Factor IV Estimation in Conditional Moment Models with an application to Inflation Dynamics

Bertille Antoine[1] and Xiaolin Sun[2]

[1]Simon Fraser University, bertille_antoine@sfu.ca
[2]Monash University, xiaolin.sun@monash.edu

January 16, 2024

Abstract

In a conditional moment model, we develop a new integrated conditional moment (ICM) estimator which directly exploits factor-based conditional moment restrictions without having to first parametrize, or estimate such restrictions. We focus on a time series framework where the large number of available instruments and associated lags is driven by a relatively small number of unobserved factors. We build on the ICM principle originally proposed by Bierens (1982) and combine it with information reduction methods to handle the large number of potential instruments which may exceed the sample size. Under the maintained validity of the true factors, but not that of observed instruments, and standard regularity assumptions, our estimator is consistent, asymptotically normally distributed, and easy to compute. In our simulation studies, we document its reliability and power in cases where the underlying relationship between the endogenous variables and the instruments may be heterogeneous, non-linear, or even unstable over time. Our estimation of the New Keynesian Phillips curve with US data reveals that forward- and backward-looking behaviors are quantitatively equally as important, while the driver's role is nil.

Keywords: Endogeneity; Conditional mean independence; Dimension reduction; Nonlinearity; Instability.

JEL Classification: C13; C12.

1 Introduction

In many econometric models with endogenous variables, the structural parameters of interest are identified through (conditional) moment restrictions. Their informativeness often depends on the quality of available instruments, and, in practice, it can be quite challenging to find *good* informative instruments from observed data. Such difficulties have been at the heart of IV-based econometrics since the early 1990s, and various alternatives are now available: some are identification-robust methods which account - and correct - for the possibility of less informative (so-called *weak*) instruments, while others exploit additional sources of information in order to *improve* the quality of the instrument. Even though progress has been made, one important question remains open, and concerns issues arising from the number of considered instruments - particularly in time series frameworks where lags of observed variables often serve as valid instruments. Since commonly used economic models rarely provide guidance for instrument choice, the number of instruments used in empirical studies can be much larger than the number of instrumented variables and, sometimes, quite large relative to the sample size. This practice uses up degrees of freedom, which is likely to cause size distortions and/or power losses. In this paper, we consider this problem from a *conditional* moment perspective, and rely on information reduction methods - including principal components and factor analysis - to get around it, while maintaining (strong) identification of the parameters of interest.

More specifically, we contribute to the second stream of above-mentioned literature by developing a flexible and convenient alternative to standard estimators such as 2SLS or GMM which directly exploits all the informational content of (factor-based) conditional moment restrictions without having to first either parametrize, or estimate such restrictions. We build on the ICM principle originally proposed by Bierens (1982) and combine it with information reduction methods to handle the large number of potential instruments and associated lags. We focus on a time series framework where the large number of available instruments (which may exceed the sample size) is driven by a relatively small number of unobserved factors. It is important to mention that the validity of the instruments is not maintained; rather, it is only the validity of the (unobserved) true factors which is required. Since our approach does not need to specify, characterize, or estimate the relationship between the endogenous variable and the instruments, we are especially interested in studying - and documenting - the reliability and power of our approach when such a relationship may be heterogenous, non-linear, or even unstable over time. Overall, our factor-based estimator is easy to compute and asymptotically normally distributed under standard regularity assumptions.

ICM-based estimation (see e.g. Dominguez and Lobato (2004), Lavergne and Patilea (2013), Antoine and Lavergne (2014), Escanciano (2018), Antoine and Sun (2022)) is appealing because it remains valid - that is, associated estimators are consistent - under a weaker condition than that of standard IV-estimation: namely, conditional mean independence, rather than uncorrelatedness, which is directly exploited without relying on its parametrization. This is in contrast with standard inference procedures such as 2SLS which often build on a linear first-stage: such a linear first-stage may *artificially* appear weak if the underlying relationship between the endogenous variable and the instrument(s) is non-linear. For further discussions - and numerical illustrations - on the potential threat to the relevance of standard IV-estimation (including 2SLS) associated with an incorrect, or inappropriate functional form for the first-stage equation, see Antoine and Lavergne (2023) and Tsyawo (2023); see also Experiment #2c) in our Monte-Carlo section. In this paper, we build on the smooth minimum distance (SMD) estimator of Lavergne and Patilea (2013) developed under the i.i.d. setup and extend their approach to the time series framework.

Information reduction methods including principal components and factor analysis are not only popular and convenient, but they have also been shown to improve standard IV methods in economics - including the 2SLS estimator, especially with time series and small samples, as occurs, for example, in macroeconomics: see e.g. Bai and Ng (2010), Kapetanios and Marcellino (2010), and references therein; see also the recent survey by Mikusheva (2021). We demonstrate that the same holds for ICM-based estimators. To do so, we follow Bai and Ng (2010), and rely on factor models as a tool for constructing a relatively small number of *higher quality instruments*. We assume that the (large) number of available instruments depends on a small number of true (unobserved) factors. The validity of the true factors is maintained throughout, but not that of observed instruments. Importantly, in our conditional moment framework, validity of the true factors means that the conditional mean of the error term on the factors is zero.

Our work is also related to alternative strategies that have been proposed in the literature to handle issues related to large dimensions (e.g. when using a large number of covariates, moment restrictions, or conditioning variables) such as regularization or penalization. On regularization, our work relates to Carrasco (2012) who propose an original approach based on regularized 2SLS to solve the problem of many instruments: several regularization schemes (such as Tikhonov and Principal Component Analysis) are considered. On penalization, our work relates to the inference procedure recently proposed by Chen et al. (2022) for parameters identified by conditional moments: it is designed to handle a large number of conditioning variables through a penalized Bierens maximum statistic, Bierens (1990). Our estimation procedure does not involve any penalty since we rely instead on information reduction methods.

Overall, in a linear framework that is realistic and well-suited for time series applications, we propose a new convenient and flexible alternative to standard estimators such as 2SLS or GMM. Our estimator exploits all the information contained in conditional moments based on instruments, allows for many available candidate instruments, and bypasses the characterization of the first-stage regression by not having to model it, linearly or otherwise. We also consider extensions to allow for invalid and/or weak instruments while maintaining the validity and relevance of the underlying (latent) factors.

In a series of simulation studies, we document the reliability and power of our proposed estimator in cases where the underlying relationship between the endogenous variables and the instruments may be heterogenous, non-linear, or even unstable over time. Finally, we revisit an important tool in recent monetary policy analysis, the New Keynesian Phillips Curve (NKPC) which explains inflation dynamics through the relation between expected inflation and marginal cost. Our empirical analysis with quarterly US data from 1960 to 2022 provides strong support for the hybrid NKPC introduced by Gali and Gertler (1999). In addition, our estimation results are relatively stable over time and quite precise. They reveal that forward- and backward-looking behaviors are quantitatively equally as important, while the driver's role is nil. Our empirical analysis also explores macro-finance linkages by augmenting traditional sets of (macro) instruments with macro-finance variables. Overall, our results remain quite similar with and without these additional variables.

Our paper is organized as follows. In section 2, we introduce and motivate our framework. In section 3, we present the asymptotic properties of our factor-based ICM estimator. In section 4, we generalize our framework to allow for weak and/or invalid instruments. In section 5, we illustrate its finite sample properties and compare its performance to standard IV estimators such as 2SLS and GMM. Our main empirical analysis of the NKPC with US data is conducted in section 6. Proofs, tables of results and graphs are collected in the Appendix. Additional empirical results are also presented in a Supplementary Appendix.

2 Framework and Motivation

We consider the (standard) linear regression model[1] with scalar dependent variable y_t and p endogenous variables Y_t,

$$y_t = Y_t'\beta_0 + u_t\,, \tag{1}$$

where β_0 is the unknown vector of p parameters of interest. We are interested in estimating β_0, and we rely on a vector W_t of weakly exogenous instruments that may include lags of the dependent variable as well as other exogenous variables such that,

$$E(u_t|\mathcal{I}(W_t)) = 0 \quad \text{with probability 1 (hereafter w.p. 1),} \tag{2}$$

where $\mathcal{I}(W_t)$ denotes the information set available at time t, that is the sigma-algebra generated by W_t and its lags. In such a framework, it is standard to derive unconditional moment restrictions from (2) using a matrix of instruments[2], say $A[\mathcal{I}(W_t)]$, and to estimate β_0 by GMM based on the following moment restrictions

$$E(A[\mathcal{I}(W_t)]u_t) = 0\,. \tag{3}$$

Under maintained homoskedasticity, one may even rely on a linear reduced form equation to explicitly - and parametrically - link the endogenous variables to (some of) the instruments such as,

$$Y_t = \Pi W_t + V_t \quad \text{with } E(W_t V_t) = 0\,, \tag{4}$$

and estimate β_0 by 2SLS.

In this paper, we develop an alternative estimation strategy which aims at directly using the informational content of (2) without having to, either discard any information, as done in (3), or rely on the parametrization and estimation of a "first-stage" equation, such as (4). To do so, we adapt and combine two approaches. First, to handle the large number of candidate instrumental variables, we extend the factor-based IV regression model which offers a convenient and parsimonious description of the cross-series dependence between instruments: see e.g. Kapetanios et al. (2016) and Mikusheva (2021). Specifically, while the instruments are assumed to be driven

[1]For simplicity and ease of exposition, we abstract - for now - of the presence of additional (exogenous) regressors that enter linearly in (1) and may be partialled out.

[2]Specifically, a matrix of instruments is built by taking measurable functions of the information set.

by a small number of common unobserved factors denoted F_t, we neither restrict, nor estimate the relationship between the endogenous variables and these common factors: said differently, the conditional mean of the endogenous variables on the factors, $E[Y_t|\mathcal{I}(F_t)]$, is not "modelled" - either parametrically (e.g. linear) or nonparametrically - as we do not aim to estimate it. Instead, we rather adapt an original idea from Bierens (1982) (see also de Jong (1996) and Bierens and Ploberger (1997) for time series extensions), and exploit the conditional mean independence of the factors by rewriting (2) as an equivalent continuum of unconditional moment restrictions based on the (complex) exponential function. Overall, our factor-IV framework can be written as:

$$y_t = Y_t'\beta_0 + u_t \,, \tag{5}$$

$$W_t = \Lambda F_t + E_t \,, \tag{6}$$

with F_t vector of k unobservable and independent factors, W_t vector of w_q (observed) instruments, and Λ the (w_q, k)-matrix of factor loadings. In our flexible framework, we do not explicitly model $E(Y_t|\mathcal{I}(F_t)]$ since we are not interested in estimating it, or characterizing it either. All we rely on is the conditional mean independence of the error term u_t with respect to the information set based on the factors F_t,

$$E[u_t|\mathcal{I}(F_t)] = 0 \text{ w.p. } 1 \,, \tag{7}$$

under the maintained assumption that $E(Y_t|\mathcal{I}(F_t)]$ is not almost surely 0,

$$E(Y_t|\mathcal{I}(F_t)] \neq 0 \text{ w.p. } 1 \,. \tag{8}$$

Notice also that, similar to Kapetanios et al. (2016), the validity of the instruments W_t is not maintained: it is only the validity of the (unobserved) true factors F_t which is required instead of the standard one (see e.g. (2) above). Finally, it is important to mention that our framework is general enough to accommodate two interesting sub-cases: (i) the case where the number of instruments w_q exceeds the number of observations T - as long as k remains small; (ii) the *hybrid* case where the factors F_t may combine a small number of (chosen) instruments with (true) latent factors[3].

[3]See Section 6 for some illustrations.

Accordingly, equation (6) would then be rewritten as,

$$W_t \equiv \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix} = \Lambda F_t + E_t \quad \text{with} \quad \Lambda = \begin{pmatrix} I & 0 \\ 0 & \tilde{\Lambda} \end{pmatrix} \quad \text{and} \quad F_t \equiv \begin{pmatrix} W_{1,t} \\ \tilde{F}_t \end{pmatrix} .$$

To directly use the informational content of the above-mentioned conditional mean independence, we rewrite (7) as an equivalent continuum of unconditional moments indexed by ξ,

$$E\left[u_t \exp\left(\iota \sum_{j=0}^{c} \xi_j' F_{t-j}\right)\right] = 0 \,, \tag{9}$$

where $\xi \in \Xi$ some compact subset of $\mathbb{R}^k$, and c some positive (finite) constant. Beyond the complex exponential, other functions have been used: in time series, see de Jong (1996) and Bierens and Ploberger (1997) who rely on the real (non-complex) exponential function; see also Stinchcombe and White (1998) for a characterization of a large class of suitable functions in the i.i.d framework. The main idea is then to combine the above continuum of restrictions into a single theoretical criterion, uniquely minimized at β_0, and convenient to compute.

Accordingly, our estimator is defined as the minimizer of

$$\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1,\ s \neq t}^{T} u_s(\beta) u_t(\beta) K\left(\frac{F_t - F_s}{h}\right) \quad \text{where } u_t(\beta) \equiv y_t - Y_t' \beta \ \forall \ t \,, \tag{10}$$

with $K(.)$ a kernel function defined on $\mathbb{R}^k$ such that

$$K(f_t) = \int_{\Xi} \exp\left(\iota \sum_{j=0}^{c} \xi_j' f_{t-j}\right) d\mu(\xi) \,, \tag{11}$$

for some strictly positive measure μ (except possibly for a set of isolated points), and h some positive (bandwidth) parameter. Our estimation procedure has a *built-in* way to capture past information through the kernel function which incorporates up to c lags of the factors[4]. In the i.i.d. framework - with $c = 0$ and observed F_t - the estimator

[4]As a result, a given factor should only enter the (conditioning) information set once - say as its current value, and not *multiple times* as its current and lagged values. Alternatively, lags of factors can be directly incorporated in the conditioning set, say (F_t, F_{t-1}), after adjusting the constant c that enters the kernel function (e.g. by setting $c = 0$). We can actually show that these two implementations

that minimizes (10) was introduced by Lavergne and Patilea (2013) who motivate it by the following equality,

$$E\left[u_s(\beta)u_t(\beta)K\left(\frac{F_t - F_s}{h}\right)\right] = \int_\Xi |E[u_t(\beta)\exp\left(\iota\sum_{j=0}^{c}\xi_j' F_{t-j}\right)]|^2 d\mu(\xi)\,,$$

where (y_s, Y_s, W_s, F_s) is an independent copy of (y_t, Y_t, W_t, F_t), and the observation that the objective function on the left-hand side is more convenient to handle as it avoids computing the derivative of the norm of a complex function.

As pointed out by Antoine et al. (2020), the above equality does not usually hold in general time series models, and they suggest combining the Law of Iterated Expectations with additional regularity assumptions that pertains to the exogeneity of the factors and the dynamics of the error terms (see Assumption 1(iv) below) to ensure

$$\begin{aligned} M_\infty(\beta) &\equiv E\left[u_s(\beta)u_t(\beta)K\left(\frac{F_t - F_s}{h}\right)\right] \\ &= E\left[E\left[u_s(\beta)|\mathcal{I}(F_s)\right] E[u_t(\beta)|\mathcal{I}(F_t)]K\left(\frac{F_t - F_s}{h}\right)\right] \\ &\quad \text{where we assume that } s < t \text{ without loss of generality.} \end{aligned} \tag{12}$$

Then, at least for h sufficiently small, minimizing the population objective function (12) amounts to searching for a value of β that is as close as possible to fulfilling the conditional moment restrictions (2), or equivalently the continuum of unconditional moments (9).

Assumption 1 below gathers all the regularity assumptions discussed so far in order to ensure β_0 is identified and that (12) is uniquely minimized at β_0.

Assumption 1. *(Regularity assumptions)*

(i) $E[u_t|\mathcal{I}(F_t)] = 0$ *with probability 1, and* u_t *has finite fourth moments.*

(ii) $E[Y_t Y_t']$ *is non-singular, and* $E[Y_t|\mathcal{I}(F_t)] \neq 0$ *with probability 1.*

---

of our estimation procedure are equivalent. For simplicity, consider the case where current and first lags of factors are relevant. Our kernel function relies on the kernel function based on F_t with $c = 1$ as defined in (11) which can be rewritten as a kernel function based on (F_t, F_{t-1}) with $c = 0$,

$$\int_\Xi \left(\exp[\iota\xi_0' f_t]\right) \times \left(\exp[\iota\xi_1' f_{t-1}]\right) d\mu(\xi)\,.$$

(iii) Let μ be a given strictly positive measure defined on Ξ a compact subset of $\mathbb{R}^k$. Let $K(.)$ be the kernel function defined on $\mathbb{R}^k$ such that:

$$K(f_t) = \int_\Xi \exp\left(\iota \sum_{j=0}^{c} \xi_j' f_{t-j}\right) d\mu(\xi)\,,$$

where $\xi \in \Xi$ some compact subset of $\mathbb{R}^k$, and c some positive (finite) constant. We assume that $K(.)$ is a symmetric bounded density function on $\mathbb{R}^k$ and that its Fourier transform is strictly positive.

(iv) Let $u_t(\beta) \equiv y_t - Y_t'\beta$ for any t. We assume that:

$$\begin{aligned} E\left[u_t(\beta)|\mathcal{I}(F_t)\right] &= E\left[u_t(\beta)|\mathcal{I}(F_t, y_{t-1}, Y_{t-1})\right] \quad \textit{for any } t. \\ E\left[u_s(\beta)|\mathcal{I}(F_s)\right] &= E\left[u_s(\beta)|\mathcal{I}(F_t)\right] \quad \textit{for any } s < t. \end{aligned}$$

Assumption 1(i) maintains the validity of the (true) factors - and associated information set, while (ii) is akin to maintaining their relevance; together, they ensure that β_0 is (non-parametrically) identified. (iii) imposes mild restrictions on the measure $\mu(.)$ and associated kernel $K(.)$. Finally, (iv) maintains that the (exogenous) factors summarize the dynamics of the errors, and ensures that the factors are strictly exogenous. When thinking about the factors as state variables, such assumptions are not uncommon in asset pricing models: see section 6 in Antoine et al. (2020) and references therein for further discussion of these additional regularity conditions, and their interpretation in the context of asset pricing models.

Proposition 1. *(Identification of β_0)*
Under Assumption 1, β_0 is the unique minimizer of (12) with $M_\infty(\beta_0) = 0$ and

$$\beta_0 = E\left[Y_t Y_s' K\left(\frac{F_t - F_s}{h}\right)\right]^{-1} E\left[Y_t y_s K\left(\frac{F_t - F_s}{h}\right)\right].$$

The invertibility of $E\left[Y_t Y_s' K\left(\frac{F_t - F_s}{h}\right)\right]$ follows from Assumptions 1(i) and (ii), as shown formally in the appendix. Intuitively, it can be interpreted as a generalization of the traditional rank condition. In practice, it implies that the p-vector β_0 of interest can be identified from a much smaller number of factors (e.g. one factor only!) as long as there does not exist a linear combination of the endogenous variables that is conditionally mean-independent of the factor; that is, $E(a'Y_t|F_t) \neq 0$ a.s. for any non-zero real vector a. For example, with two endogenous variables, say $Y_{1,t}$ and $Y_{2,t}$ and one factor $F_{1,t}$, the conditional means of $Y_{1,t}$ and $Y_{2,t}$ cannot be proportional to each other almost surely: that is, for any $a = (a_1, a_2) \neq (0, 0)$,

$$a_1 E(Y_{1,t}|F_{1,t}) + a_2 E(Y_{2,t}|F_{1,t}) \neq 0 \ a.s.$$

In our empirical analysis of the hybrid NKPC, there are two endogenous variables which can be identified - and consistently estimated - with only one factor; see also section B.1 in the appendix for additional discussions and examples.

Let F denote the (T, k) matrix with rows F_t' with $t = 1, \cdots, T$. A natural (infeasible) estimator of β_0 is defined as the minimizer of a sample analog of (12),

$$\tilde{\beta}_T = \arg\min_{\beta \in B} M_T(\beta, F) \tag{13}$$

$$\text{with} \qquad M_T(\beta, F) = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s \neq t, s=1}^{T} u_s(\beta) u_t(\beta) K\left(\frac{F_t - F_s}{h}\right). \tag{14}$$

The infeasible estimator $\tilde{\beta}_T$ defined in (13) is a special case of the Smooth Minimum Distance (SMD) estimator introduced by Lavergne and Patilea (2013) when F is observed, a fixed bandwidth h is used, and $c = 0$. In their i.i.d framework, they show that it is consistent and asymptotically normally distributed, while Antoine et al. (2020) extend these results to time series data. In our (linear) factor-IV framework, the infeasible estimator $\tilde{\beta}_T$ is available in closed-form,

$$\tilde{\beta}_T = \left[Y'\kappa Y\right]^{-1} Y'\kappa y\,,$$

with Y the (T, p)-matrix with row t as Y_t', y the $(T, 1)$-vector, and κ the (T, T)-matrix with element (t, s) as $K((F_t - F_s)/h)$.

In section 3, we introduce our (feasible) factor-based SMD (or F-SMD) estimator, and show that it shares the asymptotic properties of $\tilde{\beta}_T$ which are presented next.

3 Large sample theory of F-SMD

In this section, we first present the asymptotic properties of $\tilde{\beta}_T$, the infeasible factor-based SMD estimator of β_0 defined in (13). Then, we introduce our Factor-SMD (F-SMD hereafter) estimator $\hat{\beta}_T$, as a feasible estimator of β_0 with the same asymptotic properties as $\tilde{\beta}_T$.

3.1 The infeasible factor-based SMD estimator

The infeasible factor-based estimator $\tilde{\beta}_T$ defined in (13) is a special case of the Smooth Minimum Distance (SMD) estimator introduced by Lavergne and Patilea (2013) when F is observed, and its asymptotic properties with dependent data have been derived in Antoine et al. (2020) (see their sections 7.4 and 7.5). Before presenting these results in our factor-IV framework, we introduce our regularity assumptions on the data generating process.

Assumption 2. *(Regularity assumptions on the data generating process)*

(i) (y_t, Y_t, W_t, F_t) *is a stationary weakly dependent process.*

(ii) (y_t, Y_t, W_t, F_t) *satisfy sufficient regularity conditions so that central limit theorems for all appropriate U-statistics apply.*

Assumption 2 allows for general weak dependence in the data, while maintaining high-level restrictions (e.g. on the strength of the mixing property) to ensure CLTs apply on all relevant U-statistics. For explicit conditions, see e.g. Fan and Li (1999) for a general CLT for second order U-statistics with variable kernels for absolutely regular processes; for results beyond absolute regularity see e.g. Dehling and Wendler (2010).

Proposition 2. *(Asymptotic properties of* $\tilde{\beta}_T$*)*
Under Assumptions 1 and 2, the (infeasible) factor-based estimator defined in (13) is consistent for β_0 *and asymptotically normally distributed,*

$$\sqrt{T}(\tilde{\beta}_T - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where $\Sigma = [E(Y_t Y_s' \kappa_{t,s})]^{-1} H_\infty [E(Y_t Y_s' \kappa_{t,s})]^{-1}$ *and* H_∞ *is explicitly*[5] *defined in the Appendix (see equation (26) on page 32).*

3.2 Our proposed F-SMD estimator

In the context of our factor-IV framework (5)-(6), let $\hat{F}$ denote the (T, k) matrix of the first k principal components obtained from the (T, w_q) matrix W; these are commonly used as estimators for F and are in line with Stock and Watson (2002). To deliver our

[5]The matrix H_∞ corresponds to the long-run variance of the underlying U-statistics defined from the first-order conditions. In practice, it involves a double sum of terms such as $(Y_t u_s \kappa_{t,s})$. See the proof in the Appendix for computational details and explicit expressions.

F-SMD estimator, we aim to replace F by $\hat{F}$ in the objective function $M_T(\beta, .)$ defined in (14). The corresponding F-SMD estimator is denoted $\hat{\beta}_T$ and defined as:

$$\hat{\beta}_T = \left[Y'\hat{\kappa}Y\right]^{-1} Y'\hat{\kappa}y \,, \tag{15}$$

with Y the (T,p)-matrix with row t as Y_t', y the $(T,1)$-vector, and $\hat{\kappa}$ the (T,T)-matrix with element (t,s) as $K((\hat{F}_t - \hat{F}_s)/h)$. We show that, under mild conditions, $\hat{\beta}_T$ is asymptotically equivalent to the (infeasible) factor-based estimator $\tilde{\beta}_T$ studied in the previous section. We start with our regularity conditions on the factor structure, and associated estimated factors.

Assumption 3. *(Regularity assumptions on the factor structure)*

(i) $E\|F_t\|^4 \leq M < \infty$; $\sum_t F_t F_t'/T \xrightarrow{p} \Sigma_F$ *with* Σ_F *some* (k,k)*-positive definite matrix;* Λ *has bounded elements, and* $\|\Lambda\Lambda'/w_q - D\| \to 0$ *as* $w_q \to \infty$ *with* D *a positive definite matrix.*

(ii) $E(e_{j,t}) = 0$, $E[|e_{j,t}|^8] < \infty$, *where* $E_t = (e_{1,t}, e_{2,t}, \cdots, e_{w_q,t})'$. *The variance of* E_t *is denoted by* Σ_E. F_t *and* E_s *are independent for all* (t,s).

(iii) Let $\tau_{j,l,t,s} \equiv E[e_{j,t}e_{l,s}]$. *We assume that:*

$$(a) \qquad \sum_{j,l=1}^{w_q} |\tau_{j,l,s,s}|/w_q < \infty \quad \textit{for all } s$$

$$(b) \qquad \sum_{s,t=1}^{T}\sum_{j,l=1}^{w_q} |\tau_{j,l,t,s}|/(Tw_q) < \infty$$

$$(c) \qquad E\left[\sum_{j=1}^{w_q} |e_{j,s}e_{j,t} - \tau_{j,j,t,s}|^4/\sqrt{w_q}\right] < \infty \quad \textit{for all } (t,s).$$

(iv) $\sqrt{T}/w_q \to 0$.

Assumption 3 is rather standard in the factor literature - see e.g. Bai and Ng (2002), Stock and Watson (2002), Bai (2003), Bai and Ng (2006)) - and it is similar to Kapetanios et al. (2016). It is used to obtain consistency of the estimator of the factors and that of the parameters in factor-augmented models; (iv) further ensures that the estimation error in the estimated factors is asymptotically negligible. Thus, for large w_q, the factors can be treated as known. This is reflected in our next result.

Theorem 1. *(Asymptotic properties of F-SMD)*
Under Assumptions 1 to 3, our F-SMD estimator $\hat{\beta}_T$ defined in (15) is consistent for β_0 and asymptotically normally distributed,

$$\sqrt{T}(\hat{\beta}_T - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where $\Sigma = [E(Y_t Y_s' \kappa_{t,s})]^{-1} H_\infty [E(Y_t Y_s' \kappa_{t,s})]^{-1}$ and H_∞ is explicitly defined in the Appendix (see equation (26) on page 32).

Theorem 1 shows that, under our regularity conditions, our proposed F-SMD estimator is asymptotically equivalent to the infeasible factor-based estimator $\tilde{\beta}_T$. When $\sqrt{T}/w_q \to 0$ does not hold, the estimation effect of $\hat{F}_t$ cannot be ignored and the limiting distribution of $\hat{\beta}_T$ will also depend on the limiting distribution of $\hat{F}_t$: the treatment of this case is beyond the scope of this paper. In our Monte-Carlo experiments and in our empirical studies, such an assumption is credible.

Under our maintained assumptions, a consistent estimator of Σ can be obtained after replacing each term by its sample counterpart,

$$[\sum_{t=1}^{n} \sum_{s \neq t}^{n} \hat{\kappa}_{t,s} Y_t Y_s']^{-1} \hat{H}_{\infty,T} [\sum_{t=1}^{n} \sum_{s \neq t}^{n} \hat{\kappa}_{t,s} Y_t Y_s']^{-1} \quad = \quad [Y' \hat{\kappa} Y]^{-1} \hat{H}_{\infty,T} [Y' \hat{\kappa} Y]^{-1} \,,$$

with $\hat{H}_{\infty,T}$ a consistent estimator of H_∞ explicitly defined in Appendix; see equation (27) on page 35. For example, in absence of serial dependence, we use $\hat{H}_{\infty,T} = Y' \hat{\kappa} \Omega_T \hat{\kappa} Y$ with Ω_T a consistent estimator of the variance-covariance matrix of the associated residuals, $\hat{u}_t \equiv y_t - Y_t' \hat{\beta}_T$. See also Appendix B.2.

4 Extended Framework

In this section, we generalize our framework to allow for weak and/or invalid instruments. Specifically, we consider:

$$y_t = Y_t' \beta_0 + u_t \quad \text{with} \quad u_t = \tilde{u}_t + b_T' E_t \tag{16}$$

$$W_t = \Lambda_{w_q} F_t + E_t \tag{17}$$

Such a generalized framework encompasses the factor-IV framework introduced in equations (5)-(6), while also allowing for the instruments W_t to be weak - when Λ_{w_q} is *close to zero* - and/or invalid - when b_T is non-zero. More specifically, we start by highlighting the following features of our extended framework:

- when $b_T = 0$ and $\Lambda_{w_q} = \Lambda$, we are back to our standard Factor-IV framework from section 2 where the observed instruments W_t are valid and driven by a small number of latent (strong) factors F_t.

- when $b_T \neq 0$, at least one component of the vector b_T is non-zero and the structural error u_t is correlated[6] with the observed instruments W_t through E_t. We will consider

$$b_T = \frac{b}{T^\nu} \text{ with } b \neq 0 \text{ fixed vector and } 0 \leq \nu \leq 1/2\,.$$

  This allows us to capture instruments that are always invalid when $\nu = 0$ as well as instruments that are invalid for any T when $\nu > 0$, but remain asymptotically valid since $b_T \to 0$.

- we consider the *weak factor structure* suggested by Kapetanios and Marcellino (2010) which relies on local-to-zero factor loadings,

$$\Lambda_{w_q} = \frac{\Lambda}{w_q^\alpha} \text{ with } 0 \leq \alpha \leq 1/2\,.$$

  The associated factor structure is weak in the sense that, as the cross-sectional dimension of the dataset w_q increases, the factors explain a diminishing proportion of the variance of the data. Kapetanios and Marcellino (2010) show that such a model can accommodate a variety of weak factor loading structures that are relevant in practice; we refer to their section 2.3 for additional discussions and illustrations. Assumption 3 will need to be updated accordingly.

It is important to mention that, in our extended framework, the factors remain valid and strong throughout since Assumptions 1(i)-(ii) continue to hold, that is

$$E(u_t|\mathcal{I}(F_t)) = 0 \; a.s. \quad \text{and} \quad E(Y_t|\mathcal{I}(F_t)) \neq 0 \; a.s.$$

Accordingly, the asymptotic properties of the *infeasible* F-SMD estimator $\tilde{\beta}_T$ - which is computed using the latent factors F_t - remain the same as those established in section 3.

[6] The potential correlation between the observed instruments and the structural error is related to the work of Carrasco and Tchuente (2015) who study issues associated with using many instruments to (efficiently) estimate a linear IV model and introduce a regularized LIML-type estimator.

Next, we derive the asymptotic properties of the (feasible) F-SMD estimator $\hat{\beta}_T$ which relies on the *estimated* factors that are extracted by PCA, as previously explained.

To derive the asymptotic behavior of our proposed F-SMD estimator $\hat{\beta}_T$, we rely on Theorem 4 in Kapetanios and Marcellino (2010) which establishes a sufficient condition on α to ensure the consistent estimation of the (space spanned by the) factors using principal components. Assumptions 3(i) and (iv) are updated as follows.

Assumption 4. *(Regularity assumptions in the extended framework)*

(i) $E\|F_t\|^4 \leq M < \infty$; $\sum_t F_t F_t'/T \xrightarrow{p} \Sigma_F$ *with* Σ_F *some* (k,k)*-positive definite matrix.*

(ii) $\Lambda_{w_q} = \Lambda/w_q^{\alpha}$ *with* $0 \leq \alpha < 1/4$*,* Λ *has bounded elements, and* $\|\Lambda\Lambda'/w_q - D\| \to 0$ *as* $w_q \to \infty$ *with* D *a positive definite matrix.*

(iii) $w_q = o(T^{1/(4\alpha)})$.

Specifically, a sufficient condition for estimation in weak factor model is the presence of a "relatively strong" local-to-zero factor model with $\alpha < 1/4$. This condition allows us to show that $\hat{\beta}_T$ remains asymptotically equivalent to $\tilde{\beta}_T$.

Theorem 2. *(Asymptotic properties of* $\hat{\beta}_T$ *in the extended framework)*
Under Assumptions 1, 2, 3(ii)-(iii), and 4, the F-SMD $\hat{\beta}_T$ *defined in (15) is consistent for* β^0 *and asymptotically normally distributed as in Theorem 1.*

The condition on α is not necessary: in specific cases, such as those considered in our Monte-Carlo experiments, it is possible to obtain consistent estimators of the factors even when $\alpha < 1/2$; we refer the interested reader to section 2.3 in Kapetanios and Marcellino (2010).

5 Monte-Carlo study

We investigate the small sample properties of our F-SMD estimator in the following (linear) structural model,

$$y_t = \alpha_0 + \beta_0 Y_t + u_t$$

where y_t and Y_t are both univariate. We maintain $\alpha_0 = 0$ throughout and focus on the properties of the estimator of β_0 exclusively. We consider two main frameworks:

(i) A small number of (exogenous) instruments Z_t - or observable factors - is available. These are used *directly* - e.g. without relying on a preliminary PCA - through their conditional mean independence to implement F-SMD,

$$E(u_t|\mathcal{I}(Z_t)) = 0\,.$$

We consider cases where the first-stage (or reduced-form equation) is either heterogenous, or unstable over time and show that our F-SMD estimator is reliable and well-behaved without having to specify or estimate the first-stage equation.

(ii) A large number of instruments W_t is available. A small number of (exogenous) factors $\hat{F}_t$ is first extracted from the observed instruments by PCA before implementing our F-SMD estimator based on the conditional mean independence of the underlying true factors F_t,

$$E(u_t|\mathcal{I}(F_t)) = 0\,.$$

In all our simulation designs, performance of the competing estimators[7] (e.g. F-SMD, 2SLS, and efficient GMM) is evaluated by reporting the Monte-Carlo average bias (Bias), standard error (SE), median bias, median standard error, and median absolute deviation; we also report the Monte-Carlo average of the standard errors computed using the heteroskedasticity-robust formula from the asymptotic distribution, the average of the t-statistic when testing the true unknown parameter value, and the associated rejection rate of the t-test. All are computed over 5,000 replications.

5.1 Experiment #1: Small number of observed instruments

We first consider a time series setup with a small number of observed exogenous instruments and first-stage instability[8]. More specifically, there is a structural break in the coefficients of the underlying first-stage equation:

$$\begin{aligned} y_t &= Y_t\beta_0 + \sigma_t u_t \\ Y_t &= 10 \times (2Z_{2,t} - 1)(Z_{1,t} - 2Z_{1,t}^3/5) + v_t \end{aligned}$$

[7] Implementation details are provided in Appendix B.2.

[8] Results with a heterogenous first-stage are presented in the Supplementary Appendix.

where Z_1 is uniformly distributed over $[-2, 2]$, and Z_2 is the break point indicator: it equals 0 up to T_{break} and 1 afterwards where T_{break} is such that the corresponding break fraction, $\lfloor (T - T_{break})/T \rfloor$, is either 0.2 or 0.05. We consider three versions of the model based on the specification of σ_t:

- homoskedastic with $\sigma_t^2 = 1$;
- heteroskedastic (HET1) with $\sigma_t^2 = \sqrt{3 \times (1 + Z_{1,t}^2)/7}$;
- heteroskedastic-GARCH(1,1) (HET2) with $\sigma_t^2 = 0.1 + 0.6\sigma_{t-1}^2 u_{t-1}^2 + 0.3\sigma_{t-1}^2$.

The error terms (u_t, v_t) are independently generated according to a bivariate normal distribution with mean 0, variance 1 and correlation 0.6.

We consider different information sets with respect to the break: (i) the structural break is unknown and ignored, (ii) the existence of the break is known and its location is estimated[9], or (iii) both the existence and the location of the break are known. Accordingly, the econometrician either only observes Z_1, or $(Z_1, \hat{Z}_2)$ where $\hat{Z}_2$ is computed after estimating the break fraction, or (Z_1, Z_2). We then compare the performance of the following estimators: F-SMD, 2SLS, (efficient) GMM, as well as BGMM and B2SLS[10]. Notice that estimating the break fraction requires modelling the first-stage equation - which is precisely what our estimation approach with F-SMD is trying to avoid: as a result, we only implement F-SMD in cases (i) and (iii) when the break is either ignored or fully known.

The results are reported in Tables 1 to 6. We discuss the results from Table 1 obtained in the homoskedastic setup when the break fraction is 0.2, since results under heteroskedasticity are qualitatively similar. When the structural break is ignored (Panel A), the F-SMD estimator performs substantially better than competitors - especially in terms of average and median bias, as well as standard error and overall size control. When the sample size increases to $T = 2,000$, F-SMD still performs better. When the break is fully known, each estimator improves overall, but F-SMD is still the preferred estimator.[11]

[9]The estimated break fraction is obtained by minimizing the SSR in the first-stage equation where the endogenous variable Y_t is regressed on the observed instrument $Z_{1,t}$. Since we consider cases with a small break fraction (0.05), we expand the usual range of candidate break fractions from 0.04 to 0.96.

[10]BGMM and B2SLS, respectively Break-GMM and Break-2SLS, are two estimators introduced in Antoine and Boldea (2018) that use structural changes in the first-stage equation to estimate more efficiently the (stable) structural parameters: e.g. by interacting instruments with breaks from the first-stage. See implementation details in Appendix B.2.

[11]With a sample size of $T = 200$, 2SLS and GMM display erratic behavior and (very) large standard

5.2 Experiment #2: Large number of observed instruments

Our second simulation design involves a large number of (observed) instruments W_t driven by a small number of (unobserved) factors F_t. We consider the extended framework introduced in section 4 where the instruments may be weak and/or invalid:

$$
\begin{aligned}
y_t &= Y_t\beta_0 + \sigma_t u_t && (18)\\
Y_t &= 10 \times (2F_{2,t} - 1)(F_{1,t} - 2F_{1,t}^3/5) + v_t && (19)\\
W_t &= \frac{1}{w_q^{\alpha}}\left[\Lambda_1 F_{1,t} + \Lambda_2 F_{2,t}\right] + E_t && (20)\\
u_t &= \tilde{u}_t + \frac{b'}{T^{\upsilon}} E_t\,, && (21)\\
\sigma_t^2 &= 1 \quad \text{or} \quad 0.1 + 0.6\sigma_{t-1}^2 u_{t-1}^2 + 0.3\sigma_{t-1}^2\,, && (22)
\end{aligned}
$$

where W_t is a vector of $w_q = 50$ observed instruments driven by two unobservable independent factors: F_1 is uniformly distributed over $[-2, 2]$ and F_2 is reminiscent of a break indicator which equals 0 for the first T_{break} observations, and 1 afterwards where $T_{break} = \lfloor 0.95T \rfloor$. The factor loadings Λ_1 and Λ_2 are two vectors of size w_q whose elements are all i.i.d. drawn from a normal distribution with mean 1 and variance 1. The error terms E_t are i.i.d. standard normal, independent of the factors and of all the other errors in the model; the error terms $(\tilde{u}_t, v_t)$ are independently generated according to a bivariate normal distribution with mean 0, variance 1 and correlation 0.6. The model is either heteroskedastic of the GARCH(1,1)-type or homoskedastic. The parameter α controls the strength of the factors: when $\alpha = 0$, it is the standard case with strong factors; we also consider weaker factor structures with $\alpha = 0.125$ or 0.25. The parameter b controls the validity of the instruments: when $b = 0$, the instruments are valid; when $b \neq 0$, the instruments become invalid, either always invalid when $\upsilon = 0$, or weakly exogenous when $\upsilon \neq 0$. We consider the following cases[12]: either $b = 0$ or $b \neq 0$ when its first 10 elements are equal to 1 and the remaining 40 are set to 0 and υ is either 0, 0.25, or 0.5.

errors. These are due to a small number of runs; see section B.4 in the appendix for additional discussions and further analyses.

[12] In the Supplementary Appendix, we also consider a more severe case of invalid instruments where the first 25 elements of b are set to 1, and the remaining ones to 0.

5.2.1 Experiment #2a): strong and valid instruments

We consider the standard framework where the instruments are valid ($b = 0$) and the factor structure is strong ($\alpha = 0$) with GARCH-type heteroskedasticity.

Our results are reported in Table 8 where we consider F-SMD, 2SLS and (efficient) GMM using either one, two or three estimated factors; these observed factors are estimated using Principal Component Analysis on the matrix of observed instruments,

$$W = [W_1, W_2, \cdots, W_T].$$

Once again, the performance of F-SMD is excellent throughout, and dominates that of others in terms of bias and standard deviation. In addition, the performance of F-SMD is particularly insensitive to the number of estimated factors. This is in sharp contrast with 2SLS and GMM that are both negatively affected when the number of estimated factors is less than the true one: in such cases, both display large biases and standard deviations when the sample size is small. These issues are somewhat mitigated when the sample size increases, but these estimates are still more biased and less precise than corresponding F-SMD estimates.

As a robustness check, we also consider a design that is fully linear with strong and valid instruments: the results can be found in the Supplementary Appendix.

5.2.2 Experiment #2b): possibly weak and invalid instruments

We now consider DGPs where the instruments may either be weak or invalid. To model invalid instruments ($b \neq 0$), we consider cases where the first 10 elements of the vector b are set to 1, while the remaining elements are set to 0. We set ν to either 0, 0.25 or 0.5. To model weaker instruments, we consider cases where α ranges from 0 to 0.25. Our results with invalid instruments ($b \neq 0$) are reported in Tables 10 to 12, while the baseline results with valid instruments ($b = 0$) are collected in Table 13.

Overall, the performance of F-SMD is excellent throughout: as previously noted, it remains insensitive to the number of estimated factors, even when the instruments may be weak or invalid. This is in line with our asymptotic results. By contrast, the performance of GMM is affected by the number of estimated factors in all cases; in addition, it deteriorates when the instruments become invalid and/or weak - both in terms of bias and standard errors. Even when considering GMM computed using the correct number of factors in an ideal DGP with strong and valid instruments (as in Panel A of Table 13), it does not perform as well as F-SMD: bias and standard errors are always much smaller with F-SMD, even with median-based measures that are not sensitive to outliers.

5.2.3 Experiment #2c): *artificially* weak first-stage

We now consider a DGP where the first-stage artificially appears weakly identified. Said differently, the relationship between the endogenous variables Y_t and the factor F_t is strong and non-linear: however, when incorrectly fitting a linear model, as commonly done in practice, it appears weak. Specifically, we replace equation (19) in the above model by:

$$Y_t = 3F_{1,t} - F_{1,t}^3 + F_{1,t}^2 - 1 + v_t$$

and F_1 is normally distributed with mean 0 and variance 1; everything else remains as is. Notice than when a linear (factor) model is fitted, the solution of

$$\min_{\pi} E(Y - \pi F_1)^2 = \min_{\pi} E(3F_1 - F_1^3 + F_1^2 - 1 - \pi F_1)^2$$

is $\pi = 0$. Hence, relying on a linear (factor) model effectively destroys the explanatory power of F_1 - and that of associated instruments driven by F_1. Consequently, standard inference procedures such as 2SLS, or GMM based on the estimated factor $\hat{F}_1$ will suffer from identification issues.[13]

In Table 14, we report results obtained with F-SMD, 2SLS and GMM. In addition, we report: (i) the rejection frequencies associated with the (diagnostic) test of the null of weak identification based on the F-test, (ii) the proportion of identification-robust confidence intervals obtained with the Anderson-Rubin method that are infinitely large, and (iii) the average length of (finite) confidence intervals obtained with the Anderson-Rubin method and with F-SMD.

Overall, the performance of F-SMD is excellent throughout: as previously noted, it remains insensitive to the number of estimated factors even with a sample size of $T = 200$. It outperforms GMM in all cases in terms of bias and standard errors. As previously explained, this design is artificially weak. Indeed, the null of weak identification is rejected with probability at most 0.4 including when the sample size is large ($T = 2,000$). When using identification-robust inference with the Anderson-Rubin method, associated confidence intervals are infinitely large in at least 60% of the cases. And, even when we focus on cases where theses intervals are finite, they are

[13] Antoine and Lavergne (2023) were first to highlight this point in a standard linear IV model. Our simulation design is inspired by theirs.

much wider than those obtained with F-SMD: e.g. at least four times wider even with large sample sizes.

5.3 Experiment #3: Empirically-motivated design

In our last experiment, we consider a design that is motivated by our empirical application. Similar to the NKPC model estimated in section 6, the dependent variable y_t depends linearly on two endogenous variables, $Y_{1,t}$ and $Y_{2,t}$ (e.g. inflation and driver), and a constant,

$$y_t = c + \gamma_f Y_{1,t} + \lambda Y_{2,t} + \tilde{u}_t + \rho \tilde{u}_{t-1}$$

with $c = 0$, $\gamma_f = 0.55$ and $\lambda = 0.05$. These endogenous variables are generated from two (latent) factors, $F_{1,t}$ and $F_{2,t}$, which correspond to the two factors extracted by PCA from the large number of instruments available in our application (see e.g. section 6.1),

$$Y_{j,t} = g_j(F_{1,t}, F_{2,t}) + c_j v_{j,t} \quad \text{for} \quad j = 1, 2$$

Further, the functions g_1 and g_2 are obtained by fitting a flexible non-linear model between the two observed endogenous variables from our empirical application (e.g. inflation and driver) and these two extracted factors. Finally, our (observed) exogenous instruments W_t are generated as follows:

$$W_t = \Lambda_1 F_{1,t} + \Lambda_2 F_{2,t} + E_t$$

$(u_t, v_{1,t}, v_{2,t})$ are independently generated according to a trivariate normal distribution with means 0, variances 1 and correlations 0.6 between $v_{j,t}$ and u_t and 0 between $v_{1,t}$ and $v_{2,t}$; c_1 and c_2 are set at 0.025 and 0.05 respectively, and

$$\tilde{u}_t = 0.01 v_{1,t} + 0.01 v_{2,t} + 0.02 u_t \,.$$

E_t are independently generated as multivariate uncorrelated normals with means 0 and variances 0.1 - independently of $(u_t, v_{1,t}, v_{2,t})$. Finally, ρ is either set at 0 or 0.5 to generate MA-dynamics in the structural errors. The results are collected in Table 15 where we report estimation results obtained with F-SMD implemented using either 1 or 2 estimated factors, and GMM using 2 estimated factors. As we already explained, the coefficients of both endogenous variables can be estimated with F-SMD implemented with only one estimated factor, whereas GMM needs two.

Overall, the performance of F-SMD is excellent throughout, and, as previously noted, it remains insensitive to the number of estimated factors. In addition, the presence of MA-dynamics in the structural errors does not affect its performance either. GMM delivers reliable estimation results for both parameters in all cases. However, it is important to note that γ_f is always less precisely estimated by GMM. When the number of instruments decreases (from 50 to 6), both estimators still perform well: the standard errors are comparable, and there is only a slight increase in the bias of γ_f.

6 Inflation dynamics and the NKPC

The New Keynesian Phillips Curve (NKPC) has played an important role in recent monetary policy analysis. In its canonical form, the NKPC model expresses current inflation as a (linear) function of expected inflation and marginal costs. To respond to criticisms stemming from the model's inability to sufficiently explain the persistence of US inflation dynamics, Gali and Gertler (1999) introduced the hybrid NKPC which also includes a backward-looking component and can be written as,

$$\pi_t = \gamma_0 + \gamma_f \pi^e_{t+1} + \gamma_b \pi_{t-1} + \lambda mc_t \,, \tag{23}$$

where π_t is the rate of inflation, π^e_{t+1} is the expected inflation for $(t+1)$ at time t, and mc_t is the marginal cost of production. Notice that the hybrid NKPC (23) encompasses the canonical NKPC, as it reduces to it when $\gamma_b = 0$. Choosing between the canonical model and the hybrid one has been an important empirical issue, not only to understand inflation dynamics, but also to design effective monetary policy. Indeed, the presence of lagged inflation in (23) indicates the *lagged effect of monetary policy* by changing the real economy, while the forward-looking term captures its *direct effect* by changing economic agents' expectations.

Previous studies deliver conflicting results as to the relative importance of forward- and backward-looking behaviors, depending on the chosen empirical specification and econometric method. This paper contributes to this important issue by implementing our flexible F-SMD estimation procedure with various instrumental variables, from traditional ones (taken as lags of included variables) to additional ones, either using alternative measures (e.g. of inflation), or built as comprehensive indicators of broad macro-economic conditions. Our estimation procedure is flexible in two important ways: first, it can easily accommodate (many) instrumental variables; second, it is robust to the specification of the first stage (such as structural breaks, or non-linearities).

Using quarterly US data from 1960 to 2022, our main empirical analysis provides strong support for the hybrid NKPC. In addition, our estimation results are relatively stable over time and quite precise. They reveal that forward- and backward-looking behaviors are quantitatively equally as important, while the driver's role is nil. Our empirical analysis also explores macro-finance linkages by augmenting traditional sets of (macro) instruments with macro-finance variables. Overall, our results remain quite similar with and without these additional variables.

6.1 Main empirical analysis

In our main empirical analysis, we consider the following hybrid NKPC model where expected inflations are simply replaced by future realizations of inflation as commonly done under the maintained assumption that expectations are rational:

$$\pi_t = c + \gamma_f \pi_{t+1} + (1 - \gamma_f)\pi_{t-1} + \lambda mc_t + u_t \,. \tag{24}$$

Our set of instruments include standard instruments taken as lags of the variables included in the model, as well as lags of another common driver and alternate measures of inflation[14]: namely, one lag of inflation, marginal cost, output gap, wage inflation, spread between long and short interest rates and inflation on commodity price. We also consider more comprehensive instruments obtained from the large dataset of macro-finance variables from McCracken and Ng (2020).

We consider quarterly US data from 1960Q1 to 2022Q2 obtained from Federal Reserve Economic Data (FRED) of St. Louis. Specifically, *inflation* is defined as the percentage change of the GDP deflator (series ID: GDPDEF). For the *real marginal cost*, we use the HP-filtered series of the log of the labor income share of nonfarm business sector (series ID: PRS85006173). *Output gap* is constructed by the log deviation of real GDP (series ID: GDPC1), also measured by the HP-filter. *Wage inflation* is created by the percentage change of the unit labor cost of nonfarm business sector (series ID: ULCNFB). Our *macro-factor* is obtained by principal component analysis (PCA hereafter) from the large macro-finance dataset of McCracken and Ng (2020)).

The above-mentioned inflation series is found to be non-stationary over the sample period, an issue previously reported in the literature: see e.g. section 3.4.1 in Mavroeidis et al. (2014) for discussions and additional references. Following related

[14]In the recent empirical analysis by Choi (2021), these instruments have been found to be sufficiently strong for standard GMM estimation to be reliable; see e.g. p652.

literature, we then write our model in terms of changes in inflation,

$$\Delta\pi_t = c + \gamma_f(\pi_{t+1} - \pi_{t-1}) + \lambda mc_t + u_t\,, \tag{25}$$

and use one lag of $\Delta\pi_t$ as instrument (instead of one lag of π_t)[15]. Hereafter, the (current) marginal cost variable mc_t is not assumed to be exogenous.

Ultimately, our sample contains 247 observations. We use HAC standard errors throughout.

6.2 Empirical results and discussion

In Table 16, we estimate the model by F-SMD, GMM and B-GMM using different instrument sets, chosen as one lag of either, marginal cost, output gap, commodity inflation, spread, wage-inflation, or our macro-factor. To be clear, we are in the *hybrid* situation described on page 7 where our factors combine chosen (observed) instruments - as done in the literature - with a (true) factor extracted by PCA from the large macro-finance dataset mentioned above.

Overall, the estimation results for F-SMD are fairly consistent regardless of the selected instrument set. They reveal that forward- and backward-looking behaviors are quantitatively equally as important with estimates for γ_f close to 0.50 (or slightly above), and statistically significantly different from 0 and from 1 at 95% - but not from 0.50 - which also provides support to the hybrid NKPC. Further, the driver is systematically found to have little to no effect with the estimation of λ, the slope parameter of the marginal cost, approximately zero throughout and not statistically significant at any reasonable level.

These results are in sharp contrast with those obtained using GMM which are economically implausible, very noisy, and very sensitive to the instrument set: for example, while some estimates of γ_f are negative, all 95% confidence intervals are quite wide and always include both 0 and 1; estimates of λ are more reasonable and on par with those obtained by F-SMD.

It is quite remarkable that these issues and inconsistencies are resolved when imposing a break point at the onset of the pandemic, 2020Q1, in the first-stage equation, and using B-GMM to estimate the model instead. All B-GMM associated estimates for γ_f are now very much in line with those obtained by F-SMD[16]: estimates are close to 0.50

[15]However, such an instrument appears to be extremely noisy compared to other ones, and we choose to leave it aside.

[16]Recall that F-SMD estimates are obtained without imposing any restriction on the first-stage

(or slightly above), and statistically significantly different from 0 and from 1 at 95% - but not from 0.50. As robustness checks, we also report in Table 17 estimation results obtained with F-SMD and GMM over the first subsample obtained with 238 observations. These results are very much in line with the results obtained by F-SMD over the whole sample: this suggests that estimates of the structural parameters (γ_f and λ) remain relatively stable, whereas the first-stage equation seems to display parameter instability. In addition, it also appears appropriate to maintain the linearity of the first-stage equation after accounting for parameter instability. Indeed, while F-SMD and B-GMM estimates for the structural parameters γ_f and λ reported in Table 16 remain quite close to each other, associated standard errors are not: e.g. they can be quite a bit smaller with B-GMM, especially with larger sets of instruments. This may be interpreted as the price to implement our *robust* estimation strategy which remains immune to potential misspecification of the first-stage equation. Given the noisy and unreliable results obtained with a standard and non-robust procedure such as GMM, this appears to be a modest price to pay.

Nonetheless, to mitigate potential concerns related to the implementation of F-SMD with a larger number of instruments, our last set of results relies on using one instrument only. In Table 18, we estimate the model by F-SMD using only one instrument, chosen as one lag of either, marginal cost, output gap, commodity inflation, spread, wage-inflation, or the macro-factor, as well as the first PCA extracted from these six instruments (PCA1) and the first PCA extracted from the joint set of these five macro instruments and all macro-finance variables from McCracken and Ng (2020) (one-step PCA1). Once again, the estimation results are quite stable with many estimates of γ_f around 0.50[17]. Noticeably, standard errors associated with the generated instrument labelled one-step PCA1 are among the smallest ones.

Overall, our empirical results emphasize the convenience and reliability of our estimation strategy which does not require the specification of the first-stage equation, or its estimation, even when using modest sample sizes.

References

Antoine, B. and O. Boldea (2018). Efficient Estimation with Time-Varying Information and the New Keynesian Phillips Curve. *Journal of Econometrics 204*, 268–300.

equation: that is, without imposing a break point, linearity, or any other functional form assumption.

[17]The estimate obtained with C-Inf is noticeably smaller at 0.266 with larger standard errors of 0.169.

Antoine, B. and P. Lavergne (2014). Conditional Moment Models under Semi-Strong Identification. *Journal of Econometrics 182*(1), 59–69.

Antoine, B. and P. Lavergne (2023). Identification-robust non-parametric inference in a linear IV model. *Journal of Econometrics 235*, 1–24.

Antoine, B., K. Proulx, and E. Renault (2020). Pseudo-True SDFs in Conditional Asset Pricing Models. *Journal of Financial Econometrics 18*(4), 656–714. Invited Lecture with Discussion.

Antoine, B. and X. Sun (2022). Partially Linear Models with Endogeneity: a conditional moment based approach. *The Econometrics Journal 25*(1), 256–275.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–173.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*, 1133–1150.

Bai, J. and S. Ng (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory 26*, 1577–1606.

Bierens, H. (1982). Consistent Model Specification Tests. *Journal of Econometrics 20*(1), 105–134.

Bierens, H. (1990). A consistent conditional moment test of functional form. *Econometrica 58*, 1443–1458.

Bierens, H. and W. Ploberger (1997). Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica 65*(5), 1129–1151.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics 170*, 383–398.

Carrasco, M. and G. Tchuente (2015). Regularized LIML for many instruments. *Journal of Econometrics 186*, 427–442.

Chen, X., S. Lee, M. Seo, and M. Song (2022). Inference for parameters identified by conditional moment restrictions using a penalized Bierens maximum statistic. *Working paper*.

Choi, Y. (2021). Inflation dynamics, the role of inflation at different horizons and inflation uncertainty. *International Review of Economics and Finance 71*, 649–662.

de Jong, R. (1996). The Bierens test under data dependence. *Journal of Econometrics 72*, 1–32.

Dehling, H. and M. Wendler (2010). Central Limit Theorem and the Bootstrap for U-Statistics of Strongly Mixing Data. *Journal of Multivariate Analysis 101*, 126–137.

Dominguez, M. and I. Lobato (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica 72*, 1601–1615.

Escanciano, J. C. (2018). A simple and robust estimator for linear regression models with strictly exogenous instruments. *Econometrics Journal 21*, 36–54.

Fan, Y. and Q. Li (1999). Central limit theorem for degenerate U-statistics of Absolutely Regular Processes with Applications to Model Specification Testing. *Journal of Nonparametric Statistics 10*(3), 245–271.

Gali, J. and M. Gertler (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics 44*(2), 195–222.

Kapetanios, G., L. Khalaf, and M. Marcellino (2016). Factor-Based Identification-Robust Interference in IV Regressions. *Journal of Applied Econometrics 31*(5), 821–842.

Kapetanios, G. and M. Marcellino (2010). Factor-GMM estimation with large sets of possibly weak instruments. *Computational Statistics and Data Analysis 54*, 2655–2676.

Lavergne, P. and V. Patilea (2013). Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory. *Journal of Econometrics 177*(1), 47–59.

Mavroeidis, S., M. Plagborg-Møller, and J. Stock (2014). Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature 52*(1), 124–188.

McCracken, M. and S. Ng (2020). FRED-QD: A Quarterly Database for Macroeconomic Research. *Working Paper 2020-005B*.

Mikusheva, A. (2021). Many Weak Instruments in Time Series Econometrics. *Working Paper*. Available at https://economics.mit.edu/faculty/amikushe/working.

Stinchcombe, M. and H. White (1998). Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative. *Econometric Theory 14*, 295–325.

Stock, J. and M. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Sun, X. (2022). Estimation of Heterogeneous Treatment Effects Using a Conditional Moment Based Approach. *Working paper*.

Tsyawo, E. S. (2023). Feasible IV Regression without Excluded Instruments. *The Econometrics Journal 26*, 235–256.

A Proofs of the theoretical results

• **Proof of Proposition 1**:
Without loss of generality, take $s < t$. Under Assumption 1, we have:

$$
\begin{aligned}
& M_\infty(\beta) \\
\equiv\ & E\left[u_s(\beta)u_t(\beta)K\left(\frac{F_t - F_s}{h}\right)\right] \\
=\ & E\left[u_s(\beta)E\left[u_t(\beta)|\mathcal{I}(F_t), y_{t-1}, Y_{t-1}\right]K\left(\frac{F_t - F_s}{h}\right)\right] \\
=\ & E\left[u_s(\beta)E\left[u_t(\beta)|\mathcal{I}(F_t)\right]K\left(\frac{F_t - F_s}{h}\right)\right] \\
=\ & E\left[E\left[u_s(\beta)|\mathcal{I}(F_t)\right]E[u_t(\beta)|\mathcal{I}(F_t)]K\left(\frac{F_t - F_s}{h}\right)\right] \\
=\ & E\left[E\left[u_s(\beta)|\mathcal{I}(F_s)\right]E[u_t(\beta)|\mathcal{I}(F_t)]K\left(\frac{F_t - F_s}{h}\right)\right] \\
=\ & \int_\Xi E\left[E[u_s(\beta)\exp(-\iota\sum_{j=0}^{c}\xi_j' F_{s-j}/h)|\mathcal{I}(F_s)]E[u_t(\beta)\exp(\iota\sum_{j=0}^{c}\xi_j' F_{t-j}/h)|\mathcal{I}(F_t)]\right]d\mu(\xi) \\
=\ & \int_\Xi\left\{E[u_t(\beta)\exp(\iota\sum_{j=0}^{c}\xi_j' F_{t-j}/h)]^2\right. \\
+\ & \left.\texttt{Cov}\left[E[u_s(\beta)\exp\left(-\iota\sum_{j=0}^{c}\xi_j' F_{s-j}/h\right)|\mathcal{I}(F_s)], E[u_t(\beta)\exp(\iota\sum_{j=0}^{c}\xi_j' F_{t-j}/h)|\mathcal{I}(F_t)]\right]\right\}d\mu(\xi)
\end{aligned}
$$

Hence, we have $M_\infty(\beta) \geq 0$ since the first term on the RHS is non-negative, while the second tends to zero as $|t-s| \to \infty$; further, we have $M_\infty(\beta_0) = 0$ since $E[u_t(\beta_0)|\mathcal{I}(F_t)] = 0$ w.p.1 $\forall\ t$, and we conclude that β_0 minimizes $M_\infty(.)$.
Let $\kappa_{t,s} \equiv K\left((F_t - F_s)/h\right)$ for convenience. Then, the associated FOC write:

$$
\begin{aligned}
& E(Y_t(y_s - Y_s'\beta_0)\kappa_{t,s} + (y_t - Y_t'\beta_0)Y_s\kappa_{t,s}) = 0 \\
\Rightarrow\ & E(Y_t(y_s - Y_s'\beta_0)\kappa_{t,s}) + E((y_t - Y_t'\beta_0)Y_s\kappa_{t,s}) = 0 \\
\Rightarrow\ & \beta_0 = E[(Y_tY_s' + Y_sY_t')\kappa_{t,s}]^{-1}E[(Y_ty_s + Y_sy_t)\kappa_{t,s}] \\
\Rightarrow\ & \beta_0 = [E(Y_tY_s'\kappa_{t,s})]^{-1}E[Y_ty_s\kappa_{t,s}]
\end{aligned}
$$

where the last expression follows from $E(Y_tY_s'\kappa_{t,s})$ being positive definite hence symmetric. We now show that $E[Y_tY_s'\kappa_{t,s}]$ is positive definite.
For any a real vector of size p, we have:

$$
\begin{aligned}
& E\left[a'Y_tY_s'a\kappa_{t,s}\right] \\
= \;& E\Big[a'E\left[Y_t|\mathcal{I}(F_t)\right] E[Y_s'|\mathcal{I}(F_s)]a\kappa_{t,s}\Big] \\
= \;& \int_\Xi E\left[a'E\left[Y_t|\mathcal{I}(F_t)\right]\exp(\iota\sum_{j=0}^{c}\xi_j'F_{t-j}/h)E[Y_s'|\mathcal{I}(F_s)]a\exp(-\iota\sum_{j=0}^{c}\xi_j'F_{s-j}/h)\right] \\
= \;& \int_\Xi E\left[a'E[Y_t|\mathcal{I}(F_t)]\exp(\iota\sum_{j=0}^{c}\xi_j'F_{t-j}/h)\right]\times E\left[E[Y_s'|\mathcal{I}(F_s)]a\exp(-\iota\sum_{j=0}^{c}\xi_j'F_{s-j}/h)\right]d\mu(\xi) \\
+ \;& \int_\Xi \mathtt{Cov}\left[a'E[Y_t|\mathcal{I}(F_t)]\exp(\iota\sum_{j=0}^{c}\xi_j'F_{t-j}/h), E[Y_s'|\mathcal{I}(F_s)]a\exp(-\iota\sum_{j=0}^{c}\xi_j'F_{s-j}/h)\right]d\mu(\xi)
\end{aligned}
$$

Notice that this expression is non-negative since the first term on the RHS is non-negative, wile the second becomes negligible as s and t are further apart. To see this, it is useful to rewrite the first term as follows after introducing $\bar{F}_t = (F_t, F_{t-1}, \cdots, F_{t-l})$ with some $l \geq c$ and its density $f_{\bar{F}}(.)$:

$$
\begin{aligned}
& \int_\Xi |\left(\int a'E[Y_t|\mathcal{I}(F_t)]\exp(\iota\sum_{j=0}^{c}\xi_j'F_{t-j}/h)f_{\bar{F}}(\bar{F}_t)d(\bar{F}_t)\right)|^2 d\mu(\xi) \\
= \;& (2\pi)^{2k}\int_\Xi |\left(\mathcal{F}[a'E(Y_t|\mathcal{I}(F_t))f_{\bar{F}}(\bar{F}_t)](\xi)\right)|^2 d\mu(\xi) \\
\geq \;& 0\,,
\end{aligned}
$$

since μ strictly positive on Ξ and with $\mathcal{F}[g]$ the Fourier transform of a well-defined function $g(.)$ on Ξ formally defined as,

$$
\mathcal{F}[g](\xi) = \frac{1}{(2\pi)^k}\int \exp\left(\iota\sum_{j=0}^{c}\xi_j'u_{t-j}\right)g(u_t, u_{t-1}, \cdots, u_{t-l})d(u_t, u_{t-1}, \cdots, u_{t-l})\,.
$$

We then have:

$$
\begin{aligned}
E(a'Y_tY_s'a\kappa_{t,s}) = 0 \;&\Leftrightarrow\; \exists\, a \neq 0 \;s.t.\; a'E[Y_t|\mathcal{I}(F_t)]f_{\bar{F}}(\bar{F}_t) = 0 \;a.s. \\
&\Leftrightarrow\; \exists\, a \neq 0 \;s.t.\; a'E[Y_t|\mathcal{I}(F_t)] = 0 \;a.s.
\end{aligned}
$$

This cannot hold, since, under Assumption 1, $E(Y_t|\mathcal{I}(F_t)) \neq 0$ a.s. and $E(Y_tY_t')$ is nonsingular. ■

• **Proof of Proposition 2**:

From the FOC, we have:

$$\begin{aligned}
& \left[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,\ s\neq t}^{T}\kappa_{t,s}Y_tY_s'\right]\tilde{\beta}_T = \left[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,\ s\neq t}^{T}\kappa_{t,s}Y_ty_s\right] \\
\Leftrightarrow \quad & \left[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,\ s\neq t}^{T}\kappa_{t,s}Y_tY_s'\right]\left(\tilde{\beta}_T - \beta_0\right) = \left[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,\ s\neq t}^{T}\kappa_{t,s}Y_tu_s\right] \\
\Leftrightarrow \quad & A_T\left(\tilde{\beta}_T - \beta_0\right) = B_T
\end{aligned}$$

with obvious notations. We now show that A_T and B_T are both U-statistics, and find their asymptotic distributions by applying appropriate CLTs.

(i) To show that A_T is a U-statistic, notice that

$$\begin{aligned}
A_T &= \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,\ s\neq t}^{T}\kappa_{t,s}Y_tY_s' \\
&= \frac{1}{2}\frac{2}{T(T-1)}\sum_{1\leq s<t\leq T}(\kappa_{t,s}Y_tY_s' + \kappa_{s,t}Y_sY_t') \\
&= \frac{1}{2}\tilde{A}_T
\end{aligned}$$

Hence, A_T is a half of a U-statistics denoted $\tilde{A}_T$. Let us denote $X_t \equiv (y_t, Y_t, W_t, F_t)$ for any t. Using the Hoeffding decomposition, $\tilde{A}_T$ can be rewritten as:

$$\begin{aligned}
\tilde{A}_T &= E[h(X_t, X_s)] + \frac{2}{T}\sum_{t=1}^{T}h_1(X_t) + \frac{2}{T(T-1)}\sum_{t<s}h_2(X_t, X_s) \\
\text{with} \quad h(X_t, X_s) &\equiv Y_tY_s'\kappa_{t,s} + Y_sY_t'\kappa_{s,t} \\
h_1(x) &\equiv E[h(x, X_s)] - E[h(X_t, X_s)] \\
h_2(x, z) &\equiv h(x, z) - h_1(x) - h_1(z) - E[h(X_t, X_s)]
\end{aligned}$$

Then, a CLT applies to $(a'\tilde{A}_T)$ for any $a \in \mathbb{R}^p$ (e.g. Theorem 1.8 in Dehling and Wendler

(2010)), and implies that:

$$
\begin{aligned}
& \sqrt{T}a'(\tilde{A}_T - E[h(X_t, X_s)]) = \mathcal{O}_P(1) \\
\Rightarrow \quad & \tilde{A}_T \xrightarrow{P} E[h(X_t, X_s)] \\
& \text{with} \quad E[h(X_t, X_s)] = E(Y_t Y_s' \kappa_{ts} + Y_s Y_t' \kappa_{st}) = 2E(Y_t Y_s' \kappa_{t,s})
\end{aligned}
$$

which is a nonsingular matrix as shown previously in the proof of Proposition 1.

(ii) We follow the same steps for B_T:

$$
\begin{aligned}
B_T &= \frac{1}{2}\tilde{B}_T = \frac{1}{2}\frac{2}{T(T-1)} \sum_{1 \le s < t \le T} g(X_t, X_s) \\
\text{where} \quad \tilde{B}_T &= \frac{2}{T}\sum_{t=1}^{T} g_1(X_t) + \frac{2}{T(T-1)} \sum_{t<s} g_2(X_t, X_s) \\
\text{with} \quad g(X_t, X_s) &\equiv Y_t u_s \kappa_{t,s} + Y_s u_t \kappa_{s,t} \\
E[g(X_t, X_s)] &= 0 \quad \text{(shown at the end of the proof)} \\
g_1(x) &\equiv E[g(x, X_s)] \\
g_2(x, z) &\equiv g(x, z) - g_1(x) - g_1(z)
\end{aligned}
$$

Then, a CLT applies to $(a'\tilde{B}_T)$ for any $a \in \mathbb{R}^p$ (e.g. Theorem 1.8 in Dehling and Wendler (2010)), and we get:

$$
\sqrt{T}a'\tilde{B}_T \xrightarrow{d} \mathcal{N}(0, 4\sigma_B(a))
$$

with $\sigma_B^2(a) = Var[a'g_1(X_t)] + 2\sum_{k=1}^{\infty} Cov(a'g_1(X_t), a'g_1(X_{t+k}))$.
The asymptotic distribution of $\tilde{B}_T$ follows from the application of the Cramér-Wold theorem, $\sqrt{T}\tilde{B}_T \xrightarrow{d} \mathcal{N}(0, 4H_\infty)$, with

$$
H_\infty \equiv Var[g_1(X_t)] + \sum_{k=1}^{\infty}(Cov(g_1(X_t), g_1(X_{t+k})') + Cov(g_1(X_{t+k}), g_1(X_t)')) \,. \tag{26}
$$

The expected result follows with $\Sigma = [E(Y_t Y_s' \kappa_{t,s})]^{-1} H_\infty [E(Y_t Y_s' \kappa_{t,s})]^{-1}$. We conclude the proof by showing that $E[g(X_t, X_s)] = 0$.

$$
\begin{aligned}
E[g(X_t, X_s)] &= 2E[Y_t u_s \kappa_{t,s}] \\
&= 2\int_{\Xi} E\left(Y_t u_s \exp\left(\iota \sum_{j=0}^{c} \xi_j'(F_{t-j} - F_{s-j})/h\right)\right) d\mu(\xi) \\
&= 2\int_{\Xi} E\left(E[Y_t \exp\left(\iota \sum_{j=0}^{c} \xi_j' F_{t-j}/h\right) |\mathcal{I}(F_t)] \right. \\
&\qquad \left. \times E[u_s \exp\left(-\iota \sum_{j=0}^{c} \xi_j' F_{s-j}/h\right) |\mathcal{I}(F_s)] \right) d\mu(\xi) \\
&= 0
\end{aligned}
$$

which follows from $E[u_s \exp\left(-\iota \sum_{j=0}^{c} \xi_j' F_{s-j}/h\right) |\mathcal{I}(F_s)] = 0$ under Assumption 1(i). ■

• **Proof of Theorem 1** :

From the FOC, we have:

$$
\begin{aligned}
& \left[\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1,\ s\neq t}^{T} \hat{\kappa}_{t,s} Y_t Y_s'\right] \hat{\beta}_T = \left[\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1,\ s\neq t}^{T} \hat{\kappa}_{t,s} Y_t y_s\right] \\
\Leftrightarrow\ & \left[\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1,\ s\neq t}^{T} \hat{\kappa}_{t,s} Y_t Y_s'\right] \left(\hat{\beta}_T - \beta_0\right) = \left[\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1,\ s\neq t}^{T} \hat{\kappa}_{t,s} Y_t u_s\right] \\
\Leftrightarrow\ & A_T(\hat{F}) \left(\hat{\beta}_T - \beta_0\right) = B_T(\hat{F})
\end{aligned}
$$

with obvious notations, including $\hat{\kappa}_{t,s} \equiv K((\hat{F}_t - \hat{F}_s)/h)$. We now study the asymptotic properties of $A_T(\hat{F})$ and $B_T(\hat{F})$ and show how they relate to those of A_T and B_T defined in the proof of Proposition 2.

$$
\begin{aligned}
A_T(\hat{F}) &= \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,\ s\neq t}^{T}\hat{\kappa}_{t,s}Y_tY_s' \\
\text{where } \hat{\kappa}_{t,s} &= \int_{\Xi}\exp\left[\sum_{j=0}^{c}\xi_j'\left(\frac{\hat{F}_{t-j}-\hat{F}_{s-j}}{h}\right)\right]d\mu(\xi) \\
&= \int_{\Xi}\exp\left[\sum_{j=0}^{c}\xi_j'\left(\frac{H'(F_{t-j}-F_{s-j})}{h}+c_T\right)\right]d\mu(\xi) \\
&= \exp(c_T)\times\kappa_{t,s}\,,
\end{aligned}
$$

with H invertible matrix and c_T which follow from Lemma 1 below,

$$
c_T = \mathcal{O}_p\left(\frac{1}{\min(\sqrt{w_q},T)}\right).
$$

This implies that:

$$
A_T(\hat{F}) = \exp(c_T)\times A_T
$$

Similarly, we have:

$$
B_T(\hat{F}) \;=\; \exp(c_T)\times B_T\,.
$$

As a result, under Assumption 3(iv), it follows immediately that $\sqrt{T}(\hat{\beta}_T-\beta_0)$ is asymptotically equivalent to $\sqrt{T}(\tilde{\beta}_T-\beta_0)$.

We conclude the proof by detailing how to consistently estimate the asymptotic variance Σ of $\hat{\beta}_T$ with

$$
\Sigma = [E(Y_tY_s'\kappa_{t,s})]^{-1}H_\infty[E(Y_tY_s'\kappa_{t,s})]^{-1}\,.
$$

As explained in the main text, under our maintained assumptions, the estimation of the factors can be ignored; this also means that a consistent estimator of Σ can simply be obtained after replacing each term by its sample counterpart as follows:

$$
[\sum_{t=1}^{n}\sum_{s\neq t}^{n}\hat{\kappa}_{t,s}Y_tY_s']^{-1}\hat{H}_{\infty,T}[\sum_{t=1}^{n}\sum_{s\neq t}^{n}\hat{\kappa}_{t,s}Y_tY_s']^{-1} \;=\; [Y'\hat{\kappa}Y]^{-1}\hat{H}_{\infty,T}[Y'\hat{\kappa}Y]^{-1}\,,
$$

with $\hat{H}_{\infty,T}$ a consistent estimator of H_∞ such as

$$\begin{aligned}
&\sum_{t=1}^{T}\left[\left[\sum_{s=1}^{T}\kappa_{t,s}Y_s\right]\left[\sum_{s=1}^{T}\kappa_{t,s}Y_s\right]' u_t^2\right] \\
&+ \sum_{k=1}^{M}\sum_{t=1+k}^{T}(1-\frac{k}{M+1})\left[\left[\sum_{s=1}^{T}\kappa_{t,s}Y_s\right]\left[\sum_{s=1}^{T}\kappa_{t-k,s}Y_s\right]' u_t u_{t-k}\right] \\
&+ \sum_{k=1}^{M}\sum_{t=1+k}^{T}(1-\frac{k}{M+1})\left[\left[\sum_{s=1}^{T}\kappa_{t-k,s}Y_s\right]\left[\sum_{s=1}^{T}\kappa_{t,s}Y_s\right]' u_t u_{t-k}\right]. \qquad (27)
\end{aligned}$$

■

Lemma 1. *Under Assumptions 1 to 3(i),(ii),(iii), there exists an invertible matrix H such that*

$$\hat{F}_t - H'F_t = \mathcal{O}_p\left(\frac{1}{\min(\sqrt{w_q}, T)}\right). \qquad (28)$$

Proof of Lemma 1. Lemma 1 follows directly from Theorem 1 in Bai (2003): see e.g. the discussions that follow Theorem 1 on page 12. We first recall the maintained assumptions in Theorem 1 in Bai (2003) (Assumptions A to G). Then, we verify that Theorem 1 in Bai (2003) applies in our framework.

• Assumption A. $E\|F_t\|^4 \leq M < \infty$ and $T^{-1}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$, some (k,k) positive definite matrix.

• Assumption B. Λ has bounded elements λ_i, and $\|\Lambda'\Lambda/w_q - \Sigma_\Lambda\| \to 0$ for some (k,k) positive definite matrix.

• Assumption C. There exists a positive constant M such that for all w_q and T,

1. $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M < \infty$.
2. $E(e_s'e_t/w_q) = E(\sum_{i=1}^{w_q} e_{is}e_{it}/w_q) = \gamma_{w_q}(s,t)$, $|\gamma_{w_q}(s,t)| \leq M$ for all s and $\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{w_q}(s,t)|/T \leq M$.
3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t. In addition, $\sum_{i=1}^{w_q}\sum_{j=1}^{w_q}|\tau_{ij}| \leq M$.
4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $\sum_{i=1}^{w_q}\sum_{j=1}^{w_q}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leq M$.
5. For every (t,s), $E|\sum_{i=1}^{w_q}[e_{is}e_{it} - E(e_{is}e_{it})]/\sqrt{w_q}|^4 \leq M$.

• Assumption D.

$$E\left(\frac{1}{w_q}\sum_{i=1}^{w_q}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it}\|^2\right)\leq M$$

• Assumption E. There exists $M<\infty$ such that, for all T and w_q, and for every $t\leq T$ and every $i\leq w_q$,

1. $\sum_{s=1}^{T}|\gamma_N(s,t)|\leq M$.

2. $\sum_{k=1}^{w_q}|\tau_{ki}|\leq M$.

• Assumption F. There exists $M<\infty$ such that, for all T and w_q,

1. For each t,

$$E\|\frac{1}{\sqrt{w_q T}}\sum_{s=1}^{T}\sum_{k=1}^{w_q}F_s[e_{ks}e_{kt}-E(e_{ks}e_{kt})]\|^2\leq M$$

2. The (k,k) matrix satisfies

$$E\|\frac{1}{\sqrt{w_q T}}\sum_{t=1}^{T}\sum_{k=1}^{w_q}F_t\lambda_k' e_{kt}\|^2\leq M$$

3. For each t, as $w_q\to\infty$,

$$\frac{1}{\sqrt{w_q}}\sum_{i=1}^{w_q}\lambda_i e_{it}\xrightarrow{d}\mathcal{N}(0,\Gamma_t)\quad\text{where }\Gamma_t=\lim_{w_q\to\infty}\sum_{i=1}^{w_q}\sum_{j=1}^{w_q}\lambda_i\lambda_j' E(e_{it}e_{jt})$$

4. For each i, as $T\to\infty$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it}\xrightarrow{d}\mathcal{N}(0,\Phi_i)\quad\text{where }\Phi_i=\texttt{plim}_T\sum_{s=1}^{T}\sum_{t=1}^{T}E[F_tF_s' e_{is}e_{it}]$$

• Assumption G. The eigenvalues of the (k,k) matrix $(\Sigma_\Lambda.\Sigma_F)$ defined in Assumptions A and B are distinct.

Assumptions A and B are the same as Assumption 3(i). Assumption C is the same as Assumption 3(ii)-(iii). Assumption D maintains weak dependence between factors and idiosyncratic errors, and it is more general than Assumption 3(ii) which imposes independence between them. Assumption E maintains weak dependence on the idiosyncratic errors, and it is more general than Assumption 3(ii) which imposes their time series and cross-section independence. As explained in section 3.1 in Bai (2003), F and Λ are not separately identifiable; however, they can be estimated up to an invertible (k, k) matrix transformation. Bai (2003) relies on his Assumption G to define the invertible matrix H. We follow instead Bai and Ng (2006) (see e.g. their section 2.1) and define the invertible matrix H - without imposing Assumption G - as

$$H = \tilde{V}^{-1} \left(\frac{\hat{F}'F}{T} \right) \left(\frac{\Lambda'\Lambda}{w_q} \right) ,$$

with $\tilde{V}$ the (k, k) diagonal matrix consisting of the k largest eigenvalues of $WW'/(w_qT)$ with W the (T, w_q)-matrix with row t as W_t'. ■

• **Proof of Theorem 2** :

The proof of Theorem 2 closely follows the proof of Theorem 1. In particular, it means that we need to extend Lemma 1 to our new framework: this is done in Lemma 2 below by using several results from Kapetanios and Marcellino (2010) (e.g. Theorem 4) to update the rate of convergence, accordingly. Specifically, instead of the rate in Lemma 1, $\min\{\sqrt{w_q}, T\}$, we use the rate derived in Kapetanios and Marcellino (2010),

$$\delta_{w_qT\alpha} = w_q^{-2\alpha} \times \min\{\sqrt{w_q}, T\} .$$

Ultimately - and similarly to the proof of Theorem 1 - we end up with a new quantity c_T that mimics the updated rate and is shared by both $A_T(\hat{F})$ and $B_T(\hat{F})$. Consequently, the asymptotic properties of $\hat{\beta}_T$ are as in Theorem 1. ■

Lemma 2. *Under Assumptions 1, 2, 3(ii),(iii) and 4, there exists an invertible matrix H such that*

$$\hat{F}_t - H'F_t = \mathcal{O}_p \left(\frac{1}{w_q^{-2\alpha} \min\{\sqrt{w_q}, T\}} \right) . \tag{29}$$

Proof of Lemma 2. We start from the decomposition of $\hat{F}_t - H'F_t$ which is similar to the one used in the proof of Theorem 4 in Kapetanios and Marcellino (2010). In particular, we rely on the same normalization for Λ, that is, we use: $\Lambda'\Lambda/w_q^{1-2\alpha} = I$ which leads to:

$$\hat{F} = w_q^{-1+2\alpha} W \tilde{\Lambda} \quad \text{and} \quad \tilde{\Lambda} = T^{-1} W' \tilde{F} ,$$

with $\tilde{F}$ the solution of the optimization problem of maximizing $tr(F'(W'W)F)$ subject to $F'F/T = I$. Let $H = ((\tilde{F}'F/T)(\Lambda_{w_q}'\Lambda_{w_q}/w_q^{1-2\alpha}))'$. Then,

$$\hat{F}_t - H'F_t = w_q^{2\alpha} T^{-1} \left(\sum_{s=1}^{T} \tilde{F}_s \gamma_{w_q}(s,t) + \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + \sum_{s=1}^{T} \tilde{F}_s \eta_{st} + \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \right)$$

$$\Rightarrow \quad w_q^{-2\alpha}(\hat{F}_t - H'F_t) = T^{-1} \left(\sum_{s=1}^{T} \tilde{F}_s \gamma_{w_q}(s,t) + \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + \sum_{s=1}^{T} \tilde{F}_s \eta_{st} + \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \right)$$

$$\begin{aligned} \text{where} \qquad \zeta_{st} &= \frac{e_s' e_t}{w_q} - \gamma_{w_q}(s,t) \\ \eta_{st} &= F_s' \Lambda_{w_q} e_t / w_q \\ \xi_{st} &= F_t' \Lambda_{w_q} e_s / w_q = \eta_{ts}\,. \end{aligned}$$

We analyze each term separately to show that, under the assumptions maintained in Lemma 2, with $\delta_{w_q T} = \min\{\sqrt{w_q}, \sqrt{T}\}$, we have:

$$\begin{aligned} (a) \qquad & T^{-1} \sum_{s=1}^{T} \tilde{F}_s \gamma_{w_q}(s,t) = \mathcal{O}_p\left(\frac{1}{\sqrt{T}\delta_{w_q T}} \right) \\ (b) \qquad & T^{-1} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} = \mathcal{O}_p\left(\frac{1}{\sqrt{w_q}\delta_{w_q T}} \right) \\ (c) \qquad & T^{-1} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} = \mathcal{O}_p\left(\frac{1}{\sqrt{w_q}\delta_{w_q T}} \right) \\ (d) \qquad & T^{-1} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} = \mathcal{O}_p\left(\frac{1}{\sqrt{w_q}\delta_{w_q T}} \right) \end{aligned}$$

These results follow, in part, from results proved by Kapetanios and Marcellino (2010) in the proof of Theorem 4 as well as results proved by Bai (2003) in the proof of Lemma A.2. It then follows that the sum of these four terms is $\mathcal{O}_p(1/\min\{T, \sqrt{w_q}\})$.
(a)

$$\begin{aligned} T^{-1} \left| \sum_{s=1}^{T} \tilde{F}_s \gamma_{w_q}(t,s) \right| &\leq \left| T^{-1} \sum_{s=1}^{T} \|\tilde{F}_s\|^2 \right|^{1/2} \times T^{-1/2} \left| \sum_{s=1}^{T} \gamma_{w_q}^2(t,s) \right|^{1/2} \\ &= \mathcal{O}_p(1/\delta_{w_q T}) \times \frac{1}{\sqrt{T}} \mathcal{O}_p(1) \\ &= \mathcal{O}_p(1/(\sqrt{T}\delta_{w_q T})) \end{aligned}$$

(b)

$$\begin{aligned} T^{-1}|\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}| &\leq |T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2|^{1/2} \times |T^{-1}\sum_{s=1}^{T}\zeta_{st}^2|^{1/2} \\ &= \mathcal{O}_p(1/\delta_{w_qT}) \times |\mathcal{O}_p(1/\sqrt{w_q}) \\ &= \mathcal{O}_p(1/(\sqrt{w_q}\delta_{w_qT})) \end{aligned}$$

(c)

$$\begin{aligned} T^{-1}|\sum_{s=1}^{T}\tilde{F}_s\eta_{st}| &\leq |T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2|^{1/2} \times |T^{-1}\sum_{s=1}^{T}\eta_{st}^2|^{1/2} \\ &= \mathcal{O}_p(1/\sqrt{w_q}\delta_{w_qT}) \end{aligned}$$

since

$$\begin{aligned} T^{-1}\sum_{s=1}^{T}\eta_{st}^2 &= T^{-1}\sum_{s=1}^{T}(F_s'\Lambda e_t)^2 w_q^{-2-2\alpha} \\ &\leq T^{-1}\sum_{s=1}^{T}\|F_s\|^2 \times \|\Lambda e_t\|^2 w_q^{-2-2\alpha} \\ &= \mathcal{O}_p(w_q^{-1-2\alpha}) \end{aligned}$$

(d)

$$\begin{aligned} T^{-1}|\sum_{s=1}^{T}\tilde{F}_s\xi_{st}| &\leq |T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2|^{1/2} \times |T^{-1}\sum_{s=1}^{T}\xi_{st}^2|^{1/2} \\ &= \mathcal{O}_p(1/\sqrt{w_q}\delta_{w_qT}) \end{aligned}$$

■

B Monte-Carlo results

B.1 Identification

- Simulation designs #1 and #2

We first discuss identification of β_0 in the simulation design #1 when one instrument Z_1 is used: similar results apply to our simulation design #2 and are not explicitly discussed here. These identification properties are related to general results in Sun (2022), and are provided in our specific simulation setup for completeness. We show that β_0 is identified as long as, (i) Z_2 is conditionally mean-dependent of Z_1, or, (ii) $p_z \neq 0.5$ when Z_2 is conditionally mean-independent of Z_1.

Recall that, for the SMD estimator, the general identification condition can be written as a rank condition that ensures that $E(\kappa_{j,l} Y_j Y_l')$ is invertible. When $E(\kappa_{j,l} Y_j Y_l')$ is not invertible, it means that there exists $a \neq 0$ such that:

$$a' E(\kappa_{j,l} Y_j Y_l') a = 0 \quad \Leftrightarrow \quad a' E(Y_j | Z_j) = 0 \ a.s.$$

When implementing the SMD estimator with only one instrument, Z_1, this condition becomes,

$$\begin{aligned} a' E(Y_j | Z_{1,j}) &= a' E\left[10(2Z_{2,j} - 1)(Z_{1,j} - 2Z_{1,j}^3/5) + v_j | Z_{1,j}\right] \\ &= 10 a' E\left[(2Z_{2,j} - 1)|Z_{1,j}\right](Z_{1,j} - 2Z_{1,j}^3/5) \\ &= 10 a' \left[2E(Z_{2,j}|Z_{1,j}) - 1\right](Z_{1,j} - 2Z_{1,j}^3/5) \\ &= 0\,. \end{aligned}$$

Notice that, when Z_2 is conditionally mean-independent of Z_1 and $Pr(Z_2 = 1) = 0.5$, then, for any a, we have $a' E(Y_j | Z_{1,j}) = 0$, and, as a result, the parameter β_0 is not identified.

- Simulation design #3

We verify Assumptions 1(i) and (ii) in the simulation design #3. In this DGP, we have $(\sigma_t u_t)$ independent of F_t; hence, it follows that $E(\sigma_t u_t | \mathcal{I}(F_t)) = 0$.
We also have:

$$\begin{aligned} E(Y_t Y_t') &= E(Y_t^2) \quad \text{since } Y_t \text{ is univariate here} \\ &= E(100(2F_{2,t} - 1)^2 (F_{1,t} - 2F_{1,t}^3/5)^2 + v_t^2 + 20(2F_{2,t} - 1)(F_{1,t} - 2F_{1,t}^3/5) v_t) \\ &= 100 \times E(2F_{2,t} - 1)^2 \times E(F_{1,t} - 2F_{1,t}^3/5)^2 + 1 \\ &\quad \text{since } v_t, F_{1,t}, F_{2,t} \text{ are independent} \\ &> 0 \end{aligned}$$

and

$$
\begin{aligned}
E(Y_t|\mathcal{I}(F_t)) &= E(Y_t|F_t) \\
&= 10 \times (2F_{2,t} - 1) \times (F_{1,t} - 2F_{1,t}^3/5) \\
&\neq 0 \; a.s. \quad \text{since } F_{1,t} \sim U([-2,2]) \text{ and } 2F_{2,t} - 1 \neq 0 \; \forall \; t
\end{aligned}
$$

B.2 Implementation details

- F-SMD

All the results presented in the main paper are obtained with $c = 0$, and a Gaussian kernel. Results with other values of c are presented in the Supplementary Appendix and reveal that the value of c does not seem to play an important role in our framework even for smaller sample sizes.

- B-2SLS and B-GMM

We first provide expressions for the B-2SLS and B-GMM estimators. The B-2SLS estimator is defined as:

$$\hat{\beta}_{B2SLS} = (\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t')^{-1} \sum_{t=1}^{T} \hat{Y}_t y_t$$

with

$$\hat{Y}_t' = \begin{cases} Z_{1,t}' \hat{\Pi}_{t \leq T_{break}}, & \text{if } t \leq T_{break} \\ Z_{1,t}' \hat{\Pi}_{t > T_{break}}, & \text{if } t > T_{break}. \end{cases}$$

and a consistent estimator of its asymptotic variance is obtained as:

$$\hat{V}ar\hat{\beta}_{B2SLS} = (\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t')^{-1} (\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t' \hat{u}_t^2) (\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t')^{-1}$$

with $\hat{u}_t = y_t - \hat{Y}_t \beta_{B2SLS}$

The B-GMM estimator is defined as:

$$\hat{\beta}_{BGMM} = (Y'Z(Z'Z)^{-1}Z'Y)^{-1}Y'Z(Z'Z)^{-1}Z'y$$

with

$$Z = \begin{pmatrix} Z_{1,t \leq T_{break}} & 0 \\ 0 & Z_{1,t > T_{break}} \end{pmatrix}$$

where Z_1 is split as follows,

$$Z_1 = \begin{pmatrix} Z_{1,t \leq T_{break}} \\ Z_{1,t > T_{break}} \end{pmatrix}$$

And a consistent estimator of its asymptotic variance is obtained as:

$$\hat{V}ar\hat{\beta}_{B2SLS} = (\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t')^{-1} (\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t' \hat{u}_t^2)(\sum_{t=1}^{T} \hat{Y}_t \hat{Y}_t')^{-1}$$

The efficient B-GMM estimator is defined as:

$$\hat{\beta}_{BGMM.eff} = (Y'Z(Z'\hat{\Omega}_{1,T}Z)^{-1}Z'Y)^{-1}Y'Z(Z'\hat{\Omega}_{1,T}Z)^{-1}Z'y$$

with $\hat{\Omega}_{1,T}$ the variance-covariance matrix of the B-GMM residuals, $y_t - Y_t'\beta_{BGMM}$. A consistent estimator of its asymptotic variance is obtained as:

$$\hat{V}ar\hat{\beta}_{BGMM.eff} = (Y'Z(Z'\hat{\Omega}_T Z)^{-1}Z'Y)^{-1},$$

with $\hat{\Omega}_{B,T}$ a consistent estimator of the variance-covariance matrix of the residuals, $y_t - Y_t'\hat{\beta}_{BGMM.eff}$.

- Computation of standard errors.

We now detail how standard errors are computed throughout. In Experiment 1 in the iid and HET1 cases, we compute robust standard errors. In Experiment 1 in the HET2 case and in Experiments 2 and 3, we compute HAC standard errors. In the empirical application, we use HAC standard errors throughout.

B.3 Experiment #1: Small number of observed exogenous instruments

B.4 Erratic behavior of 2SLS and GMM

We specifically discuss the erratic behavior of 2SLS and GMM in Table 2. In this table, "Asympt.Heterosk.SE" presents the average of standard errors computed from the 5,000 rounds using the estimator of the variance formula, while "Median of SE" displays the associated median. The difference between the mean and the median suggests the presence of extreme cases for GMM and 2SLS when using only Z_1 as an instrument. Indeed, upon further inspecting the standard errors, we found that the maximum standard error is 5.193×10^6. This pattern happens in the small sample case when $T = 200$, demonstrating that extreme cases are more likely to happen when the sample size is small. When we disregard these extreme cases, such as by setting a maximum standard error and excluding simulations producing standard errors greater than 10, we eliminate 5.56% of the 5,000 replications. Consequently, the average of those standard errors becomes 0.623, and the median becomes 0.185 for both GMM and 2SLS. In this scenario, the standard deviation calculated from the estimates of the 5,000 rounds is 0.303, significantly lower than the value of 20.863 in the "SE" row of the same table. This is shown in Table 7. See also Table 9 for Experiment #2a.

PANEL A: break either ignored or estimated

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
	Break is ignored: Z_1 *only*			*Break location is estimated:* $(Z_1, \hat{Z}_2)$	
Panel A.1: sample size $T = 200$					
Bias	-0.003	-1.243	-1.243	0.018	0.018
SE	0.048	64.937	64.937	0.068	0.068
Asympt.Heterosk.SE	0.048	21336.756	21336.759	0.063	0.065
t-statistic	-0.006	0.062	0.062	0.295	0.308
Median bias	-0.002	0.010	0.010	0.017	0.017
Median Absolute Deviation	0.032	0.142	0.142	0.043	0.043
Median of SE	0.047	0.275	0.275	0.061	0.062
Rej. rate for Heterosk. SE	0.036	0.001	0.001	0.079	0.057
Panel A.2: sample size $T = 2,000$					
Bias	0.000	-0.040	-0.040	0.007	0.007
SE	0.013	2.953	2.953	0.044	0.044
Asympt.Heterosk.SE	0.013	35.552	35.552	0.043	0.044
t-statistic	0.011	0.052	0.052	0.167	0.176
Median bias	0.000	0.002	0.002	0.007	0.007
Median Absolute Deviation	0.009	0.058	0.058	0.027	0.027
Median of SE	0.013	0.088	0.088	0.041	0.042
Rej. rate for Heterosk. SE	0.054	0.005	0.005	0.050	0.032

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	-0.001	0.022	0.022	0.013	0.013
SE	0.034	0.329	0.329	0.133	0.133
Asympt.Heterosk.SE	0.034	0.530	0.511	0.113	0.152
t-statistic	-0.005	0.131	0.131	0.149	0.149
Median bias	-0.001	0.019	0.019	0.014	0.014
Median Absolute Deviation	0.023	0.096	0.096	0.065	0.065
Median of SE	0.034	0.163	0.162	0.092	0.099
Rej. rate for Heterosk. SE	0.047	0.003	0.003	0.057	0.010
Panel B.2: sample size $T = 2,000$					
Bias	0.000	0.002	0.002	0.002	0.002
SE	0.010	0.130	0.130	0.052	0.052
Asympt.Heterosk.SE	0.010	0.150	0.147	0.049	0.052
t-statistic	0.012	0.081	0.081	0.071	0.088
Median bias	0.000	0.005	0.005	0.003	0.003
Median Absolute Deviation	0.007	0.053	0.053	0.030	0.030
Median of SE	0.010	0.080	0.080	0.045	0.046
Rej. rate for Heterosk. SE	0.054	0.006	0.006	0.050	0.024

Table 1: Experiment #1: Small number of observed exogenous instruments with first-stage instability under homoskedasticity when the true break fraction is 0.2. In panel A, the break is either ignored (first 3 columns) or estimated (remaining 2 columns) with a sample size of $T = 200$ (panel A.1) or $T = 2,000$ (panel A.2); in panel B, the break is fully known with a sample of $T = 200$ (panel B.1) or $T = 2,000$ (panel B.2).

PANEL A: break either ignored or estimated

Estimator	*Break is ignored: Z_1 only* F-SMD	2SLS	GMM	*Break location is estimated: $(Z_1, \hat{Z}_2)$* B2SLS	BGMM
Panel A.1: sample size $T = 200$					
Bias	-0.001	-0.035	-0.035	0.016	0.016
SE	0.029	20.863	20.863	0.068	0.068
Asympt.Heterosk.SE	0.028	2029.220	2029.220	0.063	0.065
t-statistic	-0.007	0.060	0.060	0.268	0.282
Median bias	-0.001	0.010	0.010	0.015	0.015
Median Absolute Deviation	0.019	0.114	0.114	0.043	0.043
Median of SE	0.028	0.200	0.200	0.061	0.061
Rej. rate for Heterosk. SE	0.051	0.001	0.001	0.081	0.060
Panel A.2: sample size $T = 2,000$					
Bias	0.000	0.000	0.000	0.006	0.006
SE	0.009	0.319	0.319	0.044	0.044
Asympt.Heterosk.SE	0.008	0.609	0.609	0.042	0.044
t-statistic	0.010	0.042	0.042	0.165	0.177
Median bias	0.000	0.001	0.001	0.008	0.008
Median Absolute Deviation	0.006	0.037	0.037	0.029	0.029
Median of SE	0.008	0.056	0.056	0.041	0.042
Rej. rate for Heterosk. SE	0.055	0.014	0.014	0.054	0.033

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	-0.001	0.014	0.014	0.011	0.011
SE	0.027	0.303	0.303	0.139	0.139
Asympt.Heterosk.SE	0.027	0.683	0.993	0.108	0.141
t-statistic	-0.007	0.110	0.111	0.143	0.158
Median bias	-0.001	0.014	0.014	0.013	0.013
Median Absolute Deviation	0.018	0.087	0.087	0.066	0.066
Median of SE	0.027	0.137	0.136	0.088	0.089
Rej. rate for Heterosk. SE	0.050	0.005	0.006	0.080	0.049
Panel B.2: sample size $T = 2,000$					
Bias	0.000	0.000	0.000	0.002	0.002
SE	0.008	0.067	0.067	0.051	0.051
Asympt.Heterosk.SE	0.008	0.065	0.065	0.049	0.051
t-statistic	0.011	0.069	0.069	0.068	0.088
Median bias	0.000	0.003	0.003	0.003	0.003
Median Absolute Deviation	0.006	0.036	0.036	0.031	0.031
Median of SE	0.008	0.054	0.054	0.045	0.046
Rej. rate for Heterosk. SE	0.053	0.015	0.015	0.050	0.026

Table 2: Experiment #1: Small number of observed exogenous instruments with first-stage instability under homoskedasticity when the true break fraction is 0.05. In panel A, the break is either ignored (first 3 columns) or estimated (remaining 2 columns) with a sample size of $T = 200$ (panel A.1) or $T = 2,000$ (panel A.2); in panel B, the break is fully known with a sample of $T = 200$ (panel B.1) or $T = 2,000$ (panel B.2).

PANEL A: break either ignored or estimated

Estimator	*Break is ignored:* Z_1 *only* F-SMD	2SLS	GMM	*Break location is estimated:* $(Z_1, \hat{Z}_2)$ B2SLS	BGMM
Panel A.1: sample size $T = 200$					
Bias	-0.002	-1.401	-1.401	0.019	0.019
SE	0.047	69.124	69.124	0.076	0.076
Asympt.Heterosk.SE	0.047	22398.280	22398.296	0.071	0.071
t-statistic	-0.010	0.064	0.064	0.292	0.312
Median bias	-0.002	0.013	0.013	0.020	0.020
Median Absolute Deviation	0.031	0.172	0.172	0.049	0.049
Median of SE	0.046	0.334	0.334	0.068	0.068
Rej. rate for Heterosk. SE	0.032	0.001	0.001	0.082	0.064
Panel A.2: sample size $T = 2,000$					
Bias	0.000	-0.057	-0.057	0.008	0.008
SE	0.013	4.138	4.138	0.052	0.052
Asympt.Heterosk.SE	0.013	50.052	50.052	0.051	0.052
t-statistic	0.010	0.051	0.051	0.162	0.172
Median bias	0.000	0.003	0.003	0.008	0.008
Median Absolute Deviation	0.009	0.068	0.068	0.033	0.033
Median of SE	0.013	0.106	0.106	0.049	0.050
Rej. rate for Heterosk. SE	0.051	0.005	0.005	0.051	0.031

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	-0.001	0.022	0.022	0.013	0.013
SE	0.034	0.368	0.368	0.156	0.156
Asympt.Heterosk.SE	0.033	0.612	0.584	0.129	0.177
t-statistic	-0.009	0.130	0.131	0.142	0.145
Median bias	-0.001	0.023	0.023	0.016	0.016
Median Absolute Deviation	0.022	0.107	0.107	0.073	0.073
Median of SE	0.033	0.182	0.178	0.105	0.112
Rej. rate for Heterosk. SE	0.047	0.003	0.004	0.060	0.011
Panel B.2: sample size $T = 2,000$					
Bias	0.000	0.002	0.002	0.002	0.002
SE	0.010	0.157	0.157	0.063	0.063
Asympt.Heterosk.SE	0.010	0.186	0.178	0.058	0.062
t-statistic	0.012	0.076	0.076	0.068	0.084
Median bias	0.000	0.006	0.006	0.003	0.003
Median Absolute Deviation	0.006	0.062	0.062	0.036	0.036
Median of SE	0.010	0.095	0.094	0.054	0.055
Rej. rate for Heterosk. SE	0.050	0.006	0.007	0.051	0.023

Table 3: Experiment #1: Small number of observed exogenous instruments with first-stage instability under heteroskedasticity of type HET1 when the true break fraction is 0.2. In panel A, the break is either ignored (first 3 columns) or estimated (remaining 2 columns) with a sample size of $T = 200$ (panel A.1) or $T = 2,000$ (panel A.2); in panel B, the break is fully known with a sample of $T = 200$ (panel B.1) or $T = 2,000$ (panel B.2).

PANEL A: break either ignored or estimated

Estimator	*Break is ignored:* Z_1 *only* F-SMD	2SLS	GMM	*Break location is estimated:* $(Z_1, \hat{Z}_2)$ B2SLS	BGMM
Panel A.1: sample size $T = 200$					
Bias	-0.001	-0.056	-0.056	0.018	0.018
SE	0.028	24.319	24.319	0.077	0.077
Asympt.Heterosk.SE	0.028	2304.136	2304.136	0.071	0.071
t-statistic	-0.011	0.061	0.061	0.269	0.291
Median bias	-0.001	0.011	0.011	0.017	0.017
Median Absolute Deviation	0.019	0.139	0.139	0.048	0.048
Median of SE	0.028	0.244	0.244	0.068	0.067
Rej. rate for Heterosk. SE	0.048	0.001	0.001	0.083	0.070
Panel A.2: sample size $T = 2,000$					
Bias	0.000	0.000	0.000	0.007	0.007
SE	0.008	0.384	0.384	0.051	0.051
Asympt.Heterosk.SE	0.008	0.736	0.736	0.050	0.051
t-statistic	0.010	0.041	0.041	0.154	0.166
Median bias	0.000	0.002	0.002	0.008	0.008
Median Absolute Deviation	0.006	0.045	0.045	0.033	0.033
Median of SE	0.008	0.068	0.068	0.048	0.049
Rej. rate for Heterosk. SE	0.052	0.013	0.013	0.055	0.032

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	-0.001	0.015	0.015	0.012	0.012
SE	0.027	0.362	0.362	0.158	0.158
Asympt.Heterosk.SE	0.026	0.953	1.202	0.122	0.153
t-statistic	-0.011	0.105	0.106	0.139	0.161
Median bias	-0.001	0.015	0.015	0.014	0.014
Median Absolute Deviation	0.018	0.098	0.098	0.075	0.075
Median of SE	0.026	0.156	0.153	0.100	0.097
Rej. rate for Heterosk. SE	0.050	0.005	0.009	0.080	0.063
Panel B.2: sample size $T = 2,000$					
Bias	0.000	0.000	0.000	0.002	0.002
SE	0.008	0.080	0.080	0.061	0.061
Asympt.Heterosk.SE	0.008	0.077	0.076	0.058	0.061
t-statistic	0.011	0.062	0.061	0.062	0.081
Median bias	0.000	0.003	0.003	0.004	0.004
Median Absolute Deviation	0.005	0.043	0.043	0.037	0.037
Median of SE	0.008	0.064	0.064	0.054	0.055
Rej. rate for Heterosk. SE	0.053	0.014	0.015	0.049	0.026

Table 4: Experiment #1: Small number of observed exogenous instruments with first-stage instability under heteroskedasticity of type HET1 when the true break fraction is 0.05. In panel A, the break is either ignored (first 3 columns) or estimated (remaining 2 columns) with a sample size of $T = 200$ (panel A.1) or $T = 2,000$ (panel A.2); in panel B, the break is fully known with a sample of $T = 200$ (panel B.1) or $T = 2,000$ (panel B.2).

PANEL A: break either ignored or estimated

Estimator	*Break is ignored:* Z_1 *only* F-SMD	2SLS	GMM	*Break location is estimated:* $(Z_1, \hat{Z}_2)$ B2SLS	BGMM
Panel A.1: sample size $T = 200$					
Bias	-0.002	0.049	0.049	0.014	0.014
SE	0.050	15.569	15.569	0.065	0.065
Asympt.Heterosk.SE	0.043	1529.337	1529.337	0.054	0.050
t-statistic	0.003	0.065	0.065	0.267	0.303
Median bias	-0.001	0.010	0.010	0.012	0.012
Median Absolute Deviation	0.026	0.119	0.119	0.032	0.032
Median of SE	0.038	0.231	0.231	0.046	0.044
Rej. rate for Heterosk. SE	0.030	0.001	0.001	0.073	0.077
Panel A.2: sample size $T = 2,000$					
Bias	0.000	0.039	0.039	0.005	0.005
SE	0.012	2.101	2.101	0.042	0.042
Asympt.Heterosk.SE	0.012	14.891	14.891	0.040	0.039
t-statistic	-0.004	0.053	0.053	0.124	0.139
Median bias	0.000	0.003	0.003	0.005	0.005
Median Absolute Deviation	0.008	0.052	0.052	0.026	0.026
Median of SE	0.011	0.080	0.080	0.037	0.037
Rej. rate for Heterosk. SE	0.046	0.006	0.006	0.049	0.039

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	-0.001	0.018	0.018	0.009	0.009
SE	0.035	0.389	0.389	0.130	0.130
Asympt.Heterosk.SE	0.030	0.654	0.720	0.097	0.130
t-statistic	0.000	0.125	0.127	0.130	0.137
Median bias	-0.001	0.015	0.015	0.009	0.009
Median Absolute Deviation	0.019	0.076	0.076	0.051	0.051
Median of SE	0.027	0.136	0.132	0.074	0.075
Rej. rate for Heterosk. SE	0.043	0.002	0.004	0.055	0.018
Panel B.2: sample size $T = 2,000$					
Bias	0.000	0.005	0.005	0.001	0.001
SE	0.009	0.150	0.150	0.050	0.050
Asympt.Heterosk.SE	0.009	0.152	0.151	0.045	0.046
t-statistic	-0.009	0.084	0.084	0.046	0.064
Median bias	0.000	0.005	0.005	0.002	0.002
Median Absolute Deviation	0.006	0.047	0.047	0.028	0.028
Median of SE	0.009	0.073	0.072	0.041	0.041
Rej. rate for Heterosk. SE	0.047	0.009	0.009	0.051	0.031

Table 5: Experiment #1: Small number of observed exogenous instruments with first-stage instability under heteroskedasticity of type HET2 when the true break fraction is 0.2. In panel A, the break is either ignored (first 3 columns) or estimated (remaining 2 columns) with a sample size of $T = 200$ (panel A.1) or $T = 2,000$ (panel A.2); in panel B, the break is fully known with a sample of $T = 200$ (panel B.1) or $T = 2,000$ (panel B.2).

PANEL A: break either ignored or estimated

Estimator	*Break is ignored:* Z_1 *only* F-SMD	2SLS	GMM	*Break location is estimated:* $(Z_1, \hat{Z}_2)$ B2SLS	BGMM
Panel A.1: sample size $T = 200$					
Bias	0.000	-0.192	-0.192	0.013	0.013
SE	0.028	14.302	14.302	0.063	0.063
Asympt.Heterosk.SE	0.025	1156.288	1156.288	0.054	0.050
t-statistic	0.014	0.054	0.054	0.250	0.277
Median bias	-0.001	0.006	0.006	0.011	0.011
Median Absolute Deviation	0.016	0.092	0.092	0.033	0.033
Median of SE	0.022	0.169	0.169	0.046	0.044
Rej. rate for Heterosk. SE	0.046	0.001	0.001	0.070	0.084
Panel A.2: sample size $T = 2,000$					
Bias	0.000	-0.002	-0.002	0.006	0.006
SE	0.008	0.175	0.175	0.041	0.041
Asympt.Heterosk.SE	0.008	0.160	0.160	0.039	0.038
t-statistic	-0.002	0.045	0.045	0.149	0.166
Median bias	0.000	0.001	0.001	0.005	0.005
Median Absolute Deviation	0.005	0.034	0.034	0.025	0.025
Median of SE	0.007	0.051	0.051	0.036	0.036
Rej. rate for Heterosk. SE	0.050	0.013	0.013	0.049	0.038

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	0.000	0.012	0.012	0.008	0.008
SE	0.027	0.288	0.288	0.131	0.131
Asympt.Heterosk.SE	0.024	0.441	0.420	0.089	0.100
t-statistic	0.012	0.103	0.111	0.124	0.147
Median bias	-0.001	0.009	0.009	0.007	0.007
Median Absolute Deviation	0.015	0.064	0.064	0.048	0.048
Median of SE	0.021	0.109	0.102	0.067	0.061
Rej. rate for Heterosk. SE	0.049	0.005	0.014	0.078	0.081
Panel B.2: sample size $T = 2,000$					
Bias	0.000	0.002	0.002	0.002	0.002
SE	0.008	0.061	0.061	0.049	0.049
Asympt.Heterosk.SE	0.008	0.060	0.058	0.045	0.045
t-statistic	-0.003	0.069	0.070	0.059	0.079
Median bias	0.000	0.002	0.002	0.003	0.003
Median Absolute Deviation	0.005	0.033	0.033	0.028	0.028
Median of SE	0.007	0.049	0.048	0.041	0.040
Rej. rate for Heterosk. SE	0.049	0.015	0.018	0.046	0.030

Table 6: Experiment #1: Small number of observed exogenous instruments with first-stage instability under heteroskedasticity of type HET2 when the true break fraction is 0.05. In panel A, the break is either ignored (first 3 columns) or estimated (remaining 2 columns) with a sample size of $T = 200$ (panel A.1) or $T = 2,000$ (panel A.2); in panel B, the break is fully known with a sample of $T = 200$ (panel B.1) or $T = 2,000$ (panel B.2).

PANEL A: break either ignored or estimated

	Break is ignored: Z_1 *only*			*Break location is estimated:* $(Z_1, \hat{Z}_2)$	
Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel A.1: sample size $T = 200$					
Bias	-0.001	0.005	0.005	0.015	0.015
SE	0.028	0.303	0.303	0.067	0.067
Asympt.Heterosk.SE	0.028	0.623	0.623	0.063	0.064
t-statistic	-0.005	0.064	0.064	0.261	0.276
Median bias	-0.001	0.009	0.009	0.015	0.015
Median Absolute Deviation	0.019	0.105	0.105	0.043	0.043
Median of SE	0.028	0.185	0.185	0.060	0.061
Rej. rate for Heterosk. SE	0.051	0.001	0.001	0.079	0.060

PANEL B: break is fully known - use (Z_1, Z_2)

Estimator	F-SMD	2SLS	GMM	B2SLS	BGMM
Panel B.1: sample size $T = 200$					
Bias	-0.001	0.014	0.014	0.012	0.012
SE	0.027	0.206	0.206	0.130	0.130
Asympt.Heterosk.SE	0.027	0.267	0.270	0.104	0.126
t-statistic	-0.004	0.114	0.115	0.151	0.165
Median bias	-0.001	0.014	0.014	0.014	0.014
Median Absolute Deviation	0.018	0.083	0.083	0.064	0.064
Median of SE	0.027	0.133	0.131	0.087	0.088
Rej. rate for Heterosk. SE	0.050	0.005	0.006	0.079	0.049

Table 7: Experiment #1: this table corresponds to Table 2 after eliminating the replications that generate extreme behavior for 2SLS and GMM; see section B.4.

B.5 Experiment #2: Large number of observed instruments

	1 estim. factor			2 estim. factors			3 estim. factors		
Estimator	F-SMD	2SLS	GMM	F-SMD	2SLS	GMM	F-SMD	2SLS	GMM
PANEL A: sample size $T = 200$									
Bias	0.000	1.129	1.129	0.000	0.011	0.011	0.000	0.012	0.012
SE	0.016	78.527	78.527	0.016	0.244	0.244	0.017	0.141	0.141
Asympt.Heterosk.SE	0.015	21079.329	21079.321	0.015	0.349	0.345	0.016	0.150	0.145
t-statistic	0.003	0.045	0.045	0.000	0.097	0.097	0.000	0.133	0.136
Median bias	0.000	0.005	0.005	0.000	0.009	0.009	0.000	0.009	0.009
Median Absolute Deviation	0.009	0.091	0.091	0.009	0.066	0.066	0.010	0.057	0.057
Median of SE	0.013	0.168	0.168	0.014	0.110	0.106	0.014	0.091	0.087
Rej. rate for Heterosk. SE	0.052	0.001	0.001	0.057	0.003	0.005	0.053	0.006	0.011
PANEL B: sample size $T = 2,000$									
Bias	0.000	0.005	0.005	0.000	0.001	0.001	0.000	0.002	0.002
SE	0.006	0.394	0.394	0.006	0.070	0.070	0.007	0.061	0.061
Asympt.Heterosk.SE	0.005	0.318	0.318	0.005	0.063	0.062	0.005	0.056	0.055
t-statistic	0.009	0.032	0.032	0.011	0.061	0.061	0.016	0.077	0.077
Median bias	0.000	0.000	0.000	0.000	0.002	0.002	0.000	0.003	0.003
Median Absolute Deviation	0.003	0.034	0.034	0.003	0.033	0.033	0.003	0.032	0.032
Median of SE	0.005	0.052	0.052	0.005	0.049	0.048	0.005	0.047	0.047
Rej. rate for Heterosk. SE	0.051	0.011	0.011	0.050	0.013	0.014	0.050	0.016	0.019

Table 8: Experiment #2a: Large number of observed instruments that are strong and valid. The three estimators (F-SMD, 2SLS, and GMM) are either implemented using 1 estimated factor (first 3 columns), 2 estimated factors (middle 3 columns), or 3 estimated factors (last 3 columns). All factors are estimated by PCA with a sample size $T = 200$ (Panel A), or $T = 2,000$ (Panel B).

	1 estim. factor			2 estim. factors			3 estim. factors		
Estimator	F-SMD	2SLS	GMM	F-SMD	2SLS	GMM	F-SMD	2SLS	GMM
PANEL A: sample size $T = 200$									
Bias	0.000	0.005	0.005	0.000	0.011	0.011	0.000	0.011	0.011
SE	0.016	0.279	0.279	0.016	0.174	0.174	0.017	0.125	0.125
Asympt.Heterosk.SE	0.015	0.564	0.564	0.015	0.215	0.212	0.015	0.135	0.130
t-statistic	0.006	0.048	0.048	0.001	0.099	0.099	0.000	0.132	0.134
Median bias	0.000	0.005	0.005	0.000	0.008	0.008	0.000	0.009	0.009
Median Absolute Deviation	0.009	0.085	0.085	0.009	0.064	0.064	0.010	0.056	0.056
Median of SE	0.013	0.156	0.156	0.013	0.106	0.104	0.014	0.089	0.085
Rej. rate for Heterosk. SE	0.051	0.001	0.001	0.056	0.003	0.005	0.053	0.006	0.011

Table 9: Experiment #2a: this table corresponds to Table 8 after eliminating the replications that generate extreme behavior for 2SLS and GMM; see section B.4

.

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: $b \neq 0$**,** $\alpha = 0$**, and** $\upsilon = 0$						
Bias	0.013	-3.925	0.012	-0.067	0.012	-0.024
SE	0.057	212.728	0.058	1.923	0.061	1.047
Asympt.Heterosk.SE	0.057	56699.114	0.056	2.592	0.058	0.856
t-statistic	0.240	-0.222	0.226	-0.076	0.215	-0.035
Median bias	0.013	-0.181	0.013	-0.057	0.013	-0.028
Median Absolute Deviation	0.038	0.397	0.039	0.475	0.041	0.461
Median of SE	0.057	0.697	0.056	0.534	0.057	0.434
Rej. rate for Heterosk. SE	0.061	0.002	0.068	0.022	0.073	0.061
PANEL B: $b \neq 0$ **and** $\alpha = 0.125$**, and** $\upsilon = 0$						
Bias	0.017	0.118	0.016	-0.097	0.016	-0.074
SE	0.062	42.687	0.063	2.127	0.067	1.146
Asympt.Heterosk.SE	0.061	1733.132	0.061	3.226	0.063	0.936
t-statistic	0.289	-0.328	0.278	-0.168	0.266	-0.134
Median bias	0.018	-0.273	0.016	-0.127	0.016	-0.089
Median Absolute Deviation	0.042	0.416	0.043	0.524	0.046	0.479
Median of SE	0.061	0.738	0.060	0.557	0.063	0.441
Rej. rate for Heterosk. SE	0.066	0.003	0.069	0.031	0.073	0.069
PANEL C: $b \neq 0$ **and** $\alpha = 0.25$**, and** $\upsilon = 0$						
Bias	0.018	0.029	0.018	-0.218	0.018	-0.165
SE	0.073	17.296	0.077	2.259	0.082	1.277
Asympt.Heterosk.SE	0.071	330.265	0.073	3.581	0.076	1.211
t-statistic	0.261	-0.469	0.250	-0.358	0.240	-0.306
Median bias	0.019	-0.426	0.017	-0.259	0.017	-0.201
Median Absolute Deviation	0.049	0.482	0.053	0.509	0.056	0.477
Median of SE	0.071	0.834	0.073	0.565	0.076	0.454
Rej. rate for Heterosk. SE	0.061	0.007	0.065	0.039	0.071	0.077

Table 10: Experiment #2b: Large number of observed instruments that are possibly weak and invalid. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$. In all cases, the instruments are invalid with $b \neq 0$ and $\upsilon = 0$: its first 10 elements are set to 1 and the remaining 40 to 0. The identification strength varies with $\alpha = 0$, 0.125, or 0.25.

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: $b \neq 0$, $\alpha = 0$, **and** $\upsilon = 0.25$						
Bias	0.003	-2.011	0.003	-0.006	0.003	0.008
SE	0.022	77.300	0.023	0.591	0.024	0.309
Asympt.Heterosk.SE	0.022	20408.664	0.022	0.806	0.023	0.278
t-statistic	0.168	-0.119	0.160	0.010	0.153	0.060
Median bias	0.004	-0.042	0.003	-0.004	0.003	0.005
Median Absolute Deviation	0.015	0.153	0.015	0.148	0.016	0.140
Median of SE	0.022	0.268	0.022	0.191	0.023	0.158
Rej. rate for Heterosk. SE	0.056	0.001	0.059	0.012	0.062	0.035
PANEL B: $b \neq 0$ **and** $\alpha = 0.125$, **and** $\upsilon = 0.25$						
Bias	0.004	-0.057	0.004	-0.017	0.004	-0.005
SE	0.024	16.724	0.025	0.626	0.026	0.336
Asympt.Heterosk.SE	0.024	757.199	0.024	0.969	0.025	0.301
t-statistic	0.202	-0.194	0.199	-0.058	0.192	-0.011
Median bias	0.005	-0.067	0.004	-0.019	0.005	-0.010
Median Absolute Deviation	0.016	0.157	0.017	0.161	0.017	0.143
Median of SE	0.024	0.278	0.024	0.200	0.025	0.159
Rej. rate for Heterosk. SE	0.058	0.000	0.059	0.015	0.065	0.034
PANEL C: $b \neq 0$ **and** $\alpha = 0.25$, **and** $\upsilon = 0.25$						
Bias	0.005	0.086	0.005	-0.041	0.005	-0.027
SE	0.028	5.797	0.030	0.694	0.031	0.431
Asympt.Heterosk.SE	0.028	99.034	0.029	1.070	0.030	0.432
t-statistic	0.181	-0.303	0.176	-0.199	0.172	-0.137
Median bias	0.005	-0.105	0.004	-0.056	0.004	-0.038
Median Absolute Deviation	0.019	0.167	0.020	0.162	0.021	0.149
Median of SE	0.028	0.301	0.029	0.203	0.030	0.163
Rej. rate for Heterosk. SE	0.054	0.002	0.056	0.016	0.066	0.033

Table 11: Experiment #2b: Large number of observed instruments that are possibly weak and invalid. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$. In all cases, the instruments are (weakly) exogenous with $b \neq 0$ and $\upsilon = 0.25$: its first 10 elements are set to 1 and the remaining 40 to 0. The identification strength varies with $\alpha = 0$, 0.125, or 0.25.

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: $b \neq 0$, $\alpha = 0$, **and** $\upsilon = 0.5$						
Bias	0.001	-1.503	0.001	0.010	0.001	0.017
SE	0.018	46.906	0.018	0.313	0.018	0.161
Asympt.Heterosk.SE	0.018	11837.323	0.018	0.447	0.018	0.173
t-statistic	0.065	0.002	0.066	0.093	0.064	0.145
Median bias	0.001	-0.005	0.001	0.009	0.001	0.014
Median Absolute Deviation	0.012	0.117	0.012	0.088	0.012	0.077
Median of SE	0.017	0.203	0.018	0.138	0.018	0.113
Rej. rate for Heterosk. SE	0.053	0.001	0.059	0.006	0.060	0.012
PANEL B: $b \neq 0$ **and** $\alpha = 0.125$, **and** $\upsilon = 0.5$						
Bias	0.001	-0.103	0.001	0.004	0.001	0.013
SE	0.019	12.540	0.019	0.313	0.020	0.170
Asympt.Heterosk.SE	0.019	502.648	0.019	0.478	0.020	0.183
t-statistic	0.076	-0.024	0.083	0.067	0.081	0.119
Median bias	0.001	-0.011	0.001	0.004	0.002	0.011
Median Absolute Deviation	0.013	0.118	0.013	0.089	0.013	0.076
Median of SE	0.019	0.206	0.019	0.141	0.020	0.115
Rej. rate for Heterosk. SE	0.054	0.001	0.055	0.005	0.058	0.009
PANEL C: $b \neq 0$ **and** $\alpha = 0.25$, **and** $\upsilon = 0.5$						
Bias	0.001	0.101	0.001	0.006	0.001	0.009
SE	0.022	4.116	0.023	0.355	0.025	0.261
Asympt.Heterosk.SE	0.022	75.479	0.023	0.536	0.024	0.274
t-statistic	0.068	-0.064	0.069	0.013	0.072	0.073
Median bias	0.001	-0.019	0.001	0.003	0.002	0.006
Median Absolute Deviation	0.015	0.121	0.015	0.090	0.016	0.078
Median of SE	0.022	0.213	0.023	0.144	0.024	0.115
Rej. rate for Heterosk. SE	0.051	0.002	0.052	0.004	0.056	0.008

Table 12: Experiment #2b: Large number of observed instruments that are possibly weak and invalid. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$. In all cases, the instruments are (weakly) exogenous with $b \neq 0$ and $\upsilon = 0.5$: its first 10 elements are set to 1 and the remaining 40 to 0. The identification strength varies with $\alpha = 0$, 0.125, or 0.25.

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: Baseline model ($b = 0$ and $\alpha = 0$)						
Bias	0.000	-1.318	0.000	0.016	0.000	0.020
SE	0.017	39.039	0.017	0.271	0.018	0.145
Asympt.Heterosk.SE	0.017	9125.085	0.017	0.396	0.018	0.163
t-statistic	0.012	0.061	0.018	0.125	0.018	0.175
Median bias	0.000	0.008	0.000	0.014	0.000	0.017
Median Absolute Deviation	0.012	0.113	0.012	0.082	0.012	0.069
Median of SE	0.017	0.199	0.017	0.133	0.018	0.109
Rej. rate for Heterosk. SE	0.054	0.001	0.056	0.006	0.056	0.010
PANEL B: Baseline model ($b = 0$ and $\alpha = 0.125$)						
Bias	0.000	-0.120	0.000	0.012	0.000	0.020
SE	0.019	11.947	0.019	0.273	0.020	0.151
Asympt.Heterosk.SE	0.018	480.472	0.019	0.431	0.019	0.170
t-statistic	0.012	0.061	0.022	0.126	0.023	0.175
Median bias	0.000	0.008	0.000	0.014	0.000	0.017
Median Absolute Deviation	0.013	0.113	0.013	0.082	0.013	0.069
Median of SE	0.018	0.198	0.019	0.135	0.019	0.109
Rej. rate for Heterosk. SE	0.054	0.001	0.055	0.005	0.056	0.007
PANEL C: Baseline model ($b = 0$ and $\alpha = 0.25$)						
Bias	0.000	0.107	0.000	0.023	0.000	0.023
SE	0.022	4.062	0.023	0.295	0.024	0.231
Asympt.Heterosk.SE	0.022	83.147	0.022	0.511	0.024	0.247
t-statistic	0.008	0.062	0.013	0.126	0.019	0.179
Median bias	0.000	0.009	0.000	0.014	0.000	0.019
Median Absolute Deviation	0.015	0.115	0.015	0.083	0.016	0.070
Median of SE	0.021	0.202	0.022	0.137	0.023	0.110
Rej. rate for Heterosk. SE	0.053	0.002	0.056	0.003	0.056	0.007

Table 13: Experiment #2b: Large number of observed instruments that are possibly weak and invalid. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$. In all cases, the instruments are valid with $b = 0$. The identification strength varies with $\alpha = 0$, 0.125, or 0.25.

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: sample size $T = 200$						
Bias	-0.002	-0.092	-0.002	0.013	-0.003	0.027
SE	0.046	9.691	0.047	0.380	0.050	0.225
Asympt.Heterosk.SE	0.040	1445.113	0.041	0.611	0.043	0.238
t-statistic	-0.004	0.052	-0.008	0.132	-0.007	0.201
Median bias	-0.001	0.006	-0.001	0.016	-0.002	0.021
Median Absolute Deviation	0.026	0.122	0.027	0.098	0.028	0.086
Median of SE	0.036	0.240	0.036	0.164	0.038	0.135
Rej. rate for Heterosk. SE	0.051	0.003	0.051	0.009	0.052	0.019
Rej. rate for the null of weakness		0.430		0.383		0.357
Proportion of finite CI with AR		0.430		0.377		0.351
Length of finite CI with AR		1.907		3.580		1.917
Length of CI with FSMD		0.158		0.161		0.167
PANEL B: sample size $T = 2,000$						
Bias	0.000	0.203	0.000	0.021	0.000	0.025
SE	0.014	7.661	0.015	0.573	0.015	0.222
Asympt.Heterosk.SE	0.013	678.100	0.014	2.683	0.014	0.298
t-statistic	-0.007	0.054	-0.007	0.118	-0.005	0.161
Median bias	0.000	0.009	0.000	0.019	0.000	0.022
Median Absolute Deviation	0.009	0.133	0.009	0.105	0.010	0.092
Median of SE	0.013	0.286	0.013	0.191	0.013	0.158
Rej. rate for Heterosk. SE	0.053	0.000	0.055	0.004	0.052	0.006
Rej. rate for the null of weakness		0.402		0.350		0.324
Proportion of finite CI with AR		0.402		0.345		0.318
Length of finite CI with AR		1.895		1.871		1.806
Length of CI with F-SMD		0.053		0.053		0.055

Table 14: Experiment #2c: Large number of observed instruments with artificially weak first-stage. The two estimators (F-SMD and GMM) are implemented using either 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$ (Panel A), or $T = 2,000$ (Panel B). Additionally, we report: (i) the rejection frequencies associated with the (diagnostic) test of the null of weak identification based on the F-test, (ii) the proportion of identification-robust confidence intervals obtained with the Anderson-Rubin method that are infinitely large, (iii) the average length of the (finite) confidence intervals obtained by Anderson-Rubin and F-SMD.

B.6 Experiment #3: Empirically-motivated design

Parameter	$\gamma_f = 0.55$			$\lambda = 0.05$		
Estimator	F-SMD		GMM	F-SMD		GMM
Number of estim. factors	1	2	2	1	2	2
PANEL A: 50 Instruments and $\rho = 0$						
Bias	-0.008	-0.008	-0.001	0.000	0.000	0.000
SE	0.050	0.044	0.053	0.005	0.005	0.005
Asympt.Heterosk.SE	0.049	0.043	0.055	0.005	0.005	0.005
t-statistic	11.174	12.937	10.436	10.640	9.542	9.818
Median bias	-0.008	-0.007	0.001	0.000	0.000	0.000
Median Absolute Deviation	0.034	0.030	0.036	0.003	0.004	0.003
Median of SE	0.049	0.042	0.053	0.005	0.005	0.005
Rej. rate for Heterosk. SE	0.048	0.056	0.018	0.048	0.059	0.043
PANEL B: 50 Instruments and $\rho = 0.5$						
Bias	-0.010	-0.009	-0.001	0.000	0.000	0.000
SE	0.057	0.054	0.076	0.006	0.007	0.007
Asympt.Heterosk.SE	0.056	0.051	0.075	0.006	0.006	0.007
t-statistic	9.884	10.816	7.630	8.845	7.905	7.310
Median bias	-0.010	-0.009	0.002	0.000	0.000	0.000
Median Absolute Deviation	0.038	0.036	0.052	0.004	0.005	0.005
Median of SE	0.055	0.051	0.073	0.006	0.006	0.007
Rej. rate for Heterosk. SE	0.052	0.061	0.023	0.055	0.066	0.054
PANEL C: 6 Instruments and $\rho = 0$						
Bias	-0.017	-0.012	-0.002	-0.001	0.000	0.000
SE	0.057	0.042	0.057	0.005	0.005	0.005
Asympt.Heterosk.SE	0.055	0.042	0.057	0.005	0.005	0.005
t-statistic	10.190	13.152	10.057	9.527	10.652	9.485
Median bias	-0.014	-0.012	0.001	0.000	0.000	0.000
Median Absolute Deviation	0.037	0.028	0.038	0.003	0.003	0.004
Median of SE	0.053	0.041	0.055	0.005	0.005	0.005
Rej. rate for Heterosk. SE	0.050	0.056	0.021	0.049	0.051	0.040
PANEL D: 6 Instruments and $\rho = 0.5$						
Bias	-0.022	-0.016	-0.002	-0.001	0.000	0.000
SE	0.065	0.052	0.080	0.006	0.006	0.008
Asympt.Heterosk.SE	0.062	0.051	0.077	0.006	0.006	0.007
t-statistic	8.943	10.849	7.432	7.998	8.850	7.131
Median bias	-0.018	-0.016	0.001	-0.001	0.000	0.000
Median Absolute Deviation	0.042	0.035	0.054	0.004	0.004	0.005
Median of SE	0.060	0.050	0.075	0.006	0.006	0.007
Rej. rate for Heterosk. SE	0.059	0.060	0.025	0.054	0.058	0.051

Table 15: Experiment #3: Empirically-motivated design. The two estimators (F-SMD and GMM) are implemented using either 1 or 2 estimated factors which are estimated by PCA. In the first 3 columns, we report estimation results of the first slope parameter γ_f (set at 0.55) and in the last 3 those of the second slope parameter λ (set at 0.05). We consider either 50 instruments or 6 instruments, with ρ either 0 or 0.5, and the sample size is $T = 238$.

C Empirical results: estimation of the NKPC

	Instrument sets					
	mc	(mc,og)	(mc,og,C-Inf)	(mc,og,C-Inf,spread)	(mc,og,C-Inf,spread,W-Inf)	(mc,og,C-Inf,spread, W-Inf,Macro-Factor)
F-SMD						
γ_f	0.511	0.561	0.517	0.610	0.649	0.602
se	0.038	0.120	0.114	0.119	0.219	0.186
CI_l	0.437	0.326	0.294	0.377	0.220	0.237
CI_u	0.585	0.796	0.740	0.843	1.078	0.967
λ	0.014	0.021	0.005	0.010	0.019	0.019
se	0.018	0.019	0.019	0.020	0.022	0.022
GMM						
γ_f		-0.326	-0.427	-0.393	0.001	-0.326
se		3.731	3.308	1.753	1.095	0.703
CI_l		-7.639	-6.911	-3.829	-2.145	-1.704
CI_u		6.987	6.057	3.043	2.147	1.052
λ		0.032	0.033	0.034	0.033	0.038
se		0.060	0.059	0.036	0.023	0.017
B-GMM						
γ_f		0.611	0.604	0.612	0.559	0.564
se		0.103	0.075	0.052	0.052	0.049
CI_l		0.409	0.457	0.510	0.457	0.468
CI_u		0.813	0.751	0.714	0.661	0.660
λ		0.018	0.017	0.009	0.008	0.003
se		0.012	0.012	0.012	0.011	0.010

Table 16: Estimation of the NKPC over the whole sample with 247 observations: F-SMD (Top panel), GMM (Middle panel), and B-GMM (Bottom panel) with 1 to 6 IV, taken as one lag of marginal cost (mc), output gap (og), commodity inflation (C-Inf), spread (spread), wage inflation (W-Inf), first PCA of large macro-finance dataset (Macro-Factor). For B-GMM, one break is imposed in the first-stage equation at 2020Q1.

Instrument sets	mc	(mc,og)	(mc,og,C-Inf)	(mc,og,C-Inf,spread)	(mc,og,C-Inf,spread,W-Inf)	W-Inf,Macro-Factor)	PCA1	(PCA1, PCA2)
F-SMD								
γ_f	0.506	0.553	0.517	0.674	0.779	0.740	0.570	0.599
se	0.050	0.141	0.119	0.138	0.264	0.213	0.133	0.150
CI_l	0.408	0.277	0.284	0.404	0.262	0.323	0.309	0.305
CI_u	0.604	0.829	0.750	0.944	1.296	1.157	0.831	0.893
λ	0.019	0.030	0.012	0.017	0.023	0.020	-0.048	0.010
se	0.017	0.017	0.018	0.019	0.023	0.023	0.137	0.025
GMM								
γ_f		0.588	0.509	0.495	0.550	0.650		0.841
se		0.161	0.154	0.147	0.151	0.161		0.373
CI_l		0.272	0.207	0.207	0.254	0.334		0.110
CI_u		0.904	0.811	0.783	0.846	0.966		1.572
λ		0.023	0.022	0.022	0.025	0.023		0.034
se		0.013	0.012	0.012	0.011	0.012		0.037

Table 17: Estimation of the NKPC over the first subsample with 238 observations: F-SMD (Top panel), and GMM (Bottom panel) with 1 to 6 IV, taken as one lag of marginal cost (mc), output gap (og), commodity inflation (C-Inf), spread (spread), wage inflation (W-Inf), first PCA of large macro-finance dataset (Macro-Factor). Results in the last two columns are obtained using 1 or 2 IV generated as the first one or two PCAs (PCA1, PCA2) from the largest set of 6 IV reported in column 6.

	mc	og	C-Inf	spread	W-Inf	Macro-Factor	PCA1	one-step PCA1
γ_f	0.511	0.491	0.266	0.495	0.510	0.494	0.570	0.492
se	0.038	0.140	0.169	0.117	0.160	0.054	0.101	0.053
CI_l	0.437	0.217	-0.065	0.266	0.196	0.388	0.372	0.388
CI_u	0.585	0.765	0.597	0.724	0.824	0.600	0.768	0.596
λ	0.014	-0.053	-0.056	0.083	-0.008	-0.013	-0.099	-0.012
se	0.018	0.054	0.052	0.112	0.085	0.028	0.129	0.028

Table 18: Estimation of the NKPC over the whole sample with 247 observations using F-SMD with one IV only, taken as one lag of either marginal cost (mc), output gap (og), commodity inflation (C-Inf), spread (spread), wage inflation (W-Inf), macro-Factor, the first PCA obtained from these 6 IV (PCA1), or the first PCA obtained from the set of these 5 macro instruments and all the macro-finance variables (one-step PCA1).

Supplementary Appendix to:

Factor IV Estimation in Conditional Moment Models with an application to Inflation Dynamics

by

Bertille Antoine and Xiaolin Sun

In this Supplementary appendix, we consider alternative choices for the implementation of F-SMD, as well as additional empirical results associated with the estimation of the hybrid NKPC model highlighted in Section 6 of the main paper.

A Implementation of F-SMD

We first present additional Monte-Carlo results associated with different choices for the implementation of F-SMD: specifically, we focus throughout on Experiment #1 with HET2 highlighted in Section 5 in the main text.

In Table A.1, we consider different versions of the F-SMD estimator defined in (15) with $c = 0, 1, 2$, or 3. Results do not change much with c with the larger sample size, but there is some variation when $T = 200$. Specifically, we notice an increase in the standard error when c increases; this is to be expected since a larger c means fewer observations. That being said, it is a small sample issue as there is little to no change in the results when $T = 2,000$.

Next, we consider an alternate definition of the F-SMD objective function (14) which eliminates pairs of observations (say s and t) that are not only equal to each other, but also *too close* to each other. This is motivated by the proof of the asymptotic theory of the associated estimator, which suggests that these pairs do not contribute asymptotically. Accordingly, we consider the alternate objective function $\tilde{M}_T(\beta, \hat{F}, \tilde{c})$ and associated F-SMD estimator $\hat{\beta}_T(\tilde{c})$:

$$\begin{aligned} \tilde{M}_T(\beta, \hat{F}, \tilde{c}) &= \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{|s-t|>\tilde{c}, s=1}^{T} u_s(\beta) u_t(\beta) K\left(\frac{F_t - F_s}{h}\right), \\ \text{and} \quad \hat{\beta}(\tilde{c}) &= \arg\min \tilde{M}_T((\beta, \hat{F}, \tilde{c}). \end{aligned} \tag{30}$$

In Table A.2, we report results obtained with the alternate F-SMD estimator defined in (30) with values of $\tilde{c}$ ranging from 0 to 3. The results do not seem to depend much on $\tilde{c}$ even for the smaller sample size $T = 200$.

B Additional Monte-Carlo results

B.1 Experiment A.1: First-stage heterogeneity

We start by presenting results associated with a new experiment labelled Experiment A.1 where we consider a homoskedastic i.i.d setting with a heterogenous first-stage equation[18]. More specifically, there are two groups of individuals who respond differently to the instrument Z_1 in the sense that their underlying first-stage equation is different. The instrument Z_1 is always observed and available to the econometrician, whereas the group membership (or instrument Z_2) may or may not be known. In practice, estimation procedures will either rely on using only one instrument Z_1, or both instruments (Z_1, Z_2). Our DGP is as follows[19]:

$$\begin{aligned} y_i &= Y_i\beta_0 + u_i \\ Y_i &= 10 \times (2Z_{2,i} - 1)(Z_{1,i} - 2Z_{1,i}^3/5) + v_i \end{aligned}$$

where Z_1 is uniformly distributed over $[-2, 2]$, and Z_2 follows a Bernoulli distribution with $Pr(Z_2 = 1) = p_{z_2}$ set to either 0.2 or 0.05. The error terms (u_i, v_i) are independently generated according to a bivariate normal distribution with mean 0, variance 1 and correlation 0.6.

We compare the performance of our F-SMD estimator[20] to that of the 2SLS estimator - both implemented using either one or two instruments, respectively Z_1 or (Z_1, Z_2). The results are reported in Table A.3: in Panel A when $p_{z_2} = 0.2$ and in Panel B when $p_{z_2} = 0.05$.

Overall, the performance of the F-SMD estimators - when considering either only one instrument or both instruments - clearly dominates that of the corresponding 2SLS

[18] As suggested by a referee, this experiment does not fit well with the rest of our paper where we focus on a time series setup. We believe the results are interesting and we have chosen to present them in this Supplementary Appendix.

[19] Our DGP builds on the DGP used by Antoine and Lavergne (2023) though these authors always assume that the group membership is known. However, they do consider cases where the instrument Z_1 may be weak, whereas we always maintain that Z_1 is sufficiently strong.

[20] When the factors correspond to the observed (exogenous) instruments, the F-SMD estimator correspondonds to the SMD estimator introduced by Antoine and Lavergne (2014).

estimator. It is most noticeable when the group membership is unknown since 2SLS displays much larger biases and variances than F-SMD - even when one of the two groups is much larger than the other (e.g. 95% vs 5% of the sample); recall that 2SLS is implemented under the maintained linearity assumption of the first-stage. The performance of F-SMD remains excellent throughout even when the group membership is unknown. Our experiment emphasizes the robustness and advantages of our estimator which is not only easy to compute, but also convenient to implement without having to fully specify, characterize, or estimate the underlying first-stage equation.

B.2 Experiment #2: Large number of observed instruments

In this section, we revisit the DGP of Experiment #2 in the main text which features a large number of observed instruments that may be invalid and/or weak.

B.2.1 Fully linear design.

First, we present complementary results that are obtained in a model similar to that considered in Experiment #2a in the main text with strong and valid instruments, except that it is fully linear. More specifically, the DGP corresponds to that of section 5.2 and equations (18)-(22) after updating equation (19) as follows, to render it linear:

$$Y_t = 10 \times F_{1,t} + v_t$$

and setting $b = 0$ to ensure strong and valid instruments. The results are presented in Table A.4.

B.2.2 Weak and invalid instruments.

Second, we provide complementary results that are obtained in the model considered in Experiment #2b in the main text, that is the observed instruments may be invalid and/or weak. Specifically, we consider here a more severe case of invalid instruments where the first 25 elements of the parameter b are set to 1, and the remaining ones to 0. In addition, we set α to 0, 0.125, or 0.25 and υ to 0, 0.25, or 0.5. The results are presented in Tables A.5 to A.7.

C Additional results for the Hybrid NKPC

Next, we present additional empirical results associated with the estimation of the hybrid NKPC model highlighted in Section 6 of the main paper. These new results are obtained when considering alternative sets of instruments and specifications.

In Table A.8, we re-estimate the main model using the exogenous variable (current marginal cost) and one to six additional instruments, chosen as one lag of either, marginal cost, output gap, commodity inflation, spread, wage-inflation, or the macro-factor, after using a preliminary one-to-one transformation to ensure that each conditioning variable is bounded[21]. As suggested in Bierens (1990), we rely on the following transformation: $x \to \tan^{-1}(x)$. The results presented below are very much in line with those presented in the main paper in Table 16.

In Table A.9, we re-estimate the main model over an alternate *first subsample* ending just before the official start of the pandemic - that is from 1960Q2 to 2019Q4 using the exogenous variable (current marginal cost) and one to six additional instruments, chosen as one lag of either, marginal cost, output gap, commodity inflation, spread, wage-inflation, or the macro-factor. The results presented below are very much in line with those presented in the main paper in Table 17.

Finally, in Table A.10, we re-estimate the main model using the exogenous variable (current marginal cost) and one additional instrument, chosen as one lag of either, marginal cost, output gap, commodity inflation, spread, wage-inflation, or the macro-factor, after using the above-mentioned one-to-one $\tan^{-1}$ transformation. The results presented below are very much in line with those presented in the main paper in Table 18.

[21]Note that since we rely on the complex exponential to rewrite the conditional moments as a continuum of unconditional ones, we do not need to maintain such an assumption.

D Tables of results

c	0	1	2	3
Panel 1: sample size $T = 200$				
Bias	-0.002	-0.004	-0.006	-0.002
SE	0.050	0.068	0.100	0.413
Asympt.Heterosk.SE	0.043	0.050	0.088	1.418
t-statistic	0.003	0.000	0.015	0.040
Median bias	-0.001	-0.002	-0.001	0.000
Median Absolute Deviation	0.026	0.029	0.033	0.039
Median of SE	0.038	0.043	0.049	0.057
Rej. rate for Heterosk. SE	0.030	0.024	0.019	0.017
Panel 2: sample size $T = 2,000$				
Bias	0.000	0.000	0.000	0.000
SE	0.012	0.013	0.013	0.014
Asympt.Heterosk.SE	0.012	0.012	0.013	0.013
t-statistic	-0.004	0.000	0.000	0.003
Median bias	0.000	0.000	0.000	0.000
Median Absolute Deviation	0.008	0.008	0.009	0.009
Median of SE	0.011	0.012	0.012	0.012
Rej. rate for Heterosk. SE	0.046	0.044	0.046	0.044

Table A.1: Experiment #1 under HET2 when the true break fraction is 0.2. We report results obtained with the F-SMD estimator that ignores the break and considers $c = 0$, 1, 2 or 3 with a sample size of $T = 200$ (panel 1) or $T = 2,000$ (panel 2).

$\tilde{c}$	0	1	2	3
Panel 1: sample size $T = 200$				
Bias	-0.002	-0.002	-0.002	-0.002
SE	0.050	0.050	0.052	0.052
Asympt.Heterosk.SE	0.043	0.044	0.044	0.044
t-statistic	0.003	0.004	0.006	0.004
Median bias	-0.001	-0.001	-0.001	-0.001
Median Absolute Deviation	0.026	0.027	0.027	0.027
Median of SE	0.038	0.039	0.039	0.039
Rej. rate for Heterosk. SE	0.030	0.027	0.029	0.028
Panel 2: sample size $T = 2,000$				
Bias	0.000	0.000	0.000	0.000
SE	0.012	0.012	0.012	0.012
Asympt.Heterosk.SE	0.012	0.012	0.012	0.012
t-statistic	-0.004	-0.004	-0.004	-0.004
Median bias	0.000	0.000	0.000	0.000
Median Absolute Deviation	0.008	0.008	0.008	0.008
Median of SE	0.011	0.011	0.011	0.011
Rej. rate for Heterosk. SE	0.046	0.046	0.047	0.046

Table A.2: Experiment #1 under HET2 when the true break fraction is 0.2. We report results obtained with the alternate F-SMD estimator $\hat{\beta}(\tilde{c})$ defined in (30) that ignores the break. We consider $\tilde{c} = 0$, 1, 2 or 3 and a sample size of $T = 200$ (panel 1), or $T = 2,000$ (panel 2).

PANEL A: $Pr(Z_2 = 1) = 0.2$

Estimator	F-SMD		2SLS	
Instrument set	Z_1	(Z_1, Z_2)	Z_1	(Z_1, Z_2)
Panel A.1: sample size $T = 200$				
Bias	-0.003	-0.002	0.003	0.011
SE	0.051	0.035	16.615	0.289
Asympt.Heterosk.SE	0.050	0.034	1555.469	0.519
t-statistic	-0.019	-0.019	0.042	0.100
Median bias	-0.002	-0.001	0.008	0.013
Median Absolute Deviation	0.032	0.023	0.138	0.095
Median of SE	0.047	0.034	0.263	0.161
Rej. rate for Heterosk. SE	0.034	0.049	0.001	0.002
Panel A.2: sample size $T = 2,000$				
Bias	0.000	0.000	-0.017	0.003
SE	0.013	0.010	1.029	0.118
Asympt.Heterosk.SE	0.013	0.010	6.765	0.128
t-statistic	-0.005	-0.001	0.045	0.080
Median bias	0.000	0.000	0.002	0.005
Median Absolute Deviation	0.009	0.007	0.056	0.052
Median of SE	0.013	0.010	0.087	0.080
Rej. rate for Heterosk. SE	0.051	0.051	0.006	0.007

PANEL B: $Pr(Z_2 = 1) = 0.05$

Estimator	F-SMD		2SLS	
Instrument set	Z_1	(Z_1, Z_2)	Z_1	(Z_1, Z_2)
Panel B.1: sample size $T = 200$				
Bias	-0.002	-0.001	-0.558	0.003
SE	0.029	0.028	46.528	0.272
Asympt.Heterosk.SE	0.028	0.027	12471.831	0.425
t-statistic	-0.030	-0.027	0.026	0.074
Median bias	-0.001	-0.001	0.001	0.007
Median Absolute Deviation	0.019	0.018	0.110	0.082
Median of SE	0.028	0.027	0.197	0.134
Rej. rate for Heterosk. SE	0.053	0.055	0.001	0.007
Panel B.2: sample size $T = 2,000$				
Bias	0.000	0.000	-0.003	0.000
SE	0.008	0.008	0.129	0.068
Asympt.Heterosk.SE	0.008	0.008	0.120	0.066
t-statistic	-0.004	-0.003	0.033	0.060
Median bias	0.000	0.000	0.001	0.003
Median Absolute Deviation	0.006	0.005	0.037	0.035
Median of SE	0.008	0.008	0.056	0.054
Rej. rate for Heterosk. SE	0.053	0.052	0.016	0.018

Table A.3: Experiment A.1: Small number of observed exogenous instruments with first-stage heterogeneity when the sample size is either $T = 200$ or $T = 2,000$. We consider a setup with 2 groups with group membership Z_2 which follows a Bernoulli distribution with $Pr(Z_2 = 1)$ either equal to 0.2 (Panel A), or 0.05 (Panel B).

	1 estim. factor			2 estim. factors			3 estim. factors		
Estimator	F-SMD	2SLS	GMM	F-SMD	2SLS	GMM	F-SMD	2SLS	GMM
PANEL A: sample size $T = 200$									
Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SE	0.006	0.006	0.006	0.006	0.006	0.006	0.007	0.006	0.006
Asympt.Heterosk.SE	0.006	0.005	0.005	0.006	0.005	0.005	0.006	0.005	0.005
t-statistic	-0.001	0.001	0.001	0.006	0.004	0.004	0.009	0.008	0.008
Median bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median Absolute Deviation	0.004	0.003	0.003	0.004	0.003	0.003	0.004	0.003	0.003
Median of SE	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
Rej. rate for Heterosk. SE	0.043	0.047	0.047	0.045	0.047	0.052	0.049	0.046	0.059
PANEL B: sample size $T = 2,000$									
Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SE	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Asympt.Heterosk.SE	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
t-statistic	-0.010	-0.006	-0.006	-0.018	-0.007	-0.007	-0.025	-0.006	-0.006
Median bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median Absolute Deviation	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Median of SE	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Rej. rate for Heterosk. SE	0.051	0.052	0.052	0.050	0.051	0.052	0.050	0.051	0.055

Table A.4: Experiment #2a: Large number of observed strong and valid instruments driven by two unobserved (true) factors. The three estimators (F-SMD, 2SLS, and GMM) are either implemented using 1 estimated factor (first 3 columns), 2 estimated factors (middle 3 columns), or 3 estimated factors (last 3 columns). All factors are estimated by PCA with a sample size $T = 200$ (Panel A), or $T = 2,000$ (Panel B). In this design, the model is fully linear.

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: $b \neq 0$, $\alpha = 0$, **and** $\upsilon = 0$						
Bias	0.030	-5.465	0.027	-0.211	0.027	-0.137
SE	0.089	231.976	0.090	2.766	0.094	1.350
Asympt.Heterosk.SE	0.087	40257.247	0.086	3.711	0.089	1.167
t-statistic	0.345	-0.309	0.326	-0.164	0.310	-0.142
Median bias	0.030	-0.389	0.026	-0.185	0.027	-0.130
Median Absolute Deviation	0.059	0.636	0.060	0.694	0.061	0.622
Median of SE	0.087	1.113	0.086	0.804	0.089	0.634
Rej. rate for Heterosk. SE	0.072	0.001	0.081	0.021	0.074	0.041
PANEL B: $b \neq 0$ **and** $\alpha = 0.125$, **and** $\upsilon = 0$						
Bias	0.040	-0.764	0.038	-0.311	0.037	-0.250
SE	0.096	60.869	0.098	3.460	0.103	1.694
Asympt.Heterosk.SE	0.094	2760.718	0.094	5.135	0.097	1.409
t-statistic	0.437	-0.473	0.416	-0.308	0.396	-0.282
Median bias	0.042	-0.664	0.038	-0.374	0.037	-0.272
Median Absolute Deviation	0.064	0.715	0.066	0.800	0.069	0.692
Median of SE	0.094	1.256	0.093	0.876	0.097	0.667
Rej. rate for Heterosk. SE	0.085	0.009	0.092	0.036	0.086	0.063
PANEL C: $b \neq 0$ **and** $\alpha = 0.25$, **and** $\upsilon = 0$						
Bias	0.044	-0.358	0.043	-0.628	0.043	-0.482
SE	0.114	33.745	0.120	3.312	0.128	1.962
Asympt.Heterosk.SE	0.110	695.812	0.113	5.014	0.118	1.837
t-statistic	0.411	-0.669	0.390	-0.573	0.367	-0.534
Median bias	0.046	-1.117	0.046	-0.702	0.043	-0.523
Median Absolute Deviation	0.076	0.863	0.082	0.811	0.087	0.722
Median of SE	0.109	1.643	0.112	0.929	0.117	0.708
Rej. rate for Heterosk. SE	0.084	0.030	0.087	0.058	0.088	0.092

Table A.5: Experiment #2b: Large number of observed possibly weak and invalid instruments driven by two unobserved (true) factors. $b \neq 0$ means that its first 25 elements are equal to 1 and the remaining 25 are set to 0. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: $b \neq 0$, $\alpha = 0$, **and** $\upsilon = 0.25$						
Bias	0.008	-2.421	0.007	-0.044	0.007	-0.022
SE	0.029	80.350	0.029	0.814	0.030	0.385
Asympt.Heterosk.SE	0.028	16033.421	0.028	1.088	0.029	0.350
t-statistic	0.285	-0.231	0.272	-0.094	0.259	-0.053
Median bias	0.008	-0.102	0.007	-0.038	0.007	-0.023
Median Absolute Deviation	0.019	0.198	0.019	0.198	0.020	0.176
Median of SE	0.028	0.354	0.028	0.250	0.029	0.199
Rej. rate for Heterosk. SE	0.065	0.002	0.072	0.015	0.071	0.030
PANEL B: $b \neq 0$ **and** $\alpha = 0.125$, **and** $\upsilon = 0.25$						
Bias	0.011	-0.291	0.010	-0.074	0.010	-0.052
SE	0.031	23.331	0.032	0.973	0.033	0.476
Asympt.Heterosk.SE	0.031	1021.869	0.031	1.454	0.032	0.416
t-statistic	0.361	-0.375	0.348	-0.219	0.333	-0.175
Median bias	0.011	-0.171	0.010	-0.086	0.010	-0.056
Median Absolute Deviation	0.021	0.214	0.021	0.227	0.022	0.192
Median of SE	0.031	0.382	0.031	0.267	0.032	0.209
Rej. rate for Heterosk. SE	0.075	0.002	0.079	0.022	0.080	0.041
PANEL C: $b \neq 0$ **and** $\alpha = 0.25$, **and** $\upsilon = 0.25$						
Bias	0.011	-0.017	0.011	-0.150	0.011	-0.112
SE	0.036	9.983	0.038	0.935	0.041	0.523
Asympt.Heterosk.SE	0.036	182.266	0.037	1.398	0.039	0.472
t-statistic	0.340	-0.559	0.324	-0.453	0.308	-0.396
Median bias	0.012	-0.286	0.012	-0.168	0.012	-0.126
Median Absolute Deviation	0.024	0.254	0.026	0.231	0.027	0.204
Median of SE	0.036	0.474	0.037	0.282	0.038	0.218
Rej. rate for Heterosk. SE	0.068	0.013	0.071	0.031	0.076	0.055

Table A.6: Experiment #2b: Large number of observed possibly weak and invalid instruments driven by two unobserved (true) factors. $b \neq 0$ means that its first 25 elements are equal to 1 and the remaining 25 are set to 0. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$

	1 estim. factor		2 estim. factors		3 estim. factors	
Estimator	F-SMD	GMM	F-SMD	GMM	F-SMD	GMM
PANEL A: $b \neq 0$, $\alpha = 0$, **and** $\upsilon = 0.5$						
Bias	0.002	-1.611	0.002	0.000	0.002	0.009
SE	0.018	46.916	0.018	0.359	0.019	0.173
Asympt.Heterosk.SE	0.018	10674.087	0.018	0.501	0.019	0.183
t-statistic	0.126	-0.057	0.124	0.040	0.119	0.093
Median bias	0.002	-0.020	0.002	0.001	0.002	0.007
Median Absolute Deviation	0.012	0.121	0.012	0.095	0.013	0.081
Median of SE	0.018	0.213	0.018	0.144	0.019	0.118
Rej. rate for Heterosk. SE	0.056	0.002	0.060	0.008	0.057	0.014
PANEL B: $b \neq 0$ **and** $\alpha = 0.125$, **and** $\upsilon = 0.5$						
Bias	0.003	-0.166	0.003	-0.011	0.003	0.000
SE	0.020	14.524	0.020	0.380	0.021	0.194
Asympt.Heterosk.SE	0.020	571.643	0.020	0.571	0.021	0.200
t-statistic	0.158	-0.124	0.159	-0.020	0.152	0.033
Median bias	0.003	-0.035	0.003	-0.011	0.003	-0.002
Median Absolute Deviation	0.013	0.126	0.013	0.100	0.014	0.084
Median of SE	0.019	0.219	0.020	0.150	0.020	0.120
Rej. rate for Heterosk. SE	0.056	0.001	0.057	0.007	0.057	0.011
PANEL C: $b \neq 0$ **and** $\alpha = 0.25$, **and** $\upsilon = 0.5$						
Bias	0.003	0.074	0.003	-0.023	0.003	-0.013
SE	0.023	4.837	0.024	0.384	0.025	0.237
Asympt.Heterosk.SE	0.023	77.398	0.024	0.582	0.025	0.264
t-statistic	0.147	-0.224	0.143	-0.136	0.140	-0.073
Median bias	0.003	-0.064	0.003	-0.030	0.004	-0.017
Median Absolute Deviation	0.015	0.132	0.016	0.102	0.017	0.087
Median of SE	0.023	0.232	0.024	0.152	0.025	0.122
Rej. rate for Heterosk. SE	0.053	0.002	0.056	0.004	0.058	0.009

Table A.7: Experiment #2b: Large number of observed possibly weak and invalid instruments driven by two unobserved (true) factors. $b \neq 0$ means that its first 25 elements are equal to 1 and the remaining 25 are set to 0. The two estimators (F-SMD and GMM) are either implemented using 1 estimated factor (first 2 columns), 2 estimated factors (middle 2 columns), or 3 estimated factors (last 2 columns). All factors are estimated by PCA with a sample size $T = 200$

	mc	(mc,og)	(mc,og,C-Inf)	(mc,og,C-Inf,spread)	(mc,og,C-Inf,spread,W-Inf)	(mc,og,C-Inf,spread, W-Inf,Macro-Factor)
γ_f	0.503	0.518	0.523	0.501	0.456	0.394
se	0.036	0.059	0.081	0.073	0.179	0.167
CI_l	0.432	0.402	0.364	0.358	0.105	0.067
CI_u	0.574	0.634	0.682	0.644	0.807	0.721
λ	0.017	0.019	0.008	0.011	0.019	0.021
se	0.018	0.018	0.017	0.016	0.017	0.018

Table A.8: Estimation of the NKPC over the whole sample with 247 observations: F-SMD with 1 to 6 IV, taken as either one lag of, marginal cost (mc), output gap (og), commodity inflation (C-Inf), spread (spread), wage inflation (W-Inf), first PCA of large macro-finance dataset (Macro-Factor). All instruments are first transformed using $\tan^{-1}(.)$.

	mc	(mc,og)	(mc,og,C-Inf)	(mc,og,C-Inf,spread)	(mc,og,C-Inf,spread,W-Inf)	(mc,og,C-Inf,spread, W-Inf,Macro-Factor)	PCA1
γ_f	0.507	0.554	0.518	0.673	0.777	0.736	0.573
se	0.050	0.142	0.120	0.136	0.261	0.210	0.143
CI_l	0.409	0.276	0.283	0.406	0.265	0.324	0.293
CI_u	0.605	0.832	0.753	0.940	1.289	1.148	0.853
λ	0.020	0.031	0.013	0.017	0.022	0.019	-0.056
se	0.017	0.017	0.018	0.019	0.023	0.023	0.159

Table A.9: Estimation of the NKPC over the first subsample with 237 observations: F-SMD with 1 to 6 IV, taken as one lag of marginal cost (mc), output gap (og), commodity inflation (C-Inf), spread (spread), wage inflation (W-Inf), first PCA of large macro-finance dataset (Macro-Factor). Results in the last column (PCA1) are obtained using 1 IV generated as the first PCA from the largest set of 6 IV.

	mc	og	C-Inf	spread	W-Inf	Macro-Factor	PCA1
γ_f	0.503	0.470	0.362	0.510	0.470	0.494	0.485
se	0.036	0.093	0.101	0.116	0.109	0.054	0.076
CI_l	0.432	0.288	0.164	0.283	0.256	0.388	0.336
CI_u	0.574	0.652	0.560	0.737	0.684	0.600	0.634
λ	0.017	-0.034	-0.049	0.075	0.011	-0.013	0.032
se	0.018	0.043	0.054	0.094	0.100	0.028	0.110

Table A.10: Estimation of the NKPC over the whole sample with 247 observations: F-SMD with one IV only taken as either one lag of, marginal cost (mc), output gap (og), commodity inflation (C-Inf), spread (spread), wage inflation (W-Inf), first PCA of large macro-finance dataset (Macro-Factor). All instrumental variables are first transformed using $\tan^{-1}(.)$.