MANCHESTER 1824

The University of Manchester

Economics
Discussion Paper Series
EDP-1705

The Asymptotic Properties of GMM and Indirect Inference Under Second-Order Identification

Prosper Dovonon

Alastair R. Hall

February 2017

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

The Asymptotic Properties of GMM and Indirect Inference under Second-order Identification[1]

Prosper Dovonon

Concordia University[2]

and

Alastair R. Hall

University of Manchester[3]

November 14, 2016

[1]The paper has benefited from the comments of the Guest Editors, two referees, and participants at the Conference in honour of Jean-Marie Dufour at Cireq in Montreal (May 2016), a seminar at the University of Nottingham and the Canadian Econometric Study Group meeting in London, Ontario, Canada (October 2016).

[2]Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada. E-mail:prosper.dovonon@concordia.ca.

[3]Corresponding author. Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. E-mail: alastair.hall@manchester.ac.uk.

Abstract

This paper presents a limiting distribution theory for GMM and Indirect Inference estimators when first-order identification fails but the parameters are second-order identified. These limit distributions are shown to be non-standard, but we show that they can be easily simulated, making it possible to perform inference about the parameters in this setting. We illustrate our results in the context of a dynamic panel data model in which the parameter of interest is identified locally at second order by non-linear moment restrictions but not at first order at a particular point in the parameter space. Our simulation results indicate that our theory leads to reliable inferences in moderate to large samples in the neighbourhood of this point of first-order identification failure. In contrast, inferences based on standard asymptotic theory (derived under the assumption of first-order local identification) are very misleading in the neighbourhood of the point of first-order local identification failure.

Keywords: Moment-based estimation, First-order identification failure, Minimum-chi squared estimation, Simulation-based estimation

1 Introduction

Generalized Method of Moments (GMM) was introduced by Lars Hansen in a paper published in *Econometrica* in 1982. Since then this article has come to be recognized as one of the most influential papers in econometrics.[1] One aspect of this influence is that applications of GMM have demonstrated the power of thinking in terms of moment conditions in econometric estimation. This, in turn, can be said to have inspired the development of other moment-based approaches in econometrics, a leading example of which is Indirect Inference (II). GMM can be applied in wide variety of situations including those where the distribution of the data is unknown and those where it is known but the likelihood is intractable. In the latter scenario, it was realized in the late 1980's and early 1990's that simulation-based methods provide an alternative - and often more efficient way - to estimate the model parameters than GMM. A number of methods were proposed: Method of Simulated Moments (McFadden, 1989), Simulated Method of Moments (SMM, Duffie and Singleton, 1993), Indirect Inference (II, Gourieroux, Monfort, and Renault, 1993, Smith, 1990, 1993)[2] and Efficient Method of Moments (EMM, Gallant and Tauchen, 1996). While SMM and EMM have their distinctive elements, both can be viewed as examples of II as they have the "indirect" feature of estimating parameters of the model of interest by matching moments from a different - and often misspecified - model.

The standard first-order inference frameworks for Generalized Method of Moments (GMM) and Indirect Inference (II) rest crucially on the assumption of first-order local identification that is, a certain derivative matrix has full rank when evaluated at the true parameter value. However, it has been realized that in a number of situations first-order identification either fails or is close to failing with the result that inferences based on the standard framework are misleading. To date, this concern and its consequences have largely been explored in the context of GMM, but recently concerns about identification have been raised in dynamic stochastic general equilibrium (DSGE) models to which GMM and II have been applied.[3]

Within the GMM framework, these concerns about the consequences of identification have mostly arisen in the special case of Generalized Instrumental Variables (GIV) estimation (Hansen and Singleton, 1982) in which the moment condition derives form the orthogonality of a function $u_t(\theta)$, involving the parameter vector θ, to a vector of instruments, z_t. In this case, the condition for first-order local identification is that $\partial u_t(\theta)/\partial\theta$ (evaluated at the true parameter value, θ_0) has a sufficiently strong relationship to z_t in the population. However, if this threshold is only marginally satisfied then the standard first-order asymptotic theory can provide a very poor approximation to the finite sample behaviour of various GMM-based statistics. To help derive more accurate approximations, Staiger and Stock (1997) introduced the concept of weak identification. Statistical analyses demonstrated that key statistics behave very differently under weak identification than under the standard first-order asymptotic framework with its assumption of first-order local identification.[4] For example, Dufour (1997) demonstrated that the potential presence of weak identification renders the conventional "estimator plus/minus a multiple of the standard error" confidence interval invalid. In response, the focus shifted to developing inference techniques that

[1] For example, see The Royal Swedish Academy of Sciences (2013), p.24.

[2] Smith (1993) refers to the method as "simulated quasi-maximum likelihood" and his analysis covers a more restrictive setting than that of Gourieroux, Monfort, and Renault (1993).

[3] Applications of GMM/II to DSGE include Christiano, Eichenbaum, and Evans (2005), Coenen, Levin, and Christoffel (2007), Dupaigne, Fève, and Matheron (2007), Ruge-Murcia (2007), Le and *et al* (2011).

[4] For example, Staiger and Stock (1997) and Stock and Wright (2000) derive the properties of various estimators such as GMM in linear and nonlinear models respectively.

are valid irrespective of the quality of the identification, such as Kleibergen's (2005) K-statistic. For our purposes here, it is not necessary to summarize subsequent developments within the weak identification framework; it suffices to note that weak identification involves a situation in which *both* first-order local identification *and* global identification fail (in the limit).[5]

Canova and Sala (2009) argue that, at their time of writing, the quality of the identification in DSGE models was often neglected, and also that there are grounds for suspecting identification may fail in certain cases of interest. Iskrev (2010), Komunjer and Ng (2011) and Qu and Tkachenko (2012) derive conditions for first-order local identification using alternative representations of the model. In this context, the responses to potential identification failure have been twofold. The first approach is the same as in the GMM literature and is based on developing inference techniques that are robust to weak identification, for example see Dufour, Khalaf, and Kichian (2013) and Qu (2014). The second approach views the source of identification failure as deriving from the method used to solve the DSGE for the path of the variables. DSGE models are typically highly nonlinear, and so as a result practitioners have resorted to using approximations in solving the models. For the most part, first-order approximations have been used but Mutschler (2015) has recently demonstrated that these may be the source of identification failures, finding that the use of second-order approximations restores first-order local identification in some cases.

As is evident from the above discussion, the focus of the above analyses is on first-order local identification - understandably, as this condition is crucial for the standard first-order asymptotic framework. In linear models, first-order local and global identification are the same, but in non-linear models, they are not: local identification can fail at first order but hold at a higher order. Furthermore, in such cases, it is possible to develop a framework for inference based on large sample arguments. For the case where local identification holds at second but not first order, Sargan (1983) and Rotnitzky, Cox, Bottai, and Robins (2000) develop a limiting distribution theory for estimators obtained respectively by IV in a nonlinear in parameters model and Maximum Likelihood (ML). Dovonon and Renault (2009, 2013) derive the limiting distribution of the GMM overidentifying restrictions test statistic. This pattern of identification has been shown to arise in a number of situations in statistics and econometrics such as: ML for skew-normal distributions, *e.g.* Azzalini (2005); ML for binary response models based on skew-normal distributions, Stingo, Stanghellini, and Capobianco (2011); ML for missing not at random (MNAR) models, *e.g.* Jansen and *et al* (2006); GMM estimation of conditional heteroscedastic factor models, Dovonon and Renault (2009, 2013); GMM estimation of panel data models using second moments, Madsen (2009); ML estimation of panel data models, Kruiniger (2014).

In this paper, we consider the case where local identification fails at first order but holds at second order. Although this situation has been recognized to arise in models of interest, there are no general results available on either GMM or II estimators in this case. In this paper, we fill this gap. We present the limiting distribution of (i) the GMM estimator; (ii) the II estimator in cases where the auxiliary model is second-order but not first-order identified. These limit distributions are shown to be non-standard but easily simulated, making it possible to perform inference about the parameters in this setting. Our results for GMM cover all the cases cited in the previous paragraph, and our results for II cover cases in which any of the models cited in the previous paragraph are used as auxiliary models.[6] We conjecture our results may also be relevant to estimation of certain DSGE

[5]Subsequent developments include the introduction of asymptotics based on either nearly-weak identification or many moments; for a recent review of this literature see Hall (2015).

[6]Gourieroux, Phillips, and Yu (2010) suggest using II to bias correct ML. In this case, the auxiliary model is the ML estimator from the sample and is based on the same distributional assumption as the the simulator.

models by GMM or II, an issue to which we return at the end of the paper. We examine the accuracy of our distribution theory as an approximation to finite sample behaviour in a small simulation study involving a dynamic panel data model in which the parameter of interest is identified locally at second order by a set of non-linear moment restrictions but not at first order at a particular point in the parameter space. Our simulation results indicate that the limiting distribution theory derived in our paper leads to reliable GMM/II-based inferences in moderate to large samples in the neighbourhood of this point of first-order identification failure. In contrast, inferences based on standard asymptotic theory (derived under the assumption of first-order local identification) are very misleading in this neighbourhood. Comparing GMM and II, we find our limiting distribution theory provides a reasonable approximation to the behaviour of the GMM at smaller sample sizes than it does for the II estimator, but that II exhibits smaller bias at the point of first-order local identification failure.

An outline of the paper is as follows. Section 2 briefly reviews GMM and II estimation and their inference frameworks under first-order local identification. Section 3 defines second-order identification and provides two examples. Sections 4 and 5 present the limiting distribution for GMM and II estimators respectively. Section 6 reports the results from the simulation study, and Section 7 offers some concluding remarks. All proofs are relegated to an Appendix.

2 Identification and the first-order asymptotics of GMM and II

In this section, we briefly review the basic GMM and II inference frameworks based on first-order asymptotics, paying especial attention to the role of first-order local identification. Since both methods can be viewed as special cases of "minimum chi-squared", we use the latter to unify our presentation. Therefore, we begin by defining the GMM and II estimators, and then present the minimum chi-squared framework. To this end, we introduce the following notation. In each case the model involves random vector X which is assumed strictly stationary with distribution $P(\theta_0)$ that is indexed by a parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^p$. For some of the discussion only a subset of the parameters may be of primary interest, and so we write $\theta = (\phi', \psi')'$ where $\phi \in \Phi \subset \mathbb{R}^{p_\phi}$ and $\psi \in \Psi \subset \mathbb{R}^{p_\psi}$. Throughout, W_T denotes a positive semi-definite matrix with the dimension defined implicitly by the context.

GMM:
GMM is a partial information method in the sense that its implementation does not require knowledge of $P(\cdot)$ but only a population moment condition implied by this underlying distribution. In view of this, we suppose that ϕ_0 is of primary interest and the model implies:[7]

$$E[g(X, \phi_0)] = 0, \tag{1}$$

where $g(\cdot)$ is a $q \times 1$ vector of continuous functions. The GMM estimator of ϕ_0 based on (1) is defined as:

$$\hat{\phi}_{GMM} \;=\; argmin_{\phi\in\Phi}\, Q_T^{GMM}(\phi), \tag{2}$$

[7] If $p_\psi = 0$ then $\phi = \theta$ and our presentation covers the case when the entire parameter vector is being estimated.

where

$$Q_T^{GMM}(\phi) \; = \; T^{-1} \sum_{t=1}^{T} g(x_t, \phi)' W_T T^{-1} \sum_{t=1}^{T} g(x_t, \phi), \tag{3}$$

and $\{x_t\}_{t=1}^T$ represents the sample observations on X.

As evident from the above, GMM estimation is based on the information that the population moment $E[g(X, \phi)]$ is zero when evaluated at $\phi = \phi_0$. The form of this moment condition depends on the application: in economic models that fit within the framework of discrete dynamic programming models then the moment condition often takes the form of Euler equation times a vector of instruments;[8] in model estimated via quasi-maximum likelihood then the moment condition is the quasi-score.[9]

II:

II is essentially a full information method in the sense it provides a method of estimation of θ_0 given knowledge of $P(\,\cdot\,)$. Within II, there are two models: the "simulator" which represents the model of interest - $X \sim P(\theta)$ in our notation - and an "auxiliary model" that is introduced solely as the basis for estimation of the parameters of the simulator. Although θ_0 is unknown, data can be simulated from the simulator for any given θ. To implement II, this simulation needs to be performed a number of times, s say, and we denote these simulated series by $\{x_t^{(i)}(\theta)\}_{t=1}^T$ for $i = 1, 2, \ldots s$. The auxiliary model is estimated from the data; let $h_T = h\left(\{x_t\}_{t=1}^T\right)$ be some feature of this model, and $h_T^{(i)}(\theta) = h\left(\{x_t^{(i)}(\theta)\}_{t=1}^T\right)$. Assume $dim(h_T) = \ell > p$. The II estimator of θ_0 is:[10]

$$\hat{\theta}_{II} \; = \; argmin Q_T^{II}(\theta), \tag{4}$$

where

$$Q_T^{II}(\theta) \; = \; \left[h_T \; - \; \frac{1}{s} \sum_{i=1}^{s} h_T^{(i)}(\theta) \right]' W_T \left[h_T \; - \; \frac{1}{s} \sum_{i=1}^{s} h_T^{(i)}(\theta) \right]. \tag{5}$$

To characterize the population analog of the information being exploited here, we assume that $h_T \xrightarrow{P} h_*$, for some constant h_*. Noting that there exists a mapping from θ_0 to $h(\,\cdot\,)$ through $x_t(\theta_0)$, we can write $h_* = b(\theta_0)$ for some $b(\,\cdot\,)$, known as the binding function. Then, as Gourieroux, Monfort, and Renault (1993) observe, II exploits the information that $k(h_*, \theta_0) = h_* - b(\theta_0) = 0$ - in essence that, at the true parameter value, the simulator encompasses the auxiliary model.

The choice of $h(\,\cdot\,)$ varies, in practice, and depends on the setting. Examples include: raw data moments, such as the first two moments of macroeconomic or asset series, *e.g.* see Heaton (1995); the estimator or score vector from an auxiliary model that is in some way closely related to the simulator,[11] *e.g.* Gallant and Tauchen (1996), Garcia, Renault, and Veredas (2011); estimated moments or parameters from the auxiliary model, such as in DSGE models, *e.g.* see the references in footnote 3.

[8] For example, the consumption based asset pricing model in the seminal article by Hansen and Singleton (1982).

[9] For example, see Hamilton (1994)[p.428-9].

[10] We note that II as defined in (4)-(5) is one version of the estimator. An alternative version involves simulating a single series of length ST. For scenarios involving optimization in the auxiliary model, this second approach has the advantage of requiring only one optimization. The first-order asymptotic properties of the II estimator are the same either way; see Gourieroux, Monfort, and Renault (1993).

[11] For the first-order asymptotic equivalence of these two approaches, see Gourieroux, Monfort, and Renault (1993).

Minimum chi-squared:
As is apparent from the above definitions, both GMM and II estimation involve minimizing a quadratic form in the sample analogs to the population information about θ_0 on which they are based namely, $E[g(X, \phi_0)] = 0$ for GMM and $k(h_*, \theta_0) = 0$ for II. As such they can both be viewed as fitting within the class of minimum chi-squared. This common structure explains many of the parallels in their first-order asymptotic structure, and is also useful for highlighting the role of various identification conditions in the analyses.

Minimum chi-squared estimation is first introduced by Neyman and Pearson (1928) in the context of a specific model, but their insight is applied in more general models by Neyman (1949), Barankin and Gurland (1951) and Ferguson (1958). Suppose again that ϕ_0 is of primary interest, recalling that $p_\psi = 0$ implies $\phi = \theta$, and let $\tilde{m}_T(\phi)$ be a $n \times 1$ vector, where $n \geq p_\phi$, satisfying

Assumption 1. *$\tilde{m}_T(\phi_0) \xrightarrow{d} N(0, V_m)$, where V_m, a positive definite matrix of finite constants.*

As a result, $\tilde{m}_T(\phi_0)' V_m^{-1} \tilde{m}_T(\phi_0) \xrightarrow{d} \chi^2_n$, and this structure explains the designation of the following estimator as a minimum chi-squared:

$$argmin_{\phi \in \Phi}\, \tilde{m}_T(\phi)' \hat{V}_m^{-1} \tilde{m}_T(\phi), \tag{6}$$

where $\hat{V}_m \xrightarrow{p} V_m$.

However, for our purposes here, it is convenient to begin with the more general definition of minimum chi-squared estimator:[12]

$$\hat{\phi}_{MC} \;=\; argmin_{\phi \in \Phi}\, Q_T(\phi), \tag{7}$$

where

$$Q_T(\phi) \;=\; m_T(\phi)' W_T m_T(\phi), \tag{8}$$

and $m_T(\phi) = T^{-1/2} \tilde{m}_T(\phi)$.

To consider the first-order asymptotic properties of minimum chi-squared estimators, we introduce a number of high level assumptions.

Assumption 2. *(i) $W_T \xrightarrow{p} W$, a positive definite matrix of constants; (ii) $Q_T(\phi) \xrightarrow{p} Q(\phi) = m(\phi)' W m(\phi)$ uniformly in ϕ; (iii) $Q(\phi_0) < Q(\phi)$ $\forall\, \phi \neq \phi_0$, $\phi \in \Phi$.*

Assumption 2(iii) serves as a global identification condition. These conditions are sufficient to establish consistency; for example see Newey and McFadden (1994).

Proposition 1. *If Assumption 2 holds then $\hat{\phi}_{MC} \xrightarrow{p} \phi_0$.*

The first-order conditions of the minimization in (8) are:

$$M_T(\hat{\phi}_{MC})' W_T m_T(\hat{\phi}_{MC}) \;=\; 0, \tag{9}$$

where $M_T(\phi) = \partial m_T(\phi)/\partial \phi'$, a matrix commonly referred to as the Jacobian in this context. These conditions are the source for the standard first-order asymptotic distribution theory of the estimator, but the latter requires the Jacobian to satisfy certain restrictions. To present these conditions, define $N_\epsilon = \{\phi;\, \|\phi - \phi_0\| < \epsilon\}$.

[12] See Ferguson (1958).

Assumption 3. *(i)* $M_T(\phi) \xrightarrow{p} M(\phi)$ *uniformly in* N_ϵ*; (ii)* $M(\phi)$ *is continuous on* N_ϵ*; (iii)* $M(\phi_0)$ *is rank* p_ϕ*.*

Assumption 3(iii) is the condition for first-order local identification. It is sufficient but not necessary for local identification of θ_0 on N_ϵ, but it is necessary for the development of the standard first-order asymptotic theory. Under Assumptions 1-3, the Mean Value Theorem applied to (9) yields:

$$T^{1/2}(\hat{\phi}_{MC} - \phi_0) \simeq -\{M(\phi_0)'WM(\phi_0)\}^{-1} M(\phi_0)'W\tilde{m}_T(\phi_0),$$

where $\simeq$ denotes equality up to terms of $o_p(1)$, from which the first-order asymptotic distribution follows.

Proposition 2. *If Assumptions 1-3 hold then:*

$$T^{1/2}(\hat{\phi}_{MC} - \phi_0) \xrightarrow{d} N(0, V_\phi),$$

where

$$V_\phi = [M(\phi_0)'WM(\phi_0)]^{-1}M(\phi_0)'WV_mWM(\phi_0)[M(\phi_0)'WM(\phi_0)]^{-1}.$$

As apparent, V_ϕ depends on W. The choice of W that minimizes V_ϕ is $W = V_m^{-1}$ which yields: $V_\phi = \{M(\phi_0)'V_m^{-1}M(\phi_0)\}^{-1}$.[13] This efficiency bound can be achieved in practice by setting $W_T = \hat{V}_m^{-1}$ where $\hat{V}_m \xrightarrow{P} V_m$ to produce the version of the estimator in (6).

Identification:
Hansen (1982) provides general conditions under which the first-order asymptotic framework above goes through for GMM with

$$m_T(\phi) = T^{-1}\sum_{t=1}^{T} g(x_t, \phi).$$

Gourieroux, Monfort, and Renault (1993) prove the same results for II with[14]

$$m_T(\theta) = h_T - \frac{1}{s}\sum_{i=1}^{s} h_T^{(i)}(\theta).$$

We now turn to the nature and role of the identification conditions in the above analysis.[15] Global identification is crucial for consistency. Given Assumption 2(i), the global identification condition for GMM can be equivalently stated as $E[g(X, \theta)] = 0$ has a unique solution at $\theta = \theta_0$; likewise for II, the global identification condition is that $k(h_*, \theta) = 0$ has a unique solution at $\theta = \theta_0$. However, global identification is not sufficient for the asymptotic distribution theory in Proposition 2. For the latter, first-order local identification is needed: for GMM, the condition that $E[\partial g(X, \phi)/\partial\phi'|_{\phi=\phi_0}]$ is full column rank; for II, that $E[\partial h_T(\theta)/\partial\theta'|_{\theta=\theta_0}]$ is full column rank.

[13] This result can be established via linear algebraic arguments in Hansen (1982)[Theorem 3.2].

[14] In spite of the similarities of the two methods, the asymptotic properties of II cannot be deduced directly from the corresponding GMM analysis because the simulation-based implementation takes II outside the GMM framework; see *inter alia* Duffie and Singleton (1993) or Ghysels and Guay (2003, 2004).

[15] The minimum chi-squared structure can also be used to explain other common features of GMM and II, see Dovonon and Hall (2015).

In linear models, global and first-order local identification are equivalent. However in nonlinear models, global identification is possible without first-order local identification because local identification can be ensured by higher order derivatives of $m(\phi)$. Under such a scenario, the parameters can be consistently estimated but the standard first-order asymptotic framework described above is not valid. For the rest of this paper, we focus on the case in which the parameters are globally identified but local identification is only second-order. In the next section, we formally define second-order local identification and provide two examples of econometric models in which it occurs; Sections 4 and 5 characterize the limiting behaviour of GMM and II estimators within this framework.

3 Second-order local identification

For our analysis of GMM and II, we adopt the definition of second-order local identification originally introduced by Dovonon and Renault (2009). To present this definition, we introduce the following notations:

Let $m(\phi) = E(g(X,\phi))$ and

$$M_k^{(2)}(\phi_0) \;=\; E\left[\left.\frac{\partial^2 g_k(X,\phi)}{\partial\phi\partial\phi'}\right|_{\phi=\phi_0}\right], \qquad k = 1,2\ldots,q$$

where $g_k(X,\phi)$ is the k^{th} element of $g(X,\phi)$ and $g(\cdot)$ is defined in (1). Second-order local identification is defined as follows.

Definition 1. *The moment condition $m(\phi) = 0$ locally identifies $\phi_0 \in \Phi$ up to the second order if:*

(a) $m(\phi_0) = 0$.

(b) For all u in the range of $M(\phi_0)'$ and all v in the null space of $M(\phi_0)$, we have:

$$\left(M(\phi_0)u + \left(v' M_k^{(2)}(\phi_0)v\right)_{1\leq k\leq q} = 0\right) \Rightarrow (u = v = 0).$$

This condition is derived using a second order expansion of $m(\phi)$ around $m(\phi_0)$ and can be motivated as follows. For any non-zero $\phi - \phi_0 \in N_\epsilon$, we have $\phi - \phi_0 = c_1 u + c_2 v$ where c_1, c_2 are constants such that $c_1 \neq 0$ and/or $c_2 \neq 0$.[16] For those directions for which c_1 is non-zero then the first order term is non-zero and dominates, and for those directions in which $c_1 = 0$, then the second order term is non-zero. Thus, without requiring that the Jacobian matrix $M(\phi_0)$ to have full rank, conditions (a) and (b) in Definition 1 guarantee local identification in the sense that there is no sequence of points $\{\phi_n\}$ different from ϕ_0 but converging to ϕ_0 such that $m(\phi_n) = 0$ for all n. The difference between first-order local identification and second-order local identification (with $M(\phi_0)$ rank deficient) is how sharply $m(\phi)$ moves away from 0 in the neighborhood of ϕ_0. We now consider two examples in which local identification fails at first order but holds at second order.

[16] Recall the range of $M(\phi_0)'$ is the orthogonal complement of the null space of $M(\phi_0)$.

Example 1. ***Nonstationary panel AR(1) model with individual fixed effects.*** *Consider the standard AR(1) panel data model with individual specific effects,*

$$y_{it} = \rho y_{i,t-1} + \eta_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, 2. \tag{10}$$

Assume that the vector $(y_{i0}, \eta_i, \varepsilon_{i1}, \varepsilon_{i2})$ *is i.i.d. across* i *with mean 0 and that* $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$, $E(\varepsilon_{is}\varepsilon_{it}) = 0$ *for* $s \neq t$, $s, t = 1, 2$, $E(\eta_i^2) = \sigma_\eta^2$, $E(y_{i0}^2) = \sigma_0^2$, $E(\varepsilon_{it}\eta_i) = 0$, $E(\varepsilon_{it}y_{i0}) = 0$, $t = 1, 2$, *and* $E(y_{i0}\eta_i) = \sigma_{0\eta}$. *For this example,* $\theta = (\rho, \sigma_0^2, \sigma_\eta^2, \sigma_{0\eta}, \sigma_\varepsilon^2)'$. *Our primary focus here is on estimation of* ρ *and so we partition the parameter vector as follows:* $\theta = (\rho, \theta_2)'$.

For $|\rho| < 1$, *this model can be estimated via GMM using the moment conditions in Arellano and Bond (1991). However as pointed out by Blundell and Bond (1998), the Arellano-Bond (AB) moments only provide weak identification of* θ *as* ρ *tends to one. Blundell and Bond (1998) propose augmenting the AB moments with an additional set of moments to produce the so-called "System GMM estimator": this approach solves the weak identification problems for* ρ *less than one but is not valid for* $\rho = 1$ *because the approach exploits properties of the series that only hold for* $|\rho| < 1$. *Quasi Maximum Likelihood estimation of the model has been studied by Kruiniger (2013) for* $-1 < \rho < 1$.

An alternative solution to the identification problems with AB moments is to base estimation on higher moments. Expressing the variance of $y_i = (y_{i0}, y_{i1}, y_{i2})'$ *as a function of the model parameters,* θ *can be identified by the moment condition restriction:*

$$E\left[g(y_i) - H(\rho)\theta_2\right] = 0, \tag{11}$$

where $g(\cdot) = [g_1(\cdot)', g_2(\cdot)']'$, $H(\cdot) = [H_1(\cdot)', H_2(\cdot)']'$,

$$g_1(y_i) = \begin{pmatrix} y_{i0}y_{i1} \\ y_{i0}y_{i2} \end{pmatrix}, \quad g_2(y_i) = \begin{pmatrix} y_{i0}^2 \\ y_{i1}^2 \\ y_{i1}y_{i2} \\ y_{i2}^2 \end{pmatrix},$$

$$H_1(\rho) = \begin{pmatrix} \rho & 0 & 1 & 0 \\ \rho^2 & 0 & 1+\rho & 0 \end{pmatrix}, \quad H_2(\rho) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \rho^2 & 1 & 2\rho & 1 \\ \rho^3 & 1+\rho & \rho(1+2\rho) & \rho \\ \rho^4 & (1+\rho)^2 & 2\rho^2(1+\rho) & 1+\rho^2 \end{pmatrix}.$$

Note that $H_2(\rho)$ *is nonsingular for all* $\rho \neq 0$ *and we have:*

$$\theta_2(\rho) = H_2^{-1}(\rho)E[g_2(y_i)].$$

Using (11) *involving* $g_1(\cdot)$, *we can therefore consider the moment condition:*

$$E\left[g_1(y_i) - H_1(\rho)(H_2(\rho))^{-1}g_2(y_i)\right] = 0 \tag{12}$$

for inference about ρ, *our main parameter of interest. The true parameter value that we consider for the data generating process is* $\theta^* = (1, \theta_2^*)$, *with* $\theta_2^* = (\sigma_0^{*^2}, 0, 0, \sigma_\varepsilon^{*^2})'$. *In the appendix, it is shown that* (12) *globally identifies* ρ *but local identification fails at first order but holds at second order.*

While the above discussion has concentrated on GMM, we note that II methods have also been proposed for dynamic panel data models. Gourieroux, Phillips, and Yu (2010) propose an II estimator in which the function h_T is the MLE under normality. They note that the II approach can be based on nonlinear moments and, following their suggestion, II can be applied using the moments in (11) as the auxiliary model. In Section 6, we report simulation results that compare GMM based on (11) with II using (11) as the auxiliary model.

Example 2. ***A conditionally heteroscedastic factor model***
Consider the conditionally heteroskedastic factor (CHF) model of two asset returns:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} f_t + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \tag{13}$$

with

$$\begin{gathered} E\left[(f_t, u_t')'|\mathfrak{F}_{t-1}\right] = 0,\ Var\left[f_t|\mathfrak{F}_{t-1}\right] = \sigma_{t-1}^2, \\ Var[(u_{1t}, u_{2t})'|\mathfrak{F}_{t-1}] = Diag(\Omega_1, \Omega_2),\ Cov[f_t, u_t|\mathfrak{F}_{t-1}] = 0. \end{gathered} \tag{14}$$

In this model, f_t is the latent common GARCH factor, u_t is the vector of idiosyncratic shocks and σ_{t-1}^2 is the time varying conditional variance of f_t where the conditioning set $\mathfrak{F}_t$ is an increasing filtration containing current and past values of f_t and y_t. In addition to this specification, it is assumed that $\gamma_1 \neq 0$ and $\gamma_2 \neq 0$, meaning that the two asset return processes are conditionally heteroskedastic. Conditions for the identification of the factor structure[17] (13)-(14) *can be found in Doz and Renault (2004). The parameter vector of interest is $\theta \equiv (\gamma_1, \gamma_2, \Omega_1, \Omega_2)'$.*

This model has been introduced by Diebold and Nerlove (1989) and further studied by Fiorentini, Sentana, and Shephard (2004) and Doz and Renault (2006). Fiorentini, Sentana, and Shephard (2004) impose additional structure on the model and propose a Kalman filter approach to estimation. Doz and Renault (2006) propose a GMM approach based on moment conditions that identify the parameters up to one (say, γ_1) that is given a 'reasonable' value. This partial identification is the cost of allowing $Var[(u_{1t}, u_{2t})'|\mathfrak{F}_{t-1}]$ to be non-diagonal. Here, we consider an II estimator for the model. The simulator is (13)-(14) and the assumption that $(f_t, u_t')'$ is conditionally normally distributed.

The auxiliary model is defined as:

$$\begin{aligned} E\left[\begin{pmatrix} 1 \\ z_{t-1} \end{pmatrix}\left\{(y_{1t} - \delta y_{2t}\right\}^2 - c\right] &= 0 \\ E[y_{1t}^2] &= b_1 \\ E[y_{2t}^2] &= b_2 \\ E[y_{1t}y_{2t}] &= b_3. \end{aligned} \tag{15}$$

where $z_{t-1} \in \mathfrak{F}_{t-1}$, (e.g lagged square returns), $\delta = \gamma_1/\gamma_2$, $b_1 = \gamma_1^2 + \Omega_1$, $b_2 = \gamma_2^2 + \Omega_2$, $b_3 = \gamma_1\gamma_2$, $c = \Omega_1 + \delta^2\Omega_2$, and $c = b_1 + \delta^2 b_2 - 2\delta b_3$. The parameter vector in the auxiliary model,

[17]The conditionally heteroskedastic factor representation (13)-(14) for y_t is uniquely determined if we restrict to decompositions such that the factor has unit variance and non degenerate conditional variance with no positive lower bound, that is: $E(\sigma_t^2) = 1$, $Var(\sigma_t^2) > 0$ and $P(\sigma_t^2 > \sigma^2) < 1$ for all $\sigma^2 > 0$. In this case, Ω is uniquely determined and $\gamma = (\gamma_1, \gamma_2)'$ is identified up to the sign. Restricting $\gamma_1 > 0$ completes the identification of the factor structure. We refer to Doz and Renault (2004) for a more detailed discussion on the identification of CHF models.

$h = (b_1, b_2, b_3, \delta, c)'$, is globally identified. In addition, the parameter θ of the structural model can be determined from h (enforcing $\gamma_1 > 0$)as follows:

$$\theta_1 \equiv \gamma_1 = \sqrt{\delta b_3}, \quad \theta_2 \equiv \gamma_2 = \sqrt{\frac{b_3}{\delta}}, \quad \theta_3 \equiv \Omega_1 = b_1 - \delta b_3, \quad \theta_4 \equiv \Omega_2 = b_2 - \frac{b_3}{\delta}.$$

However, as shown in the appendix, h is not locally identified at first order but is at second order.

4 The limiting distribution of the GMM estimator

In this section, we consider the moment condition model (1) and study the asymptotic behaviour of the GMM estimator when ϕ_0 is second-order locally identified because the moment condition exhibits the properties in Definition 1 but the standard local identification condition (Assumption 3(iii)) fails.[18]

We study the asymptotic behaviour of the GMM estimator by restricting ourselves to the case of one-dimension rank deficiency, i.e. rank of $M(\phi_0)$ is equal to $p_\phi - 1$, since this seems to be the only case that is analytically tractable. We start by considering the case where the rank deficiency is due to the last column of the Jacobian being null. To this end, we partition ϕ into $(\phi'_{1:p_\phi-1}, \phi_{p_\phi})'$ where $\phi_{1:p_\phi-1}$ is the vector consisting of the first $p_\phi - 1$ elements of ϕ and ϕ_{p_ϕ} is the ${p_\phi}^{th}$ element of ϕ. For ease of presentation below, we shorten the subscript and write ϕ_1 for $\phi_{1:p_\phi-1}$. Thus $\phi_0 = (\phi'_{0,1}, \phi_{0,p_\phi})'$ where $\phi_{0,1}$ is a $(p_\phi - 1) \times 1$ vector containing the true values of $\phi_{1:p_\phi-1}$ and ϕ_{0,p_ϕ} is the true value of ϕ_{p_ϕ}. If $M(\phi_0)$ has rank $p_\phi - 1$ with $\frac{\partial m}{\partial \phi_{p_\phi}}(\phi_0) = 0$, second-order identification is equivalent to:

$$Rank \begin{pmatrix} \frac{\partial m}{\partial \phi'_1}(\phi_0) & \frac{\partial^2 m}{\partial \phi^2_{p_\phi}}(\phi_0) \end{pmatrix} = p_\phi.$$

This is the setting studied by Sargan (1983) for the instrumental variables estimator in nonlinear in parameters model.

Letting $D = \frac{\partial m}{\partial \phi'_1}(\phi_0)$ and $G = \frac{\partial^2 m}{\partial \phi^2_{p_\phi}}(\phi_0)$, we next derive the asymptotic distribution of the GMM estimator under the following condition.

Assumption 4. *(i)* $\frac{\partial m}{\partial \phi_{p_\phi}}(\phi_0) = 0$.

(ii) $Rank\,(D \;\; G) = p_\phi$.

We also require the following stronger assumption than Assumptions 1 and 3 in Section 2:

Assumption 5. *(i) $m_T(\phi)$ has partial derivatives up to order 3 in a neighborhood N_ϵ of ϕ_0 and the derivatives of $m_T(\phi)$ converge in probability uniformly over N_ϵ to those of $m(\phi)$.*

(ii) $\sqrt{T} \begin{pmatrix} m_T(\phi_0) \\ \frac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbb{Z}_0 \\ \mathbb{Z}_1 \end{pmatrix}$.

[18]Lee and Liao (2016) and Sentana (2015) propose an alternative approach using GMM estimation based on an augmented set of moment conditions that involve both the original moments and restrictions on their derivatives. While this approach allows the application of standard first-order asymptotic theory, the augmentation substantially increases the number of moment conditions used in the estimation which tends to lead to a deterioration in the quality of the first-order asymptotic theory as an approximation to finite sample behaviour; for example see Newey and Smith (2004).

(iii) $W_T - W = o_P(T^{-1/4})$, $\frac{\partial m_T}{\partial \phi_1'}(\phi_0) - D = O_P(T^{-1/2})$,
$\frac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_0) - G = O_P(T^{-1/2})$, $\frac{\partial^2 m_T}{\partial \phi_1' \partial \phi_{p_\phi}}(\phi_0) - G_{1p_\phi} = o_P(1)$, *and* $\frac{\partial^3 m_T}{\partial \phi_{p_\phi}^3}(\phi_0) - L = o_P(1)$,
with $G_{1p_\phi} = \frac{\partial^2 m}{\partial \phi_1' \partial \phi_{p_\phi}}(\phi_0)$ *and* $L = \frac{\partial^3 m}{\partial \phi_{p_\phi}^3}(\phi_0)$.

Assumption 5 is useful to derive the asymptotic distribution of the GMM estimator under the second-order local identification setting of Assumption 4. These conditions are slightly stronger than the standard ones. The derivation of the asymptotic distribution of the GMM estimator requires a mean value expansion of $m_T(\phi)$ up to the third order and the uniform convergence guaranteed by Assumption 5(i) is, in particular, useful to control the remainder of our expansions. Assumption 5(ii) gives the joint asymptotic distribution of $m_T(\phi_0)$ and $\frac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0)$. Under mild assumptions on $g(x, \phi_0)$ and $\frac{\partial g}{\partial \phi_{p_\phi}}(x, \phi_0)$, both having zero mean, the central limit theorem guarantees that $\begin{pmatrix} \mathbb{Z}_0 \\ \mathbb{Z}_1 \end{pmatrix} \sim N(0, v)$, with $v = \lim_{T\to\infty} Var\sqrt{T} \begin{pmatrix} m_T(\phi_0) \\ \frac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0) \end{pmatrix}$.

Assumption 5(iii) imposes the asymptotic order of magnitude of the difference between some sample dependent quantities and their probability limits. These orders of magnitude are enough to make these differences negligible in the expansions. It is worth mentioning that Assumption 5(iii) is not particularly restrictive since most of the orders of magnitude imposed are guaranteed by the central limit theorem.

In preparation for our asymptotic theory result, we define the following quantities. Let M_d be the matrix of the orthogonal projection on the orthogonal complement of $W^{1/2}D$:

$$M_d = I_q - W^{1/2}D(D'WD)^{-1}D'W^{1/2},$$

where I_q is the identity matrix of size q, let P_g be the matrix of the orthogonal projection on $M_d W^{1/2} G$:

$$P_g = M_d W^{1/2} G \left(G' W^{1/2} M_d W^{1/2} G \right)^{-1} G' W^{1/2} M_d,$$

and let M_{dg} be the matrix of the orthogonal projection on the orthogonal complement of $\begin{pmatrix} W^{1/2}D & W^{1/2}G \end{pmatrix}$:

$$M_{dg} = M_d - P_g.$$

Let

$$\begin{aligned} \mathbb{R}_1 \;=\; & \left(\mathbb{Z}_0' W^{1/2} P_g W^{1/2} \mathbb{Z}_0 G' - G' W^{1/2} M_d W^{1/2} \mathbb{Z}_0 \mathbb{Z}_0' \right) \\ & \times \; W^{1/2} M_d W^{1/2} \left(\tfrac{1}{3} L + G_{1p_\phi} H G \right) / \sigma_G \qquad (16) \\ & + \; \mathbb{Z}_0' W^{1/2} M_{dg} W^{1/2} (\mathbb{Z}_1 + G_{1p_\phi} H \mathbb{Z}_0), \end{aligned}$$

with $\sigma_G = G' W^{1/2} M_d W^{1/2} G$, and $H = -(D'WD)^{-1} D'W$.

The following result gives the asymptotic distribution of the GMM estimator $\hat{\phi}$ as defined by (3):

Theorem 1. *Under Assumptions 2, 4, and 5, we have:*

(a) $\hat{\phi}_1 - \phi_{0,1} = O_P(T^{-1/2})$ *and* $\hat{\phi}_{p_\phi} - \phi_{0,p_\phi} = O_P(T^{-1/4})$.

(b) If in addition, ϕ_0 *is interior to* Φ*, then*

$$\sqrt{T}\begin{pmatrix} \hat{\phi}_1 - \phi_{0,1} \\ (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} H\mathbb{Z}_0 + HG\mathbb{V}/2 \\ \mathbb{V} \end{pmatrix},$$

with $\mathbb{V} = -2\frac{\mathbb{Z}\mathbb{I}(\mathbb{Z}<0)}{G'W^{1/2}M_dW^{1/2}G}$ *and* $\mathbb{Z} = G'W^{1/2}M_dW^{1/2}\mathbb{Z}_0$. $\mathbb{I}(\cdot)$ *is the usual indicator function.*

(c) If in addition, $\mathbb{R}_1$ *does not have an atom of probability at 0, then:*

$$\begin{pmatrix} \sqrt{T}(\hat{\phi}_1 - \phi_{0,1}) \\ T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) \end{pmatrix} \xrightarrow{d} \mathbb{X} \equiv \begin{pmatrix} H\mathbb{Z}_0 + HG\mathbb{V}/2 \\ (-1)^{\mathbb{B}}\sqrt{\mathbb{V}} \end{pmatrix},$$

with $\mathbb{B} = \mathbb{I}(\mathbb{R}_1 \geq 0)$.

The proof of this theorem is provided in Appendix. Part (a) is due to Dovonon and Renault (2009). We however provide a proof since our conditions are slightly different from theirs. Part (b) gives the asymptotic distribution of $(\hat{\phi}_1 - \phi_{0,1}, (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2)$. This result is obtained by eliciting the $O_P(T^{-1})$ terms of $m_T'(\hat{\phi})W_Tm_T(\hat{\phi})$ which are collected into $K_T(\phi)$ as given by (47) in the Appendix. The fact that $K_T(\phi_{p_\phi})$ is a quadratic function of $(\phi_{p_\phi} - \phi_{0,p_\phi})^2$ gives an intuition of the fact that only the asymptotic distribution of $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2$ can be obtained from this leading term of the expansion of the GMM objective function. The distribution of $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$ can be obtained from Part (b) up to the sign which cannot be deduced from this leading term but rather is obtainable from the higher order, $O_P(T^{-5/4})$, term of the objective function's expansion. We actually obtain:

$$m_T'(\hat{\phi})W_Tm_T(\hat{\phi}) = K_T(\hat{\phi}_{p_\phi}) + (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})R_{1T} + o_P(T^{-5/4})$$

showing that the minimum is reached at $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$ having opposite sign to R_{1T}. See (48) in Appendix for the expression of R_{1T}. So long as TR_{1T}, with limit distribution $\mathbb{R}_1$, does not vanish asymptotically, the sign of $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$ can be identified by this higher order term in the expression leading to Part (c) of the theorem.

Remark 1. *The continuity condition for* $\mathbb{R}_1$ *at 0 is not expected to be restrictive in general since* $\mathbb{R}_1$ *is a quadratic function of the Gaussian vector* $(\mathbb{Z}_0', \mathbb{Z}_1')'$*. However, when* $q = p = p_\phi = 1$ *(one moment restriction with one non first-order locally identified parameter), we can see that* $\mathbb{R}_1 = 0$*. In this case, the characterization of the asymptotic distribution of* $T^{1/4}(\hat{\phi} - \phi_0)$ *may be problematic if the estimating function is quadratic in* ϕ*. Actually,* $T^{1/4}(\hat{\phi} - \phi_0)$ *may not have a proper asymptotic distribution in this case whereas* $\sqrt{T}(\hat{\phi} - \phi_0)^2$ *does have one as given by Theorem 1(b).*

Remark 2. *The asymptotic distributions in Parts (b) and (c) of Theorem 1 are both non-standard but easy to simulate. The source of randomness is* $(\mathbb{Z}_0', \mathbb{Z}_1')'$ *which is typically a Gaussian vector with zero mean and asymptotic variance* $v = \lim_{T\to\infty} TVar\begin{pmatrix} m_T(\phi_0) \\ \frac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0) \end{pmatrix}$ *which can be consistently estimated by sample variance if there are no serial correlation or by heteroskedasticity and* *autocorrelation consistent procedures if there are serial correlations (see Andrews, 1991). Letting $\hat{v}$ be a consistent estimate of v, drawing randomly copies of $(\mathbb{Z}_0', \mathbb{Z}_1')'$ from $N(0, \hat{v})$ and using consistent estimators of D, W, G, L and G_{1p_ϕ} shall give reasonable approximation of copies from these limiting distributions.*

Remark 3. *When the moment condition model has a single parameter that is not locally identified at the first order but is at the second order, asymptotically correct $(1-\alpha)$-confidence interval, $1-\alpha > 1/2$, for ϕ_0 can be obtained without simulation using the following formula deduced from Theorem 1(b). The proof of asymptotic correctness is provided in Appendix.*

$$CI_{1-\alpha}(\phi_0) = \hat{\phi} \pm \frac{1}{T^{1/4}} \cdot \left(\frac{2\sqrt{\hat{G}'W_T\hat{\Omega}W_T\hat{G}}}{\hat{G}'W_T\hat{G}} z_\alpha \right)^{\frac{1}{2}}, \tag{17}$$

where $\hat{\phi}$ is the GMM estimator of ϕ_0, $\hat{G}$ is a consistent estimator of G, $\hat{\Omega}$ a consistent estimator of the long run variance of the estimating function, i.e. $Var(\mathbb{Z}_0)$ and z_α is the $(1-\alpha)$-quantile of the standard normal distribution.

Assumption 4 requires that the rank deficiency occurs in a particular way as one column of the Jacobian matrix of the moment function vanishes whereas the other columns are linearly independent. This is only a particular form of lack of first-order identification that does not fit exactly our second example in Section 3. However, as mentioned by Sargan (1983), up to a rotation of the parameter space, all rank deficient problems can be brought into this configuration as we can see below.

Let $M_0 = \frac{\partial m}{\partial \phi'}(\phi_0)$ and assume that the moment condition model (1) is such that $Rank\,(M_0) = p_\phi - 1$ without having a column that is equal to 0.

Let R be any nonsingular (p_ϕ, p_ϕ)-matrix such that $M_0 R_{\bullet p_\phi} = 0$, where $R_{\bullet p_\phi}$ represents the last column of R. We can write (1) in terms of the parameter vector η: $\phi = R\eta$ and consider the model:

$$E\left(g(X, R\eta)\right) = 0. \tag{18}$$

By the chain rule, it is not hard to see that Model (18) identifies $\eta_0 = R^{-1}\phi_0$ with local identification properties matching Assumption 4. More precisely, setting $\eta = (\eta_1', \eta_{p_\phi})'$ where η_1 is $p_\phi - 1$, we have:

$$\left.\frac{\partial m(R\eta)}{\partial \eta_{p_\phi}}\right|_{\eta_0} = M_0 R_{\bullet p_\phi} = 0 \text{ and } Rank\left(\left.\frac{\partial m(R\eta)}{\partial \eta_1'}\right|_{\eta_0} \right) = Rank(M_0 R_1) = p_\phi - 1,$$

where R_1 is the sub-matrix of the first $p_\phi - 1$ columns of R. We can therefore claim that the asymptotic distribution, $\tilde{\mathbb{X}}$, of $\begin{pmatrix} \sqrt{T}(\hat{\eta}_1 - \eta_{0,1}) \\ T^{1/4}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi}) \end{pmatrix}$ is obtained by Theorem 1 with D, G, L, and G_{1p_ϕ} replaced respectively by:

$$\tilde{D} = M_0 R_1;\ \tilde{G} = \left(R'_{\bullet p_\phi} \frac{\partial^2 m_k}{\partial \phi \partial \phi'}(\phi_0) R_{\bullet p_\phi} \right)_{1 \le k \le q};\ \tilde{L} = \left(R'_{\bullet p_\phi} A_k R_{\bullet p_\phi} \right)_{1 \le k \le q},$$

$$A_k = \left(\frac{\partial^3 m_k}{\partial \phi_i \partial \phi_j \partial \phi'}(\phi_0) R_{\bullet p_\phi} \right)_{1 \le i,j \le p_\phi};$$

and $\tilde{G}_{1p_\phi}$, the $(q, p_\phi - 1)$-matrix with its k^{th} row equal to $R'_{\bullet p_\phi} \frac{\partial^2 m_k}{\partial\phi\partial\phi'}(\phi_0) R_1$.

We use the fact that $\hat{\phi} - \phi_0 = R(\hat{\eta} - \eta_0)$ to obtain the asymptotic distribution of $\hat{\phi} - \phi_0$. Specifically, letting $B_T = \begin{pmatrix} \sqrt{T} I_{p_\phi - 1} & 0 \\ 0 & T^{1/4} \end{pmatrix}$, we obtain the asymptotic distribution of $B_T R^{-1}(\hat{\phi} - \phi_0)$ as that of $B_T(\hat{\eta} - \eta_0)$.

Feasible inference is possible by replacing R by a consistent estimate $\hat{R}$. However, because all the components of $R^{-1}(\hat{\phi} - \phi_0)$ are not converging at the same rate, one needs to exercise some caution in claiming the asymptotic equivalence between $B_T \hat{R}^{-1}(\hat{\phi} - \phi_0)$ and $B_T R^{-1}(\hat{\phi} - \phi_0)$. Clearly,

$$B_T \hat{R}^{-1}(\hat{\phi} - \phi_0) = B_T R^{-1}(\hat{\phi} - \phi_0) + \epsilon_T \tag{19}$$

$\epsilon_T = -B_T \hat{R}^{-1}(\hat{R} - R) R^{-1}(\hat{\phi} - \phi_0)$. But ϵ_T does not always vanish asymptotically. We distinguish two cases:

Case 1: $\hat{R} - R = o_P(T^{-1/4})$. This is the case, for example, if R does not depend on ϕ_0 and $\hat{R}$ is a smooth function of sample means of the data (and does not depend on ϕ_0). In such a case we typically have $\hat{R} - R = O_P(T^{-1/2})$. By the Cauchy-Schwarz inequality, we have:

$$\|\epsilon_T\| \leq \|\hat{R}^{-1}\| \|T^{1/4}(\hat{R} - R)\| \|T^{1/4} R^{-1}(\hat{\phi} - \phi_0)\| = O_P(1) o_P(1) O_P(1)$$

and this remainder is negligible so that:

$$B_T \hat{R}^{-1}(\hat{\phi} - \phi_0) \overset{d}{\to} \tilde{\mathbb{X}}. \tag{20}$$

Case 2: $\hat{R} - R = O_P(T^{-1/4})$. This is expected for example if R is a function of ϕ_0, *i.e.* $R \equiv R(\phi_0)$. If $R(\cdot)$ is continuously differentiable in a neighborhood of ϕ_0, we can show (see Appendix) that:

$$\epsilon_T = -A\sqrt{T}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi})^2 + o_P(1), \tag{21}$$

with

$$A = \begin{pmatrix} I_{p_\phi - 1} & 0 \\ 0 & 0 \end{pmatrix} R^{-1} \frac{\partial R_{\bullet p_\phi}}{\partial \phi'}(\phi_0) R_{\bullet p_\phi}.$$

Hence, we have:

$$B_T \hat{R}^{-1}(\hat{\phi} - \phi_0) \overset{d}{\to} \tilde{\mathbb{X}} - A(\tilde{\mathbb{X}}_{p_\phi})^2. \tag{22}$$

To perform inference about ϕ_0 in the current local identification set up, it is essential to estimate a relevant reparameterization matrix R that may depend on the model parameter ϕ.

In many cases, including Examples 1 and 2, it is possible, by examining the population Jacobian matrix at the true value, to figure out that it is of rank $p_\phi - 1$ at ϕ_0 because a non-zero vector $R_{p_\phi}(\phi_0)$ is found in its null space. At most, $R_{p_\phi}(\phi_0)$ is a moment function depending on ϕ_0 that can be consistently estimated by sample moments evaluated at $\hat{\phi}$ providing, therefore, a consistent estimator $\hat{R}_{p_\phi}(\hat{\phi})$. This vector can be completed by the first $p_\phi - 1$ elements of the canonical basis of $\mathbb{R}^{p_\phi}$ that do not linearly determine $\hat{R}_{p_\phi}(\hat{\phi})$ to obtain an estimate $\hat{R}(\hat{\phi})$ of the full reparameterization matrix $R(\phi_0)$. Examples 1 and 2 are further studied below to determine their respective reparameterization matrices.

In other cases, one may rely on testing to determine whether the rank of the Jacobian matrix at ϕ_0 is $p_\phi - 1$. The rank test of Wright (2003) can serve this purpose. In this case, the estimation of an element of the null space of the Jacobian is a bit more involved. Assuming that the first $p_\phi - 1$ columns of $E\left(\frac{\partial g(X,\phi_0)}{\partial \phi'}\right)$ are linearly independent. Then, in a similar spirit as the augmented regression of Arellano, Hansen, and Sentana (2012), we can find an element R_{p_ϕ} in its null space such that $R_{p_\phi} = (r', -1)' \in \mathbb{R}^{p_\phi - 1} \times \mathbb{R}$, with r defined by:

$$E\left(\frac{\partial g(X,\phi_0)}{\partial \phi_1'} r - \frac{\partial g(X,\phi_0)}{\partial \phi_{p_\phi}}\right) = 0.$$

Letting $\Gamma_1(\phi_0) = E\left(\frac{\partial g(X,\phi_0)}{\partial \phi_1'}\right)$, and $\Gamma_{p_\phi}(\phi_0) = E\left(\frac{\partial g(X,\phi_0)}{\partial \phi_{p_\phi}}\right)$, we have: $r = \mathbf{r}(\phi_0)$ with

$$\mathbf{r}(\phi) = \left(\Gamma_1(\phi)'\Gamma_1(\phi)\right)^{-1} \Gamma_1(\phi)'\Gamma_{p_\phi}(\phi).$$

Hence, r can be estimated by plug-in where ϕ_0 is replaced by $\hat{\phi}$ and the population means are replaced by sample means: $\hat{r} = \hat{\mathbf{r}}(\hat{\phi})$. This leads to an estimate $\hat{R}_{p_\phi}(\hat{\phi})$ that can be completed as described above to obtain an estimate $\hat{R}$ of R. If $\hat{R} - R = o_P(T^{-1/4})$, then valid inference can be carried out about ϕ_0 using (20). This is the case in particular when the last column of $\frac{\partial \mathbf{r}}{\partial \phi'}(\phi_0)R$ is nil. However, one shall in general expect that $\hat{R} - R = O_P(T^{-1/4})$ and inference shall be done using (22). In this case, a consistent estimate of A can be obtained by plug-in. Note that $\frac{\partial R_{\bullet p_\phi}}{\partial \phi'}(\phi_0) = \left(\frac{\partial \mathbf{r}'}{\partial \phi}(\phi_0) \;\; 0\right)'$. Letting ϕ_k be the kth component of ϕ,[19] $\Gamma_j \equiv \Gamma_j(\phi)$ and $\Gamma_{j,k} \equiv \frac{\partial \Gamma_j}{\partial \phi_k}(\phi)$, for $j = 1, p_\phi$ and $k = 1, \dots, p_\phi$, we have:

$$\frac{\partial \mathbf{r}}{\partial \phi_k}(\phi) = \left(\Gamma_1'\Gamma_1\right)^{-1}\left\{-\left(\Gamma_{1,k}'\Gamma_1 + \Gamma_1'\Gamma_{1,k}\right)\left(\Gamma_1'\Gamma_1\right)^{-1}\Gamma_1\Gamma_{p_\phi} + \Gamma_{1,k}'\Gamma_{p_\phi} + \Gamma_1'\Gamma_{p_\phi,k}\right\}.$$

Returning to our examples we can see that in Example 1 of Section 3, the Jacobian matrix of the moment function (12) is null, hence, Theorem 1 applies to the GMM estimator of this model without a need for reparameterization. In Example 2 of the same section, the Jacobian matrix of the auxiliary model at the true parameter value is:

$$\begin{pmatrix} 0 & 0 & 0 & -2(b_3 - \delta b_2) & -1 \\ 0 & 0 & 0 & -2(b_3 - \delta b_2)E(z_{t-1}) & -E(z_{t-1}) \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}.$$

Recall that $b_3 - \delta b_2 = -\Omega_2 \neq 0$. The null space of this matrix contains $R_5 = (0, 0, 0, -1, 2(b_3 - \delta b_2))'$ and, by completing this vector by the elements of the canonical basis of $\mathbb{R}^5$, we obtain a relevant reparameterization matrix along with a consistent estimator as:

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 2(b_3 - \delta b_2) \end{pmatrix} \quad \text{and} \quad \hat{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 2(\hat{b}_3 - \hat{\delta}\hat{b}_2) \end{pmatrix},$$

[19] This involves a slight abuse of notation: in the formula for $\Gamma_{j,k}$ $\partial \mathbf{r}/\partial \phi_k$, ϕ_1 denotes the first element of ϕ and not $\phi_{1:p_\phi - 1}$ as above.

respectively. The asymptotic distribution for the auxiliary estimator is given by (22) which can be used to build inference about the indirect inference estimator along with Theorem 2 in the next section.

We now describe how the above results can be used to perform inference about the parameters. Our discussion focuses on Case 1 above but the methods can also be applied in settings covered by Case 2. Suppose first it is desired to construct a confidence set for ϕ_0. From (20), it follows via the Continuous Mapping Theorem that

$$lim_{T\to\infty} P\Big(\|B_T \hat{R}^{-1}(\hat{\phi}-\phi_0)\|^2 < c(1-\alpha) \Big) \;=\; 1-\alpha \tag{23}$$

where $c(1-\alpha)$ is the $100(1-\alpha)^{th}$ percentile of $\|\tilde{\mathbb{X}}\|^2$. Thus,

$$\{\phi,\ \|B_T \hat{R}^{-1}(\hat{\phi}-\phi_0)\|^2 < c_{1-\alpha}\}$$

represents a $100(1-\alpha)\%$ confidence set for ϕ_0. To derive a confidence interval for the j^{th} element of ϕ_0, $\phi_{0,j}$, [20] we re-write (20) as

$$\hat{\phi}-\phi_0 \;\overset{a}{\sim}\; \hat{R}B_T^{-1}\tilde{\mathbb{X}}. \tag{24}$$

A confidence interval for $\phi_{0,j}$ can be built using simulation of $\tilde{\mathbb{X}}$ as follows: Simulate $\tilde{\mathbb{X}}$ to obtain the simulated distribution of $[\hat{R}B_T^{-1}\tilde{\mathbb{X}}]_j$. To construct an equal-tailed confidence interval, calculate the $c_j(\alpha/2)$ and $c_j(1-\alpha/2)$, respectively the $100(\alpha/2)^{th}$ and $100(1-\alpha/2)^{th}$ percentiles of the simulated distribution of $[\hat{R}B_T^{-1}\tilde{\mathbb{X}}]_j$. A $100(1-\alpha)\%$ confidence interval is:

$$\{\phi_j,\ c_j(\alpha/2) < \phi_j < c_j(1-\alpha/2)\}.$$

The results in Theorem 1 can also be used to test hypotheses of the form $H_0 : f(\phi_0) = 0$ versus $H_1 : f(\phi_0) \neq 0$ where $f(\cdot)$ is a $K \times 1$ vector of two times continuously differentiable functions. In the appendix, it is shown that:

$$f(\hat{\phi}) - f(\phi_0) \;\overset{a}{\sim}\; \frac{\partial f}{\partial \phi'}(\hat{\phi})\hat{R}B_T^{-1}\tilde{\mathbb{X}} - \frac{1}{2}\frac{1}{\sqrt{T}}\left(\hat{R}'_{\bullet p_\phi}\frac{\partial^2 f_k}{\partial\phi\partial\phi'}(\hat{\phi})\hat{R}_{\bullet p_\phi}\tilde{\mathbb{X}}^2_{p_\phi}\right)_{1\le k\le K} \equiv V_T(\hat{\phi},\tilde{\mathbb{X}}). \tag{25}$$

Quantiles of $V_T(\hat{\phi},\tilde{\mathbb{X}})$ can be simulated from copies of $\tilde{\mathbb{X}}$. Note that if f is a linear function then the asymptotic then

$$V_T(\hat{\phi},\tilde{\mathbb{X}}) = \frac{\partial f}{\partial \phi'}(\hat{\phi})\hat{R}B_T^{-1}\tilde{\mathbb{X}}.$$

5 The limiting distribution of the II estimator

In this section, we derive the asymptotic distribution of the indirect inference estimator as defined by (4) and (5) when the auxiliary model is given by moment conditions that are first-order locally under-identified.

Let us consider the auxiliary model to be the following moment condition:

$$E[g(x,h)] = 0, \tag{26}$$

[20] This involves a slight abuse of notation as here $\phi_{0,1}$ denotes the first element of $\phi_{0,1}$ and not the $(p_\phi - 1 \times 1)$ sub-vector of ϕ_0 as above.

where $g(\cdot)$ a $q \times 1$ vector of continuous functions and h is the $\ell \times 1$ vector of parameters. As described in Section 2, h is estimated based on (26) using the data and simulated series providing the sequences h_T and $h_T^{(i)}(\theta)$, $i = 1, \dots, s$ that are the auxiliary features used to estimate the parameter of interest θ by the quadratic optimization (5).

We assume that (26) satisfies the local identification property in Assumption 5 in terms of the parameter h and derive the asymptotic distribution of the indirect estimator $\hat{\theta}_{II}$ in this framework. We use Ω_T to denote the sequence of weighting matrices that determine the indirect estimator in (5) and keep W_T as sequence of weighting matrices that determine $\hat{h}_T$. We assume that Ω_T converges in probability to Ω that is symmetric positive definite.

Proposition 2 ensures that the indirect estimator is consistent under Assumption 2 which continues to hold even when the auxiliary model is not first-order locally identified. If θ_0 is interior to Θ, the indirect estimator solves with probability approaching 1 the first-order condition (9):

$$M_{IT}(\hat{\theta}_{II})'\Omega_T m_{IT}(\hat{\theta}_{II}) = 0,$$

with $m_{IT}(\theta) = h_T - \frac{1}{s}\sum_{i=1}^{s} h_T^{(i)}(\theta)$ and $M_{IT}(\theta) = \frac{\partial m_{IT}}{\partial \theta'}(\theta)$. By a first-order mean value expansion of m_{IT} around θ_0, we have:

$$M_{IT}(\hat{\theta}_{II})\Omega_T \left(m_{IT}(\theta_0) + M_{IT}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0) \right) = 0,$$

with $\dot{\theta}_T \in (\hat{\theta}_{II}, \theta_0)$ and may differ from row to row. We deduce that:

$$\hat{\theta}_{II} - \theta_0 = \dot{F}_T \left(h_T - \frac{1}{s}\sum_{i=1}^{s} h_T^{(i)}(\theta_0) \right), \tag{27}$$

with

$$\dot{F}_T = -\left(M_{IT}(\hat{\theta}_{II})'\Omega_T M_{IT}(\dot{\theta}_T) \right)^{-1} M_{IT}(\hat{\theta}_{II})'\Omega_T.$$

The asymptotic distribution of $\hat{\theta}_{II} - \theta_0$ depends on that of $h_T - \frac{1}{s}\sum_{i=1}^{s} h_T^{(i)}(\theta_0)$. Under the conditions of Theorem 1 for the auxiliary moment condition model,

$$B_T(h_T - h_0) \xrightarrow{d} \mathbb{X}, \quad \text{and} \quad B_T(h_T^{(i)} - h_0) \xrightarrow{d} \mathbb{X},$$

for all $i = 1, \dots, s$ with B_T the diagonal $\ell \times \ell$ matrix of rates of convergence with all its diagonal elements equal $\sqrt{T}$ except for the last one which is $T^{1/4}$.

Hence, assuming that $h_T^{(i)}(\theta_0)$ are independent across i and independent of h_T[21], we have:

$$B_T \left(h_T - \frac{1}{s}\sum_{i=1}^{s} h_T^{(i)}(\theta_0) \right) \xrightarrow{d} \mathbb{Y} \equiv \mathbb{X}_0 - \frac{1}{s}\sum_{i=1}^{s} \mathbb{X}_i,$$

where $\mathbb{X}_0, \mathbb{X}_1, \dots \mathbb{X}_s$ are independent with the same distribution as $\mathbb{X}$.

The fact that the rates of convergence in the diagonal of B_T are not all equal make the determination of the rate of convergence of $\hat{\theta}_{II} - \theta_0$ from that of $m_{IT}(\theta_0)$ more complicated than in the standard case. Pre-multiplying (27) by $T^{1/4}$, we have:

$$T^{1/4}(\hat{\theta}_{II} - \theta_0) = \dot{F}_{T,\bullet\ell} T^{1/4} m_{IT,\ell}(\theta_0) + o_P(1) = F_{\bullet\ell} T^{1/4} m_{IT,\ell}(\theta_0) + o_P(1), \tag{28}$$

[21] This is the case when there are no state variables so that the simulated samples are independent across $i = 1, \dots, s$. (See Gourieroux, Monfort, and Renault, 1993).)

where F is the probability limit of $\dot{F}_T$ and $\dot{F}_{T,\bullet\ell}$ and $F_{\bullet\ell}$ are the ℓ^{th} column of $\dot{F}_T$ and F, respectively. Hence:

$$T^{1/4}(\hat{\theta}_{II} - \theta_0) \xrightarrow{d} F_{\bullet\ell}\mathbb{Y}_\ell,$$

where $F_{\bullet\ell}$ is defined similarly to $\dot{F}_{T,\bullet\ell}$ and $\mathbb{Y}_\ell$ is the ℓ^{th} component of $\mathbb{Y}$.

This asymptotic distribution represents a p-dimensional sample dependent random vector that converges in distribution to a random vector that has only one dimension of randomness. In fact, $T^{1/4}$ appears to be the slowest rate of convergence of $(\hat{\theta}_{II} - \theta_0)$ in any direction in the space. However asymptotic inference on θ_0 would benefit from a further characterization of the asymptotic distribution. We expect that some linear combinations of $\hat{\theta}_{II} - \theta_0$ converge faster than others that converge at the rate $T^{1/4}$.

To derive this asymptotic distribution, we will rely on a second-order expansion of $m_{IT}(\hat{\theta}_{II})$ around θ_0. Such higher order expansion is required by the fact that $(\hat{\theta}_{II} - \theta_0)$ has the rate of convergence $T^{1/4}$ in some directions and therefore, its quadratic function is a non-negligible component of $m_{IT}(\hat{\theta}_{II})$. We make the following assumption:

Assumption 6. $\Delta_{IT,k}(\theta) \equiv \frac{\partial^2 m_{IT,k}(\theta)}{\partial\theta\partial\theta'}$ *converges in probability uniformly over* N_ϵ *to* $\Delta_{I,k}(\theta) \equiv \frac{\partial^2 m_{I,k}(\theta)}{\partial\theta\partial\theta'}$ *for* $k = 1, \ldots, \ell$.

By a second-order mean value expansion of $m_{IT}(\theta_0)$ around $\hat{\theta}_{II}$, and after re-arranging, we have:

$$\begin{aligned} & m_{IT}(\hat{\theta}_{II}) \\ = \;& m_{IT}(\theta_0) + M_{IT}(\hat{\theta}_{II})(\hat{\theta}_{II} - \theta_0) - \tfrac{1}{2}\left((\hat{\theta}_{II} - \theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0)\right)_{1\le k\le \ell}, \end{aligned}$$

where $\dot{\theta}_T \in (\theta_0, \hat{\theta}_{II})$ may differ from row to row. Solving this in $(\hat{\theta}_{II} - \theta_0)$ yields:

$$\hat{\theta}_{II} - \theta_0 = \hat{F}_T\left(m_{IT}(\theta_0) - \frac{1}{2}\left((\hat{\theta}_{II} - \theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0)\right)_{1\le k\le \ell}\right), \tag{29}$$

with

$$\hat{F}_T = -\left(M_{IT}(\hat{\theta}_{II})'\Omega_T M_{IT}(\hat{\theta}_{II})\right)^{-1} M_{IT}(\hat{\theta}_{II})'\Omega_T.$$

To characterize the directions of fast convergence of $\hat{\theta}_{II} - \theta_0$, let $\hat{S}$ be the $p\times p$ matrix with unit and pairwise orthogonal p-vectors as rows with the last row equal to the last column of $\hat{F}_T$ normalized and $\hat{S}_1$ be the $(p-1)\times p$ sub-matrix of the first $(p-1)$ rows of $\hat{S}$. The last remark in this section gives how the matrix $\hat{S}$ can be determined as a continuous function of the last column of $\hat{F}_T$. By construction, $\hat{S}_1\hat{F}_T m_{IT}(\theta_0)$ does not depend on the slow converging component, $m_{IT,\ell}(\theta_0)$, of $m_{IT}(\theta_0)$. We therefore have:

$$\begin{aligned} & \sqrt{T}\hat{S}_1\left(\hat{\theta}_{II} - \theta_0\right) \\ = \;& \hat{S}_1\hat{F}_T B_T\left(m_{IT}(\theta_0) - \tfrac{1}{2}\left((\hat{\theta}_{II} - \theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0)\right)_{1\le k\le \ell}\right). \end{aligned} \tag{30}$$

By combining (28) and (30) and letting S be the probability limit of $\hat{S}$ and $B_{IT} = \begin{pmatrix} \sqrt{T}I_{p-1} & 0 \\ 0 & T^{1/4} \end{pmatrix}$, we have the following result:

Theorem 2. *Assume that the indirect estimator's program satisfies Assumptions 2, 3 and 6 with θ_0 interior to Θ. Assume that the auxiliary model satisfies Assumptions 2, 4 and 5, and that h_0 is interior to the auxiliary parameter set and that the related random variable $\mathbb{R}_1$ as defined by* (16) *has no atom of probability at 0. If the s indirect inference samples are generated independently and the last column of F is different from 0, then:*

$$B_{IT}\hat{S}\left(\hat{\theta}_{II}-\theta_0\right) \xrightarrow{d} \begin{pmatrix} S_1 F\left(\mathbb{Y}-\frac{(\mathbb{Y}_\ell)^2}{2}\left(F'_{\bullet\ell}\Delta_{I,k}(\theta_0)F_{\bullet\ell}\right)_{1\leq k\leq\ell}\right) \\ \\ S_{p\bullet}F_{\bullet\ell}\mathbb{Y}_\ell \end{pmatrix},$$

where S_1 is the sub-matrix of the first $(p-1)$ rows of S, $S_{p\bullet}$ is the last row of S, $F_{\bullet\ell}$ is the last column of F, $\mathbb{Y}=\mathbb{X}_0-\frac{1}{s}\sum_{i=1}^{s}\mathbb{X}_i$, with $\mathbb{X}_j$'s independently and identically distributed as $\mathbb{X}$, and $\mathbb{Y}_\ell$ is the ℓ^{th} component of $\mathbb{Y}$.

The proof is relegated to the Appendix. The asymptotic distribution of $B_{IT}\hat{S}(\hat{\theta}_{II}-\theta_0)$ can be simulated by replacing S, F and $\Delta_{I,k}(\theta_0)$, $k=1,\ldots,\ell$ by their estimates, $\hat{S}$, $\hat{F}$ and $\Delta_{IT,k}(\hat{\theta}_{II})$, $k=1,\ldots,\ell$. The simulation of $\mathbb{Y}$ is based on that of $\mathbb{X}$ which is described in the previous section.

Remark 4. *In the case where the rank deficiency in the auxiliary model appears in a way that no column of the Jacobian matrix is nil, we can get the asymptotic distribution of the indirect estimator as follows. The asymptotic distribution of $B_T\hat{R}^{-1}(h_T-h_0)$ is derived in the previous section. Let $\tilde{\mathbb{X}}$ denote this asymptotic distribution in either Case 1 or Case 2. From* (27)*, we can show that:*

$$T^{1/4}(\hat{\theta}_{II}-\theta_0)=FR_{\bullet\ell}T^{1/4}(\hat{R}^{-1}m_{IT}(\theta_0))_\ell+o_P(1),$$

where $R_{\bullet\ell}$ is the last column of R and $(\hat{R}^{-1}m_{IT}(\theta_0))_\ell$ is the last component of $\hat{R}^{-1}m_{IT}(\theta_0)$. Also, from (29)*, we have*

$$\hat{\theta}_{II}-\theta_0=\hat{F}_T\hat{R}\left(\hat{R}^{-1}m_{IT}(\theta_0)-\frac{1}{2}\hat{R}^{-1}\left((\hat{\theta}_{II}-\theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II}-\theta_0)\right)_{1\leq k\leq\ell}\right).$$

Letting $\hat{S}_R$ be row-wise, the orthonormal basis obtained by completing the last column of $\hat{F}_T\hat{R}$ according to Remark 5 below,and S_R its probability limit, we have that:

$$B_{IT}\hat{S}_R\left(\hat{\theta}_{II}-\theta_0\right) \xrightarrow{d} \begin{pmatrix} S_{R,1}FR\left(\tilde{\mathbb{Y}}-\frac{(\tilde{\mathbb{Y}}_\ell)^2}{2}R^{-1}\left(R'_{\bullet\ell}F'\Delta_{I,k}(\theta_0)FR_{\bullet\ell}\right)_{1\leq k\leq\ell}\right) \\ \\ S_{R,p\bullet}FR_{\bullet\ell}\tilde{\mathbb{Y}}_\ell \end{pmatrix},$$

where $\tilde{\mathbb{Y}}=\tilde{\mathbb{X}}_0-\frac{1}{s}\sum_{i=1}^{s}\tilde{\mathbb{X}}_i$, with $\tilde{\mathbb{X}}_j$'s are independent and identically distributed as $\tilde{\mathbb{X}}$ and $S_{R,1}$, $S_{R,p\bullet}$ are defined similarly to S_1 and $S_{p\bullet}$ in Theorem 2.

Remark 5. *We now describe a procedure that can be used to determine the matrix of orthogonal directions $\hat{S}$ from $\hat{F}_T$.*

Let u be a p-vector different from 0. Take the first $p-1$ vectors from the canonical basis $(e_1,e_2,\ldots,e_p)$ of $\mathbb{R}^p$, the span of which does not contain u. Assume without loss of generality that *these elements are* $e_1, e_2, \ldots, e_{p-1}$ *in this order (the order of elements in the bases are important to guarantee uniqueness of the outcome).*

Consider the basis $(u, e_1, e_2, \ldots, e_{p-1})$ *and determine an orthonormal basis from this basis using the Gram-Schmidt orthonormalization process. Let* $(\tilde{u}, \tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_{p-1})$ *be the resulting orthonormal basis. Take*

$$S(u) = (\tilde{e}_1 \quad \tilde{e}_2 \quad \ldots \quad \tilde{e}_{p-1} \quad \tilde{u})'.$$

We can verify that this procedure gives a unique S *and is a continuous function of* u*. The continuity of this procedure allows the application of the continuous mapping theorem as we do in the proof of Theorem 2.*

In this subsection we have studied the asymptotic behaviour of the indirect estimator when the auxiliary moment condition model is only second-order locally identified, but the indirect inference estimation program (5) is standard in the sense that it satisfies Assumptions 1, 2 and 3. It is worth mentioning the possibility of the indirect inference program suffering local identification issues in its own right. This would be the case if $M_I(\theta_0)$ has rank $r < p$. second-order identification would be warranted if $m_I(\theta)$ satisfies Definition 1 at θ_0. If, in particular, the rank of $M_I(\theta_0)$ is $p - 1$ and the conditions of Assumption 5 apply to $m_I(\cdot)$ and Ω_T, then the asymptotic distribution of the indirect estimator is readily available by applying Theorem 1. Note however that the investigation of local identification properties of the indirect inference program may be difficult particularly as $m_I(\cdot)$ and $M_I(\cdot)$ are often obtained by simulation.

In this section, we have focussed on the case where the auxiliary model is a GMM estimator because we believe this relevant for practitioners. Recently, Frazier and Renault (2016) have considered the more general situation in which the auxiliary model is estimated from a linear combination of moment conditions. They show that under certain conditions choosing the linear combination of moments in the auxiliary model implicitly associated with GMM is asymptotically inefficient for estimation of the parameters of the simulator. However, the conditions for their analysis include that the parameters of the auxiliary model are first order identified. The situation considered in our paper is therefore outside their framework.

6 Monte Carlo results

This section illustrates the finite sample performance of the asymptotic results derived in this paper through Monte Carlo simulations. We are mainly concerned with coverage probabilities of confidence intervals (CI) based on the asymptotic distribution derived for the GMM and II estimators in Theorems 1 and 2, respectively.

Our simulations are based on the dynamic panel data model in Example 1 in Section 3. The simulated data are generated from $\varepsilon_{it} \sim NID(0, \sigma_\varepsilon^2)$, $i = 1, \ldots, n$, $t = 1, 2$ independent of $(\eta_i, y_{i0}) \sim NID(0, \Sigma)$, with $\Sigma_{11} = \sigma_\eta^2$, $\Sigma_{22} = \sigma_0^2$, and $\Sigma_{12} = \sigma_{0\eta}$. The autoregressive dynamic in (10) is then used to obtain samples $\{y_i = (y_{i0}, y_{i1}, y_{i2}) : i = 1, \ldots, n\}$ for various values of ρ.

The parameter of interest ρ is estimated by 2-step GMM and II using the moment restriction in (12). The 2-step GMM estimator $\hat{\rho}$ is obtained using the identity matrix as weighting matrix at the first step and the estimated optimal weighting matrix at the second step. We obtain the indirect inference estimator as follows. We fix the indirect inference sample parameters at $\tilde{\sigma}_0^2$, $\tilde{\sigma}_\eta^2$, $\tilde{\sigma}_{0\eta}$, and $\tilde{\sigma}_\varepsilon^2$. Then, for each ρ, s samples of size n: $\{y_i^k(\rho) : i = 1, \ldots, n\}$ $(k = 1, \ldots, s)$ are simulated and the 2-step GMM estimator of ρ, $h_k(\rho)$, is obtained for each sample $k = 1, \ldots, s$. The estimated binding function $b_s(\rho) = \frac{1}{s}\sum_{k=1}^{s} h_k(\rho)$ is used to determine the II estimator $\hat{\rho}^{II}$ of ρ:

$$\hat{\rho}^{II} = \arg\min_{\rho} |\hat{\rho} - b_s(\rho)|^2 .$$

We follow Gourieroux, Phillips, and Yu (2010) by setting $\tilde{\sigma}_0^2$, $\tilde{\sigma}_\eta^2$, $\tilde{\sigma}_{0\eta}$, and $\tilde{\sigma}_\varepsilon^2$ to the set of values σ_0^2, σ_η^2, $\sigma_{0\eta}$, and σ_ε^2 that govern the dynamics of the original sample. We later relax this for robustness checking. We set $s = 50$ throughout.

One of the main interests of the II estimator, as established by Gourieroux, Phillips, and Yu (2010), is its ability to reduce potential finite sample bias from the original estimator. However, if interest lies with CI's, in the case where $\rho = 1$ and $\sigma_\eta^2 = \sigma_{0\eta} = 0$, the standard asymptotic distribution fails and inference must be based on Theorem 2.

In each of our Monte Carlo experiments, CI's from the standard theory and that from our theory are considered. For the GMM and as already mentioned (see Remark 3), when the moment condition model has a single parameter that is not locally identified at the first order but rather at the second order, the asymptotic distribution of $n^{1/4}(\hat{\rho}-\rho)$ is a simple function of a Gaussian variable and CI's can be derived analytically using quantiles from the standard Gaussian distribution. However, in general, the asymptotic distribution of $n^{1/4}(\tilde{\rho}-\rho)$ (with $\tilde{\rho}$ being the GMM or II estimator) can be simulated and one can consider symmetric CI's based on the quantiles of the asymptotic distribution of $n^{1/4}(\tilde{\rho}-\rho)$ or the so-called equal-tailed CI's that use $\alpha/2$-quantile and $(1-\alpha/2)$-quantile of this asymptotic distribution. Throughout this section, only results from symmetric CI's are reported. Equal-tailed CI's have very similar performance and have not been reported. Simulated quantiles are obtained from 1,000 draws from the estimated asymptotic distribution of $n^{1/4}(\tilde{\rho}-\rho)$ where G and W are replaced by their estimates.

Table 1 gives the results related to GMM estimation and coverage rate of CI's based on GMM using the standard asymptotics (Cov-1) and our results (Cov-2 and Cov-3 using analytic and simulated quantiles, respectively). We take $\sigma_0^2 = \sigma_\varepsilon^2 = 1$ and $\sigma_\eta^2 = \sigma_{0\eta} = 0$ and consider $\rho = 0.2, 0.3, 0.5, 0.75, 0.8, 0.9, 0.95, 0.97, 0.98, 1.0, 1.1, 1.2, 1.3$ and 1.5. Even though first-order local identification issues arise at $\rho = 1.0$, this range of values for ρ allows us to investigate the finite sample performance of the non-standard CI near $\rho = 1.0$, i.e. near singularity of the Jacobian matrix of the moment function. The Euclidean norm of the simulated mean of this Jacobian, $D(\rho)$ in (31), and the second derivative matrix, $G(\rho)$ in (32), are also reported as $|\bar{\hat{D}}|$ and $|\bar{\hat{G}}|$ respectively.

First, we observe that Cov-2 and Cov-3 have approximately the same values meaning that the non-standard CI's based on simulation or on quantiles from the standard normal distribution are almost identical. Besides, for values of ρ 'far' from singularity point ($\rho = 0.2, 0.3, 0.5, 1.3$ and 1.5), the coverage rates of the standard CI (Cov-1), seems to converge to the nominal level 95% as n becomes large whereas the non-standard CI substantially over-covers at those values for n large with coverage rates larger than 99%.

However, near $\rho = 1.0$, the non-standard CI has coverage rates of about nominal 95% while the standard CI substantially under-covers at around 82% for $n = 5,000$. Specifically, the standard CI performs poorly for ρ ranging from 0.8 through 1.2 even in large samples. In small samples ($n = 50, 100, 200$), except for $\rho = 1.3$ and 1.5, the non-standard CI seems to outperform the standard one as it delivers coverage rates substantially closer to nominal. It is worth mentioning that the smaller the Jacobian norm ($|\bar{\hat{D}}|$), the better the non-standard CI performs.

Figure 1 reports, for $n = 5,000$, histograms of the simulated GMM estimators for $\rho = 0.3, 1.0, 1.3$ and also QQ-plots of these distributions against the standard normal distribution. These reveal that the GMM estimator has a very different distribution for $\rho = 1$, the point of first-order identification failure than at the other two points at which ρ is first-order locally identified. The distribution for $\rho = 1$ is also evidently non-normal.

To explore the behaviour of the estimator in a different neighbourhood to the point of first-order local identification failure, we fix $\rho = 1$ and set $\sigma_{0\eta} = \lambda$, $\sigma^2_\eta = |\lambda|$, with $\lambda = 0, \pm 0.1, 0.2, 0.3, 0.5$. These results are reported in Table 2. Qualitatively, the results are the same as in Table 1: the CI's based on Theorem 1 have approximately the nominal coverage level for λ values close to 0, the point of first-order local identification failure but the coverage is too high outside this neighbourhood. In contrast, the coverage of the CI based on the standard theory is well below the nominal 95% level in this neighbourhood: for example at $n = 5,000$, the coverage is between 78%, 82% and 84% for $\lambda = -0.1, 0, 0.1$, respectively.

Table 3 reports analogous results for II to those for Table 1 for GMM. These results indicate that the CI's based on standard asymptotic theory are too low at and in the neighbourhood of the point of first-order local identification failure whereas the coverage for the CI's based on Theorem 2 are closer to the nominal level although only achieve the nominal level at the largest sample size. Comparing the GMM and II CI's based on our theory for parameter values in the neighbourhood of the point of first-order local identification failure, it can be seen that the coverage rates for GMM tend to be closer to the nominal level than those for II. As noted above, one reason for employing II is bias reduction. In Table 3 we report the simulated bias and RMSE of the two estimators. From Table 3, it can be seen that for $\rho \leq 1$, our II estimator exhibits lower bias; but for $\rho > 1$ GMM exhibits less bias.

The simulated distribution of the II estimator is displayed by Figure 2 which is the II analogue of Figure 1. We can see that at singularity ($\rho = 1$), II is also clearly non-normal whereas as $\rho = 0.3$ or 1.3, related histograms and QQ-plots reveal a behaviour of II in line with normality.

We conclude this simulation experiments by investigating the robustness of the properties of the II estimator. In this experiment, we assume the researcher uses (11) as the auxiliary model but calibrates the value of θ_2. So for the true data generation process: $\rho = 1$, $\sigma^2_\varepsilon = \sigma^2_0 = 1$ and $\sigma_{0\eta} = \sigma^2_\eta = 0$; but the calibrated values of θ_2 are: $\tilde{\sigma}^2_\varepsilon = \tilde{\sigma}^2_0 = 1$, $\tilde{\sigma}_{0\eta} = \lambda$ and $\tilde{\sigma}^2_\eta = |\lambda|$, $\lambda = 0, \pm 0.1, 0.2$. Note that due to the calibration, only ρ is estimated via II. The results are displayed in Table 4. The CI based on Theorem 2 outperforms the standard CI for all values of λ and for all the sample sizes considered. We can also see that for $n = 5,000$, the coverage rates from the non-standard CI are all close to nominal except for $\lambda = -0.1$ where the coverage rate is 80.1%. Note that this still outperforms the standard CI by about 4 percentage points. Besides, for each sample size, we measure the stability of coverage across λ's by the mean absolute deviation (MAD) from nominal, 95%. Thanks to this metric, we can see that the non-standard CI has a better robustness property since its MAD varies from 13.5 ($n = 200$) to 3.36 ($n = 5,000$) and is always smaller than that of the standard CI which lies between 10.3 and 20.5. It is also worth mentioning that the bias reduction property expected for II is also robust to the deviations considered for the II samples since this estimator has a smaller bias than GMM across λ's.

7 Concluding remarks

In this paper, we provide new results on the limiting behaviour of GMM and II estimators when first-order identification fails but the parameters are second-order identified. These limit distributions are shown to be non-standard, but we show that they can be easily simulated, making it possible to perform inference about the parameters in this setting. An implication of our results is that the limiting distributions of GMM and II are different under first-order and second-order identification. While first-order local identification may only fail at a point in the parameter space, our simulation results indicate that our theory based on second-order identification provide a better approximation to finite sample behaviour of GMM and II estimators than standard first order asymptotic theory in a neighbourhood of the point of first-order local identification failure. Our simulation study further reveals that the limiting distribution theory derived in our paper leads to reliable GMM/II-based inferences in moderate to large samples in the neighbourhood of the point of first-order identification failure. Comparing GMM and II, we find our limiting distribution theory provides a reasonable approximation to the behaviour of the GMM at smaller sample sizes than it does for the II estimator, but that II exhibits smaller bias at the point of first-order local identification failure.

The choice of limit theory then requires knowledge of the quality of the identification but this may be difficult to assess *a priori*. It would be interesting to explore diagnostics for cases when local identification fails at first order but not at second order. Such diagnostics for local identification have recently been receiving some attention in the context of DSGE models. Iskrev (2010) and Qu and Tkachenko (2012) develop methods for evaluating the first-order local identification based on numerical evaluation of the Jacobian over the parameter space. An attractive feature of such analyses is that can reveal areas of the parameter space where first-order identification fails. By their nature, these methods focus on first-order identification. However, we conjecture that, given the complexity of the models and the need for approximations to their solutions, parameters of DSGE models may be second-order but not first-order locally identified in some cases of interest. The results presented in our paper provide a basis for performing inference about the parameters in this context. It would therefore be interesting to explore extensions of these diagnostics to look for evidence of second-order local identification.

Alternatively, it may be of interest to explore ways to generate confidence sets based on GMM and II estimators that are robust to first- or second- order identification. One possible approach may be the use of bootstrap methods, building from recent work on bootstrapping the GMM overidentification test by Dovonon and Gonçalves (2014).

References

Andrews, D. W. K. (1991). 'Heteroscedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, 59: 817–858.

Arellano, M., and Bond, S. R. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies*, 58: 277–297.

Arellano, M., Hansen, L. P., and Sentana, E. (2012). 'Underidentification?', *Journal of Econometrics*, 138: 256–280.

Azzalini, A. (2005). 'The skew-normal distribution and related multivariate families', *Scandanavian Journal of Statistics*, 32: 159–188.

Barankin, E., and Gurland, J. (1951). 'On asymptotically normal efficient estimators: I', *University of California Publications in Statistics*, 1: 86–130.

Blundell, R., and Bond, S. R. (1998). 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics*, 87: 115–143.

Canova, F., and Sala, L. (2009). 'Back to square one: identification issues in DSGE models', *Journal of Monetary Economics*, 56: 431–449.

Christiano, L., Eichenbaum, M., and Evans, C. (2005). 'Nominal rigidities and the dynamic effects of a shock to monetary policy', *Journal of Political Economy*, 113: 1–45.

Coenen, G., Levin, A. T., and Christoffel, K. (2007). 'Identifying the influences of nominal and real rigidities in aggregate price-setting behavior', *Journal of Monetray Economics*, 54: 2439–2466.

Diebold, F. X., and Nerlove, M. (1989). 'The dynamics of exchange rate volatility: a multivariate latent factor ARCH model', *Journal of Applied Econometrics*, 4: 1–22.

Dovonon, P., and Gonçalves, S. (2014). 'Bootstrapping the GMM overidentification test under first-order underidentification', Discussion paper, Department of Economics, Concordia University, Montreal, Canada.

Dovonon, P., and Hall, A. R. (2015). 'GMM and Indirect Inference - an appraisal of their connections and new results on their properties under second order identification', Discussion paper, Department of Economics, University of Manchester, Discussion paper EDP-1505.

Dovonon, P., and Renault, E. (2009). 'GMM overidentification test with first order underidentification', Discussion paper, Department of Economics, Concordia University, Montreal, Canada.

——— (2013). 'Testing for common conditionally heteroscedastic factors', *Econometrica*, 81: 2561–2586.

Doz, C., and Renault, E. (2004). 'Conditionally heteroskedastic factor models: Identification and instrumental variables estimation', Discussion paper, University Cergy-Pontoise, THEMA, France, 2004-13.

——— (2006). ‘Factor volatility in mean models: a GMM approach’, *Econometric Reviews*, 25: 275–309.

Duffie, D., and Singleton, K. J. (1993). ‘Testing for common conditionally heteroscedastic factors’, *Econometrica*, 61: 929–952.

Dufour, J.-M. (1997). ‘Some impossibility theorems in econometrics with applications to structural and dynamic models’, *Econometrica*, 65: 1365–1387.

Dufour, J.-M., Khalaf, L., and Kichian, M. (2013). ‘Identification-robust analysis of DSGE and structural macroeconomic models’, *Journal of Monetary Economics*, 60: 340–350.

Dupaigne, M., Fève, P., and Matheron, J. (2007). ‘Avoiding pitfalls in using structural VARs to estimate economic models’, *Review of Economic Dynamics*, 10: 238–255.

Ferguson, T. S. (1958). ‘A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities’, *Annals of Mathematical Statistics*, 29: 1046–1062.

Fiorentini, G., Sentana, E., and Shephard, N. (2004). ‘Likelihood-based estimation of generalised ARCH structures’, *Econometrica*, 72: 1481–1517.

Frazier, D. T., and Renault, E. (2016). ‘Indirect inference with(out) constraints’, Discussion paper, Department of Economics, University of Econometrics and Business Statistics, Monash University, Melbourne, New South Wales, Australia.

Gallant, A. R., and Tauchen, G. (1996). ‘Which moments to match?’, *Econometric Theory*, 12: 657–681.

Garcia, R., Renault, E., and Veredas, D. (2011). ‘Estimation of stable distributions by indirect inference’, *Journal of Econometrics*, 161: 325–337.

Ghysels, E., and Guay, A. (2003). ‘Structural change tests for simulated method of moments’, *Journal of Econometrics*, 115: 91–123.

Gourieroux, C., Monfort, A., and Renault, E. (1993). ‘Indirect inference’, *Journal of Applied Econometrics*, 8: S85–S118.

Gourieroux, C., Phillips, P. C. B., and Yu, J. (2010). ‘Indirect Inference for dynamic panel models’, *Journal of Econometrics*, 157: 68–77.

Hall, A. R. (2015). ‘Econometricians have their moments: GMM at 32’, *Economic Record*, 91, S1: 1–24.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press, Princeton, NJ, U. S. A.

Hansen, L. P. (1982). ‘Large sample properties of Generalized Method of Moments estimators’, *Econometrica*, 50: 1029–1054.

Hansen, L. P., and Singleton, K. S. (1982). ‘Generalized instrumental variables estimation of nonlinear rational expectations models’, *Econometrica*, 50: 1269–1286.

Heaton, J. (1995). ‘An empirical investigation of asset pricing with temporally dependent preference specifications’, *Econometrica*, 63: 681–717.

Iskrev, N. (2010). ‘Local identification in DSGE models’, *Journal of Monetary Economics*, 57: 189–202.

Jansen, I., and *et al* (2006). ‘The nature of sensitivity in monotone missing not at random models’, *Computational Statistics and Data Analysis*, 50: 830–858.

Kleibergen, F. (2005). ‘Testing parameters in GMM without assuming that they are ideintified’, *Econometrica*, 73: 1103–1124.

Komunjer, I., and Ng, S. (2011). ‘Dynamic identification of dynamic stochastic general equilibrium models’, *Econometrica*, 79: 1995–2032.

Kruiniger, H. (2013). ‘Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions’, *Journal of Econometrics*, 173: 175–188.

——— (2014). ‘A further look at Modied ML estimation of the panel AR(1) model with xed eects and arbitrary initial conditions’, Discussion paper, University of Durham, unpublished mimeo.

Le, V. P. M., and *et al* (2011). ‘How much nominal rigidity is there in the US economy? Testing a new Keynesian DSG Emodel using indirect inference’, *Journal of Economic Dynamics and Control*, 35: 2078–2104.

Lee, J. H., and Liao, Z. (2016). ‘On Standard Inference for GMM with Local Identication Failure of Known Forms’, Discussion paper, unpublished mimeo, Department of Economics, University of Illinois, Urbana IL, USA.

Madsen, E. (2009). ‘GMM-based inference in the AR(1) panel data model for parameter values where local idntification fails’, Discussion paper, Centre for Applied Microeconometrics, Department of Economics, University of Copenhagen, Copenhagen, Denmark.

McFadden, D. (1989). ‘A method of simulated moments for estimation of discrete response models without numerical integration’, *Econometrica*, 57: 995–1026.

Mutschler, W. (2015). ‘Identification of DSGE models - the effect of higher-order approximation and pruning’, *Journal of Economic Dynamics and Control*, 56: 34–54.

Newey, W. K., and McFadden, D. L. (1994). ‘Large sample estimation and hypothesis testing’, in R. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2247. Elsevier Science Publishers, Amsterdam, The Netherlands.

Newey, W. K., and Smith, R. J. (2004). ‘Higher order properties of GMM and generalized empirical likelihood estimators’, *Econometrica*, 72: 219–256.

Neyman, J. (1949). ‘Contribution to the theory of the χ^2 test’, in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 239–273. University of California Press, Berkeley, CA, USA.

Neyman, J., and Pearson, E. S. (1928). ‘On the use and interpretation of certain test criteria for purposes of statistical inference: part II’, *Biometrika*, 20A: 263–294.

Pearson, K. S. (1894). ‘Contributions to the mathematical theory of evolution’, *Philosophical transactions of the Royal Society of London (A)*, 185: 71–110.

——— (1895). 'Contributions to the mathematical theory of evolution, II: skew variation', *Philosophical transactions of the Royal Society of London (A)*, 186: 343–414.

Qu, Z. (2014). 'Inference in dynamic stochastic general equilibrium models with possible weak identification', *Quantitative Economics*, 3: 95–132.

Qu, Z., and Tkachenko, D. (2012). 'Identification and frequency domain quasi-mximum likelihood estimation of linearized dynamic stochastic general equilibrium models', *Quantitative Economics*, 3: 95–132.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). 'Likelihood-based inference with singular information matrix', *Bernouilli*, 6: 243–284.

Ruge-Murcia, F. J. (2007). 'Methods to estimate dynamic stochastic general equilibrium models', *Journal of Economic Dynamics and Control*, 31: 2599–2636.

Sargan, J. D. (1983). 'Identification and lack of identification', *Econometrica*, 51: 1605–1633.

Sentana, E. (2015). 'Finite Underidentification', Discussion paper, CEMFI Working Paper 1508, Madrid, Spain.

Smith, A. A. (1993). 'Estimating nonlinear time series models using simulated vector autoregressions', *Journal of Applied Econometrics*, 8: S63–S84.

Staiger, D., and Stock, J. (1997). 'Instrumental variables regression with weak instruments', *Econometrica*, 65: 557–586.

Stingo, F. C., Stanghellini, E., and Capobianco, R. (2011). 'On the estimation of a binary response model in a selected population', *Journal of Statistical Planning and Inference*, 141: 3293–3303.

Stock, J., and Wright, J. (2000). 'GMM with weak identification', *Econometrica*, 68: 1055–1096.

The Royal Swedish Academy of Sciences (2013). *Prizes in Economic Sciences 2013, Scientific Background.* http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/advanced-economicsciences2013.pdf.

Wright, J. H. (2003). 'Detecting Lack of Identification in GMM', *Econometric Theory*, 19: 322–330.

A Examples from Section 3

Example 1: panel data model
The moment condition in (11) can be derived using the assumptions in the text along with the following equation, implied by (10), that holds for all $i = 1, \dots, n$ and $t = 1, 2$:

$$y_{it} = \rho^t y_{i0} + \sum_{s=0}^{t-1} \rho^s \eta_i + \sum_{s=1}^{t} \rho^{t-s} \varepsilon_{is}.$$

It can be shown that the moment condition in (11) globally identify θ if the data generating process is such that the true parameter value θ^* satisfies $\sigma^*_{0\eta} \neq (1-\rho^*)\sigma^{*^2}_0$. Nevertheless, (11) also ensures global identification of θ if $\rho^* = 1$ and $\sigma^*_{0\eta} = \sigma^*_\eta = 0$; that is when the AR(1) panel dynamics has unit root and no fixed effects.

The Jacobian matrix of this moment function is:

$$\begin{pmatrix} -\dfrac{\partial H(\rho)}{\partial \rho}\theta_2 & -H(\rho) \end{pmatrix}.$$

As shown by Madsen (2009), see also Dovonon and Gonçalves (2015), if $\rho^* = 1$ and $\sigma^*_{0\eta} = \sigma^{*^2}_\eta = 0$, this Jacobian matrix has rank $4 < 5$ at the true parameter value so that the moment condition model (11) is first-order locally under identified. In fact, it can be seen that $H(\rho)$ is of rank 4 for any ρ and $\frac{\partial H(1)}{\partial \rho}\theta^*_2 = H(1)\delta$ with $\delta = (0, \sigma^{*^2}_\varepsilon, \sigma^{*^2}_0, -\sigma^{*^2}_\varepsilon)'$.

The statements about the identification of ρ based on (12) can be justified as follows. It can be shown that (12) globally identifies ρ so long as (11) globally identifies θ (see above). The Jacobian matrix associated with these moment conditions is:

$$D(\rho) = -\left(H^{(1)}_{1,\rho} - H_{1,\rho}H^{-1}_{2,\rho}H^{(1)}_{2,\rho}\right)\theta_2(\rho), \tag{31}$$

where $H_{k,\rho} = H_k(\rho)$ and $H^{(j)}_{k,\rho} = \frac{\partial^j H_k(\rho)}{\partial \rho^j}$. Since $\frac{\partial H(1)}{\partial \rho}\theta^*_2 = H(1)\delta$, we have: $D(1) = 0$. Some straightforward calculations show that the second-order derivative of the moment function in (12) is:

$$G(\rho) = -\left(H^{(2)}_{1,\rho} - H_{1,\rho}H^{-1}_{2,\rho}H^{(2)}_{2,\rho}\right)\theta_2(\rho) + 2\left(H^{(1)}_{1,\rho} - H_{1,\rho}H^{-1}_{2,\rho}H^{(1)}_{2,\rho}\right)H^{-1}_{2,\rho}H^{(1)}_{2,\rho}\theta_2(\rho), \tag{32}$$

and $G(1) \neq 0$ in general. $\diamond$

Example 2: a conditional heteroscedastic factor model
Auxiliary Model: There exists δ such that $\begin{pmatrix} 1 & -\delta \end{pmatrix}\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = 0$. Hence, $y_{1t} - \delta y_{2t} = u_{1t} - \delta u_{2t}$. We therefore have:

$$E\left[(y_{1t} - \delta y_{2t})^2 | \mathfrak{F}_{t-1}\right] = c (= \Omega_1 + \delta^2 \Omega_2).$$

Taking an instrument z_{t-1} from $\mathfrak{F}_{t-1}$ such that $Cov[z_{t-1}, y^2_{2t}] \neq 0$ and $E[z_{t-1}] \neq 0$, (e.g lagged square returns), we have:

$$m_0(z_{t-1}, y_t, \delta, c) = 0,$$

with $m_0(z_{t-1}, y_t, \delta, c) = E\left[\begin{pmatrix} 1 \\ z_{t-1} \end{pmatrix}\left\{(y_{1t} - \delta y_{2t})^2 - c\right\}\right]$.

We can show that this model identifies globally both δ and c. We also have:

$$E[y_{1t}^2] = \gamma_1^2 + \Omega_1 \equiv b_1, \quad E[y_{1t}^2] = \gamma_1^2 + \Omega_2 \equiv b_2, \quad \text{and} \quad E[y_{1t}y_{1,t-1}] = \gamma_1\gamma_2 \equiv b_3.$$

The auxiliary model is defined as:

$$\begin{array}{rcl} m_0(z_{t-1}, y_t, \delta, c) & = & 0 \\ E\left[y_{1t}^2\right] & = & b_1 \\ E\left[y_{2t}^2\right] & = & b_2 \\ E\left[y_{1t}y_{2t}\right] & = & b_3. \end{array} \tag{33}$$

The parameter vector $h = (b_1, b_2, b_3, \delta, c)'$ of this model is globally identified. In addition, the parameter θ of the structural model can be determined from h. In fact, we can use the relations:

$$b_1 = \gamma_1^2 + \Omega_1, \quad b_2 = \gamma_2^2 + \Omega_2, \quad b_3 = \gamma_1\gamma_2, \quad c = \Omega_1 + \delta^2\Omega_2, \quad \text{and} \quad c = b_1 + \delta^2 b_2 - 2\delta b_3$$

to obtain:

$$\theta_1 \equiv \gamma_1 = \sqrt{\delta b_3}, \quad \theta_2 \equiv \gamma_2 = \sqrt{\frac{b_3}{\delta}}, \quad \theta_3 \equiv \Omega_1 = b_1 - \delta b_3, \quad \theta_4 \equiv \Omega_2 = b_2 - \frac{b_3}{\delta}.$$

The auxiliary model is first-order locally underidentified: The Jacobian matrix of

$$m_0(z_{t-1}, y_t, \delta, c)$$

at the true parameter value is:

$$-2E\left[\begin{pmatrix} 1 \\ z_{t-1} \end{pmatrix} y_{2t}\left(y_{1t} - \delta y_{2t}\right)\right] \quad -\begin{pmatrix} 1 \\ E[z_t] \end{pmatrix}.$$

At the true parameter value, $y_{1t} - \delta y_{2t} = u_{1t} - \delta u_{2t}$. Therefore, $E\left[y_{2t}(y_{1t} - \delta y_{2t})|\mathfrak{F}_{t-1}\right] = -\delta\Omega_2$. (Since $y_{2t}(y_{1t} - \delta y_{2t}) = \gamma_2 f_t(u_{1t} - \delta u_{2t}) + u_{2t}(u_{1t} - \delta u_{2t})$.) Thus, By the law of iterated expectations, this Jacobian matrix is:

$$\left(2\delta\Omega_2 \begin{pmatrix} 1 \\ E[z_{t-1}] \end{pmatrix} \quad - \begin{pmatrix} 1 \\ E[z_{t-1}] \end{pmatrix}\right)$$

which is of rank 1. In total, the Jacobian matrix of the auxiliary model is

$$\begin{pmatrix} 0 & 0 & 0 & 2\delta\Omega_2 & -1 \\ 0 & 0 & 0 & 2\delta\Omega_2 E[z_{t-1}] & -E[z_{t-1}] \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}$$

which is of rank 4 instead of 5.

The auxiliary model is second-order locally identified. To see this, we can check Condition (b) of Definition 1 by focusing solely on the first equality of (33). Let $\phi = (\delta, c)'$. The range space of $\frac{\partial m_0'}{\partial \phi}(\phi_0)$ is determined by $u = a\,(2\delta\Omega_2, -1)'$: $a \in \mathbb{R}$ and the null space of its transpose is determined by $v = b\,(E(z_{t-1}), -1)'$: $b \in \mathbb{R}$. Also,

$$\frac{\partial^2 m_{0,1}}{\partial \delta^2} \equiv 2E[y_{2t}^2], \quad \frac{\partial^2 m_{0,1}}{\partial \delta \partial c} \equiv 0, \quad \frac{\partial^2 m_{0,1}}{\partial c^2} \equiv 0$$

and

$$\frac{\partial^2 m_{0,2}}{\partial \delta^2} \equiv 2E[z_{t-1} y_{2t}^2], \quad \frac{\partial^2 m_{0,2}}{\partial \delta \partial c} \equiv 0, \quad \frac{\partial^2 m_{0,2}}{\partial c^2} \equiv 0.$$

Hence,

$$\begin{aligned} & \tfrac{\partial m_0}{\partial \phi'}(\phi_0) u + \left(v' \tfrac{\partial^2 m_{0,k}}{\partial \phi \partial \phi'} v \right)_{k=1,2} \\ = & \begin{pmatrix} a\left(4\delta^2 \Omega_2^2 + 1\right) + 2b^2 \left(E[z_{t-1}]\right)^2 E[y_{2t}^2] \\ \\ a\left(4\delta^2 \Omega_2^2 + 1\right) E[z_{t-1}] + 2b^2 (E[z_{t-1}])^2 E[y_{2t}^2 z_{t-1}] \end{pmatrix} \end{aligned}$$

which is equal to 0 if and only if $a = b = 0$, i.e. $u = v = 0$; so long as $Cov[y_{2t}^2, z_{t-1}] \neq 0$ and $E[z_{t-1}] \neq 0$. $\diamond$

B Proofs

Proof of Theorem 1 (a) We write $m_T(\hat{\phi}) = m_T(\hat{\phi}_1, \hat{\phi}_{p_\phi})$. A first-order mean value expansion of $\phi_1 \mapsto m_T(\phi_1, \hat{\phi}_{p_\phi})$ around $\phi_{0,1}$ yields:

$$m_T(\hat{\phi}_1, \hat{\phi}_{p_\phi}) = m_T(\phi_{0,1}, \hat{\phi}_{p_\phi}) + \frac{\partial m_T}{\partial \phi_1'}(\bar{\phi}_1, \hat{\phi}_{p_\phi})(\hat{\phi}_1 - \phi_{0,1}),$$

where $\bar{\phi}_1 \in (\phi_{0,1}, \hat{\phi}_1)$ and may differ from row to row. Next, a second-order mean value expansion of $\phi_{p_\phi} \mapsto m_T(\phi_{0,1}, \phi_{p_\phi})$ around ϕ_{0,p_ϕ} that we plug back in the expression of $m_T(\hat{\phi})$ yields:

$$\begin{aligned} m_T(\hat{\phi}) &= m_T(\phi_0) + \tfrac{\partial m_T}{\partial \phi_1'}(\bar{\phi}_1, \hat{\phi}_{p_\phi})(\hat{\phi}_1 - \phi_{0,1}) + \tfrac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) \\ & \quad + \tfrac{1}{2} \tfrac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_{0,1}, \bar{\phi}_{p_\phi})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2, \end{aligned}$$

where $\bar{\phi}_{p_\phi} \in (\phi_{0,p_\phi}, \hat{\phi}_{p_\phi})$ and may differ from row to row.

Since $\frac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0) = O_P(T^{-1/2})$ and $\hat{\phi}_{p_\phi} - \phi_{0,p_\phi} = o_P(1)$ (from Proposition 1), we have:

$$\begin{aligned} m_T(\hat{\phi}) &= m_T(\phi_0) + \tfrac{\partial m_T}{\partial \phi_1'}(\bar{\phi}_1, \hat{\phi}_{p_\phi})(\hat{\phi}_1 - \phi_{0,1}) \\ & \quad + \tfrac{1}{2} \tfrac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_{0,1}, \bar{\phi}_{p_\phi})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(T^{-1/2}). \end{aligned} \tag{34}$$

Let us define $\bar{D} = \frac{\partial m_T}{\partial \phi_1'}(\bar{\phi}_1, \hat{\phi}_{p_\phi})$ and $\bar{G} = \frac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_{0,1}, \bar{\phi}_{p_\phi})$. Pre-multiplying (34) by $\bar{D}' W_T$, we get

$$\begin{aligned} \hat{\phi}_1 - \phi_{0,1} &= \left(\bar{D}' W_T \bar{D}\right)^{-1} \bar{D}' W_T \left(m_T(\hat{\phi}) - m_T(\phi_0) \right) \\ & \quad - \tfrac{1}{2} \left(\bar{D}' W_T \bar{D}\right)^{-1} \bar{D}' W_T \bar{G} (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(T^{-1/2}). \end{aligned} \tag{35}$$

The $o_P(T^{-1/2})$ term stays with the same order because $\bar{D}$ and W_T are both $O_P(1)$. Plugging this back into (34), we get:

$$\begin{aligned} m_T(\hat{\phi}) &= m_T(\phi_0) + \bar{D} \left(\bar{D}' W_T \bar{D}\right)^{-1} \bar{D}' W_T \left(m_T(\hat{\phi}) - m_T(\phi_0) \right) \\ & \quad + \tfrac{1}{2} W_T^{-1/2} \bar{M}_d W_T^{1/2} \bar{G} (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(T^{-1/2}), \end{aligned}$$

with $\bar{M}_d = I_q - W_T^{1/2}\bar{D}\left(\bar{D}'W_T\bar{D}\right)^{-1}\bar{D}'W_T^{1/2}$.

Hence,

$$
\begin{aligned}
& m_T'(\hat{\phi})W_T m_T(\hat{\phi}) \\
= \quad & m_T'(\phi_0)W_T m_T(\phi_0) + \tfrac{1}{4}\bar{G}'W_T^{1/2}\bar{M}_d W_T^{1/2}\bar{G}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 \\
& + (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 O_P(T^{-1/2}) + O_P(T^{-1})
\end{aligned}
\tag{36}
$$

The orders of magnitude in (36) follow from the fact that $\bar{M}_d$ converges in probability to M_d and therefore is $O_P(1)$ and the fact that both $m_T(\phi_0)$ and $m_T(\hat{\phi})$ are $O_P(T^{-1/2})$.
The latter comes from the fact that $m_T'(\hat{\phi})W_T m_T(\hat{\phi}) \leq m_T'(\phi_0)W_T m_T(\phi_0)$ (by definition of GMM estimator). Since W_T converges in probability to W symmetric positive definite, we can claim that $m_T(\hat{\phi})$ is $O_P(T^{-1/2})$ as is $m_T(\phi_0)$. Again, by the definition of the GMM estimator, the left hand side of (36) is less or equal to $m_T'(\phi_0)W_T m_T(\phi_0)$ and this gives:

$$
\begin{aligned}
& \tfrac{1}{4}G'W^{1/2}M_d W^{1/2}GT(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 + o_P(1)T(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 \\
\leq \quad & O_P(1) + \sqrt{T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 O_P(1)
\end{aligned}
\tag{37}
$$

Thanks to Assumption 4(ii) and the fact that W is nonsingular, $M_d W^{1/2}G \neq 0$. As a consequence, $G'W^{1/2}M_d W^{1/2}G \neq 0$ which is sufficient to deduce from (37) that $T(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 = O_P(1)$; or equivalently that $T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) = O_P(1)$. We obtain $\hat{\phi}_1 - \phi_{0,1} = O_P(T^{-1/2})$ from (35).

(b) From (a) and (34), we have

$$
m_T(\hat{\phi}) = m_T(\phi_0) + D(\hat{\phi}_1 - \phi_{0,1}) + \frac{1}{2}G(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(T^{-1/2}).
$$

The first-order condition for interior solution is given by:

$$
\frac{\partial m_T}{\partial \phi'}(\hat{\phi})W_T m_T(\hat{\phi}) = 0.
$$

In the direction of ϕ_1, this amounts to

$$
(D' + o_P(1))W\left(\sqrt{T}m_T(\phi_0) + D\sqrt{T}(\hat{\phi}_1 - \phi_{0,1}) + \frac{1}{2}G\sqrt{T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(1)\right) = 0.
$$

This gives:

$$
\sqrt{T}(\hat{\phi}_1 - \phi_{0,1}) = -(D'WD)^{-1}D'W\left(\sqrt{T}m_T(\phi_0) + \frac{1}{2}G\sqrt{T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2\right) + o_P(1). \tag{38}
$$

In the direction of ϕ_{p_ϕ}, the first-order condition amounts to

$$
\begin{aligned}
& \left(G'T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + o_P(1)\right) \\
\times \quad & W\left(\sqrt{T}m_T(\phi_0) + D\sqrt{T}(\hat{\phi}_1 - \phi_{0,1}) + \tfrac{1}{2}G\sqrt{T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(1)\right) = 0.
\end{aligned}
\tag{39}
$$

The terms in the first parentheses are obtained by a first-order mean value expansion of $\frac{\partial m_T}{\partial \phi_{p_\phi}}(\hat{\phi})$ around ϕ_0 and taking the limit. Plugging (38) into (39), we get:

$$
\begin{aligned}
& T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) \\
\times \quad & \left(G'W^{1/2}M_d W^{1/2}\sqrt{T}m_T(\phi_0) + \tfrac{1}{2}G'W^{1/2}M_d W^{1/2}G\sqrt{T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2\right) = o_P(1).
\end{aligned}
\tag{40}
$$

Since $\sqrt{T}m_T(\phi_0)$ and $T^{1/4}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})$ are $O_P(1)$, the pair is jointly $O_P(1)$ and by the Prohorov's theorem, any subsequence of them has a further subsequence that jointly converges in distribution towards, say, $(\mathbb{Z}_0,\mathbb{V}_0)$. From (40), $(\mathbb{Z}_0,\mathbb{V}_0)$ satisfies:

$$\mathbb{V}_0\left(\mathbb{Z}+\frac{1}{2}G'W^{1/2}M_dW^{1/2}G\mathbb{V}_0^2\right)=0,$$

almost surely with $\mathbb{Z}=G'W^{1/2}M_dW^{1/2}\mathbb{Z}_0$. Clearly, if $\mathbb{Z}\geq 0$, then, $\mathbb{V}_0=0$, almost surely. Conversely, following the proof of Dovonon and Renault (2013, Proposition 3.2), we can show that if $\mathbb{Z}<0$, then $\mathbb{V}_0\neq 0$, almost surely, and hence $\mathbb{V}_0^2=-2\mathbb{Z}/G'W^{1/2}M_dW^{1/2}G$.

In either case, $\mathbb{V}_0^2=-2\frac{\mathbb{Z}\mathbb{I}(\mathbb{Z}<0)}{G'W^{1/2}M_dW^{1/2}G}(\equiv\mathbb{V})$ and is the limit distribution of the relevant subsequence of $\sqrt{T}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^2$. Hence, that subsequence of $(\sqrt{T}m_T(\phi_0),\sqrt{T}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^2)$ converges in distribution towards $(\mathbb{Z}_0,\mathbb{V})$. The fact that this limit does not depend on a specific subsequence means that the whole sequence converges in distribution to that limit. We use (38) to conclude.

Next, we establish (c). We recall that the result in (b) gives the asymptotic distribution of $\sqrt{T}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^2$. To get the asymptotic distribution of $T^{1/4}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})$, it suffices to characterize its sign. Following the approach of Rotnitzky, Cox, Bottai, and Robins (2000) for MLE, we can do this by expanding $m_T'(\hat{\phi})W_Tm_T(\hat{\phi})$ up to $o_P(T^{-5/4})$. Being of order $O_P(T^{-1})$, its $O_P(T^{-5/4})$ terms actually provide the sign of $(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})$; leading to the asymptotic distribution of $(\sqrt{T}(\hat{\phi}_1-\phi_{0,1}),\,T^{1/4}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi}))$. By a mean value expansion of $m_T(\hat{\phi})$ up to the third order, we have:

$$\begin{aligned}&m_T(\hat{\phi})\\ =\;&m_T(\phi_0)+\tfrac{\partial m_T}{\partial\phi_1'}(\phi_0)(\hat{\phi}_1-\phi_{0,1})+\tfrac{\partial m_T}{\partial\phi_p}(\phi_0)(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})+\tfrac{1}{2}\tfrac{\partial^2 m_T}{\partial\phi_{p_\phi}^2}(\phi_0)(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^2\\ &+\tfrac{\partial^2 m_T}{\partial\phi_{p_\phi}\partial\phi_1'}(\phi_0)(\hat{\phi}_1-\phi_{0,1})(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})+\tfrac{1}{6}\tfrac{\partial^3 m_T}{\partial\phi_{p_\phi}^3}(\dot{\phi})(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^3+O_P(T^{-1}),\end{aligned}$$

where $\dot{\phi}\in(\phi_0,\hat{\phi})$ and may differ from row to row. From Assumption 5(i), we obtain:

$$\begin{aligned}m_T(\hat{\phi})\;&=\;m_T(\phi_0)+D(\hat{\phi}_1-\phi_{0,1})+\tfrac{\partial m_T}{\partial\phi_{p_\phi}}(\phi_0)(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})+\tfrac{1}{2}G(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^2\\ &\quad+G_{1p_\phi}(\hat{\phi}_1-\phi_{0,1})(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})+\tfrac{1}{6}L(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^3+o_P(T^{-3/4}).\end{aligned}$$

Hence, defining $Z_{0T}=m_T(\phi_0)$ and $Z_{1T}=\frac{\partial m_T}{\partial\phi_{p_\phi}}(\phi_0)$, it follows that:

$$\begin{aligned}m_T(\hat{\phi})\;&=\;Z_{0T}+D(\hat{\phi}_1-\phi_{0,1})+Z_{1T}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})+\tfrac{1}{2}G(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^2\\ &\quad+G_{1p_\phi}(\hat{\phi}_11-\phi_{0,1})(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})+\tfrac{1}{6}L(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})^3+o_P(T^{-3/4}).\end{aligned}\tag{41}$$

The first-order condition for the $\hat{\phi}$ in the direction of ϕ_1 is:

$$0=\tfrac{\partial m_T'}{\partial\phi_1}(\hat{\phi})W_Tm_T(\hat{\phi})=\left(D+G_{1p_\phi}(\hat{\phi}_{p_\phi}-\phi_{0,p_\phi})\right)'Wm_T(\hat{\phi})+o_P(T^{-3/4})\tag{42}$$

Substituting (41) into (42), we obtain an equation of the form:

$$D'WZ_{0T}+D'WD(\hat{\phi}_1-\phi_{0,1})+\text{other terms}\;=\;o_P(T^{-3/4}).\tag{43}$$

This equation can be solved to yield:

$$\hat{\phi}_1-\phi_{0,1}\;=\;-(D'WD)^{-1}\left(D'WZ_{0T}+\text{other terms}\right)+o_P(T^{-3/4}).\tag{44}$$

Note that "other terms" contains $\hat{\phi}_1 - \phi_{0,1}$ involved in quadratic functions of $\hat{\phi} - \phi_0$. Replacing in "other terms" $\hat{\phi}_1 - \phi_{0,1}$ by this expression and keeping only the leading terms (up to $O_P(T^{-3/4})$), we obtain the following expression for $\hat{\phi}_1 - \phi_{0,1}$:

$$
\begin{aligned}
\hat{\phi}_1 - \phi_{0,1} &= H\Big(Z_{0T} + (Z_{1T} + G_{1p_\phi} H Z_{0T})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{2}G(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
&\quad + \left(\tfrac{1}{2}G_{1p_\phi} HG + \tfrac{1}{6}L\right)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3\Big) \\
&\quad + H_1\Big((Z_{0T} + DHZ_{0T})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{2}(DHG + G)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3\Big) \\
&\quad + o_P(T^{-3/4}) \\
&= H\Big(Z_{0T} + \tfrac{1}{2}G(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2\Big) \\
&\quad + \left(HZ_{1T} + HG_{1p_\phi} HZ_{0T} + H_1 Z_{0T} + H_1 DHZ_{0T}\right)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) \\
&\quad + \tfrac{1}{2}\left(H(G_{1p_\phi} HG + \tfrac{L}{3}) + H_1(DHG + G)\right)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + o_P(T^{-3/4}).
\end{aligned}
$$

with $H = -(D'WD)^{-1}D'W$ and $H_1 = -(D'WD)^{-1}G'_{1p_\phi}W$. Hence, for a natural definition of A_1, B_1 and C_1, $(\hat{\phi}^1 - \phi_0^1)$ has the form:

$$
(\hat{\phi}_1 - \phi_{0,1}) = A_1 + B_1(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + C_1(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + o_P(T^{-3/4}) \tag{45}
$$

Using (41), we have:

$$
\begin{aligned}
& m'_T(\hat{\phi})W_T m_T(\hat{\phi}) = m'_T(\hat{\phi})W m_T(\hat{\phi}) + o_P(T^{-5/4}) \\
= \; & Z'_{0T}WZ_{0T} + (\hat{\phi}_1 - \phi_{0,1})'D'WD(\hat{\phi}_1 - \phi_{0,1}) + \tfrac{1}{4}G'WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 \\
& + 2Z'_{0T}WD(\hat{\phi}_1 - \phi_{0,1}) + 2Z'_{0T}WZ_{1T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + Z'_{0T}WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
& + 2Z'_{0T}WG_{1p_\phi}(\hat{\phi}_1 - \phi_{0,1})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{3}Z'_{0T}WL(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 \\
& + 2(\hat{\phi}_1 - \phi_{0,1})'D'WZ_{1T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + (\hat{\phi}_1 - \phi_{0,1})'D'WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
& + 2(\hat{\phi}_1 - \phi_{0,1})'D'WG_{1p_\phi}(\hat{\phi}_1 - \phi_{0,1})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{3}(\hat{\phi}_1 - \phi_{0,1})'D'WL(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 \\
& + Z'_{1T}WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + G'WG_{1p_\phi}(\hat{\phi}_1 - \phi_{0,1})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 \\
& + \tfrac{1}{6}G'WL(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^5 + o_P(T^{-5/4}).
\end{aligned} \tag{46}
$$

Replacing $\hat{\phi}_1 - \phi_{0,1}$ by its expression from (45) into (46), the leading $O_P(T^{-1})$ term of $m'_T(\hat{\phi})W_T m_T(\hat{\phi})$ is obtained as $K_T(\hat{\phi}_{p_\phi})$ with

$$
\begin{aligned}
& K_T(\phi_{p_\phi}) \\
= \; & Z'_{0T}WZ_{0T} + \left(Z_{0T} + \tfrac{1}{2}G(\phi_{p_\phi} - \phi_{0p})^2\right)' H'D'WDH\left(Z_{0T} + \tfrac{1}{2}G(\phi_{p_\phi} - \phi_{0,p_\phi})^2\right) \\
& + \tfrac{1}{4}G'WG(\phi_{p_\phi} - \phi_{0,p_\phi})^4 + 2Z'_{0T}WDH(Z_{0T} + \tfrac{1}{2}G(\phi_{p_\phi} - \phi_{0,p_\phi})^2) \\
& + Z'_{0T}WG(\phi_{p_\phi} - \phi_{0,p_\phi})^2 + \left(Z_{0T} + \tfrac{1}{2}G(\phi_{p_\phi} - \phi_{0,p_\phi})^2\right)' H'D'WG(\phi_{p_\phi} - \phi_{0,p_\phi})^2.
\end{aligned}
$$

Hence,

$$K_T(\phi_{p_\phi}) = Z_{0T}'W^{1/2}M_dW^{1/2}Z_{0T} + Z_{0T}'W^{1/2}M_dW^{1/2}G(\phi_{p_\phi}-\phi_{0,p_\phi})^2 + \tfrac{1}{4}G'W^{1/2}M_dW^{1/2}G(\phi_{p_\phi}-\phi_{0,p_\phi})^4. \tag{47}$$

The next leading term in the expansion of $m_T'(\hat\phi)W_Tm_T(\hat\phi)$ is of order $O_P(T^{-5/4})$ and given by:

$$\begin{aligned}
&R_T = (\hat\phi_{p_\phi}-\phi_{0,p_\phi})\times\\
&\Big\{2A_1'D'WDB_1 + 2Z_{0T}'WDB_1 + 2Z_{0T}'WZ_{1T}\\
&+2Z_{0T}'WG_{1p_\phi}A_1 + 2A_1'D'WZ_{1T} + 2A_1'D'WG_{1p_\phi}A_1\\
&+(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2\Big(2A_1'D'WDC_1 + 2Z_{0T}'WDC_1 + \tfrac{1}{3}Z_{0T}'WL + B_1'D'WG\\
&+\tfrac{1}{3}A_1'D'WL + Z_{1T}'WG + G'WG_{1p_\phi}A_1\Big)\\
&+(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^4\left(C_1'D'WG + \tfrac{1}{6}G'WL\right)\Big\}
\end{aligned}$$

$$\begin{aligned}
&R_T = (\hat\phi_{p_\phi}-\phi_{0,p_\phi})\times\\
&\Big\{2Z_{0T}'H'D'WDB_1 + 2Z_{0T}'WDB_1 + 2Z_{0T}'WZ_{1T} + 2Z_{0T}'WG_{1p_\phi}HZ_{0T}\\
&+2Z_{0T}'H'D'WZ_{1T} + 2Z_{0T}'H'D'WG_{1p_\phi}HZ_{0T}\\
&+(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2\quad\Big(2Z_{0T}'H'D'WDC_1 + 2Z_{0T}'WDC_1 + \tfrac{1}{3}Z_{0T}'WL + B_1'D'WG\\
&+\tfrac{1}{3}Z_{0T}'H'D'WL + Z_{1T}'WG + G'WG_{1p_\phi}HZ_{0T} + G'H'D'WDB_1\\
&+G'H'D'WZ_{1T} + Z_{0T}'WG_{1p_\phi}HG + Z_{0T}'H'D'WG_{1p_\phi}HG + G'H'D'WG_{1p_\phi}HZ_{0T}\Big)\\
&+(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^4\quad\Big(C_1'D'WG + \tfrac{1}{6}G'WL + \tfrac{1}{2}G'H'D'WG_{1p_\phi}HG + G'H'D'WDC_1\\
&+\tfrac{1}{6}G'H'D'WL + \tfrac{1}{2}G'WG_{1p_\phi}HG\Big)\Big\} \equiv (\hat\phi_{p_\phi}-\phi_{0,p_\phi})\times 2R_{1T}.
\end{aligned}$$

Re-arranging the terms and using the fact that $M_dW^{1/2}D = 0$, we have:

$$\begin{aligned}
2R_{1T} = {}& 2Z_{0T}'W^{1/2}M_dW^{1/2}Z_{1T} + 2Z_{0T}'W^{1/2}M_dW^{1/2}G_{1p_\phi}HZ_{0T}\\
&+(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2\Big(\tfrac{1}{3}Z_{0T}'W^{1/2}M_dW^{1/2}L + Z_{1T}'W^{1/2}M_dW^{1/2}G\\
&+G'W^{1/2}M_dW^{1/2}G_{1p_\phi}HZ_{0T} + Z_{0T}'W^{1/2}M_dW^{1/2}G_{1p_\phi}HG\Big)\\
&+(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^4\Big(\tfrac{1}{6}G'W^{1/2}M_dW^{1/2}L + \tfrac{1}{2}G'W^{1/2}M_dW^{1/2}G_{1p_\phi}HG\Big).
\end{aligned} \tag{48}$$

We can check that the GMM estimator $\hat\phi_{p_\phi}$ as given by the first-order condition (40) is minimizer of

$K_T(\phi_{p_\phi})$. When $T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$ is not $o_P(1)$, this first-order condition determines

$$(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 = -2\frac{G'W^{1/2}M_dW^{1/2}Z_{0T}}{G'W^{1/2}M_dW^{1/2}G} + o_P(T^{-1/2})$$

but not the sign of $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$. Following the analysis of Rotnitzky, Cox, Bottai, and Robins (2000) for the maximum likelihood estimator, the sign of $\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}$ can be determined by the remainder R_T of the expansion of $m_T'(\hat{\phi})W_Tm_T(\hat{\phi})$. At the minimum, we expect R_T to be negative; i.e. $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$ and R_{1T} have opposite sign.

Hence,

$$T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) = (-1)^{B_T}T^{1/4}|\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}|,$$

with $B_T = \mathbb{I}(TR_{1T} \geq 0)$.

Plugging the expression of $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2$ into (48) and scaling by T, we can see, using the continuous mapping theorem, that TR_{1T} converges in distribution towards $\mathbb{R}_1$:

$$\begin{aligned}
\mathbb{R}_1 \quad = \quad & \mathbb{Z}_0'W^{1/2}M_{dg}W^{1/2}(\mathbb{Z}_1 + G_{1p_\phi}H\mathbb{Z}_0) \\
& + \left(\mathbb{Z}_0'W^{1/2}(M_d - M_{dg})W^{1/2}\mathbb{Z}_0G' - G'W^{1/2}M_dW^{1/2}\mathbb{Z}_0\mathbb{Z}_0'\right) \\
& \times W^{1/2}M_dW^{1/2}\left(\tfrac{1}{3}L + G_{1p_\phi}HG\right)/\sigma_G,
\end{aligned} \tag{49}$$

with $\sigma_G = G'W^{1/2}M_dW^{1/2}G$ and
$M_{dg} = M_d - M_dW^{1/2}G(G'W^{1/2}M_dW^{1/2}G)^{-1}G'W^{1/2}M_d$, the matrix of the orthogonal projection on the orthogonal of $\begin{pmatrix} W^{1/2}D & W^{1/2}G \end{pmatrix}$.

We actually have that: $(\sqrt{T}Z_{0T}, \sqrt{T}Z_{1T}, TR_{1T})$ converges in distribution towards $(\mathbb{Z}_0, \mathbb{Z}_1, \mathbb{R}_1)$. Applying Lemma 1, we have $(\sqrt{T}Z_{0T}, \sqrt{T}Z_{1T}, (-1)^{B_T}) \xrightarrow{d} (\mathbb{Z}_0, \mathbb{Z}_1, (-1)^{\mathbb{B}})$, where $\mathbb{B} = \mathbb{I}(\mathbb{R}_1 \geq 0)$.

Since $\left(\sqrt{T}(\hat{\phi}_1 - \phi_{0,1}),\ T^{1/4}|\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}|,\ (-1)^{B_T}\right) = O_P(1)$, any subsequence of the left hand side has a further subsequence that converges in distribution. Using (b), such subsequence satisfies:

$$\left(\sqrt{T}(\hat{\phi}_1 - \phi_{0,1}),\ T^{1/4}|\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}|,\ (-1)^{B_T}\right) \xrightarrow{d} \left(H\mathbb{Z}_0 + HG\mathbb{V}/2, \sqrt{\mathbb{V}}, (-1)^{\mathbb{B}}\right).$$

(We keep T to index the subsequence for simplicity.)

Since the limit distribution does not depend on the subsequence, the whole sequence converges towards that limit. By the continuous mapping theorem, we deduce that:

$$\left(\sqrt{T}(\hat{\phi}_1 - \phi_{0,1}),\ T^{1/4}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})\right) \xrightarrow{d} \left(H\mathbb{Z}_0 + HG\mathbb{V}/2, (-1)^{\mathbb{B}}\sqrt{\mathbb{V}}\right).$$

□

Lemma 1. *Let $(X_T)_T$ and $(Y_T)_T$ be two sequences of random variables and $B_T = \mathbb{I}(X_T \geq 0)$. If $(X_T, Y_T) \xrightarrow{d} (X, Y)$ and $P(X = 0) = 0$, then*

$$\left((-1)^{B_T}, Y_T\right) \xrightarrow{d} \left((-1)^B, Y\right),$$

with $B = \mathbb{I}(X \geq 0)$.

Proof of Lemma 1: Using the Cramer-Wold device, it suffices to show that: for all $(\lambda_1, \lambda_2) \in \mathbb{R} \times \mathbb{R}$,

$$\lambda_1(-1)^{B_T} + \lambda_2 Y_T \xrightarrow{d} \lambda_1(-1)^B + \lambda_2 Y.$$

Let $x \in \mathbb{R}$ be a continuity point of $F(x) = P(\lambda_1(-1)^B + \lambda_2 Y \leq x)$. We show that:

$$P\left(\lambda_1(-1)^{B_T} + \lambda_2 Y_T \leq x\right) \to F(x), \quad as \quad T \to \infty.$$

We have:

$$P\left(\lambda_1(-1)^{B_T} + \lambda_2 Y_T \leq x\right) = P(\lambda_2 Y_T \leq x - \lambda_1, X_T < 0) + P(\lambda_2 Y_T \leq x + \lambda_1, X_T \geq 0).$$

To complete the proof, it suffices to show that, as $T \to \infty$,

$$\begin{aligned} &P(\lambda_2 Y_T \leq x - \lambda_1, X_T < 0) \to P(\lambda_2 Y \leq x - \lambda_1, X < 0) \qquad and \\ &P(\lambda_2 Y_T \leq x + \lambda_1, X_T \geq 0) \to P(\lambda_2 Y \leq x + \lambda_1, X \geq 0) \end{aligned} \tag{50}$$

since $F(x) = P(\lambda_2 Y \leq x - \lambda_1, X < 0) + P(\lambda_2 Y \leq x + \lambda_1, X \geq 0)$.

We now establish the first condition in (50). The second one is obtained along the same lines. Note that $P(\lambda_2 Y_T \leq x - \lambda_1, X_T < 0) = P((\lambda_2 Y_T, X_T) \in A)$ with boundary of A given by: $\partial A = ((-\infty, x - \lambda_1] \times \{0\}) \bigcup (\{x - \lambda_1\} \times (-\infty, 0])$. Since (X_T, Y_T) converge jointly in distribution towards (X, Y), it suffices to show that

$$P((\lambda_2 Y, X) \in \partial A) = 0.$$

We have:

$$P\left((\lambda_2 Y, X) \in (-\infty, x - \lambda_1] \times \{0\}\right) \leq P(X = 0) = 0.$$

Besides,

$$P\left((\lambda_2 Y, X) \in \{x - \lambda_1\} \times (-\infty, 0]\right) = P(\lambda_2 Y = x - \lambda_1, X \leq 0).$$

By continuity of F at x, $P\left(\lambda_1(-1)^B + \lambda_2 Y = x\right) = 0$, i.e.

$$P(\lambda_2 Y = x + \lambda_1, X \geq 0) + P(\lambda_2 Y = x - \lambda_1, X < 0) = 0.$$

Thus, $P(\lambda_2 Y = x - \lambda_1, X < 0) = 0$. Since $P(X = 0) = 0$, we can claim that

$$P(\lambda_2 Y = x + \lambda_1, X \leq 0) = 0.$$

This completes the proof. $\square$

Proof of Equation (17)**:** First, we observe that the asymptotic distribution $\mathbb{V}$ of $T^{1/2}(\hat{\phi} - \phi_0)^2$ is continuous at any $c > 0$ (with $P(\mathbb{V} = 0) = 1/2$). Let us search for $c_{1-\alpha}$ such that $P(\mathbb{V} \leq c_{1-\alpha}) = 1 - \alpha$. Since $1 - \alpha > 1/2$, we have $c_{1-\alpha} > 0$ and $P(T^{1/4}|\hat{\phi} - \phi_0| \leq \sqrt{c_{1-\alpha}}) \to 1 - \alpha$, as $T \to \infty$. Hence, $\sqrt{c_{1-\alpha}}$ defines an asymptotically correct confidence interval for ϕ_0. To obtain $c_{1-\alpha}$, we recall that:

$$\begin{aligned} &P(\mathbb{V} \leq c_{1-\alpha}) = P\left(-\tfrac{2\mathbb{Z}\mathbb{I}(\mathbb{Z}<0)}{\sigma_G} \leq c_{1-\alpha}\right) \\ = \quad &P\left(-\tfrac{2\mathbb{Z}\mathbb{I}(\mathbb{Z}<0)}{\sigma_G} \leq c_{1-\alpha}, \mathbb{Z} \geq 0\right) + P\left(-\tfrac{2\mathbb{Z}\mathbb{I}(\mathbb{Z}<0)}{\sigma_G} \leq c_{1-\alpha}, \mathbb{Z} < 0\right) = P\left(\mathbb{Z} \leq \tfrac{\sigma_G}{2} c_{1-\alpha}\right). \end{aligned}$$

The last equality uses the fact that $\mathbb{Z}$ has a symmetric distribution about 0 as a zero mean Gaussian variable. Since $Z \sim N(0, G'W\Omega WG)$, $c_{1-\alpha}$ solves:

$$\frac{\sigma_G}{2\sqrt{G'W\Omega WG}} c_{1-\alpha} = z_\alpha$$

giving $c_{1-\alpha} = \frac{2\sqrt{G'W\Omega WG}}{\sigma_G} z_\alpha$.

When a consistent estimator $\hat{c}_{1-\alpha}$ of $c_{1-\alpha}$ is used as in the statement of (17), we can rely on the following two facts to conclude: (1) $\mathbb{V}$ has a continuous distribution at all $c > 0$ and (2) if X_n is a sequence of random variables converging in distribution to X with cumulative distribution function F_X that is continuous on an interval $[a, b]$, then the sequence of cumulative distributions F_{X_n} of X_n converges uniformly over $[a, b]$ to F_X. $\square$

Proof of Equation (21)**:** Since $\hat{\phi} - \phi_0 = R(\hat{\eta} - \eta_0)$, we have

$$T^{1/4}(\hat{\phi} - \phi_0) = R_{\bullet p_\phi} T^{1/4}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi}). \tag{51}$$

We also have

$$\epsilon_T = -B_T \hat{R}^{-1}(\hat{R} - R)R^{-1}(\hat{\phi} - \phi_0) = -B_T \hat{R}^{-1}(\hat{R} - R)(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi}).$$

But $\hat{R} \equiv R(\hat{\phi})$ and $R \equiv R(\phi_0)$. By mean value expansions, for $j = 1, \ldots, p_\phi$,

$$\hat{R}_{\bullet j} - R_{\bullet j} = \frac{\partial R_{\bullet j}}{\partial \phi'}(\dot{\phi}_j)(\hat{\phi} - \phi_0),$$

where $\dot{\phi}_j \in (\phi_0, \hat{\phi})$ and may differ from row to row and $R_{\bullet j}$ denotes the column vector corresponding to the jth column of the matrix R. We also use $R_{h\bullet}$ to denote the row vector corresponding to the hth row of R.

For $h = 1, \ldots, p_\phi - 1$,

$$\begin{aligned}
\epsilon_{T,h} &= -\sqrt{T}\left(\hat{R}^{-1}\right)_{h\bullet} \sum_{j=1}^{p_\phi} \left(\frac{\partial R_{\bullet j}}{\partial \phi'}(\dot{\phi}_j)(\hat{\phi} - \phi_0)\right)(\hat{\eta}_j - \eta_{0,j}) \\
&= -\left(\hat{R}^{-1}\right)_{h\bullet} \sum_{j=1}^{p_\phi} \left(\frac{\partial R_{\bullet j}}{\partial \phi'}(\dot{\phi}_j) T^{1/4}(\hat{\phi} - \phi_0)\right) T^{1/4}(\hat{\eta}_j - \eta_{0,j}) \\
&= -\left(R^{-1}\right)_{h\bullet} \frac{\partial R_{\bullet p_\phi}}{\partial \phi'}(\phi_0) R_{\bullet p_\phi} \left(T^{1/4}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi})\right)^2 + o_P(1),
\end{aligned}$$

where the last equality uses (51) and the fact that $\hat{R}$ and $\frac{\partial R_{\bullet j}}{\partial \phi'}(\dot{\phi}_j)$ converge in probability towards R and $\frac{\partial R_{\bullet j}}{\partial \phi'}(\phi_0)$, respectively and the fact that $T^{1/4}(\hat{\eta}_j - \eta_{0,j}) = o_P(1)$ for $j = 1, \ldots, p_\phi - 1$.

Besides, we have

$$\epsilon_{T,p_\phi} = -\left(\hat{R}^{-1}\right)_{p_\phi \bullet} \sum_{j=1}^{p_\phi} \left(\frac{\partial R_{\bullet j}}{\partial \phi'}(\dot{\phi}_j) T^{1/4}(\hat{\phi} - \phi_0)\right)(\hat{\eta}_j - \eta_{0,j}) = o_P(1).$$

Putting together these last two equalities, we get:

$$\epsilon_T = -A\left(T^{1/4}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi})\right)^2 + o_P(1)$$

as expected. $\square$

Proof of Equation (25): A second-order mean value expansion yields:

$$f(\phi_0) = f(\hat{\phi}) + \frac{\partial f}{\partial \phi'}(\hat{\phi})(\phi_0 - \hat{\phi}) + \frac{1}{2}\left((\phi_0 - \hat{\phi})' \frac{\partial^2 f_k}{\partial \phi \partial \phi'}(\dot{\phi})(\phi_0 - \hat{\phi})\right)_{1 \le k \le K},$$

with $\dot{\phi} \in (\phi_0, \hat{\phi})$ and may vary from row to row. This can be re-written:

$$f(\hat{\phi}) - f(\phi_0) = \frac{\partial f}{\partial \phi'}(\hat{\phi})(\hat{\phi} - \phi_0) - \frac{1}{2}\left((\hat{\phi} - \phi_0)' \frac{\partial^2 f_k}{\partial \phi \partial \phi'}(\dot{\phi})(\hat{\phi} - \phi_0)\right)_{1 \le k \le K}.$$

From (24), we can claim that:

$$f(\hat{\phi}) - f(\phi_0) \overset{a}{\sim} \frac{\partial f}{\partial \phi'}(\hat{\phi})\hat{R}B_T^{-1}\tilde{\mathbb{X}} - \frac{1}{2}\left(\tilde{\mathbb{X}}'B_T^{-1}\hat{R}'\frac{\partial^2 f_k}{\partial\phi\partial\phi'}(\dot{\phi})\hat{R}B_T^{-1}\tilde{\mathbb{X}}\right)_{1\leq k\leq K}.$$

Note that the first term on the RHS is of order $O_P(T^{-1/4})$ or $O_P(T^{-1/2})$ and cannot vanish trivially at a rate faster than $O_P(T^{-1/2})$ if $\frac{\partial f}{\partial \phi'}(\phi_0) \neq 0$ because R and $Var(\tilde{\mathbb{X}})$ are non singular. The only element from the expansion of the second term that does not vanish at a rate faster than $O_P(T^{-1/2})$ is, with $\hat{R}_{\bullet p_\phi}$ denoting the p_ϕ^{th} column of $\hat{R}$,

$$-\frac{1}{2}\frac{1}{\sqrt{T}}\left(\hat{R}'_{\bullet p_\phi}\frac{\partial^2 f_k}{\partial\phi\partial\phi'}(\dot{\phi})\hat{R}_{\bullet p_\phi}\tilde{\mathbb{X}}^2_{p_\phi}\right)_{1\leq k\leq K}.$$

Thus, we have

$$f(\hat{\phi}) - f(\phi_0) \overset{a}{\sim} \frac{\partial f}{\partial \phi'}(\hat{\phi})\hat{R}B_T^{-1}\tilde{\mathbb{X}} - \frac{1}{2}\frac{1}{\sqrt{T}}\left(\hat{R}'_{\bullet p_\phi}\frac{\partial^2 f_k}{\partial\phi\partial\phi'}(\dot{\phi})\hat{R}_{\bullet p_\phi}\tilde{\mathbb{X}}^2_{p_\phi}\right)_{1\leq k\leq K}. \tag{52}$$

Under standard conditions, we have:

$$\frac{\partial^2 f_k}{\partial\phi\partial\phi'}(\dot{\phi}) - \frac{\partial^2 f_k}{\partial\phi\partial\phi'}(\hat{\phi}) = o_P(1)$$

and using this result in (52) yields (25). $\square$

Proof of Theorem 2: We have:

$$B_{IT}\hat{S}\left(\hat{\theta}_{II} - \theta_0\right) = \begin{pmatrix} \sqrt{T}\hat{S}_1(\hat{\theta}_{II} - \theta_0) \\ T^{1/4}\hat{S}_{p\bullet}(\hat{\theta}_{11} - \theta_0) \end{pmatrix},$$

where $\hat{S}_{p\bullet}$ is the p^{th} row of $\hat{S}$. From (30), we have

$$\sqrt{T}\hat{S}_1\left(\hat{\theta}_{II} - \theta_0\right) = \hat{S}_1\hat{F}_T\left(B_T m_{IT}(\theta_0) - \frac{1}{2}z_T\right),$$

with $z_T = B_T\left((\hat{\theta}_{II} - \theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0)\right)_{1\leq k\leq \ell}$. For $k = 1, \ldots, \ell - 1$,

$$z_{T,k} = \sqrt{T}(\hat{\theta}_{II} - \theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0) = T^{1/4}(\hat{\theta}_{II} - \theta_0)'\Delta_{IT,k}(\dot{\theta}_T)T^{1/4}(\hat{\theta}_{II} - \theta_0)$$

and

$$z_{T,\ell} = T^{1/4}(\hat{\theta}_{II} - \theta_0)'\Delta_{IT,\ell}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0).$$

From (28), we have $T^{1/4}(\hat{\theta}_{II} - \theta_0) = F_{\bullet\ell}T^{1/4}m_{IT,\ell}(\theta_0) + o_P(1)$. In addition, the fact that $\Delta_{IT,k}(\dot{\theta}_T)$ converges in probability towards $\Delta_{I,k}(\theta_0)$ for all $k = 1, \ldots, \ell$, allows us to claim that: for $1 \leq k \leq \ell - 1$,

$$z_{T,k} = F'_{\bullet\ell}\Delta_{I,k}(\theta_0)F_{\bullet\ell}\left(T^{1/4}m_{IT,\ell}(\theta_0)\right)^2 + o_P(1)$$

and

$$z_{T,\ell} = O_P(1)O_P(1)o_P(1) = o_P(1).$$

Thus,

$$\begin{aligned} & \sqrt{T}\hat{S}_1\left(\hat{\theta}_{II} - \theta_0\right) \\ = \; & \hat{S}_1\hat{F}_T\left(B_T m_{IT}(\theta_0) - \tfrac{1}{2}\begin{pmatrix} (F'_{\bullet\ell}\Delta_{I,k}(\theta_0)F_{\bullet\ell})_{1\leq k\leq\ell-1} \\ 0 \end{pmatrix}\left(T^{1/4}m_{IT,\ell}(\theta_0)\right)^2\right) + o_P(1). \end{aligned}$$

Since the last column of $\hat{S}_1 \hat{F}_T$ is nil, we can write:

$$\begin{aligned} & \sqrt{T}\hat{S}_1 \left(\hat{\theta}_{II} - \theta_0\right) \\ = \quad & \hat{S}_1 \hat{F}_T \left(B_T m_{IT}(\theta_0) - \frac{1}{2} \left(F'_{\bullet\ell} \Delta_{I,k}(\theta_0) F_{\bullet\ell}\right)_{1 \le k \le \ell} \left(T^{1/4} m_{IT,\ell}(\theta_0)\right)^2 \right) + o_P(1). \end{aligned} \tag{53}$$

Using again (28), we have

$$T^{1/4} \hat{S}_{p\bullet} (\hat{\theta}_{II} - \theta_0) = \hat{S}_{p\bullet} F_{\bullet\ell} T^{1/4} m_{IT,\ell}(\theta_0) + o_P(1). \tag{54}$$

By the continuous mapping theorem, $\hat{S}_1 \hat{F}_T$ converges in probability towards $S_1 F$ with nil last column and $\hat{S}_{p\bullet}$ converges in probability towards $S_{p\bullet}$. Since $B_T m_{IT}(\theta_0)$ converges in distribution towards $\mathbb{Y}$, we can deduce from (53) and (54) that:

$$\begin{pmatrix} \sqrt{T}\hat{S}_1(\hat{\theta}_{II} - \theta_0) \\ \\ T^{1/4} \hat{S}_{p\bullet}(\hat{\theta}_{11} - \theta_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_1 F \left(\mathbb{Y} - \frac{(\mathbb{Y}_\ell)^2}{2} \left(F'_{\bullet\ell} \Delta_{I,k}(\theta_0) F_{\bullet\ell}\right)_{1 \le k \le \ell} \right) \\ \\ S_{p\bullet} F_{\bullet\ell} \mathbb{Y}_\ell \end{pmatrix}.$$

□

C Tables and graphs

Table 1: GMM estimation of panel data model

ρ	$\hat{\rho}$	RMSE	Coverage probabilities Cov-1	Cov-2	Cov-3	$\lvert\bar{\tilde{D}}\rvert$	$\lvert\bar{\tilde{G}}\rvert$	$\frac{\lvert\bar{\tilde{G}}\rvert}{\lvert\bar{\tilde{D}}\rvert}$
				$n = 50$				
0.20	0.928	6.203	67.88	83.98	83.99	0.643	89.291	138.89
0.30	0.756	2.566	73.45	87.77	87.74	0.561	34.497	61.49
0.50	0.730	1.160	79.34	92.31	92.30	0.428	30.346	70.86
0.75	0.909	0.698	78.29	91.31	91.25	0.299	4.937	16.51
0.80	0.966	1.042	78.21	90.44	90.37	0.269	2.813	10.46
0.90	1.044	0.482	78.34	88.84	88.83	0.198	3.010	15.20
0.95	1.079	0.540	78.46	87.69	87.66	0.137	2.832	20.67
0.97	1.097	0.369	78.59	87.19	87.16	0.117	2.908	24.85
0.98	1.104	0.359	78.78	86.87	86.86	0.105	2.911	27.72
1.00	1.119	0.363	78.78	86.16	86.16	0.089	2.997	33.67
1.10	1.197	0.355	79.35	83.12	83.07	0.118	3.322	28.15
1.20	1.277	0.356	79.28	79.66	79.50	0.241	3.558	14.76
1.30	1.361	0.349	79.21	77.65	77.55	0.366	3.851	10.52
1.50	1.546	0.313	81.57	78.08	78.03	0.544	4.385	8.06
				$n = 100$				
0.20	0.503	2.157	72.82	87.82	87.80	0.663	21.984	33.13
0.30	0.502	4.404	79.12	91.97	91.97	0.617	96.243	155.90
0.50	0.618	0.504	82.48	95.27	95.24	0.416	22.801	54.78
0.75	0.884	0.336	80.52	93.77	93.76	0.293	2.788	9.52
0.80	0.931	0.325	80.30	93.31	93.33	0.267	2.248	8.42
0.90	1.015	0.293	80.44	91.78	91.78	0.180	2.333	12.96
0.95	1.047	0.291	80.56	90.30	90.30	0.115	2.539	22.08
0.97	1.059	0.289	80.66	89.70	89.67	0.088	2.592	29.45
0.98	1.066	0.289	80.72	89.44	89.39	0.078	2.626	33.67
1.00	1.078	0.289	80.63	88.54	88.49	0.062	2.738	44.16
1.10	1.160	0.296	81.17	84.86	84.73	0.131	3.066	23.40
1.20	1.242	0.297	80.87	81.22	81.16	0.265	3.402	12.84
1.30	1.330	0.288	81.37	80.15	80.07	0.382	3.702	9.69
1.50	1.531	0.245	84.99	83.83	83.79	0.515	4.230	8.21
				$n = 200$				
0.20	0.334	1.149	80.68	93.16	93.16	0.669	5.359	8.01
0.30	0.381	0.462	86.09	96.62	96.59	0.567	95.114	167.85
0.50	0.574	0.226	85.87	96.99	97.00	0.438	4.809	10.97
0.75	0.844	0.248	81.51	94.70	94.71	0.275	2.044	7.43
0.80	0.897	0.250	81.60	94.51	94.44	0.248	2.025	8.17
0.90	0.989	0.246	81.08	93.39	93.40	0.168	2.171	12.92
0.95	1.022	0.244	80.93	92.19	92.14	0.101	2.351	23.28
0.97	1.034	0.242	81.11	91.36	91.34	0.072	2.428	33.72
0.98	1.039	0.241	81.25	90.99	90.99	0.058	2.479	42.74
1.00	1.051	0.241	81.45	90.02	89.97	0.044	2.574	58.50
1.10	1.125	0.250	81.49	84.85	84.76	0.169	2.998	17.74
1.20	1.211	0.250	81.17	81.71	81.65	0.300	3.336	11.12
1.30	1.316	0.235	82.78	83.10	83.05	0.361	3.554	9.84
1.50	1.527	0.184	88.21	90.25	90.20	0.471	4.103	8.71

Continued over

Table 1 (continued): GMM estimation of panel data model

ρ	$\hat{\rho}$	RMSE	Coverage probabilities Cov-1	Cov-2	Cov-3	$\lvert\bar{\hat{D}}\rvert$	$\lvert\bar{\hat{G}}\rvert$	$\frac{\lvert\bar{\hat{G}}\rvert}{\lvert\bar{\hat{D}}\rvert}$
				$n = 1000$				
0.20	0.226	0.069	91.07	99.63	99.63	0.724	2.015	2.78
0.30	0.320	0.070	92.98	99.94	99.94	0.657	1.191	1.81
0.50	0.521	0.096	91.68	99.82	99.82	0.484	1.156	2.39
0.75	0.784	0.154	83.61	96.78	96.72	0.253	1.718	6.79
0.80	0.839	0.163	82.15	96.24	96.17	0.215	1.815	8.44
0.90	0.961	0.166	81.96	97.11	97.08	0.168	1.928	11.48
0.95	1.005	0.162	81.91	96.18	96.11	0.118	2.076	17.59
0.97	1.014	0.160	82.22	95.35	95.28	0.079	2.177	27.56
0.98	1.019	0.160	82.13	94.96	94.93	0.060	2.227	37.12
1.00	1.025	0.159	82.52	93.77	93.74	0.019	2.351	123.74
1.10	1.092	0.168	83.07	87.35	87.23	0.174	2.818	16.20
1.20	1.198	0.158	83.28	87.18	87.12	0.251	3.079	12.27
1.30	1.312	0.129	88.66	93.32	93.23	0.293	3.308	11.29
1.50	1.512	0.088	93.03	98.47	98.46	0.470	4.028	8.57
				$n = 5000$				
0.20	0.204	0.026	94.27	100.00	100.00	0.787	0.616	0.78
0.30	0.302	0.029	94.52	100.00	100.00	0.692	0.623	0.90
0.50	0.503	0.042	93.67	100.00	100.00	0.496	0.950	1.92
0.75	0.760	0.083	87.93	99.29	99.29	0.250	1.565	6.26
0.80	0.812	0.096	84.66	98.35	98.35	0.202	1.710	8.47
0.90	0.927	0.114	81.06	97.43	97.41	0.127	1.943	15.30
0.95	0.992	0.113	81.72	97.74	97.71	0.113	2.013	17.81
0.97	1.009	0.111	81.59	97.02	97.00	0.089	2.083	23.40
0.98	1.013	0.110	81.51	96.53	96.49	0.069	2.134	30.93
1.00	1.016	0.108	82.17	94.87	94.84	0.014	2.268	162.00
1.10	1.084	0.115	81.18	87.80	87.77	0.164	2.726	16.62
1.20	1.207	0.089	87.69	95.47	95.43	0.194	2.922	15.06
1.30	1.308	0.065	92.14	99.23	99.20	0.285	3.255	11.42
1.50	1.504	0.040	94.20	99.98	99.98	0.488	4.031	8.26

Notes: Simulated mean and root-mean-squared-error of the GMM estimator of ρ; coverage probability of confidence intervals (in %) based on: the standard asymptotic theory assuming first-order local identification (Cov-1); the asymptotic distribution in Theorem 1(b), using asymptotic critical values (Cov-2) and simulated critical values (Cov-3); $\lvert\bar{\hat{D}}\rvert$ is the norm of the simulated mean of the Jacobian; $\lvert\bar{\hat{G}}\rvert$ is the norm of the second-order derivative of the moment function. The true underlying data set has a dynamic panel structure (Example 1) with $\sigma_0^2 = \sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = \sigma_{0\eta} = 0$ and ρ as in the table. (10,000 runs)

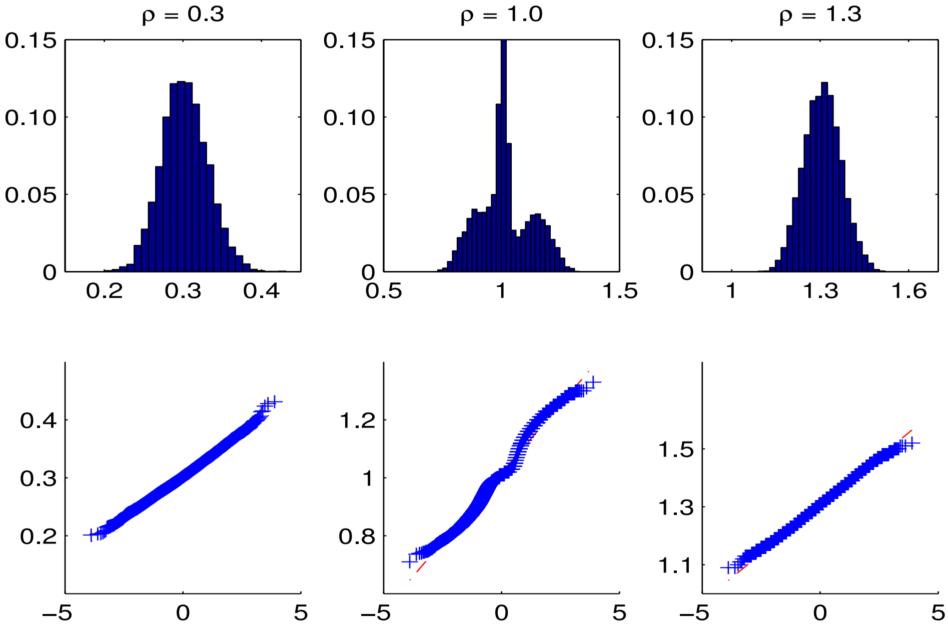

Figure 1: Histogram of simulated GMM estimator of ρ; $\rho = 0.3, 1.0$ and 1.3, respectively and their QQ-plot versus the standard normal distribution. The true underlying data set has a dynamic panel structure (Example 1) with $\sigma_0^2 = \sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = \sigma_{0\eta} = 0$ and ρ. Simulated sample size $n = 5,000$. (10,000 runs)

Table 2: GMM estimation of panel data model

λ	$\hat{\rho}$	RMSE	Coverage probabilities Cov-1	Cov-2	Cov-3	$\lvert\bar{\tilde{D}}\rvert$	$\lvert\bar{\tilde{G}}\rvert$	$\frac{\lvert\bar{\tilde{G}}\rvert}{\lvert\bar{\tilde{D}}\rvert}$
				$n = 50$				
-0.50	1.131	0.439	78.96	89.25	89.23	0.694	3.019	4.35
-0.30	1.154	0.447	76.48	85.32	85.29	0.410	3.037	7.41
-0.20	1.152	0.397	76.55	84.91	84.87	0.279	2.994	10.73
-0.10	1.138	0.387	77.72	85.25	85.20	0.164	2.960	18.05
0.00	1.119	0.363	78.78	86.16	86.16	0.089	2.997	33.67
0.10	1.109	0.373	79.25	80.08	80.01	0.386	4.352	11.27
0.20	1.095	0.384	79.20	74.45	74.35	0.787	5.688	7.23
0.30	1.084	0.382	78.75	70.64	70.57	1.172	6.955	5.93
0.50	1.067	0.350	79.39	69.56	69.54	1.826	8.937	4.89
				$n = 100$				
-0.50	1.105	0.406	82.62	93.77	93.71	0.652	2.663	4.08
-0.30	1.139	0.346	77.56	88.78	88.77	0.344	2.594	7.54
-0.20	1.139	0.338	76.91	87.15	87.01	0.220	2.599	11.81
-0.10	1.115	0.314	78.23	87.51	87.41	0.128	2.633	20.57
0.00	1.078	0.289	80.63	88.54	88.49	0.062	2.738	44.16
0.10	1.064	0.311	80.92	80.98	80.92	0.407	4.036	9.92
0.20	1.050	0.317	81.23	75.90	75.84	0.804	5.297	6.59
0.30	1.052	0.313	80.92	73.66	73.56	1.113	6.278	5.64
0.50	1.052	0.275	82.41	75.82	75.66	1.657	7.919	4.78
				$n = 200$				
-0.50	1.069	0.227	86.54	97.30	97.25	0.657	2.545	3.87
-0.30	1.114	0.284	79.79	91.96	91.86	0.324	2.407	7.43
-0.20	1.122	0.293	77.01	89.05	89.00	0.190	2.399	12.63
-0.10	1.101	0.272	77.49	88.97	88.93	0.101	2.437	24.13
0.00	1.051	0.241	81.45	90.02	89.97	0.044	2.574	58.50
0.10	1.027	0.258	82.13	81.93	81.86	0.436	3.880	8.90
0.20	1.031	0.264	82.19	78.19	78.21	0.755	4.901	6.49
0.30	1.042	0.255	81.73	77.51	77.49	1.007	5.715	5.68
0.50	1.044	0.209	85.35	82.39	82.39	1.525	7.268	4.77
				$n = 1000$				
-0.50	1.014	0.087	94.20	99.92	99.94	0.694	2.496	3.60
-0.30	1.047	0.152	88.35	97.22	97.24	0.365	2.322	6.36
-0.20	1.073	0.188	82.07	94.02	93.96	0.192	2.235	11.64
-0.10	1.077	0.196	77.61	91.66	91.61	0.070	2.220	31.71
0.00	1.025	0.159	82.52	93.77	93.74	0.019	2.351	123.74
0.10	1.004	0.169	85.25	87.35	87.23	0.384	3.481	9.07
0.20	1.031	0.168	85.00	86.69	86.62	0.597	4.220	7.07
0.30	1.031	0.145	86.89	89.10	88.97	0.878	5.063	5.77
0.50	1.016	0.090	93.12	96.85	96.79	1.517	6.894	4.54
				$n = 5000$				
-0.50	1.002	0.035	94.86	100.00	100.00	0.707	2.504	3.54
-0.30	1.008	0.057	94.77	99.90	99.89	0.415	2.375	5.72
-0.20	1.024	0.096	89.69	97.18	97.14	0.248	2.287	9.22
-0.10	1.052	0.137	78.20	91.44	91.50	0.072	2.185	30.35
0.00	1.016	0.108	82.17	94.87	94.84	0.014	2.268	162.00
0.10	1.016	0.116	84.14	90.46	90.27	0.300	3.216	10.72
0.20	1.018	0.094	88.13	92.52	92.50	0.588	4.083	6.94
0.30	1.008	0.060	93.40	98.37	98.36	0.927	5.045	5.44
0.50	1.003	0.037	93.76	99.89	99.89	1.569	6.936	4.42

Notes: Definitions as Table 1 except that the true underlying data set has a dynamic panel structure (Example 1) with $\rho = 1$, $\sigma_0^2 = \sigma_\varepsilon^2 = 1$, $\sigma_{0\eta} = \lambda$, $\sigma_\eta^2 = |\lambda|$, with λ as in the table.

Table 3: GMM and II estimation of panel data model

ρ	Bias GMM	Bias II	RMSE GMM	RMSE II	Cov. probability Cov-1ii	Cov. probability Cov-2ii	$\lvert\bar{\hat{D}}\rvert$	$\lvert\bar{\hat{G}}\rvert$	$\frac{\lvert\bar{\hat{G}}\rvert}{\lvert\bar{\hat{D}}\rvert}$
					$n = 50$				
0.30	0.486	0.349	2.402	2.383	82.70	87.50	0.511	1.656	3.240
0.80	0.140	-0.035	0.548	0.511	83.50	89.96	0.249	3.442	13.806
0.90	0.131	-0.054	0.483	0.471	75.78	78.12	0.469	4.561	9.729
1.00	0.101	-0.088	0.366	0.388	67.12	67.94	0.802	5.909	7.370
1.20	0.073	-0.105	0.355	0.426	67.18	66.26	1.451	8.179	5.638
1.30	0.063	-0.109	0.348	0.432	66.06	63.28	1.678	8.719	5.196
					$n = 100$				
0.30	0.118	0.080	0.611	0.575	90.40	97.30	0.653	1.238	1.896
0.80	0.127	0.038	0.299	0.265	79.00	86.34	0.190	3.305	17.427
0.90	0.112	0.017	0.294	0.262	80.66	84.38	0.181	3.369	18.614
1.00	0.081	-0.019	0.289	0.268	82.14	84.50	0.315	3.745	11.902
1.20	0.039	-0.066	0.298	0.299	77.92	78.20	0.703	4.655	6.622
1.30	0.040	-0.062	0.288	0.294	79.10	74.42	0.792	4.787	6.042
					$n = 200$				
0.30	0.079	0.034	0.403	0.390	61.30	71.40	0.558	3.257	5.839
0.80	0.101	0.032	0.252	0.217	84.50	93.68	0.174	2.382	13.693
0.90	0.094	0.019	0.248	0.216	84.74	90.18	0.094	2.510	26.664
1.00	0.055	-0.022	0.243	0.227	81.92	83.62	0.198	3.003	15.144
1.20	0.006	-0.074	0.251	0.261	71.60	70.20	0.597	3.922	6.575
1.30	0.010	-0.069	0.236	0.252	75.04	72.54	0.683	4.103	6.006
					$n = 1000$				
0.30	0.019	-0.004	0.069	0.068	94.30	99.50	0.634	1.556	2.453
0.80	0.036	-0.004	0.162	0.149	85.48	92.04	0.158	2.025	12.796
0.90	0.058	0.018	0.165	0.159	79.48	91.70	0.097	2.122	21.909
1.00	0.021	-0.017	0.158	0.170	75.66	85.00	0.107	2.587	24.218
1.20	0.001	-0.017	0.158	0.179	75.36	78.32	0.315	3.222	10.233
1.30	0.015	0.008	0.130	0.138	86.66	89.72	0.311	3.347	10.771
					$n = 5000$				
0.30	0.002	0.002	0.030	0.030	94.30	100.00	0.691	0.645	0.933
0.80	0.012	0.005	0.095	0.088	87.40	97.82	0.193	1.730	8.987
0.90	0.026	0.008	0.112	0.096	90.12	96.38	0.098	1.999	20.426
1.00	0.016	-0.009	0.107	0.100	86.68	94.06	0.043	2.360	55.217
1.20	0.008	-0.016	0.088	0.098	82.60	86.48	0.263	3.033	11.536
1.30	0.008	-0.009	0.064	0.067	91.56	97.66	0.337	3.325	9.860

Notes: Simulated mean and root-mean-squared-error of the GMM and II estimators of ρ; coverage probability of II-based confidence intervals for ρ using: (i) the standard II asymptotic theory assuming first-order local identification (Cov-1ii); and the result of Theorem 2 (Cov-2ii). We set $s = 50$ for the estimated II binding function. For other definitions see the notes to Table 1. (5,000 runs)

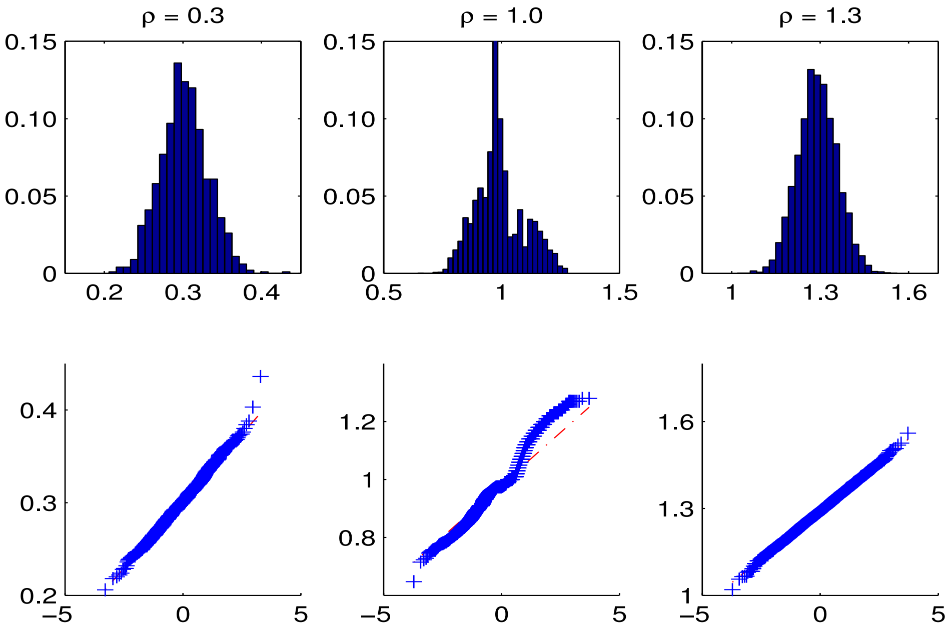

Figure 2: Histogram of simulated II estimator of ρ; $\rho = 0.3, 1.0$ and 1.3, respectively and their QQ-plot versus the standard normal distribution. The true underlying data set has a dynamic panel structure (Example 1) with $\sigma_0^2 = \sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = \sigma_{0\eta} = 0$ and ρ. Simulated sample size $n = 5,000$. (5,000 runs)

Table 4: Robustness of II simulating model

λ	Bias GMM	Bias II	RMSE GMM	RMSE II	Cov. probability Cov-1ii	Cov. probability Cov-2ii	$\lvert\bar{\tilde{D}}\rvert$	$\lvert\bar{\tilde{G}}\rvert$	$\frac{\lvert\bar{\tilde{G}}\rvert}{\lvert\bar{\tilde{D}}\rvert}$
					$n = 200$				
-0.20	0.051	-0.009	0.244	0.235	80.10	83.50	0.171	2.977	17.436
-0.10	0.051	-0.041	0.244	0.237	78.60	80.90	0.256	3.202	12.483
0.00	0.051	-0.025	0.244	0.229	80.60	83.10	0.201	3.018	15.048
0.10	0.051	-0.052	0.244	0.256	74.10	79.10	0.325	3.479	10.702
0.20	0.051	-0.060	0.244	0.269	74.60	81.30	0.401	3.895	9.724
MAD					16.65	13.35			
					$n = 1000$				
-0.20	0.022	-0.022	0.159	0.176	73.60	79.80	0.122	2.625	21.460
-0.10	0.022	-0.015	0.159	0.181	73.40	82.70	0.112	2.618	23.370
0.00	0.022	-0.015	0.159	0.171	76.10	85.20	0.104	2.582	24.801
0.10	0.022	0.008	0.159	0.174	74.70	82.60	0.067	2.545	38.220
0.20	0.022	0.008	0.159	0.160	87.60	95.90	0.049	2.473	50.663
MAD					20.55	12.43			
					$n = 5000$				
-0.20	0.019	0.006	0.107	0.099	87.00	95.10	0.011	2.299	215.015
-0.10	0.019	0.004	0.107	0.110	76.50	80.10	0.018	2.323	128.947
0.00	0.019	-0.006	0.107	0.100	87.40	94.90	0.036	2.351	65.579
0.10	0.019	0.004	0.107	0.097	87.70	93.70	0.013	2.304	183.863
0.20	0.019	0.000	0.107	0.106	84.90	94.60	0.025	2.334	95.251
MAD					10.30	3.36			

Notes: The true underlying data set has a dynamic panel structure (Example 2) with $\rho = 1$, $\sigma_0^2 = \sigma_\varepsilon^2 = 1$, $\sigma_{0\eta} = \sigma_\eta^2 = 0$. II estimation of ρ uses (10) for simulated samples and (11) as auxiliary model with θ_2 calibrated as follows: $\tilde{\sigma}_0^2 = \tilde{\sigma}_\varepsilon^2 = 1$, $\tilde{\sigma}_{0\eta} = \lambda$ and $\tilde{\sigma}_\eta^2 = |\lambda|$, with λ as in the table. All figures in the table relate to the estimators of ρ. 'MAD' is the mean absolute deviation of the coverage probabilities from the nominal (95%). For other definitions see the notes to Table 3. (1,000 runs)