DISCUSSION PAPER SERIES

No. 7726

FACTOR-GMM ESTIMATION WITH LARGE SETS OF POSSIBLY WEAK INSTRUMENTS

George Kapetanios and
Massimiliano Marcellino

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP7726.asp

ISSN 0265-8003

FACTOR-GMM ESTIMATION WITH LARGE SETS OF POSSIBLY WEAK INSTRUMENTS

George Kapetanios, Queen Mary University
Massimiliano Marcellino, European University Institute, Bocconi University and CEPR

Discussion Paper No. 7726
March 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 7726

March 2010

ABSTRACT

Factor-GMM Estimation with Large Sets of Possibly Weak Instruments

This paper analyses the use of factor analysis for instrumental variable estimation when the number of instruments tends to infinity. In particular, we focus on situations where many weak instruments exist and/or the factor structure is weak. Theoretical results, simulation experiments and empirical applications highlight the relevance of Factor-GMM estimation, which is also easily implemented.

JEL Classification: C32, C51 and E52
Keywords: factor models, GMM, instrumental variables, principal components, weak instruments and DSGE models

George Kapetanios
Department of Economics
Queen Mary University of London
Mile End Road
London E1 4NS
UK

Email: g.kapetanios@qmul.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157789

Massimiliano Marcellino
Department of Economics
European University Institute
Via della Piazzuola 43
50133 Florence
ITALY

Email:
massimiliano.marcellino@eui.eu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=139608

Submitted 26 February 2010

We are grateful to two anonymous Referees, Jushan Bai, Domenico Giannone, Marco Lippi, Helmut Lutkepohl, Serena Ng, Hashem Pesaran, Jim Stock and participants at conferences at the European University Institute, the Bank of England and the 2008 NBER Summer Institute for helpful comments on a previous draft. The usual disclaimers apply.

Factor-GMM Estimation with Large Sets of Possibly Weak Instruments*

George Kapetanios†
Queen Mary University of London

Massimiliano Marcellino‡
European University Institute, Bocconi University and CEPR

First Version: October 2006
This Version: December 2009

Abstract

This paper analyses the use of factor analysis for instrumental variable estimation when the number of instruments tends to infinity. In particular, we focus on situations where many weak instruments exist and/or the factor structure is weak. Theoretical results, simulation experiments and empirical applications highlight the relevance of Factor-GMM estimation, which is also easily implemented.

J.E.L. Classification: C32, C51, E52

Keywords: Factor models, Principal components, Instrumental variables, GMM, weak instruments, DSGE models

1 Introduction

The paradigm of a factor model is very appealing and has been used extensively in economic analyses. Underlying the factor model is the idea that a large number of economic variables can be adequately explained by a small number of indicator variables or shocks. Most factor analyses were either based on a limited number of variables, N, or used the assumption of i.i.d. variables, which is rather unrealistic for most economic time series. Recently, Stock and Watson (2002b) have put forward the case for using all the information in large datasets, where

*We are grateful to two anonymous Referees, Jushan Bai, Domenico Giannone, Marco Lippi, Helmut Lutkepohl, Serena Ng, Hashem Pesaran, Jim Stock and participants at conferences at the European University Institute, the Bank of England and the 2008 NBER Summer Institute for helpful comments on a previous draft. The usual disclaimers apply.

†Department of Economics, Queen Mary, University of London, Mile End Rd., London E1 4NS. Email: G.Kapetanios@qmul.ac.uk

‡EUI - Department of Economics, Via della Piazzuola 43, 50133, Milano, Italy. E-mail: massimiliano.marcellino@eui.eu

N is allowed to tend to infinity and temporal dependence is taken into consideration. Stock and Watson (2002b) suggest the use of static principal components for estimating factors in this context, Forni, Hallin, Lippi, and Reichlin (2000) and Forni, Hallin, Lippi, and Reichlin (2004) propose dynamic principal components, while Kapetanios and Marcellino (2009) develop a parametric estimator. From an empirical point of view, these new techniques have been mostly applied to provide a more adequate reduced form modelling tool in various contexts, such as forecasting.

Recently, there has a been an interest in more structural applications of factor analysis. In particular, Stock and Watson (2005), Giannone, Reichlin, and Sala (2002) and Bernanke, Boivin, and Eliasz (2005) have shown that it is possible to obtain more realistic impulse response functions in a structural factor model. However, overall, factor analysis has been less widely considered for the estimation of structural relationships. The latter typically requires to address the problem of endogeneity of the regressors, and the use of instruments is the standard solution to provide valid estimation and inference. Typically, few valid instruments are available, but there are also cases where many instruments exist. For example, due to interdependence, the contemporaneous value of a macroeconomic variable can be related to past developments in a large set of other variables, which are also orthogonal to the error term in the structural equation of interest. Some authors have analyzed the properties of instrumental variable (IV) estimators when the number of instruments tends to infinity as the sample size grows; eminent examples are Morimune (1983) and Bekker (1994). Clearly once allowance is made for a large and possibly increasing number of instruments, tools that parsimoniously summarize them, such as factor models, become important.

From an empirical point of view, Favero, Marcellino, and Neglia (2005) show that using factors extracted from a large set of macroeconomic variables as additional instruments in GMM estimation of forward looking Taylor rules for the US and Europe, substantially improves the efficiency of the parameter estimators. Beyer, Farmer, Henry, and Marcellino (2008) extend the analysis to a system context, where a Taylor rule is jointly estimated with a forward looking output equation and a hybrid Phillips curve, along the lines of Galí and Gertler (1999), finding again substantial gains in the GMM estimator's efficiency when adding factors to the instrument set. The present paper provides a theoretical explanation for such empirical findings, and more generally a theory for Factor-GMM estimation in the presence of a large set of instruments.

Another paper that analyzes the interface of factor models and instrumental variable estimation is Bai and Ng (2009), while related earlier references on the use of principal components for IV estimation are Kloek and Mennes (1960) and Amemiya (1966). In a similar vein to our paper, Bai and Ng (2009) consider the case where the endogenous regressors are linear functions of a set of unobserved factors, which are also underlying an expanding set of observed instruments. Clearly, under these circumstances, the use of the true but unknown factors as instruments would provide a superior GMM estimator with respect to the one based on the observed set of instruments. Factor analysis can provide an estimate of the factors, and thereby enable feasible Factor-GMM estimation. Bai and Ng (2009) and our paper independently analyse the properties of Factor-GMM estimation in this context.

With respect to Bai and Ng (2009), our analysis of Factor-GMM focuses on a second important aspect of IV estimation that has been recently explored in the literature. This is the possibility that instruments are weak, in the sense that their relation to the endogenous regressors is local-to-zero. A key reference in the context of a finite set of instruments is Staiger and Stock (1997). In that paper the strength of the correlation of the regressors and the instruments is measured in terms of what is refereed to as a concentration parameter. In standard IV estimation this parameters diverges at a rate equal to the number of observations. Staiger and Stock (1997) consider the case of a constant concentration parameter, which implies that the IV estimator is no longer consistent. The work of Staiger and Stock (1997) has been extended in a variety of ways. Some interesting examples of recent work include Florens, Johannes, and Van Bellegem (2006), Chao, Hausman, Newey, Swanson, and Woutersen (2006), Dufour, Khalaf, and Kichian (2006a) and Dufour, Khalaf, and Kichian (2006b). In our view, the most interesting generalization relates to combining the framework of many instruments with the framework of weak instruments. One of the first papers in the literature to do this was Chao and Swanson (2005). This work was subsequently generalised extensively by Han and Phillips (2006). Other relevant references are Stock and Yogo (2003), Hansen, Hausman, and Newey (2009) and Newey (2004). We consider a number of elements of such an analysis in the context of Factor-GMM estimation, which provides a richer framework for analysing weak instruments than those previously adopted.

Our third contribution in this paper is the evaluation of the presence of a weak factor structure, a topic considered in some detail in Onatski (2006). Our paper, unlike Onatski (2006), assumes that the factor structure although weak is still discernible in terms of the asymptotic properties of the covariance matrix of the data. First, we show that, under certain conditions, it is still possible to obtain a consistent estimator of the (space spanned by the) factors using principal components. Second, we assess the consequences of a weak factor structure for the properties of Factor-GMM estimators, possibly combined with a weak instrument situation.

The fourth contribution of this paper is an extensive Monte Carlo study of the finite sample properties of the Factor-IV estimator for a wide variety of settings, including ones where the instruments are weak or many, and the factor structure is strong or weak. Not only the results are in line with the theoretical findings, they also clearly indicate the superiority of Factor-GMM over standard GMM estimation in finite samples in most situations. Such robust performance supports the general applicability of Factor-GMM for empirical applications.

Related to the previous point, our fifth and final contribution is an empirical analysis where Factor-GMM is applied to estimate the parameters of a forward looking New Keynesian Phillips Curve. The results show first that both a small set of macroeconomic indicators and factors extracted from a much larger set of variables are jointly useful instruments; second, that the new estimation method is easily implemented; third, that it is empirically relevant since it substantially reduces estimation uncertainty with respect to standard GMM; and, fourth, that simple methods to strengthen the factor structure can further improve the efficiency of Factor-GMM..

The paper is structured as follows: Section 2 develops the theoretical properties of factor-based instrumental variable estimators. Section 3 generalizes the results to the GMM context. Section 4 studies the finite sample properties of Factor-GMM estimation using Monte Carlo experiments. Section 5 presents the empirical example. Finally, Section 6 summarizes and concludes. All proofs are contained in the Appendix.

2 Factor-IV estimation

In this Section we study the properties of factor-based Instrumental Variable (IV) estimators with uncorrelated and homoskedastic errors, which is useful to provide insights on the working of factors as instruments. In the first subsection we derive results for the standard case of strong instruments and strong factor structure. In the second subsection we consider weak instruments. In the final subsection we allow for a weak factor structure, possibly combined with weak instruments.

2.1 Strong instruments and strong factors

Let the equation of interest be

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where the k regressors in x_t (or a subset of them) are possibly correlated with the error term ϵ_t. A standard source of correlation in the IV literature is measurement error, which could be widespread in macroeconomic applications, where the variables are typically expressed as deviations from an unobservable equilibrium value. Another source of endogeneity is, of course, simultaneity, which is again widespread in applied macroeconomic applications based on single equation estimation. A more specific source of endogeneity in forward looking models, such as the new generation of DSGE models, is the use of expectations of future variables as regressors, which are then typically replaced by their true values for estimation, see for example the literature on Taylor rules or hybrid Phillips curves (e.g., Clarida, Galí, and Gertler (1998) or Galí and Gertler (1999), and our empirical application in Section 5).

Let us assume that there exist N instrumental variables, s_t, generated by a factor model with $r \geq k$ unobservable factors:

$$s_t = \Lambda^{0\prime} f_t + v_t, \tag{2}$$

where r is much smaller than N. Therefore, each instrumental variable can be decomposed into a common component (an element of $\Lambda^{0\prime} f_t$) that is driven by a few common forces, the factors, and an idiosyncratic component (an element of v_t). When the latter is small compared to the former, the information in the large set of N instrumental variables s_t can be efficiently summarized by the r factors f_t.

As the data generation mechanisms for x_t we consider

$$x_t = A_Z^{0\prime} z_t + A^{0\prime} f_t + u_t, \tag{3}$$

where $E(u_t' \epsilon_t) \neq 0$ to introduce simultaneity in (1), and z_t is a finite (and small) subset of s_t, so that

$$z_t = \Lambda_z^{0\prime} f_t + v_{zt}.$$

Equation (3) appears to be the most interesting case from an economic point of view. For example, future inflation can be expected to depend on a set of key macroeconomic indicators, such as monetary policy, oil prices and unit labor costs, on the past values of inflation itself due to persistence, but also on the behaviour of a large set of other variables, such as developments at the sectoral or regional level, that can be well summarized by a few factors (see Beck, Hubrich, and Marcellino (2009)). A similar reasoning holds for unobservable variables, such as the output gap. Moreover, (3) seems the most common case also in empirical analyses, e.g. it is what turns out to be relevant in the empirical application of Section 5 where a subset of s_t and the factors are jointly significant to explain x_t. More generally, in the forecasting literature adding factors to an autoregressive or vector autoregressive model for the variable of interest improves the performance.

It is also worth mentioning that (3) nests the case considered by Bai and Ng (2009), where $A_Z^0 = 0$, which represents the most favourable situation for Factor-IV estimation, since the original instruments s_t become irrelevant, conditional on the factors. Interestingly, Bai and Ng (2009) show that in this case the factors can be valid instruments also when the s variables are correlated with the error term in the structural equation. Equation (3) nests also the standard case where $A^0 = 0$, which is considered in detail in Kapetanios and Marcellino (2006). When $A^0 = 0$, z_t are the optimal instruments and, conditional on z_t, the factors are irrelevant. However, when $z_t = s_t$ and N is very large, possibly larger than T, the full asymptotic properties of the standard IV estimator are in general unknown, but it can be expected to perform poorly and in some cases is even inconsistent, as discussed in Bekker (1994) and Chao and Swanson (2005). In this context, the factors become useful again, since they provide a concise summary of the information in s_t.

Stacking observations across time for the model presented above gives:

$$y = X\beta + \epsilon \tag{4}$$

$$S = F\Lambda^0 + v \tag{5}$$

$$Z = F\Lambda_Z^0 + v_z \tag{6}$$

$$X = ZA_Z^0 + FA^0 + u \tag{7}$$

where $y = (y_1, ..., y_T)'$, $X = (x_1, ..., x_T)'$, $S = (s_1, ..., s_T)'$, $Z = (z_1, ..., z_T)'$, $F = (f_1, ..., f_T)'$, $u = (u_1, ..., u_T)'$, $v = (v_1, ..., v_T)'$ and $\epsilon = (\epsilon_1, ..., \epsilon_T)'$.

Let $\hat{F}$ denote the Stock and Watson (2002b) principal component estimator of F. We define the Factor-IV estimator as:

$$\hat{\beta} = \left(X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'X\right)^{-1} X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'y, \tag{8}$$

while the infeasible factor estimator is given by

$$\bar{\beta} = \left(X'F(F'F)^{-1}F'X\right)^{-1} X'F(F'F)^{-1}F'y. \tag{9}$$

To study the properties of the estimators, we make the following assumptions:

Assumption 1 *1. $E||f_t||^4 \leq M < \infty$, $T^{-1}\sum_{t=1}^T f_t f_t' \xrightarrow{p} \Sigma_f$ for some $r \times r$ positive definite matrix Σ_f. Λ^0 has bounded elements. Further $||\Lambda^{0'}\Lambda^0/N - D|| \to 0$, as $N \to \infty$, where D is a positive definite matrix.*

2. $E(v_{i,t}) = 0$, $E|v_{i,t}|^8 \leq M$ where $v_t = (v_{1,t}, ..., v_{N,t})'$ The variance of v_t is denoted by Σ_v. f_s and v_t are independent for all s, t.

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

- $(NT)^{-1}\sum_{s=1}^T\sum_{t=1}^T |\sum_{i=1}^N \tau_{i,i,t,s}| \leq M$
- $|1/N\sum_{i=1}^N \tau_{i,i,s,s}| \leq M$ *for all* s
- $N^{-1}\sum_{i=1}^N\sum_{j=1}^N |\tau_{i,j,s,s}| \leq M$
- $(NT)^{-1}\sum_{s=1}^T\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N |\tau_{i,j,t,s}| \leq M$
- *For every* (t, s), $E|(N)^{-1/2}\sum_{i=1}^N (v_{i,s}v_{i,t} - \tau_{i,i,s,t})|^4 \leq M$

Assumption 2 *ϵ_t is a martingale difference sequence with finite fourth moment and $E(\epsilon_t^2|\mathcal{F}_t) = \sigma^2 < \infty$ where $\mathcal{F}_t$ is the σ-field generated by (f_s, z_s), $s \leq t$.*

Assumption 3 *$E(s_{it}\epsilon_t) = 0$, $i = 1, ..., N$. $E(s_t s_t')$ is nonsingular for all N and t. $E(s_t x_t')$ has full column rank k. x_t and s_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 guarantees that standard IV estimation using (possibly a subset of) s_t as instruments is feasible, and Assumption 2 that it is efficient. Assumption 2 will be relaxed in the next Section, while Assumptions 1 and 3 are assumed to hold throughout the paper, unless otherwise stated or modified.

The next theorem provides the asymptotic distribution of the Factor-IV estimator under (4)-(7).

Theorem 1 *Assuming that f_t is known, $\sqrt{T}(\bar{\beta}-\beta)$ has an asymptotically Normal distribution, with zero mean and variance covariance matrix given, up to a scalar constant of proportionality, by*

$$Avar\left(\sqrt{T}(\bar{\beta}-\beta)\right) = \left(\left(A_Z^{0\prime}\Lambda_Z^{0\prime} + A^{0\prime}\right)\Sigma_f\left(\Lambda_Z^0 A_Z^0 + A^0\right)\right)^{-1} \tag{10}$$

If $\sqrt{T}/N = o(1)$ then

$$\sqrt{T}(\bar{\beta}-\beta) - \sqrt{T}(\hat{\beta}-\beta) = o_p(1) \tag{11}$$

Three additional comments are required. First, similar results hold when z_t are added to the factors as instruments; actually the efficiency of the estimator improves, but we do not report the details of this case since the notation is even more complicated. Second, comparable results on the performance of the standard IV estimator when the number of instruments tends to infinity as the sample size grows ($A^{0\prime} = 0$ and $s_t = z_t$ in (3)) are provided by Morimune (1983) and Bekker (1994). With respect to that literature, the analysis of the Factor IV estimator is much simpler, because the number of factors remains fixed even if the number of instruments diverges. Moreover, there are several cases where consistency is lost for the standard-IV estimator when N diverges, as we will also see in the simulation experiments of Section 4, while the Factor IV estimator remains consistent under the mild Assumptions 1-3. Finally, notice that the condition $\sqrt{T}/N = o(1)$ is needed for the variance of the factor estimator to become negligible when computing the asymptotic variance of the Factor-IV estimator, but the latter remains consistent even if that condition is not satisfied.

2.2 Weak instruments and strong factors

To analyze the weak instrument case, we substitute equation (3) with

$$x_t = A_{(T)Z}^{0\prime} z_t + A_{(T)}^{0\prime} f_t + u_t. \tag{12}$$

Formally, the instruments are referred to as weak when $A^{0\prime}_{(T)Z} Z' Z A^0_{(T)Z}$ or $A^{0\prime}_{(T)} F' F A^0_{(T)}$ is less than $O_p(T)$ (see, e.g., Chao and Swanson (2005)). This implies that the explanatory power for the endogenous variables x of either the factors or the Z variables or both vanishes asymptotically.

A key reference for the analysis of the standard IV estimator in the context of a finite set of weak instruments is Staiger and Stock (1997), while Chao and Swanson (2005) and Han and Phillips (2006) allow the number of weak instruments to diverge. The following theorem provides results for our Factor-IV estimator in the case of many weak instruments.

Theorem 2 *Assume that: a)* $N = O(T^{\gamma})$, $\gamma > 1/2$; *b) either every element of* $\Lambda^0_Z A^0_{(T)Z}$ *is* $O(T^{-\psi})$, *where* $0 \le \psi < 1/2$ *or every element of* $A^0_{(T)}$ *is* $O(T^{-\vartheta})$, $0 \le \vartheta < 1/2$ *(or both). Then, under (12), if* $\vartheta < \psi$

$$T^{1/2-\vartheta}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_{\epsilon}\left(\Psi' \Sigma_f \Psi\right)^{-1}\right) \tag{13}$$

if $\vartheta > \psi$,

$$T^{1/2-\psi}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_{\epsilon}\left(\Upsilon' \Sigma_f \Upsilon\right)^{-1}\right) \tag{14}$$

and if $\vartheta = \psi$,

$$T^{1/2-\psi}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_{\epsilon}\left(\left(\Upsilon + \Psi\right)' \Sigma_f \left(\Upsilon + \Psi\right)\right)^{-1}\right) \tag{15}$$

where

$$\lim_{T\to\infty} \frac{\Lambda^0_Z A^0_{(T)Z}}{T^{-\psi}} = \Upsilon, \quad \lim_{T\to\infty} \frac{A^0_{(T)}}{T^{-\vartheta}} = \Psi \tag{16}$$

It is worth making a number of comments on the assumptions of Theorem 2. First, the condition $N = O(T^{\gamma})$, $\gamma > 1/2$, guarantees that the convergence rate of the estimated factors is fast enough to avoid generated regressor problems for the computation of the variance of the Factor IV estimator (the larger γ, the slower T increases with respect to N). Second, in the requirement that every element of $\Lambda^0_Z A^0_{(T)Z}$ is $O(T^{-\psi})$, $0 \le \psi < 1/2$, the parameter ψ controls how fast the instruments become weak when T increases (the larger ψ, the faster the instruments become weak, the slower the speed of convergence of the Factor-IV estimator). The condition that every element of $A^0_{(T)}$ be $O(T^{-\vartheta})$, $0 \le \vartheta < 1/2$ has a similar interpretation. If both ψ and ϑ are too large (i.e. assumption b is violated), the Factor-IV estimator is no longer consistent. Third, and related to the previous comment, the relative size of ϑ and ψ determines whether the correlation of x with the z variables or the factors decreases faster. For example, when $\vartheta > \psi$, the factors "become" weak instruments faster than the z variables. In this context we note the relevant work of Hahn and Kuersteiner (2002). Finally, the assumption that the elements of $A^0_{(T)Z}$ and $A^0_{(T)}$ are deterministic can be relaxed to allow for the possibility of random elements that are independent of F, ϵ, u and v. Then, the conditions in (16) would be modified to ones involving stochastic convergence.

Three more general comments on the results in this subsection are the following. First, in our context the concentration parameter is growing but slower than the sample size. This differs from Staiger and Stock (1997) where the concentration parameter is constant. As a consequence, we obtain asymptotically normal IV estimators, even though the speed of convergence is slower than $T^{1/2}$. Second, and related to the previous point, the results in Theorem 2 can be also used to justify the use of standard t or F-tests to verify hypotheses of interest on the parameters. In analogy to the standard case of IV estimation and testing, where there is no need to explicitly normalise estimates by the $T^{1/2}$ parametric rate of convergence to obtain t and F-tests, there is no need to know ψ or ϑ in order to construct tests in the present context. Finally, if every element of $A^0_{(T)}$ is $O(T^{-1/2})$, i.e., the factors are very weak instruments, we are in the context analysed in Staiger and Stock (1997). In this case, if condition *a)* and Assumption 1 (1) hold, it can be shown that the factors can be estimated consistently, the 2SLS estimator based on the true factors has the same probability limit as the 2SLS estimator based on the estimated factor (the Factor-IV estimator), the Factor-IV estimator is no longer consistent, and the results of Theorem 1 of Staiger and Stock (1997) for the 2SLS estimator are valid for the Factor-IV estimator.

2.3 Weak factors

The key idea underlying a factor model such as (2) is that all variables are driven by the same limited number of factors. In this context, the larger the number of variables, the better the estimators of the factors, which has led to the use of larger and larger datasets in empirical analyses, see, e.g., Stock and Watson (2002a). However, it can be expected that while a large but limited set of key macroeconomic variables have a strong common component, the common factors are less and less relevant with respect to the idiosyncratic component for the additional variables that are added to the basic dataset to achieve a very large value for N.

In particular, we might consider a situation such as

$$z_{it} = \frac{\lambda_i^{0\prime}}{i^{\alpha_i}} f_t + e_{it}, \qquad (17)$$
$$\alpha_i = \begin{cases} 0 \text{ for } i < N^* \\ \alpha \text{ for } i \geq N^* \end{cases},$$

for $i = 1, ..., N$ and $\alpha > 0$, so that, from N^* onwards, the larger is i the smaller the fraction of variance of z_i explained by the common factors.

If we write the model in a compact notation as

$$z_t = \Lambda_N^{0\prime} f_t + e_t, \qquad (18)$$
$$\Lambda_N^{0\prime} = \left[\frac{\lambda_1^0}{1}, \frac{\lambda_2^0}{2^{\alpha_2}}, ..., \frac{\lambda_N^0}{N^{\alpha_N}} \right]',$$

the loading matrix Λ_N^0 does not necessarily satisfy the condition in Assumption 1 (1). Therefore, this model cannot be amenable to standard factor analysis. To see this, notice that for Assumption 1 (1) to be satisfied the all rows sum of Λ_N^0 must be $O(N)$. But if $\alpha > 0$, it follows that $\sum_{i=1}^{N} \frac{\lambda_i^{0\prime}}{i^{\alpha_i}} \leq c_1 N^* + c_2 N^{1-\alpha} = o(N)$ as long as $N^* = o(N)$, for some finite vectors c_1 and c_2.

Another factor specification that can violate Assumption 1 (1) is

$$z_t = \Lambda_N^{0\prime} f_t + e_t, \tag{19}$$

$$\Lambda_N^{0\prime} = \left[\frac{\lambda_1^0}{N^\alpha}, \frac{\lambda_2^0}{N^\alpha}, ..., \frac{\lambda_N^0}{N^\alpha} \right]' = \frac{\Lambda^{0\prime}}{N^\alpha}, \tag{20}$$

so that the factor loadings are decreasing in N for each variable. Such a model may be considered unrealistic from an economic point of view, but it is analytically tractable and if we can find consistent factor estimators in (19), the same estimators will remain consistent in (17). Actually, for all $i = [b_i N]$, where [.] denotes integer part and $0 < b_i \leq 1$, (19) is equivalent to (17) since $\frac{\lambda_i^{0\prime}}{i^\alpha} = \frac{\lambda_i^{0\prime}}{b_i^\alpha N^\alpha} = \frac{\tilde{\lambda}_i^{0\prime}}{N^\alpha}$ where $\tilde{\lambda}_i^{0\prime} = \frac{\lambda_i^{0\prime}}{b_i^\alpha}$.

Further justification can be provided for the setup in (19). First, this setup, which essentially defines a sequence of models, is similar to that of Chao and Swanson (2005), which we have used in the previous subsections to consider the case of many weak instruments. From the literature on many weak instruments, it is clear that defining a sequence of models whereby parameters depend on the sample size (either N or T) is the most common way of exploring instrument weakness in IV settings. We simply transfer this device to a factor setting. Second, the dependence of the loadings matrix Λ_N^0 on N is implicit in all factor models, since the dimension of the loadings matrix changes with the number of variables N. Here we just make the loadings of each equation also dependent on N. Third, the main point of departure of (19) from (17) is that loadings for all variables get weaker as N increases rather that only for additional variables. While this can be considered more restrictive, it addresses the problem of the dependence of (17) on the ordering of the variables. Moreover, (19) can be considered as a device for allowing the proportion of the variance explained by the factors to go to zero as N tends to infinity. It is likely that this effect is pervasive within the dataset rather than specific to a particular subset of the variables, thereby motivating (19) instead of (17).

The first result we present analyses conditions under which a local-to-zero factor loading matrix in (19) leads to a model that loses its defining factor characteristic, which is commonly taken to imply that the largest r eigenvalues of the variance covariance of z_t tend to infinity.

Theorem 3 *Let $\Lambda_N^0 = \Lambda^0 / N^\alpha$ as in (20). The eigenvalues of the population variance covariance matrix of Z are bounded for $\alpha \geq 1/2$ for all N.*

In the case considered in the above Theorem, the factor model is no longer identifiable and common and idiosyncratic components cannot be distinguished. This possibility is studied in some detail in Onatski (2006). The next result concerns estimation of the factors for a local-to-zero factor loading matrix.

Theorem 4 *Let $\Lambda_N^0 = \Lambda^0/N^\alpha$ as in (20), where $0 \leq \alpha < 1/4$. Then, as long as $N = o(T^{1/4\alpha})$,*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_t - Hf_t\right\|^2 = O_p\left(\min\left(N^{-4\alpha}T, N^{1-4\alpha}\right)^{-1}\right) \tag{21}$$

Therefore, a sufficient condition for estimation in the local to zero case is the presence of a relatively strong local-to-zero factor model $(\alpha < 1/4)$. Note also the tradeoff between α and the allowable rate of increase for N (the bigger α, the slower the allowed rate of increase for N). Finally, the condition on α is not necessary: in specific cases, such as those considered in the Monte Carlo, it is possible to obtain consistent estimators for the factors even when $\alpha < 1/2$.

We now want to evaluate the properties of the Factor IV estimator in the presence of a weak factor structure, possibly combined with weak instruments. The starting point is the following lemma.

Lemma 1 *Let $\Lambda_N^0 = \Lambda^0/N^\alpha$ as in (20). Let $\alpha < 1/4$ and $N^{2\alpha} = o(T^{1/2})$. Let q_t be any multivariate sequence of random variables. Then,*

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = O_p\left(C_{NT}^{-1}\right) \tag{22}$$

where C_{NT} is defined in (54), as long as q_t has finite fourth moments, nonsingular covariance matrix and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q_t$ satisfies a central limit theorem.

Then, we have the following theorem, which allows for both a weak factor structure and, possibly, weak instruments.

Theorem 5 *Let $\Lambda_N^0 = \Lambda^0/N^\alpha$ as in (20). Let $\alpha < 1/4$, $N = o(T^{1/4\alpha})$ and $C_{NT}^{-1}T^{1/2} = o(1)$ where C_{NT} is defined in (54). Then, Theorem 1 follows.*
Further, if either every element of $\Lambda_{Z,N}^0 A_{(T)Z}^0$ is $O(T^{-\psi})$, $0 \leq \psi < 1/2$ or every element of $A_{(T)}^0$ is $O(T^{-\vartheta})$, $0 \leq \vartheta < 1/2$, then (13)-(15) of Theorem 2 follow, where

$$\lim_{T\to\infty}\frac{\Lambda_N^0 A_{(T)Z}^0}{T^{-\psi}} = \Upsilon, \quad \lim_{T\to\infty}\frac{A_{(T)}^0}{T^{-\vartheta}} = \Psi \tag{23}$$

and Υ and Ψ are nonsingular matrices.

Note that as $\alpha \to 0$, namely, the factor structure becomes strong, the condition $C_{NT}^{-1}T^{1/2} = o(1)$ becomes equivalent to $\sqrt{T}/N = o(1)$, which was the condition in Theorem 1 for asymptotic equivalence of the infeasible and feasible Factor-IV estimators. The other requirements are similar to those in Theorem 2, and limit the decrease in the explanatory power of the observable instruments and factors.

A further issue that arises in a weak factor model is the determination of the number of factors. The use of information criteria has been suggested by Bai and Ng (2002) in the strong factor case. Specifically, they suggest criteria of the form

$$V_r + rg(N,T) \tag{24}$$

where

$$V_r = (NT)^{-1} \sum_{i=1}^{N} \sum_{i=1}^{T} \left(z_{i,t} - \sum_{j=1}^{r} \hat{\lambda}_{j,i} \hat{f}_{j,t} \right)^2 \tag{25}$$

$\hat{\lambda}_{j,i}$ denotes the j,i-th element of the estimate of Λ_N^0 and $g(N,T)$ denotes a penalty term that depends on N and T. The following Theorem provides a condition on the penalty term $g(N,T)$ that ensures consistency of the estimated number of factors even under a weak factor structure.

Theorem 6 *Let the number of factors be determined by minimising (24) over $1 \leq r \leq r_{\max}$ for some constant $r_{\max}$. Let*

$$g(N,T) = \ln\left(\min(N,T)\right)^{-1} \tag{26}$$

Then, the estimated number of factors is consistent for the true number of factors for all $0 \leq \alpha < 1/4$.

The penalty term in Theorem 6 is smaller than that suggested by Bai and Ng (2002) so that, when $\alpha = 0$ and N and T are finite, our criterion is expected to select a larger number of factors. However, in the weak factor case, it provides the correct choice with probability one in large samples, while the criteria by Bai and Ng (2002) would underestimate the number of factors.

A final issue that deserves discussion is what happens to the Factor-IV estimator when Theorem 3 holds, i.e., the factor structure is so loose that the factors cannot be identified and consistently estimated. In this case the Factor-IV estimator is no longer consistent. Kapetanios and Marcellino (2008) discuss this case in more detail and introduce an alternative IV estimator based on averages of the instruments that, under certain conditions, is consistent in the complete absence of a factor structure. Note that this estimator is of use, in the absence of a factor structure, only if the large set of instruments enters directly in (3), since otherwise the large set of instruments becomes irrelevant for x_t. Of course, the question remains as to how to decide if we are in a situation with very weak factors where Factor-IV is not consistent or in a situation where Factor-IV is consistent. We suggest that information criteria can be used for that. In particular, the case of no factors is a valid case to consider when searching for the number of factors using the information criteria discussed in Theorem 6. If zero factors are selected then Factor-IV should not be used. We examine some of these issues through our Monte Carlo study.

In summary, in this Section we have introduced the Factor-IV estimator, which can be used in the presence of a large set of instruments whose generating mechanism can be well represented by a factor model, and derived its asymptotic properties. We have shown that when N tends to infinity it becomes feasible to estimate the unobserved factors f_t consistently, even for local-to-zero factor models (Theorem 4). Under mild conditions, estimation of the factors does not matter for the asymptotic properties of the Factor-IV estimator (Theorems 1 and 5). Moreover, whereas the Factor-IV estimator can remain consistent and asymptotically normal even in the case where z_t are weak instruments (Theorems 2 and 5), standard IV estimation can be inconsistent if the number of instrument increases fast enough, as discussed by Bekker (1994) and Chao and Swanson (2005). Finally, the Factor-IV estimator is inconsistent either when the factors are very weak instruments, along the lines of Staiger and Stock (1997), or when the factor structure is very loose, and in this case the estimator of Kapetanios and Marcellino (2008) can be preferable.

3 Factor-GMM estimation

We now relax assumption 2 and allow for correlation and heteroskedasticity in the errors ϵ of equation (1). We formalise this with the following assumption, which substitutes assumption 2:

Assumption 4 *ϵ_t is a zero mean process with finite variance. The process $s_t\epsilon_t$ and, by implication, $f_t\epsilon_t$, satisfies the conditions for the application of some central limit theorem for weakly dependent processes, with a zero mean asymptotic normal limit.*

$$S_{f\epsilon} = \lim_{T\to\infty}\left[E\left(\left[T^{-1}\sum_{t=1}^{T}\epsilon_t^2 f_t f_t'\right]\left[T^{-1}\sum_{t=1}^{T}\epsilon_t^2 f_t f_t'\right]'\right)\right]$$

exists and is nonsingular.

We further add the following regularity condition.

Assumption 5 *$E[(s_{ti}x_{tj})^2]$ exists and has finite elements for $t = 1, ..., T$, $i = 1, ..., N$ and $j = 1, ..., k$.*

Remark 1 *Assumption 4 is a high level assumption. It is given in this form for generality. More primitive conditions on ϵ_t such as, e.g., mixing with polynomially declining mixing coefficients or near epoque dependence (see, e.g, Davidson (1994)) are sufficient for Assumption 4 to hold.*

As long as the instruments remain uncorrelated with the errors at all leads and lags, the Factor-IV estimator remains consistent and asymptotically normal, but it is no longer efficient. In fact, the efficient estimator in this context is obtained by GMM estimation with $S_{f\epsilon}^{-1}$ as the weighting matrix. Using standard methods, the resulting estimator is $\bar{b} = \left(X'FS_{f\epsilon}^{-1}F'X\right)^{-1} X'FS_{f\epsilon}^{-1}F'y$, and when the errors are uncorrelated and homoskedastic, this expressions simplifies to (9). In practice, $S_{f\epsilon}$ can be estimated by a HAC procedure, such as that developed in Newey and West (1987). For example, using a Bartlett kernel, we have $\widehat{S}_{f\epsilon,h} = \widehat{\Phi}_0 + \sum_{j=1}^{h}(1 - \frac{j}{h+1})(\widehat{\Phi}_j + \widehat{\Phi}_j'),\ \widehat{\Phi}_j = T^{-1} \sum_{T=j+1}^{T} \widehat{\epsilon}_t\widehat{\epsilon}_{t-j}\hat{f}_t\hat{f}_{t-j}'$, where h is the length of the window (bandwidth), $\widehat{\epsilon}_t = y_t - x_t'b$, and b is a consistent estimator for β. The resulting Factor-GMM estimator is denoted by $\widehat{b}$. Some properties of the Factor-GMM estimator are given in the following Theorems.

Theorem 7 *Under Assumptions 3,4 and 5 and assuming that f_t is observed, then under (3)*

$$var\left(\sqrt{T}(\bar{b} - \beta)\right) = \left(\left(A_Z^{0\prime}\Lambda_Z^{0\prime} + A^{0\prime}\right)\Sigma_f S_{f\epsilon}^{-1}\Sigma_f\left(\Lambda_Z^0 A_Z^0 + A^0\right)\right)^{-1} \tag{27}$$

If $\sqrt{T}/N = o(1)$ then

$$\sqrt{T}(\bar{b} - \beta) - \sqrt{T}(\widehat{b} - \beta) = o_p(1) \tag{28}$$

Theorem 8 *Under Assumptions 1,3,4 and 5, the assumptions of Theorem 2 and assuming that the bandwidth, h, of the HAC variance estimator is $o(T^{1/2})$, then, if $\vartheta < \psi$*

$$T^{1/2-\vartheta}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left(\Psi'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f\Psi\right)^{-1}\right) \tag{29}$$

if $\vartheta > \psi$,

$$T^{1/2-\psi}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left(\Upsilon'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f\Upsilon'\right)^{-1}\right) \tag{30}$$

and if $\vartheta = \psi$,

$$T^{1/2-\psi}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left((\Psi + \Upsilon)'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f(\Psi + \Upsilon)\right)^{-1}\right) \tag{31}$$

where $\lim_{T\to\infty} \frac{\Lambda_Z^0 A_{(N,T)Z}^0}{T^{-\psi}} = \Upsilon \quad \lim_{T\to\infty} \frac{A_{(T)}^0}{T^{-\vartheta}} = \Psi.$

Finally, the results of Theorem 5 on the use of factors in the presence of weak instruments and a weak factor structure follow straightforwardly also for the GMM case.

4 Monte Carlo Study

This section presents a detailed Monte Carlo study of the relative performance in finite samples of the standard and Factor-IV estimators, with a focus on the weak instruments and/or weak factor cases. In the first subsection, we consider the setup and estimators we have proposed in Section 2. In the second subsection, we present additional results on the role of variable preselection and on the Factor-GMM estimation of Section 3.

4.1 Factor-IV estimation in finite samples

The basic setup of the Monte Carlo experiments is:

$$y_t = \sum_{i=1}^{k} x_{it} + \epsilon_t \tag{32}$$

$$s_{it} = \sum_{j=1}^{r} N^{-p} f_{jt} + e_{it}, \quad i = 1, ..., N \tag{33}$$

$$x_{it} = \sum_{j=1}^{r} T^{-\theta} r^{-1/2} j f_{jt} + u_{it}, \quad i = 1, ..., k \tag{34}$$

where $e_{it} \sim i.i.d.N(0,1)$, $f_{it} \sim i.i.d.N(0,1)$ and $cov(e_{it}, e_{sj}) = 0$ for $i \neq s$. Let $\kappa_t = (\epsilon_t, u_{1t}, ..., u_{kt})'$. Then, $\kappa_t = P\eta_t$, where $\eta_t = (\eta_{1,t}, ..., \eta_{k+1,t})'$, $\eta_{i,t} \sim i.i.d.N(0,1)$ and $P = [p_{ij}]$, $p_{ij} \sim i.i.d.N(0,1)$. The errors e_{it} and u_{is} are independent for each i and s.

We do not consider heteroskedastic and/or correlated errors, in the main part of the Monte Carlo study, because we want to compare the standard and Factor-IV estimators without the possible complications arising from estimation of the HAC variance covariance matrix of the errors. We will consider the case of serial correlation in the second subsection. Also, we only report results for $k = 1$, $r = 1$, since there are no qualitative changes by increasing the number of endogenous variables or of factors (results available upon request).

The instrumental variables s_{it} are generated by the factor model in (33). The parameter p controls the "strength" of the factor structure. We consider the values $p = 0, 0.25, 0.33, 0.5$. When $p = 0$ we are in the standard case analyzed by Stock and Watson (2002b), Stock and Watson (2002a), Bai (2003), and Bai and Ng (2002). When $p > 0$ we are in the weak factor structure case, analyzed in Theorems 4 and 5 of Section 2.3. In particular, we know from Theorem 4 that it is still possible to estimate consistently the (space spanned by the) factors using the principal component based estimator when $p < 0.25$. In this particularly simple set up, it can be easily shown that the principal component based estimator remains consistent when $p < 0.5$. Actually, from (33), it is $\sum_{i=1}^{N} N^{p-1} z_{it} = f_t + \sum_{i=1}^{N} N^{p-1} e_{it}$ and $\sum_{i=1}^{N} N^{p-1} e_{it}$ goes to zero when $p < 0.5$.

The parameter θ measures the "strength" of the factors as instruments, which decreases for higher values of θ. In fact, from (34), it is

$$var(x_t) = T^{-2\theta} + 1. \tag{35}$$

We consider values $\theta = 0, 0.25, 0.5$. When $\theta = 0$ we are in the standard strong instrument case; when $\theta = 0.25$ we are in the weak instrument case covered by Theorem 2 of Section 2.2; when $\theta = 0.5$ we are in the very weak instrument context analysed in Staiger and Stock (1997) and discussed for the factor set-up at the end of Section 2.2.

Note that (34) differs from (3) for the omission of the z regressors. However, as long as the dimension of z is finite and z are strong instruments, as in the empirical application we will present below, there are no major changes in the Monte Carlo results. The cases where z has dimension N, and possibly x depends on z only and not on the factors, are analyzed in Kapetanios and Marcellino (2008).

We consider the following combinations of N and T: $N = (30, 50, 100, 200)$, $T = (30, 50, 100, 200)$. In Table 1 we report the mean squared errors (MSE) of the alternative estimators (the biases are in general small for both estimators, results available upon request). In Table 1 we, further report the coverage probability of the Factor-IV estimator (computed as the percentage of times the null hypothesis that the coefficient is one is not rejected by a t-test).

Let us start with the case $p = 0$ and $\theta = 0$. In this context, the estimator of the factor is simply the average of the s variables, while the pseudo true value of the coefficient of s_{it} in a regression of x_t on s_{it}, $i = 1, ..., N$, is $1/N$. Hence, when T is large, and larger than N, the two IV estimators should produce similar results. This is indeed the case in Table 1: when $T = 200$ and $N = 30$ the MSE for standard IV is 0.109 versus 0.075 for Factor-IV. But when T is small relative to N, it is more efficient to use the estimated factor rather than estimating the coefficients of z_{it}, i.e. using the standard IV estimator. For example, when $T = 50$ and $N = 30$ the MSE for standard IV is 0.280 versus 0.159 for Factor-IV. Note also that for both estimators the MSE decreases with T when N is fixed. However, when both T and N increase (i.e. we move along the diagonal of the relevant subpanels of Table 1) the MSE of the standard IV does not decrease, signalling inconsistency, while that of the Factor-IV does decrease. Also, and as expected, for given T the MSE of the standard IV increases with the number of instruments N, while it is stable for the Factor-IV estimator. The final comment on this base case concerns the coverage rates in Table 1, which are remarkably close to the 95% value.

The findings so far confirm the theoretical results in our Section 2.1 and in Bai and Ng (2009), namely, that when the instruments are many and strong, and the factor structure is strong, Factor-IV is systematically better than standard IV. Next, we analyze the, perhaps more interesting, cases where either the factors are weak instruments ($\theta > 0$), or the factor structure is weak ($p > 0$), or both.

Figures for the case $p = 0, \theta > 0$ are reported in the upper parts of the upper and middle panels of Table 1. When $\theta = 0.25$, there is an increase in the MSE of the Factor-IV estimator, which becomes larger than that of the standard IV for $T < 100$. However, the MSE of the Factor-IV still decreases with T, which confirms consistency, in line with Theorem 2. When instead $\theta = 0.5$, the MSE of the Factor-IV further increases and becomes rather stable with T, suggesting inconsistency, in line with the comments we made in Section 2.2 and with the results in Staiger and Stock (1997). Note that this finding does not imply that we should use standard IV in this context, even if its MSE is lower it is also inconsistent. Rather, the implication is that instrument weakness can be a problem also in the factor context, as we will also see in the empirical application in the next Section. In terms of coverage probability, the corresponding panels of Table 1 show that there is a deterioration proportional to the size of θ, with values in the range $0.88 - 0.93$ when $\theta = 0.5$.

Figures for the case $p > 0, \theta = 0$ are reported in the left parts of the upper and middle panels of Table 1. Theory predicts that there should be no effects on the standard IV estimator, since the fit of the first stage regression converges quickly to $\sum_{i=1}^{N} N^{p-1} s_{it}$, which in turn is a close estimate of the true factor. Indeed, the values of the MSE and coverage rates are very similar to those for $p = 0$. Instead, the Factor-IV estimator should loose consistency when p is too large (larger than 0.5 in this specially simple Monte Carlo design, as discussed above). When $p = 0.25$ there are very small changes in the MSE of the Factor-IV, while the MSE starts increasing substantially for $p \geq 0.33$, and for $p = 0.5$ the MSE increases with N rather than remaining stable. In this case using s_t, which contains information on the true factors f_t, is better than estimating f_t with the principal component based estimator. However, the coverage probability is less affected.

Finally, figures for the cases $p > 0, \theta > 0$ are presented in the remaining panels of Table 1. Two major features deserve a comment. First, the effects of higher θ are in general stronger for $p > 0$ than for $p = 0$, and the joint presence of $p > 0, \theta > 0$ often increases the MSE more than the sum of the separate increases for $p > 0$ and $\theta > 0$. Second, when $\theta = 0.5$ (very weak instruments) changes in p have small effects, while when $p = 0.5$ (very weak factor structure) changes in θ still matter, and in this sense weak instruments seems to matter more than weak factors. In the worst case of $p = 0.5, \theta = 0.5$, the MSEs are 10 to 30 times higher than for $p = 0, \theta = 0$, while the coverage probabilities decrease to about $0.83 - 0.88$.

Overall, the Factor-IV estimator is better than the standard IV estimator in finite samples, as long as the factor model remains strong. When the parameter p increases, the performance of the estimator of the factor $(\widehat{f_t})$ deteriorates, causing a larger variance for the factor-IV estimator of the parameters of the structural equation. A weaker link between the endogenous variables and the factors (large θ) also increases the variance of the Factor-IV estimator, but in general even more so for the standard IV estimator, unless p is large.

4.2 Additional results

The first issue we address is the effect of the use of the efficient GMM estimators of Section 3 in the presence of serial correlation of the errors of the structural equation. We let ϵ_t in (32) be either an $AR(1)$ process with AR coefficient equal to 0.5 or an $MA(1)$ process with MA coefficient equal to 0.5. These two cases are interesting because the errors of structural equations in DSGE models often have an AR structure, while an MA component emerges when expected future values of the regressors are substituted with their actual values. Results for the MSE in the AR case are presented in Table 2, those for MA case are similar and available upon request, as well as the coverage probabilities.

Reassuringly, the resulting ranking of the estimators is extremely similar to the case of no serial correlation in Table 1, with Factor-GMM having a smaller variance than standard GMM in most cases.

The second additional issue we evaluate is the role of variable preselection (i.e. selecting the variables that enter the factor analysis), since it may be conducive to better results in various modelling situations, such as forecasting macroeconomic variables (see Boivin and Ng (2006)). To assess whether such a procedure may have some relevance to our work, we consider the setup (32), (33) and (34) but, prior to using the instruments s_t either for factor estimation or to compute the standard IV estimator, we preselect the 50% of the instruments with the highest correlation with x_t. Then, we carry out standard and Factor-IV estimation as usual. Results are reported in Table 3, focusing for simplicity on a subset of the experiments ($\theta = 0$ only) .

Comparing Tables 3 and 1, it turns out that standard IV estimation is basically unaffected by instrument preselection. The MSE of Factor-IV is also not affected when $p = 0$, but it improves substantially when $p > 0$ (weak factor structure), and the coverage probability remains good. When $p > 0$, variable preselection plays a double role: it selects instruments correlated with the target, but because of this the selected instruments are also more correlated among themselves and therefore will present a stronger factor structure. Overall, with variable preselection Factor-IV estimation becomes superior to standard IV also in the case of a weak factor structure, and in the next Section we will see that this can matter in practice.

Next, we carry out a robustness check, associated with our assumption of a single factor. To check whether our analysis extends to multiple factors, we assume that there are two factors and replace (33) with

$$s_{it} = \sum_{j=1}^{r} N^{-p} \nu_{ij} f_{jt} + e_{it}, \quad i = 1, ..., N \tag{36}$$

where $\nu_{ij} \sim i.i.d.U(0,1)$. We set $p = 0$ and $\theta = 0$ and run a Monte Carlo exercise leaving the rest of the Monte Carlo settings as before. We report RMSE results in Table 4. We see a pattern of results that is similar to the single factor case. If anything, Factor IV performs better that standard IV, compared to the single factor case.

Further, we consider the performance of the estimator proposed by Kapetanios and Marcellino (2008) in our setting. Kapetanios and Marcellino (2008) suggest using the cross-sectional averages (CSA) of s_{it} as instruments. If $k = 1$, then the CSA estimator is consistent if the coefficients of s_{it} in the regression model of x_t do not have zero mean, while if $k > 1$, more restrictive conditions are needed. We know from the Monte Carlo study of Kapetanios and Marcellino (2008) that the estimator works better than Factor-IV if there is no factor structure but s_{it} enters directly in x_t. Our current setting is different. s_{it} only enters via the factors in x_t. It is of interest to examine the performance of CSA IV in this context when the factor structure gets weaker. We expect that in this case where s_{it} has information for x_t only via the factors, the two estimators will perform similarly.We report RMSE results for CSA-IV in Table 5. We see that, as expected given the above discussion, CSA-IV and Factor-IV perform very similarly.

Finally, we consider the ability of our modified infomation criterion, analysed in Theorem 6, to provide evidence on the weakness of the factor structure, and thereby on the validity of the estimated factors as instruments for Factor-IV. We use the same setup as that used to produce the results in Table 1 but consider $\theta = 0$ and $p = 0, 0.25, 0.5, 0.75$. We use the information criterion analysed in Theorem 6, to estimate the number of factors while allowing that zero factors be selected. The average number of factors selected are reported in Table 6. We see that while the criterion always selects 1 factor when $p = 0$, it selects a smaller number than 1 for $p = 0.25$, while for the extreme cases, where $p = 0.5, 0.75$, and Factor IV should not be used, the criterion correctly suggests that no factor structure exists.

5 Empirical Application

In this Section we discuss two empirical applications of the Factor-GMM procedure, and compare the results with the CSA GMM estimation of Kapetanios and Marcellino (2008). The first application concerns estimation of a forward looking Taylor rule, along the lines of Clarida, Galí, and Gertler (1998) (CGG), Clarida, Galí, and Gertler (2000) (CGG2)) and Favero, Marcellino, and Neglia (2005). The second application focuses on estimation of a New-Keynesian Phillips curve, along the lines of Galí and Gertler (1999) (GG 1999) and Beyer, Farmer, Henry, and Marcellino (2008).

5.1 Taylor Rule

For the Taylor rule, we adopt the following specification :

$$r_t = \alpha + (1-\rho)\beta(\pi_{t+12} - \pi_t^*) + (1-\rho)\gamma(y_t - y_t^*) + \rho r_{t-1} + \epsilon_t, \quad (37)$$

where $\epsilon_t = (1-\rho)\beta(\pi_{t+12}^e - \pi_{t+12}) + v_t$, and v_t is an i.i.d. error. We use the federal funds rate for r_t, annual cpi inflation for π_t, 2% as a measure of the inflation target π_t^*, and the potential output y_t^* is the Hodrick Prescott filtered version of the IP series. Since π_{t+12} and $y_t - y_t^*$ are correlated with the error term ϵ_t, and the error term has an MA structure, we adopt GMM estimation with a correction for the MA component in the error ϵ_t and a proper choice of instruments.

In particular, we use a HAC estimator for the weighting matrix, based on a Bartlett kernel with Newey and West (1994) automatic bandwith selection. For the set of instruments, in the base case the choice is similar to that in CGG and CGG2. We use one lag of the output gap, inflation, commodity price index, unemployment and interest rate. We focus on the period 1985-2003, since Beyer, Farmer, Henry, and Marcellino (2008) have detected instability in Phillips curves and Taylor rules estimated on a longer sample with an earlier start date.

For the Factor GMM estimator, we add to the set of instruments the (one period lagged) factors extracted from a large dataset of 132 monthly macroeconomic and financial variables for the US, extracted from the dataset in Stock and Watson (2005). The number of factors is eight, as indicated by the Bai and Ng (2002) criteria, which suggests that the factor structure is rather weak. As in the Monte Carlo experiments, we also consider a subset of 12 of the 132 variables, those with an absolute correlation with inflation higher than 0.40, since this can strenghten the factor structure and improve the information content of the factors for future inflation. In fact, in this case one factor explains over 60% of the variance of all variables, and we use one to twelve lags of this factor as instruments, in addition to those in the base case.[1]

For the CSA GMM, we add to the basic set of instruments the simple average of either all the (standardized) 132 macroeconomic variables, or of only the subset of 12 variables mostly correlated with inflation. In both cases, we included one to twelve lags of the averages as instruments.

Finally, we also considered one lag of the 12 selected macroeconomic variables as instruments, to compare the performance of Factor and CSA IV with a relatively small set of instruments.

The results from the six estimation methods (Base, Factor-GMM All data, CSA-GMM All data, Factor GMM Select data, CSA GMM Select data, and Select data as instruments)

[1] We select the variables using inflation as the target rather than the output gap since the first stage equation for the gap is unproblematic, with an adjusted R^2 close to 0.90, mostly due to the high explanatory power of lagged gap.

are reported in Table 7. For the base case, the estimated values for β and γ are, respectively, about 2.3 and 1, and the fact that the output gap matters less than inflation is not surprising. The persistence parameter, ρ, is about 0.88, in line with other studies. An LM test for the null hypothesis of no correlation in the residuals of an MA(11) model for $\widehat{\epsilon}_t$ does not reject the null hypothesis, which provides evidence in favor of the correct dynamic specification of the Taylor rule in (37). The p-value of the J-statistic for instrument validity is 0.11, so that the null hypothesis is not rejected at the conventional level of 10%.

Adding the "All data" factors to the instrument set does not improve the precision of the estimators of ρ, γ and β. Instead, the CSA GMM using "All data" produces a major reduction in the variance of the estimators. Using the "Select data" factors, the precision of the Factor GMM improves substantially, becoming much lower than for the base case and comparable to that of the CSA GMM based on the "Select data". Using directly the lagged "Select data" as additional instruments produces bad results in terms of variances of the estimators, much worse than in the base case for γ and β. The point parameter estimates are also fairly different from the other five cases. These findings indicate that GMM estimation based on 18-20 macroeconomic instruments can already be problematic.

Finally, a regression of future (12 months ahead) inflation on the alternative sets of instruments indicates that each set of factors is significant at the 10% level when added to the macro variables, while the CSA from the "All data" are not, and those from "Select data" only marginally so. However, a few of the lagged CSA variables are strongly significant in both cases. Moreover, the values of the adjusted R^2 in these equations are all of comparable size. Instead, the additional regressors are often not significant in the first stage regression for the output gap, since the latter is mostly explained by its own lag.

5.2 New Keynesian Phillips Curve

For the second empirical example, the New-Keynesian Phillips curve is specified as,

$$\pi_t = c + \gamma\pi_{t+1} + \alpha x_t + \rho\pi_{t-1} + \epsilon_t, \tag{38}$$

where $\epsilon_t = \gamma(\pi^e_{t+1} - \pi_{t+1}) + v_t$, and v_t is an i.i.d. error. Moreover, π_t is annual CPI inflation, π^e_{t+1} is the forecast of π_{t+1} made in period t, and x_t is unemployment, with reference to Okun's law, as in e.g. Beyer and Farmer (2004). As for the Taylor rule, π_{t+1} and x_t are correlated with the error term ϵ_t, which in turn is correlated over time. Hence, we estimate the parameters of (38) by GMM, with a correction for the MA component in the error ϵ_t, and the same six sets of instruments as for the Taylor rule.

The results are reported in Table 8. For the base case, the coefficient of the forcing variable is not statistically significant (though it has the correct sign), while the coefficients of the backward and forward looking components of inflation, ρ and γ, are similar and close to 0.5.

Adding the "All data" factors to the instrument set improves the precision of the estimators of all parameters, but the gains are much larger with the "Select" data factors. For the latter, the gains are about 10% for α and 120% for γ and ρ. Moreover, a regression of future (1 month ahead) inflation on the instruments indicates that only the Select data factors are strongly significant when added to the set of macroeconomic regressors. As for the Taylor rule, the CSA GMM based on "All data" performs better than the corresponding Factor GMM. However, CSA and Factor GMM based on "Select data" produce very similar results in terms of both point estimates of the parameters, and the variances of the estimates. The CSA from "Select data" are also strongly jointly significant in a regression of future (1 month ahead) inflation on the instruments.

We should mention that, as for the Taylor rule, there is little role for the factors in the first step regression for the other potentially endogenous variable, the unemployment rate. The adjusted R^2 in this equation is about 0.98, moslty due to the strong explanatory power of lagged unemployment. Finally, for estimation of the New Keynesian Phillips curve, using directly the "Select data" as instruments is slightly better in terms of efficiency than the base case, but much worse than either CSA or Factor GMM.

In summary, these two empirical examples confirm the practical relevance of factors as additional instruments to enhance the efficiency of GMM estimation, possibly after variable preselection when the factors are weak.

6 Conclusions

In this paper, and in a related independent article by Bai and Ng (2009), we develop the theoretical underpinnings of Factor-GMM estimation, which was used by Favero, Marcellino, and Neglia (2005) and Beyer, Farmer, Henry, and Marcellino (2008) to estimate structural forward looking equations, such as those typically encountered in DSGE models. We show that when the endogenous variables in a structural equation are explained by a set of unobservable factors, which are also the driving forces of a larger set of instrumental variables, using the estimated factors as instruments rather than the large set of instrumental variables yields sizeable efficiency gains. Bai and Ng (2009) show that a similar finding remains valid in a system framework, and the same would be true for our methodology.

We then extend the basic results in two directions. First, we assess the consequences of a weak factor structure. In this case, the by now standard principal component based estimators of the factors can be no longer consistent, basically because the factor model is no longer identified. However, we show that these factor estimators remain consistent even if the factor loadings in the factor model converge to zero, but at a sufficiently slow rate as a function of N. In this case, it is still possible to use Factor-IV estimators with well defined asymptotic properties.

Second, we evaluate what happens when the instruments are weak, possibly combined with a weak factor structure. It is still possible to derive standard and Factor-IV estimators with well defined asymptotic properties, when the parameters in the equation that relates the instruments (or the factors) to the endogenous variables converge to zero at a sufficiently slow rate. Both types of "weaknesses", in the factor structure and/or in the instruments, imply a slower convergence rate of the instrumental variable estimators.

To evaluate the finite sample properties of the Factor-IV estimators, we conduct an extensive set of simulation experiments. The results indicate that Factor-IV estimation is in general more effective than standard IV estimation, intuitively because the information in a large set of weak instruments in condensed in just a few variables. Variable preselection is helpful to strengthen the factor structure and further increase the efficacy of the Factor-IV estimator. Similar results hold for Factor-GMM estimators.

Finally, we apply Factor-GMM for the estimation of a New Keynesian Phillips curve for the US, using factors extracted from a large set of macroeconomic variables. The findings confirm the empirical relevance of the theoretical results in this paper, in particular when the instrumental variables are pre-selected in a first stage, based on their correlation with the endogenous variable(s). Variable pre-selection can in fact alleviate both the weak instrument problem, since only instruments correlated with the target variable(s) are retained, and the weak factor structure problem, since more homogeneous variables are retained. In such a context, the gains from Factor-GMM estimation with respect to standard GMM estimation can be fully exploited.

References

Amemiya, T. (1966): "On the Use of Principal Components of Independent Variables in Two-Stage Least-Squares Estimation," *International Economic Review*, 7, 283–303.

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.

——— (2009): "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, Forthcoming.

Beck, G., K. Hubrich, and M. Marcellino (2009): "Regional inflation dynamics within and across euro area countries and a comparison with the US," *Economic Policy*, 24, 141–184.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

Bernanke, B., J. Boivin, and P. S. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387–422.

Beyer, A., and R. Farmer (2004): "On the indeterminacy of New-Keynesian economics," *Computing in Economics and Finance*, 152.

Beyer, A., R. Farmer, J. Henry, and M. Marcellino (2008): "Factor analysis in a new-Keynesian model," *Econometrics Journal*, 11(2), 271–286.

Boivin, J., and S. Ng (2006): "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 127, 169–194.

Chao, J., J. Hausman, W. Newey, N. R. Swanson, and T. M. Woutersen (2006): "IV Estimation with Heteroscedasticity and Many Instruments," *mimeo*.

Chao, J. C., and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692.

Clarida, R., J. Galí, and M. Gertler (1998): "Monetary policy rules in practice: Some international evidence," *European Economic Review*, 42, 1033–1067.

——— (2000): "Monetary policy rules and macroeconomic stability: evidence and some theory," *Quarterly Journal of Economics*, 115, 147–180.

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

Dufour, J. M., L. Khalaf, and M. Kichian (2006a): "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis," *Journal of Economic Dynamics and Control*, 30, 1707–1727.

——— (2006b): "Structural Estimation and Evaluation of Calvo-Style Inflation Models," *Mimeo, University of Montreal.*

Favero, C., M. Marcellino, and F. Neglia (2005): "Principal components at work: the empirical analysis of monetary policy with large datasets," *Journal of Applied Econometrics*, 20, 603–620.

Florens, J. P., J. Johannes, and S. Van Bellegem (2006): "Instrumental Regression in Partially Linear Models," *CORE Discussion Paper No. 2006/25.*

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalised Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554.

——— (2004): "The Generalised Factor Model: Consistency and Rates," *Journal of Econometrics*, Forthcoming.

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

Giannone, D., L. Reichlin, and L. Sala (2002): "Tracking Greenspan: systematic and unsystematic monetary policy revisited," *CEPR Working Paper No. 3550.*

Hahn, J., and G. Kuersteiner (2002): "Discontinuities of Weak Instrument Limiting Distributions," *Economics Letters*, 75, 325–331.

Han, C., and P. C. B. Phillips (2006): "GMM with Many Moment Conditions," *Econometrica*, 74, 147–182.

HANSEN, C., J. HAUSMAN, AND W. K. NEWEY (2009): "Many Instruments, Weak Instruments and Microeconometric Practice," *Journal of Business and Economic Statistics*, Forthcoming.

KAPETANIOS, G., AND M. MARCELLINO (2006): "Factor-GMM Estimation with Large Sets of Possibly Weak Instruments," *Queen Mary University of London Working Paper No. 577.*

——— (2008): "Cross-sectional Averaging and Instrumental Variable Estimation with Many Weak Instruments," *Queen Mary University of London Working Paper No. 627.*

——— (2009): "A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions," *Journal of Time Series Analysis*, 30, 208–238.

KLOEK, T., AND L. MENNES (1960): "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables," *Econometrica*, 28, 46–61.

LUTKEPOHL, H. (1996): *Handbook of Matrices.* Wiley.

MORIMUNE, K. (1983): "Approximate Distributions of k-class Estimators when the Degree of Overidentification is Large Compared with the Sample Size," *Econometrica*, 51, 821–842.

NEWEY, W., AND K. WEST (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

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. (2004): "Many Weak Moment Asymptotics for the Continuously Updated GMM Estimator," *Working Paper, MIT.*

ONATSKI, A. (2006): "Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak," *mimeo.*

SCHWARZ, H., H. R. RUTISHAUSER, AND E. STIEFEL (1973): *Numerical Analysis of Symmetric Matrices.* Prentice Hall.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

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 M. Yogo (2003): “Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments,” in *Identification and Inference for Econometric Models: Essays in Honour of Thomas J. Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge University Press.

Appendix

Proof of Theorem 1

Asymptotic normality for the estimators follows straightforwardly from the martingale difference central limit theorem given Assumption 2, while unbiasedness follows from $\frac{F'\epsilon}{T} \xrightarrow{p} 0$ and $\frac{Z'\epsilon}{T} \xrightarrow{p} 0$. We then examine the asymptotic variances. The general expressions for the covariance matrices of $\sqrt{T}(\bar{\beta}-\beta)$ are given by the probability limits as $T \to \infty$ of $\left(\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T}\right)^{-1}$. The following probability limits, using standard laws of large numbers and the uncorrelatedness of u_t and v_t, give the required ingredients for the results

$$\frac{F'F}{T} \xrightarrow{p} \Sigma_f \tag{39}$$

$$\frac{Z'Z}{T} = \frac{\left(F\Lambda_Z^0 + v_Z\right)'\left(F\Lambda_Z^0 + v_Z\right)}{T} \xrightarrow{p} \Lambda_Z^{0\prime}\Sigma_f\Lambda_Z^0 + \Sigma_{vZ}$$

$$\frac{X'F}{T} = \frac{\left(A_Z^{0\prime}\left(\Lambda_Z^{0\prime}F' + v_Z'\right) + A^{0\prime}F' + u'\right)F}{T} \xrightarrow{p} \left(A_Z^{0\prime}\Lambda_Z^{0\prime} + A^{0\prime}\right)\Sigma_f$$

$$\frac{X'Z}{T} = \frac{\left(A_Z^{0\prime}\left(\Lambda_Z^{0\prime}F' + v_Z'\right) + A^{0\prime}F' + u'\right)\left(F\Lambda_Z^0 + v_Z\right)}{T} \xrightarrow{p}$$

$$A_Z^{0\prime}\Lambda_Z^{0\prime}\Sigma_f\Lambda_Z^0 + A_Z^{0\prime}\Sigma_{vZ} + A^{0\prime}\Sigma_f\Lambda_Z^0$$

Then, the first part of the Theorem follows. For the second part, it is sufficient to prove that

$$\sqrt{T}\left(\left(X'F(F'F)^{-1}F'X\right)^{-1}X'F(F'F)^{-1}F'\epsilon - \left(X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'X\right)^{-1}X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'\epsilon\right) = o_p(1) \tag{40}$$

(40) follows if

$$\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T} - \frac{X'\hat{F}}{T}\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{\hat{F}'X}{T} = o_p(1) \tag{41}$$

and

$$\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\epsilon}{T^{1/2}} - \frac{X'\hat{F}}{T}\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{\hat{F}'\epsilon}{T^{1/2}} = o_p(1) \tag{42}$$

(41) and (42) follow if

$$\frac{X'F}{T} - \frac{X'\hat{F}}{T} = o_p(1), \quad \frac{F'F}{T} - \frac{\hat{F}'F}{T} = o_p(1) \tag{43}$$

$$\sqrt{T}\left(\frac{F'\epsilon}{T} - \frac{\hat{F}'\epsilon}{T}\right) = o_p(1) \tag{44}$$

hold. We examine (43)-(44). They can all be written as

$$A_T \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = o_p(1) \tag{45}$$

where A_T is $1, 1$ and $\sqrt{T}$ and q_t is x_t, f_t and ϵ_t respectively for (43)-(44). By Lemma A.1 of Bai and Ng (2006) we have that

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = O_p\left(\min(N,T)^{-1}\right) \tag{46}$$

as long as q_t has finite fourth moments, nonsingular covariance matrix and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q_t$ satisfies a central limit theorem. These conditions are satisfied for x_t, f_t and ϵ_t via assumptions 2 and 3. Hence, (43) follows, while (44) follows if $\sqrt{T}/N = o(1)$.

Proof of Theorem 2

For the sake of brevity, we assume that $A^0_{(T)} = 0$, and we show that in this case $T^{1/2-\psi}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_\epsilon \left(\Upsilon'\Sigma_f\Upsilon\right)^{-1}\right)$ .Then, (13)-(15) follow similarly but with a more complex notation. To start with, we examine the asymptotic distribution of $T^{1/2-\psi}\left(X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'X\right)^{-1}X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'\epsilon$. Using (45) for $A_T = T^{\psi}$ and since by assumption $N = O(T^{\gamma})$, $\gamma > 1/2$, it is sufficient to examine the asymptotic distribution of $\left(\frac{X'F}{T^{1-\psi}}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T^{1-\psi}}\right)^{-1}\frac{X'F}{T^{1-\psi}}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\epsilon}{T^{1/2}}$. A standard central limit theorem suffices to show that under assumptions 1-3, $T^{-1/2}F'\epsilon \xrightarrow{d} N\left(0, \sigma^2_\epsilon\Sigma_f\right)$. We examine the limits of $\frac{F'F}{T}$ and $\frac{X'F}{T^{1-\alpha}}$. The first is given in (39). We examine the second. We have

$$\frac{X'F}{T^{1-\psi}} = \frac{\left(A^{0\prime}_{(T)Z}\left(\Lambda^{0\prime}_Z F' + v_Z'\right) + u'\right)F}{T^{1-\psi}} = \frac{A^{0\prime}_{(T)Z}\Lambda^{0\prime}_Z}{T^{-\psi}}\frac{F'F}{T} + \frac{A^{0\prime}_{(T)Z}v_Z'F}{T^{1-\psi}} + \frac{u'F}{T^{1-\psi}} \tag{47}$$

The second and third terms of the RHS of (47) tend to zero since $1 - \psi > 1/2$. The first term tends to $\Upsilon'\Sigma_f$. Hence, the result follows.

Proof of Theorem 3

The covariance matrix of Z, Σ_Z, is given by $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N + \Sigma_v$ where Σ_f and Σ_v are the covariance matrices of F and v respectively. By Weyl's theorem (see 5.3.2(9) of Lutkepohl (1996)) the eigenvalues of Σ_Z are bounded if the eigenvalues of $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N$ and Σ_v are bounded. By assumption the eigenvalues of Σ_v are bounded. Hence we examine $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N$. By Schwarz, Rutishauser, and Stiefel (1973), the eigenvalues of $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N$, will be bounded if the column sum norm of $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N$ is bounded. But every element of $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N$ is $O(N^{-2\alpha})$. Hence the column sum norm of $\Lambda^{0\prime}_N\Sigma_f\Lambda^0_N$ is O(1) for all $\alpha \geq 1/2$. Hence the result follows.

Proof of Theorem 4

We follow the proof of Theorem 1 of Bai and Ng (2002). A crucial difference is that because of the local nature of Λ_N^0 we use a different normalisation for Λ. Therefore, rather than the normalisation $\Lambda'\Lambda/N = I$, we use $\Lambda'\Lambda/N^{1-2\alpha} = I$. This leads to the mathematical identities $\hat{F} = N^{-1+2\alpha}X\tilde{\Lambda}$ and $\tilde{\Lambda} = T^{-1}X'\tilde{F}$ where $\tilde{F}$ is the solution to the optimisation problem of maximising $tr(F'(X'X)F)$ subject to $F'F/T = I$. Let $H = ((\tilde{F}'F/T)(\Lambda_N^{0\prime}\Lambda_N^0/N^{1-2\alpha}))'$. Then,

$$\hat{f}_t - Hf_t = N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t) + N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\zeta_{st} + N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\eta_{st} + N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\xi_{st}$$

where $\gamma_N(s,t) = E(e_s'e_t/N)$

$$\zeta_{st} = e_s'e_t/N - \gamma_N(s,t) \tag{48}$$

$$\eta_{st} = f_s'^0\Lambda_N^0 e_t/N \tag{49}$$

$$\xi_{st} = f_t'^0\Lambda_N^0 e_s/N = \eta_{ts} \tag{50}$$

It is easy to see that $||\hat{f}_t - Hf_t||^2 \leq 4(a_t + b_t + c_t + d_t)$, where $a_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t)\right\|^2$, $b_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\zeta_{st}\right\|^2$, $c_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right\|^2$, $d_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\xi_{st}\right\|^2$. It follows that $1/T\sum_{t=1}^{T}||\hat{f}_t - Hf_t||^2 \leq c/T\sum_{t=1}^{T}(a_t + b_t + c_t + d_t)$, for some constant c. Now, $\left\|\sum_{s=1}^{T}\tilde{f}_t\gamma_N(s,t)\right\|^2 \leq \left(\sum_{s=1}^{T}\left\|\tilde{f}_s\right\|^2\right)\left(\sum_{s=1}^{T}\gamma_N^2(s,t)\right)$, which implies that $1/T\sum_{t=1}^{T}a_t = O_p(N^{4\alpha}T^{-1})$. Following the analysis for $N^{-4\alpha}b_t$ given in the proof of Bai and Ng (2002) we see that $1/\left(N^{4\alpha}T\right)\sum_{t=1}^{T}b_t = O_p(N^{-1})$, and hence $1/T\sum_{t=1}^{T}b_t = O_p(N^{-1+4\alpha}) = o_p(1)$, as long as $a < 1/4$. Finally we look at c_t. d_t can be treated similarly.

$$c_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right\|^2 = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s f_s'^0\Lambda_N^0 e_t/N\right\|^2 = N^{2\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s f_s'^0\Lambda^0 e_t/N\right\|^2$$

$$\leq N^{-2+2\alpha}\left\|\Lambda^0 e_t\right\|^2\left(T^{-1}\sum_{s=1}^{T}\left\|\tilde{f}_s\right\|^2\right)\left(T^{-1}\sum_{s=1}^{T}\|f_s\|^2\right) = N^{-2+2\alpha}\left\|\Lambda^0 e_t\right\|^2 O_p(1)$$

So

$$1/T\sum_{t=1}^{T}c_t = O_p(1)N^{-1+2\alpha}T^{-1}\sum_{t=1}^{T}\left\|\frac{\Lambda^0 e_t}{N^{1/2}}\right\|^2 = O_p(N^{-1+2\alpha})$$

Proof of Lemma 1

We follow the proof of Lemma A.1 of Bai and Ng (2006). Using (48)-(50) we get

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t)\right)q_t' + N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\zeta_{st}\right) + \tag{51}$$

$$N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right)q_t' + N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\xi_{st}\right)q_t'$$

The first two terms of (51) apart from the normalisation $N^{2\alpha}$ are the same as those analysed in Lemma A.1 of Bai and Ng (2006). Thus, under the assumption of the Lemma for q_t, we immediately get that they are $O_p\left(N^{2\alpha}T^{-1/2}\min(N,T)^{-1/2}\right)$ and $O_p\left(N^{2\alpha-1/2}\min(N,T)^{-1/2}\right)$ respectively. Thus, the sum of these two terms is $O_p\left(N^{2\alpha}\min(N,T)^{-1}\right)$. The third and the fourth term of (51) are analysed similarly. We focus on the third term. We have

$$N^{2\alpha}T^{-2}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right)q_t' = N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}Hf_s\eta_{st}\right)q_t' + N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\left(\tilde{f}_s - Hf_s\right)\eta_{st}\right)q_t'. \tag{52}$$

The first term on the RHS of (52) can be written as

$$N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}Hf_s\eta_{st}\right)q_t' = N^{2\alpha}\left(H\frac{1}{T}\sum_{t=1}^{T}f_sf_s'\right)\frac{1}{NT}\sum_{t=1}^{T}\Lambda_N^0 e_t q_t' =$$

$$N^{2\alpha}\left(H\frac{1}{T}\sum_{t=1}^{T}f_sf_s'\right)\frac{N^{-2\alpha}}{NT}\sum_{t=1}^{T}\Lambda^0 e_t q_t' = \left(H\frac{1}{T}\sum_{t=1}^{T}f_sf_s'\right)\frac{1}{NT}\sum_{t=1}^{T}\Lambda^0 e_t q_t' = O_p\left((NT)^{-1/2}\right)$$

For the second term of (52) we have

$$\left\|N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\left(\tilde{f}_s - Hf_s\right)\eta_{st}\right)q_t'\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{f}_s - Hf_s\right\|^2\right)^{1/2}\left(N^{4\alpha}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\eta_{st}q_t'\right\|^2\right)^{1/2}$$

But $\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{f}_s - Hf_s\right\|^2\right)^{1/2} = O_p\left(\min\left(N^{-2\alpha}T^{1/2}, N^{1/2-2\alpha}\right)^{-1}\right)$, by Theorem 4. Then,

$$N^{4\alpha}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\eta_{st}q_t'\right\|^2 = N^{4\alpha}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{f_s'^0\Lambda_N^0 e_t}{N}q_t'\right\|^2 = N^{2\alpha}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{f_s'^0\Lambda^0 e_t}{N}q_t'\right\|^2 \tag{53}$$

But, $\frac{1}{T}\sum_{t=1}^{T}\frac{f_s'^0\Lambda^0 e_t}{N}q_t' = O_p(N^{-1/2})$ and so the RHS of (53) is also $O_p(N^{-1+2\alpha})$. So the second term of (52) is

$$O_p\left(\min\left(N^{-2\alpha}T^{1/2}, N^{1/2-2\alpha}\right)^{-1}\left(N^{1/2-\alpha}\right)^{-1}\right) = O_p\left(\min\left(N^{1/2-3\alpha}T^{1/2}, N^{1-3\alpha}\right)^{-1}\right)$$

As a result, the third term of (51) is $O_p\left(\min\left(N^{1/2-3\alpha}T^{1/2}, N^{1-3\alpha}, (NT)^{1/2}\right)^{-1}\right)$. Thus, overall

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = O_p\left(\min\left(N^{1/2-3\alpha}T^{1/2}, N^{1-3\alpha}, N^{-2\alpha}\min(N,T), (NT)^{1/2}\right)^{-1}\right) = O_p\left(C_{NT}^{-1}\right)$$

where

$$C_{NT} \equiv \min\left(N^{1/2-3\alpha}T^{1/2}, N^{1-3\alpha}, N^{-2\alpha}\min(N,T)\right) \tag{54}$$

since $(NT)^{1/2}$ is always larger than $N^{-2\alpha}\min(N,T)$

Proof of Theorem 5

The results follow from Lemma 1, (45) and the proofs of Theorems 1 and 2. More particularly, the results of Theorem 1 require a result similar to (45) which is now provided by Lemma 1. Then, the results of Theorem 1 are obtained immediately. Similarly, once Lemma 1 is used (13)-(15) of Theorem 2 follow.

Proof of Theorem 6

The theorem is proved if we show that the sufficient conditions for consistency given in Theorem 2 of Bai and Ng (2002) are satisfied for all $0 \leq \alpha < 1/4$. Let $\tilde{C}_{NT} = \min\left(N^{-4\alpha}T, N^{1-4\alpha}\right)$ denote the rate derived in Theorem 3.. The two conditions of Theorem 2 of Bai and Ng (2002) are firstly that $g(N,T) \to 0$ (Condition 1) and secondly that $\tilde{C}_{NT}g(N,T) \to \infty$ (Condition 2). First, note that Corollary 2 of Bai and Ng (2002) implies that the results of Theorem 2 of Bai and Ng (2002) hold for a an unspecified estimator with unspecified rate $\bar{C}_{NT}$. Condition 1 is easily seen to be satisfied for $g(N,T)$ in (26). Since $\tilde{C}_{NT}$ grows polynomially in $\min(N,T)$, for all $0 \leq \alpha < 1/4$, Condition 2 is also seen to be satisfied for all $0 \leq \alpha < 1/4$. Hence, the Theorem holds.

Proof of Theorem 7

The general expression for the covariance matrices of $\sqrt{T}(\bar{b} - \beta)$ is given by the probability limits as $T \to \infty$ of $\left(\frac{X'F}{T} S_{f\epsilon}^{-1} \frac{F'X}{T}\right)^{-1}$. The results follow from those in the Proof of Theorem 1 and consistency of the HAC estimator of S. Next, we need to prove that

$$\left(\frac{X'F}{T}S_{f\epsilon}^{-1}\frac{F'X}{T}\right)^{-1}\frac{X'F}{T}S_{f\epsilon}^{-1}\frac{F'\epsilon}{T^{1/2}} - \left(\frac{X'\hat{F}}{T}\widehat{S}_{\hat{f}\epsilon}^{-1}\frac{\hat{F}'X}{T}\right)^{-1}\frac{X'\hat{F}}{T}\widehat{S}_{\hat{f}\epsilon}^{-1}\frac{\hat{F}'\epsilon}{T^{1/2}} = o_p(1)$$

From the proof of Theorem 3 we already know that $\frac{X'F}{T} - \frac{X'\hat{F}}{T} = o_p(1)$, $\sqrt{T}\left(\frac{F'\epsilon}{T} - \frac{\hat{F}'\epsilon}{T}\right) = o_p(1)$. Then we have, $\widehat{S}_{\hat{f}\epsilon,h} = \widehat{\Phi}_0 + \sum_{j=1}^{h}\left(1 - \frac{j}{h+1}\right)(\widehat{\Phi}_j + \widehat{\Phi}_j')$, $\widehat{\Phi}_j = T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_{t-j}\widehat{f}_t\widehat{f}_{t-j}'$, $\Phi_j = T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_{t-j}f_tf_{t-j}'$, so that $\Phi_j - \widehat{\Phi}_j = T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_{t-j}\left(f_tf_{t-j}' - \widehat{f}_t\widehat{f}_{t-j}'\right)$. 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

$$\left\|T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t^2\left(f_tf_t' - \widehat{f}_t\widehat{f}_t'\right)\right\| \leq C_1\left\|T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t^2 f_t'\left(Hf_t - \widehat{f}_t'\right)\right\| \leq$$

$$C_2\left\|T^{-1}\sum_{T=j+1}^{T}\epsilon_t^2 f_t'\left(Hf_t - \widehat{f}_t'\right)\right\| + C_3\left\|T^{-1}\sum_{T=j+1}^{T}(\widehat{\epsilon}_t - \epsilon_t)f_t'\left(Hf_t - \widehat{f}_t'\right)\right\|$$

for some constants C_1,C_2 and C_3. But, by (45) and $\sqrt{T}/N = o(1)$, $\left\|T^{-1}\sum_{T=j+1}^{T}\epsilon_t^2 f_t'\left(Hf_t - \widehat{f}_t'\right)\right\| = o_p\left(T^{-1/2}\right)$ as long as ϵ_t has finite eighth moments. Then,

$$\left\|T^{-1}\sum_{T=j+1}^{T}(\widehat{\epsilon}_t - \epsilon_t)f_t'\left(Hf_t - \widehat{f}_t'\right)\right\| \leq C_5\|b - \beta\|\left\|T^{-1}\sum_{T=j+1}^{T}x_tf_t'\left(Hf_t - \widehat{f}_t'\right)\right\|$$

for some constant C_5. Again by (45) and $\sqrt{T}/N = o(1)$, $\left\|T^{-1}\sum_{T=j+1}^{T}x_tf_t'\left(Hf_t - \widehat{f}_t'\right)\right\| = o_p\left(T^{-1/2}\right)$. By consistency of b, $\|b - \beta\| = o_p(1)$. Hence $\Phi_0 - \widehat{\Phi}_0 = o_p(h^{-1})$ as long as $h = o(T^{1/2})$.

Proof of Theorem 8

It follows from the proof of Theorem 2, given Theorem 7 and consistency of $\widehat{S}_{\hat{f}\epsilon}$.

Table Appendix

Table 1. Results for Factor and Standard IV estimators													
RMSE Results for Standard IV													
	θ	0				0.25				0.5			
p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
0	30	0.370	0.375	0.370	0.369	0.620	0.590	0.618	0.604	0.688	0.696	0.694	0.689
0	50	0.284	0.366	0.364	0.371	0.587	0.624	0.629	0.626	0.686	0.685	0.686	0.683
0	100	0.180	0.250	0.359	0.378	0.542	0.592	0.626	0.658	0.700	0.697	0.700	0.700
0	200	0.111	0.150	0.243	0.352	0.500	0.566	0.629	0.655	0.710	0.699	0.704	0.710
0.25	30	0.378	0.377	0.372	0.376	0.601	0.607	0.610	0.599	0.700	0.683	0.691	0.678
0.25	50	0.295	0.364	0.369	0.372	0.594	0.607	0.630	0.625	0.683	0.697	0.686	0.699
0.25	100	0.189	0.246	0.355	0.364	0.552	0.603	0.648	0.653	0.697	0.691	0.695	0.693
0.25	200	0.120	0.164	0.243	0.359	0.516	0.565	0.612	0.654	0.697	0.685	0.713	0.713
0.33	30	0.369	0.368	0.371	0.377	0.597	0.612	0.601	0.610	0.696	0.699	0.696	0.692
0.33	50	0.301	0.366	0.370	0.367	0.602	0.626	0.617	0.613	0.693	0.693	0.693	0.695
0.33	100	0.197	0.264	0.360	0.361	0.563	0.612	0.627	0.653	0.693	0.694	0.699	0.705
0.33	200	0.127	0.170	0.260	0.356	0.531	0.584	0.619	0.663	0.702	0.700	0.706	0.691
0.5	30	0.377	0.371	0.379	0.378	0.615	0.609	0.613	0.616	0.704	0.713	0.686	0.717
0.5	50	0.316	0.368	0.369	0.361	0.601	0.624	0.618	0.637	0.684	0.712	0.712	0.691
0.5	100	0.241	0.290	0.367	0.361	0.594	0.611	0.643	0.642	0.718	0.691	0.688	0.704
0.5	200	0.164	0.212	0.286	0.358	0.564	0.600	0.642	0.657	0.702	0.718	0.722	0.709
RMSE Results for Factor IV													
0	30	0.229	0.217	0.210	0.227	1.069	1.005	0.969	0.794	1.853	1.881	1.986	1.617
0	50	0.170	0.163	0.161	0.160	0.777	0.729	0.754	0.768	1.828	1.741	1.982	1.790
0	100	0.108	0.105	0.108	0.103	0.695	0.561	0.586	0.475	1.570	1.858	1.962	1.903
0	200	0.075	0.072	0.071	0.075	0.327	0.409	0.469	0.444	2.003	1.771	1.858	1.909
0.25	30	0.286	0.304	0.241	0.224	0.962	1.110	0.930	1.187	1.779	1.818	1.772	1.846
0.25	50	0.176	0.182	0.162	0.172	0.979	0.904	0.769	0.861	1.954	1.815	1.832	1.779
0.25	100	0.118	0.114	0.111	0.105	0.762	0.662	0.679	0.493	1.845	1.761	1.805	1.674
0.25	200	0.080	0.078	0.075	0.076	0.413	0.455	0.477	0.389	1.844	1.801	1.910	1.664
0.33	30	0.532	0.411	0.499	0.436	1.387	1.154	1.309	1.385	1.802	1.898	2.071	1.881
0.33	50	0.190	0.381	0.199	0.190	0.919	0.967	0.980	1.072	1.749	1.887	1.804	1.990
0.33	100	0.126	0.119	0.117	0.119	0.648	0.758	0.614	0.660	1.869	1.824	1.813	1.755
0.33	200	0.085	0.084	0.082	0.081	0.534	0.574	0.508	0.438	1.918	1.977	1.932	1.700
0.5	30	1.311	1.548	1.651	1.759	2.041	2.156	1.984	2.087	2.026	2.154	2.147	1.925
0.5	50	1.009	1.320	1.462	1.521	1.685	2.042	1.927	1.864	2.184	1.954	1.956	2.054
0.5	100	0.541	0.792	1.096	1.463	1.481	1.586	1.859	2.141	2.014	1.996	2.155	2.201
0.5	200	0.117	0.230	0.430	0.869	0.998	1.122	1.449	1.868	1.971	2.034	1.855	2.054
Estimated Coverage Probability for Factor IV													
0	30	0.946	0.937	0.938	0.928	0.933	0.932	0.940	0.914	0.906	0.882	0.906	0.898
0	50	0.943	0.934	0.947	0.950	0.953	0.949	0.942	0.931	0.885	0.899	0.910	0.921
0	100	0.950	0.952	0.950	0.955	0.942	0.939	0.941	0.949	0.886	0.898	0.910	0.896
0	200	0.947	0.949	0.956	0.941	0.942	0.941	0.937	0.951	0.904	0.909	0.903	0.916
0.25	30	0.940	0.939	0.944	0.942	0.908	0.928	0.930	0.927	0.894	0.895	0.891	0.891
0.25	50	0.947	0.953	0.951	0.951	0.934	0.942	0.931	0.929	0.899	0.902	0.912	0.909
0.25	100	0.947	0.956	0.961	0.956	0.950	0.932	0.945	0.951	0.890	0.910	0.911	0.887
0.25	200	0.945	0.945	0.953	0.948	0.944	0.938	0.948	0.944	0.895	0.903	0.913	0.895
0.33	30	0.946	0.952	0.947	0.953	0.938	0.933	0.903	0.921	0.904	0.874	0.880	0.893
0.33	50	0.940	0.940	0.954	0.940	0.933	0.937	0.942	0.933	0.891	0.882	0.895	0.895
0.33	100	0.949	0.967	0.959	0.958	0.942	0.945	0.933	0.925	0.905	0.892	0.897	0.881
0.33	200	0.957	0.944	0.948	0.942	0.946	0.953	0.940	0.945	0.903	0.905	0.888	0.913
0.5	30	0.960	0.961	0.949	0.965	0.908	0.897	0.893	0.888	0.829	0.839	0.842	0.818
0.5	50	0.959	0.954	0.949	0.968	0.910	0.904	0.902	0.869	0.861	0.856	0.824	0.854
0.5	100	0.942	0.945	0.952	0.969	0.936	0.925	0.911	0.915	0.864	0.870	0.869	0.840
0.5	200	0.950	0.951	0.940	0.968	0.948	0.922	0.923	0.906	0.901	0.857	0.850	0.845

The table reports the Root Mean Squared Error for the standard and Factor-IV estimators and the Estimated Coverage Probability for the Factor-IV Estimator. The Monte Carlo design is as in (32)-(34), with r=1, k=1.

Table 2. RMSE Results for Factor and Standard GMM when errors are autocorrelated

Standard IV													
	θ	0				0.25				0.5			
p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
0	30	0.375	0.367	0.372	0.371	0.605	0.598	0.600	0.584	0.665	0.680	0.668	0.666
0	50	0.281	0.369	0.367	0.363	0.586	0.628	0.635	0.631	0.686	0.684	0.681	0.690
0	100	0.189	0.246	0.355	0.360	0.532	0.602	0.640	0.651	0.696	0.697	0.678	0.691
0	200	0.115	0.153	0.241	0.357	0.483	0.560	0.625	0.650	0.706	0.700	0.692	0.702
0.25	30	0.373	0.374	0.374	0.367	0.607	0.608	0.601	0.597	0.673	0.682	0.686	0.686
0.25	50	0.299	0.367	0.366	0.366	0.581	0.619	0.616	0.621	0.688	0.700	0.700	0.678
0.25	100	0.197	0.259	0.361	0.360	0.564	0.598	0.642	0.650	0.695	0.691	0.695	0.695
0.25	200	0.121	0.164	0.241	0.352	0.510	0.568	0.614	0.671	0.687	0.688	0.706	0.696
0.33	30	0.382	0.366	0.372	0.374	0.605	0.611	0.591	0.583	0.677	0.681	0.682	0.663
0.33	50	0.305	0.362	0.366	0.366	0.609	0.618	0.612	0.607	0.678	0.694	0.688	0.677
0.33	100	0.213	0.268	0.353	0.356	0.564	0.593	0.643	0.647	0.709	0.709	0.692	0.690
0.33	200	0.131	0.172	0.251	0.357	0.531	0.569	0.630	0.665	0.690	0.692	0.706	0.709
0.5	30	0.377	0.373	0.376	0.362	0.612	0.604	0.588	0.605	0.669	0.673	0.678	0.673
0.5	50	0.325	0.361	0.369	0.360	0.596	0.615	0.602	0.633	0.695	0.692	0.694	0.689
0.5	100	0.247	0.295	0.360	0.360	0.585	0.639	0.647	0.637	0.697	0.701	0.705	0.704
0.5	200	0.165	0.210	0.287	0.357	0.565	0.607	0.627	0.650	0.702	0.707	0.685	0.708
Factor IV													
0	30	0.249	0.238	0.236	0.238	1.140	1.116	1.079	0.891	2.062	2.126	1.861	1.998
0	50	0.187	0.170	0.177	0.173	0.735	0.817	0.937	0.894	1.999	2.102	2.071	2.034
0	100	0.118	0.123	0.122	0.117	0.496	0.737	0.546	0.544	2.124	2.032	1.985	2.141
0	200	0.085	0.083	0.083	0.081	0.425	0.443	0.451	0.391	2.132	1.946	2.044	2.120
0.25	30	0.325	0.300	0.283	0.277	1.259	0.916	1.153	1.183	1.970	2.020	1.917	1.873
0.25	50	0.196	0.180	0.190	0.181	0.901	0.912	1.057	0.833	2.137	1.971	2.014	2.112
0.25	100	0.134	0.128	0.127	0.126	0.815	0.698	0.606	0.636	2.165	2.198	2.179	2.148
0.25	200	0.088	0.089	0.091	0.083	0.515	0.524	0.413	0.400	2.088	1.967	2.155	2.096
0.33	30	0.409	0.442	0.484	0.548	1.445	1.375	1.470	1.292	2.151	1.926	2.113	1.978
0.33	50	0.223	0.207	0.213	0.251	1.347	1.052	1.049	1.072	2.389	2.002	2.167	2.252
0.33	100	0.145	0.140	0.136	0.136	0.785	0.862	0.731	0.832	2.337	2.142	1.961	2.109
0.33	200	0.095	0.098	0.094	0.091	0.524	0.559	0.568	0.491	2.221	2.312	2.029	2.056
0.5	30	1.398	1.541	1.862	1.825	1.926	2.146	2.162	2.201	2.040	2.222	2.081	2.237
0.5	50	0.936	1.155	1.630	2.002	1.822	1.754	2.040	2.096	2.296	2.265	2.249	2.175
0.5	100	0.458	0.704	1.098	1.553	1.457	1.660	1.875	2.128	2.398	2.394	2.423	2.264
0.5	200	0.152	0.159	0.431	1.027	0.943	1.222	1.539	1.968	2.120	2.340	2.317	2.483

The table reports the Root Mean Squared error for the standard and Factor-GMM estimators. The Monte Carlo design is as in (32)-(34), with r=1, k=1, and $\epsilon_t$in (32) is AR(1) with AR coefficient equal to 0.5.

Table 3. MSE and coverage of Factor- IV estimator with variable preselection													
		MSE Results								Coverage Probability			
		Factor IV				Standard IV				Factor IV			
p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
0	30	0.216	0.211	0.225	0.209	0.368	0.379	0.372	0.376	0.932	0.932	0.937	0.955
0	50	0.158	0.157	0.155	0.160	0.282	0.362	0.361	0.368	0.947	0.959	0.958	0.942
0	100	0.102	0.105	0.105	0.104	0.179	0.238	0.358	0.355	0.941	0.953	0.956	0.959
0	400	0.070	0.072	0.071	0.073	0.106	0.145	0.235	0.351	0.946	0.947	0.960	0.945
0.25	30	0.209	0.209	0.224	0.227	0.358	0.363	0.371	0.370	0.941	0.944	0.938	0.936
0.25	50	0.151	0.156	0.155	0.157	0.271	0.360	0.369	0.363	0.953	0.929	0.947	0.952
0.25	100	0.103	0.107	0.105	0.108	0.169	0.241	0.358	0.361	0.956	0.937	0.935	0.954
0.25	400	0.070	0.072	0.071	0.072	0.101	0.144	0.238	0.357	0.957	0.947	0.964	0.943
0.33	30	0.218	0.215	0.217	0.210	0.361	0.357	0.371	0.368	0.943	0.934	0.943	0.937
0.33	50	0.142	0.152	0.160	0.147	0.258	0.350	0.358	0.366	0.937	0.959	0.953	0.953
0.33	100	0.101	0.096	0.102	0.103	0.164	0.231	0.345	0.353	0.945	0.961	0.955	0.960
0.33	400	0.068	0.073	0.072	0.068	0.096	0.142	0.227	0.351	0.953	0.936	0.954	0.953
0.5	30	0.394	0.456	0.435	0.545	0.478	0.495	0.517	0.535	0.936	0.932	0.936	0.944
0.5	50	0.223	0.251	0.276	0.348	0.397	0.488	0.496	0.537	0.944	0.954	0.944	0.935
0.5	100	0.149	0.157	0.179	0.203	0.263	0.374	0.517	0.529	0.951	0.964	0.941	0.946
0.5	400	0.103	0.109	0.115	0.131	0.170	0.252	0.385	0.527	0.937	0.946	0.950	0.950

The table reports the MSE and coverage probability of the Factor-IV estimator. The Monte Carlo design is as in (32)-(34), with r=1, k=1, and we preselect the 50% of the instruments with the highest correlation with x_t

Table 4. RMSE Results for Factor-IV with 2 factors ($\theta = 0$)									
		Standard IV				Factor IV			
0	30	0.220	0.966	1.010	1.064	0.134	0.142	0.125	0.131
0	50	0.159	0.214	0.845	1.327	0.100	0.094	0.095	0.089
0	100	0.096	0.133	0.211	1.171	0.068	0.065	0.066	0.064
0	200	0.059	0.076	0.123	0.206	0.046	0.046	0.047	0.047

The table reports the Root Mean Squared Error for the standard and Factor-IV estimators. The Monte Carlo design is as in (32), (36) and (34), with r=2, k=1.

Table 5. RMSE Results for CSA IV													
	θ	0				0.25				0.5			
p	T/N	30	50	100	200	30	50	100	200				
0	30	0.221	0.215	0.204	0.231	1.110	0.981	0.957	0.822	1.848	1.874	2.008	1.729
0	50	0.167	0.162	0.157	0.160	0.796	0.752	0.785	0.778	1.766	1.758	1.879	1.767
0	100	0.108	0.105	0.107	0.102	0.735	0.559	0.552	0.472	1.724	1.815	1.963	1.954
0	200	0.075	0.072	0.071	0.075	0.327	0.418	0.448	0.393	1.956	1.820	1.926	1.855
0.25	30	0.270	0.262	0.229	0.218	1.103	1.044	0.905	1.023	1.895	1.873	1.651	1.685
0.25	50	0.173	0.175	0.158	0.165	0.844	0.824	0.792	0.870	1.894	1.791	1.888	1.746
0.25	100	0.117	0.111	0.110	0.103	0.769	0.636	0.590	0.479	1.732	1.739	1.800	1.736
0.25	200	0.080	0.078	0.075	0.076	0.416	0.442	0.463	0.386	1.869	1.906	1.848	1.695
0.33	30	0.304	0.289	0.244	0.233	1.121	0.978	1.205	1.137	1.788	1.972	1.969	1.976
0.33	50	0.178	0.198	0.182	0.166	0.880	0.831	0.796	0.691	1.987	1.769	1.842	2.043
0.33	100	0.123	0.115	0.113	0.112	0.599	0.708	0.593	0.596	1.998	1.937	1.895	1.685
0.33	200	0.084	0.082	0.081	0.079	0.518	0.515	0.516	0.386	1.967	1.977	1.904	1.832
0.5	30	0.476	0.547	0.463	0.440	1.522	1.492	1.583	1.490	2.057	2.191	2.101	2.078
0.5	50	0.229	0.278	0.270	0.339	1.233	1.311	1.472	1.189	1.889	2.228	1.968	1.915
0.5	100	0.156	0.163	0.161	0.159	1.042	0.937	1.002	1.176	2.173	2.101	1.965	2.084
0.5	200	0.102	0.101	0.109	0.104	0.805	0.772	0.778	0.707	1.886	1.880	1.835	2.092

The table reports the Root Mean Squared Error for the CSA-IV estimator of Kapetanios and Marcellino (2008). The Monte Carlo design is as in (32)-(34), with r=1, k=1.

Table 6. Results on factor number selection using modified information criterion								
	$p=0$				$p=0.25$			
T/N	30	50	100	200	30	50	100	200
30	1.000	1.000	1.000	1.000	0.852	0.302	0.008	0.000
50	1.000	1.000	1.000	1.000	0.699	0.255	0.000	0.000
100	1.000	1.000	1.000	1.000	0.471	0.041	0.000	0.000
200	1.000	1.000	1.000	1.000	0.262	0.000	0.000	0.000
	$p=0.5$				$p=0.75$			
30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The table reports the average number of factors over Monte Carlo replications, using the modified information criterion analysed in Theorem 6. The Monte Carlo design is as in (32)-(34), with r=1, k=1.

Table 7. Results for Taylor rule											
								1st stage (π_{1+12})		**1st stage ($y_t - y_t^*$)**	
		ρ	γ	β	$\overline{\mathbf{R}}^2$	**S.E.**	**P. J**	$\overline{\mathbf{R}}^2$	**P. F**	$\overline{\mathbf{R}}^2$	**P. F**
Base		0.883	0.993	2.310	0.98	0.27	0.11	0.12		0.86	
	st. err	0.037	0.241	0.278							
Factors		0.908	1.261	2.905	0.98	0.27	0.13	0.15	0.05	0.97	0.00
All data	st. err	0.024	0.291	0.394							
Average		0.901	1.102	2.206	0.98	0.24	0.47	0.20	0.47	0.87	0.00
All data	st. err	0.018	0.189	0.228							
Factors		0.884	1.122	2.251	0.98	0.27	0.52	0.15	0.08	0.85	0.99
Select data	st. err	0.028	0.233	0.204							
Average		0.877	1.086	2.353	0.98	0.28	0.50	0.15	0.10	0.85	0.98
Select data	st. err	0.030	0.227	0.216							
All		0.940	1.723	2.953	0.99	0.23	0.18	0.14	0.14	0.86	0.30
Select data	st. err	0.019	0.412	0.479							

Notes: The estimated equation is $r_t = \alpha + (1-\rho)\beta(\pi_{t+12} - \pi_t^*) + (1-\rho)\gamma(y_t - y_t^*) + \rho r_{t-1} + \epsilon_t$ (see text for details). The parameters are estimated by GMM over 1986.01-2003.12. In the base case (no factors) the set of instruments used includes lags of the output gap, unemployment, inflation, interest rate and commodity price index. In the Factors cases, the SW factors are added to the instruments. In particular, in "All data" the (8) factors are extracted from the whole dataset; in "Select data" the (1) factor extracted from.a subset of the variables selected with the Boivin and Ng (2006) criterion. The number of factors is based on the Bai and Ng (2002) criteria for "All data", while it is set to one for "Select data". We use one lag of each factor, but 12 lags for the "Select data" factor. In the Average cases, the instruments are one to 12 lags of the simple average of the standardized variables in "All data" or in "Select data". In the "All select data" case, the instruments are one lag of all the variables selected with the Boivin and Ng (2006) criterion. The last four columns contain statistics related to the first-stage regression of the one-year ahead expected inflation or the gap on the set of instruments used. In particular, we report the adjusted $R^2$and the p-value of an F-test for the joint significance of the coefficients on factors, when factors are added to the baseline model.

Table 8. Results for New Keynesian Phillips curve

								1st stage (π_{l+1})		1st stage (ur_t)	
		α	γ	ρ	$\overline{\mathbf{R^2}}$	**S.E.**	**P. J**	$\overline{\mathbf{R^2}}$	**P. F**	$\overline{\mathbf{R^2}}$	**P. F**
Base		-0.002	0.538	0.462	0.98	0.16	0.62	0.12		0.98	
	st. err	0.007	0.048	0.047							
Factors		-0.000	0.513	0.492	0.98	0.16	0.30	0.11	0.48	0.98	0.23
All data	st. err	0.006	0.038	0.038							
Average		-0.002	0.473	0.532	0.98	0.16	0.29	0.12	0.37	0.98	0.01
All data	st. err	0.006	0.030	0.029							
Factors		-0.002	0.500	0.509	0.98	0.15	0.12	0.23	0.00	0.98	0.50
Select data	st. err	0.006	0.021	0.020							
Average		-0.002	0.501	0.509	0.98	0.16	0.10	0.16	0.03	0.98	0.85
Select data	st. err	0.006	0.021	0.020							
All		-0.000	0.551	0.459	0.98	0.16	0.27	0.11	0.64	0.98	0.21
Select data	st. err	0.006	0.043	0.042							

Notes: The estimated equation is $\pi_t = c + \alpha(ur_t) + \gamma(\pi_{t+1}) + \rho\pi_{t-1} + \epsilon_t$ (see text for details). The parameters are estimated by GMM over 1986.01-2003.12. In the base case (no factors) the set of instruments used includes lags of the output gap, unemployment, inflation, interest rate and commodity price index. In the Factors cases, the SW factors are added to the instruments. In particular, in ”All data” the 8 factors are extracted from the whole dataset; in ”Select data” the (1) factor extracted from a subset of the variables selected with the Boivin and Ng (2006) criterion. The number of factors is based on the Bai and Ng (2002) criteria for ”All data”, while it is set to one for ”Select data”. We use one lag of each factor, but 12 lags for the ”Select data” factor. In the Average cases, the instruments are one to 12 lags of the simple average of the standardized variables in ”All data” or in ”Select data”. In the ”All select data” case, the instruments are one lag of all the variables selected with the Boivin and Ng (2006) criterion. The last four columns contain statistics related to the first-stage regression of the one-year ahead expected inflation or the unemployment rate on the set of instruments used. In particular, we report the adjusted $R^2$and the p-value of an F-test for the joint significance of the coefficients on factors, when factors are added to the baseline model.