EFWW
7/10/08
9:45 a.m.

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

George Kapetanios†
Queen Mary University of London

Massimiliano Marcellino‡
IEP-Bocconi University, IGIER and CEPR

First Version: October 2006
This Version: March 2008

Abstract

This paper analyses the use of factor analysis for instrumental variable estimation when the number of instruments tends to infinity. We consider cases where the unobserved factors are the optimal instruments but also cases where the factors are not necessarily the optimal instruments but can provide a summary of a large set of instruments. Further, the situations where many weak instruments exist and/or the factor structure is weak are also considered. 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

*We are grateful to Jushan Bai, Domenico Giannone, Marco Lippi, Helmut Lutkepohl, Hashem Pesaran, Jim Stock and participants at conferences at the European University Institute and the Bank of England 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

‡IGIER - Università Bocconi, Via Salasco 5, 20136, Milano, Italy. Phone: +39-02-5836-3327. Fax: +39-02-5836-3302. E-mail: massimiliano.marcellino@uni-bocconi.it

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. Factor analysis has been used fruitfully to model, for example, asset returns, macroeconomic aggregates and Engel curves (see, e.g., Chamberlain and Rothschild (1983), Stock and Watson (1989), Lewbel (1991)).

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 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 (2006b) 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), Bernanke, Boivin, and Eliasz (2005) and Kapetanios and Marcellino (2006a) 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 (2005) 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 (2006b), 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 (2006b) 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 (2006b) and our paper independently analyse the properties of Factor-GMM estimation in this context.

However, assuming that the endogenous regressors are only functions of the factors can be a restrictive assumption. We therefore, generalise our and the analysis of Bai and Ng (2006b) to the case where regressors are either only functions of the observed instruments or both of the observed instruments and the unobserved factors. In these two more general cases the superiority of a Factor-GMM estimator is not obvious. In fact we show that, when regressors are only functions of a large but finite set of observed instruments, standard GMM estimation is preferable to using the factors as instruments, even when the factors are known. In the more general case where the endogenous variables depend both on a small set of key observed instruments and on the factors, the ranking of Factor-GMM and standard GMM estimation depends on the parameter values.

Our analysis of Factor-GMM also evaluates another 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), Hausman, Newey, 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 (2006) 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 fourth 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 fifth 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, the regressors depend on either the unobserved factors or the observed instruments or both, and the factor structure is strong or weak. The results are in line with the theoretical findings and clearly indicate the superiority of Factor-GMM over standard GMM estimation also in finite samples.

Our sixth and final contribution is an empirical analysis where Factor-GMM is applied to estimate the parameters of forward looking equations for macroeconomic variables, specifically, Taylor rules and New Keynesian Phillips curves. The results show that the new estimation method is also easily implemented and empirically relevant.

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 examples. 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' are possibly correlated with the error term ϵ_t.[1] 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)

[1] In practice, only a subset of the k regressors could be correlated with the error term, but for the sake of simplicity we will assume that all of them are endogenous.

or Galí and Gertler (1999)).

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

$$z_t = \Lambda^{0'} f_t + v_t, \tag{2}$$

where r is much smaller than N. Therefore, each instrumental variable can be decomposed into a common component (an element of $\Lambda^{0'} f_t$) 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 z_t can be efficiently summarized by the r factors f_t.

We consider three different data generation mechanisms for x_t that allow for non-zero correlation between x_t and ϵ_t, and a different degree of efficiency of the instruments z_t and of the factors f_t. They are given by

$$x_t = A_Z^{0'} z_t + u_t, \tag{3}$$

$$x_t = A^{0'} f_t + u_t, \tag{4}$$

and

$$x_t = A_Z^{0'} z_t + A^{0'} f_t + u_t, \tag{5}$$

with $E(u_t' \epsilon_t) \neq 0$ to introduce simultaneity in (1). In (3) the endogenous variables depend directly on the instruments. Therefore, the optimal instruments in this case are z_t rather than f_t and, conditional on z_t, the factors are irrelevant. However, when N is very large, possibly larger than T, the properties of the standard IV estimator are in general unknown, but it can be expected to become inefficient and in some cases even inconsistent, as discussed by Bekker (1994) and Chao and Swanson (2005). In this context, the factors could become useful again, since they provide a concise summary of the information in z_t.

In (4), the endogenous variables depend directly on the factors. This is the case also considered by Bai and Ng (2006b), and it represents the most favourable situation for Factor-IV estimation, since the original instruments z_t become irrelevant, conditional on the factors.[2]

[2]More specifically, Bai and Ng (2006b) assume that only a subset of the regressors in (1) are endogenous, say x_{2t}. The x_{2t} variables depend on a few factors, say f_t, which are a subset of the set of factors driving the N exogenous variables z_t, say F_t. They also discuss procedures for selecting f_t from F_t. Moreover, Bai and Ng (2006b) show that in this case the factors can be valid instruments also when the z variables are correlated with the error term in the structural equation.

In (5), the endogenous variables depend both on (possibly a subset of) the instrumental variables and on (possibly a subset of) the factors. This 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 (2006)). A similar reasoning holds for unobservable variables, such as the output gap. Unfortunately, we will see that under (5) it is not possible to provide a unique ranking of the standard IV and of the Factor-IV estimators, since the ranking depends on the loading matrices in (5).[3]

Stacking observations across time for all models presented above gives:

$$y = X\beta + \epsilon \tag{6}$$

$$Z = F\Lambda^0 + v \tag{7}$$

$$X = ZA_Z^0 + u \tag{8}$$

$$X = FA^0 + u \tag{9}$$

$$X = ZA_Z^0 + FA^0 + u \tag{10}$$

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

Let $\hat{F}$ denote the Stock and Watson (2002b) principal component estimator of F. We consider two alternative two stages least squares estimators for the parameters of interest, β. The Factor-IV estimator is:

$$\hat{\beta} = \left(X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'X\right)^{-1} X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'y, \tag{11}$$

while the standard IV estimator is

$$\tilde{\beta} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'y. \tag{12}$$

We also define the infeasible factor estimator given by

$$\bar{\beta} = \left(X'F(F'F)^{-1}F'X\right)^{-1} X'F(F'F)^{-1}F'y. \tag{13}$$

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

[3] We do not consider the analysis of IV estimators when N diverges, but focus on the asymptotic properties of Factor-IV estimators.

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* $||\Lambda^0 \Lambda^0/N - D|| \rightarrow 0$ *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 (x_t', z_t') *are jointly stationary.* z_t *is predetermined, so that* $E(z_{it}\epsilon_t) = 0$, $i = 1, ..., N$. *The probability limit of* $\frac{z_t z_t'}{T}$ *is finite and nonsingular.* $E(z_t x_t')$ *has full column rank* k. x_t *and* z_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 (2006a) to show consistency of the parameter estimators in factor augmented regressions. Assumption 3 guarantees that standard IV estimation using z_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 alternative IV estimators for the case where N is finite, and shows that, as expected, their relative efficiency depends on the generating mechanism of the data.

Theorem 1 *Assuming that* f_t *is observed and* N *is finite,* $\sqrt{T}(\bar{\beta} - \beta)$ *has an asymptotically Normal distribution, with zero mean and variance covariance matrix under (3)-(5) given, up* *to the same scalar constant of proportionality, by*

$$Avar\left(\sqrt{T}(\bar{\beta}-\beta)\right) = \left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 A_Z^0\right)^{-1} \tag{14}$$

$$Avar\left(\sqrt{T}(\bar{\beta}-\beta)\right) = \left(A^{0\prime}\Sigma_f A^0\right)^{-1} \tag{15}$$

and

$$Avar\left(\sqrt{T}(\bar{\beta}-\beta)\right) = \left(\left(A_Z^{0\prime}\Lambda^{0\prime} + A^{0\prime}\right)\Sigma_f\left(\Lambda^0 A_Z^0 + A^0\right)\right)^{-1} \tag{16}$$

respectively. For the standard IV estimator, $\sqrt{T}(\tilde{\beta}-\beta)$ has an asymptotically Normal distribution, with zero mean and variance covariance matrix under (3)-(5) given, up to the same scalar constant of proportionality, by

$$Avar\left(\sqrt{T}(\tilde{\beta}-\beta)\right) = \left(A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right)A_Z^0\right)^{-1} \tag{17}$$

$$Avar\left(\sqrt{T}(\tilde{\beta}-\beta)\right) = \left(A^{0\prime}\Sigma_f\Lambda^{0\prime}\left(\Lambda^0\Sigma_f\Lambda^{0\prime} + \Sigma_v\right)^{-1}\Lambda^0\Sigma_f A^0\right)^{-1} \tag{18}$$

and

$$Avar\left(\sqrt{T}(\tilde{\beta}-\beta)\right) = \left(\left(A^{0\prime}\left(\Sigma_f\Lambda^0 + \Sigma_v\right) + A^{0\prime}\Sigma_f\Lambda^{0\prime}\right)\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right)^{-1}\right. \tag{19}$$

$$\left.\left(A^{0\prime}\left(\Sigma_f\Lambda^0 + \Sigma_v\right) + A^{0\prime}\Sigma_f\Lambda^{0\prime}\right)'\right)^{-1}$$

respectively. The difference between the RHS of (14) and (17) is a positive semidefinite matrix. The difference between the RHS of (18) and (15) is a positive semidefinite matrix.

The next result provides asymptotic equivalence between the feasible and infeasible Factor-IV estimators in the case of diverging N.

Theorem 2 *If $\sqrt{T}/N = o(1)$ then*

$$\sqrt{T}(\bar{\beta}-\beta) - \sqrt{T}(\hat{\beta}-\beta) = o_p(1) \tag{20}$$

Comparable results on the performance of the standard IV estimator when the number of instruments tends to infinity as the sample size grows 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 and 3. Actually, notice that the condition $\sqrt{T}/N = o(1)$ in Theorem 2 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 equations (3)-(5) with

$$x_t = A^{0\prime}_{(N,T)Z} z_t + u_t, \tag{21}$$

$$x_t = A^{0\prime}_{(T)} f_t + u_t, \tag{22}$$

and

$$x_t = A^{0\prime}_{(N,T)Z} z_t + A^{0\prime}_{(T)} f_t + u_t. \tag{23}$$

Formally, the instruments are referred to as weak when $A^{0\prime}_{(N,T)Z} Z'Z A^{0}_{(N,T)Z}$ 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.

In our context, we have to distinguish two cases. When the data are generated according to (21), the explanatory power of z_t can decrease because the matrix of coefficients A vanishes when T increases, when N increases, or both. When instead the data are generated according to (22), increasing N has no effects and the instruments can be weak only if the matrix of coefficients A vanishes when T increases.

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 3 *Let one of (21), (22) or (23) hold. Let $N = O(T^{\gamma})$, $\gamma > 1/2$. Further, let every element of $\Lambda^0 A^0_{(N,T)Z}$ be $O(N^{-\beta}) = O(T^{-\psi})$, $\beta \geq 0$, where $\psi = \gamma\beta$, $0 \leq \psi < 1/2$. Also let every element of $A^0_{(T)}$ be $O(T^{-\vartheta})$, $0 \leq \vartheta < 1/2$. Then, under (21),*

$$T^{1/2-\psi}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_{\epsilon}\left(\Upsilon'\Sigma_f\Upsilon\right)^{-1}\right) \tag{24}$$

Under (22)

$$T^{1/2-\psi}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_{\epsilon}\left(\Psi'\Sigma_f\Psi\right)^{-1}\right) \tag{25}$$

and under (23), if $\vartheta < \psi$

$$T^{1/2-\vartheta}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2_{\epsilon}\left(\Psi'\Sigma_f\Psi\right)^{-1}\right) \tag{26}$$

if $\vartheta > \psi$,

$$T^{1/2-\psi}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma_\epsilon^2 \left(\Upsilon' \Sigma_f \Upsilon\right)^{-1}\right) \tag{27}$$

and if $\vartheta = \psi$,

$$T^{1/2-\psi}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma_\epsilon^2 \left(\left(\Upsilon+\Psi\right)' \Sigma_f \left(\Upsilon+\Psi\right)\right)^{-1}\right) \tag{28}$$

where

$$\lim_{T\to\infty} \frac{\Lambda^0 A^0_{(N,T)Z}}{T^{-\psi}} = \Upsilon \tag{29}$$

$$\lim_{T\to\infty} \frac{A^0_{(T)}}{T^{-\vartheta}} = \Psi \tag{30}$$

and Υ *and* Ψ *are nonsingular matrices.*

It is worth making a few comments on the rates in the assumptions of Theorem 3. 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).

In the requirement that every element of $\Lambda^0 A^0_{(N,T)Z}$ is $O(N^{-\beta}) = O(T^{-\psi})$, $\beta \geq 0$, $\psi = \gamma\beta$, $0 \leq \psi < 1/2$, the parameters β and ψ control how fast the instruments become weak when, respectively, N and T increase (the larger β and ψ, the faster the instruments become weak, the slower the speed of convergence of the Factor IV estimator). If β and ψ are too large, the Factor IV estimator is no longer consistent. The condition $\psi = \gamma\beta$ allows for a larger value of ψ for a given β when γ is large, since in this case the rate of increase of T with respect to N is slow. The condition that every element of $A^0_{(T)}$ be $O(T^{-\vartheta})$, $0 \leq \vartheta < 1/2$ has a similar interpretation, but in this case it is necessary to only control the temporal dimension since the longitudinal dimension has no effects on the weakness of the factors as instruments.

When the data are generated according to equation (23), 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, and therefore the asymptotic distribution of the Factor IV estimator is the same as when the data are generated according to (3).

Similar to the previous case, in the presence of a diverging number of both strong and weak instruments of the same type, e.g. factors, the former dominate and the asymptotic distribution of the Factor IV estimator is as in the previous Section.

The assumption that the elements of $A^0_{(N,T)Z}$ 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 (29) and (30) would be modified to ones involving stochastic convergence.

Notice also that 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 $T^{1/2}$.

Finally, the results in Theorem 3 can be also use 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.

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

$$z_{it} = \frac{\lambda_i^{0\prime}}{i^{\alpha_i}} f_t + e_{it}, \tag{31}$$

$$\alpha_i = \begin{cases} 0 \text{ for } i < N^* \\ \alpha \text{ for } i \geq N^* \end{cases},$$

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

$$z_t = \Lambda_N^{0\prime} f_t + e_t, \tag{32}$$

$$\Lambda_N^{0\prime} = \left[\frac{\lambda_1^0}{1}, \frac{\lambda_2^0}{2^{\alpha_2}}, ..., \frac{\lambda_N^0}{N^{\alpha_N}} \right]', \tag{33}$$

the loading matrix Λ_N^0 does not necessarily satisfy the condition in Assumption 1 (1). Therefore, this model can be not 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 $\sum_{i=1}^{N} \frac{\lambda_i^{0'}}{i^{\alpha_i}} \leq c_1 N^* + c_2 N^{1-\alpha} = o(N)$ 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

$$z_t = \Lambda_N^{0'} f_t + e_t, \tag{34}$$

$$\Lambda_N^{0'} = \left[\frac{\lambda_1^0}{N^\alpha}, \frac{\lambda_2^0}{N^\alpha}, ..., \frac{\lambda_N^0}{N^\alpha} \right]' = \frac{\Lambda^{0'}}{N^\alpha}, \tag{35}$$

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 (34), the same estimators will remain consistent in (31). Actually, for all $i = [b_i N]$, where $[.]$ denotes integer part and $0 < b_i \leq 1$, (34) is equivalent to (31) since $\frac{\lambda_i^{0'}}{i^\alpha} = \frac{\lambda_i^{0'}}{b_i^\alpha N^\alpha} = \frac{\tilde{\lambda}_i^{0'}}{N^\alpha}$ where $\tilde{\lambda}_i^{0'} = \frac{\lambda_i^{0'}}{b_i^\alpha}$.

Further justification can be provided for the setup in (34). 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 (34) from (31) 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 (31) on the ordering of the variables. Moreover, (34) 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 (34) instead of (31).

Notice that (34) also implies that if (3) holds, i.e. the endogenous variables depend on the z variables, then the true (or, by extension, estimated) factors are weak instruments, while the z variables are strong instruments. Further, if (4) holds, i.e. the endogenous variables depend on the factors, then the observed instruments, z_t, are weak instruments since their relationship with the factors vanishes when N increases. The true factors are strong instruments, but there is an issue whether the principal component based estimator of the factors remains consistent in this context, and therefore whether the estimated factors are strong instruments or not.

The first result we present analyses conditions under which a local-to-zero factor loading matrix in (34) 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 4 *Let $\Lambda_N^0 = \Lambda^0/N^\alpha$ as in (35). 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 5 *Let $\Lambda_N^0 = \Lambda^0/N^\alpha$ as in (35), where $0 \leq \alpha < 1/4$. Then, $\hat{F} - FH' = o_p(1)$ for some nonsingular matrix H, as long as $N = o(T^{1/4\alpha})$. Further,*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_t - Hf_t\right\|^2 = O_p\left(\min\left(N^{-4\alpha}T, N^{1-4\alpha}\right)^{-1}\right) \tag{36}$$

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 (35). Let* $\alpha < 1/4$ *and* $N^{2\alpha} = o(T^{1/2})$ *. Then,*

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = O_p\left(C_{NT}^{-1}\right) \tag{37}$$

where C_{NT} *is defined in (129), as long as* q_t *has finite fourth moments, nonsingular covariance matrix and satisfies a central limit theorem.*

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

Theorem 6 *Let* $N = O(T^\gamma)$*. Let* $\Lambda_N^0 = \Lambda^0/N^\alpha$ *as in (35). 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 (129). Then, Theorem 2 follows. Further, let every element of* $\Lambda_N^0 A_{(N,T)Z}^0$ *be* $O(N^{-\beta}) = O(T^{-\psi})$, $\beta \geq 0$, *where* $\psi = \gamma\beta$, $0 \leq \psi < 1/2$*. Also let every element of* $A_{(T)}^0$ *be* $O(T^{-\vartheta})$, $0 \leq \vartheta < 1/2$*. Then, (24)-(28) of Theorem 3 follow, where*

$$\lim_{T\to\infty}\frac{\Lambda_{N(T)}^0 A_{(N,T)Z}^0}{T^{-\psi}} = \Upsilon \tag{38}$$

$$\lim_{T\to\infty}\frac{A_{(T)}^0}{T^{-\vartheta}} = \Psi \tag{39}$$

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 2 for asymptotic equivalence on the infeasible and feasible Factor-IV estimators. The other requirements are similar to those in Theorem 3, 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

$$V_r + rg(N,T) \tag{40}$$

where

$$V_r = (NT)^{-1}\sum_{i=1}^{N}\sum_{i=1}^{T}\left(z_{i,t} - \sum_{j=1}^{r}\hat{\lambda}_{j,i}\hat{f}_{j,t}\right)^2 \tag{41}$$

$\hat{\lambda}_{j,i}$ 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 7 *Let the number of factors be determined by minimising (40) over* $1 \leq r \leq r_{\max}$ *for some constant* $r_{\max}$. *Let*

$$g(N,T) = \ln\left(\min(N,T)\right)^{-1} \tag{42}$$

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 7 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.

In summary, in this Section we have shown that for some data generating processes, using z_t rather than f_t is preferable in the case of finite N, as detailed in Theorem 1. However, this result is reversed when N tends to infinity. First, as $N \to \infty$, it becomes feasible to estimate the unobserved factors f_t consistently even for local-to-zero factor models, as shown in Theorem 5, and estimation of the factors does not to matter for the asymptotic properties of the Factor-IV estimators, as shown in Theorems 2 and 6. Moreover, whereas IV estimation based on the estimated factors remains consistent and asymptotically normal even in the case where z_t are weak instruments, as shown in Theorems 3 and 6, standard IV estimation can be inconsistent if the number of instrument increases fast enough, as discussed by Bekker (1994) and Chao and Swanson (2005).

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* $z_t\epsilon_t$ *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. The probability limits of* $\frac{F'\epsilon\epsilon' F}{T}$ *and* $\frac{Z'\epsilon\epsilon' Z}{T}$, *denoted by* $S_{f\epsilon}$ *and* $S_{z\epsilon}$ *exist and are nonsingular.*

We further add the following regularity condition.

Assumption 5 $E[(z_{ti}x_{tj})^2]$ *exists and is finite for 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 estimators $\hat{\beta}$ and $\widetilde{\beta}$ in (11) and (12) remain consistent and asymptotically normal. Furthermore, we have:

Theorem 8 *For finite N, the asymptotic variance covariance matrix of $\sqrt{T}(\bar{\beta}-\beta)$ under (3)-(4) are given by*

$$Avar\left(\sqrt{T}(\bar{\beta}-\beta)\right) = \left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 A_Z^0\right)^{-1} A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f^{-1}S_{f\epsilon}\Sigma_f^{-1}\Lambda^0 A_Z^0 \left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 A_Z^0\right)^{-1} \quad (43)$$

$$Avar\left(\sqrt{T}(\bar{\beta}-\beta)\right) = \left(A^{0\prime}\Sigma_f A^0\right)^{-1} A^{0\prime}\Sigma_f^{-1}S_{f\epsilon}\Sigma_f^{-1}A^0 \left(A^{0\prime}\Sigma_f A^0\right)^{-1} \quad (44)$$

where $S_{f\epsilon} = E(F'\epsilon\epsilon'F)$. The asymptotic variance covariance matrix of $\sqrt{T}(\tilde{\beta}-\beta)$ under (3)-(4) are given by

$$Avar\left(\sqrt{T}(\tilde{\beta}-\beta)\right) = \left(A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0+\Sigma_v\right)A_Z^0\right)^{-1} \quad (45)$$

$$A_Z^{0\prime}S_{z\epsilon}A_Z^0\left(A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0+\Sigma_v\right)A_Z^0\right)^{-1}$$

$$Avar\left(\sqrt{T}(\tilde{\beta}-\beta)\right) = \left(A^{0\prime}\Sigma_f\Lambda^{0\prime}\left(\Lambda^0\Sigma_f\Lambda^{0\prime}+\Sigma_v\right)^{-1}\Lambda^0\Sigma_f A^0\right)^{-1} \quad (46)$$

$$\left(A^{0\prime}\Sigma_f\Lambda^{0\prime-1}S_{z\epsilon}^{-1}\Lambda^0\Sigma_f A^0\right)\left(A^{0\prime}\Sigma_f\Lambda^{0\prime}\left(\Lambda^0\Sigma_f\Lambda^{0\prime}+\Sigma_v\right)^{-1}\Lambda^0\Sigma_f A^0\right)^{-1}$$

Notice that when the errors ϵ are uncorrelated and homoskedastic, it is (up to a scalar constant) $S_{f\epsilon} = E(F'F) = \Sigma_f$ and $S_{z\epsilon} = E(Z'Z) = \Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v$. Therefore, the variance covariance matrices of $\hat{\beta}$ and $\widetilde{\beta}$ reduce to those derived in Theorem 1. In practice, $S_{f\epsilon}$ and $S_{z\epsilon}$ can be estimated by a HAC procedure, such as that developed in Newey and West (1987). For example, using a Bartlett kernel, we have

$$\widehat{S}_{z\epsilon,h} = \widehat{\Phi}_0 + \sum_{j=1}^{h}(1-\frac{j}{h+1})(\widehat{\Phi}_j + \widehat{\Phi}_j')$$

$$\widehat{\Phi}_j = T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_{t-j}z_t z_{t-j}',$$

where h is the length of the window, $\widehat{\epsilon}_t = y_t - x_t'b$, and b is a consistent estimator for β. We focus on the Newey and West (1987) HAC procedure using the Bartlett kernel in the rest of the section.

A remaining problem with the two stage least square estimators $\hat{\beta}$ and $\widetilde{\beta}$ is that they are not efficient in the presence of a general error structure. In fact, the efficient estimators in this context are obtained by GMM estimation with either $S_{f\epsilon}^{-1}$ or $S_{z\epsilon}^{-1}$ as the weighting matrix. Using standard methods, the resulting estimators are

$$\widehat{b} = \left(X'\hat{F}\widehat{S}_{\widehat{f}\epsilon}^{-1}\hat{F}'X\right)^{-1} X'\hat{F}\widehat{S}_{\widehat{f}\epsilon}^{-1}\hat{F}'y \tag{47}$$

$$\widetilde{b} = \left(X'Z\widehat{S}_{z\epsilon}^{-1}Z'X\right)^{-1} X'Z\widehat{S}_{z\epsilon}^{-1}Z'y \tag{48}$$

and

$$\bar{b} = \left(X'FS_{f\epsilon}^{-1}F'X\right)^{-1} X'FS_{f\epsilon}^{-1}F'y, \tag{49}$$

When the errors are uncorrelated and homoskedastic, these expressions simplify to those in (11)-(13).

Theorem 9 *Assuming that f_t is observed, then for finite N the asymptotic variance covariance matrix of $\sqrt{T}(\bar{b}-\beta)$ under (3)-(5) are given by*

$$var\left(\sqrt{T}(\bar{b}-\beta)\right) = \left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f S_{f\epsilon}^{-1}\Sigma_f\Lambda^0 A_Z^0\right)^{-1} \tag{50}$$

$$var\left(\sqrt{T}(\bar{b}-\beta)\right) = \left(A^{0\prime}\Sigma_f S_{f\epsilon}^{-1}\Sigma_f A^0\right)^{-1} \tag{51}$$

and

$$var\left(\sqrt{T}(\bar{b}-\beta)\right) = \left(\left(A_Z^{0\prime}\Lambda^{0\prime} + A^{0\prime}\right)\Sigma_f S_{f\epsilon}^{-1}\Sigma_f\left(\Lambda^0 A_Z^0 + A^0\right)\right)^{-1} \tag{52}$$

respectively. The asymptotic variance covariance matrix of $\sqrt{T}(\widetilde{b}-\beta)$ under (3)-(5) are given by

$$var\left(\sqrt{T}(\widetilde{b}-\beta)\right) = \left(A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right)S_{z\epsilon}^{-1}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right)A_Z^0\right)^{-1} \tag{53}$$

$$var\left(\sqrt{T}(\widetilde{b}-\beta)\right) = \left(A^{0\prime}\Sigma_f\Lambda^{0\prime}S_{z\epsilon}^{-1}\Lambda^0\Sigma_f A^0\right)^{-1} \tag{54}$$

and

$$var\left(\sqrt{T}(\widetilde{b}-\beta)\right) = \left(\left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 + A^{0\prime}\left(\Sigma_f\Lambda^0 + \Sigma_v\right)\right)S_{z\epsilon}^{-1}\right. \tag{55}$$
$$\left.\left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 + A^{0\prime}\left(\Sigma_f\Lambda^0 + \Sigma_v\right)\right)'\right)^{-1}$$

respectively.

Notice that the higher precision of $\bar{b}$ with respect to β when the model is (4) (and of $\widetilde{b}$ for (3)) follows from the choice of the weighting matrix.

Theorem 10 *If* $\sqrt{T}/N = o(1)$ *then*

$$\sqrt{T}(\bar{b} - \beta) - \sqrt{T}(\widehat{b} - \beta) = o_p(1) \tag{56}$$

Next, we have

Theorem 11 *Under the same assumptions of Theorem 3, under (3),*

$$T^{1/2-\psi}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left(\Upsilon'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f \Upsilon'\right)^{-1}\right) \tag{57}$$

Under (4)

$$T^{1/2-\psi}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left(\Psi'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f \Psi\right)^{-1}\right) \tag{58}$$

and under (5), if $\vartheta < \psi$

$$T^{1/2-\vartheta}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left(\Psi'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f \Psi\right)^{-1}\right) \tag{59}$$

if $\vartheta > \psi$,

$$T^{1/2-\psi}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left(\Upsilon'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f \Upsilon'\right)^{-1}\right) \tag{60}$$

and if $\vartheta = \psi$,

$$T^{1/2-\psi}\left(\widehat{b} - \beta\right) \xrightarrow{d} N\left(0, \left((\Psi + \Upsilon)'\Sigma_f S_{f\epsilon}^{-1}\Sigma_f (\Psi + \Upsilon)\right)^{-1}\right) \tag{61}$$

where

$$\lim_{T\to\infty} \frac{\Lambda^0 A^0_{(N,T)Z}}{T^{-\psi}} = \Upsilon \tag{62}$$

$$\lim_{T\to\infty} \frac{A^0_{(T)}}{T^{-\vartheta}} = \Psi \tag{63}$$

and Υ *and* Ψ *are nonsingular matrices.*

Finally, note that that the results of Theorem 6 on the use of factors in the presence of weak instruments and a weak factor structure follow straightforwardly also for the GMM case.[4]

[4]It would be of interest in future research to generalise the results to the case where the factors are estimated by dynamic principal components, as in Forni, Hallin, Lippi, and Reichlin (2000) and Forni, Hallin, Lippi, and Reichlin (2004). In this respect it is interesting to note the empirical results of Favero, Marcellino, and Neglia (2005), who find that static principal components work better than dynamic principal components when used as additional instruments in the estimation of forward looking Taylor rules.

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. In the first subsection, we consider all setups and estimators we have proposed in Section 2. In the second subsection, we present additional results on the role of variable preselection, on weak instruments with strong factors, and on Factor-GMM estimation.

4.1 Factor-IV estimation in finite samples

The basic setup of the Monte Carlo experiments is:

$$y_t = \sum_{i=1}^{k} x_{it} + \epsilon_t \tag{64}$$

$$z_{it} = \sum_{j=1}^{r} N^{-p} f_{jt} + c_2 e_{it}, \quad i = 1, ..., N \tag{65}$$

$$x_{it} = \sum_{j=1}^{N} N^{-q} z_{jt} + u_{it}, \quad i = 1, ..., k \tag{66}$$

$$x_{it} = \sum_{j=1}^{r} c_1^{-1} r^{-1/2} f_{jt} + u_{it}, \quad i = 1, ..., k \tag{67}$$

and

$$x_{it} = \sum_{j=1}^{N} N^{-1} z_{jt} + \sum_{j=1}^{r} c_1^{-1} r^{-1/2} f_{jt} + u_{it}, \quad i = 1, ..., k \tag{68}$$

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.[5]

[5] Results for k and r larger than one are available upon request.

The instrumental variables z_{it} are generated by the factor model in (65). The parameter p controls the "strength" of the factor structure. We consider the values $p = 0, 0.1, 0.25, 0.33, 0.45, 0.5$. 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 5 and 6. In particular, we know from Theorem 5 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 $p < 0.5$.[6] The parameter c_2 controls the relative size of the idiosyncratic component, so that a larger value of c_2 makes factor estimation harder, at least for small values of N. In the base case we set $c_2 = 1$, but we also consider experiments with $c_2 = 0.5$ or 4.

As in Section 2, we consider three different generating mechanisms for the endogenous variable x_t. First, for the setup corresponding to (3), where x_t depends directly on the instrumental variables z_{it}, we use (64), (65) and (66). In this case, $\Lambda^0 A^0_{NZ}$ is $O(N^{1-p-q})$, so that we are in the context of Theorems 3 and 6. Notice also that it is

$$var(x_t) = N^{2(1-p-q)} + N^{1-2q} c_2 + 1. \tag{69}$$

Hence, when $q = 1$, $p = 0$, the z_{it} are strong instruments, $\sum_{j=1}^{N} N^{-1} z_{jt}$ in (66) is $O_p(1)$ as $N \to \infty$. However, when $q = 1$ but $p > 0$, the contribution of z_{it} to the variance of x_{it} vanishes as $N \to \infty$, as well as the correlation between z_{it} and x_t for each i. Therefore, the instruments become weak. More precisely, when $q = 1$, $p > 0$ we have a combination of weak factor structure and weak instruments. To evaluate the case of weak instruments but strong factor structure, we will also run experiments with $q = 1.2$, $p = 0$ in the next subsection. In that case, $\sum_{j=1}^{N} N^{-q} z_{jt} = O_p(N^{1-q})$, implying weak instruments.

In this first experimental design, the standard IV estimator should perform well, at least for limited values of N, large sample size T, and $p = 0$. When N is large, possibly larger than T, the properties of the standard-IV estimator are mostly unknown (with the exception of the cases studied by Bekker (1994) and Chao and Swanson (2005)). When T is small, the estimators in the first step regression could be imprecise, due to the relatively large number

[6] From (65), it is

$$\sum_{i=1}^{N} N^{p-1} z_{it} = f_t + c_2 \sum_{i=1}^{N} N^{p-1} e_{it}$$

and $c_2 \sum_{i=1}^{N} N^{p-1} e_{it}$ goes to zero when $p < 0.5$.

of regressors. When $p > 0$, we have seen that we enter into the weak instrument situation, at least for large values of N.

The Factor-IV estimator should produce reasonable results also in this first experimental design, since the factors are a proxy for the z variables. In particular, the loss in explanatory power for the endogenous variable x when using the factors instead of the instrumental variables z is $N^{1-2q}c_2$, which vanishes quickly as N grows for $q = 1$. Moreover, the Factor-IV estimator could be even better than the standard IV estimator for large values of N, small T, and $0 < p < 0.5$, when the z are weak instruments but the factors can still be consistently estimated.

In the second setup, corresponding to (4), the endogenous variable depends on the factor. We use (64), (65) and (67), where $c_2 = 1$ and $c_1 = 0.5, 1, 4$. The parameter c_1 measures the "strength" of the factors as instruments, which decreases for higher values of c_1. In fact, it is

$$var(x_t) = c_1^{-2} + 1. \tag{70}$$

The parameter p is also important in this context, since it controls the strength of the factor model and for values of $p \geq 0.5$ it is no longer possible to obtain consistent estimators of the factors. This can be considered as a particular case of weak instruments, in the sense that the true factors are strong instruments, but the estimated factors will be in general weak instruments.

To evaluate the expected performance of the standard IV estimator in this context, let us assume that $p = 0$, $c_1 = 1$, $c_2 = 1$. In this context, the estimator of the factor is simply the average of the z variables, while the pseudo true value of the coefficient of z_{it} in a regression of x_t on z_{it}, $i = 1, ..., N$, is $1/N$. Hence, when T is large, and larger than N, the two IV estimators should produce similar results. 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. However, when p increases and the factor structure becomes weak, the ranking of the two estimators is no longer obvious.

Finally, for the setup corresponding to (5), where the endogenous variable depends both on the factors and on the instrumental variables, we use (64), (65) and (68) where $c_2 = 0.5, 1$ and $c_1 = 0.5, 1$. We have

$$var(x_t) = N^{-2p} + c_1^{-2} + 1, \tag{71}$$

so that for the base case of $p = 0$ and $c_1 = 1$ the factor and the instrumental variables have the same explanatory power. When p increases, the z variables become weak instruments, but the factor structure also becomes weak, and we wish to evaluate the relative importance of these two types of weaknesses.

In all cases $p = 0, 0.1, 0.25, 0.33, 0.45, 0.5$, and we consider the following combinations of N and T: $N = (30, 50, 100, 200)$, $T = (30, 50, 100, 200)$.[7] The biases of the alternative estimators are systematically small, detailed tables are available upon request. Therefore, we focus on the variances of the estimators under the different simulation designs.

Results make interesting reading. Looking first at the figures for the experiment using (64), (65) and (67), where the endogenous variable depends on the factor, we note that the performance of the standard IV estimator improves with T, but the extent of that improvement diminishes with N, as expected since the estimator is adversely affected by large N (see Table 1). Actually, when $N = 200$, the variance seems to remain stable when T increases, suggesting that the estimator could be inconsistent. When c_1 increases, the dependence between the endogenous variables and the factors decreases, as well as that between the variables and the z_t that are a proxy for the factors. Therefore, the variance of the IV estimator increases. Interestingly, when p increases, even though the link between the factors and z_t is weaker, there is only a slight increase in the variance of the standard IV estimator. The intuition is that the fit of the first stage regression converges quickly to $\sum_{i=1}^{N} N^{p-1} z_{it}$, which in turn is a close estimate of the true factor.

From Table 2, when $p = 0$, the Factor-IV estimator improves with the sample size T, and the degree of the improvement is not affected by N. Instead, as expected. the Factor-IV estimator is adversely affected by an increase in p and c_1. When $p = 0$, the Factor-IV estimator is better in terms of variance than the standard IV estimator for all values of c_1 and combinations of N and T (except, $T = 30$ and $c_1 = 4$, namely, in a very small sample and with a low explanatory power for the factor). However, when p increases, the sample size T must be larger and larger for the Factor-IV to have a lower variance than standard IV estimator, even for small values of c_1. In particular, notice that when p is closer or equal to 0.5, increasing N increases the variance of the estimator, which remains large even for $T = 200$. Actually, we know that in this case the Factor-estimator may no longer be consistent. Paradoxically, in this case using z_t, which contains information on the true factors f_t,

[7] Notice that when $N > T$ we use generalized inverses in the computation of the first step of the standard (two-stages) IV-estimator.

is better than estimating f_t with the principal component based estimator.

Overall, the Factor-IV estimator is better than the standard IV estimator 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 c_1) also increases the variance of the factor-IV estimator, but more so for the standard IV estimator, unless p is large.

Moving on to the first experimental design, (64), (65) and (66), where the endogenous variable depends on the instrumental variables, a comparison of the standard and Factor-IV estimators yields two main results (see Tables 3 and 4). First, when $p = 0$, the variance of the standard IV estimator increases with N, and for $N = 200$ it does not decrease with T, suggesting that also in this case the estimator could be inconsistent when N is large. Instead, the variance of the Factor-IV estimator is stable with N and always decreases with T, in line with Theorem 3. Moreover, the Factor-IV estimator is systematically more efficient than the standard IV estimator.

Second, when p increases, the variance of the standard IV estimator increases. This could seem surprising, since p is a parameter of the factor model. However, from (69), when p increases the explanatory power of equation (66) for x_t decreases. However, the variance of the Factor-IV estimator increases even more, in particular when p is close or equal to 0.5. In those cases, the standard IV estimator is better. Notice in particular than when $p = 0.5$ and $N = 200$, the variance of the Factor-IV estimator decreases very slowly with T.

The results obtained for the experiments using the third design, (64), (65) and (68), where the endogenous variable depends both on the z and on the factor, are reported in Tables 5 and 6. Overall, the figures are in line with those observed in the first two designs. In particular, as long as p is small, the Factor-IV estimator has a lower variance that the standard IV estimator for virtually any combination of c_1, c_2, N and T. However, when p is large, $0.45 - 0.50$, the T dimension and the value of c_1 become relevant, and the Factor-IV estimator is worse in a variety of cases.

4.2 Additional results

The first 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 (64), (65) and (67), where the endogenous variable depends on the factors, but, prior to using the instruments z_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 (since we consider experiments with one x_t variable only). Then, we carry out standard and Factor-IV estimation as usual. Results are reported in Tables 7-8, focusing for simplicity on a subset of the experiments ($c_1 = 0.5$ only) .

Comparing Tables 7 and 1, it turns out that standard IV estimation is in general improved when instrument preselection occurs. This is, of course, intuitive as the best instruments are retained. A second reason is that variable preselection lowers N, and we have seen in Table 1 that this generally lowers the variance of the standard IV estimator.

Comparing Tables 8 and 2, Factor-IV estimation improves as well, and to a larger extent than standard IV estimation in many cases. Interestingly, the improvement is most apparent for high values of p. When p is high, 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 is superior to standard IV also in the case of a weak factor structure (values of $p > 0$).

Another issue worth exploring is the possibility of weak instruments in the presence of a standard factor model, which corresponds to the setup of Theorem 3. We set $p = 0$ in (65), and substitute (66) and (68) with, respectively,

$$x_t = \sum_{j=1}^{N} N^{-1.2} z_{jt} + u_t, \tag{72}$$

and

$$x_t = \sum_{j=1}^{N} N^{-1.2} z_{jt} + c_1^{-1} f_t + u_t, \tag{73}$$

so that now the instruments are weak. Results for these designs are reported in Tables 9-12.

Starting with (72), a comparison of Tables 9 and 3 for $p = 0$ reveals, as expected, a substantial increase in the variance of the standard IV estimator, in particular for large values of T. From Tables 10 and 4, there is also deterioration in the performance of the Factor-IV estimator, since the factor is also a weak instrument in this case. However the Factor-IV estimator is still preferable to the standard IV estimator when $T \geq 100$. The ranking is instead reversed when p increases, since now not only the factor is a weak instrument but also the factor structure becomes weak.

The case of (73) is perhaps even more interesting. Actually, from (73), conditional on f, the correlation between z_{it} and x_t goes to zero when N increases but, unconditional on f, the correlation does not go to zero, since z_{it} approximates the factor. Hence, the z are no longer weak instruments in the latter case. However, we expect the performance of the standard IV estimator to worsen with respect to the design in (68), since the explanatory power of (73) for x_t is lower.

Comparing Table 11 with 9, we see that indeed the variance of the standard IV estimator decreases, while it is systematically higher than that in Table 5 (based on (68)), and of that in Table 12 based on the Factor-IV estimator (when $p < 0.45$).

A final issue worth exploring 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 focus on the data generating process (65), and introduce serial correlation by letting ϵ_t 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 AR and MA cases are presented in Tables 13-14 and 15-16 respectively.

Reassuringly, the resulting ranking of the estimators is extremely similar to the case of no serial correlation, both for AR and MA errors, with Factor-GMM having a smaller variance than standard GMM in most cases.

In summary, the simulation results indicate that, as long as there is a well-defined factor structure, the Factor-IV (or GMM) estimator is preferable to the standard IV (or GMM) estimator, even when the endogenous variable depends on the instrumental variables rather than on the factors. When the factor structure is loose, other parameters, such as the dimension of the sample, the size of the idiosyncratic component in the factor model or the parameters in the equation for the endogenous variable, become important. Finally, when the factor structure is very weak, the standard IV estimator has a lower variance than the Factor-IV estimator. However, variable preselection strengthens the factor structure, improving the performance of the Factor-IV estimator and making it again often the first best.

5 Empirical Applications

In this Section we discuss two empirical applications of the factor GMM estimation. The former 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 latter 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 (2005).

For the Taylor rule, we adopt the following specification :

$$r_t = \alpha + (1-\rho)\beta(\pi_{t+12} - \pi_t^*) + (1-\rho)\gamma(y_t - y_t^*) + \rho r_{t-1} + \epsilon_t, \tag{74}$$

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.

Estimation of equation (74) presents several problems. First π_{t+12} is correlated with the error term ϵ_t. Second, the error term is correlated over time. In particular, under correct specification of the model, ϵ_t should be an MA(11) since it contains the forecast error $\pi_{t+12}^e - \pi_{t+12}$. Finally, the output gap is likely measured with error, so that $y_t - y_t^*$ can be also correlated with the error term. These problems can be handled by 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. Then we also include factors extracted from a large dataset of macroeconomic and financial variables for the US, the same used in Stock and Watson (2005) that contains 132 time series for the period 1959-2003. If the factors contain useful information, more precise estimates of the parameters should be obtained, as we have seen from a theoretical point of view and in the Monte Carlo simulations.[8]

We focus on the period 1986-2003, since Beyer, Farmer, Henry, and Marcellino (2005) have detected instability in Phillips curves and Taylor rules estimated on a longer sample with an earlier start date. We consider factors estimated in three ways. First, from the whole dataset (All data). Second. from subsets of nominal, real and financial variables (Split data). Third, from variables pre-selected with the Boivin and Ng (2006) criterion (Select data). Pre-selecting the variables has a double effect in this context. First, by only retaining series related to the endogenous variable it attenuates the weak instrument problem. Second, the similarity among the retained series can be expected to be higher than that among all the series, so that the common component can be expected to be dominant with respect to the idiosyncratic component, which decreases problems of a weak factor structure.

The number of factors is determined by the Bai and Ng (2002) criteria, that suggest 8 factors for all data, 2 for nominal and financial variables, and 8 for real variables. Hence, the largest variability in the data seems to be in the real series. For the Select data we use just one factor. Actually, in this case only 13 variables remain in the dataset, those with a correlation with yearly inflation higher than 0.4 in absolute value, and one factor already explains more than 60% of the variability of these 13 variables, suggesting a strong factor structure. For comparison, the first factor explains only 14% of the variability of all the 132 series, and about 20-30% of the subsets of nominal, real and financial variables, so that the factor structure can be expected to be weaker in these cases, and the performance of the Factor-GMM estimator worse. We use one lag of each factor for All data and Split data, 12 lags for the single factor from Select data to use a comparable number of instruments. Adding additional lags of the variables or factors, or additional factors, is not helpful in this application.

The results are reported in Table 17. For the base case, the estimated values for β and γ

[8]We also assume that lagged factors have been used by all the agents in the economy to form expectations on future inflation, which ensures the orthogonality of the factors with the forecast error component of the error in (74).

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 (74). 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 factors to the instrument set does not improve the precision of the estimators of ρ, γ and β when using All data or Split Data. However, when the Select data factor is used, there is a reduction in the variance of the estimator of ρ of about 20%, and of about 4% and 30% for γ and β. With this set of factors there are also no significant changes in the parameter estimates, while there is an improvement in the p-value of the J-statistic. It is also worth mentioning that when only factors are used as instruments, the precision of the GMM estimators decreases substantially, which suggests that a combined use of key macro variables and factors is the optimal solution. Finally, a regression of future (12 months ahead) inflation on the instruments indicates that each set of factors is significant at the 10% level when added to the macro variables.

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

$$\pi_t = c + \gamma\pi_{t+1} + \alpha x_t + \rho\pi_{t-1} + \epsilon_t, \tag{75}$$

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 a real forcing variable (unemployment, with reference to Okun's law, as in e.g. Beyer and Farmer (2003). [9]

As for the Taylor rule, π_{t+1} is correlated with the error term ϵ_t, which in turn is correlated over time. Hence, we estimate the parameters of (75) by GMM, with a correction for the MA component in the error ϵ_t, and the same four sets of instruments as for the Taylor rule (but using the second lag of inflation).

The results are reported in Table 18. 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 factors to the instrument set improves the precision of the estimators of

[9]The results are qualitatively similar using the output gap.

all parameters, with the best results again from 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.

Finally, since the number of variables in the "Select data" set is relatively small, they could be directly used as additional instrumental variables instead of the factor that drives all of them. However, it turns out that the resulting parameter estimators are substantially less efficient than the factor-GMM estimators for both the Taylor rule and the hybrid Phillips curve.

In summary, these two examples confirm the relevance of factors as additional instruments for GMM estimation. Moreover, and in line with the results for forecasting, variable pre-selection appears to be relevant for the extraction of the factors to be used as (additional) instruments in GMM estimation.

6 Conclusions

The use of factor models has become very popular in the last few years, following the seminal work of Stock and Watson (2002b) and Forni, Hallin, Lippi, and Reichlin (2000). Paralleling the developments in the VAR literature in the '80s and '90s, so far factor models have been mainly used for reduced form modelling and forecasting. However, recently there has a been an interest in more structural applications of factor analysis. Stock and Watson (2005), Giannone, Reichlin, and Sala (2002) and Kapetanios and Marcellino (2006a) have shown that it is possible to obtain more realistic impulse response functions in a structural factor model. Favero, Marcellino, and Neglia (2005) and Beyer, Farmer, Henry, and Marcellino (2005) have estimated structural forward looking equations, such as those typically encountered in DSGE models, by means of factor augmented GMM estimation.

In this paper, and in a related independent article by Bai and Ng (2006b), we develop the theoretical underpinnings of Factor-GMM estimation. 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 (2006b) 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 three directions. First, we evaluate what happens when the endogenous variables depend on a large set of instrumental variables rather than on the factors, or on a combination of them. We show theoretically that in this case the ranking of the standard and Factor-IV estimators is no longer clear-cut, since it depends on the specific parameter values. However, in an extensive set of simulation experiments, we have found that Factor-IV estimation seems to be more efficient also in this context.

Second, 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.

Third, 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, at least in our designs, a weak factor structure is more relevant than a weak instrument situation. Moreover, in the presence of a well defined factor structure, and even with weak instruments, Factor-IV estimation is in general more efficient 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 efficiency of the Factor-IV estimators. Similar results hold for Factor-GMM estimators

Finally, we apply Factor-GMM for the estimation of a Taylor rule and 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.

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

——— (2006b): "Instrumental Variable Estimation in a Data Rich Environment," *Mimeo.*

Beck, G., K. Hubrich, and M. Marcellino (2006): "Regional inflation dynamics within and across euro area countries and a comparison with the US," *ECB Working Paper No. 681.*

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 (2003): "On the indeterminacy of New-Keynesian economics," *ECB Working Paper No. 323.*

Beyer, A., R. Farmer, J. Henry, and M. Marcellino (2005): "Factor analysis in a new-Keynesian model," *ECB Working Paper No. 510.*

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

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure and Man-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

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*, Forthcoming.

Florens, J. P., J. Johannes, and S. Van Bellegem (2006): "Instrumental Regression in Partially Linear Models," *mimeo.*

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.*

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 (2006): "Many Instruments, Weak Instruments and Microeconometric Practice," *Working Paper, MIT.*

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

Kapetanios, G., and M. Marcellino (2006a): "Factor-GMM Estimation with Large Sets of Possibly Weak Instruments," *Queen Mary Working Paper No. 577.*

——— (2006b): "A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions," *CEPR Working Paper No. 5620.*

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

Lewbel, A. (1991): "The Rank of Demand Systems: Theory and Nonparametric Estimation," *Econometrica*, 59, 711–730.

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 (1989): "New Indices of Coincident and Leading Indicators," in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard, and S. Fischer. Cambridge, M.I.T. Press.

——— (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 $\sqrt{T}(\bar{\beta} - \beta)$ and $\sqrt{T}(\tilde{\beta} - \beta)$ are given by the probability limits as $T \to \infty$ of $\left(\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T}\right)^{-1}$ and $\left(\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T}\right)^{-1}$. We begin by deriving results under (4) as it is more straightforward. 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

$$\frac{X'F}{T} = \frac{\left(A^{0\prime}F' + u'\right)F}{T} \xrightarrow{p} A^{0\prime}\Sigma_f \tag{76}$$

$$\frac{F'F}{T} \xrightarrow{p} \Sigma_f \tag{77}$$

$$\frac{X'Z}{T} = \frac{\left(A^{0\prime}F' + u'\right)\left(F\Lambda^0 + v\right)}{T} \xrightarrow{p} A^{0\prime}\Sigma_f\Lambda^0 \tag{78}$$

$$\frac{Z'Z}{T} = \frac{\left(F\Lambda^0 + v\right)'\left(F\Lambda^0 + v\right)}{T} \xrightarrow{p} \Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v \tag{79}$$

Then, (15) and (18) easily follow. Similarly, under (3), (76) and (79) become

$$\frac{X'F}{T} = \frac{\left(A_Z^{0\prime}\left(\Lambda^{0\prime}F' + v'\right) + u'\right)F}{T} \xrightarrow{p} A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f \tag{80}$$

$$\frac{X'Z}{T} = \frac{\left(A_Z^{0\prime}\left(\Lambda^{0\prime}F' + v'\right) + u'\right)\left(F\Lambda^0 + v\right)}{T} \xrightarrow{p} A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 + A_Z^{0\prime}\Sigma_v = A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right) \tag{81}$$

Hence, (14) and (17) easily follow. Finally, under (5), (76) and (79) become

$$\frac{X'F}{T} = \frac{\left(A_Z^{0\prime}\left(\Lambda^{0\prime}F' + v'\right) + A^{0\prime}F' + u'\right)F}{T} \xrightarrow{p} \left(A_Z^{0\prime}\Lambda^{0\prime} + A^{0\prime}\right)\Sigma_f \tag{82}$$

$$\frac{X'Z}{T} = \frac{\left(A_Z^{0\prime}\left(F'\Lambda^{0\prime} + v'\right) + A^{0\prime}F' + u'\right)\left(F\Lambda^0 + v\right)}{T} \xrightarrow{p} A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 + A^{0\prime}\left(\Sigma_v + \Sigma_f\Lambda^0\right) \tag{83}$$

We next examine

$$\left(A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 A_Z^0\right)^{-1} - \left(A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right)A_Z^0\right)^{-1} \tag{84}$$

which is positive semidefinite (psd) if

$$A_Z^{0\prime}\left(\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v\right)A_Z^0 - A_Z^{0\prime}\Lambda^{0\prime}\Sigma_f\Lambda^0 A_Z^0 \tag{85}$$

is psd. But (85) is psd since

$$\Lambda^{0\prime}\Sigma_f\Lambda^0 + \Sigma_v - \Lambda^{0\prime}\Sigma_f\Lambda^0 = \Sigma_v \tag{86}$$

Hence the result follows. Finally we examine

$$\left(A^{0\prime}\Sigma_f\Lambda^{0\prime}\left(\Lambda^0\Sigma_f\Lambda^{0\prime} + \Sigma_v\right)^{-1}\Lambda^0\Sigma_f A^0\right)^{-1} - \left(A^{0\prime}\Sigma_f A^0\right)^{-1} \tag{87}$$

which is psd if

$$A^{0\prime}\Sigma_f A^0 - A^{0\prime}\Sigma_f\Lambda^{0\prime}\left(\Lambda^0\Sigma_f\Lambda^{0\prime} + \Sigma_v\right)^{-1}\Lambda^0\Sigma_f A^0 \tag{88}$$

is. But (88) is psd if

$$A^{0\prime}\Sigma_f A^0 - A^{0\prime}\Sigma_f\Lambda^{0\prime}\left(\Lambda^0\Sigma_f\Lambda^{0\prime}\right)^{-1}\Lambda^0\Sigma_f A^0 \tag{89}$$

is. Define $\tilde{A}^0 = \Sigma_f^{1/2}A^0$ and $\tilde{\Lambda}^0 = \Lambda^0\Sigma_f^{1/2}$. Then, (89) becomes

$$\tilde{A}^{0\prime}\tilde{A}^0 - \tilde{A}^{0\prime}\tilde{\Lambda}^{0\prime}\left(\tilde{\Lambda}^0\tilde{\Lambda}^{0\prime}\right)^{-1}\tilde{\Lambda}^0\tilde{A}^0 = \tilde{A}^{0\prime}\left(I - \tilde{\Lambda}^{0\prime}\left(\tilde{\Lambda}^0\tilde{\Lambda}^{0\prime}\right)^{-1}\tilde{\Lambda}^0\right)\tilde{A}^0 \tag{90}$$

which is psd since $I - \tilde{\Lambda}^{0\prime}\left(\tilde{\Lambda}^0\tilde{\Lambda}^{0\prime}\right)^{-1}\tilde{\Lambda}^0$ is. Hence the result follows.

Proof of Theorem 2

We need to prove that

$$\sqrt{T}\left(\left(X'F(F'F)^{-1}F'X\right)^{-1}X'F(F'F)^{-1}F'\epsilon - \left(X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'X\right)^{-1}X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'\epsilon\right) = o_p(1) \tag{91}$$

or

$$\left(\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T}\right)^{-1}\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\epsilon}{T^{1/2}} - \tag{92}$$

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

(92) follows if

$$\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T} - \frac{X'\hat{F}}{T}\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{\hat{F}'X}{T} = o_p(1) \tag{93}$$

and

$$\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\epsilon}{T^{1/2}} - \frac{X'\hat{F}}{T}\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{\hat{F}'\epsilon}{T^{1/2}} = o_p(1) \tag{94}$$

(93) and (94) follow if

$$\frac{X'F}{T} - \frac{X'\hat{F}}{T} = o_p(1) \tag{95}$$

$$\frac{F'F}{T} - \frac{\hat{F}'F}{T} = o_p(1) \tag{96}$$

and

$$\sqrt{T}\left(\frac{F'\epsilon}{T} - \frac{\hat{F}'\epsilon}{T}\right) = o_p(1) \tag{97}$$

hold. We examine (95)-(97). They can all be written as

$$A_T \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = o_p(1) \tag{98}$$

where A_T is $1, 1$ and $\sqrt{T}$ and q_t is x_t, f_t and ϵ_t respectively for (95)-(97). By Lemma A.1 of Bai and Ng (2006a) we have that

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = O_p\left(\min(N,T)^{-1}\right) \tag{99}$$

as long as q_t has finite fourth moments, nonsingular covariance matrix and satisfies a central limit theorem. These conditions are satisfied for x_t, f_t and ϵ_t via assumptions 2 and 3. Hence, (95)-(96) follow, while (97) follows if $\sqrt{T}/N = o(1)$.

Proof of Theorem 3

We establish (24). (25)-(28) follow similarly. We examine the asymptotic distribution of

$$T^{1/2-\psi}\left(X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'X\right)^{-1} X'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'\epsilon \tag{100}$$

By theorem 4 and (98) for $A_T = o_p(T^{1/2})$ and since by assumption $N = O(T^\gamma)$, $\gamma > 1/2$, it is sufficient to examine the asymptotic distribution of

$$\left(\frac{X'F}{T^{1-\psi}}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T^{1-\psi}}\right)^{-1}\frac{X'F}{T^{1-\psi}}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\epsilon}{T^{1/2}} \tag{101}$$

A standard central limit theorem suffices to show that under assumptions 1-3

$$T^{-1/2}F'\epsilon \xrightarrow{d} N\left(0, \sigma_\epsilon^2\Sigma_f\right) \tag{102}$$

We examine the limits of $\frac{F'F}{T}$ and $\frac{X'F}{T^{1-\alpha}}$. The first is given in (77). We examine the second. We have

$$\frac{X'F}{T^{1-\psi}} = \frac{\left(A^{0\prime}_{(N,T)Z}\left(\Lambda^{0\prime}F' + v'\right) + u'\right)F}{T^{1-\psi}} = \tag{103}$$
$$\frac{A^{0\prime}_{(N,T)Z}\Lambda^{0\prime}}{T^{-\psi}}\frac{F'F}{T} + \frac{A^{0\prime}_{(N,T)Z}v'F}{T^{1-\psi}} + \frac{u'F}{T^{1-\psi}}$$

The second and third terms of the RHS of (103) tend to zero since $1-\psi > 1/2$. The first term tends to $\Upsilon'\Sigma_f$. Hence, the result follows.

Proof of Theorem 4

The covariance matrix of Z, Σ_Z, is given by $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0+\Sigma_v$ 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 $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0$ and Σ_v are bounded. By assumption the eigenvalues of Σ_v are bounded. Hence we examine $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0$. By Schwarz, Rutishauser, and Stiefel (1973), the eigenvalues of $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0$, will be bounded if the column sum norm of $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0$ is bounded. But every element of $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0$ is $O(N^{-2\alpha})$. Hence the column sum norm of $\Lambda_N^{0\prime}\Sigma_f\Lambda_N^0$ is O(1) for all $\alpha \geq 1/2$. Hence the result follows.

Proof of Theorem 5

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

$$H = (\tilde{F}'F/T)(\Lambda_N^{0\prime}\Lambda_N^0/N^{1-2\alpha}) \tag{104}$$

Then,

$$\hat{f}_t - Hf_t = N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t) + N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\zeta_{st} + \tag{105}$$
$$N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\eta_{st} + N^{2\alpha}T^{-1}\sum_{s=1}^{T}\tilde{f}_s\xi_{st}$$

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

$$\zeta_{st} = e_s'e_t/N - \gamma_N(s,t) \tag{106}$$

$$\eta_{st} = f_s'^0\Lambda_N^0 e_t/N \tag{107}$$

$$\xi_{st} = f_t'^0\Lambda_N^0 e_s/N = \eta_{ts} \tag{108}$$

It is easy to see that

$$||\hat{f}_t - Hf_t||^2 \leq 4(a_t + b_t + c_t + d_t) \tag{109}$$

where

$$a_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t)\right\|^2 \tag{110}$$

$$b_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\zeta_{st}\right\|^2 \tag{111}$$

$$c_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right\|^2 \tag{112}$$

$$d_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\xi_{st}\right\|^2 \tag{113}$$

It follows that

$$1/T\sum_{t=1}^{T}||\hat{f}_t - Hf_t||^2 \leq c/T\sum_{t=1}^{T}(a_t + b_t + c_t + d_t) \tag{114}$$

for some constant c. Now

$$\left\|\sum_{s=1}^{T}\tilde{f}_t\gamma_N(s,t)\right\|^2 \leq \left(\sum_{s=1}^{T}\left\|\tilde{f}_s\right\|^2\right)\left(\sum_{s=1}^{T}\gamma_N^2(s,t)\right) \tag{115}$$

which implies that

$$1/T\sum_{t=1}^{T}a_t = O_p(N^{4\alpha}T^{-1}) \tag{116}$$

Following the analysis for $N^{-4\alpha}b_t$ given in the proof of Bai and Ng (2002) we see that

$$1/N^{4\alpha}T\sum_{t=1}^{T}b_t = O_p(N^{-1}) \tag{117}$$

and hence

$$1/T\sum_{t=1}^{T}b_t = O_p(N^{-1+4\alpha}) = o_p(1) \tag{118}$$

as long as $a < 1/4$. Finally we look at c_t. d_t can be treated similarly.

$$c_t = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right\|^2 = N^{4\alpha}T^{-2}\left\|\sum_{s=1}^{T}\tilde{f}_s f_s'^0\Lambda_N^0 e_t/N\right\|^2 = \tag{119}$$

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

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

$$N^{-2+2\alpha}\left\|\Lambda^0 e_t\right\|^2 O_p(1)$$

So

$$1/T\sum_{t=1}^{T}c_t = O_p(1)N^{-1+2\alpha}T^{-1}\sum_{t=1}^{T}\left\|\frac{\Lambda^0 e_t}{N^{1/2}}\right\|^2 = O_p(N^{-1+2\alpha}) \tag{120}$$

Proof of Lemma 1

We follow the proof of Lemma A.1 of Bai and Ng (2006a). Using (106)-(108) we get

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\gamma_N(s,t)\right)q_t' + N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\zeta_{st}\right) + \quad (121)$$

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

The first two terms of (121) apart from the normalisation $N^{2\alpha}$ are the same as those analysed in Lemma A.1 of Bai and Ng (2006a). 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 (121) are analysed similarly. We focus on the third term. We have

$$N^{2\alpha}T^{-2}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{f}_s\eta_{st}\right)q_t' = N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}Hf_s\eta_{st}\right)q_t' + \quad (122)$$

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

The first term on the RHS of (122) can be written as

$$N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}Hf_s\eta_{st}\right)q_t' = N^{2\alpha}\left(H\frac{1}{T}\sum_{t=1}^{T}f_sf_s'\right)\frac{1}{NT}\sum_{t=1}^{T}\Lambda_N^0 e_t q_t' = \quad (123)$$

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

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

For the second term of (122) we have

$$\left\|N^{2\alpha}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\left(\tilde{f}_s - Hf_s\right)\eta_{st}\right)q_t'\right\| \leq \quad (124)$$

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

But

$$\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{f}_s - Hf_s\right\|^2\right)^{1/2} = O_p\left(\min\left(N^{-2\alpha}T^{1/2}, N^{1/2-2\alpha}\right)^{-1}\right) \quad (125)$$

by Theorem 5. Then,

$$N^{4\alpha}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\eta_{st}q_t'\right\|^2 = N^{4\alpha}\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{f_s'^0\Lambda_N^0 e_t}{N}q_t'\right\|^2 = \tag{126}$$

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

But,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{f_s'^0\Lambda^0 e_t}{N}q_t' = O_p(N^{-1/2}) \tag{127}$$

and so the RHS of (126) is also $O_p(N^{-1+2\alpha})$. So the second term of (122) 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 (121) 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

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - Hf_t)q_t' = O_p\left(\min\left(N^{1/2-3\alpha}T^{1/2}, N^{1-3\alpha}, N^{-2\alpha}\min(N,T), (NT)^{1/2}\right)^{-1}\right) = O_p\left(C_{NT}^{-1}\right) \tag{128}$$

where

$$C_{NT} \equiv \min\left(N^{1/2-3\alpha}T^{1/2}, N^{1-3\alpha}, N^{-2\alpha}\min(N,T)\right) \tag{129}$$

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

Proof of Theorem 6

The results follow from Lemma 1, (98) and the proofs of Theorems 2 and 3. More particularly, the results of Theorem 2 require a result similar to (98) which is now provided by Lemma 1. Then, the results of Theorem 2 are obtained immediately. Similarly, once Lemma 1 is used (24)-(28) of Theorem 3 follow.

Proof of Theorem 7

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 4.. 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 (42). 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 8

The general expressions for the covariance matrices of $\sqrt{T}(\bar{\beta}-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta)$ are given by the probability limits as $T \to \infty$ of

$$\left(\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T}\right)^{-1}\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'\epsilon\epsilon'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T}\left(\frac{X'F}{T}\left(\frac{F'F}{T}\right)^{-1}\frac{F'X}{T}\right)^{-1}$$

and

$$\left(\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T}\right)^{-1}\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'\epsilon\epsilon'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T}\left(\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T}\right)^{-1}.$$

The probability limits of $\frac{X'F}{T}$, $\frac{F'F}{T}$, $\frac{X'Z}{T}$ and $\frac{Z'Z}{T}$ are as in Theorem 3, while

$$\frac{F'\epsilon\epsilon'F}{T} \xrightarrow{p} S_{f\epsilon}, \tag{130}$$

$$\frac{Z'\epsilon\epsilon'Z}{T} \xrightarrow{p} S_{z\epsilon}. \tag{131}$$

Proof of Theorem 9

The general expressions for the covariance matrices of $\sqrt{T}(\bar{b}-\beta)$ and $\sqrt{T}(\widetilde{b}-\beta)$ are given by the probability limits as $T \to \infty$ of $\left(\frac{X'F}{T}S_{f\epsilon}^{-1}\frac{F'X}{T}\right)^{-1}$ and $\left(\frac{X'Z}{T}\widehat{S}_{z\epsilon}^{-1}\frac{Z'X}{T}\right)^{-1}$. The results follow from those in the Proof of Theorem 1 and consistency of the HAC estimator of S.

Proof of Theorem 10

We need to prove that

$$\sqrt{T}\left(\left(X'FS_{f\epsilon}^{-1}F'X\right)^{-1}X'FS_{f\epsilon}^{-1}F'\epsilon-\left(X'\hat{F}\widehat{S}_{\hat{f}\epsilon}^{-1}\hat{F}'X\right)^{-1}X'\hat{F}\widehat{S}_{\hat{f}\epsilon}^{-1}\hat{F}'\epsilon\right)=o_p(1) \tag{132}$$

or

$$\left(\frac{X'F}{T}S_{f\epsilon}^{-1}\frac{F'X}{T}\right)^{-1}\frac{X'F}{T}S_{f\epsilon}^{-1}\frac{F'\epsilon}{T^{1/2}}- \tag{133}$$
$$\left(\frac{X'\hat{F}}{T}\widehat{S}_{\hat{f}\epsilon}^{-1}\frac{\hat{F}'X}{T}\right)^{-1}\frac{X'\hat{F}}{T}\widehat{S}_{\hat{f}\epsilon}^{-1}\frac{\hat{F}'\epsilon}{T^{1/2}}=o_p(1)$$

From the proof of Theorem 4 we already know that

$$\frac{X'F}{T}-\frac{X'\hat{F}}{T}=o_p(1) \tag{134}$$

and

$$\sqrt{T}\left(\frac{F'\epsilon}{T}-\frac{\hat{F}'\epsilon}{T}\right)=o_p(1). \tag{135}$$

Then we have,

$$\widehat{S}_{\widehat{f\epsilon},h}=\widehat{\Phi}_0+\sum_{j=1}^{h}(1-\frac{j}{h+1})(\widehat{\Phi}_j+\widehat{\Phi}_j') \tag{136}$$

$$\widehat{\Phi}_j=T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_{t-j}\widehat{f}_t\widehat{f}_{t-j}',$$

so that

$$\Phi_j-\widehat{\Phi}_j=T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_{t-j}\left(f_tf_{t-j}'-\widehat{f}_t\widehat{f}_{t-j}'\right). \tag{137}$$

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

$$\left\|T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t^2\left(f_tf_t'-\widehat{f}_t\widehat{f}_t'\right)\right\|\leq C_1\left\|T^{-1}\sum_{T=j+1}^{T}\widehat{\epsilon}_t^2f_t'\left(Hf_t-\widehat{f}_t'\right)\right\|\leq \tag{138}$$

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

for some constants C_1,C_2 and C_3. But, by (98) and $\sqrt{T}/N=o(1)$,

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

as long as ϵ_t has finite eighth moments. Then,

$$\left\|T^{-1}\sum_{T=j+1}^{T}(\widehat{\epsilon}_t-\epsilon_t)f_t'\left(Hf_t-\widehat{f}_t'\right)\right\|\leq C_4\left\|T^{-1}(b-\beta)\sum_{T=j+1}^{T}x_tf_t'\left(Hf_t-\widehat{f}_t'\right)\right\| \tag{139}$$

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

for some constants C_4 and C_5. Again by (98) 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 11

It follows from the proof of theorem 3, given Theorem 10 and consistency of $\widehat{S}_{\widehat{f\epsilon}}$.

Table Appendix

Setup of equation (4)

Table 1. Variance of standard IV estimator														
		c_1	0.5				1				4			
r	p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
	0	30	0.251	0.252	0.255	0.256	0.372	0.371	0.376	0.371	0.576	0.579	0.601	0.586
	0	50	0.179	0.246	0.251	0.248	0.284	0.370	0.371	0.368	0.522	0.577	0.574	0.580
	0	100	0.112	0.154	0.239	0.243	0.182	0.248	0.365	0.370	0.405	0.484	0.584	0.570
	0	200	0.066	0.089	0.147	0.243	0.108	0.152	0.241	0.356	0.285	0.356	0.476	0.575
	0.1	30	0.251	0.250	0.249	0.250	0.366	0.381	0.371	0.374	0.580	0.575	0.579	0.579
	0.1	50	0.181	0.242	0.241	0.252	0.277	0.366	0.371	0.367	0.528	0.573	0.576	0.573
	0.1	100	0.109	0.150	0.234	0.239	0.181	0.240	0.358	0.356	0.412	0.491	0.568	0.574
	0.1	200	0.069	0.091	0.149	0.236	0.111	0.161	0.240	0.348	0.291	0.370	0.472	0.581
	0.25	30	0.256	0.250	0.262	0.253	0.371	0.372	0.368	0.381	0.579	0.584	0.591	0.574
	0.25	50	0.187	0.251	0.245	0.248	0.289	0.362	0.366	0.366	0.533	0.575	0.575	0.566
	0.25	100	0.120	0.161	0.242	0.237	0.193	0.259	0.364	0.357	0.415	0.485	0.567	0.563
1	0.25	200	0.073	0.097	0.154	0.247	0.122	0.159	0.245	0.356	0.303	0.379	0.487	0.567
	0.33	30	0.257	0.253	0.261	0.255	0.378	0.374	0.373	0.364	0.572	0.585	0.589	0.575
	0.33	50	0.192	0.251	0.246	0.247	0.295	0.365	0.362	0.361	0.538	0.561	0.568	0.574
	0.33	100	0.123	0.165	0.240	0.238	0.201	0.266	0.359	0.363	0.423	0.490	0.564	0.567
	0.33	200	0.077	0.103	0.158	0.241	0.127	0.174	0.248	0.359	0.314	0.389	0.497	0.572
	0.45	30	0.259	0.258	0.253	0.254	0.369	0.374	0.371	0.374	0.582	0.576	0.589	0.596
	0.45	50	0.201	0.248	0.254	0.251	0.318	0.362	0.366	0.363	0.540	0.580	0.588	0.572
	0.45	100	0.140	0.182	0.240	0.243	0.227	0.279	0.356	0.355	0.463	0.508	0.568	0.567
	0.45	200	0.089	0.125	0.169	0.237	0.148	0.195	0.273	0.359	0.364	0.425	0.508	0.564
	0.5	30	0.251	0.252	0.257	0.251	0.368	0.371	0.385	0.373	0.585	0.579	0.568	0.580
	0.5	50	0.205	0.240	0.252	0.249	0.326	0.364	0.350	0.358	0.550	0.582	0.568	0.579
	0.5	100	0.155	0.189	0.240	0.244	0.238	0.292	0.359	0.361	0.476	0.526	0.567	0.567
	0.5	200	0.097	0.129	0.183	0.234	0.162	0.209	0.290	0.357	0.366	0.446	0.511	0.564

Table 2. Variance of Factor-IV estimator														
		c_1		0.5				1				4		
r	p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
	0	30	0.146	0.147	0.148	0.147	0.217	0.241	0.215	0.210	0.764	0.868	0.720	0.857
	0	50	0.105	0.107	0.101	0.106	0.156	0.143	0.162	0.161	0.438	0.528	0.474	0.433
	0	100	0.070	0.073	0.075	0.070	0.106	0.107	0.104	0.108	0.247	0.242	0.227	0.242
	0	200	0.052	0.050	0.052	0.050	0.072	0.074	0.074	0.073	0.154	0.160	0.153	0.149
	0.1	30	0.153	0.148	0.145	0.150	0.233	0.250	0.225	0.223	0.829	0.795	0.772	0.941
	0.1	50	0.109	0.110	0.113	0.108	0.163	0.162	0.163	0.154	0.463	0.468	0.452	0.411
	0.1	100	0.074	0.074	0.071	0.073	0.104	0.102	0.103	0.106	0.262	0.248	0.232	0.238
	0.1	200	0.051	0.053	0.054	0.050	0.075	0.074	0.072	0.075	0.162	0.151	0.151	0.147
	0.25	30	0.205	0.158	0.175	0.157	0.262	0.293	0.258	0.272	1.129	0.810	0.917	0.858
	0.25	50	0.116	0.116	0.122	0.119	0.168	0.162	0.173	0.174	0.548	0.468	0.436	0.497
	0.25	100	0.079	0.083	0.078	0.076	0.117	0.117	0.114	0.110	0.253	0.253	0.251	0.248
1	0.25	200	0.054	0.053	0.054	0.054	0.079	0.077	0.073	0.073	0.175	0.166	0.158	0.160
	0.33	30	0.303	0.238	0.296	0.365	0.409	0.391	0.487	0.480	1.105	0.975	1.132	1.228
	0.33	50	0.126	0.129	0.134	0.134	0.202	0.209	0.186	0.216	0.547	0.607	0.695	0.656
	0.33	100	0.089	0.085	0.082	0.085	0.125	0.119	0.117	0.122	0.345	0.283	0.311	0.281
	0.33	200	0.056	0.055	0.057	0.057	0.083	0.083	0.080	0.080	0.177	0.172	0.171	0.171
	0.45	30	0.800	1.021	1.169	1.348	1.053	1.184	1.436	1.565	1.753	1.711	1.872	1.775
	0.45	50	0.433	0.678	0.878	1.125	0.787	0.872	1.195	1.355	1.347	1.459	1.787	1.866
	0.45	100	0.108	0.211	0.388	0.698	0.160	0.249	0.579	0.944	0.489	0.672	0.936	1.375
	0.45	200	0.071	0.069	0.073	0.157	0.100	0.105	0.110	0.148	0.242	0.329	0.245	0.376
	0.5	30	1.134	1.237	1.353	1.451	1.217	1.468	1.570	1.796	1.758	1.975	1.994	1.919
	0.5	50	0.749	1.004	1.475	1.562	0.833	1.103	1.429	1.780	1.816	1.592	1.764	1.938
	0.5	100	0.361	0.611	0.846	1.183	0.484	0.703	1.103	1.355	0.964	1.210	1.430	1.872
	0.5	200	0.080	0.106	0.346	0.738	0.123	0.182	0.508	0.840	0.274	0.418	0.869	1.203

Setup of equation (3)

Table 3. Variance of standard IV estimator

		c_2	0.5				1				4			
r	p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
	0	30	0.362	0.367	0.373	0.374	0.362	0.368	0.383	0.376	0.357	0.358	0.373	0.366
	0	50	0.281	0.375	0.359	0.365	0.274	0.360	0.363	0.362	0.263	0.347	0.363	0.361
	0	100	0.172	0.250	0.356	0.364	0.171	0.237	0.354	0.361	0.160	0.236	0.358	0.351
	0	200	0.108	0.149	0.237	0.356	0.107	0.144	0.237	0.356	0.098	0.142	0.233	0.357
	0.1	30	0.485	0.488	0.524	0.544	0.478	0.494	0.509	0.538	0.446	0.471	0.511	0.530
	0.1	50	0.390	0.491	0.510	0.534	0.380	0.486	0.516	0.529	0.354	0.459	0.497	0.530
	0.1	100	0.272	0.377	0.514	0.528	0.264	0.366	0.507	0.524	0.247	0.357	0.496	0.511
	0.1	200	0.172	0.257	0.393	0.531	0.166	0.257	0.392	0.525	0.144	0.230	0.374	0.523
	0.25	30	0.612	0.629	0.649	0.660	0.594	0.616	0.649	0.676	0.541	0.598	0.635	0.663
	0.25	50	0.547	0.606	0.630	0.665	0.543	0.610	0.652	0.674	0.473	0.575	0.622	0.660
	0.25	100	0.428	0.546	0.641	0.651	0.424	0.533	0.642	0.677	0.359	0.503	0.629	0.658
1	0.25	200	0.312	0.448	0.597	0.654	0.310	0.438	0.580	0.659	0.244	0.387	0.563	0.649
	0.33	30	0.637	0.674	0.672	0.676	0.628	0.657	0.683	0.697	0.600	0.609	0.672	0.681
	0.33	50	0.609	0.654	0.668	0.682	0.586	0.641	0.685	0.701	0.528	0.615	0.665	0.687
	0.33	100	0.522	0.617	0.672	0.682	0.494	0.590	0.678	0.681	0.406	0.548	0.656	0.672
	0.33	200	0.405	0.538	0.636	0.666	0.372	0.519	0.641	0.675	0.288	0.437	0.598	0.677
	0.45	30	0.675	0.697	0.696	0.714	0.664	0.689	0.700	0.703	0.612	0.646	0.669	0.696
	0.45	50	0.640	0.682	0.698	0.698	0.619	0.678	0.700	0.699	0.554	0.642	0.670	0.688
	0.45	100	0.599	0.654	0.680	0.694	0.573	0.648	0.696	0.692	0.448	0.582	0.666	0.689
	0.45	200	0.517	0.622	0.683	0.696	0.474	0.595	0.685	0.705	0.332	0.492	0.636	0.702
	0.5	30	0.677	0.692	0.695	0.707	0.675	0.686	0.704	0.704	0.626	0.661	0.692	0.719
	0.5	50	0.646	0.700	0.694	0.703	0.639	0.683	0.702	0.706	0.565	0.654	0.693	0.704
	0.5	100	0.619	0.668	0.707	0.710	0.583	0.665	0.695	0.689	0.461	0.600	0.674	0.699
	0.5	200	0.541	0.631	0.697	0.705	0.498	0.619	0.678	0.708	0.342	0.519	0.645	0.711

Table 4. Variance of Factor-IV estimator														
		c_2	0.5				1				4			
r	p	T/N	30	50	100	200	30	50	100	200	30	50	100	200
	0	30	0.209	0.199	0.221	0.206	0.211	0.229	0.219	0.223	0.202	0.224	0.228	0.227
	0	50	0.169	0.153	0.154	0.163	0.149	0.163	0.151	0.156	0.148	0.141	0.149	0.155
	0	100	0.105	0.105	0.103	0.104	0.102	0.103	0.103	0.104	0.101	0.107	0.104	0.104
	0	200	0.074	0.072	0.072	0.074	0.071	0.069	0.075	0.069	0.069	0.068	0.069	0.074
	0.1	30	0.489	0.492	0.402	0.626	0.499	0.439	0.490	0.457	0.443	0.426	0.417	0.595
	0.1	50	0.226	0.282	0.347	0.329	0.218	0.258	0.291	0.322	0.225	0.233	0.296	0.294
	0.1	100	0.149	0.159	0.176	0.186	0.149	0.156	0.168	0.202	0.133	0.153	0.174	0.185
	0.1	200	0.107	0.110	0.118	0.129	0.100	0.106	0.119	0.124	0.093	0.101	0.111	0.124
	0.25	30	1.091	1.303	1.485	1.573	0.914	1.206	1.332	1.786	1.272	1.488	1.583	1.886
	0.25	50	0.634	0.653	0.959	1.293	0.735	0.828	0.897	1.297	0.814	1.105	1.355	1.388
	0.25	100	0.280	0.348	0.544	0.904	0.305	0.352	0.486	0.708	0.236	0.362	0.459	0.843
1	0.25	200	0.167	0.199	0.266	0.418	0.168	0.212	0.277	0.506	0.137	0.172	0.206	0.304
	0.33	30	1.450	1.538	1.733	1.905	1.465	1.587	1.955	1.984	1.691	1.807	1.902	1.993
	0.33	50	0.910	1.175	1.590	1.710	0.960	1.120	1.369	1.682	1.533	1.446	1.868	2.187
	0.33	100	0.511	0.627	0.994	1.423	0.474	0.617	1.060	1.307	0.809	1.105	1.549	1.821
	0.33	200	0.252	0.316	0.467	0.819	0.231	0.287	0.511	0.707	0.204	0.390	0.693	1.052
	0.45	30	1.825	2.087	2.281	2.101	1.937	2.027	1.994	2.034	2.009	2.012	1.966	1.970
	0.45	50	1.600	1.780	2.018	2.096	1.462	1.916	2.086	1.986	1.835	2.168	2.008	1.980
	0.45	100	0.857	1.328	1.592	1.909	0.900	1.230	1.811	2.029	1.542	1.903	2.143	2.204
	0.45	200	0.475	0.783	1.140	1.591	0.346	0.648	1.207	1.760	1.265	1.501	1.844	1.971
	0.5	30	1.798	2.108	1.932	2.156	2.101	2.103	1.911	2.166	2.019	1.996	2.150	2.157
	0.5	50	1.633	1.906	2.171	2.117	1.708	2.072	2.055	2.015	1.796	1.978	2.223	2.147
	0.5	100	1.182	1.469	2.035	2.069	1.168	1.740	2.028	2.045	1.753	1.933	1.823	2.074
	0.5	200	0.534	0.907	1.651	2.010	0.470	1.013	1.596	2.121	1.227	1.652	1.924	1.996

Setup of equation (5)

Table 5. Variance of standard IV estimator

			c_2	0.5				1			
r	p	c_1	T/N	30	50	100	200	30	50	100	200
	0	0.5	30	0.122	0.124	0.122	0.124	0.121	0.124	0.122	0.125
	0	0.5	50	0.083	0.114	0.113	0.118	0.083	0.114	0.115	0.116
	0	0.5	100	0.053	0.069	0.105	0.110	0.051	0.066	0.107	0.111
	0	0.5	200	0.033	0.039	0.063	0.104	0.035	0.039	0.059	0.106
	0	1	30	0.162	0.164	0.157	0.159	0.160	0.161	0.161	0.158
	0	1	50	0.109	0.152	0.149	0.153	0.111	0.154	0.150	0.155
	0	1	100	0.068	0.090	0.149	0.154	0.065	0.089	0.146	0.159
	0	1	200	0.041	0.052	0.083	0.142	0.043	0.052	0.084	0.143
	0.1	0.5	30	0.148	0.149	0.152	0.157	0.151	0.151	0.153	0.159
	0.1	0.5	50	0.099	0.140	0.148	0.157	0.101	0.142	0.149	0.153
	0.1	0.5	100	0.062	0.081	0.142	0.149	0.061	0.083	0.143	0.147
1	0.1	0.5	200	0.039	0.050	0.082	0.145	0.039	0.048	0.081	0.143
	0.1	1	30	0.196	0.203	0.212	0.213	0.194	0.200	0.208	0.218
	0.1	1	50	0.141	0.191	0.206	0.215	0.134	0.196	0.201	0.204
	0.1	1	100	0.082	0.116	0.203	0.206	0.081	0.117	0.202	0.205
	0.1	1	200	0.050	0.067	0.116	0.201	0.052	0.069	0.118	0.199
	0.25	0.5	30	0.177	0.187	0.195	0.208	0.178	0.186	0.199	0.203
	0.25	0.5	50	0.128	0.175	0.184	0.197	0.119	0.180	0.189	0.197
	0.25	0.5	100	0.077	0.106	0.182	0.189	0.076	0.108	0.181	0.194
	0.25	0.5	200	0.048	0.063	0.106	0.185	0.046	0.063	0.110	0.188
	0.25	1	30	0.254	0.264	0.272	0.292	0.249	0.260	0.272	0.288
	0.25	1	50	0.173	0.250	0.272	0.284	0.179	0.256	0.270	0.285
	0.25	1	100	0.111	0.159	0.263	0.283	0.108	0.157	0.264	0.275
	0.25	1	200	0.069	0.096	0.164	0.275	0.067	0.094	0.163	0.274
	0.33	0.5	30	0.194	0.203	0.207	0.221	0.196	0.200	0.208	0.223
	0.33	0.5	50	0.136	0.196	0.202	0.209	0.140	0.194	0.199	0.212
	0.33	0.5	100	0.081	0.119	0.197	0.206	0.084	0.121	0.199	0.208
	0.33	0.5	200	0.051	0.071	0.116	0.203	0.052	0.072	0.120	0.198
	0.33	1	30	0.272	0.285	0.310	0.321	0.275	0.292	0.301	0.314
	0.33	1	50	0.203	0.276	0.292	0.310	0.204	0.282	0.294	0.306
	0.33	1	100	0.127	0.179	0.287	0.302	0.123	0.180	0.285	0.304
	0.33	1	200	0.077	0.112	0.184	0.303	0.078	0.112	0.192	0.303
	0.45	0.5	30	0.213	0.216	0.229	0.238	0.207	0.222	0.227	0.240
	0.45	0.5	50	0.157	0.209	0.221	0.220	0.158	0.209	0.225	0.232
	0.45	0.5	100	0.095	0.139	0.213	0.223	0.099	0.140	0.216	0.221
1	0.45	0.5	200	0.061	0.085	0.136	0.221	0.060	0.083	0.141	0.220
	0.45	1	30	0.300	0.314	0.330	0.343	0.297	0.319	0.325	0.343
	0.45	1	50	0.233	0.307	0.322	0.339	0.235	0.307	0.324	0.332
	0.45	1	100	0.149	0.217	0.315	0.326	0.154	0.221	0.313	0.328
	0.45	1	200	0.099	0.138	0.227	0.326	0.096	0.139	0.226	0.326
	0.5	0.5	30	0.217	0.223	0.230	0.238	0.223	0.226	0.237	0.237
	0.5	0.5	50	0.163	0.221	0.223	0.230	0.166	0.214	0.226	0.229
	0.5	0.5	100	0.107	0.150	0.222	0.225	0.107	0.148	0.217	0.229
	0.5	0.5	200	0.067	0.090	0.148	0.223	0.067	0.092	0.149	0.221
	0.5	1	30	0.309	0.329	0.347	0.353	0.306	0.322	0.341	0.344
	0.5	1	50	0.250	0.308	0.331	0.343	0.240	0.313	0.338	0.330
	0.5	1	100	0.170	0.229	0.324	0.343	0.169	0.238	0.312	0.334
	0.5	1	200	0.109	0.154	0.247	0.329	0.109	0.158	0.244	0.329

Table 6. Variance of factor IV estimator

			c_2	0.5				1			
r	p	c_1	T/N	30	50	100	200	30	50	100	200
	0	0.5	30	0.082	0.081	0.084	0.081	0.081	0.079	0.083	0.080
	0	0.5	50	0.063	0.062	0.059	0.060	0.059	0.060	0.059	0.060
	0	0.5	100	0.043	0.041	0.041	0.043	0.043	0.044	0.041	0.041
	0	0.5	200	0.029	0.028	0.029	0.030	0.030	0.029	0.029	0.029
	0	1	30	0.099	0.100	0.098	0.102	0.099	0.102	0.099	0.096
	0	1	50	0.076	0.076	0.073	0.072	0.076	0.073	0.076	0.073
	0	1	100	0.051	0.052	0.052	0.051	0.051	0.051	0.054	0.051
	0	1	200	0.036	0.036	0.037	0.036	0.038	0.035	0.036	0.035
	0.1	0.5	30	0.094	0.098	0.096	0.101	0.092	0.096	0.099	0.100
	0.1	0.5	50	0.068	0.072	0.071	0.076	0.070	0.072	0.074	0.077
	0.1	0.5	100	0.047	0.049	0.050	0.052	0.048	0.050	0.049	0.051
1	0.1	0.5	200	0.034	0.036	0.036	0.036	0.033	0.033	0.035	0.036
	0.1	1	30	0.117	0.122	0.124	0.129	0.120	0.120	0.127	0.131
	0.1	1	50	0.089	0.091	0.093	0.095	0.088	0.089	0.091	0.092
	0.1	1	100	0.059	0.064	0.067	0.065	0.058	0.061	0.062	0.064
	0.1	1	200	0.043	0.044	0.045	0.045	0.042	0.042	0.044	0.044
	0.25	0.5	30	0.110	0.118	0.118	0.124	0.112	0.119	0.116	0.118
	0.25	0.5	50	0.086	0.088	0.090	0.088	0.079	0.086	0.089	0.091
	0.25	0.5	100	0.059	0.057	0.063	0.061	0.057	0.059	0.063	0.060
	0.25	0.5	200	0.040	0.041	0.041	0.041	0.039	0.042	0.040	0.044
	0.25	1	30	0.154	0.162	0.173	0.212	0.146	0.172	0.167	0.174
	0.25	1	50	0.112	0.117	0.119	0.130	0.114	0.111	0.121	0.128
	0.25	1	100	0.075	0.078	0.082	0.087	0.071	0.078	0.081	0.081
	0.25	1	200	0.053	0.054	0.056	0.060	0.051	0.055	0.056	0.058
	0.33	0.5	30	0.129	0.132	0.132	0.134	0.121	0.132	0.134	0.145
	0.33	0.5	50	0.088	0.095	0.099	0.099	0.093	0.094	0.098	0.099
	0.33	0.5	100	0.063	0.064	0.065	0.066	0.062	0.069	0.065	0.065
	0.33	0.5	200	0.042	0.045	0.047	0.046	0.043	0.046	0.044	0.049
	0.33	1	30	0.219	0.339	0.347	0.511	0.213	0.284	0.277	0.446
	0.33	1	50	0.132	0.146	0.142	0.150	0.206	0.135	0.161	0.226
	0.33	1	100	0.083	0.091	0.093	0.095	0.084	0.090	0.091	0.098
	0.33	1	200	0.057	0.061	0.063	0.066	0.058	0.063	0.060	0.067
	0.45	0.5	30	0.342	0.393	0.573	0.778	0.265	0.485	0.429	0.979
	0.45	0.5	50	0.103	0.131	0.253	0.343	0.117	0.198	0.207	0.441
	0.45	0.5	100	0.073	0.078	0.076	0.092	0.076	0.074	0.082	0.089
1	0.45	0.5	200	0.050	0.049	0.053	0.054	0.050	0.050	0.051	0.058
	0.45	1	30	0.747	1.098	1.172	1.528	0.687	1.179	1.456	1.509
	0.45	1	50	0.519	0.752	0.929	1.152	0.549	0.765	0.959	1.421
	0.45	1	100	0.129	0.214	0.366	0.956	0.120	0.175	0.440	0.704
	0.45	1	200	0.074	0.078	0.092	0.214	0.069	0.079	0.089	0.333
	0.5	0.5	30	0.451	0.666	0.990	1.220	0.510	0.759	1.036	1.177
	0.5	0.5	50	0.247	0.307	0.832	0.969	0.335	0.278	0.710	0.974
	0.5	0.5	100	0.078	0.089	0.211	0.375	0.077	0.087	0.372	0.426
	0.5	0.5	200	0.054	0.056	0.060	0.070	0.053	0.054	0.058	0.071
	0.5	1	30	1.061	1.233	1.558	1.808	1.163	1.382	1.569	1.829
	0.5	1	50	0.934	1.121	1.423	1.591	0.858	0.939	1.222	1.685
	0.5	1	100	0.241	0.538	0.879	1.308	0.354	0.667	0.967	1.314
	0.5	1	200	0.083	0.140	0.320	0.853	0.078	0.119	0.575	0.989

Setup of equation (4) with homogeneous factor loadings and with variable preselection

Table 7. Variance of standard IV estimator with preselection						
r	p	T/N	30	50	100	200
	0	30	0.170	0.230	0.263	0.260
	0	50	0.125	0.148	0.251	0.246
	0	100	0.085	0.104	0.141	0.241
	0	200	0.060	0.066	0.083	0.131
	0.1	30	0.182	0.230	0.256	0.254
	0.1	50	0.132	0.161	0.242	0.246
	0.1	100	0.092	0.105	0.148	0.236
	0.1	200	0.061	0.068	0.089	0.141
	0.25	30	0.192	0.239	0.253	0.248
	0.25	50	0.150	0.171	0.250	0.247
	0.25	100	0.099	0.120	0.160	0.241
1	0.25	200	0.071	0.080	0.108	0.155
	0.33	30	0.194	0.236	0.251	0.253
	0.33	50	0.152	0.188	0.244	0.249
	0.33	100	0.110	0.130	0.169	0.243
	0.33	200	0.080	0.090	0.116	0.164
	0.45	30	0.205	0.242	0.253	0.257
	0.45	50	0.171	0.199	0.242	0.250
	0.45	100	0.128	0.150	0.185	0.239
	0.45	200	0.095	0.111	0.136	0.178
	0.5	30	0.219	0.250	0.259	0.254
	0.5	50	0.177	0.203	0.252	0.245
	0.5	100	0.139	0.160	0.191	0.239
	0.5	200	0.098	0.118	0.147	0.190

Table 8. Variance of Factor-IV estimator with preselection						
r	p	T/N	30	50	100	200
	0	30	0.130	0.133	0.135	0.139
	0	50	0.105	0.103	0.109	0.099
	0	100	0.074	0.077	0.072	0.073
	0	200	0.055	0.054	0.053	0.053
	0.1	30	0.140	0.135	0.128	0.136
	0.1	50	0.107	0.107	0.102	0.104
	0.1	100	0.078	0.076	0.076	0.074
	0.1	200	0.054	0.054	0.052	0.055
	0.25	30	0.146	0.147	0.143	0.132
	0.25	50	0.116	0.112	0.117	0.114
	0.25	100	0.080	0.086	0.086	0.084
1	0.25	200	0.061	0.061	0.065	0.066
	0.33	30	0.158	0.156	0.153	0.149
	0.33	50	0.126	0.127	0.121	0.125
	0.33	100	0.092	0.094	0.096	0.097
	0.33	200	0.070	0.067	0.070	0.075
	0.45	30	0.246	0.490	0.281	0.397
	0.45	50	0.218	0.195	0.400	0.274
	0.45	100	0.112	0.112	0.116	0.121
	0.45	200	0.082	0.086	0.087	0.094
	0.5	30	0.417	0.644	0.549	0.817
	0.5	50	0.309	0.344	0.510	0.726
	0.5	100	0.161	0.266	0.374	0.244
	0.5	200	0.088	0.097	0.105	0.157

Setup of equation (3) with weak instruments

Table 9. Variance of standard IV estimator										
		c_2	0.5				1			
r	p	T/N	30	50	100	200	30	50	100	200
	0	30	0.568	0.596	0.626	0.649	0.574	0.581	0.621	0.640
	0	50	0.516	0.590	0.618	0.636	0.505	0.582	0.605	0.645
	0	100	0.393	0.507	0.607	0.631	0.382	0.498	0.607	0.639
	0	200	0.270	0.384	0.539	0.631	0.266	0.384	0.535	0.640
	0.1	30	0.621	0.662	0.675	0.672	0.628	0.650	0.666	0.697
	0.1	50	0.587	0.633	0.664	0.687	0.591	0.656	0.672	0.684
	0.1	100	0.504	0.613	0.671	0.697	0.501	0.601	0.678	0.685
	0.1	200	0.379	0.511	0.641	0.687	0.374	0.514	0.638	0.684
	0.25	30	0.680	0.700	0.719	0.717	0.687	0.693	0.715	0.715
	0.25	50	0.657	0.674	0.684	0.707	0.645	0.708	0.701	0.709
	0.25	100	0.623	0.672	0.698	0.701	0.598	0.690	0.694	0.706
1	0.25	200	0.537	0.632	0.688	0.709	0.527	0.633	0.689	0.707
	0.33	30	0.691	0.710	0.697	0.718	0.690	0.696	0.714	0.714
	0.33	50	0.671	0.696	0.723	0.717	0.685	0.709	0.711	0.718
	0.33	100	0.646	0.702	0.720	0.709	0.633	0.692	0.701	0.698
	0.33	200	0.591	0.662	0.693	0.690	0.584	0.664	0.698	0.690
	0.45	30	0.710	0.728	0.714	0.709	0.704	0.699	0.714	0.720
	0.45	50	0.687	0.705	0.717	0.716	0.704	0.711	0.692	0.730
	0.45	100	0.677	0.711	0.691	0.722	0.654	0.697	0.699	0.706
	0.45	200	0.651	0.678	0.697	0.714	0.619	0.680	0.705	0.710
	0.5	30	0.712	0.711	0.717	0.731	0.697	0.707	0.715	0.716
	0.5	50	0.712	0.711	0.706	0.723	0.704	0.700	0.717	0.699
	0.5	100	0.688	0.686	0.715	0.699	0.678	0.697	0.698	0.709
	0.5	200	0.648	0.696	0.709	0.703	0.662	0.699	0.698	0.717

Table 10. Variance of Factor-IV estimator										
		c_2	0.5				1			
r	p	T/N	30	50	100	200	30	50	100	200
	0	30	0.891	0.872	1.241	1.492	0.796	0.973	1.120	1.169
	0	50	0.518	0.497	0.771	0.848	0.414	0.505	0.933	0.962
	0	100	0.240	0.274	0.342	0.399	0.264	0.253	0.327	0.387
	0	200	0.152	0.165	0.187	0.250	0.152	0.163	0.197	0.253
	0.1	30	1.190	1.358	1.738	1.962	1.224	1.377	1.715	1.910
	0.1	50	0.857	1.169	1.290	1.607	0.794	1.213	1.389	1.606
	0.1	100	0.385	0.609	0.794	1.221	0.395	0.533	0.801	1.149
	0.1	200	0.222	0.287	0.398	0.695	0.210	0.257	0.395	0.800
	0.25	30	1.848	1.877	2.015	2.112	1.885	2.001	2.029	2.125
	0.25	50	1.731	1.713	1.983	2.109	1.798	1.822	2.196	2.015
	0.25	100	1.158	1.303	1.724	1.987	0.929	1.366	1.677	1.974
1	0.25	200	0.543	0.885	1.415	1.702	0.444	0.796	1.279	1.792
	0.33	30	1.902	2.024	2.120	2.139	1.747	2.064	2.183	2.227
	0.33	50	1.822	2.179	2.102	2.060	1.850	2.003	1.974	2.020
	0.33	100	1.299	2.048	1.965	2.082	1.275	1.696	2.129	2.271
	0.33	200	0.876	1.387	1.677	1.963	0.823	1.090	1.756	2.204
	0.45	30	2.302	1.966	2.115	1.946	2.010	2.089	2.155	2.044
	0.45	50	2.067	2.068	2.086	2.059	1.999	2.126	2.071	2.162
	0.45	100	1.722	2.062	1.991	2.214	1.632	2.023	1.846	2.074
	0.45	200	1.372	1.803	1.902	2.135	1.369	1.649	1.969	2.081
	0.5	30	2.124	2.026	2.175	1.952	1.924	2.006	2.002	2.094
	0.5	50	2.082	2.209	2.063	1.979	2.189	2.091	2.163	2.050
	0.5	100	1.979	2.059	2.007	2.199	1.943	2.146	2.028	2.083
	0.5	200	1.581	1.867	2.166	2.069	1.514	1.793	2.041	2.014

Setup of equation (5) with weak instruments

Table 11. Variance of standard IV estimator

			c_2	0.5				1			
r	p	c_1	T/N	30	50	100	200	30	50	100	200
	0	0.5	30	0.171	0.176	0.184	0.191	0.168	0.178	0.184	0.191
	0	0.5	50	0.117	0.168	0.173	0.183	0.114	0.166	0.175	0.181
	0	0.5	100	0.072	0.100	0.173	0.180	0.071	0.098	0.173	0.179
	0	0.5	200	0.045	0.056	0.098	0.175	0.044	0.057	0.099	0.176
	0	1	30	0.235	0.246	0.252	0.272	0.236	0.243	0.255	0.269
	0	1	50	0.166	0.239	0.255	0.266	0.162	0.234	0.250	0.265
	0	1	100	0.099	0.146	0.246	0.270	0.101	0.143	0.244	0.258
	0	1	200	0.060	0.085	0.148	0.256	0.060	0.082	0.148	0.255
	0.1	0.5	30	0.193	0.196	0.211	0.214	0.190	0.193	0.207	0.219
	0.1	0.5	50	0.132	0.186	0.199	0.208	0.130	0.191	0.199	0.205
	0.1	0.5	100	0.079	0.113	0.187	0.202	0.082	0.112	0.195	0.203
1	0.1	0.5	200	0.050	0.065	0.111	0.198	0.050	0.065	0.116	0.199
	0.1	1	30	0.262	0.275	0.296	0.306	0.263	0.280	0.291	0.313
	0.1	1	50	0.185	0.269	0.285	0.302	0.187	0.272	0.287	0.295
	0.1	1	100	0.116	0.168	0.289	0.295	0.118	0.169	0.284	0.293
	0.1	1	200	0.069	0.102	0.173	0.291	0.072	0.099	0.176	0.291
	0.25	0.5	30	0.213	0.223	0.220	0.238	0.211	0.220	0.225	0.238
	0.25	0.5	50	0.152	0.210	0.216	0.229	0.151	0.215	0.222	0.224
	0.25	0.5	100	0.092	0.130	0.219	0.226	0.092	0.129	0.215	0.224
	0.25	0.5	200	0.056	0.076	0.129	0.215	0.056	0.077	0.130	0.215
	0.25	1	30	0.309	0.314	0.331	0.345	0.301	0.314	0.327	0.340
	0.25	1	50	0.219	0.314	0.320	0.335	0.220	0.308	0.320	0.328
	0.25	1	100	0.141	0.202	0.320	0.325	0.142	0.200	0.321	0.331
	0.25	1	200	0.086	0.124	0.206	0.330	0.088	0.123	0.207	0.324
	0.33	0.5	30	0.226	0.228	0.236	0.242	0.219	0.227	0.238	0.244
	0.33	0.5	50	0.162	0.229	0.226	0.235	0.158	0.222	0.230	0.234
	0.33	0.5	100	0.096	0.139	0.219	0.232	0.099	0.138	0.221	0.230
	0.33	0.5	200	0.063	0.082	0.137	0.226	0.062	0.081	0.140	0.228
	0.33	1	30	0.317	0.335	0.335	0.348	0.318	0.332	0.347	0.354
	0.33	1	50	0.243	0.323	0.328	0.353	0.240	0.319	0.336	0.337
	0.33	1	100	0.156	0.220	0.331	0.337	0.157	0.218	0.332	0.335
	0.33	1	200	0.099	0.136	0.224	0.340	0.100	0.137	0.229	0.336
	0.45	0.5	30	0.233	0.239	0.245	0.247	0.231	0.238	0.246	0.247
	0.45	0.5	50	0.172	0.230	0.235	0.241	0.171	0.225	0.242	0.241
	0.45	0.5	100	0.111	0.154	0.232	0.233	0.110	0.152	0.231	0.239
1	0.45	0.5	200	0.067	0.094	0.155	0.233	0.070	0.093	0.154	0.232
	0.45	1	30	0.335	0.343	0.356	0.367	0.327	0.343	0.362	0.366
	0.45	1	50	0.269	0.337	0.339	0.361	0.275	0.330	0.345	0.348
	0.45	1	100	0.182	0.245	0.345	0.342	0.185	0.251	0.342	0.343
	0.45	1	200	0.118	0.171	0.256	0.344	0.116	0.169	0.254	0.347
	0.5	0.5	30	0.236	0.240	0.242	0.250	0.236	0.243	0.249	0.249
	0.5	0.5	50	0.179	0.236	0.239	0.236	0.185	0.232	0.239	0.239
	0.5	0.5	100	0.115	0.163	0.232	0.238	0.116	0.163	0.237	0.236
	0.5	0.5	200	0.074	0.102	0.160	0.232	0.072	0.101	0.161	0.235
	0.5	1	30	0.337	0.346	0.359	0.365	0.338	0.355	0.357	0.362
	0.5	1	50	0.278	0.337	0.343	0.350	0.281	0.340	0.353	0.357
	0.5	1	100	0.200	0.262	0.349	0.349	0.203	0.262	0.349	0.346
	0.5	1	200	0.131	0.185	0.268	0.352	0.133	0.187	0.267	0.349

Table 12. Variance of factor IV estimator

			c_2	0.5				1			
r	p	c_1	T/N	30	50	100	200	30	50	100	200
	0	0.5	30	0.109	0.110	0.112	0.112	0.097	0.105	0.106	0.119
	0	0.5	50	0.082	0.078	0.084	0.087	0.077	0.078	0.085	0.087
	0	0.5	100	0.053	0.055	0.055	0.059	0.053	0.056	0.057	0.059
	0	0.5	200	0.039	0.039	0.039	0.039	0.038	0.037	0.040	0.040
	0	1	30	0.139	0.139	0.150	0.148	0.138	0.135	0.152	0.157
	0	1	50	0.101	0.100	0.107	0.113	0.100	0.104	0.107	0.109
	0	1	100	0.068	0.069	0.072	0.078	0.072	0.070	0.072	0.075
	0	1	200	0.046	0.049	0.053	0.053	0.048	0.050	0.053	0.054
	0.1	0.5	30	0.122	0.118	0.119	0.121	0.117	0.120	0.123	0.127
	0.1	0.5	50	0.085	0.090	0.093	0.092	0.084	0.090	0.090	0.092
	0.1	0.5	100	0.058	0.064	0.062	0.064	0.059	0.059	0.060	0.062
1	0.1	0.5	200	0.041	0.043	0.044	0.045	0.041	0.041	0.045	0.044
	0.1	1	30	0.157	0.156	0.173	0.177	0.154	0.155	0.169	0.173
	0.1	1	50	0.113	0.122	0.122	0.126	0.117	0.120	0.129	0.128
	0.1	1	100	0.074	0.081	0.085	0.085	0.078	0.083	0.082	0.089
	0.1	1	200	0.051	0.056	0.057	0.060	0.053	0.057	0.059	0.061
	0.25	0.5	30	0.127	0.130	0.128	0.138	0.126	0.134	0.134	0.141
	0.25	0.5	50	0.094	0.100	0.101	0.101	0.099	0.098	0.104	0.099
	0.25	0.5	100	0.065	0.066	0.071	0.070	0.067	0.069	0.069	0.071
	0.25	0.5	200	0.045	0.044	0.046	0.050	0.045	0.046	0.048	0.049
	0.25	1	30	0.196	0.198	0.208	0.201	0.204	0.198	0.262	0.230
	0.25	1	50	0.135	0.134	0.134	0.148	0.138	0.138	0.143	0.153
	0.25	1	100	0.091	0.091	0.094	0.097	0.090	0.095	0.097	0.097
	0.25	1	200	0.061	0.066	0.067	0.065	0.062	0.065	0.069	0.069
	0.33	0.5	30	0.140	0.145	0.146	0.158	0.144	0.144	0.149	0.156
	0.33	0.5	50	0.100	0.104	0.107	0.108	0.100	0.108	0.106	0.107
	0.33	0.5	100	0.068	0.069	0.071	0.073	0.068	0.072	0.070	0.076
	0.33	0.5	200	0.049	0.049	0.048	0.049	0.050	0.049	0.049	0.051
	0.33	1	30	0.297	0.366	0.344	0.579	0.445	0.309	0.351	0.439
	0.33	1	50	0.200	0.161	0.171	0.185	0.154	0.166	0.171	0.183
	0.33	1	100	0.099	0.108	0.106	0.110	0.103	0.107	0.107	0.103
	0.33	1	200	0.068	0.070	0.073	0.074	0.069	0.070	0.070	0.076
	0.45	0.5	30	0.281	0.405	0.557	0.911	0.229	0.378	0.604	0.923
	0.45	0.5	50	0.125	0.153	0.241	0.511	0.125	0.134	0.194	0.339
	0.45	0.5	100	0.081	0.086	0.087	0.090	0.079	0.080	0.083	0.093
1	0.45	0.5	200	0.054	0.056	0.060	0.058	0.055	0.056	0.057	0.060
	0.45	1	30	0.755	1.279	1.306	1.412	0.892	1.126	1.410	1.549
	0.45	1	50	0.516	0.645	0.925	1.166	0.603	0.826	0.920	1.242
	0.45	1	100	0.135	0.188	0.435	0.852	0.138	0.407	0.383	0.698
	0.45	1	200	0.085	0.088	0.104	0.315	0.085	0.094	0.097	0.217
	0.5	0.5	30	0.531	0.792	0.963	1.301	0.552	0.741	1.066	1.298
	0.5	0.5	50	0.211	0.579	0.635	0.979	0.180	0.418	0.612	0.875
	0.5	0.5	100	0.084	0.091	0.115	0.420	0.086	0.093	0.158	0.417
	0.5	0.5	200	0.058	0.061	0.063	0.091	0.058	0.059	0.065	0.078
	0.5	1	30	1.259	1.425	1.634	1.758	1.133	1.557	1.519	1.614
	0.5	1	50	0.916	1.302	1.520	1.571	0.961	1.032	1.445	1.634
	0.5	1	100	0.457	0.504	1.073	1.496	0.250	0.690	1.115	1.598
	0.5	1	200	0.115	0.169	0.416	0.843	0.138	0.144	0.573	0.811

Setup of equation (4) with serial correlation in ϵ_t

Table 13. Variance of standard GMM estimator when ϵ_t is an AR process										
		c_1	0.5				1			
r	p	T/N	30	50	100	200	30	50	100	200
	0	30	0.260	0.260	0.254	0.256	0.365	0.364	0.369	0.360
	0	50	0.187	0.247	0.249	0.252	0.282	0.356	0.378	0.358
	0	100	0.116	0.154	0.245	0.254	0.184	0.244	0.361	0.362
	0	200	0.073	0.093	0.150	0.240	0.118	0.154	0.244	0.350
	0.1	30	0.258	0.253	0.256	0.254	0.373	0.363	0.365	0.368
	0.1	50	0.187	0.242	0.255	0.246	0.286	0.368	0.362	0.369
	0.1	100	0.115	0.155	0.243	0.239	0.180	0.250	0.355	0.357
	0.1	200	0.075	0.094	0.148	0.236	0.119	0.156	0.240	0.353
	0.25	30	0.257	0.255	0.258	0.256	0.369	0.380	0.369	0.365
	0.25	50	0.191	0.246	0.253	0.245	0.295	0.364	0.355	0.357
	0.25	100	0.125	0.162	0.243	0.239	0.192	0.256	0.359	0.359
1	0.25	200	0.078	0.100	0.156	0.234	0.126	0.165	0.247	0.354
	0.33	30	0.247	0.250	0.251	0.248	0.373	0.359	0.372	0.363
	0.33	50	0.191	0.244	0.246	0.254	0.307	0.364	0.367	0.363
	0.33	100	0.128	0.164	0.240	0.238	0.206	0.261	0.355	0.353
	0.33	200	0.083	0.104	0.162	0.234	0.133	0.172	0.260	0.353
	0.45	30	0.255	0.255	0.255	0.254	0.377	0.367	0.368	0.370
	0.45	50	0.205	0.245	0.246	0.247	0.320	0.360	0.369	0.352
	0.45	100	0.147	0.179	0.241	0.240	0.227	0.288	0.355	0.349
	0.45	200	0.098	0.117	0.174	0.238	0.157	0.196	0.271	0.354
	0.5	30	0.260	0.253	0.253	0.254	0.377	0.372	0.375	0.362
	0.5	50	0.215	0.249	0.249	0.250	0.315	0.362	0.370	0.362
	0.5	100	0.153	0.190	0.235	0.245	0.242	0.289	0.359	0.361
	0.5	200	0.103	0.127	0.183	0.237	0.165	0.208	0.278	0.353

Table 14. Variance of Factor-GMM estimator when ϵ_t is an AR process										
		c_1	0.5				1			
r	p	T/N	30	50	100	200	30	50	100	200
	0	30	0.164	0.157	0.161	0.158	0.262	0.238	0.237	0.230
	0	50	0.120	0.125	0.116	0.118	0.179	0.172	0.178	0.173
	0	100	0.087	0.082	0.081	0.082	0.119	0.121	0.121	0.115
	0	200	0.060	0.059	0.059	0.058	0.085	0.084	0.085	0.084
	0.1	30	0.160	0.162	0.160	0.157	0.267	0.246	0.253	0.269
	0.1	50	0.123	0.119	0.119	0.116	0.185	0.174	0.178	0.176
	0.1	100	0.087	0.083	0.080	0.082	0.126	0.123	0.121	0.121
	0.1	200	0.060	0.060	0.058	0.059	0.087	0.087	0.085	0.082
	0.25	30	0.194	0.186	0.169	0.167	0.340	0.323	0.350	0.248
	0.25	50	0.133	0.128	0.131	0.121	0.189	0.199	0.183	0.189
	0.25	100	0.090	0.090	0.088	0.084	0.134	0.128	0.128	0.124
1	0.25	200	0.060	0.061	0.062	0.065	0.092	0.091	0.089	0.084
	0.33	30	0.275	0.252	0.399	0.494	0.392	0.374	0.387	0.508
	0.33	50	0.153	0.144	0.144	0.154	0.229	0.211	0.199	0.240
	0.33	100	0.096	0.090	0.093	0.090	0.134	0.138	0.136	0.135
	0.33	200	0.068	0.063	0.065	0.061	0.094	0.099	0.093	0.093
	0.45	30	0.999	1.175	1.264	1.542	1.091	1.284	1.392	1.715
	0.45	50	0.506	0.793	0.971	1.140	0.568	0.892	1.221	1.463
	0.45	100	0.124	0.295	0.528	0.661	0.201	0.434	0.556	0.809
	0.45	200	0.081	0.080	0.085	0.115	0.114	0.122	0.122	0.301
	0.5	30	1.035	1.391	1.431	1.691	1.422	1.455	1.792	2.000
	0.5	50	0.894	1.213	1.329	1.469	1.021	1.333	1.554	1.865
	0.5	100	0.411	0.355	0.872	1.349	0.424	0.691	1.170	1.522
	0.5	200	0.093	0.109	0.448	0.777	0.136	0.168	0.405	0.996

Table 15. Variance of standard GMM estimator when ϵ_t is an MA process										
		c_1	0.5				1			
r	p	T/N	30	50	100	200	30	50	100	200
	0	30	0.264	0.258	0.264	0.262	0.376	0.383	0.385	0.369
	0	50	0.188	0.249	0.256	0.259	0.291	0.369	0.373	0.367
	0	100	0.118	0.158	0.245	0.250	0.188	0.247	0.359	0.369
	0	200	0.074	0.094	0.147	0.240	0.114	0.154	0.242	0.353
	0.1	30	0.262	0.266	0.260	0.264	0.380	0.372	0.381	0.375
	0.1	50	0.186	0.250	0.256	0.249	0.290	0.370	0.361	0.371
	0.1	100	0.117	0.157	0.243	0.244	0.189	0.253	0.363	0.359
	0.1	200	0.074	0.096	0.150	0.235	0.121	0.159	0.246	0.363
	0.25	30	0.265	0.254	0.261	0.265	0.379	0.372	0.379	0.385
	0.25	50	0.197	0.254	0.250	0.247	0.295	0.364	0.373	0.362
	0.25	100	0.125	0.160	0.243	0.244	0.202	0.258	0.364	0.356
1	0.25	200	0.079	0.102	0.154	0.239	0.125	0.162	0.254	0.351
	0.33	30	0.264	0.260	0.261	0.258	0.382	0.378	0.386	0.378
	0.33	50	0.200	0.251	0.255	0.251	0.308	0.371	0.370	0.371
	0.33	100	0.131	0.166	0.244	0.240	0.207	0.267	0.366	0.363
	0.33	200	0.084	0.103	0.157	0.241	0.136	0.172	0.250	0.353
	0.45	30	0.259	0.267	0.270	0.264	0.375	0.378	0.379	0.386
	0.45	50	0.213	0.246	0.253	0.247	0.309	0.366	0.369	0.363
	0.45	100	0.149	0.181	0.248	0.241	0.233	0.286	0.361	0.359
	0.45	200	0.096	0.122	0.174	0.237	0.157	0.192	0.274	0.362
	0.5	30	0.259	0.262	0.265	0.264	0.381	0.381	0.381	0.381
	0.5	50	0.224	0.248	0.248	0.252	0.322	0.368	0.369	0.371
	0.5	100	0.154	0.187	0.245	0.244	0.243	0.296	0.361	0.365
	0.5	200	0.111	0.130	0.185	0.245	0.167	0.216	0.288	0.357

Table 16. Variance of Factor-GMM estimator when ϵ_t is an MA process										
		c_1	0.5				1			
r	p	T/N	30	50	100	200	30	50	100	200
	0	30	0.156	0.154	0.156	0.159	0.235	0.255	0.242	0.232
	0	50	0.119	0.124	0.120	0.128	0.182	0.186	0.178	0.168
	0	100	0.085	0.083	0.083	0.087	0.124	0.115	0.117	0.120
	0	200	0.058	0.060	0.061	0.060	0.088	0.083	0.085	0.082
	0.1	30	0.165	0.163	0.161	0.166	0.267	0.243	0.242	0.237
	0.1	50	0.128	0.122	0.122	0.121	0.181	0.185	0.179	0.173
	0.1	100	0.083	0.087	0.086	0.085	0.124	0.121	0.118	0.125
	0.1	200	0.056	0.058	0.060	0.056	0.084	0.087	0.087	0.084
	0.25	30	0.192	0.180	0.174	0.170	0.295	0.294	0.284	0.263
	0.25	50	0.133	0.129	0.127	0.134	0.205	0.191	0.184	0.183
	0.25	100	0.093	0.087	0.087	0.086	0.131	0.131	0.128	0.127
1	0.25	200	0.063	0.064	0.061	0.062	0.089	0.088	0.087	0.085
	0.33	30	0.210	0.413	0.352	0.276	0.552	0.423	0.504	0.554
	0.33	50	0.299	0.144	0.141	0.147	0.239	0.217	0.209	0.222
	0.33	100	0.097	0.097	0.092	0.094	0.145	0.140	0.134	0.133
	0.33	200	0.067	0.064	0.064	0.065	0.097	0.094	0.092	0.092
	0.45	30	0.904	0.851	1.265	1.616	1.027	1.389	1.753	1.682
	0.45	50	0.469	0.606	0.936	0.973	0.764	0.737	1.282	1.485
	0.45	100	0.128	0.239	0.515	0.630	0.181	0.293	0.533	0.961
	0.45	200	0.079	0.079	0.088	0.123	0.118	0.117	0.179	0.235
	0.5	30	1.238	1.413	1.580	1.627	1.405	1.602	1.744	1.913
	0.5	50	0.808	0.916	1.273	1.593	0.969	1.233	1.453	1.713
	0.5	100	0.351	0.614	0.858	1.289	0.378	0.657	1.078	1.482
	0.5	200	0.131	0.377	0.590	0.939	0.127	0.239	0.518	1.011

Empirical Results

Table 17. Results for Taylor rule										
								First stage regression (infl+12)		
		ρ	γ	β	**R2-adj**	**S.E. regr**	**Pval J-stat**	**R2-adj**	**S.E. regr**	**Pval F-stat**
Base		0.883	0.993	2.310	0.98	0.27	0.11	0.12	0.002	
	st. err	0.037	0.241	0.278						
Factors		0.908	1.261	2.905	0.98	0.27	0.13	0.15	0.002	0.05
All data	st. err	0.024	0.291	0.394						
Factors		0.929	1.291	3.346	0.98	0.26	0.14	0.18	0.002	0.01
Split data	st. err	0.021	0.335	0.567						
Factors		0.884	1.122	2.251	0.98	0.27	0.52	0.15	0.002	0.08
Select data	st. err	0.028	0.233	0.204						

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 "Split data" the factors are extracted from separate datasets for nominal (2), real (8) and financial variables (2); 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 each dataset, except "Select data" where it is set to one. We use one lag of each factor, but 12 lags for the Select data factor. The last three columns contain statistics related to the first-stage regression of the one-year ahead expected inflation on the set of instruments used. In particular, we report the adjusted R^2, the standard error of the regression and the F-test for the joint significance of the coefficients on factors, when factors are added to the baseline model.

Table 18. Results for New Keynesian Phillips curve										
								First stage regression (infl+1)		
		α	γ	ρ	**R2-adj**	**S.E. regr**	**Pval J-stat**	**R2-adj**	**S.E. regr**	**Pval F-stat**
Base		-0.002	0.538	0.462	0.98	0.16	0.62	0.12	0.002	
	st. err	0.007	0.048	0.047						
Factors		0.000	0.513	0.492	0.98	0.16	0.30	0.11	0.002	0.48
All data	st. err	0.006	0.038	0.038						
Factors		0.003	0.527	0.478	0.98	0.16	0.28	0.14	0.002	0.25
Split data	st. err	0.006	0.038	0.037						
Factors		-0.002	0.500	0.509	0.98	0.15	0.12	0.23	0.002	0.00
Select data	st. err	0.006	0.021	0.020						

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 (first lag of the) SW factors are added to the instruments. In particular, in "All data" the (8) factors are extracted from the whole dataset; in "Split data" the factors are extracted from separate datasets for nominal (2), real (8) and financial variables (2); in "Select data" the (1) factor is 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 each dataset, except "Select data" where it is set to one. We use one lag of each factor, but 12 lags for the Select data factor. The last three columns contain statistics related to the first-stage regression of the one-month ahead expected inflation on the set of instruments used. In particular, we report the adjusted R^2, the standard error of the regression and the F-test for the joint significance of the coefficients on factors, when factors are added to the baseline model.