The Econometrics Journal

Econometrics Journal (2009), volume **12**, pp. S135–S171.
doi: 10.1111/j.1368-423X.2009.00279.x

Efficient GMM with nearly-weak instruments

BERTILLE ANTOINE† AND ERIC RENAULT‡

†*Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada*
E-mail: bertille_antoine@sfu.ca

‡*Department of Economics, University of North Carolina at Chapel Hill, CB 3305, Chapel Hill, NC 27599, USA*
E-mail: renault@email.unc.edu

First version received: July 2008; final version accepted: December 2008

Summary This paper is in the line of the recent literature on weak instruments, which, following the seminal approach of Stock and Wright captures weak identification by drifting population moment conditions. In contrast with most of the existing literature, we do not specify *a priori* which parameters are strongly or weakly identified. We rather consider that weakness should be related specifically to instruments, or more generally to the moment conditions. In addition, we focus here on the case dubbed *nearly-weak identification* where the drifting DGP introduces a limit rank deficiency reached at a rate slower than root-T. This framework ensures the consistency of Generalized Method of Moments (GMM) estimators of all parameters, but at a rate possibly slower than usual. It also validates the GMM-LM test with standard formulas. We then propose a comparative study of the power of the LM test and its modified version, or K-test proposed by Kleibergen. Finally, after a well-suited rotation in the parameter space, we identify and estimate directions where root-T convergence is maintained. These results are all directly relevant for practical applications without requiring the knowledge or the estimation of the slower rate of convergence.

Keywords: *GMM*, *Instrumental variables*, *Weak identification*.

1. INTRODUCTION

In this paper, we revisit the Generalized Method of Moments (GMM) of Hansen (1982) when classical identification assumptions are only barely satisfied. Following Hansen (1982), we imagine an economic model with structural parameter of interest $\theta \in \Theta \subseteq \mathbb{R}^p$. The econometrician's information about the true unknown value θ^0 of θ comes through moment conditions. For some stationary ergodic process (Y_t), $\phi(Y_t, \theta)$ is a K-dimensional function, integrable for all $\theta \in \Theta$, and the underlying economic model states that these moment conditions are satisfied at the true unknown value of the parameter:

$$E[\phi(Y_t, \theta^0)] = 0. \tag{1.1}$$

Moment conditions (1.1) *strongly globally identify* θ^0 if they do not admit any other solution:

$$E\left[\phi(Y_t, \theta)\right] = 0, \qquad \theta \in \Theta \Leftrightarrow \theta = \theta^0. \tag{1.2}$$

Hansen (1982) maintains (1.2) to prove consistency of a GMM estimator, defined as:

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta}[\bar{\phi}'_T(\theta)\Omega_T\bar{\phi}_T(\theta)]. \tag{1.3}$$

$\bar{\phi}_T(\theta)$ is the sample mean of $\phi(Y_t, \theta)$; Ω_T is a sequence of random non-negative matrices with positive definite probability limit. To address the asymptotic distribution of a GMM estimator, Hansen (1982) extends the above definition and considers any sequence $\hat{\theta}_T$ such that:

$$\text{Plim}[T^{1/2}A_T\bar{\phi}_T(\hat{\theta}_T)] = 0. \tag{1.4}$$

(A_T) is a sequence of (p, K) random matrices converging in probability to a constant full-row rank matrix A_0. The strong global identification condition (1.2) may be replaced by a local one: moment conditions (1.1) *strongly locally identify* θ^0 if θ^0 belongs to the interior of Θ where $\phi(Y_t, \theta)$ is continuously differentiable, and $E[\partial\phi(Y_t, \theta^0)/\partial\theta']$ has full-column rank. Under this strong local identification condition, the GMM estimator $\hat{\theta}_T$ defined by (1.4) is consistent and asymptotically normal. Moreover, the GMM estimator defined by (1.3) is a special case of (1.4) (through the first-order conditions), when strong local identification holds.

Both global and local strong identification conditions have been questioned in the literature during the last ten years. Stock and Wright (2000) relax the strong global identification when considering a drifting Data Generating Process (DGP) with:

$$E[\phi(Y_t, \theta)] = \frac{m_{1T}(\theta)}{T^{1/2}} + m_2(\theta_1) \quad \text{for some given subvector } \theta_1 \text{ of } \theta. \tag{1.5}$$

Then, only possibly θ_1 is identified, since, for the other components of θ, the relevant moment information vanishes at rate square-root T, the speed at which information is accumulated with a larger sample size. This case has been referred to as (global) weak identification.

Kleibergen (2005) focuses on the GMM score-type test of a null hypothesis: $\underline{H}_0: \theta = \theta_0$. For such a problem, only local identification is relevant, and Kleibergen (2005) refers to as (local) weak identification the case where:

$$E\left[\frac{\partial\phi(Y_t, \theta^0)}{\partial\theta'}\right] = \frac{C}{T^{1/2}} \quad \text{with full-column rank matrix } C. \tag{1.6}$$

We revisit the issue of GMM estimators and score-type tests in the nearly-weak identification case, first introduced by Hahn and Kuersteiner (2002): through a drifting DGP approach, information now vanishes when sample size T increases, but at a rate δ_T slower than $T^{1/2}$. In the context of Stock and Wright (2000), *global nearly-weak identification* would mean:

$$E\,[\phi(Y_t, \theta)] = \frac{m_{1T}(\theta)}{\delta_T} + m_2(\theta_1), \tag{1.7}$$

while in the context of Kleibergen (2005), *local nearly-weak identification* would mean:

$$E\left[\frac{\partial\phi(Y_t, \theta^0)}{\partial\theta'}\right] = \frac{C}{\delta_T}. \tag{1.8}$$

The possible case of nearly-weak identification has been quite overlooked in the literature while, after all, it makes sense to study a variety of asymptotic behaviours when δ_T may be associated to any rate between $\mathcal{O}(1)$ and $\mathcal{O}(T^{1/2})$. Weak identification is only a limit case where identification is completely lost. So far, only Hahn and Kuersteiner (2002) in a linear context, and Caner (2007) in a non-linear one, have considered nearly-weak identification. Our contribution as concerns nearly-weak identification is to imagine that, in realistic circumstances, nearly-weak identification may occur for some moments while strong identification is still guaranteed for others. This new point of view paves the way for new results as follows.

In terms of GMM score-type testing, the partition between locally strongly- and locally nearly-weakly identifying moment conditions determines the different rates of convergence associated with specific directions in the parameter space against which the test has power.[1] As a result, the GMM score test has power even in quite weak directions, where the weakness degree δ_T may be arbitrarily close to $T^{1/2}$. We show that, by contrast, Kleibergen's modified score test is more likely to waste some power in such directions. It is the price to pay to be robust to weak identification ($\delta_T = T^{1/2}$) when, as shown by Kleibergen (2005), the standard GMM score test does not work. We show that the GMM score test and Kleibergen's modified score test are actually asymptotically equivalent under relevant sequences of local alternatives, but only in cases of moderate weakness of identification: we refer to *nearly-strong identification*, when δ_T goes to infinity slower than $T^{1/4}$. This equivalence is tightly related to a stronger equivalence result between the standard two-step (efficient) GMM and the continuously updated GMM of Hansen et al. (1996) for efficient estimation of all directions. Such a result can only be embraced after extending the pioneered setting introduced by Stock and Wright (2000): we now consider that some moment conditions are globally-identifying, while some others are weakly-identifying. In other words, the vector $\phi(Y_t, \theta)$ is partitioned into two subvectors $\phi(Y_t, \theta) = [\phi_1(Y_t, \theta)' \vdots \phi_2(Y_t, \theta)']'$ such that:

$$E\left[\phi_1(Y_t, \theta)\right] = \rho_1(\theta) \quad \text{and} \quad E\left[\phi_2(Y_t, \theta)\right] = \rho_2(\theta)/\delta_T \tag{1.9}$$

with the *global nearly-weak identification* condition:

$$\rho(\theta) = 0 \Leftrightarrow \theta = \theta^0 \quad \text{where} \quad \rho(\theta) = [\rho_1'(\theta) \vdots \rho_2'(\theta)]'.$$

Identification is nearly-weak because δ_T goes to infinity, but we rather call it *nearly-strong* when associated to a rate slower than $T^{1/4}$. By contrast with Stock and Wright (2000), we have no prior knowledge on the subset of parameters that are weakly identified. Intuitively, the first set of moment conditions (respectively the second one) identifies strong (respectively weak) directions in the parameter space. Through a convenient rotation in the parameter space in the spirit of Phillips (1989), we define a reparametrization such that the first components of this new parameter are estimated at standard rate square-root T, while the others are estimated only at slower rate $\lambda_T = T^{1/2}/\delta_T$. Asymptotic covariances come with standard GMM-like formulas, but only with *nearly-strong identification*, i.e. rate λ_T faster than $T^{1/4}$. Interpreting this latter condition is germane to Andrews' (1994) study of MINPIN estimators.[2] In our case, the nuisance parameter is not infinite dimensional. However, due to nearly-weak identification, it is associated to a rate of convergence slower than the standard parametric square-root T. As in Andrews (1994), the slow rate of convergence needs to be faster than $T^{1/4}$ to avoid contamination of the well-identified estimated directions by the nearly-weak ones. We also show that the nearly-strong identification condition is exactly needed to ensure that all directions are equivalently estimated

[1] As far as the size properties are concerned, we know from results in Andrews and Guggenberger (2007) that nearly-weak identification does not offer additional insights. Only genuine weak identification is important in investigating size properties of testing procedures.

[2] These estimators are defined as MINimizing a criterion function that might depend on a Preliminary Infinite dimensional Nuisance parameter estimator.

by efficient two-step GMM and continuously updated GMM. This explains the aforementioned partial equivalence between GMM score and Kleibergen's modified score tests. More generally, our unified setting for mixed strong/nearly-strong identification incorporates coherently both global and local points of view.

Ultimately, evidence of weak identification should not always lead to renounce meaningful estimation and testing, as the alleged weak identification may only be nearly-weak, or even nearly-strong. Overlooking these cases could lead to wasting some relevant information. Moreover, possible weakness should be assigned to some specific instruments (or to some moment conditions) as in (1.9) rather than to specific parameters as in (1.7). It is the econometrician's duty to determine the different directions in the parameter space where she has more or less accurate information. We illustrate this new point of view with a Monte Carlo study on the well-known example of consumption-based CAPM, already extensively studied in the literature: see Stock and Wright (2000) among others.

The paper is organized as follows. In Section 2, we discuss GMM-based tests of a simple null hypothesis $\underline{H}_0 : \theta = \theta_0$. We compare the asymptotic behaviour of the standard GMM score test (Newey and West, 1987), and the Kleibergen modified score test. When complete weak identification is precluded, both tests work. Our framework allows us to display relevant sequences of local alternatives with heterogeneous rates of convergence depending on the direction of departure in the parameter space. By contrast with Kleibergen (2005), different degrees of nearly-weak identification are simultaneously considered: this opens the door for non-equivalence, even asymptotically, between standard and modified score tests. In Section 3, consistency and rate of convergence of any GMM estimator are analysed in a nearly-weak identification setting. The special case of nearly-strong identification allows us in Section 4 to discuss efficient estimation with various rates of convergence in various directions, and to check equivalence between two-step efficient GMM and continuously updated GMM in all directions. These last results bridge the gap between estimation and score tests as discussed in Section 2. The practical relevance of our new asymptotic theory is checked in Section 5 in a consumption-based intertemporal asset pricing model. It validates our point of view of nearly-strong identification with different rates of convergence in different directions for realistic simulated parameter configurations. Section 6 concludes. Proofs are gathered in the Appendix; Table 1 summarizes the different concepts of identification.

2. GMM SCORE-TYPE TESTING

We want to test the null hypothesis: $\underline{H}_0 : \theta = \theta_0$. Our information about parameter θ comes from the following moment conditions,

$$E[\phi(Y_t, \theta^0)] = 0, \tag{2.1}$$

always assumed to be fulfilled at least by the true unknown value of the parameter θ^0. Observed time series $(Y_t)_{1 \le t \le T}$ of a stationary ergodic process are available, and such that sample counterparts of the moment conditions satisfy a Central Limit Theorem (CLT) at the true value:

ASSUMPTION 2.1. *(CLT at the true value θ^0). With $\bar{\phi}_T(\theta^0) = \frac{1}{T}\sum_{t=1}^{T} \phi(Y_t, \theta^0)$:*

(i) $\sqrt{T}\bar{\phi}_T(\theta^0)$ is asymptotically normally distributed with zero mean and covariance matrix S^0.

Table 1. Global, partial and local identification.

		Global Identification	
Strong		$E[\phi(Y_t, \theta)] = 0 \Leftrightarrow \theta = \theta^0$	
		on parameters (SW)	on moment conditions (AR)
Nearly-strong	$1 \ll \delta_T \ll T^{1/4}$	$E[\phi(Y_t, \theta)] = m_1(\theta_1) + m_{2T}(\theta)/\delta_T$	$\begin{cases} E[\phi_1(Y_t, \theta)] = \rho_1(\theta) \\ E[\phi_2(Y_t, \theta)] = \rho_2(\theta)/\delta_T \end{cases}$
Nearly-weak	$1 \ll \delta_T \ll T^{1/2}$	with θ_1 known subset of θ	with ϕ_1 known subset of ϕ
Weak	$\delta_T = T^{1/2}$	and $\begin{cases} m_1(\theta_1) = 0 \Leftrightarrow \theta_1 = \theta_1^0 \\ m_2(\theta_1^0, \theta_2) = 0 \Leftrightarrow \theta_2 = \theta_2^0 \end{cases}$	and $\begin{pmatrix} \rho_1(\theta) \\ \rho_2(\theta) \end{pmatrix} = 0 \Leftrightarrow \theta = \theta^0$
		Partial Identification	
		General case with $\phi(Y_t, \theta) = A(Y_t)\theta$ and rank conditions on $E[A(Y_t)]$	
		Rank deficiencies nest the cases $\delta_T = \infty$ (both for (SW) and (AR))	
		Local Identification	
Strong		$E\left[\frac{\partial \phi(Y_t, \theta^0)}{\partial \theta'}\right]$ full column rank	
		on parameters (SW)	on moment conditions (AR)
Nearly-strong	$1 \ll \delta_T \ll T^{1/4}$	$E\left[\frac{\partial \phi(Y_t, \theta^0)}{\partial \theta'}\right] = \left[\frac{\partial m_1(\theta_1^0)}{\partial \theta_1'} \vdots 0\right] + \frac{1}{\delta_T}\left[\frac{\partial m_2(\theta^0)}{\partial \theta_1'} \vdots \frac{\partial m_2(\theta^0)}{\partial \theta_2'}\right]$	$\text{Plim}\begin{pmatrix} \frac{\partial \bar{\phi}_{1T}(\theta^0)}{\partial \theta'} \\ \delta_T \frac{\partial \bar{\phi}_{2T}(\theta^0)}{\partial \theta'} \end{pmatrix} = C$
Nearly-weak	$1 \ll \delta_T \ll T^{1/2}$	$\frac{\partial m_1(\theta_1^0)}{\partial \theta_1'}$ full column rank	with C full column rank
Weak	$\delta_T = T^{1/2}$		

(ii) *HAC estimator $S_T(\theta)$ of the long-term covariance matrix S^0 is available and such that:* $S^0 = \text{Plim}[S_T(\theta^0)]$.

We focus here on a case where local nearly-weak identification of some directions in the parameter space may occur simultaneously as strong identification of other directions. More precisely, we assume:

ASSUMPTION 2.2. (*Nearly-weak local identification*).

(i) *θ^0 belongs to the interior of Θ, and $\phi(Y_t, \theta)$ is continuously differentiable on Θ.*
(ii) *There exists a (K, p) matrix C, with full-column rank such that:*

$$\text{Plim}\left[\frac{\partial \bar{\phi}_{1T}(\theta^0)}{\partial \theta'}\right] = C_1 \quad \textit{and} \quad \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_{2T}(\theta^0)}{\partial \theta'}\right] = C_2,$$

where $\bar{\phi}_T(\theta) = [\bar{\phi}'_{1T}(\theta) \vdots \bar{\phi}'_{2T}(\theta)]'$, $C = [C'_1 \vdots C'_2]'$, $\lambda_T \overset{T}{\to} \infty$, *and* $\frac{\lambda_T}{\sqrt{T}} \overset{T}{\to} 0$.

The following toy example illustrates our focus of interest.

EXAMPLE 2.1. (Toy example). Consider the moment conditions: $E[Y_{1t}] = g(\theta^0)$ and $E[Z_t \otimes (Y_{2t} - X_{2t}\theta^0)] = 0$, where the general functions $\phi_1(.)$ and $\phi_2(.)$ are defined as follows:

$$\phi_1(Y_t, \theta) = Y_{1t} - g(\theta) \quad \text{and} \quad \phi_2(Y_t, \theta) = -Z_t \otimes (Y_{2t} - X_{2t}\theta).$$

The instruments Z_t introduced in ϕ_2 are only nearly-weak instruments since

$$E\left[Z_t \otimes X_{2t}\right] = \frac{C_2}{\delta_T} \quad \text{with} \quad \delta_T = \frac{\sqrt{T}}{\lambda_T} \overset{T}{\to} \infty \quad \text{and} \quad \frac{\delta_T}{\sqrt{T}} = \frac{1}{\lambda_T} \overset{T}{\to} 0.$$

Then the associated Jacobian matrix is:

$$\text{Plim}\left[\frac{\partial \bar{\phi}_{1T}(\theta^0)}{\partial \theta'}\right] = \frac{\partial g(\theta^0)}{\partial \theta'} = C_1$$

$$\text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_{2T}(\theta^0)}{\partial \theta'}\right] = \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{1}{T}\sum_{t=1}^{T}(Z_t \otimes X_{2t})\right] = \lim_T \left[\frac{\sqrt{T}}{\lambda_T} E\left(Z_t \otimes X_{2t}\right)\right] = C_2$$

and we assume that the following matrix has full-column rank:

$$\left[\frac{\partial g'(\theta^0)}{\partial \theta} \vdots E\left(Z'_t \otimes X'_{2t}\right)\right]'.$$

GMM score-type testing asks whether the test value θ_0 is closed to fulfil the first-order conditions of the (efficient) two-step GMM minimization, $\min_{\theta \in \Theta}[\bar{\phi}'_T(\theta) S_T^{-1}(\theta_0)\bar{\phi}_T(\theta)]$, that is, whether the score vector is close to zero. It is defined at the test value θ_0 as:

$$V_T(\theta_0) = \frac{\partial \bar{\phi}'_T(\theta_0)}{\partial \theta} S_T^{-1}(\theta_0)\bar{\phi}_T(\theta_0).$$

The GMM score test statistic (Newey and West, 1987) is then a suitable norm of $V_T(\theta_0)$:

$$\xi_T^{NW} = T V'_T(\theta_0)\left[\frac{\partial \bar{\phi}'_T(\theta_0)}{\partial \theta} S_T^{-1}(\theta_0)\frac{\partial \bar{\phi}_T(\theta_0)}{\partial \theta'}\right]^{-1} V_T(\theta_0).$$

Kleibergen (2005) considers instead the first-order conditions of the continuously updated GMM minimization: $\min_{\theta\in\Theta}[\bar{\phi}'_T(\theta)S_T^{-1}(\theta)\bar{\phi}_T(\theta)]$. The corresponding score vector can be computed either by direct differentiation (see equations (15) and (16) and appendix in Kleibergen, 2005), or with even simpler computations within a Euclidean Empirical Likelihood approach (see Antoine et al., 2007). The score vector computed at the test value θ_0 is:

$$\tilde{V}_T(\theta_0) = \frac{\partial \tilde{\phi}'_T(\theta_0)}{\partial\theta} S_T^{-1}(\theta_0)\bar{\phi}_T(\theta_0),$$

where each column $\left[\frac{\partial \tilde{\phi}_T^{(j)}(\theta_0)}{\partial\theta}\right]$ of $\left[\frac{\partial \tilde{\phi}'_T(\theta_0)}{\partial\theta}\right]$ is the residual of the long-term affine regression of $\left[\frac{\partial \bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}\right]$ on $\bar{\phi}_T(\theta_0)$:[3]

$$\frac{\partial \tilde{\phi}_T^{(j)}(\theta_0)}{\partial\theta} = \frac{\partial \bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta} - \mathrm{Cov}_{as}\left(\sqrt{T}\frac{\partial \bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta_0)\right)\mathrm{Var}_{as}\left(\sqrt{T}\bar{\phi}_T(\theta_0)\right)^{-1}\bar{\phi}_T(\theta_0),$$

where $\mathrm{Var}_{as}(\sqrt{T}\bar{\phi}_T(\theta_0)) = S^0$ is the long-term covariance matrix of the moment conditions $\phi(Y_t, \theta_0)$, and $\mathrm{Cov}_{as}(\sqrt{T}\frac{\partial \bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta_0))$ is the long-term covariance between $\left[\frac{\partial \phi^{(j)}(Y_t,\theta_0)}{\partial\theta}\right]$ and $\phi(Y_t, \theta_0)$. This long-term covariance is assumed to be well defined.[4]

ASSUMPTION 2.3. (*Long-term covariance*).

$$\mathrm{Cov}_{as}\left(\sqrt{T}\frac{\partial \bar{\phi}_T^{(j)}(\theta^0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta^0)\right) \equiv \lim_T\left\{T\mathrm{Cov}\left(\frac{\partial \bar{\phi}_T^{(j)}(\theta^0)}{\partial\theta}, \bar{\phi}_T(\theta^0)\right)\right\}$$

is a well-defined (p, K) *matrix.*

Kleibergen (2005) maintains Assumption 2.3, and in addition assumes that it corresponds to the asymptotic covariance matrix of the (assumed) joint asymptotic normal distribution of $\left[\sqrt{T}\frac{\partial \bar{\phi}_T^{(j)}(\theta^0)}{\partial\theta}\right]$ and $\left[\sqrt{T}\bar{\phi}_T(\theta^0)\right]$. In our nearly-weak identification setting, what we really need, albeit almost equivalent for all practical purposes, is only the following regularity condition:

ASSUMPTION 2.4. (*Well-behaved Jacobian matrix for the strong subset of moment conditions*)

$$\sqrt{T}\left[\frac{\partial \bar{\phi}_{1T}(\theta^0)}{\partial\theta'} - C_1\right] = \mathcal{O}_P(1).$$

From the above discussion, replacing $[\partial \bar{\phi}_T^{(j)}(\theta^0)/\partial\theta]$ by $[\partial \tilde{\phi}_T^{(j)}(\theta^0)/\partial\theta]$ amounts to removing the finite sample correlation between $[\partial \bar{\phi}_T^{(j)}(\theta^0)/\partial\theta]$ and $\bar{\phi}_T(\theta^0)$. As extensively discussed in Antoine et al. (2007), while this correlation may be responsible for the finite sample bias of standard two-stage GMM, the well-documented improved bias performance of continuously updated GMM is precisely due to this correction. When considering genuine weak instruments ($\lambda_T = 1$ in Assumption 2.2), this correlation remains asymptotic, and

[3] For any vector $\psi \in \mathbb{R}^K$, we distinguish between the following notations. (i) $\psi = [\psi'_1\ \psi'_2]'$ refers to the partition introduced in Assumption 2.2: ψ_1 (respectively ψ_2) is a subvector of ψ with the same dimension as the strong (respectively nearly-weak) group of moment conditions. (ii) $\psi = [\psi^{(j)}]_{1\leq j\leq K}$ refers to the (single) components of the vector ψ: each $\psi^{(j)}$ is a real number. The latter (cumbersome) notation is not used often.

[4] These notations are precisely defined in Assumption 2.3.

Kleibergen (2005) introduces a modified version of the Newey and West (1987) score test statistic:

$$\xi_T^K = T\tilde{V}_T'(\theta_0)\left[\frac{\partial\tilde{\phi}_T'(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\frac{\partial\tilde{\phi}_T(\theta_0)}{\partial\theta'}\right]^{-1}\tilde{V}_T(\theta_0).$$

By contrast, we show that with nearly-weak instruments, the correlation is immaterial, and the standard GMM score test statistic ξ_T^{NW} works. It is actually asymptotically equivalent to the modified Kleibergen's score test statistic under the null:

PROPOSITION 2.1. (*Equivalence under the null*). *Under Assumptions 2.1–2.4, under the null* $\underline{H}_0: \theta = \theta_0$, *we have:* $\text{Plim}[\xi_T^{NW} - \xi_T^K] = 0$ *and both* ξ_T^{NW} *and* ξ_T^K *converge in distribution towards a chi-square with p degrees of freedom.*

The main contribution of this paper is to characterize the heterogeneity of the informational content of moment conditions, along different directions in the parameter space. While a proper assessment of this heterogeneity is crucial for efficient estimation (see Section 4), it also matters when considering power of tests under sequences of local alternatives.

EXAMPLE 2.2. (Toy example, continued). Consider a sequence of local alternatives defined by a given deterministic sequence $(\gamma_T)_{T\geq 0}$ in $\mathbb{R}^p$, going to zero when T goes to infinity, and such that the true unknown value θ_0 is drifting as: $\theta_T = \theta_0 + \gamma_T$. For large T: $g(\theta_T) \sim g(\theta_0) + [\partial g(\theta_0)/\partial\theta']\gamma_T$.

Therefore, the strongly-identified moment restrictions $E[Y_{1t} - g(\theta_T)] = 0$ are informative with respect to the violation of the null ($\theta_T \neq \theta_0$) if and only if: $[\partial g(\theta_0)/\partial\theta']\gamma_T \neq 0$.

As a consequence, we expect GMM-based tests of $\underline{H}_0: \theta = \theta_0$ to have power against sequences of local alternatives converging at standard rate square-root T, $\theta_T = \theta_0 + \gamma/\sqrt{T}$, if and only if $[\partial g(\theta_0)/\partial\theta']\gamma \neq 0$, or, when γ does not belong to the null space of $C_1 = [\partial g(\theta_0)/\partial\theta']$. By contrast, if $C_1\gamma = 0$, violations of the null can only be built from the nearly-weakly identifying conditions:

$$E\left[Z_t \otimes Y_t\right] = \frac{C_2}{\delta_T}\theta_T.$$

We will show that relevant sequences of local alternatives to characterize non-trivial power are necessarily such that: $\theta_T = \theta_0 + \delta_T\frac{\gamma}{\sqrt{T}} = \theta_0 + \frac{\gamma}{\lambda_T}$.

In other words, the degree of weakness of the moment conditions δ_T downplays the standard rate $[\gamma/\sqrt{T}]$ of sequences of local alternatives against which the tests have non-trivial local power. The intuition is quite clear. Under such a sequence of local alternatives,

$$E\left[Z_t \otimes Y_t\right] = \frac{C_2\theta_0}{\delta_T} + \frac{C_2\gamma}{\sqrt{T}}$$

differs from its value under the null by the standard scale $1/\sqrt{T}$.

We need to reinforce Assumptions 2.1(ii) and 2.2(ii) for the study of sequences of local alternatives:

ASSUMPTION 2.5. (*Reinforced assumptions for study of local power*). *For any sequence* θ_T *in the interior of* Θ *such that* $\text{Plim}[\theta_T - \theta^0] = 0$, *(i)* *HAC estimator $S_T(\theta)$ of the long-term covariance matrix S^0 is available and such that:* $S^0 = \text{Plim}[S_T(\theta_T)]$.

(ii)

$$\text{Plim}\left[\frac{\partial\bar{\phi}_{1T}(\theta_T)}{\partial\theta'}\right] = C_1 \quad \textit{and} \quad \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta_T)}{\partial\theta'}\right] = C_2.$$

(iii) *$\phi_1(Y_t, \theta)$ is twice continuously differentiable, and for all components $\phi_{1T}^{(j)}$ (see also footnote 3),* $\text{Plim}[\partial^2\phi_{1T}^{(j)}(\theta_T)/\partial\theta\partial\theta']$ *is a well-defined matrix.*

PROPOSITION 2.2. (*Local power of GMM score tests*). *Under Assumptions 2.1–2.5:*

(i) *With a drifting true unknown value, $\theta_T = \theta_0 + \gamma/\sqrt{T}$, for some $\gamma \in \mathbb{R}^p$, we have* $\text{Plim}[\xi_T^{NW} - \xi_T^K] = 0$, *and both ξ_T^{NW} and ξ_T^K converge in distribution towards a non-central chi-square with p degrees of freedom and non-centrality parameter* $\mu = (\gamma' C_1' \vdots 0)[S^0]^{-1}\binom{C_1\gamma}{0}$.

(ii) *When $\lambda_T^2/\sqrt{T} \overset{T}{\to} \infty$, with a drifting true unknown value $\theta_T = \theta_0 + \gamma/\lambda_T$, for some $\gamma \in \mathbb{R}^p$ such that $C_1\gamma = 0$, we have* $\text{Plim}[\xi_T^{NW} - \xi_T^K] = 0$, *and both ξ_T^{NW} and ξ_T^K converge in distribution towards a non-central chi-square with p degrees of freedom and non-centrality parameter* $\mu = (0 \vdots \gamma' C_2')[S^0]^{-1}\binom{0}{C_2\gamma}$.

Two additional conclusions implicitly follow from Proposition 2.2.

First, if $C_1\gamma \neq 0$, the two GMM score tests behave more or less as usual against sequences of local alternatives in the direction γ. They are asymptotically equivalent, and both consistent against sequences converging slower than square-root T. They both follow asymptotically a non-central chi-square against sequences with exactly the rate square-root T. However, the non-centrality parameter (and hence the power of the test) does not really depend on the size of the departure γ from the null, but only on the size of its orthogonal projection of the space spanned by the columns of C_1' (orthogonal space of the null space of C_1); that is, by the columns of the Jacobian matrix corresponding to the strong moment conditions.

Second, if $C_1\gamma = 0$, the two GMM score tests have no power against sequences of local alternatives $\theta_T = \theta_0 + \gamma/\sqrt{T}$. They may have power against sequences $\theta_T = \theta_0 + \gamma/\lambda_T$ (or slower); their behaviour is pretty much the standard one, but only in the nearly-strong case where λ_T goes to infinity faster than $T^{1/4}$. The case $C_1\gamma = 0$ corresponds to the study of Guggenberger and Smith (2005) (see theorem 3, p. 680). In the setting of Stock and Wright (2000) that they adopt, with $\theta = (\alpha' \ \beta')'$, α weakly identified and β strongly identified, $C_1\gamma = 0$ means that β is fixed at its value β_0 under the null. However, since this approach does not disentangle partition of moments and partition of parameters, it amounts for us to consider that $\phi = \phi_2$. This simplifies the study of local power (see our Proposition 2.3 below) and makes immaterial the condition of nearly-strong identification.

We now explain why non-standard asymptotic behaviour of both score tests may arise when we consider sequences of local alternatives in the weak directions ($\theta_T = \theta_0 + \gamma/\lambda_T$ with $C_1\gamma = 0$) while the nearly-weak identification problem is so severe that even $\lambda_T^2/\sqrt{T}$ goes to zero. Recall that the genuine weak identification usually considered in the literature ($\lambda_T = 1$) is a limit case, since we always maintain the nearly-weak identification condition $\lambda_T \overset{T}{\to} \infty$. Under such a sequence of local alternatives, while by Assumption 2.1, $\sqrt{T}\bar{\phi}_T(\theta_T)$ is asymptotically normal with zero mean, the key to get a common non-central chi-square for the asymptotic distribution of a score test statistic is to ensure that $\sqrt{T}\bar{\phi}_T(\theta_0)$ is asymptotically normal with non-zero mean if and only if γ is not zero. This result should follow from the Taylor approximation:

$$\begin{aligned}\sqrt{T}\bar{\phi}_T(\theta_T) &\approx \sqrt{T}\bar{\phi}_T(\theta_0) + \sqrt{T}\frac{\partial \bar{\phi}_T(\theta_T^*)}{\partial \theta'}(\theta_T - \theta_0) \\ &= \sqrt{T}\bar{\phi}_T(\theta_0) + \frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_T(\theta_T^*)}{\partial \theta'}\gamma \\ &\approx \sqrt{T}\bar{\phi}_T(\theta_0) + \begin{pmatrix} 0 \\ C_2\gamma \end{pmatrix}.\end{aligned}$$

Assumption 2.5 justifies this approximation insofar as we can show that $C_1\gamma = 0$ implies:

$$\text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_{1T}(\theta_T^*)}{\partial \theta'}\gamma\right] = 0. \tag{2.2}$$

This question is irrelevant if, as in Kleibergen (2005), two different degrees of identification are never simultaneously considered.[5] In other words, we can easily state:

PROPOSITION 2.3. (*Special case with only one degree of weakness*). *Consider the special case where the same degree of weakness is assumed for all moment conditions at hand ($\phi(.) = \phi_2(.)$). Under Assumptions 2.1–2.5, with a drifting true unknown value $\theta_T = \theta_0 + \gamma/\lambda_T$, for some $\gamma \in \mathbb{R}^p$, we have* $\text{Plim}[\xi_T^{NW} - \xi_T^K] = 0$, *and both ξ_T^{NW} and ξ_T^K converge in distribution towards a non-central chi-square with p degrees of freedom and non-centrality parameter $\mu = \gamma' C' [S^0]^{-1} C\gamma$.*

Needless to say that a result similar to Proposition 2.3 holds when $\phi(.) = \phi_1(.)$ (standard strong identification). Hence, the interesting case is precisely the mixture of strong and nearly-weak identification, or non-empty subsets of components ϕ_1 and ϕ_2. In this case (2.2) should follow from:

$$C_1\gamma = 0 \Rightarrow \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_{1T}(\theta_T)}{\partial \theta'}\gamma\right] = 0. \tag{2.3}$$

Note that (2.3) is a direct consequence of Assumption 2.4 applied to the drifting true value $\theta^0 = \theta_T$. Of course, conditions (2.2) and (2.3) are identical in the special case where the moment conditions $\phi_1(.)$ are linear. In other words, we can also easily state:

PROPOSITION 2.4. (*Special case with linear strong moment conditions*). *Consider the special case where $\phi_1(Y_t, \theta)$ is linear with respect to θ. Then conclusions of Proposition 2.3 hold.*

By contrast, in the general case of non-linear moment restrictions, we would like to be able to deduce (2.2) from (2.3) through a Taylor argument for each component $\phi_1^{(j)}$:

$$\frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_{1T}^{(j)}(\theta_T^*)}{\partial \theta'} \approx \frac{\sqrt{T}}{\lambda_T}\frac{\partial \bar{\phi}_{1T}^{(j)}(\theta_T)}{\partial \theta'} + \frac{\sqrt{T}}{\lambda_T}\frac{\partial^2 \bar{\phi}_{1T}^{(j)}(\theta_T^{**})}{\partial \theta \partial \theta'}(\theta_T^* - \theta_T). \tag{2.4}$$

[5] The fact that with only one rate of convergence, nearly-weak identification does not modify the standard equivalence between tests (that all have trivial asymptotic power) has already been noticed by Smith (2007): see footnote 3, p. 244.

The problem is then that, since we only know that $(\theta_T^* - \theta_T) = \mathcal{O}_P(1/\lambda_T)$, we can neglect the second term in the RHS of (2.4) only when $(\sqrt{T}/\lambda_T^2)$ goes to zero, or nearly-strong identification. Otherwise, there is no guarantee that $\sqrt{T}\bar{\phi}_T(\theta_0)$ is asymptotically normal under a sequence of 'weak' local alternatives $\theta_T = \theta_0 + \gamma/\lambda_T$, even when $C_1\gamma = 0$. This explains why we no longer get a characterization of local power through standard non-central chi-square. Fortunately, this problem may not invalidate the consistency of the standard GMM-score test against sufficiently slow sequences of weak alternatives. Ideally, one would like to prove that:

(A hypothetical claim of consistency) The GMM-score test is consistent against any sequence of local alternatives $\theta_T = \theta_0 + \gamma/\delta_T$, *for all* $\gamma \in \mathbb{R}^p$, *when* $\delta_T/\lambda_T \overset{T}{\to} 0:\ \text{Plim}[\xi_T^{NW}] = +\infty$.

This claim is trivially deduced from former results if either the direction γ is strong ($C_1\gamma \neq 0$) or with nearly-strong identification ($(\sqrt{T}/\lambda_T^2)$ goes to zero). The novelty would be to maintain consistency even in case of nearly-weak identification arbitrarily close to weak identification (λ_T goes to infinity arbitrarily slowly) insofar as the sequence of alternatives converges even more slowly than the sequence λ_T. We develop in the Appendix an argument to show that the consistency is likely, but not granted. In other words, the information may be quite weak but efficiently used for testing. This result is consistent with our estimation result in Section 3 below: we show that in any case, with nearly-weak global identification, a GMM estimator will converge at least at rate λ_T. This consistency property is no longer likely for the Kleibergen's modified score test, because its modification may waste this fragile part of information. This is the price to pay for a correct asymptotic size even in the limit case of no-identification. Kleibergen (2005) overlooked this problem since, as pointed out by Proposition 2.3, it may occur only when considering simultaneously two different rates of identification. Fragile identification may be wasted by Kleibergen's modification precisely because it comes with another piece of information which is stronger. To see this, the key is the aforementioned lack of logical implication from (2.3) to (2.2). As a result, the modified score statistic and the original one may have quite different asymptotic behaviours since, as reminded above:

$$\sqrt{T}\frac{\partial\tilde{\phi}_T^{(j)}(\theta_0)}{\partial\theta'} = \sqrt{T}\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta'} - \text{Cov}_{as}\left(\sqrt{T}\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta_0)\right)\text{Var}_{as}(\sqrt{T}\bar{\phi}_T(\theta_0))^{-1}\sqrt{T}\bar{\phi}_T(\theta_0). \quad (2.5)$$

It is quite evident from (2.5) that, when $\sqrt{T}\bar{\phi}_T(\theta_0)$ is not $\mathcal{O}_P(1)$, the modified score statistic may have an arbitrarily nasty asymptotic behaviour.

3. CONSISTENT ESTIMATION WITH NEARLY-WEAK INSTRUMENTS

3.1. General framework

In this section, we provide consistent estimation of the true (unknown) parameter θ^0. Standard GMM estimation defines its estimator $\hat{\theta}_T$ as follows:

DEFINITION 3.1. *Let* Ω_T *be a sequence of symmetric positive definite random matrices of size* K *which converges in probability towards a positive definite matrix* Ω. *A GMM estimator* $\hat{\theta}_T$ *of*

θ^0 is then defined as:

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} Q_T(\theta) \quad \textit{where} \qquad Q_T(\theta) \equiv \bar{\phi}'_T(\theta)\Omega_T\bar{\phi}_T(\theta), \tag{3.1}$$

with $\bar{\phi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\phi(Y_t,\theta)$, the empirical mean of the moment restrictions.

We consider here k_1 standard moment restrictions such that[6]

$$\sqrt{T}[\bar{\phi}_{1T}(\theta) - \rho_1(\theta)] = \mathcal{O}_P(1) \tag{3.2}$$

and k_2 $(=K - k_1)$ weaker moment restrictions such that

$$\sqrt{T}\left[\bar{\phi}_{2T}(\theta) - \frac{\lambda_T}{\sqrt{T}}\rho_2(\theta)\right] = \mathcal{O}_P(1) \quad \text{where} \quad \lambda_T = o(\sqrt{T}) \quad \text{and} \quad \lambda_T \overset{T}{\to} \infty. \tag{3.3}$$

λ_T measures the degree of weakness of the second group of moment restrictions: its corresponding component $\rho_2(.)$ is squeezed to zero and $\text{Plim}[\bar{\phi}_{2T}(\theta)] = 0$ for all $\theta \in \Theta$. Therefore, the probability limit of $\bar{\phi}_T(\theta)$ does not allow to discriminate between θ^0 and any $\theta \in \Theta$. In such a context, identification is a combined property of the functions $\phi(Y_t, .)$, and $\rho(.)$ and the asymptotic behaviour of λ_T. Assumption 3.1 below reinforces the standard CLT stated in Assumption 2.1 for moment conditions evaluated under the null by maintaining a functional CLT on the whole parameter set Θ. In this respect, we follow Stock and Wright (2000).[7]

ASSUMPTION 3.1. (*Identification*)

(i) *$\rho(.)$ is a continuous function from a compact parameter space $\Theta \subset \mathbb{R}^p$ into $\mathbb{R}^K$:*

$$\rho(\theta) = \begin{bmatrix}\rho_1(\theta)\\ \rho_2(\theta)\end{bmatrix} = 0 \iff \theta = \theta^0.$$

(ii) *The empirical process $(\Psi_T(\theta))_{\theta\in\Theta}$ obeys a functional CLT:*

$$\Psi_T(\theta) \equiv \sqrt{T}\begin{bmatrix}\bar{\phi}_{1T}(\theta) - \rho_1(\theta)\\ \bar{\phi}_{2T}(\theta) - \dfrac{\lambda_T}{\sqrt{T}}\rho_2(\theta)\end{bmatrix} \Rightarrow \Psi(\theta),$$

where $\Psi(\theta)$ is a Gaussian stochastic process on Θ with mean zero.

(iii) *λ_T is a deterministic sequence of positive real numbers with*

$$\lim_{T\to\infty}\lambda_T = \infty \quad \textit{and} \quad \lim_{T\to\infty}\frac{\lambda_T}{\sqrt{T}} = 0.$$

Our framework is different from the seminal paper by Stock and Wright (2000) in two ways. First, we rather consider nearly-weak identification, as introduced in a linear setting by Hahn and Kuersteiner (2002), than weak identification. In this sense, we are closer to Caner (2007). Second, we do not assume the *a priori* knowledge of a partition $\theta = (\alpha' \vdots \beta')'$, where α is strongly identified and β (near)-weakly identified. Our framework conveys the idea that identification

[6] Functions ρ_1 and ρ_2 are introduced in equation (1.9). See also Assumption 3.1 below for a formal definition.

[7] As stressed by Stock and Wright (2000), the uniformity in θ provided by the functional CLT is crucial in case of non-linear non-separable moment conditions, that is when the occurrences of θ and the observations in the moment conditions are not additively separable. By contrast, Hahn and Kuersteiner (2002) (linear case) and Lee (2004) (separable case) do not need to resort to a functional CLT.

is a matter of the moment conditions: nearly-weak identification is produced by the rates of convergence of the moment conditions. More precisely, Assumption 3.1 implies that, for the first set of moment conditions, we have (as for standard GMM),

$$\rho_1(\theta) = \text{Plim}[\bar{\phi}_{1T}(\theta)],$$

whereas we only have for the second set of moment conditions

$$\rho_2(\theta) = \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\bar{\phi}_{2T}(\theta)\right].$$

The above identification assumption allows for consistent GMM estimator, even in the case of nearly-weak identification.

THEOREM 3.1. (*Consistency of* $\hat{\theta}_T$). *Under Assumption 3.1, any GMM estimator* $\hat{\theta}_T$ *like (3.1) is weakly consistent.*

Under additional (local) identification assumption, we get consistency at the slowest available rate λ_T.

THEOREM 3.2. (*Rate of convergence*). *Under Assumptions 2.2, and 2.4–3.1, we have:* $\|\hat{\theta}_T - \theta^0\| = \mathcal{O}_P(1/\lambda_T)$.

In Section 4, we introduce a convenient rotation in the parameter space that allows us to identify some strongly identified directions at rate $\sqrt{T}$.

3.2. Single-equation linear IV model

We already pointed out a major difference between our framework and the existing literature: possible weakness is assigned to some specific instruments (or moment conditions), and not to the structural parameters of the model. The following single-equation linear IV model with two structural parameters, two orthogonal instruments (and no exogenous variables for convenience) sheds some light on the consequences of such distinction:

$$\begin{cases} \underset{(T,1)}{y} = \underset{(T,2)}{Y}\ \underset{(2,1)}{\theta} + \underset{(T,1)}{u} \\ \underset{(T,2)}{Y} = \underset{(T,2)}{[X_1\ X_2]}\ \underset{(2,2)}{\Pi} + \underset{(T,2).}{[V_1\ V_2]} \end{cases} \tag{3.4}$$

As commonly done in the literature the matrix of coefficients Π is artificially linked to the sample size T in order to introduce some (nearly)-weak identification issues. Depending on the focus of interest, several matrices $\Pi = \Pi_T$ may be considered.

(i) Staiger and Stock (1997) consider a framework with the same genuine weak identification pattern for all the parameters: $\Pi_T^{SS} = C/\sqrt{T}$.

(ii) Stock and Wright (2000) reinterpret this framework to consider simultaneously strong and weak identification patterns. This distinction is done at the parameter level and the structural parameter θ is (*a priori*) partitioned: $\theta = [\theta_1' \vdots \theta_2']'$, with θ_1 strongly-identified and θ_2 weakly-identified. Consequently, they introduce

$$\Pi_T^{SW} = \begin{bmatrix} c_{11} & c_{12}/\delta_T \\ c_{21} & c_{22}/\delta_T \end{bmatrix} \quad \text{with} \quad \delta_T = \sqrt{T}.$$

(iii) We consider simultaneously strong and near-weak identification patterns. This distinction is done at the moment condition level (or the instrument level), and we suppose the available moment conditions $E[\phi(.)]$ to be (*a priori*) partitioned: $\phi = [\phi_1' \vdots \phi_2']'$, with ϕ_1 strongly-identified and ϕ_2 nearly-weakly-identified. Consequently, we introduce

$$\Pi_T^{AR} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21}/\delta_T & c_{22}/\delta_T \end{bmatrix} \quad \text{with} \quad \delta_T \overset{T}{\to} \infty, \ \delta_T = o(\sqrt{T}).$$

Besides considering different degrees of weakness, modelling with Π_T^{SW} or Π_T^{AR} has an important implication.[8] Π_T^{AR} modifies the explanatory power of the second instrument X_2 only. As a result, one strong moment condition (associated with X_1) and one less informative (associated with X_2) naturally emerge. Intuitively, the standard restriction should identify one standard direction in the parameter space: however, this direction is still unknown and does not necessarily correspond to one of the structural parameters. On the other hand, modelling with Π_T^{SW} amounts to treat θ_2 as weakly identified, and to alter the explanatory powers of both instruments. In the linear model (3.4) the two moment conditions write as, $E[(y_t - Y_t'\theta^0)X_t] = 0$. When Π is replaced respectively by Π_T^{SW} and Π_T^{AR}, these moments can be rewritten as:[9]

$$\rho_{1s}^{SW}(\theta_1) + \rho_{1w}^{SW}(\theta_2)/\delta_T = 0 \quad \text{and} \quad \rho_{2s}^{SW}(\theta_1) + \rho_{2w}^{SW}(\theta_2)/\delta_T = 0 \tag{3.5}$$

$$\rho_1^{AR}(\theta_1, \theta_2) = 0 \quad \text{and} \quad \rho_2^{AR}(\theta_1, \theta_2)/\delta_T = 0. \tag{3.6}$$

Finally, modelling weak identification with Π_T^{SW} or Π_T^{AR} does not change the concentration parameter μ (or matrix), which is the well-accepted measure of the strength of the instruments in the literature. In the linear model (3.4) it is well-defined as:

$$\mu = \Sigma_V^{-1/2\prime}\Pi' X' X \Pi \Sigma_V^{-1/2} \quad \text{with} \quad \text{Var}(V) \equiv \Sigma_V$$

The determinant of this concentration matrix, when Π is replaced by Π_T^{SW} or by Π_T^{AR}, writes:

$$\det[\mu^{SW}] = \det[\mu^{AR}] = \frac{1}{\delta_T^2} \det[X'X] \det[\Sigma_V^{-1}] \det[C]^2. \tag{3.7}$$

Therefore, the same features of partial identification can be captured with both approaches. While with standard weak asymptotics, $\delta_T^2 = T$ and the concentration matrix has a finite limit (see also Andrews and Stock, 2007), nearly-weak asymptotics allow an infinite limit for the determinant of the concentration matrix but at a rate smaller than $\det[X'X] = \mathcal{O}(T)$. In this respect, there is no difference between Π_T^{SW} and Π_T^{AR}: only the rate of convergence to zero of respectively a row or a column of the matrix Π matters.

[8] This difference is relatively obvious, and not extensively discussed here. This also justifies why the same parameter δ_T may refer to different rates of convergence depending on the context.

[9] See the Appendix for exact definitions of $\rho_{1s}^{SW}(.)$, $\rho_{1w}^{SW}(.)$, $\rho_{2s}^{SW}(.)$, $\rho_{2w}^{SW}(.)$, $\rho_1^{AR}(.)$ and $\rho_2^{AR}(.)$.

Partial identification (Phillips, 1989) refers to Π matrices that may not be of full rank.[10] Generalization to asymptotic rank condition failures (at rate δ_T) comes at the price of having to specify which row or column asymptotically goes to zero. At least, our approach with ρ^{AR} remains true to the partial identification approach by working with "estimable functions" of the structural parameters, or functions that can be identified and square-root T consistently estimated. In the former example, it is clearly the case for $g(\theta) = \pi_{11}\theta_1 + \pi_{12}\theta_2$.[11] By contrast, the approach ρ^{SW} implies directly a partition of the structural parameters between θ_1 which is strongly identified and θ_2 which is not.

In the literature, T/δ_T^2 plays the role of the effective number of observations available for (nearly)-weak identification and implies a slow rate of convergence $T^{1/2}/\delta_T$ for functions of the parameters that are not strongly identified, unlike $g(\theta)$, which is endowed with a square-root T consistent estimator. By contrast with Choi and Phillips (1992), we even get a normal asymptotic distribution for this estimator, while they only got a mixture of normals: by assuming that δ_T goes to infinity slower than square-root T, we keep the possibility of consistent estimation for all structural parameters. However, when considering (in the next section) more general non-linear settings of nearly-weak identification, we will realize that the price to pay for standard square-root T asymptotic normality of strongly-identified functions of the parameters may be even higher: for instance, we will need to refer to nearly-strong identification, or assuming that the analogue of δ_T goes to infinity slower than $T^{1/4}$ (see also Section 2).

4. ASYMPTOTIC DISTRIBUTION THEORY

4.1. *Asymptotic distribution theory of 2S-GMM*

First, we introduce a convenient *rotation* in the parameter space to disentangle the rates of convergence. Special linear combinations of θ can actually be estimated at the standard rate $\sqrt{T}$, while others are still estimated at the slower rate λ_T. This is formalized by a CLT which allows the practitioner to apply usual GMM formulas without knowing *a priori* the identification pattern. We consider the following situation:

(i) Only k_1 equations (defined by ρ_1) have sample counterparts converging at rate $\sqrt{T}$. Unfortunately, we have (in general) a reduced rank problem, since $[\partial\rho_1(\theta^0)/\partial\theta']$ is not full column rank: its rank s_1 is strictly smaller than p. Intuitively, the first set of equations can only identify s_1 directions in the p-dimensional parameter space.

(ii) The other k_2 equations (defined by ρ_2) should be used to identify the remaining s_2 ($s_1 + s_2 = p$) directions.[12] However, this identification comes at the slower rate λ_T.

The parameter space will be separated into two subspaces, each of them characterized as the range of a full column rank matrix: respectively (p, s_1)-matrix R_1 and (p, s_2)-matrix R_2. Since R_2 characterizes the set of slow directions, it is naturally defined via the null space of $[\partial\rho_1'(\theta^0)/\partial\theta]$,

[10] The case considered and studied by Phillips (1989) and Choi and Phillips (1992) is even more general since they also address identification issues for coefficients of exogenous variables in the structural equation.

[11] It is shown in the Appendix that $\rho_1^{AR}(\theta) = (EX_{1t}^2)[g(\theta^0) - g(\theta)]$.

[12] By assumption, our set of moment conditions identifies the entire vector of parameters θ.

i.e. the directions not identified in a standard way:

$$\frac{\partial \rho_1(\theta^0)}{\partial \theta'} R_2 = 0. \tag{4.1}$$

Then, the remaining s_1 directions are defined as follows: $R = [R_1\ R_2]$, where R is a non-singular $(p,\ p)$-matrix with rank p that can be used as a matrix of a change of basis in $\mathbb{R}^p$. The new parameter is defined as $\eta = R^{-1}\theta$, that is

$$\theta = [R_1 \quad R_2] \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \begin{matrix} \updownarrow s_1 \\ \updownarrow s_2 \end{matrix}.$$

In the next subsection, we show that this reparametrization defines two subsets of directions, each associated with a specific rate of convergence. In general, there is no hope to get standard (non-degenerate) asymptotic normality of some components of the estimator $\hat{\theta}_T$ of θ^0: after a standard expansion of the first-order conditions, $\hat{\theta}_T$ now appears as asymptotically equivalent to some linear transformations of $\bar{\phi}_T(\theta)$ which are likely to mix up the two rates. Hence, all components of $\hat{\theta}_T$ might be contaminated by the slow rate of convergence. The advantage of the reparametrization is precisely to separate these two rates. In Section 4.3, we carefully compare our theory with Stock and Wright (2000) (in the linear case), and provide conditions under which some components of $\hat{\theta}_T$ converge (by chance) at the standard rate. These correspond to what is assumed *a priori* by Stock and Wright (2000) when they separate the structural parameters into one *standard-converging* group and one *slower-converging* one. The reparametrization may not be feasible in practice since the matrix R depends on the true unknown value of the parameter θ^0. However, we can still deduce a feasible inference strategy: technical details can be found in the companion paper by Antoine and Renault (2008).

Albeit with a mixture of different rates, the Jacobian matrix of moment conditions has a consistent sample counterpart (see Lemma A.1 in the proof of Proposition 2.1):

$$\sqrt{T}\frac{\partial \bar{\phi}_T(\theta^0)}{\partial \theta'} R \Lambda_T^{-1} \xrightarrow{P} J \quad \text{with} \quad J \equiv \frac{\partial \rho(\theta^0)}{\partial \theta'} R \quad \text{and} \quad \Lambda_T = \begin{pmatrix} \sqrt{T} Id_{s_1} & 0 \\ 0 & \lambda_T Id_{s_2} \end{pmatrix}. \tag{4.2}$$

Λ_T is the $(p,\ p)$ block diagonal scaling matrix, where Id_r denotes the identity matrix of size r; J is the $(K,\ p)$ block diagonal matrix with its two blocks respectively defined as the $(k_i,\ s_i)$ matrices $[\partial \rho_i(\theta^0)/\partial \theta'\ R_i]$ for $i = 1, 2$. Note that the coexistence of two rates of convergence (λ_T and $\sqrt{T}$) implies zero northeast and southwest blocks for J.

Moreover, to derive the asymptotic distribution of the GMM estimator $\hat{\theta}_T$, convergence result (4.2) needs to be fulfilled even when θ^0 is replaced by some preliminary consistent estimator θ_T^*. Hence, Taylor expansions must be robust to a λ_T-consistent estimator, the only rate guaranteed by Theorem 3.2. This situation is rather similar to the one met in Section 2 when introducing nearly-strong identification. Hence, if moments are non-linear, we need (similarly to Andrews, 1995, for non-parametric estimates) to assume that our near-weakly identified directions are estimated at a rate faster than $T^{1/4}$.[13,14] In addition, we want as usual uniform convergence of sample Hessian matrices. This leads us to maintain the following assumption:

ASSUMPTION 4.1. (*Taylor expansions*).

[13] The link between Andrews (1994, 1995) and our setting is further discussed in Antoine and Renault (2008).

[14] As already noted, this nearly-strong condition is irrelevant when the same degree of weakness is assumed for all moment conditions ($\phi = \phi_2$ or ϕ_1 is linear with respect to θ).

(i) $\phi_1(Y_t, \theta)$ *is linear with respect to* θ, *or* $\lim_{T\to\infty}[\lambda_T^2/\sqrt{T}] = \infty$.
(ii) $\bar{\phi}_T(\theta)$ *is twice continuously differentiable on the interior of* Θ *and is such that:*

$$\forall\, 1 \le k \le k_1 \; \frac{\partial^2 \bar{\phi}_{1T,k}(\theta)}{\partial\theta\partial\theta'} \xrightarrow{P} H_{1,k}(\theta) \quad \textit{and} \quad \forall\, 1 \le k \le k_2 \; \frac{\sqrt{T}}{\lambda_T}\frac{\partial^2 \bar{\phi}_{2T,k}(\theta)}{\partial\theta\partial\theta'} \xrightarrow{P} H_{2,k}(\theta)$$

uniformly on θ *in some neighbourhood of* θ^0, *for some* (p, p) *matricial function* $H_{i,k}(\theta)$ *for* $i = 1, 2$ *and* $1 \le k \le k_i$.

Up to unusual rates of convergence, we get a standard asymptotic normality result for the new parameter $\eta = R^{-1}\,\theta$:[15]

THEOREM 4.1. (*Asymptotic normality*).

(i) *Under Assumptions 2.1, 2.2 and 2.4–4.1, the GMM estimator* $\hat{\theta}_T$ *defined by (3.1) is such that:*

$$\Lambda_T R^{-1}\left(\hat{\theta}_T - \theta^0\right) \xrightarrow{d} \mathcal{N}\big(0, [J'\Omega J]^{-1} J'\Omega S(\theta^0)\Omega J[J^0\Omega J]^{-1}\big).$$

(ii) *Under Assumptions 2.1, 2.2 and 2.4–4.1, the asymptotic variance displayed in (i) is minimal when the GMM estimator* $\hat{\theta}_T$ *is defined with a weighting matrix* Ω_T *being a consistent estimator of* $\Omega = [S(\theta^0)]^{-1}$:

$$\Lambda_T R^{-1}\left(\hat{\theta}_T - \theta^0\right) \xrightarrow{d} \mathcal{N}(0, [J'[S(\theta^0)]^{-1} J]^{-1}).$$

Since $\hat{\theta}_T = (R_1\hat{\eta}_{1,T} + R_2\hat{\eta}_{2,T})$, a linear combination $a'\hat{\theta}_T$ of the estimated parameters of interest is endowed with $\sqrt{T}$ rate of convergence as $\hat{\eta}_{1,T}$ if and only if $a'R_2 = 0$, or when a belongs to the orthogonal space of the range of R_2. By equation (4.1), it also means that a is spanned by the columns of the matrix $[\partial\rho_1'(\theta^0)/\partial\theta]$. Hence, $a'\theta$ is strongly identified if and only if it is identified by the first set of moment conditions $\rho_1(\theta) = 0$.

As far as inference about θ is concerned, several practical implications of Theorem 4.1 are worth mentioning. Up to the unknown matrix R and the unknown rate of convergence λ_T, a consistent estimator of the asymptotic covariance matrix $(J'[S(\theta^0)]^{-1}J)^{-1}$ is[16]

$$T^{-1}\Lambda_T R^{-1}\left[\frac{\partial\bar{\phi}_T'(\hat{\theta}_T)}{\partial\theta} S_T^{-1} \frac{\partial\bar{\phi}_T(\hat{\theta}_T)}{\partial\theta'}\right]^{-1} [R']^{-1}\Lambda_T, \tag{4.3}$$

where S_T is a standard consistent estimator of the long-term covariance matrix.[17] From Theorem 4.1, for large T, $[\Lambda_T R^{-1}(\hat{\theta}_T - \theta^0)]$ behaves like a Gaussian random variable with mean zero and variance (4.3). One may be tempted to deduce that $\sqrt{T}(\hat{\theta}_T - \theta^0)$ behaves like a Gaussian random variable with mean 0 and variance

$$\left[\frac{\partial\bar{\phi}_T'(\hat{\theta}_T)}{\partial\theta} S_T^{-1} \frac{\partial\bar{\phi}_T(\hat{\theta}_T)}{\partial\theta'}\right]^{-1}. \tag{4.4}$$

And this would give the feeling that we are back to standard GMM formulas of Hansen (1982). As far as practical purposes are concerned, this intuition is correct: in particular, estimation of R

[15] Note that efficiency in Theorem 4.1(ii) is implicitly considered for the given set of moment restrictions $\bar{\phi}_T(.)$.
[16] This directly follows from Lemma A.5 in the Appendix.
[17] Recall that a consistent estimator of S_T of the long-term covariance matrix $S(\theta^0)$ can be built in the standard way from a preliminary inefficient GMM estimator of θ.

is not necessary to perform inference.[18] However, from a theoretical point of view, this is a bit misleading. First, since in general all components of $\hat{\theta}_T$ converge at the slow rate, $\sqrt{T}(\hat{\theta}_T - \theta^0)$ has no limit distribution. In other words, considering the asymptotic variance (4.4) refers to the inverse of an asymptotically singular matrix. Second, for the same reason, (4.4) is not an estimator of the standard population matrix

$$\left[\frac{\partial \rho'(\theta^0)}{\partial \theta}[S(\theta^0)]^{-1}\frac{\partial \rho(\theta^0)}{\partial \theta'}\right]^{-1}. \tag{4.5}$$

To conclude, if inference about θ is technically more involved than one may believe at first sight, it is actually very similar to standard GMM from a pure practical point of view. In other words, if a practitioner is not aware of the specific framework with moment conditions associated with several rates of convergence (coming, say, from the use of instruments of different qualities) then she can still provide reliable inference by using standard GMM formulas. In this respect, we generalize Kleibergen's (2005) result that inference can be performed without *a priori* knowledge of the identification setting. However, we are more general than Kleibergen (2005) since we allow moment conditions to display simultaneously different identification patterns.[19]

Finally, the score test defined in Section 2 is completed by the classical overidentification test:

THEOREM 4.2. (*J-test*). *Under Assumptions 2.1, 2.2, 2.4–4.1, with Ω_T consistent estimator of $[S(\theta^0)]^{-1}$, $T Q_T(\hat{\theta}_T) \xrightarrow{d} \chi^2(K-p)$.*

4.2. About the (non)-equivalence of 2S-GMM and CU-GMM

We now show that the nearly-strong identification condition is exactly needed to ensure that both strong and weak directions are equivalently estimated by efficient two-step GMM and by continuously updated GMM. This explains the aforementioned case of equivalence between the GMM score test and Kleibergen's modified score test. Hansen et al. (1996) define the continuously updated GMM estimator $\hat{\theta}_T^{CU}$ as:

DEFINITION 4.1. *Let $S_T(\theta)$ and $\bar{\phi}_T(\theta)$ be defined as in Assumption 2.1. The continuously updated GMM estimator $\hat{\theta}_T^{CU}$ of θ^0 is then defined as:*

$$\hat{\theta}_T^{CU} = \arg\min_{\theta \in \Theta} Q_T^{CU}(\theta) \quad \textit{where} \quad Q_T^{CU}(\theta) \equiv \bar{\phi}_T'(\theta) S_T^{-1}(\theta) \bar{\phi}_T(\theta). \tag{4.6}$$

PROPOSITION 4.1. (*Equivalence between CU-GMM and efficient GMM*). *Both strong and weak directions are equivalently estimated by efficient two-step GMM and continuously updated GMM, if the nearly-strong identification condition is ensured. That is,*

$$\Lambda_T R^{-1}\left(\hat{\theta}_T - \hat{\theta}_T^{CU}\right) = o_P(1) \quad \textit{when} \quad \frac{\lambda_T^2}{\sqrt{T}} \xrightarrow{T} \infty.$$

[18] Antoine and Renault (2008) provide feasible asymptotic distribution. It simply involves plugging in a consistent estimator of R.

[19] For notational simplicity, we only consider here one speed of nearly-weak identification λ_T. The more general framework with an arbitrary number of speeds is considered by Antoine and Renault (2008).

In the special case where the same degree of weakness is assumed for all moment conditions (see Proposition 2.3), CU-GMM and efficient GMM are always equivalent insofar as $\lambda_T \to \infty$.

Several comments are in order.

First, since non-degenerate asymptotic normality is obtained for $\Lambda_T R^{-1}(\hat{\theta}_T - \theta^0)$ (and not for $\sqrt{T}(\hat{\theta}_T - \theta^0)$), the relevant (non-trivial) equivalence result between two-step efficient GMM and continuously updated GMM relates to the suitably rescaled and rotated difference $\Lambda_T R^{-1}(\hat{\theta}_T - \hat{\theta}_T^{CU})$. As already mentioned, a naive reading of the asymptotic theory may spuriously lead to believe that standard formulas are maintained for asymptotic distribution of estimators of θ^0.

Second, the case with nearly-weak (and not nearly-strong) identification ($\lambda_T^2/\sqrt{T} \overset{T}{\to} 0$) breaks down the standard theory of efficient GMM, and the proof shows that there is no reason to believe that continuously updated GMM may be an answer. Two-step GMM and continuously updated GMM, albeit no longer equivalent, are both perturbed by higher-order terms with ambiguous effects on asymptotic distributions. The intuition given by higher-order asymptotics in standard identification settings cannot be extended to the case of nearly-weak identification. While the latter approach shows that continuously updated GMM is, in general, higher-order efficient than the standard two-step one (see Newey and Smith, 2004, and Antoine et al., 2007), there is no clear ranking of asymptotic performances under weak identification.

Third, it is important to keep in mind that all these difficulties are due to the fact that we consider realistic circumstances where two different degrees of identification are simultaneously involved. Standard results (equivalence, or rankings between different approaches) carry on when only one rate of convergence is considered.

4.3. *Single-equation linear IV model (continued)*

First, we define the reparametrization in the linear model of Section 3.4. The derivative of the standard moment restriction is

$$J_1 = \frac{\partial \rho_1(\theta^0)}{\partial \theta'} = \left[-E(Y_{1t}X_{1t}) \vdots - E\big(Y_{2t}X_{1t}\big) \right] = \left[-E\big(X_{1t}^2\big)\pi_{11} \vdots - E\big(X_{1t}^2\big)\pi_{12} \right].$$

Hence, the null space of J_1 is spanned by the vector $[-\pi_{12} \vdots \pi_{11}]'$ and its orthogonal by $[\pi_{11} \vdots \pi_{12}]'$. The legitimate matrix of change of basis R in the parameter space $\mathbb{R}^2$ and associated new parameter η can be defined as:

$$R = \frac{1}{\Delta} \begin{bmatrix} \pi_{11} & -\pi_{12} \\ \pi_{12} & \pi_{11} \end{bmatrix} \quad \text{with} \quad \Delta = \pi_{11}^2 + \pi_{12}^2,$$

$$\eta = R^{-1}\theta, \quad \eta_1 = \pi_{11}\theta_1 + \pi_{12}\theta_2 \quad \text{and} \quad \eta_2 = -\pi_{12}\theta_1 + \pi_{11}\theta_2.$$

Only the strong direction is completely determined, since the direction orthogonal to the null space of J_1 is uniquely determined and defines the first column of R. Any other direction could define a valid second column of R and a linear combination of θ_1 and θ_2 estimated at the slower rate.

Strictly speaking, the linear re-interpretation of Staiger and Stock (1997) by Stock and Wright (2000) is not nested in our setting, because each of their moment condition contains a strong part (that only depends on a subvector of parameter) and a weak one. This setting (through the definition of the matrix Π_T^{SW}) is conveniently built so as to know *a priori* which subset of the parameters is strongly identified. Now, if we pretend that we did not realize that the set of strongly-identified parameter was known and still perform the change of variables, we get: $J_1 \simeq [-\pi_{11} \vdots 0]$ with orthogonal space spanned by $[0 \vdots 1]'$. The change of basis is defined as:

$$R = \begin{bmatrix} a & 0 \\ b & 1 \end{bmatrix} \quad \text{with} \quad a \neq 0 \Rightarrow \eta = \begin{bmatrix} 1 & 0 \\ -b & a \end{bmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

As expected, we identify the strongly-identified direction as being parallel to θ_1. In other words, even the Stock and Wright approach can be accommodated within our general framework.

5. MONTE CARLO STUDY

We now report some Monte Carlo evidence about the intertemporally separable consumption capital asset pricing model (CCAPM) with constant relative risk-aversion (CRRA) preferences. Artificial data are generated to mimic the dynamic properties of the historical data.

5.1. Moment conditions

The Euler equations lead to the following moment conditions:

$$E\left[h_{t+1}(\theta)|\mathcal{I}_t\right] = 0 \quad \text{with} \quad h_t(\theta) = \delta r_t c_t^{-\gamma} - 1.$$

Our parameter of interest is then $\theta = [\delta\ \gamma]'$, with δ the discount factor and γ the preference parameter; (r_t, c_t) denote respectively a vector of asset returns and the consumption growth at time t. To estimate this model, our K instruments $Z_t \in \mathcal{I}_t$ include the constant as well as some lagged variables. We then rewrite the above moment conditions as[20]

$$E_0\left[\phi_{t,T}(\theta)\right] = E_0\left[h_{t+1}(\theta) \otimes Z_{t,T}\right].$$

5.2. Data generation

Our Monte Carlo design follows Tauchen (1986), Kocherlakota (1990), Hansen et al. (1996) and, more recently, Stock and Wright (2000). More precisely, the artificial data are generated by the method discussed in Tauchen and Hussey (1991). This method fits a 16-state Markov chain to the law of motion of the consumption and the dividend growths, so as to approximate a beforehand-calibrated Gaussian VAR(1) model (see Kocherlakota, 1990). The CCAPM-CRRA model is then used to price the stocks and the risk-free bond in each time period, yielding a time series of asset returns.

Since the data are generated from a general equilibrium model, even the econometrician does not know whether (δ, γ) are (near)-weakly identified or not. In a similar study, Stock and Wright (2000) impose a different treatment for the parameters δ and γ: typically, δ is taken as strongly identified whereas γ is not. We do not make such an assumption. Through our convenient reparametrization, we identify some directions of the parameter space that are strongly identified and some others that are not.

[20] To stress the potential weakness of the instruments, we add the subscript T to refer to the instruments.

5.3. Strong and weak moment conditions

We consider here three instruments: the constant, the centred lagged asset return and the centred lagged consumption growth. To apply our nearly-weak GMM estimation, we first need to separate the instruments (and the associated moment conditions) according to their *strength*. Typically, a moment restriction $E[\phi_t(\theta)]$ is (nearly)-weak when $E[\phi_t(\theta)]$ is *close* to zero for all θ. This means that the restriction does not permit to (partially) identify θ. Hence, we evaluate each moment restriction for a grid of parameter values. If the moment is uniformly *close* to 0 then we conclude to its weakness. This study can always be performed and is not specifically related to the Monte Carlo setting; the Monte Carlo setting is simply convenient to get rid of the simulation noise by averaging over the many simulated samples.

Figure 1 has been built with a sample size of 100 and 2500 Monte Carlo replications: top figures for set 1 with (a) constant instrument; (b) lagged asset return and (c) lagged consumption rate; bottom figures of set 2.[21] This study clearly reveals two groups of moment restrictions: (i) with the constant instrument, the associated restriction varies quite substantially with the parameter θ; (ii) with the lagged instruments, both associated restrictions remain fairly small when θ vary over the grid. The set of instruments, and accordingly of moment conditions, is then separated as follows:

$$\phi_{t,T}(\theta) = \begin{pmatrix} (\delta r_t c_t^{-\gamma} - 1) \\ (\delta r_t c_t^{-\gamma} - 1) \otimes \begin{bmatrix} r_{t-1} - \bar{r} \\ c_{t-1} - \bar{c} \end{bmatrix} \end{pmatrix},$$

$$\bar{\phi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \phi_{t,T}(\theta) \quad \text{with} \quad \sqrt{T} E[\bar{\phi}_T(\theta)] = \begin{pmatrix} \sqrt{T} & 0_{1,2} \\ 0_{2,1} & \lambda_T Id_2 \end{pmatrix} \begin{pmatrix} \rho_1(\theta) \\ \rho_2(\theta) \end{pmatrix}.$$

As emphasized earlier, our Monte Carlo study simulates a general equilibrium model. So, even the econometrician does not know in advance which moment conditions are weak and the level of this weakness. Hence, $\sqrt{T}$ and λ_T must be chosen so as to fulfil the following conditions, $\lambda_T = o(\sqrt{T})$ and $\sqrt{T} = o(\lambda_T^2)$.

In their theoretical considerations (Section 4.3), Stock and Wright (2000) also treat differently the covariances of the moment conditions. The strength of the constant instrument is actually used to provide some intuition on their identification assumptions (δ strongly-identified and γ weakly-identified). However, we think that if γ is weakly-identified, then it affects the covariance between r_t and $c_t^{-\gamma}$, and hence the identification of δ should be altered too. This actually matches some asymptotic results of Stock and Wright (2000) where the weak parameter affects the strong one, by preventing it to converge to a standard Gaussian random variable.

5.4. Reparametrization

To identify the standard directions in the parameter space, we first define the matrix of the change of basis. Recall that it is defined through the null space of the following matrix:

$$J_1 = \frac{\partial \rho_1(\theta^0)}{\partial \theta'}.$$

[21] Note that the conclusions are not affected when larger sample sizes are considered.

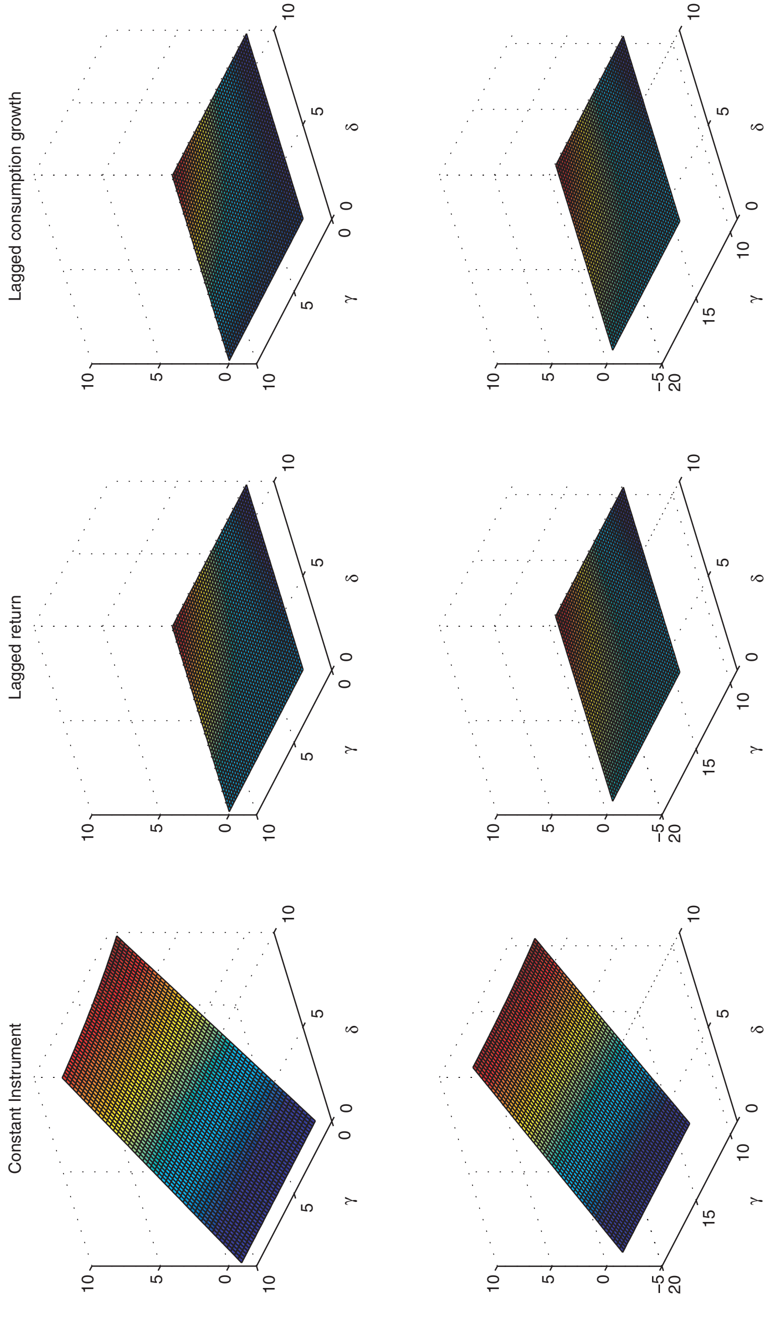

Figure 1. CCAPM: moment restrictions as a function of the parameter values θ.

Straightforward calculations lead to:

$$\left[\frac{\partial \phi_{1,t}(\theta)}{\partial \theta'}\right] = \left[\frac{\partial \phi_{1,t}(\theta)}{\partial \delta} \quad \frac{\partial \phi_{1,t}(\theta)}{\partial \gamma}\right] = \left[r_t c_t^{-\gamma} \quad \vdots \quad -\delta r_t \ln(c_t) c_t^{-\gamma}\right]$$

and J_1 is then defined as follows:

$$J_1 = \frac{\partial \rho_1(\theta^0)}{\partial \theta'} = \left[E\left[r_t c_t^{-\gamma^0}\right] \quad \vdots \quad -E\left[\delta^0 r_t \ln(c_t) c_t^{-\gamma^0}\right]\right].$$

The null space of J_1 is spanned by the vector $[-J_{12} \vdots J_{11}]'$ and its orthogonal by $[J_{11} \vdots J_{12}]'$. A convenient change of basis is then defined by the matrix:

$$R = \frac{1}{\Delta}\begin{bmatrix} J_{11} & -J_{12} \\ J_{12} & J_{11} \end{bmatrix} \quad \text{where} \quad \Delta = J_{11}^2 + J_{12}^2$$

and the new set of parameter is then obtained as,

$$\eta = R^{-1}\theta \Leftrightarrow \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} J_{11}\delta + J_{12}\gamma \\ -J_{12}\delta + J_{11}\gamma \end{pmatrix}.$$

The standard direction η_1 is completely determined: that is, the relative weights on δ and γ are known. As a convention, we normalize all vectors to unity and we also ensure that the subspaces defined respectively by the columns of R_2 and of R_1 are orthogonal.

5.5. *Asymptotic result*

The adapted asymptotic convergence result writes:

$$\begin{bmatrix} \sqrt{T}\left(\hat{\eta}_{1T} - \eta_1^0\right) \\ \lambda_T\left(\hat{\eta}_{2T} - \eta_2^0\right) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, (J'S(\theta^0)^{-1}J)^{-1}) \quad \text{with} \quad J = \begin{bmatrix} \frac{\partial \rho_1(\theta^0)}{\partial \theta'}R_1 & 0 \\ 0 & \frac{\partial \rho_2(\theta^0)}{\partial \theta'}R_2 \end{bmatrix}.$$

The approximation of J can easily be deduced from what has been done above.

5.6. *Results*

Monte Carlo results are provided for three instruments, constant, lagged asset return and lagged consumption growth, and two sets of parameter: set 1 (or model M1a as in Stock and Wright, 2000) where $\theta^{0\prime} = [0.97\ 1.3]$; set 2 (or model M1b) where $\theta^{0\prime} = [1.139\ 13.7]$. Model M1b has previously been found to produce non-normal estimator distributions.

First, as previously emphasized, the matrix of reparametrization is not known (even in our Monte Carlo setting) and is actually data dependent. We then investigate the variability of the true new parameter η^0. We found that even with small sample size ($T = 100$), the (estimated) true new parameter is really stable and does not depend much on the realization of the sample.

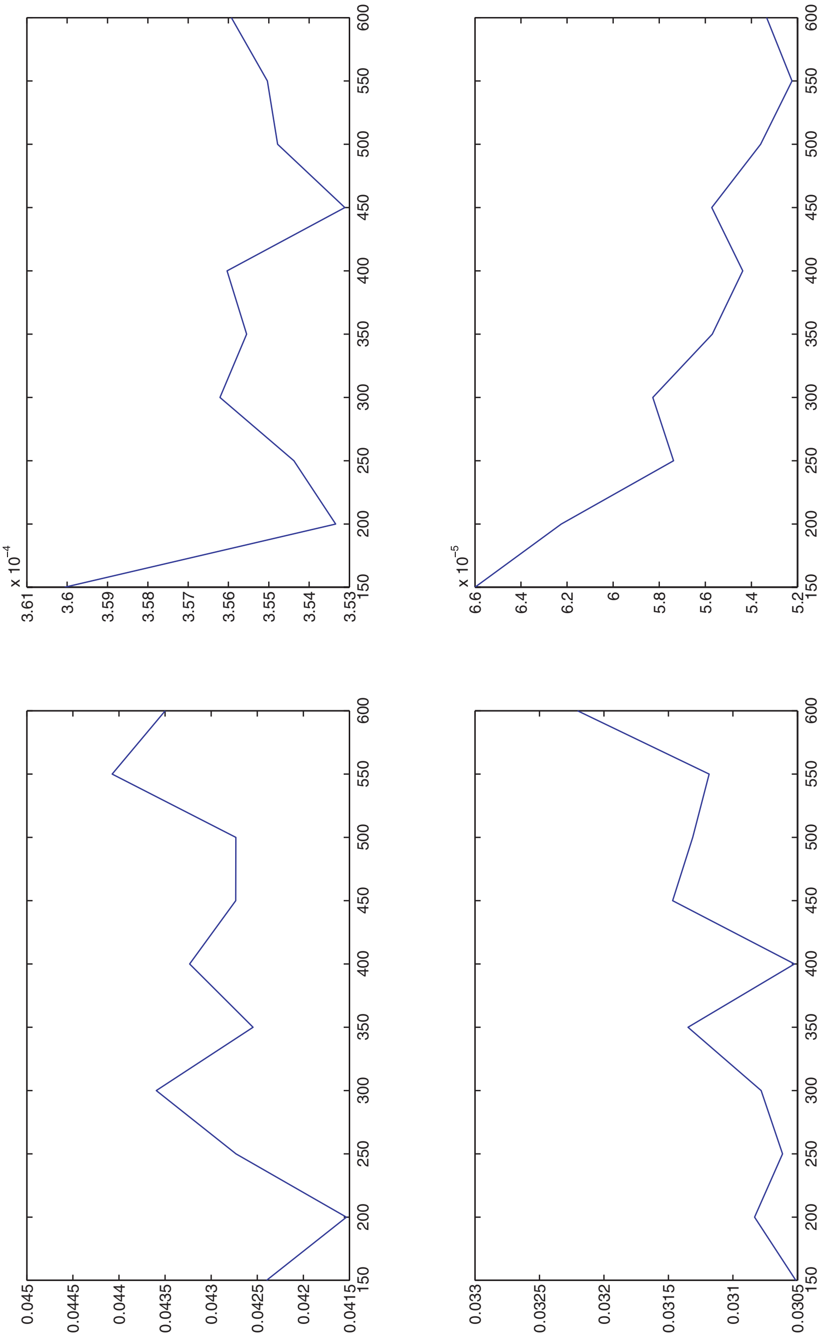

Figure 2. CCAPM: ratio for the variances as a function of the sample size.

For our two models, we find the following true new parameter:

$$\text{Set 1:} \quad \eta^0 = [0.9449 \quad 1.3183]; \qquad R^{-1} = \begin{bmatrix} 0.999 & -0.019 \\ 0.019 & 0.999 \end{bmatrix}$$

$$\text{Set 2:} \quad \eta^0 = [1.0356 \quad 13.7082]; \quad R^{-1} = \begin{bmatrix} 0.999 & -0.007 \\ 0.007 & 0.999 \end{bmatrix}.$$

Note that $\eta_1 = 0.999\delta - 0.019\gamma$ (for set 1) and $\eta_1 = 0.999\delta - 0.007\gamma$ (for set 2): in other words, we confirm Stock and Wright's (2000) intuition that the parameter η_1 which is estimated at the standard rate $\sqrt{T}$ is almost equal to δ. However, by contrast with Stock and Wright (2000), this point of view was not a prior belief but only an empirical conclusion.

Our findings are: (i) all the estimators are consistent; (ii) the variances of the estimators (for both $\hat{\eta}_T$ and $\hat{\theta}_T$) decrease to 0 with the sample size and (iii) according to our asymptotic results, in case of nearly-weak identification, the asymptotic variance of the new parameter $\hat{\eta}_{1T}$ should decrease faster with the sample size than the one of $\hat{\eta}_{2T}$. Figure 2 investigates this feature by plotting the evolution of the ratio of the Monte Carlo variance of $\hat{\eta}_{2T}$ and the Monte Carlo variance of $\hat{\eta}_{1T}$ with the sample size: top panels for set 1, left-hand panels for $\text{Var}(\hat{\eta}_{2T})/\text{Var}(\hat{\eta}_{1T})$ and right-hand panels for $\text{Var}(\hat{\theta}_{1T})/\text{Var}(\hat{\theta}_{2T})$; bottom panels for set 2. This ratio $\text{Var}(\hat{\eta}_{2T})/\text{Var}(\hat{\eta}_{1T})$ increases with T, especially for the second set of parameter values. This supports previous findings in the literature that the first set of parameter values leads to a less severe weak identification problem.

6. CONCLUSION

In a GMM context, this paper proposes a general framework to account for potentially weak instruments. In contrast with existing literature, the weakness is directly related to the moment conditions (through the instruments) and not to the parameters. More precisely, we consider two groups of moment conditions: the standard one associated with the standard rate of convergence $\sqrt{T}$ and the nearly-weak one associated with the slower rate λ_T. In this framework, the standard GMM-score-type test proposed by Newey and West (1987) is valid, and we do not need to resort to the correction proposed by Kleibergen (2005). Our comparative power study reveals that the above correction does have asymptotic consequences, especially with heterogeneous identification patterns. Hence, we recommend carefulness, especially when instruments of heterogeneous quality are used. Our proposed framework is not much more involved in terms of specifying the identification issues, and also helps identify the directions against which the tests have power.

Moreover, this framework ensures that GMM estimators of all parameters are consistent, but at rates possibly slower than usual. In addition, we identify and estimate efficiently (with non-degenerate asymptotic normality) the relevant directions, respectively strongly- or nearly-weakly identified, in the parameter space. This asymptotic distributional theory is practically relevant, since it allows inference without the knowledge, or estimation of the slow rate of identification. It allows in particular the application of the general Wald testing theory developed in a companion paper, Antoine and Renault (2008) (see also Lee, 2004). Moreover, we show that both for estimation and testing, continuous updated GMM is not always an answer to identification issues.

For notational and expositional simplicity, we focus here on two groups of moment conditions only. The extension to considering several degrees of weakness (think of a practitioner using several instruments of different informational qualities) is quite natural. Antoine and Renault (2008) specifically consider multiple groups of moment conditions associated with specific rates of convergence. However they do not explicitly consider any applications to identification issues, but rather applications in kernel, unit-root, extreme values or continuous time environments.

ACKNOWLEDGMENTS

We would like to thank O. Boldea, A. Inoue, P. Lavergne, L. Magee, W. Newey, R. Smith, and seminar participants at University of British Columbia, University of Indiana at Bloomington, Tilburg University and Yale University for helpful discussions.

REFERENCES

Andrews, D. W. K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica 62*, 43–72.

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric econometric models. *Econometric Theory 11*, 560–96.

Andrews, D. W. K. and P. Guggenberger (2007). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Cowles Foundation Discussion Paper 1605, Yale University.

Andrews, D. W. K. and J. H. Stock (2007). Inference with weak instruments. In R. Blundell, W. K. Newey and T. Persson (Eds.), *Advances in Economics and Econometrics, Volume III*, Econometric Society Monograph *43*, 122–73. Cambridge: Cambridge University Press.

Antoine, B., H. Bonnal and E. Renault (2007). On the efficient use of the informational content of estimating equations: implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics 138*, 461–87.

Antoine, B. and E. Renault (2008). Efficient minimum distance estimation with multiple rates of convergence. Working paper, Simon Fraser University.

Caner, M. (2007). Testing, estimation and higher order expansions in GMM with nearly-weak instruments. Working paper, North Carolina State University.

Choi, I. and P. C. B. Phillips (1992). Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics 51*, 113–50.

Guggenberger, P. and R. J. Smith (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory 21*, 667–709.

Hahn, J. and G. Kuersteiner (2002). Discontinuities of weak instruments limiting distributions. *Economics Letters 75*, 325–31.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–54.

Hansen, L. P., J. Heaton and A. Yaron (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics 14*, 262–80.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica 73*, 1103–23.

Kocherlakota, N. (1990). On tests of representative consumer asset pricing models. *Journal of Monetary Economics 26*, 285–304.

Lee, L. (2004). Pooling estimates with different rates of convergence—a minimum χ^2 approach with an emphasis on a social interaction model. Working paper, Ohio State University.

Newey, W. K. and R. J. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica 72*, 219–55.

Newey, W. K. and K. D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review 28*, 777–87.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory 5*, 181–240.

Smith, R. J. (2007). Weak instruments and empirical likelihood: a discussion of the papers by D. W. K. Andrews and J. H. Stock and Y. Kitamura. In R. Blundell, W. K. Newey and T. Persson (Eds.), *Advances in Economics and Econometrics, Volume III*, Econometric Society Monograph *43*, 238–60. Cambridge: Cambridge University Press.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–86.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*, 1055–96.

Tauchen, G. (1986). Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics 4*, 397–425.

Tauchen, G. and R. Hussey (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica 59*, 371–96.

APPENDIX

LEMMA A.1. *Under the assumptions of Proposition 2.1, there exists a (K, p)-matrix J of rank p,*

$$J = \text{Plim}\left[\sqrt{T}\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}R\Lambda_T^{-1}\right],$$

where Λ_T is (p, p) diagonal matrix whose first s_1 (resp. last s_2) coefficients are $\sqrt{T}$ (resp. λ_T).

Proof:

$$\sqrt{T}\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}R\Lambda_T^{-1} = \begin{bmatrix} \frac{\partial\bar{\phi}_{1T}(\theta_0)}{\partial\theta'}R_1 & \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T}(\theta_0)}{\partial\theta'}R_2 \\ \frac{\partial\bar{\phi}_{2T}(\theta_0)}{\partial\theta'}R_1 & \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta_0)}{\partial\theta'}R_2 \end{bmatrix}.$$

From Assumption 2.2, we have:

$$\text{Plim}\left[\frac{\partial\bar{\phi}_{1T}(\theta_0)}{\partial\theta'}R_1\right] = C_1R_1, \qquad \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta_0)}{\partial\theta'}R_2\right] = C_2R_2$$

and

$$\text{Plim}\left[\frac{\partial\bar{\phi}_{2T}(\theta_0)}{\partial\theta'}R_1\right] = \text{Plim}\left[\frac{\lambda_T}{\sqrt{T}}C_2R_1\right] = 0.$$

From Assumption 2.4, we have: $\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T}(\theta_0)}{\partial\theta'}R_2 = \frac{\sqrt{T}}{\lambda_T}C_1R_2 + \frac{U_TR_2}{\lambda_T}$ where $U_T = \mathcal{O}_P(1)$.

Since $C_1R_2 = 0$ (by definition of R_2), and $\lambda_T \overset{T}{\to} \infty$, we get: $\text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T}(\theta_0)}{\partial\theta'}R_2\right] = 0$.

We can then define the matrix J as: $J = \begin{bmatrix} C_1R_1 & 0 \\ 0 & C_2R_2 \end{bmatrix}$.

Note that J is of rank p since C_1R_1 (respectively C_2R_2) is of rank s_1 (respectively $s_2 = p - s_1$). The proof of Lemma A.1 is now completed. □

We need another preliminary result:

LEMMA A.2. *Under the assumptions of Proposition 2.1,*

$$T\Lambda_T^{-1}R'V_T(\theta_0) = J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) + o_P(1) = T\Lambda_T^{-1}R'\tilde{V}_T(\theta_0) + o_P(1).$$

Proof: From Assumption 2.1, $S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) = \mathcal{O}_P(1)$; combined with Lemma A.1 we get:

$$T\Lambda_T^{-1}R'V_T(\theta_0) = \sqrt{T}\Lambda_T^{-1}R'\frac{\partial\bar{\phi}_T'(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) = J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) + o_P(1).$$

We now show that: $T\Lambda_T^{-1}R'[V_T(\theta_0) - \tilde{V}_T(\theta_0)] = o_P(1)$. We have:

$$T\Lambda_T^{-1}R'\left[V_T(\theta_0) - \tilde{V}_T(\theta_0)\right] = \sqrt{T}\left[\left(\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'} - \frac{\partial\tilde{\phi}_T(\theta_0)}{\partial\theta'}\right)R\Lambda_T^{-1}\right]' \times S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0).$$

Since $S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) = \mathcal{O}_P(1)$, we only need to show that:

$$\sqrt{T}\left[\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'} - \frac{\partial\tilde{\phi}_T(\theta_0)}{\partial\theta'}\right]R\Lambda_T^{-1} = o_P(1).$$

The jth row of the above matrix writes:

$$\begin{aligned}&\sqrt{T}\left[\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta'} - \frac{\partial\tilde{\phi}_T^{(j)}(\theta_0)}{\partial\theta'}\right]R\Lambda_T^{-1}\\ &= -\operatorname{Cov}_{as}\left[\sqrt{T}\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta_0)\right](\operatorname{Var}_{as}[\sqrt{T}\bar{\phi}_T(\theta_0)])^{-1}\sqrt{T}\bar{\phi}_T(\theta_0)R\Lambda_T^{-1},\end{aligned}$$

which is $o_P(1)$ since $\sqrt{T}\bar{\phi}_T(\theta_0) = \mathcal{O}_P(1)$ and $R\Lambda_T^{-1} = o_P(1)$. This concludes the proof of Lemma A.2. □

Proof of Proposition 2.1 (*Equivalence under the null*): Consider the notations: $\text{rk}[C_1] = s_1 \leq p$; R_1 is a (p, s_1)-matrix such that: $\text{rk}[R_1] = s_1$, $\text{col}[R_1] = \text{col}[C_1']$; R_2 is a (p, s_2)-matrix such that $\text{rk}[R_2] = s_2$ and $\text{col}[R_2]$ is the null space of C_1.[22] By Assumption 2.2, $p = s_1 + s_2$, and $R = [R_1 \vdots R_2]$ is a non-singular (p, p)-matrix. We do not exclude the special case: $s_1 = p$, $s_2 = 0$ and $R = R_1$. From Lemma A.2 and Assumption 2.1, $[T\Lambda_T^{-1}R'V_T(\theta_0)]$ is asymptotically normal with zero mean and covariance matrix $J'[S^0]^{-1}J$.

Hence, $[T^2V_T'(\theta_0)\,R\Lambda_T^{-1}\,[J'[S^0]^{-1}J]^{-1}\Lambda_T^{-1}R'V_T(\theta_0)]$ converges in distribution towards a chi-square with p degrees of freedom. From Lemma A.1 and Assumption 2.1, this is also the case for:

$$TV_T'(\theta_0)R\Lambda_T^{-1}\left[\Lambda_T^{-1}R'\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}R\Lambda_T^{-1}\right]^{-1}\Lambda_T^{-1}R'V_T(\theta_0).$$

After simplifications, we find that:

$$\xi_T^{NW} = TV_T'(\theta_0)\left[\frac{\partial\bar{\phi}_T'(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}\right]^{-1}V_T(\theta_0) \xrightarrow{d} \chi^2(p),$$

[22] For any $(m \times n)$-matrix M, $\text{rk}[M]$ denotes its rank and $\text{col}[M]$ represents the subspace of $\mathbb{R}^n$ generated by the column vectors of M.

where $\chi^2(p)$ denotes the chi-square distribution with p degrees of freedom. When $\left[\frac{\partial\bar{\phi}'_T(\theta_0)}{\partial\theta}\right]$ is replaced by $\left[\frac{\partial\tilde{\phi}_T(\theta_0)}{\partial\theta}\right]$, the same argument leads to: $\xi_T^K \xrightarrow{d} \chi^2(p)$ and $\xi_T^K - \xi_T^{NW} \xrightarrow{P} 0$. □

LEMMA A.3. *Under the assumptions of Proposition 2.2,* $\sqrt{T}\bar{\phi}_T(\theta_0) \xrightarrow{d} \mathcal{N}(m, S^0)$ *with* $m = [-\gamma' C_1'\ 0]'$ *in case (i), and* $m = [0 - \gamma' C_2']'$ *in case (ii).*

Proof:

$$\sqrt{T}\bar{\phi}_T(\theta_T) = \sqrt{T}\bar{\phi}_T(\theta_0) + \sqrt{T}\frac{\partial\bar{\phi}_T(\theta_T^*)}{\partial\theta'}(\theta_T - \theta_0) \quad \text{with} \quad \theta_T^* = \theta_0 + \beta_T\gamma, \text{ where } \beta_T \in [0, \alpha_T].$$

From Assumption 2.1, we have: $\sqrt{T}\bar{\phi}_T(\theta_T) \xrightarrow{d} \mathcal{N}(0, S^0)$. Therefore, we only need to show:

$$\text{Plim}\left[\sqrt{T}\frac{\partial\bar{\phi}_T(\theta_T^*)}{\partial\theta'}(\theta_T - \theta_0)\right] = -m.$$

More precisely, in case (i), we need to show:

$$\text{Plim}\left[\frac{\partial\bar{\phi}_T(\theta_T^*)}{\partial\theta'}\gamma\right] = \begin{pmatrix} C_1\gamma \\ 0 \end{pmatrix}.$$

From Assumption 2.5, since $\text{Plim}[\theta_T^* - \theta_T] = 0$, we have:

$$\text{Plim}\left[\frac{\partial\bar{\phi}_{1T}(\theta_T^*)}{\partial\theta'}\right] = C_1 \quad \text{and} \quad \text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'}\right] = C_2 \Rightarrow \text{Plim}\left[\frac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'}\right] = 0.$$

And in case (ii), we need to show:

$$\text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_T(\theta_T^*)}{\partial\theta'}\gamma\right] = \begin{pmatrix} 0 \\ C_2\gamma \end{pmatrix}.$$

From Assumption 2.5:

$$\text{Plim}\left[\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'}\gamma\right] = C_2\gamma,$$

while for all components $\bar{\phi}_{1T}^{(j)}$, we have:

$$\begin{aligned} \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T}^{(j)}(\theta_T^*)}{\partial\theta'}\gamma &= \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T}^{(j)}(\theta_T)}{\partial\theta'}\gamma + \frac{\sqrt{T}}{\lambda_T}H_1^{(j)}(\theta_T^* - \theta_T) + o_P(1) = o_P(1), \\ \text{since} \quad \sqrt{T}\frac{\partial\bar{\phi}_{1T}^{(j)}(\theta_T)}{\partial\theta'} &= \sqrt{T}C_1 + \mathcal{O}_P(1) \\ \text{with} \quad C_1\gamma = 0, \quad \theta_T^* - \theta_T &= \mathcal{O}_P\left(\frac{1}{\lambda_T}\right) \text{ and } \frac{\lambda_T^2}{\sqrt{T}} \xrightarrow{T} \infty. \end{aligned}$$ □

Proof of Proposition 2.2 (*Local power of GMM score tests*): First, we get the asymptotic distribution result for ξ_T^{NW} in cases (i) and (ii). We study first the asymptotic distribution of $[\sqrt{T}\bar{\phi}_T(\theta_0)]$ when the true value is: $\theta_T = \theta_0 + \alpha_T\gamma$, with $\alpha_T = 1/\sqrt{T}$ in case (i), and $\alpha_T = 1/\lambda_T$ in case (ii). We need the following preliminary result: From Assumption 2.5, and Lemma A.1 the rescaled score can be written:

$$T\Lambda_T^{-1}R'V_T(\theta_0) = \sqrt{T}\Lambda_T^{-1}R'\frac{\partial\bar{\phi}'_T(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) = J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) + o_P(1),$$

since $S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) = \mathcal{O}_P(1)$. Thus, from Lemma A.3, we get:

$$T\Lambda_T^{-1}R'V_T(\theta_0) \xrightarrow{d} \mathcal{N}\left(J'[S^0]^{-1}m,\, J'[S^0]^{-1}J\right).$$

Therefore: $T^2V'_T(\theta_0)R\Lambda_T^{-1}[J'[S^0]^{-1}J]^{-1}\Lambda_T^{-1}R'V_T(\theta_0)$ converges in distribution towards a non-central chi-square with p degrees of freedom and non-centrality parameter:

$$\mu = m'[S^0]^{-1}J\left[J'[S^0]^{-1}J\right]^{-1}J'[S^0]^{-1}m.$$

From Assumption 2.5 and Lemma A.1, this also holds for:

$$\xi_T^{NW} = TV_T'(\theta_0)R\Lambda_T^{-1}\left[\Lambda_T^{-1}R'\frac{\partial\bar{\phi}_T'(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}R\Lambda_T^{-1}\right]^{-1}\Lambda_T^{-1}R'V_T(\theta_0). \tag{A.1}$$

We now only need to check the formula for the non-centrality parameter μ. It is convenient to introduce the following notations:

$$S^0 = \left(S^{1/2}\right)\left(S^{1/2}\right)', \qquad S^{-1/2} = \left(S^{1/2}\right)^{-1} \quad \text{and} \quad m_S = S^{-1/2}m, \qquad X = S^{-1/2}J.$$

Then: $\mu = m_S'X[X'X]^{-1}X'm_S$. We get the announced result if we can check that: $\mu = m_S'm_S$, that is, $P_Xm_S = m_S$, where $P_X = X[X'X]^{-1}X'$ is the orthogonal projection matrix on the column space of X. Therefore, we want to show that m_S belongs to the column space of X, or equivalently that m belongs to the column space of J. From the proof of Lemma A.1, we see that the column space of J is the set of vectors of $\mathbb{R}^K$ that can be written: $\begin{bmatrix} C_1R_1a \\ C_2R_2b \end{bmatrix}$ for some $a \in \mathbb{R}^{s_1}$, and $b \in \mathbb{R}^{s_2}$.

In case (i), $m = \begin{bmatrix} C_1\gamma \\ 0 \end{bmatrix}$ when choosing $b = 0$ and a as follows. First, $C_1\gamma = C_1\gamma^*$ where γ^* is the orthogonal projection of γ onto the orthogonal of the null space of C_1, or $\mathrm{col}[C_1'] = \mathrm{col}[R_1]$. Then, there exists some a such that $\gamma^* = R_1a$.

In case (ii), $m = \begin{bmatrix} 0 \\ C_2\gamma \end{bmatrix}$ when choosing $a = 0$. By definition, m belongs to the null space of C_1, which is spanned by the columns of R_2.

We now show that $\mathrm{Plim}[\xi_T^{NW} - \xi_T^K] = 0$ in cases (i) and (ii). We show that: $T\Lambda_T^{-1}R'[V_T(\theta_0) - \tilde{V}_T(\theta_0)] = o_P(1)$. We have:

$$T\Lambda_T^{-1}R'[V_T(\theta_0) - \tilde{V}_T(\theta_0)] = \sqrt{T}\left[\left(\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'} - \frac{\partial\tilde{\phi}_T(\theta_0)}{\partial\theta'}\right)R\Lambda_T^{-1}\right]' \times S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0).$$

Since $S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) = \mathcal{O}_P(1)$ from Lemma A.3, we only need to show that:

$$\sqrt{T}\left[\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'} - \frac{\partial\tilde{\phi}_T(\theta_0)}{\partial\theta'}\right]R\Lambda_T^{-1} = o_P(1).$$

The jth row of the above matrix writes:

$$\begin{aligned}&\sqrt{T}\left[\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta'} - \frac{\partial\tilde{\phi}_T^{(j)}(\theta_0)}{\partial\theta'}\right]R\Lambda_T^{-1}\\ &\quad = -\mathrm{Cov}_{as}\left[\sqrt{T}\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta_0)\right](\mathrm{Var}_{as}[\sqrt{T}\bar{\phi}_T(\theta_0)])^{-1}\sqrt{T}\bar{\phi}_T(\theta_0)R\Lambda_T^{-1},\end{aligned}$$

which is $o_P(1)$ since $\sqrt{T}\bar{\phi}_T(\theta_0) = \mathcal{O}_P(1)$, $R\Lambda_T^{-1} = o_P(1)$, and from Assumption 2.3:

$$B^{(j)} = \begin{bmatrix} B_1^{(j)} & B_2^{(j)} \end{bmatrix} \equiv \mathrm{Cov}_{as}\left(\sqrt{T}\frac{\partial\bar{\phi}_T^{(j)}(\theta_0)}{\partial\theta}, \sqrt{T}\bar{\phi}_T(\theta_0)\right)\left[\mathrm{Var}_{as}(\sqrt{T}\bar{\phi}_T(\theta_0))\right]^{-1} = \mathcal{O}_P(1).$$

And, we conclude: $\text{Plim}[\xi_T^K - \xi_T^{NW}] = 0$. □

Proof of Proposition 2.3 (*Special case with only one degree of weakness*): When $\phi = \phi_2$, the proof of Lemma A.3 does not require anymore the condition $\lambda_T^2/\sqrt{T} \xrightarrow{T} \infty$, and the result of Lemma A.3 remains valid in any case. The proof of Proposition 2.2 can then be rewritten based on this lemma to conclude that $\text{Plim}[\xi_T^{NW} - \xi_T^K] = 0$, and both score test statistics converge towards a non-central chi-square as announced. □

About the consistency of GMM score test (see Section 2): As in the proof of Proposition 2.2 (see equation (A.1)), ξ_T^{NW} can be rewritten as:

$$\xi_T^{NW} = TV_T'(\theta_0)R\Lambda_T^{-1}\sqrt{T}\left[\sqrt{T}\Lambda_T^{-1}R'\frac{\partial\bar{\phi}_T'(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}R\Lambda_T^{-1}\sqrt{T}\right]^{-1}\sqrt{T}\Lambda_T^{-1}R'V_T(\theta_0)$$

and we also have:

$$\begin{aligned} T\Lambda_T^{-1}R'V_T(\theta_0) &= \sqrt{T}\Lambda_T^{-1}R'\frac{\partial\bar{\phi}_T'(\theta_0)}{\partial\theta}S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_0) \\ &= J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_T) - J'S_T^{-1}(\theta_0)\sqrt{T}\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}(\theta_T - \theta_0) + o_P(1). \end{aligned} \tag{A.2}$$

The first term on the RHS of (A.2) is asymptotically normal with zero mean and variance $J'[S^0]^{-1}J$. The second term on the RHS of (A.2) can be written $(-\lambda_T/\delta_T)\pi_T$ with:

$$\pi_T = J'S_T^{-1}(\theta_0)\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_T(\theta_0)}{\partial\theta'}\gamma \equiv \pi_{1T} + \pi_{2T},$$

where the additive decomposition comes from the decomposition of $\partial\bar{\phi}_T(\theta_0)/\partial\theta'\gamma$ according to the two subsets of moment conditions ϕ_1 and ϕ_2. By Assumption 2.5:

$$\pi_{2T} = \{J'[S^0]^{-1}\}_2 C_2\gamma + o_P(1),$$

where, for any matrix M with K columns, $\{M\}_2$ is obtained by keeping only the second subset of columns of M (which corresponds to the second subset of moment conditions). Hence:

$$\begin{aligned} T\Lambda_T^{-1}R'V_T(\theta_0) &= J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_T) - \frac{\lambda_T}{\delta_T}\pi_{1T} - \frac{\lambda_T}{\delta_T}(\{J'[S^0]^{-1}\}_2 C_2\gamma + o_P(1)) \\ &= J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_T) + Z_T, \end{aligned}$$

where $\|Z_T\|$ is likely to go to infinity when (λ_T/δ_T) goes to infinity and $C_2\gamma \neq 0$. When $\|Z_T\|$ goes to infinity, since $J'S_T^{-1}(\theta_0)\sqrt{T}\bar{\phi}_T(\theta_T)$ is $\mathcal{O}_P(1)$, ξ_T^{NW} which is a positive definite quadratic function of $T\Lambda_T^{-1}R'V_T(\theta_0)$ goes to infinity and the test is consistent. The only case where this consistency is lost is when: $\pi_{1T} + \{J'[S^0]^{-1}\}_2\, C_2\gamma = o_P(\delta_T/\lambda_T)$. □

The consistency of the minimum distance estimator $\hat{\theta}_T$ is a direct implication of the identification Assumption 3.1 jointly with the following lemma:

LEMMA A.4. $\|\rho(\hat{\theta}_T)\| = \mathcal{O}_P(1/\lambda_T)$.

Proof: From (3.1), the objective function is written as follows:

$$Q_T(\theta) = \left[\frac{\Psi_T(\theta)}{\sqrt{T}} + \frac{\Lambda_T}{\sqrt{T}}\rho(\theta)\right]'\Omega_T\left[\frac{\Psi_T(\theta)}{\sqrt{T}} + \frac{\Lambda_T}{\sqrt{T}}\rho(\theta)\right], \quad \text{where } \Lambda_T = \begin{bmatrix} Id_{k_1} & 0 \\ 0 & \frac{\lambda_T}{\sqrt{T}}Id_{k_2} \end{bmatrix}.$$

Since $\hat{\theta}_T$ is the minimizer of $Q(.)$ we have in particular:

$$Q_T(\hat{\theta}_T) \leq Q(\theta^0) \Longrightarrow \left[\frac{\Psi_T(\hat{\theta}_T)}{\sqrt{T}} + \frac{\Lambda_T}{\sqrt{T}}\rho(\hat{\theta}_T)\right]' \Omega_T \left[\frac{\Psi_T(\hat{\theta}_T)}{\sqrt{T}} + \frac{\Lambda_T}{\sqrt{T}}\rho(\hat{\theta}_T)\right] \leq \frac{\Psi_T'(\theta^0)}{\sqrt{T}}\Omega_T\frac{\Psi_T(\theta^0)}{\sqrt{T}}.$$

Denoting $d_T = \Psi_T'(\hat{\theta}_T)\Omega_T\Psi_T(\hat{\theta}_T) - \Psi_T'(\theta^0)\Omega_T\Psi_T(\theta^0)$, we get:

$$[\Lambda_T\rho(\hat{\theta}_T)]'\Omega_T[\Lambda_T\rho(\hat{\theta}_T)] + 2[\Lambda_T\rho(\hat{\theta}_T)]'\Omega_T\Psi_T(\hat{\theta}_T) + d_T \leq 0.$$

Let μ_T be the smallest eigenvalue of Ω_T. The former inequality implies:

$$\mu_T\|\Lambda_T\rho(\hat{\theta}_T)\|^2 - 2\|\Lambda_T\rho(\hat{\theta}_T)\| \times \|\Omega_T\Psi_T(\hat{\theta}_T)\| + d_T \leq 0.$$

In other words, $x_T = \|\Lambda_T\rho(\hat{\theta}_T)\|$ solves the inequality:

$$x_T^2 - \frac{2\|\Omega_T\Psi_T(\hat{\theta}_T)\|}{\mu_T}x_T + \frac{d_T}{\mu_T} \leq 0$$

$$\Rightarrow \frac{\|\Omega_T\Psi_T(\hat{\theta}_T)\|}{\mu_T} - \sqrt{\Delta_T} \leq x_T \leq \frac{\|\Omega_T\Psi_T(\hat{\theta}_T)\|}{\mu_T} + \sqrt{\Delta_T} \text{ with } \Delta_T = \frac{\|\Omega_T\Psi_T(\hat{\theta}_T)\|^2}{\mu_T^2} - \frac{d_T}{\mu_T}.$$

Since $x_T \geq \lambda_T\|\rho_T(\hat{\theta}_T)\|$ we want to show that $x_T = \mathcal{O}_P(1)$, that is

$$\frac{\|\Omega_T\Psi_T(\hat{\theta}_T)\|}{\mu_T} = \mathcal{O}_P(1) \quad \text{and} \quad \Delta_T = \mathcal{O}_P(1) \Leftrightarrow \frac{\|\Omega_T\Psi_T(\hat{\theta}_T)\|}{\mu_T} = \mathcal{O}_P(1) \quad \text{and} \quad \frac{d_T}{\mu_T} = \mathcal{O}_P(1).$$

Note that since $\det(\Omega_T) \stackrel{P}{\rightarrow} \det(\Omega) > 0$, no subsequence of Ψ_T can converge in probability towards zero and thus we can assume (for T sufficiently large) that μ_T remains lower bounded away from zero with asymptotic probability one. Therefore, we just have to show that:

$$\|\Omega_T\Psi_T(\hat{\theta}_T)\| = \mathcal{O}_P(1) \quad \text{and} \quad d_T = \mathcal{O}_P(1).$$

Since $\text{tr}(\Omega_T) \stackrel{P}{\rightarrow} \text{tr}(\Omega)$ (where $\text{tr}[M]$ denotes the trace of any square matrix M), and the sequence $\text{tr}(\Omega_T)$ is upper bounded in probability, so are all the eigenvalues of Ω_T. Therefore, the required boundedness in probability just results from our Assumption 3.1(ii) ensuring that:

$$\sup_{\theta\in\Theta}\|\Psi_T(\theta)\| = \mathcal{O}_P(1). \qquad \square$$

Proof of Theorem 3.1 (*Consistency of* $\hat{\theta}_T$): Let us then deduce the weak consistency of $\hat{\theta}_T$ by a contradiction argument. If $\hat{\theta}_T$ is not consistent, there exists some positive ϵ such that:

$$P[\|\hat{\theta}_T - \theta^0\| > \epsilon]$$

does not converge to zero. Then we can define a subsequence $(\hat{\theta}_{T_n})_{n\in\mathbb{N}}$ such that, for some positive η: $P[\|\hat{\theta}_{T_n} - \theta^0\| > \epsilon] \geq \eta$ for $n \in \mathbb{N}$. Let us denote $\alpha = \inf_{\|\theta-\theta^0\|>\epsilon}\|\rho(\theta)\| > 0$ by Assumption 3.1(i).

Then for all $n \in \mathbb{N}$: $P[\|\rho(\hat{\theta}_{T_n})\| \geq \alpha] > 0$. When considering the identification Assumption 3.1(iii), this last inequality contradicts Lemma *A*.4. This completes the proof of consistency. $\square$

Proof of Theorem 3.2 (*Rate of convergence*): From Lemma A.4 $\|\rho(\hat{\theta}_T)\| = \|\rho(\hat{\theta}_T) - \rho(\theta^0)\| = \mathcal{O}_P(1/\lambda_T)$ and by application of the Mean-Value Theorem, for some $\tilde{\theta}_T$ between $\hat{\theta}_T$ and θ^0 component by component, we get:

$$\left\| \frac{\partial \rho(\tilde{\theta}_T)}{\partial \theta'} \left(\hat{\theta}_T - \theta^0\right) \right\| = \mathcal{O}_P\left(\frac{1}{\lambda_T}\right).$$

Note that, by a common abuse of notation, we omit to stress that $\tilde{\theta}_T$ actually depends on the component of $\rho(.)$. The key point is that since $\rho(.)$ is continuously differentiable and $\tilde{\theta}_T$, as $\hat{\theta}_T$, converges in probability towards θ^0, we have:

$$\frac{\partial \rho(\tilde{\theta}_T)}{\partial \theta'} \xrightarrow{P} \frac{\partial \rho(\theta^0)}{\partial \theta'} \Rightarrow \frac{\partial \rho(\theta^0)}{\partial \theta'} \times \left(\hat{\theta}_T - \theta^0\right) = z_T,$$

with $\|z_T\| = \mathcal{O}_P(1/\lambda_T)$. Since $\partial\rho(\theta^0)/\partial\theta'$ is full column rank, we deduce that:

$$\left(\hat{\theta}_T - \theta^0\right) = \left[\frac{\partial \rho'(\theta^0)}{\partial \theta}\frac{\partial \rho(\theta^0)}{\partial \theta'}\right]^{-1} \frac{\partial \rho'(\theta^0)}{\partial \theta} z_T \quad \text{also fulfils: } \left\|\hat{\theta}_T - \theta^0\right\| = \mathcal{O}_P\left(1/\lambda_T\right).$$

□

Proof of Equations (3.5) and (3.6) (*Moments in the linear IV model*): In the linear model (3.4),

$$y_t = X_{1t}\pi_{11}\theta_1^0 + X_{2t}\pi_{21}\theta_1^0 + X_{1t}\pi_{21}\theta_2^0 + X_{2t}\pi_{22}\theta_2^0 + V_{1t}\theta_1^0 + V_{2t}\theta_2^0 + u_t$$

and the two moment conditions write as follows (assuming orthogonal instruments):

$$E[(y_t - Y_t'\theta)X_t] = \begin{pmatrix} E\left(X_{1t}^2\right)\pi_{11}\left(\theta_1^0 - \theta_1\right) + E\left(X_{1t}^2\right)\pi_{12}\left(\theta_2^0 - \theta_2\right) \\ E\left(X_{2t}^2\right)\pi_{21}\left(\theta_1^0 - \theta_1\right) + E\left(X_{2t}^2\right)\pi_{22}\left(\theta_2^0 - \theta_2\right) \end{pmatrix},$$

When Π is replaced by Π_T^{SW}, the moment conditions are:

$$\begin{pmatrix} \left[E\left(X_{1t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{11} + \left[E\left(X_{1t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{12}/\delta_T \\ \left[E\left(X_{2t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{21} + \left[E\left(X_{2t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{22}/\delta_T \end{pmatrix} \equiv \begin{pmatrix} \rho_{1s}^{SW}\left(\theta_1^0\right) + \rho_{1w}^{SW}\left(\theta_2^0\right)/\delta_T \\ \rho_{2s}^{SW}\left(\theta_1^0\right) + \rho_{2w}^{SW}\left(\theta_2^0\right)/\delta_T \end{pmatrix}$$

with

$$\begin{aligned} \rho_{1s}^{SW}(\theta_1) &= \left[E\left(X_{1t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{11} \quad \rho_{1w}^{SW}(\theta_2) = \left[E\left(X_{1t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{12} \\ \rho_{2s}^{SW}(\theta_1) &= \left[E\left(X_{2t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{21} \quad \rho_{2w}^{SW}(\theta_2) = \left[E\left(X_{2t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{22}. \end{aligned}$$

When Π is replaced by Π_T^{AR}, the moment conditions are:

$$\begin{pmatrix} \left[E\left(X_{1t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{11} + \left[E\left(X_{1t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{12} \\ \left[E\left(X_{2t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{21}/\delta_T + \left[E\left(X_{2t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{22}/\delta_T \end{pmatrix} \equiv \begin{pmatrix} \rho_1^{AR}(\theta_1, \theta_2) \\ \rho_2^{AR}(\theta_1, \theta_2)/\delta_T \end{pmatrix}$$

with

$$\begin{aligned} \rho_1^{AR}(\theta_1, \theta_2) &= \left[E\left(X_{1t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{11} + \left[E\left(X_{1t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{12} \\ \rho_2^{AR}(\theta_1, \theta_2) &= \left[E\left(X_{2t}^2\right)\left(\theta_1^0 - \theta_1\right)\right]c_{21} + \left[E\left(X_{2t}^2\right)\left(\theta_2^0 - \theta_2\right)\right]c_{22}. \end{aligned}$$

□

Proof of Equation (3.7) (*Determinant of the concentration matrix*): By definition, $\mu = \Sigma_V^{-1/2'} \Pi' X' X \Pi \Sigma_V^{-1/2}$. Standard calculation rules for determinant yield to:

$$\det[\mu] = \det\left[\Sigma_V^{-1/2'}\Pi' X' X \Pi \Sigma_V^{-1/2}\right] = \det\left[\Sigma_V^{-1}\right]\det\left[X'X\right](\det[\Pi])^2$$

with $\det[\Pi] = \pi_{11}\pi_{22} - \pi_{12}\pi_{21}$. When Π is replaced respectively by Π_T^{SW} and Π_T^{AR}, we have:

$$\det\left[\Pi_T^{SW}\right] = \frac{\det[C]}{\delta_T} \quad \text{and} \quad \det\left[\Pi_T^{AR}\right] = \frac{\det[C]}{\delta_T}.$$

Hence: $\det[\mu^{SW}] = \det[\mu^{AR}] = d/\delta_T^2$ with $d = \det[\Sigma_V^{-1}]\det[X'X](\det[C])^2$. □

First, we need a preliminary result which naturally extends Lemma A.1 when the true value is replaced by some preliminary consistent estimator θ_T^*.

LEMMA A.5. *Under Assumptions 3.1–4.1, if θ_T^* is such that $\|\theta_T^* - \theta^0\| = \mathcal{O}_P(1/\lambda_T)$, then*

$$\sqrt{T}\frac{\partial\bar{\phi}_T\left(\theta_T^*\right)}{\partial\theta'}R\Lambda_T^{-1} \xrightarrow{P} J \quad \textit{when} \quad T \to \infty.$$

Proof: First, note that

$$\sqrt{T}\frac{\partial\bar{\phi}_T\left(\theta_T^*\right)}{\partial\theta'}R\tilde{\Lambda}_T^{-1} = \begin{bmatrix} \dfrac{\partial\bar{\phi}_{1T}(\theta_T^*)}{\partial\theta'}R_1 & \dfrac{\sqrt{T}}{\lambda_T}\dfrac{\partial\bar{\phi}_{1T}(\theta_T^*)}{\partial\theta'}R_2 \\ \dfrac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'}R_1 & \dfrac{\sqrt{T}}{\lambda_T}\dfrac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'}R_2 \end{bmatrix}.$$

To get the results, we have to show the following:

$$\text{(i)}\frac{\partial\bar{\phi}_{1T}\left(\theta_T^*\right)}{\partial\theta'} \xrightarrow{P} \frac{\partial\rho_1(\theta^0)}{\partial\theta'} \qquad \text{(ii)}\ \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}\left(\theta_T^*\right)}{\partial\theta'} \xrightarrow{P} \frac{\partial\rho_2(\theta^0)}{\partial\theta'}$$

$$\text{(iii)}\ \frac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'}R_1 \xrightarrow{P} 0 \qquad \text{(iv)}\ \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T}(\theta_T^*)}{\partial\theta'}R_2 \xrightarrow{P} 0.$$

(i) From Assumption 2.2(ii): $\frac{\partial\bar{\phi}_{1T}(\theta^0)}{\partial\theta'} - \frac{\partial\rho_1(\theta^0)}{\partial\theta'} = o_P(1)$. The Mean-Value Theorem applies to the kth component of $[\partial\bar{\phi}_{1T}/\partial\theta']$ for $1 \le k \le k_1$. For some $\tilde{\theta}$ between θ^0 and θ_T^*:

$$\frac{\partial\bar{\phi}_{1T,k}\left(\theta_T^*\right)}{\partial\theta'} - \frac{\partial\bar{\phi}_{1T,k}(\theta^0)}{\partial\theta'} = \left(\theta_T^* - \theta^0\right)'\frac{\partial^2\bar{\phi}_{1T,k}(\tilde{\theta}_T)}{\partial\theta\partial\theta'} = o_P(1),$$

where the last equality follows from Assumption 4.1(ii) and the assumption on θ_T^*.

(ii) From Assumption 2.2(ii):

$$\sqrt{T}\frac{\partial\bar{\phi}_{2T}(\theta^0)}{\partial\theta'} - \lambda_T\frac{\partial\rho_2(\theta_0)}{\partial\theta'} = \mathcal{O}_P(1) \Rightarrow \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta^0)}{\partial\theta'} - \frac{\partial\rho_2(\theta^0)}{\partial\theta'} = o_P(1)$$

because $\lambda_T \xrightarrow{T} \infty$. The Mean-Value Theorem applies to the kth component of $\partial\bar{\phi}_{2T}/\partial\theta'$ for $1 \le k \le k_2$. For some $\tilde{\theta}_T$ between θ^0 and θ_T^*, we have:

$$\frac{\sqrt{T}}{\lambda_T}\left(\frac{\partial\bar{\phi}_{2T,k}\left(\theta_T^*\right)}{\partial\theta'} - \frac{\partial\bar{\phi}_{2T,k}(\theta^0)}{\partial\theta'}\right) = (\theta_T^* - \theta^0)\frac{\sqrt{T}}{\lambda_T}\frac{\partial^2\bar{\phi}_{2T,k}(\tilde{\theta}_T)}{\partial\theta\partial\theta'} = o_P(1),$$

where the last equality follows from Assumption 4.1(ii) and the assumption on θ_T^*.

(iii) $\frac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'} = \frac{\lambda_T}{\sqrt{T}} \times \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{2T}(\theta_T^*)}{\partial\theta'} = o_P(1)$ because of (ii) and $\lambda_T = o(\sqrt{T})$.

(iv) Recall the Mean-Value Theorem from (i). For $1 \le k \le k_1$ and $\tilde{\theta}_T$ between θ^0 and θ_T^*:

$$\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T,k}\left(\theta_T^*\right)}{\partial\theta'} = \frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T,k}(\theta^0)}{\partial\theta'} + \lambda_T(\theta_T^* - \theta^0)'\frac{1}{\lambda_T}\frac{\sqrt{T}}{\lambda_T}\frac{\partial^2\bar{\phi}_{1T,k}(\tilde{\theta}_T)}{\partial\theta\partial\theta'}.$$

The second member of the RHS is $o_P(1)$ because of Assumptions 3.1(iii), 4.1(i) and 4.1(ii) and the assumption on θ_T^*. Now we just need to show that the first member of the RHS is $o_P(1)$. Recall from Assumption 2.4 that

$$\sqrt{T}\left[\frac{\partial\bar{\phi}_{1T}(\theta^0)}{\partial\theta'} - \frac{\partial\rho_1(\theta^0)}{\partial\theta'}\right] = \mathcal{O}_P(1) \Rightarrow \frac{\sqrt{T}}{\lambda_T}\left[\frac{\partial\bar{\phi}_{1T}(\theta^0)}{\partial\theta'} - \frac{\partial\rho_1(\theta^0)}{\partial\theta'}\right] R_2 = \mathcal{O}_P\left(\frac{1}{\lambda_T}\right).$$

By definition R_2 is such that $\frac{\partial\rho_1(\theta^0)}{\partial\theta'} R_2 = 0$. Hence we get

$$\frac{\sqrt{T}}{\lambda_T}\frac{\partial\bar{\phi}_{1T,k}(\theta^0)}{\partial\theta'} R_2^0 = \mathcal{O}_P\left(\frac{1}{\lambda_T}\right) = o_P(1). \qquad \square$$

Proof of Theorem 4.1 (*Asymptotic normality*): From the optimization problem (3.1), the first-order conditions for $\hat{\theta}_T$ are written as:

$$\frac{\partial\bar{\phi}_T'(\hat{\theta}_T)}{\partial\theta}\Omega\bar{\phi}_T(\hat{\theta}_T) = 0.$$

A mean-value expansion yields to:

$$\frac{\partial\bar{\phi}_T'(\hat{\theta}_T)}{\partial\theta}\Omega\bar{\phi}_T(\theta^0) + \frac{\partial\bar{\phi}_T'(\hat{\theta}_T)}{\partial\theta}\Omega\frac{\partial\bar{\phi}_T(\tilde{\theta}_T)}{\partial\theta'} \times \left(\hat{\theta}_T - \theta^0\right) = 0,$$

where $\tilde{\theta}_T$ is between $\hat{\theta}_T$ and θ^0. Pre-multiplying the above equation by the non-singular matrix $T\Lambda_T^{-1}R'$ yields to an equivalent set of equations:

$$\hat{J}_T'\Omega\left[\sqrt{T}\bar{\phi}_T(\theta^0)\right] + \hat{J}_T'\Omega\tilde{J}_T \times \Lambda_T R^{-1}\left(\hat{\theta}_T - \theta^0\right) = 0$$

after defining:

$$\hat{J}_T = \sqrt{T}\frac{\partial\bar{\phi}_T(\hat{\theta}_T)}{\partial\theta'}R\Lambda_T^{-1} \quad \text{and} \quad \tilde{J}_T = \sqrt{T}\frac{\partial\bar{\phi}_T(\tilde{\theta}_T)}{\partial\theta}R\Lambda_T^{-1}.$$

From Theorem 3.2 and Lemma A.5, we can deduce that: Plim $\tilde{J}_T = J$ and Plim $\hat{J}_T = J$. Hence: $\hat{J}_T'\Omega\tilde{J}_T \xrightarrow{P} J'\Omega J$ non-singular by assumption. Recall now that by Assumption 3.1(ii), $\Psi_T(\theta^0) = \sqrt{T}\bar{\phi}_T(\theta^0)$ converges to a normal distribution with mean 0. We then get the announced result. $\square$

Proof of Theorem 4.2 (*J-test*): A Taylor expansion of the moment conditions gives:

$$\begin{aligned}\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) &= \sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T}\frac{\partial\bar{\phi}_T(\hat{\theta}_T)}{\partial\theta'}\left(\hat{\theta}_T - \theta^0\right) + o_P(1)\\ &= \sqrt{T}\bar{\phi}_T(\theta^0) + \hat{J}_T\Lambda_T R^{-1}\left(\hat{\theta}_T - \theta^0\right) + o_P(1)\end{aligned}$$

with $\hat{J}_T = \sqrt{T}\partial\bar{\phi}_T(\hat{\theta}_T)/\partial\theta' R\Lambda_T^{-1}$. A Taylor expansion of the FOC gives:

$$\begin{aligned}\Lambda_T R^{-1}\left(\hat{\theta}_T - \theta^0\right) = &-\left[\left(\sqrt{T}\frac{\partial\bar{\phi}_T(\hat{\theta}_T)}{\partial\theta'}R\Lambda_T^{-1}\right)' S_T^{-1}\left(\sqrt{T}\frac{\partial\bar{\phi}_T(\hat{\theta}_T)}{\partial\theta'}R\Lambda_T^{-1}\right)\right]^{-1}\\ &\times\left(\sqrt{T}\frac{\partial\bar{\phi}_T(\hat{\theta})}{\partial\theta'}R\Lambda_T^{-1}\right)' S_T^{-1}\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1)\end{aligned}$$

with S_T a consistent estimator of the asymptotic covariance matrix of the process $\Psi(\theta)$. Combining the two above results leads to:

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) - \hat{J}_T\left[\hat{J}_T'S_T^{-1}\hat{J}_T\right]^{-1}\hat{J}_T'S_T^{-1}\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1).$$

Use the previous result to rewrite the criterion function:

$$
\begin{aligned}
T Q_T(\hat{\theta}_T) &= [\sqrt{T}\bar{\phi}_T(\hat{\theta}_T)]' S_T^{-1} \sqrt{T}\bar{\phi}_T(\hat{\theta}_T) \\
&= \left[\sqrt{T}\bar{\phi}_T(\theta^0) - \hat{J}_T \left[\hat{J}_T' S_T^{-1} \hat{J}_T\right]^{-1} \hat{J}_T' S_T^{-1} \sqrt{T}\bar{\phi}_T(\theta^0)\right]' S_T^{-1} \\
&\quad \times \left[\sqrt{T}\bar{\phi}_T(\theta^0) - \hat{J}_T \left[\hat{J}_T' S_T^{-1} \hat{J}_T\right]^{-1} \hat{J}_T' S_T^{-1} \sqrt{T}\bar{\phi}_T(\theta^0)\right] + o_P(1) \\
&= [\sqrt{T}\bar{\phi}_T(\theta^0)]' S_T^{-1} \sqrt{T}\bar{\phi}_T(\theta^0) - \sqrt{T}\bar{\phi}_T(\theta^0) S_T^{-1} \hat{J}_T \left[\hat{J}_T' S_T^{-1} \hat{J}_T\right]^{-1} \hat{J}_T' S_T^{-1} \sqrt{T}\bar{\phi}_T(\theta^0) \\
&\quad + o_P(1) = \sqrt{T}\bar{\phi}_T(\theta^0)' S_T'^{-1/2} [I - M]^{-1} S_T^{1/2} \sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1),
\end{aligned}
$$

where $S_T^{1/2}$ is such that $S_T = S_T'^{-1/2} S_T^{-1/2}$ and $M = S_T^{-1/2} \hat{J}_T [\hat{J}_T' S_T^{-1} \hat{J}_T]^{-1} \hat{J}_T' S_T'^{-1/2}$ which is a projection matrix, hence idempotent and of rank $(K - p)$. The expected result follows. □

Proof of Proposition 4.1 (*Equivalence between CU-GMM and efficient GMM*): The first-order conditions of the CU-GMM optimization problem can be written as follows (see Antoine et al., 2007):

$$
\sqrt{T} \frac{\partial \bar{\phi}_T'\left(\hat{\theta}_T^{CU}\right)}{\partial \theta} S_T^{-1}\left(\hat{\theta}_T^{CU}\right) \sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right) - P\sqrt{T} \frac{\partial \bar{\phi}_T'\left(\hat{\theta}_T^{CU}\right)}{\partial \theta} S_T^{-1}\left(\hat{\theta}_T^{CU}\right) \sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right) = 0,
$$

where P is the projection matrix onto the moment conditions. Recall that:

$$
P\sqrt{T} \frac{\partial \bar{\phi}_T^{(j)}\left(\hat{\theta}_T^{CU}\right)}{\partial \theta} = \mathrm{Cov}\left(\frac{\partial \bar{\phi}_T^{(j)}\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right) S_T^{-1}\left(\hat{\theta}_T^{CU}\right) \sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right).
$$

With a slight abuse of notation, we define conveniently the matrix of size (p, K^2) built by stacking horizontally the K matrices of size (p, K), $\mathrm{Cov}(\partial \bar{\phi}_{j,T}(\hat{\theta}_T^{CU})/\partial\theta, \bar{\phi}_T(\hat{\theta}_T^{CU}))$, as

$$
\begin{aligned}
\mathrm{Cov}\left(\frac{\partial \bar{\phi}_T'\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right) \equiv \Bigg[& \mathrm{Cov}\left(\frac{\partial \bar{\phi}_T^{(1)}\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right) \\
& \cdots \mathrm{Cov}\left(\frac{\partial \bar{\phi}_T^{(j)}\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right) \cdots \mathrm{Cov}\left(\frac{\partial \bar{\phi}_T^{(K)}\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right)\Bigg].
\end{aligned}
$$

Then, we can write:

$$
P\sqrt{T} \frac{\partial \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)}{\partial \theta} = \mathrm{Cov}\left(\frac{\partial \bar{\phi}_T'\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right)\left(Id_K \otimes \left[S_T^{-1}\left(\hat{\theta}_T^{CU}\right)\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right]\right).
$$

Pre-multiply the above FOC equations by the invertible (p, p)-matrix $\Lambda_T^{-1} R'$ to get:

$$
\begin{aligned}
& \Lambda_T^{-1} R' \sqrt{T} \frac{\partial \bar{\phi}_T'\left(\hat{\theta}_T^{CU}\right)}{\partial \theta} S_T^{-1}\left(\hat{\theta}_T^{CU}\right) \sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right) - \Lambda_T^{-1} R' \,\mathrm{Cov}\left(\frac{\partial \bar{\phi}_T'\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right) \\
& \times \left(Id_K \otimes \left[S_T^{-1}\left(\hat{\theta}_T^{CU}\right)\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right]\right) S_T^{-1}\left(\hat{\theta}_T^{CU}\right)\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right) = 0. \qquad \text{(A.3)}
\end{aligned}
$$

(i) Special case with the same weakness (λ_T) or all moment conditions: $\sqrt{T}\bar{\phi}_T(\hat{\theta}_T^{CU}) = \lambda_T \rho(\hat{\theta}_T^{CU}) + \Psi_T(\hat{\theta}_T^{CU})$, where $\Psi_T(\theta) \Rightarrow \Psi(\theta)$, a Gaussian stochastic process on Θ. Consider a Gaussian random variable

U to rewrite the second term of the LHS of equation (A.3) as follows:

$$R' \operatorname{Cov}\left(\frac{\partial \bar{\phi}'_T\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right)\left(Id_K \otimes \left[S_T^{-1}\left(\hat{\theta}_T^{CU}\right)\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right]\right) S_T^{-1}\left(\hat{\theta}_T^{CU}\right)$$
$$\times \left[\rho\left(\hat{\theta}_T^{CU}\right) - \frac{1}{\lambda_T}U\right] = o_P(1),$$

because we have: $(Id_K \otimes [S_T^{-1}(\hat{\theta}_T^{CU})\sqrt{T}\bar{\phi}_T(\hat{\theta}_T^{CU})]) = \mathcal{O}_P(1)$ and $[\rho(\hat{\theta}_T^{CU}) - \frac{1}{\lambda_T}U] = o_P(1)$. And, we conclude that CU-GMM is equivalent to efficient GMM.

(ii) Consider now two groups of moment conditions with rates $\sqrt{T}$ and λ_T:

$$\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right) = \begin{pmatrix} \sqrt{T}Id_{k_1} & 0 \\ 0 & \lambda_T Id_{k_2} \end{pmatrix} \rho\left(\hat{\theta}_T^{CU}\right) + \Psi_T\left(\hat{\theta}_T^{CU}\right).$$

The second term of the LHS of equation (A.3) can be separated into two pieces: the first one involves a Gaussian random variable V and the second one $\rho(.)$. First:

$$\Lambda_T^{-1} \underbrace{R' \operatorname{Cov}\left(\frac{\partial \bar{\phi}'_T\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right)\left(Id_K \otimes \left[S_T^{-1}\left(\hat{\theta}_T^{CU}\right)\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right]\right) S_T^{-1}\left(\hat{\theta}_T^{CU}\right)}_{\mathcal{O}_P(1)} V = o_P(1).$$

Second, recall that we have:

$$\begin{pmatrix} \sqrt{T}Id_{k_1} & 0 \\ 0 & \lambda_T Id_{k_2} \end{pmatrix} \rho\left(\hat{\theta}_T^{CU}\right) = \begin{pmatrix} \sqrt{T}Id_{k_1} & 0 \\ 0 & \lambda_T Id_{k_2} \end{pmatrix} \rho\left(\hat{\theta}_T^{CU}\right) = \begin{bmatrix} \sqrt{T}\rho_1\left(\hat{\theta}_T^{CU}\right) \\ \lambda_T\rho_2\left(\hat{\theta}_T^{CU}\right) \end{bmatrix}$$

and define the matrix

$$R' \operatorname{Cov}\left(\frac{\partial \bar{\phi}'_T\left(\hat{\theta}_T^{CU}\right)}{\partial \theta}, \bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right)\left(Id_K \otimes \left[S_T^{-1}\left(\hat{\theta}_T^{CU}\right)\sqrt{T}\bar{\phi}_T\left(\hat{\theta}_T^{CU}\right)\right]\right) S_T^{-1}\left(\hat{\theta}_T^{CU}\right)$$
$$\equiv A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where $A_{ij} = \mathcal{O}_P(1)$ for any $1 \leq i, j \leq 2$ with respective sizes: (s_1, k_1) for A_{11}; (s_1, k_2) for A_{12}; $(p - s_1, k_1)$ for A_{21}; and $(p - s_1, k_2)$ for A_{22}. Then the second term writes:

$$\Lambda_T^{-1}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{bmatrix} \sqrt{T}\rho_1\left(\hat{\theta}_T^{CU}\right) \\ \lambda_T\rho_2\left(\hat{\theta}_T^{CU}\right) \end{bmatrix} = \begin{bmatrix} A_{11}\rho_1\left(\hat{\theta}_T^{CU}\right) + \dfrac{\lambda_T}{\sqrt{T}}A_{12}\rho_2\left(\hat{\theta}_T^{CU}\right) \\ \dfrac{\sqrt{T}}{\lambda_T}A_{21}\rho_1\left(\hat{\theta}_T^{CU}\right) + A_{22}\rho_2\left(\hat{\theta}_T^{CU}\right) \end{bmatrix}.$$

And we have:

$$A_{11}\rho_1\left(\hat{\theta}_T^{CU}\right) = \mathcal{O}_P(1) \times \mathcal{O}_P\left(\frac{1}{\lambda_T}\right) = o_P(1)$$
$$\frac{\lambda_T}{\sqrt{T}}A_{12}\rho_2\left(\hat{\theta}_T^{CU}\right) = \frac{\lambda_T}{\sqrt{T}} \times \mathcal{O}_P(1) \times \mathcal{O}_P\left(\frac{1}{\lambda_T}\right) = o_P(1)$$
$$A_{22}\rho_2\left(\hat{\theta}_T^{CU}\right) = \mathcal{O}_P(1) \times \mathcal{O}_P\left(\frac{1}{\lambda_T}\right) = o_P(1)$$
$$\frac{\sqrt{T}}{\lambda_T}A_{21}\rho_1\left(\hat{\theta}_T^{CU}\right) = \frac{\sqrt{T}}{\lambda_T} \times \mathcal{O}_P(1) \times \mathcal{O}_P\left(\frac{1}{\lambda_T}\right) = \mathcal{O}_P\left(\frac{\sqrt{T}}{\lambda_T^2}\right).$$

Hence, to get the equivalence between CU-GMM and efficient GMM, we need the nearly-strong identification condition, that is: $\lambda_T^2/\sqrt{T} \stackrel{T}{\to} \infty$. □