cases, under the null hypothesis the Wald test starts overrejecting for more than five extracted factors and this overrejection reaches extremely high levels when more than seven factors are extracted (>40%).

Another experiment considers the size of the factor and standard tests when $\Lambda_k = 0$ in (12), namely, we use principal components as instruments in the absence of a factor structure, a case considered theoretically in Section 4. We report rejection probabilities under the null in Table 5. It turns out that all factor tests behave very well even though the dataset used to extract factors does not in fact contain any factor structure. On the other hand, standard tests have problematic rejection probabilities as expected.

A further set of experiments with results reported in Table 6 relates to the question of how many factors to use, which we addressed in subsection 5.3. Here we consider the case of strong instruments ($\rho = 1$) with an extra missing instrument ($\delta = 1$). For power results we set $x = 0.1$. We allow the true (r) and assumed ($\hat{r}$) number of factors to vary. In particular, we consider three experiments: Experiment FN_A: $r = 3$, $\hat{r} = 3$, Experiment FN_B: $r = 3$, $\hat{r} = 2$ and Experiment FN_C: $r = 1$, $\hat{r} = 2$. $G = 2$ throughout. As discussed in subsection 5.3 our preferred strategy is to set $\hat{r} = 2$, since $G = 2$. For experiments FN_A and FN_B we only present results for the factor tests while for the experiment FN_C we also present comparative results for the standard tests. While for FN_A we get slight size distortions in the form of underrejections for a few tests, these problems are in general reduced in experiments FN_B. Power results remain good throughout, irrespective of the true number of factors. From experiments FN_C, it turns out that the generally better size and power performance of the factor statistics with respect to the standard statistics remains unchanged when the number of factors is overestimated ($r = 1$, $\hat{r} = 2$). Overall, it is clear from Table 6 that our preferred strategy of setting $\hat{r} = G$ makes sense, once one takes into account the overall performance of the test both under the null and alternative hypotheses.

To further interpret these results, recall the definition of a point-optimal instrument from (5). In this context, given any $T \times G$ instrument matrix $\mathbf{Z}$ and any full-rank estimate (denoted $\mathbf{\Pi}^*$) of $\mathbf{\Pi}$, $M(\mathbf{Z}\mathbf{\Pi}^*) = M(\mathbf{Z})$ so $AR(\beta^0|\bar{Z}) = AR(\beta^0|\mathbf{Z})$ for the point-optimal instrument $\bar{Z} = \mathbf{Z}\mathbf{\Pi}^*$. This implies [on referring to (4)] that the associated Kleibergen statistic will numerically coincide with its ARS counterpart if (7) produces an invertible estimate. Furthermore, the minimum in (9) underlying the Moreira statistic will be zero; the LR_1 criterion will thus also coincide with AR up to a monotonic transformation. The latter in this case will be point-optimal in the sense of Dufour (2003) [see also Dufour and Taamouti (2006)] with strong instruments, again because $AR(\beta^0|\mathbf{Z}\mathbf{\Pi}) = AR(\beta^0|\mathbf{Z})$ for any full-rank $\mathbf{\Pi}$. This suggests that setting $\hat{r} = G$ makes sense from a point-optimal perspective, which supports our simulation evidence.

We carry out two final sets of experiments as robustness checks of our Monte Carlo design. The first one reexamines the weak factor power results of Table 3 where there is no omitted variable and factors are used as instruments. In that case the instruments were too weak for any test to have power. In the new experiment we raise ρ from 0.01 to 0.5. Results are presented in Table 7. We see that power increases when x is either 0.1 or 0.5 with the latter value leading to a very substantial increase in power as expected. The second robustness check reexamines the results of Table 4 where there is an omitted variable and factors are used as instruments. In the new experiment we allow a correlation of 0.5 between the first factor and the omitted variable. We expect that this will not affect rejection probabilities under the null hypothesis but will lead to a substantial rise in rejection probabilities under the alternative since now the factor instruments can capture some of the effects of the omitted variable. Results reported in Table 8 verify this conjecture.

In summary, these experiments indicate very clearly that using factors as instruments can solve the size problems of all the identification robust statistics, and increase their power, including that of the AR-statistic. Factor-reduction thus provides a promising win/win solution to the size/power trade-offs arising from over-instrumentation, even with small samples.

7 Empirical Analysis

For our empirical analysis, we consider the prototypical New Keynesian model [see Clarida, Galí, and Gertler (1999), Linde (2005) and Dufour, Khalaf, and Kichian (2013)]

$$\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda_\pi y_t + \varepsilon_{\pi,t} \,, \tag{24}$$

$$y_t = \beta_f E_t y_{t+1} + \sum_{j=1}^{4} (1 - \beta_f) \beta_{y,j} y_{t-j} - \beta_r^{-1} (R_t - E_t \pi_{t+1}) + \varepsilon_{y,t} \,, \tag{25}$$

$$R_t = \gamma_\pi \left(1 - \sum_{j=1}^{3} \rho_{R,j}\right) \pi_t + \gamma_y \left(1 - \sum_{j=1}^{3} \rho_{R,j}\right) y_t + \sum_{j=1}^{3} \rho_{R,j} R_{t-j} + \varepsilon_{R,t} \,, \tag{26}$$

where, for $t = 1, \ldots, T$, π_t is aggregate inflation, y_t is the output gap, R_t is the nominal interest rate, and $\varepsilon_{\pi,t}$, $\varepsilon_{y,t}$, and $\varepsilon_{R,t}$ are zero-mean possibly contemporaneously correlated disturbances.

The past decade has seen important changes in estimating related models, many motivated by a desire to minimize non-theory-based assumptions. These include assumptions on disturbances examples of which we study below, and, perhaps more to the point in this paper, the definition of conditioning information. Many popular specifications including DSGE systems exclude empirically relevant information because of their specificity. Concrete examples discussed in *e.g.* Consolo, Favero, and Paccagnini (2009) include the following: fiscal policy variables in a DSGE model focusing on monetary policy, foreign sector variables in a DSGE model for a closed economy, or financial indicators and term structure variables that affect the decision of policy makers. Reviews of DSGE-VAR methods raise the following different although related concerns.[10] DSGE-VAR methods broadly assess the considered DSGE against an unrestricted VAR where - regardless of statistical fit - the variables that are included in the latter [here π_t, y_t and R_t] are exactly determined by those that enter the former. More general benchmarks are gaining popularity, including factor based specifications. For all these reasons, and in response to the serious empirical concerns summarized in particular by the surveys of Schorfheide (2013) and Mavroeidis, Plagborg-Moller, and Stock (2014), we focus on this model to assess the potential contribution of factors to its identification.

Equation (24) is a New Keynesian Phillips Curve (NKPC) which is perhaps one of the first structural macroeconomic models that have been estimated using identification-robust methods. The restriction

$$\gamma_b = (1 - \gamma_f) \tag{27}$$

[10]See *e.g.* Del Negro, Schorfheide, Smets, and Wouters (2007) and discussions by Christiano L. J., Gallant A. R., Sims C. A., Faust J. and L. Kilian. See also Boivin and Giannoni (2006), Consolo, Favero, and Paccagnini (2009).

is often maintained in theoretical and applied works. Equation (26) is a Taylor rule. On the latter, identification-robust methods applied in particular by Mavroeidis (2010) on the U.S. reveal instabilities in determinacy, an important issue that seems to have escaped notice via traditional estimation. The output equation (25) serves to close the system which may then be viewed as a DSGE model. In contrast to full information methods which require and impose a unique rational expectation solution and given recent works that challenge such requirements [see Cochrane (2011) or Mavroeidis (2010)], we analyze the system using limited information IV-based methods.

Whether treated as a DSGE or not, the following features characterize the considered system: (i) (26) is not expectation-based; (ii) (24) is linear in parameters; (iii) lags of endogenous variables in (24) and (26) target *i.i.d.* disturbances; (iv) the absence of cross-equation restrictions on parameters permits single equation estimation within the system. For the latter reason, in addition to the multi-equation limited and full-information methods[11], Dufour, Khalaf, and Kichian (2013) provide a single-equation AR-based analysis using enough instruments in each case (at least six here) to satisfy the basic order condition for system identification. We proceed in this way, with further variants described next.

In the absence of a consensus view in this literature, the set of assumptions (i)-(iv) is viewed as illustrative. However, we revisit two fundamental restrictions within this set. First, we consider, in addition to the linear (24), the Galí and Gertler (1999) and Galí, Gertler, and Lopez-Salido (2001) set-up in which

$$\lambda_\pi = \frac{(1-\omega_\pi)(1-\theta_\pi)(1-\phi\theta_\pi)}{\theta_\pi + \omega_\pi[1-\theta_\pi(1-\phi)]}, \tag{28}$$

$$\gamma_f = \frac{\phi\theta_\pi}{\theta_\pi + \omega_\pi[1-\theta_\pi(1-\phi)]}, \quad \gamma_b = \frac{\omega_\pi}{\theta_\pi + \omega_\pi[1-\theta_\pi(1-\phi)]}, \tag{29}$$

where θ_π measures the degree of price stickiness, ω_π the degree of "backwardness" in price setting, and ϕ is the discount factor which we set to .99. In this specification, Galí and Gertler (1999) represent y_t by marginal costs. Furthermore, we estimate, in addition to (26), a forward-looking Taylor rule with fewer lags following Mavroeidis (2010):

$$R_t = \gamma_0 + \sum_{j=1}^{2} \rho_{R,j} R_{t-j} + \left(1 - \sum_{j=1}^{2} \rho_{R,j}\right) \gamma_\pi E_t \pi_{t+1} + \left(1 - \sum_{j=1}^{2} \rho_{R,j}\right) \gamma_y y_t + \varepsilon_{R,t}. \tag{30}$$

Each equation is represented as an IV regression, where expectations of variables in (24) and (25) are replaced with actual values plus errors:

$$\pi_{t+1} = E_t \pi_{t+1} + \upsilon_{\pi,t+1}, \quad y_{t+1} = E_t y_{t+1} + \upsilon_{y,t+1} \tag{31}$$

[11] These methods are not covered by the theoretical set-up of the present paper.

leading to

$$\pi_t = \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda_\pi y_t + (\varepsilon_{\pi,t} - \gamma_f \upsilon_{\pi,t+1}) \ , \tag{32}$$

$$y_t = \beta_f y_{t+1} + \sum_{j=1}^{4} (1-\beta_f)\,\beta_{y,j} y_{t-j} - \beta_r^{-1}(R_t - \pi_{t+1}) + \left(\varepsilon_{y,t} - \beta_f \upsilon_{y,t+1} - \beta_r^{-1}\upsilon_{\pi,t+1}\right) \tag{33}$$

and, in the case of (30), to

$$R_t = \gamma_0 + \sum_{j=1}^{2} \rho_{R,j} R_{t-j} + \left(1 - \sum_{j=1}^{2} \rho_{R,j}\right) \gamma_\pi E_t \pi_{t+1} + \left(1 - \sum_{j=1}^{2} \rho_{R,j}\right) \gamma_y y_t$$
$$+ \left(\varepsilon_{R,t} - \left(1 - \sum_{j=1}^{2} \rho_{R,j}\right) \gamma_\pi \upsilon_{\pi,t+1}\right).$$

In this context, even if $\varepsilon_{\pi,t}$, $\varepsilon_{y,t}$, and $\varepsilon_{R,t}$ are contemporaneously uncorrelated and *i.i.d.* across time, the compounded disturbances $(\varepsilon_{\pi,t} - \gamma_f \upsilon_{\pi,t+1})$, $(\varepsilon_{y,t} - \beta_f \upsilon_{y,t+1} - \beta_r^{-1}\upsilon_{\pi,t+1})$ and, in the case of (30), $(\varepsilon_{R,t} - (1 - (\rho_{R,1} + \rho_{R,2}))\gamma_\pi \upsilon_{\pi,t+1})$, can be autocorrelated at lag 1 if:

$$\text{(a): } \varepsilon_{\pi,t} \text{ is contemporaneously correlated with } \upsilon_{\pi,t}, \tag{34}$$

$$\text{(b): } \varepsilon_{y,t} \text{ is contemporaneously correlated with } \upsilon_{\pi,t} \text{ or } \upsilon_{y,t}, \tag{35}$$

$$\text{(c): } \varepsilon_{R,t} \text{ is contemporaneously correlated with } \upsilon_{\pi,t}, \tag{36}$$

respectively. Our single equation analysis thus considers both *i.i.d.* and dependent disturbance assumptions, the former subsuming

$$\text{(d): } \varepsilon_{\pi,t},\ \varepsilon_{y,t} \text{ and } \varepsilon_{R,t} \text{ are contemporaneously uncorrelated with } \upsilon_{\pi,t} \text{ and } \upsilon_{y,t}, \tag{37}$$

and the latter via HAC weighting matrices.

We use the same data as Dufour, Khalaf, and Kichian (2013): a quarterly sample for the U.S. from 1962Q1 to 2005Q3. The GDP deflator and the Fed Funds rate are used for the price level and the short-run interest rate respectively, and the real-time output gap introduced by Dufour, Khalaf, and Kichian (2013) is considered to ensure that lags of y_t remain valid instruments. In our single equation analysis of the NKPC, we also use the marginal cost as a driving variable since we revisit the Galí, Gertler, and Lopez-Salido (2001) specification. Finally, as in Linde (2005) and Dufour, Khalaf, and Kichian (2013), all our data is demeaned prior to estimation, which implies in particular that equation constants (if any) are partialled out as free parameters. We report a sub-sample analysis pre and post 1985, to reflect instability (of models and policy) issues raised, in particular, by Benati (2008) and Mavroeidis (2010).

As the "baseline" specification, we use the instrument set from Dufour, Khalaf, and Kichian (2013), which consists of lag one of π_t, lag one to four of y_t and lag one to three of R_t, as well as lag two and three of both wage and commodity price inflation. This set provides a standard base case consistent with the literature. Kapetanios and Marcellino (2010) use factors extracted via principle components from the 132 variables in Stock and Watson (2005). We also adopt the latter data-rich perspective. We thus assume that relevant instruments are driven by a few common forces, on which factors can provide an exhaustive summary. In this sense, factors may parsimoniously capture all relevant model-based and statistical information excluded from the theoretical New Keynesian model. We consider the first six factors and use the first lags of these as our instrument set.

We rely on the *i.i.d.* and AR-HAC [in (22)] statistics which we invert, in turn. That is, we collect the set of parameter values that are not rejected at the 5% level, over a model-relevant search set (reported below in each case). If this set is empty, then the model is rejected at the 5% level.[12] Rejection, in the following discussion relates to a 5% level. Rather than scrutinizing results in details, we analyse results with focus on the subject-matter of this paper, namely the contribution of factors.

1. The *i.i.d.* confidence intervals are more informative than their HAC counterparts. As (37) is relaxed, some precision cost is not unexpected. It is unlikely to be the whole story though, as the precision differences we observe are often sizable. Recall that an *i.i.d.* assumption is compatible with a model variant given (37), yet equilibrium principles provide no guidance on whether (34)-(36) must be considered instead. Our results suggest that this dilemma matters importantly and must be treated as an identification issue. While Kleibergen and Mavroeidis (2009) for example [see Mavroeidis, Plagborg-Moller, and Stock (2014) for further references] adopt (34)-(36) with a linear NKPC, Schorfheide (2013) supports (37) with a prototype Taylor rule, in the following sense: an *i.i.d.* assumption implies that lags of commonly used predictors are valid instruments, and less restrictive assumptions complicate finite sample distributions of available statistics. These same arguments are discussed by Cochrane (2011). We find this problem to be serious beyond the special cases discussed by these authors. Can a case be made that fundamentals rather than methodological issues drive results? Said differently, can one differentiate between the statistical properties of the considered HAC statistic in small samples, and the time dependence assumption *per se?*. Relevant simulation studies based on sample sizes consistent with macro-economics are scarce;

[12]The largest test p-value over the search set provides a measure of model fit, in the spirit of a J-type test; the associated parameter values (that is, the parameter values that maximize this p-value) can be considered as point estimates. We obsered that the p-value function was quite flat in parameters, which is expected given documented identification difficulties in this literature. We thus prefer to rely on the projection-based confidence sets that adequately reflect estimation uncertainty.

for example, Kleibergen and Mavroeidis (2009) use $T = 1000$ in reported experiments.[13] Our empirical findings thus raise an important econometric theory question for further research. Deeper issues regarding the Taylor rule itself are discussed below.

2. In contrast to the main point in Dufour, Khalaf, and Kichian (2013), namely that a forward looking NKPC cannot be confirmed with single-equation methods, we find that using factors sharpens inference on γ_f with *i.i.d.* specifications.[14] In particular, in the post-1985 sample, factors confirm that the output equation is more forward than backward looking: in sharp contrast with the baseline set, a confidence set exceeding .5 is obtained with factors despite the underlying small sample. Partial intuition for this result is that the large data set we draw our factors from bridges the gap between single equation and systems inference, since it synthesizes information that policy makers effectively use.

3. In contrast to one of the results in Kleibergen and Mavroeidis (2009), namely that using the gap or the marginal cost produces similar results for inference on the NKPC, we find that the driving variable matters with regards to γ_f. In particular, we cannot confirm that inflation is more forward than backward looking with the marginal cost variable. From the substantive side, we do not aim to take a stand on the controversial implications of this finding. In fact, methodological rather than fundamental issues may partly explain observed discrepancies. The specific gap measure we use is a real-time filter, in the sense that the time-t value does not use data beyond time t; its lags are thus valid instruments. Predetermined variables are often taken for granted as valid instruments. The feedback counter-example we raised in point 1 calls for caution in this regard. For different although related reasons, filtering may also undermine weak-exogeneity of lags. See Basturk, Cakmakli, Ceyhan, and van Dijk (2014) for a recent perspective on filtering.

4. With non-rejected specifications, we note that a zero value for the coefficient of driving variables in all three equations, except when it is restricted out as in the Galí, Gertler, and Lopez-Salido (2001) model, cannot be refuted even when the estimated confidence set is tight. In addition to conclusions in *e.g.* Mavroeidis, Plagborg-Moller, and Stock (2014), addressing this enduring puzzle is a valuable research direction since flat curves challenge theory fundamentally.

5. Our results on the Taylor rule differ from Mavroeidis (2010) who argues that: (i) identification problems are much more prominent post-Volcker which suggests that policy did not necessarily satisfy the Taylor principle in this sub-period, as originally shown by Clarida,

[13] Alternative identification-robust IV-based methods that permit dependence and provably perform well in small samples are not available.

[14] The multivariate test in question from Dufour, Khalaf, and Kichian (2013) is also *i.i.d.* based.

Galí, and Gertler (2000), and may have remained inactive instead; (ii) in contrast, estimates are quite precise pre-Volcker and support those of Clarida, Galí, and Gertler (2000) pointing to inactive policy. In our case, results with HAC suggest weak identification throughout the sample (not just post 80s) with both forms of the Rule we studied. However, with benchmark instruments and an *i.i.d.* test on (30), our confidence sets on γ_π and γ_y appear relatively sharp in both sub-samples and their intersection can be interpreted along Taylor principles and conventional assumptions on determinacy. These findings may challenge the conclusion of Mavroeidis (2010), but we can and must ask whether underlying assumptions are reasonable.[15] Two plausible interpretations of results, taken collectively, can be considered. On the one hand, it could be the case that instabilities in determinacy undermine identification, as argued in Mavroeidis (2010). On the other hand, fragility of empirical estimates may also be linked to the usual culprits, that is, may have more to do with an unsuitable lag or dependence structure. In the absence of other evidence, econometric methods have difficulties distinguishing between these two interpretations. Cochrane (2011) discusses an alternative interpretation suggesting that inflation is indeterminate under "active" as well as "passive" policy. Empirically, this should translate into global identification failure, unless counterfactual assumptions are imposed. Our results with factors seem to support Cochrane's perspective: using factors, our *i.i.d.* sets are empty (that is, statistically counterfactual) while their HAC-counterparts return basically the full search set, a contrast we did not obtain with benchmark instruments.

These results must be interpreted recognizing that instrument selection is unlikely to solve all challenges that plague the New Keynesian model. On balance, our results suggest that factor-based instruments, despite imperfect identification, extract useful information which improves our understanding of important substantive dilemmas in this literature.

8 Conclusions

In this paper we focus on identification robust inferential procedures and make four main contributions to the literature. First, we introduce the factor-based counterparts of the AR method as well as of the K-test from Kleibergen (2002) and Moreira (2003). In our framework the endogenous regressors depend, weakly or strongly, on a number of unobservable factors,

[15] To be clear, in addition to different instrument sets, a key difference with Mavroeidis (2010) is that we use a real-time gap. Mavroeidis (2010) also splits the sample in 1979 rather than 1985. Robustness to break dates is beyond the scope of the present analysis as well as the analysis in Mavroeidis (2010) for that matter; see Magnusson and Mavroeidis (2014) for a recent perspective on stability restrictions. Formal tests of stability require a statistical framework beyond the scope of the present paper; see for example Coroneo, Corradi, and Santos Monteiro (2011).

and the possibly large set of available instruments depends, weakly or strongly, on these factors. In this context, we show that our proposed factor-based procedures achieve size control asymptotically, given commonly used regularity assumptions.

Second, we demonstrate that the factor-based AR statistic remains finite sample exact under usual assumptions. Specifically, we show that the statistic remains pivotal whether instruments are weak or strong, and whether the underlying factor structure is weak or strong. In addition, this result does not require any assumptions on the relative size of T, the available sample size, and k, the overall number of available instruments, as long as the number of retained factors is reasonably smaller than T. Our factor-AR test is thus useful even if k is larger than T, a case where the original AR test would be infeasible.

Third, using a set of Monte Carlo experiments, we show that our factor-based approach circumvents the size problems associated with Kleibergen's and Moreira's statistics, and improves the power of the AR statistic. Overall, no one test uniformly dominates the others following instrument reduction. Information-reduction thus provides an appealing way to tackle the instruments proliferation problem.

Finally, with an empirical study on New Keynesian macroeconomic models we provide evidence that our factor-based methods are easily implemented and can bridge a number of gaps between structural and statistical macroeconomic models.

References

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752.

Andrews, D. W. K., and J. H. Stock (2007): "Testing with Many Weak Instruments," *Journal of Econometrics*, 138, 24–46.

Andrews, I., and A. Mikusheva (2014): "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Quantitative Economics*, forthcoming.

Antoine, B., and P. Lavergne (2014): "Conditional Moment Models under Semi-Strong Identification," *Journal of Econometrics*, 182, 59–69.

Ashley, R. (2009): "Assessing the Credibility of Instrumental Variables Inference with Imperfect Instruments via Sensitivity Analysis," *Journal of Applied Econometrics*, 24, 325–337.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dinensions," *Econometrica*, 71, 135–173.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

——— (2010): "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26, 1577–1606.

Basturk, N., C. Cakmakli, P. Ceyhan, and H. K. van Dijk (2014): "Posterior-Predictive Evidence on US Inflation using Extended New Keynesian Phillips Curve Models with Non-filtered Data," *Journal of Applied Econometrics*, forthcoming.

Beaulieu, M.-C., J.-M. Dufour, and L. Khalaf (2007): "Multivariate Tests of Mean-Variance Efficiency with Possibly Non-Gaussian Errors: An Exact Simulation-Based Approach," *Journal of Business and Economic Statistics*, 25, 398–410.

——— (2010): "Asset-Pricing Anomalies and Spanning: Multivariate Multifactor Tests with Heavy Tailed distributions," *Journal of Empirical Finance*, 17, 763–782.

——— (2013): "Identification-Robust Estimation and Testing of the Zero-Beta CAPM," *The Review of Economics Studies*, 80, 892–924.

Benati, L. (2008): "Investigating Inflation Persistence across Monetary Regimes," *Quarterly Journal of Economics*, 123, 1005–60.

Berkowitz, D., M. Caner, and Y. Fang (2008): "Are Nearly Exogenous Instruments reliable," *Economics Letters*, 101, 20–23.

Boivin, J., and M. Giannoni (2006): "DSGE Models in a Data-Rich Environment," Discussion paper, NBER Working Paper No. 12772.

Boivin, J., and S. Ng (1981): "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 132, 169–194.

Canova, F., and L. Sala (2009): “Back to Square One: Identification Issues in DSGE Models,” *Journal of Monetary Economics*, 54, 431–49.

Carrasco, M., and G. Tchuente (2012): “Regularized LIML with Many Instruments,” Discussion paper, Université de Montréal, Montréal, Canada.

Chao, J., and N. R. Swanson (2005): “Consistent Estimation with a large number of Weak Instruments,” *Econometrica*, 73, 1673–1692.

Chaudhuri, S., T. Richardson, J. Robins, and E. Zivot (2010): “A New Split-Sample Score Test in Linear Instrumental Variables Regression,” *Econometric Theory*, 26, 1820–1837.

Chetty, R., J. Friedman, E. Glaeser, G. Imbens, and M. Kolesar (2011): “Identification and Inference with Many Invalid Instruments,” Discussion paper, Harvard University.

Clarida, R., J. Galí, and M. Gertler (1999): “The Science of Monetary Policy: A New Keynesian Perspective,” *Journal of Economic Literature*, 37, 1661–1707.

——— (2000): “Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory,” *The Quarterly Journal of Economics*, 115, 147–180.

Cochrane, J. H. (2011): “Determinacy and Identification with Taylor Rules,” *Journal of Political Economy*, 119, 565–615.

Consolo, A., C. Favero, and A. Paccagnini (2009): “On the Statistical Identification of DSGE Models,” *Journal of Econometrics*, 150, 99–115.

Coroneo, L., V. Corradi, and P. Santos Monteiro (2011): “Testing for the degree of commitment via set-identification,” Discussion paper, University of Surrey.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): “On the Fit of New Keynesian Models,” *Journal of Business and Economic Statistics*, 25, 123–43.

Doko-Tchatoka, F., and J.-M. Dufour (2008): “Instrument Endogeneity and Identification-Robust Tests: some Analytical Results,” *Journal of Statistical Planning and Inference*, 138, 2649–2661.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

——— (2003): "Identification, Weak Instruments and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767–808.

Dufour, J.-M., and J. Jasiak (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42, 815–843.

Dufour, J.-M., and L. Khalaf (2002): "Simulation Based Finite and Large Sample Tests in Multivariate Regressions," *Journal of Econometrics*, 111, 303–322.

Dufour, J.-M., L. Khalaf, and M. Kichian (2006): "Inflation Dynamics and the New Keynesian Phillips Curve: an Identification Robust Econometric Analysis," *Journal of Economic Dynamics and Control*, 30, 1707–1728.

——— (2010): "On the Precision of Calvo Parameter Estimates in Structural NKPC Models," *Journal of Economic Dynamics and Control*, 34, 1582–1595.

——— (2013): "Identification-Robust Analysis of DSGE and Structural Macroeconomic Models," *Journal of Monetary Economics*, 60, 340–350.

Dufour, J.-M., and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

——— (2006): "Point-Optimal Instruments and Generalized Anderson-Rubin Procedures for Nonlinear Models," Discussion paper, McGill University.

——— (2007): "Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments," *Journal of Econometrics*, 139, 133–153.

Galí, J., and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 44, 195–222.

Galí, J., M. Gertler, and J. D. Lopez-Salido (2001): "European Inflation Dynamics," *European Economic Review*, 45, 1237–1270.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Inference in Weakly Identified DSGE Models," *Quantitative Economics*, 4, 197–229.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012): "On the asymptotic sizes of subset Anderson-Rubin and Lagrange Multipliers tests in linear instrumental variables regression," *Econometrica*, 80, 2649–2666.

Guggenberger, P., and R. J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667–709.

——— (2008): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," *Journal of Econometric*, 142, 134–161.

Hahn, J., and J. Hausman (2005): "IV Estimation with Valid and Invalid Instruments," *Annales d'Économie et de Statistique*, 79/80, 25–57.

Hansen, C., J. Hausman, and W. Newey (2008): "Estimation With Many Instrumental Variables," *Journal of Business and Economic Statistics*, 26, 398–422.

Iskrev, N. (2010): "Local Identification in DSGE Models," *Journal of Monetary Economics*, 57, 189–202.

Kapetanios, G., and M. Marcellino (2010): "Factor-GMM estimation with large sets of possibly weak instruments," *Computational Statistics and Data Analysis*, 54(11), 2655–2675.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

——— (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103–1123.

Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics*, 27, 293–311.

Komunjer, I., and S. Ng (2011): "Dynamic Identification of DSGE Models," *Econometrica*, 79, 1995–2032.

Linde, J. (2005): "Estimating New Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach," *Journal of Monetary Economics*, 52, 1135–49.

Magnusson, L., and S. Mavroeidis (2010): "Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve," *Journal of Money Credit and Banking*, 42, 1582–1595.

——— (2014): "Identification Using Stability Restrictions," *Econometrica*, forthcoming.

Mavroeidis, S. (2010): "Monetary policy rules and macroeconomic stability: some new evidence," *American Economic Review*, 100, 491–503.

Mavroeidis, S., M. Plagborg-Moller, and J. H. Stock (2014): "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve," *Journal of Economic Literature*, 52, 124–88.

Mikusheva, A. (2009): "Comment on invited paper by Kleibergen and Mavroeidis "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve"," *Journal of Business and Economic Statistics*, 27, 293–311.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Nason, J., and G. Smith (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics*, 23, 525–51.

Newey, W., and K. D. West (1987): "A Simple, Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newey, W. K., and F. Windmeijer (2009): "GMM with Many Weak Moment Conditions," *Econometrica*, 77, 687–719.

Onatski, A. (2009): "A formal Statistical Test for the Number of Factors in the Approximate Factor Models," *Econometrica*, 77, 1447–1479.

——— (2012): "Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors," *Journal of Econometrics*, 168, 244–258.

Schorfheide, F. (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, ed. by D. Acemoglu, M. Arrelano, and E. Deckel, pp. 184–230. Cambridge University Press, Cambridge (UK).

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H. (2010): "The Other Transformation in Econometric Practice: Robust Tools for Inference," *The Journal of Economic Perspectives*, 24, 83–94.

Stock, J. H., and M. W. Watson (2002a): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indices," *Journal of Business and Economic Statistics*, 20, 147–162.

——— (2005): "Implications of dynamic factor models for VAR analysis," *mimeo*.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1097–1126.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

Stock, J. H., and M. Yogo (2005): "Asymptotic Distributions of Instrumental Variables Statistics With Many Instruments," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by J. H. Stock, and D. W. K. Andrews, pp. 109–120. Cambridge University Press, Cambridge, U.K.

Wang, J., and E. Zivot (1998): "Inference on Structural Parameters in Instrumental Variable Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

Appendix

Proof of Theorem 1

We need to show that

$$\frac{(T-r)(Y-\mathbf{Y}_1\beta^0)'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'(Y-\mathbf{Y}_1\beta^0)}{r(Y-\mathbf{Y}_1\beta^0)'\left(I-\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right)(Y-\mathbf{Y}_1\beta^0)}-\frac{(T-r)(Y-\mathbf{Y}_1\beta^0)'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'(Y-\mathbf{Y}_1\beta^0)}{r(Y-\mathbf{Y}_1\beta^0)'\left(I-\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\right)(Y-\mathbf{Y}_1\beta^0)}=o_p(1). \tag{38}$$

This holds if

$$U'\mathbf{F}\mathbf{H}(\mathbf{H}'\mathbf{F}'\mathbf{F}\mathbf{H})^{-1}\mathbf{H}'\mathbf{F}'U-U'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'U=U'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'U-U'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'U=o_p(1),$$

under the null hypothesis, and

$$\mathbf{Y}_1'\mathbf{F}\mathbf{H}(\mathbf{H}'\mathbf{F}'\mathbf{F}\mathbf{H})^{-1}\mathbf{H}'\mathbf{F}'\mathbf{Y}_1-\mathbf{Y}_1'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\mathbf{Y}_1=\mathbf{Y}_1'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Y}_1-\mathbf{Y}_1'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\mathbf{Y}_1=o_p(1),$$

under the alternative hypothesis, for any nonsingular rotation matrix $\mathbf{H}$. These are, in turn, satisfied if

$$\frac{\mathbf{F}'\mathbf{F}}{T}-\frac{\hat{\mathbf{F}}'\mathbf{F}}{T}=o_p(1) \tag{39}$$

$$\sqrt{T}\left(\frac{\mathbf{F}'\mathbf{Y}_1}{T}-\frac{\hat{\mathbf{F}}'\mathbf{Y}_1}{T}\right)=o_p(1) \tag{40}$$

and

$$\sqrt{T}\left(\frac{\mathbf{F}'U}{T}-\frac{\hat{\mathbf{F}}'U}{T}\right)=o_p(1) \tag{41}$$

hold. We examine (39)-(41). The left hand side of (39), (40) and that of (41) divided by $\sqrt{T}$ can be re-written as

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-Hf_t)q_t' \tag{42}$$

where q_t is either f_t, $y_{1,it}$ or u_t. By Lemma A.1 of Bai and Ng (2006) we have that

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-Hf_t)q_t'=O_p\left(\min(k,T)^{-1}\right) \tag{43}$$

as long as q_t has finite fourth moments, nonsingular covariance matrix and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(q_t-E(q_t))$ satisfies a central limit theorem. These conditions are satisfied for f_t, $y_{1,it}$ and u_t via assumptions 1-3. Hence, (39) and (40) follow, while (41) follows if $\sqrt{T}/k=o(1)$. Note, for later use, that a similar result holds for $q_t=y_t$.

Proof of Theorem 3

Given the consistency of $\hat{\mathbf{F}}$ for $\mathbf{FH}$ and the invariance of $N(\mathbf{F})$ to the use of $\mathbf{F}$ or $\mathbf{FH}$, we only need to show that

$$\hat{\lambda} - \lambda = o_p(1) \tag{44}$$

where

$$\lambda = \min_b \frac{b'\mathbf{Y}'\left(\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right)\mathbf{Y}b}{b'\Omega b}$$

and

$$\Omega = \mathbf{Y}'\left(I - \left(\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right)\right)\mathbf{Y}/(T-k).$$

Note that $\hat{\lambda}$ and λ correspond to the smallest eigenvalues of $\hat{\Omega}^{-1/2}\mathbf{Y}'\left(\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\right)\mathbf{Y}\hat{\Omega}^{-1/2}$ and $\Omega^{-1/2}\mathbf{Y}'\left(\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right)\mathbf{Y}\Omega^{-1/2}$ respectively Note further, that $\hat{\Omega}^{-1/2}\mathbf{Y}'\left(\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\right)\mathbf{Y}\hat{\Omega}^{-1/2}$ is a finite dimensional matrix whose every element converges in probability to the respective element of $\Omega^{-1/2}\mathbf{Y}'\left(\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right)\mathbf{Y}\Omega^{-1/2}$, if we can show that

$$\frac{\mathbf{Y}'\mathbf{F}}{T} - \frac{\mathbf{Y}'\hat{\mathbf{F}}}{T} = o_p(1), \quad \frac{\mathbf{F}'\mathbf{F}}{T} - \frac{\hat{\mathbf{F}}'\mathbf{F}}{T} = o_p(1) \tag{45}$$

hold. Let us assume for the moment that (45) holds. Then, given that the eigenvalues of a matrix, being polynomial roots, are continuous functions of the elements of the matrix, (44) follows by Slutsky's theorem. Hence, to complete the proof we need to demonstrate (45). But it follows immediately from (42)-(43) in the proof of Theorem 1.

Proof of Theorem 4

We need to show

$$AR(\beta^0|\mathbf{F}) = O_p\left(T^{1-2\vartheta}\right) \tag{46}$$

and

$$AR(\beta^0|\hat{\mathbf{F}}) - AR(\beta^0|\mathbf{F}) = o_p\left(T^{1-2\vartheta}\right) \tag{47}$$

To see (46), we have that

$$AR(\beta^0|\mathbf{F}) = \frac{T-r}{r}\frac{\left(U + (\mathbf{F\Pi}_T + \mathbf{V})\beta^{10}\right)'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\left(U + (\mathbf{F\Pi}_T + \mathbf{V})\beta^{10}\right)}{\left(U + \mathbf{F\Pi}_T\beta^{10}\right)'\left[I - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right]\left(U + \mathbf{F\Pi}_T\beta^{10}\right)}$$

where $\beta^{10} = \beta^1 - \beta^0$ But, since,

$$\|U'U\| = O_p(T), \quad \left\|U'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'U\right\| = O_p(1), \quad \left\|\mathbf{V}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}\right\| = O_p(1),$$

$$\left\|\beta^{10\prime}\mathbf{\Pi}_T'\mathbf{F}'\mathbf{F\Pi}_T\beta^{10}\right\| = O_p\left(T^{1-2\vartheta}\right),$$

it follows that

$$\frac{T-r}{r}\frac{1}{(U+\mathbf{F\Pi}_T\beta^{10})'\left[I-\mathbf{F}(\mathbf{F'F})^{-1}\mathbf{F}'\right](U+\mathbf{F\Pi}_T\beta^{10})}=O_p(1)$$

and

$$\left(U+\mathbf{F\Pi}_T\beta^{10}\right)'\mathbf{F}(\mathbf{F'F})^{-1}\mathbf{F}'\left(U+\mathbf{F\Pi}_T\beta^{10}\right)=O_p\left(T^{1-2\vartheta}\right)$$

thus proving (46). To prove (47), we have to show that

$$\frac{(T-r)((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})'\mathbf{F}(\mathbf{F'F})^{-1}\mathbf{F}'((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})}{r((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})'\left(I-\mathbf{F}(\mathbf{F'F})^{-1}\mathbf{F}'\right)((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})}-$$

$$\frac{(T-r)((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})}{r((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})'\left(I-\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\right)((U-\mathbf{V}\beta^{10})-\mathbf{F\Pi}_T\beta^{10})}=o_p(T^{1-2\vartheta}).$$

Given the proof of Theorem 1, this holds if

$$\left\|\mathbf{\Pi}_T'\mathbf{F'F}(\mathbf{F'F})^{-1}\mathbf{F'F\Pi}_T-\mathbf{\Pi}_T'\mathbf{F}'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\mathbf{F\Pi}_T\right\|=o_p(T^{1-2\vartheta}) \tag{48}$$

But

$$\left\|\mathbf{\Pi}_T'\mathbf{F'F}(\mathbf{F'F})^{-1}\mathbf{F'F\Pi}_T-\mathbf{\Pi}_T'\mathbf{F}'\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\mathbf{F\Pi}_T\right\|\leq\|\mathbf{\Pi}_T\|^2\,T\left\|\frac{\mathbf{F'F}}{T}-\frac{\mathbf{F}'\hat{\mathbf{F}}}{T}\left(\frac{\hat{\mathbf{F}}'\hat{\mathbf{F}}}{T}\right)^{-1}\frac{\hat{\mathbf{F}}'\mathbf{F}}{T}\right\|$$

Since,

$$\|\mathbf{\Pi}_T\|^2\,T=O_p(T^{1-2\vartheta})$$

(48) holds if

$$\left\|\frac{\mathbf{F'F}}{T}-\frac{\mathbf{F}'\hat{\mathbf{F}}}{T}\left(\frac{\hat{\mathbf{F}}'\hat{\mathbf{F}}}{T}\right)^{-1}\frac{\hat{\mathbf{F}}'\mathbf{F}}{T}\right\|=o_p(1)$$

which holds by (39), proving (47) and the Theorem.

Proof of Theorem 6

Consider (13) and impose the null hypothesis which yields

$$Y-\mathbf{Y}_1\beta^0=U. \tag{49}$$

Plugging the latter into (14), we are led back to (15), which will provide the desired result provided $\mathbf{F}$ and $\mathbf{E}$ are independent from U so that $\mathbf{X}$ as well as its principle components are independent from U.

Proof of Theorem 7

In this context

$$Y - \mathbf{Y}_1\beta^0 = \mathbf{Y}_1\left(\beta - \beta^0\right) + U = \mathbf{X\Pi}\left(\beta - \beta^0\right) + U + \mathbf{V}\left(\beta - \beta^0\right) \tag{50}$$

which, under the null hypothesis, and regardless of the dimension or rank of $\mathbf{X}$, and regardless of the rank of $\mathbf{\Pi}$, yields $Y-\mathbf{Y}_1\beta^0 = U$. Plugging the latter into (14), then for any $\mathbf{Z}$, $AR(\beta^0|\mathbf{Z})$ is distributed, under the null hypothesis, like (15) which provides the desired result.

Proof of Theorem 8

We need to show that

$$T^{-1}\left(Y - \mathbf{Y}_1\beta^0\right)'\mathbf{F}\hat{S}_{fu}^{-1}\mathbf{F}'\left(Y - \mathbf{Y}_1\beta^0\right) -$$

$$T^{-1}\left(Y - \mathbf{Y}_1\beta^0\right)'\hat{\mathbf{F}}\hat{S}_{\hat{f}u}^{-1}\hat{\mathbf{F}}'\left(Y - \mathbf{Y}_1\beta^0\right) = o_p(1)$$

or, introducing normalization terms,

$$\frac{U'\hat{\mathbf{F}}}{T^{1/2}}\hat{S}_{\hat{f}u}^{-1}\frac{\hat{\mathbf{F}}'U}{T^{1/2}} - \frac{U'\mathbf{F}}{T^{1/2}}\hat{S}_{fu}^{-1}\frac{\mathbf{F}'U}{T^{1/2}} = o_p(1)$$

where $\widehat{S}_{fu,h} = \Phi_0 + \sum_{j=1}^{h}\left(1 - \frac{j}{h+1}\right)\left(\Phi_j + \Phi_j'\right)$, $\Phi_j = T^{-1}\sum_{T=j+1}^{T}\hat{u}_t^f\hat{u}_{t-j}^f f_t f_{t-j}'$. In the proofs of previous Theorems we have shown that $\sqrt{T}\left(\frac{\hat{\mathbf{F}}'U}{T} - \frac{\mathbf{F}'U}{T}\right) = o_p(1)$. The proof of the theorem is complete if we show formally that $\Phi_j - \widehat{\Phi}_j = o_p(h^{-1})$. We show below that $\Phi_0 - \widehat{\Phi}_0 = o_p(h^{-1})$. The result for $j > 0$ follows similarly. We have

$$\left\|T^{-1}\sum_{t=j+1}^{T}\left(\widehat{u}_t^f\right)^2 f_t f_t' - T^{-1}\sum_{t=j+1}^{T}\left(\widehat{u}_t^{\hat{f}}\right)^2 \hat{f}_t\hat{f}_t'\right\| \leq C_1\left\|T^{-1}\sum_{t=j+1}^{T}\left(\widehat{u}_t^f\right)^2 f_t'\left(Hf_t - \widehat{f}_t\right)\right\| + \tag{51}$$

$$C_2\left\|T^{-1}\sum_{t=j+1}^{T}\widehat{u}_t^f\left(\widehat{u}_t^f - \widehat{u}_t^{\hat{f}}\right)f_t f_t'\right\|$$

for some constants C_1,C_2. First, we consider the first term of the RHS of (51). We have

$$\left\|T^{-1}\sum_{t=j+1}^{T}\left(\widehat{u}_t^f\right)^2 f_t'\left(Hf_t - \widehat{f}_t\right)\right\| \leq$$

$$C_3\left\|T^{-1}\sum_{t=j+1}^{T}u_t^2 f_t'\left(Hf_t - \widehat{f}_t\right)\right\| + C_4\left\|T^{-1}\sum_{t=j+1}^{T}(\widehat{u}_t^f - u_t)f_t'\left(Hf_t - \widehat{f}_t\right)\right\|$$

for some constants C_3,C_4. But, by Lemma A.1 of Bai and Ng (2006), if $\sqrt{T}/k = o(1)$,

$$\left\| T^{-1} \sum_{t=j+1}^{T} u_t^2 f_t^{'} \left(H f_t - \widehat{f}_t \right) \right\| = o_p \left(T^{-1/2} \right)$$

as long as u_t has finite eighth moments. Then,

$$\left\| T^{-1} \sum_{t=j+1}^{T} (\widehat{u}_t^f - u_t) f_t^{'} \left(H f_t - \widehat{f}_t \right) \right\| \leq C_5 \left\| \hat{\vartheta}^f - \vartheta \right\| \left\| T^{-1} \sum_{t=j+1}^{T} f_t f_t^{'} \left(H f_t - \widehat{f}_t \right) \right\|$$

for some constant C_5. Again, by Lemma A.1 of Bai and Ng (2006), if $\sqrt{T}/k = o(1)$,

$$\left\| T^{-1} \sum_{t=j+1}^{T} f_t f_t^{'} \left(H f_t - \widehat{f}_t \right) \right\| = o_p \left(T^{-1/2} \right).$$

By consistency of $\hat{\vartheta}^f$ (note that ϑ is equal to zero under the null hypothesis), $\left\| \hat{\vartheta}^f - \vartheta \right\| = o_p(1)$. Next, we consider the second term on the RHS of (51). We have

$$\left\| T^{-1} \sum_{t=j+1}^{T} \widehat{u}_t^f \left(\widehat{u}_t^f - \widehat{u}_t^{\hat{f}} \right) f_t f_t^{'} \right\| \leq C_6 \left\| T^{-1} \sum_{t=j+1}^{T} u_t f_t f_t^{'} \left(H f_t - \widehat{f}_t \right) \right\| \tag{52}$$

for some constant C_6. By similar arguments to those above the RHS of (52) is $o_p \left(T^{-1/2} \right)$ if $\sqrt{T}/k = o(1)$. Hence, $\Phi_0 - \widehat{\Phi}_0 = o_p(h^{-1})$ as long as $h = o(T^{1/2})$.

Table 1. Variables as instruments, no omitted instruments

	AR	ARS	K	LM	LR	LR1	LR2		AR	ARS	K	LM	LR	LR1	LR2
K2								K2							
			Weak instruments								Strong instruments				
			Size								Size				
2	0.039	0.043	0.043	0.038	0.043	0.043	0.043	2	0.057	0.063	0.063	0.057	0.061	0.061	0.063
3	0.054	0.057	0.055	0.051	0.046	0.053	0.054	3	0.048	0.054	0.051	0.018	0.020	0.052	0.054
4	0.047	0.053	0.058	0.044	0.037	0.054	0.057	4	0.049	0.054	0.056	0.011	0.011	0.058	0.060
5	0.046	0.057	0.061	0.041	0.035	0.055	0.056	5	0.050	0.057	0.062	0.006	0.006	0.058	0.063
10	0.048	0.063	0.074	0.040	0.016	0.070	0.074	10	0.043	0.063	0.050	0.000	0.001	0.049	0.052
20	0.058	0.093	0.110	0.034	0.008	0.111	0.106	20	0.043	0.069	0.064	0.000	0.000	0.066	0.069
40	0.055	0.116	0.170	0.015	0.003	0.141	0.137	40	0.057	0.124	0.082	0.000	0.000	0.085	0.082
50	0.053	0.144	0.229	0.003	0.002	0.192	0.181	50	0.040	0.124	0.069	0.000	0.000	0.077	0.081
			Power (size corrected) x=0.1								Power (size corrected) x=0.1				
2	0.064	0.064	0.064	0.064	0.064	0.064	0.064	2	0.147	0.147	0.147	0.147	0.147	0.147	0.147
3	0.037	0.037	0.045	0.037	0.046	0.046	0.044	3	0.179	0.179	0.225	0.185	0.221	0.221	0.224
4	0.061	0.061	0.062	0.060	0.062	0.062	0.062	4	0.142	0.142	0.166	0.117	0.167	0.167	0.165
5	0.041	0.041	0.046	0.041	0.037	0.037	0.034	5	0.199	0.199	0.253	0.156	0.241	0.241	0.241
10	0.047	0.047	0.053	0.047	0.046	0.046	0.047	10	0.138	0.138	0.288	0.034	0.273	0.273	0.272
20	0.044	0.044	0.045	0.042	0.036	0.036	0.036	20	0.127	0.127	0.301	0.001	0.284	0.284	0.294
40	0.054	0.054	0.036	0.052	0.046	0.046	0.060	40	0.078	0.078	0.228	0.002	0.198	0.198	0.187
50	0.035	0.035	0.052	0.029	0.058	0.058	0.057	50	0.109	0.109	0.279	0.000	0.256	0.256	0.247
			Power (size corrected) x=0.5								Power (size corrected) x=0.5				
2	0.073	0.073	0.073	0.073	0.073	0.073	0.073	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.043	0.043	0.045	0.038	0.051	0.051	0.049	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.061	0.061	0.061	0.038	0.062	0.062	0.062	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.057	0.057	0.053	0.038	0.049	0.049	0.047	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.056	0.056	0.058	0.021	0.054	0.054	0.056	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20	0.046	0.046	0.045	0.010	0.038	0.038	0.039	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.050	0.050	0.047	0.000	0.047	0.047	0.058	40	1.000	1.000	1.000	0.989	1.000	1.000	1.000
50	0.051	0.051	0.041	0.000	0.058	0.058	0.057	50	1.000	1.000	1.000	0.868	1.000	1.000	1.000

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 2. Variables as instruments, omitted instrument, δ=1

K2	AR	ARS	K	LM	LR	LR1	LR2	K2	AR	ARS	K	LM	LR	LR1	LR2
			Weak instruments								Strong instruments				
			Size								Size				
2	0.045	0.048	0.048	0.043	0.047	0.047	0.048	2	0.040	0.043	0.043	0.040	0.042	0.042	0.043
3	0.069	0.076	0.090	0.063	0.074	0.076	0.076	3	0.045	0.052	0.039	0.012	0.020	0.042	0.042
4	0.048	0.050	0.104	0.045	0.050	0.052	0.054	4	0.047	0.053	0.051	0.009	0.010	0.048	0.049
5	0.040	0.052	0.127	0.033	0.047	0.049	0.050	5	0.051	0.060	0.054	0.007	0.007	0.059	0.062
10	0.043	0.053	0.305	0.035	0.057	0.059	0.059	10	0.047	0.055	0.069	0.001	0.001	0.065	0.066
20	0.046	0.086	0.514	0.029	0.111	0.120	0.092	20	0.043	0.078	0.070	0.000	0.000	0.075	0.077
40	0.047	0.126	0.826	0.007	0.348	0.379	0.142	40	0.052	0.130	0.117	0.000	0.000	0.156	0.154
50	0.041	0.131	0.856	0.003	0.536	0.571	0.150	50	0.056	0.154	0.177	0.000	0.000	0.183	0.192
			Power (size corrected) x=0.1								Power (size corrected) x=0.1				
2	0.056	0.056	0.056	0.056	0.056	0.056	0.056	2	0.293	0.293	0.293	0.293	0.293	0.293	0.293
3	0.030	0.030	0.039	0.031	0.030	0.030	0.030	3	0.188	0.188	0.226	0.214	0.222	0.222	0.221
4	0.040	0.040	0.062	0.039	0.040	0.040	0.040	4	0.152	0.152	0.177	0.129	0.192	0.192	0.196
5	0.053	0.053	0.082	0.053	0.053	0.053	0.053	5	0.105	0.105	0.141	0.081	0.145	0.145	0.135
10	0.041	0.041	0.143	0.040	0.041	0.041	0.041	10	0.121	0.121	0.131	0.016	0.161	0.161	0.168
20	0.042	0.042	0.189	0.040	0.043	0.043	0.043	20	0.110	0.110	0.198	0.000	0.198	0.198	0.190
40	0.033	0.033	0.078	0.033	0.033	0.033	0.033	40	0.053	0.053	0.046	0.000	0.045	0.045	0.046
50	0.039	0.039	0.086	0.034	0.044	0.044	0.041	50	0.056	0.056	0.033	0.000	0.039	0.039	0.046
			Power (size corrected) x=0.5								Power (size corrected) x=0.5				
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3	1.000	1.000	0.932	1.000	1.000	1.000	1.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4	1.000	1.000	0.730	1.000	1.000	1.000	1.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5	1.000	1.000	0.714	1.000	1.000	1.000	1.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	10	1.000	1.000	0.483	1.000	1.000	1.000	1.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	20	1.000	1.000	0.307	1.000	0.986	0.986	0.989
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	40	0.372	0.372	0.118	0.777	0.580	0.580	0.564
50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	50	0.000	0.000	0.007	0.006	0.003	0.003	0.000

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 3. Factors as instruments, no omitted instruments

	AR	ARS	K	LM	LR	LR1	LR2		AR	ARS	K	LM	LR	LR1	LR2
K2								K2							
			Weak instruments								Strong instruments				
			Size								Size				
2	0.051	0.054	0.054	0.051	0.053	0.053	0.054	2	0.049	0.050	0.050	0.046	0.049	0.049	0.050
3	0.033	0.037	0.037	0.033	0.037	0.037	0.037	3	0.036	0.037	0.037	0.036	0.036	0.036	0.037
4	0.052	0.056	0.056	0.049	0.055	0.055	0.056	4	0.047	0.053	0.053	0.046	0.053	0.053	0.053
5	0.048	0.053	0.053	0.044	0.051	0.051	0.053	5	0.035	0.038	0.038	0.034	0.036	0.036	0.038
10	0.051	0.055	0.055	0.050	0.053	0.053	0.055	10	0.052	0.060	0.060	0.050	0.056	0.056	0.060
20	0.052	0.054	0.054	0.051	0.052	0.052	0.054	20	0.041	0.045	0.045	0.041	0.045	0.045	0.045
40	0.042	0.048	0.048	0.041	0.045	0.045	0.048	40	0.042	0.048	0.048	0.039	0.046	0.046	0.048
50	0.045	0.051	0.051	0.045	0.050	0.050	0.051	50	0.053	0.057	0.057	0.052	0.055	0.055	0.057
			Power (size corrected) x=0.1								Power (size corrected) x=0.1				
2	0.047	0.050	0.050	0.045	0.050	0.050	0.050	2	0.206	0.218	0.218	0.203	0.214	0.214	0.218
3	0.046	0.049	0.049	0.044	0.047	0.047	0.049	3	0.273	0.280	0.280	0.268	0.277	0.277	0.280
4	0.056	0.057	0.057	0.054	0.056	0.056	0.057	4	0.222	0.237	0.237	0.216	0.234	0.234	0.237
5	0.040	0.043	0.043	0.039	0.041	0.041	0.043	5	0.273	0.287	0.287	0.271	0.284	0.284	0.287
10	0.057	0.063	0.063	0.055	0.062	0.062	0.063	10	0.308	0.327	0.327	0.302	0.322	0.322	0.327
20	0.052	0.058	0.058	0.052	0.055	0.055	0.058	20	0.281	0.290	0.290	0.277	0.288	0.288	0.290
40	0.053	0.056	0.056	0.053	0.056	0.056	0.056	40	0.304	0.321	0.321	0.298	0.317	0.317	0.321
50	0.048	0.052	0.052	0.048	0.050	0.050	0.052	50	0.291	0.306	0.306	0.287	0.303	0.303	0.306
			Power (size corrected) x=0.5								Power (size corrected) x=0.5				
2	0.047	0.048	0.048	0.044	0.047	0.047	0.048	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.057	0.061	0.061	0.055	0.059	0.059	0.061	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.062	0.064	0.064	0.061	0.064	0.064	0.064	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.039	0.044	0.044	0.038	0.043	0.043	0.044	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.046	0.054	0.054	0.043	0.053	0.053	0.054	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20	0.042	0.049	0.049	0.041	0.045	0.045	0.049	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.065	0.070	0.070	0.063	0.068	0.068	0.070	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	0.053	0.057	0.057	0.052	0.055	0.055	0.057	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 4. Factors as instruments, omitted instrument, δ=1

K2	AR	ARS	K	LM	LR	LR1	LR2	K2	AR	ARS	K	LM	LR	LR1	LR2
	Weak instruments								Strong instruments						
	Size								Size						
2	0.060	0.068	0.068	0.059	0.065	0.065	0.068	2	0.065	0.069	0.069	0.063	0.069	0.069	0.069
3	0.049	0.053	0.053	0.047	0.050	0.050	0.053	3	0.049	0.052	0.052	0.047	0.052	0.052	0.052
4	0.053	0.059	0.059	0.052	0.058	0.058	0.059	4	0.052	0.059	0.059	0.048	0.055	0.055	0.059
5	0.055	0.059	0.059	0.053	0.056	0.056	0.059	5	0.059	0.063	0.063	0.057	0.061	0.061	0.063
10	0.051	0.054	0.054	0.050	0.053	0.053	0.054	10	0.043	0.045	0.045	0.041	0.045	0.045	0.045
20	0.052	0.060	0.060	0.052	0.058	0.058	0.060	20	0.055	0.057	0.057	0.054	0.056	0.056	0.057
40	0.053	0.058	0.058	0.053	0.058	0.058	0.058	40	0.052	0.055	0.055	0.051	0.055	0.055	0.055
50	0.039	0.042	0.042	0.036	0.042	0.042	0.042	50	0.055	0.056	0.056	0.054	0.055	0.055	0.056
	Power (size corrected) x=0.1								Power (size corrected) x=0.1						
2	0.035	0.040	0.040	0.035	0.038	0.038	0.040	2	0.129	0.140	0.140	0.126	0.133	0.133	0.140
3	0.043	0.051	0.051	0.041	0.049	0.049	0.051	3	0.271	0.283	0.283	0.268	0.281	0.281	0.283
4	0.026	0.030	0.030	0.026	0.029	0.029	0.030	4	0.223	0.238	0.238	0.219	0.231	0.231	0.238
5	0.042	0.044	0.044	0.040	0.044	0.044	0.044	5	0.227	0.237	0.237	0.221	0.236	0.236	0.237
10	0.042	0.045	0.045	0.041	0.045	0.045	0.045	10	0.277	0.292	0.292	0.270	0.286	0.286	0.292
20	0.049	0.051	0.051	0.047	0.050	0.050	0.051	20	0.295	0.301	0.301	0.290	0.298	0.298	0.301
40	0.041	0.049	0.049	0.038	0.045	0.045	0.049	40	0.231	0.245	0.245	0.228	0.237	0.237	0.245
50	0.040	0.044	0.044	0.039	0.041	0.041	0.044	50	0.273	0.289	0.289	0.267	0.283	0.283	0.289
	Power (size corrected) x=0.5								Power (size corrected) x=0.5						
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 5. Factors/Variables as instruments, no omitted instrument, $\Lambda_k = 0$ in (12)

	AR	ARS	K	LM	LR	LR1	LR2		AR	ARS	K	LM	LR	LR1	LR2
K2								K2							
								Factors as instruments							
				Weak instruments							Strong instruments				
				Size							Size				
2	0.047	0.053	0.053	0.046	0.052	0.052	0.053	2	0.068	0.073	0.073	0.065	0.071	0.071	0.073
3	0.046	0.051	0.051	0.044	0.049	0.049	0.051	3	0.049	0.054	0.054	0.047	0.051	0.051	0.054
4	0.054	0.057	0.057	0.053	0.057	0.057	0.057	4	0.053	0.055	0.055	0.052	0.054	0.054	0.055
5	0.050	0.055	0.055	0.048	0.054	0.054	0.055	5	0.041	0.044	0.044	0.041	0.043	0.043	0.044
10	0.054	0.057	0.057	0.052	0.057	0.057	0.057	10	0.051	0.059	0.059	0.051	0.056	0.056	0.059
20	0.048	0.053	0.053	0.046	0.052	0.052	0.053	20	0.052	0.056	0.056	0.052	0.053	0.053	0.056
40	0.068	0.074	0.074	0.068	0.073	0.073	0.074	40	0.044	0.047	0.047	0.043	0.047	0.047	0.047
50	0.059	0.064	0.064	0.058	0.062	0.062	0.064	50	0.058	0.062	0.062	0.056	0.059	0.059	0.062
								Variables as instruments							
				Weak instruments							Strong instruments				
				Size							Size				
2	0.051	0.058	0.058	0.048	0.057	0.057	0.058	2	0.043	0.050	0.050	0.043	0.050	0.050	.050
3	0.037	0.044	0.039	0.035	0.031	0.047	0.047	3	0.058	0.063	0.064	0.021	0.048	0.069	0.072
4	0.059	0.066	0.053	0.056	0.048	0.064	0.067	4	0.052	0.063	0.045	0.008	0.032	0.057	0.061
5	0.059	0.070	0.060	0.048	0.040	0.065	0.069	5	0.047	0.058	0.057	0.002	0.025	0.057	0.058
10	0.058	0.074	0.070	0.046	0.021	0.077	0.082	10	0.057	0.074	0.077	0.001	0.022	0.069	0.075
20	0.051	0.083	0.075	0.027	0.006	0.086	0.093	20	0.041	0.076	0.079	0.000	0.002	0.090	0.095
40	0.059	0.142	0.204	0.012	0.002	0.168	0.167	40	0.047	0.119	0.094	0.000	0.001	0.134	0.134
50	0.032	0.121	0.239	0.001	0.003	0.172	0.156	50	0.047	0.135	0.123	0.000	0.001	0.159	0.168

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 6. Factors/Variables as instruments, omitted instrument, δ=1, results for different numbers of factors

	AR	ARS	K	LM	LR	LR1	LR2		AR	ARS	K	LM	LR	LR1	LR2
K2								K2							
								$r=3,\ \hat{r}=3$							
			Factor Tests: Size								Factor Tests: Power x=0.1				
3	0.047	0.052	0.046	0.016	0.032	0.046	0.046	3	0.111	0.115	0.131	0.039	0.076	0.130	0.133
4	0.047	0.051	0.052	0.011	0.025	0.051	0.053	4	0.203	0.221	0.253	0.106	0.160	0.238	0.246
5	0.050	0.054	0.055	0.016	0.027	0.055	0.056	5	0.174	0.191	0.212	0.081	0.117	0.211	0.213
10	0.053	0.061	0.049	0.018	0.026	0.047	0.050	10	0.190	0.205	0.229	0.087	0.139	0.237	0.238
20	0.044	0.049	0.066	0.020	0.024	0.063	0.063	20	0.222	0.235	0.260	0.126	0.165	0.259	0.265
40	0.040	0.047	0.051	0.013	0.017	0.053	0.053	40	0.231	0.247	0.297	0.140	0.184	0.296	0.303
50	0.049	0.057	0.062	0.016	0.022	0.063	0.063	50	0.264	0.283	0.322	0.162	0.216	0.325	0.331
								$r=3,\ \hat{r}=2$							
			Factor Tests: Size								Factor Tests: Power x=0.1				
3	0.057	0.060	0.060	0.057	0.059	0.059	0.060	3	0.194	0.201	0.201	0.190	0.199	0.199	0.201
4	0.053	0.056	0.056	0.050	0.055	0.055	0.056	4	0.237	0.254	0.254	0.231	0.246	0.246	0.254
5	0.035	0.038	0.038	0.032	0.037	0.037	0.038	5	0.094	0.100	0.100	0.092	0.096	0.096	0.100
10	0.049	0.052	0.052	0.047	0.052	0.052	0.052	10	0.270	0.285	0.285	0.266	0.282	0.282	0.285
20	0.050	0.053	0.053	0.050	0.052	0.052	0.053	20	0.239	0.257	0.257	0.234	0.249	0.249	0.257
40	0.037	0.043	0.043	0.036	0.041	0.041	0.043	40	0.218	0.230	0.230	0.212	0.229	0.229	0.230
50	0.038	0.051	0.051	0.038	0.045	0.045	0.051	50	0.142	0.152	0.152	0.140	0.146	0.146	0.152
								$r=1,\ \hat{r}=2$							
			Factor Tests: Size								Factor Tests: Power x=0.1				
2	0.049	0.051	0.051	0.047	0.051	0.051	0.051	2	0.128	0.139	0.139	0.127	0.137	0.137	0.139
3	0.042	0.045	0.045	0.042	0.044	0.044	0.045	3	0.121	0.128	0.128	0.116	0.127	0.127	0.128
4	0.046	0.051	0.051	0.046	0.050	0.050	0.051	4	0.134	0.142	0.142	0.132	0.139	0.139	0.142
5	0.048	0.053	0.053	0.047	0.052	0.052	0.053	5	0.146	0.156	0.156	0.143	0.152	0.152	0.156
10	0.054	0.058	0.058	0.052	0.057	0.057	0.058	10	0.149	0.160	0.160	0.146	0.156	0.156	0.160
20	0.041	0.047	0.047	0.041	0.043	0.043	0.047	20	0.140	0.153	0.153	0.138	0.148	0.148	0.153
40	0.046	0.054	0.054	0.046	0.053	0.053	0.054	40	0.149	0.157	0.157	0.144	0.153	0.153	0.157
50	0.055	0.059	0.059	0.055	0.058	0.058	0.059	50	0.130	0.142	0.142	0.126	0.140	0.140	0.142
								$r=1,\ \hat{r}=2$							
			Standard Tests: Size								Standard Tests: Power x=0.1				
2	0.049	0.054	0.054	0.047	0.052	0.052	0.054	2	0.120	0.120	0.120	0.120	0.120	0.120	0.120
3	0.052	0.058	0.061	0.045	0.053	0.061	0.062	3	0.136	0.136	0.160	0.129	0.135	0.135	0.135
4	0.042	0.049	0.051	0.026	0.036	0.049	0.051	4	0.102	0.102	0.138	0.087	0.100	0.100	0.100
5	0.053	0.057	0.068	0.026	0.044	0.059	0.061	5	0.091	0.091	0.129	0.093	0.099	0.099	0.093
10	0.042	0.067	0.074	0.008	0.030	0.071	0.074	10	0.076	0.076	0.169	0.053	0.081	0.081	0.081
20	0.053	0.088	0.141	0.004	0.048	0.114	0.098	20	0.060	0.060	0.097	0.033	0.047	0.047	0.050
40	0.041	0.093	0.226	0.000	0.051	0.194	0.103	40	0.059	0.059	0.184	0.019	0.068	0.068	0.059
50	0.050	0.163	0.389	0.000	0.226	0.425	0.179	50	0.036	0.036	0.113	0.009	0.031	0.031	0.030

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments. r denotes the true number of factors and $\hat{r}$ denotes the assumed number of factors.

Table 7. Factors as intermediately weak instruments, no omitted instruments

	AR	ARS	K	LM	LR	LR1	LR2
K2							
Intermediately weak instruments							
Power (size corrected) x=0.1							
2	0.096	0.108	0.108	0.094	0.105	0.105	0.108
3	0.095	0.102	0.102	0.092	0.099	0.099	0.102
4	0.114	0.121	0.121	0.113	0.120	0.120	0.121
5	0.102	0.108	0.108	0.100	0.107	0.107	0.108
10	0.100	0.107	0.107	0.097	0.107	0.107	0.107
20	0.083	0.087	0.087	0.083	0.085	0.085	0.087
40	0.112	0.119	0.119	0.109	0.114	0.114	0.119
50	0.104	0.112	0.112	0.099	0.107	0.107	0.112
Power (size corrected) x=0.5							
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 8. Factors as instruments, omitted instrument correlated with first factor, δ=1

	AR	ARS	K	LM	LR	LR1	LR2		AR	ARS	K	LM	LR	LR1	LR2
K2								K2							
			Weak instruments								Strong instruments				
			Size								Size				
2	0.045	0.051	0.051	0.044	0.049	0.049	0.051	2	0.054	0.055	0.055	0.054	0.055	0.055	0.055
3	0.055	0.058	0.058	0.054	0.057	0.057	0.058	3	0.045	0.053	0.053	0.043	0.049	0.049	0.053
4	0.047	0.050	0.050	0.045	0.050	0.050	0.050	4	0.030	0.033	0.033	0.030	0.032	0.032	0.033
5	0.049	0.053	0.053	0.049	0.050	0.050	0.053	5	0.045	0.049	0.049	0.045	0.048	0.048	0.049
10	0.043	0.045	0.045	0.042	0.044	0.044	0.045	10	0.052	0.059	0.059	0.051	0.054	0.054	0.059
20	0.060	0.063	0.063	0.057	0.062	0.062	0.063	20	0.051	0.056	0.056	0.050	0.053	0.053	0.056
40	0.042	0.047	0.047	0.040	0.047	0.047	0.047	40	0.042	0.047	0.047	0.042	0.047	0.047	0.047
50	0.055	0.060	0.060	0.054	0.058	0.058	0.060	50	0.046	0.049	0.049	0.044	0.049	0.049	0.049
			Power (size corrected) x=0.1								Power (size corrected) x=0.1				
2	0.247	0.259	0.259	0.243	0.256	0.256	0.259	2	0.813	0.821	0.821	0.810	0.817	0.817	0.821
3	0.048	0.050	0.050	0.048	0.050	0.050	0.050	3	0.924	0.931	0.931	0.924	0.928	0.928	0.931
4	0.299	0.320	0.320	0.297	0.316	0.316	0.320	4	0.878	0.884	0.884	0.871	0.882	0.882	0.884
5	0.623	0.641	0.641	0.622	0.634	0.634	0.641	5	0.943	0.949	0.949	0.940	0.946	0.946	0.949
10	0.596	0.614	0.614	0.591	0.607	0.607	0.614	10	0.848	0.863	0.863	0.846	0.858	0.858	0.863
20	0.564	0.576	0.576	0.554	0.573	0.573	0.576	20	0.934	0.935	0.935	0.932	0.934	0.934	0.935
40	0.639	0.648	0.648	0.634	0.644	0.644	0.648	40	0.949	0.951	0.951	0.948	0.949	0.949	0.951
50	0.640	0.656	0.656	0.636	0.650	0.650	0.656	50	0.975	0.977	0.977	0.975	0.976	0.976	0.977
			Power (size corrected) x=0.5								Power (size corrected) x=0.5				
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: The statistics under analysis are the factor based versions of the AR test, of its asymptotic version relying on the $\chi^2(k_2)/k_2$ distribution (ARS), of the LR and LM tests proposed by Wang and Zivot (1998), of the K-test by Kleibergen (2002), and of the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003). The Data Generating Process is described in Section 6. K2 indicates the number of instruments.

Table 9. Inflation Equations					
s_t	Parameter	Baseline Instruments		Factors	
		i.i.d.	HAC	i.i.d.	HAC
Equation: (24)-(27), 1962 Q1 - 1984 Q4					
Output gap	λ_π	[0.000,0.042]	[0.000,0.170]	[0.000,0.055]	[0.000,0.990]
	γ_f	[0.238,0.990]	[0.000,0.990]	[0.376,0.990]	[0.000,0.990]
Marginal cost	λ_π	[0.000,0.109]	[0.000,0.128]	[0.000,0.217]	[0.000,0.214]
	γ_f	[0.118, 0.990]	[0.000,0.990]	[0.267, 0.990]	[0.000,0.990]
Equation: (24)-(27), 1985 Q1 - 2005 Q3					
Output gap	λ_π	[0.000,0.071]	[0.000,0.113]	[0.000,0.046]	[0.000,0.101]
	γ_f	[0.000,0.990]	[0.000,0.990]	[0.533,0.990]	[0.000,0.990]
Marginal cost	λ_π	[0.000,0.088]	[0.000,0.252]	[0.000,0.125]	[0.000,0.990]
	γ_f	[0.000, 0.990]	[0.000,0.990]	[0.404,0.990]	[0.000,0.990]
Equations: (24)-(28)-(29), 1962 Q1 - 1984 Q4					
Output gap	ω_π	[0.010,0.900]	[0.010,0.900]	[0.010,0.900]	[0.010,0.900]
	θ_π	[0.383,0.990]	[0.177,0.990]	[0.528,0.990]	[0.010,0.990]
	γ_f	[0.296,0.980]	[0.202,0.979]	[0.376,0.980]	[0.012,0.980]
Marginal cost	ω_π	[0.010,0.900]	[0.010,0.900]	[0.010,0.760]	[0.010,0.900]
	θ_π	[0.129,0.990]	[0.016,0.990]	[0.269,0.990]	[0.133,0.990]
	γ_f	[0.119,0.980]	[0.011,0.980]	[0.261,0.980]	[0.129,0.980]
(24)-(28)-(29), 1985 Q1 - 2005 Q3					
Output gap	ω_π	[0.010,0.900]	[0.010,0.900]	[0.010,0.881]	[0.010,0.900]
	θ_π	[0.158,0.990]	[0.011,0.990]	[0.731,0.990]	[0.010,0.990]
	γ_f	[0.147,0.980]	[0.013,0.980]	[0.519,0.980]	[0.011,0.980]
Marginal cost	ω_π	[0.010,0.900]	[0.010,0.900]	[0.010,0.862]	[0.010,0.900]
	θ_π	[0.095,0.990]	[0.010,0.990]	[0.310,0.990]	[0.010,0.990]
	γ_f	[0.095,0.980]	[0.011,0.980]	[0.318,0.980]	[0.011,0.980]

Note: The models and estimation procedures are described in Section 7.

Table 10. Output Equations and Taylor Rules					
s_t	Parameter	Baseline Instruments		Factors	
		i.i.d.	HAC	i.i.d.	HAC
Equation: (25), 1962 Q1 - 1984 Q4					
Output gap	β_f	$\emptyset$	[0.000,1.000]	[0.517,0.660]	$\emptyset$
	β_r	$\emptyset$	[1.001,39.99]	[7.353,40.0]	$\emptyset$
Equation: (25), 1985 Q1 - 2005 Q3					
Output gap	β_f	[0.000,1.0]	[0.000,0.997]	[0.509,0.970]	[0.199,0.995]
	β_r	[1.789,40.0]	[1.000,40.0]	[5.495,40.0]	[1.006,39.934]
Equation: (26), 1962 Q1 - 1984 Q4					
Output gap	γ_π	[0.000,2.000]	[0.000,2.000]	$\emptyset$	[0.000,2.000]
	γ_y	[0.004,1.0]	[0.000,1.0]	$\emptyset$	[0.000,1.0]
	$\sum_{j=1}^{3}\rho_{R,j}$	[0.745,0.988]	[0.000,1.0]	$\emptyset$	[0.000,1.0]
Equation: (26), 1985 Q1 - 2005 Q3					
Output gap	γ_π	[0.000,2.000]	[0.000,2.000]	$\emptyset$	[0.000,2.000]
	γ_y	[0.022,0.423]	[0.000,1.0]	$\emptyset$	[0.000,1.0]
	$\sum_{j=1}^{3}\rho_{R,j}$	[0.631,0.966]	[0.000,1.0]	$\emptyset$	[0.000,1.0]
Equations: (30), 1962 Q1 - 1984 Q4					
Output gap	γ_π	[0.772,1.240]	[0.000,2.000]	$\emptyset$	[0.000,2.000]
	γ_y	[0.183,0.390]	[0.000,1.0]	$\emptyset$	[0.000,1.0]
	$\sum_{j=1}^{2}\rho_{R,j}$	[0.848,0.905]	[0.000,1.0]	$\emptyset$	[0.000,0.981]
Equations: (30), 1985 Q1 - 2005 Q3					
Output gap	γ_π	[1.107,2.0]	[0.000,2.000]	$\emptyset$	[0.021,2.000]
	γ_y	[0.065,0.284]	[0.000,1.0]	$\emptyset$	[0.104,0.904]
	$\sum_{j=1}^{2}\rho_{R,j}$	[0.695,0.873]	[0.000,1.0]	$\emptyset$	[0.000,0.923]