MANCHESTER 1824

The University of Manchester

Economics Discussion Paper Series EDP-1505

GMM and Indirect Inference:

An appraisal of their connections and new results on their properties under second order identification

Prosper Donovon

Alastair R. Hall

March 2015

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

GMM and Indirect Inference[1]

- an appraisal of their connections and new results on their properties under second order identification

Prosper Dovonon

Concordia University[2]

and

Alastair R. Hall

University of Manchester[3]

January 15, 2015

[1]This paper is based in part on material in an invited talk entitled "GMM and Indirect Inference" presented by Hall at the Conference on Indirect Estimation Methods in Finance and Economics, Abbey Hegne, Allensbach, Germany, May 30-31, 2014.

[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 makes two contributions. First, we provide a review of the similarities and difference between Generalized Method of Moments and Indirect Inference, focusing particularly on issues of moment selection, identification failure and model misspecification. Secondly, 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.

1 Introduction

Lars Hansen was awarded the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel for 2013 jointly with Eugene Fama and Robert Shiller for their "empirical analyses of asset prices".[1] While recognizing Hansen's many contributions to this field, the award is primarily in recognition of his introduction of the Generalized Method of Moments (GMM) framework for inference. As noted in the "Scientific Background" to the announcement of the award, Hansen's 1982 article in *Econometrica* that introduced the method as "one of the most influential papers in econometrics".[2] 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)[3] 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.

Given the recent award of the Nobel prize to Hansen, it seems timely to explore the connections between GMM and II, highlighting both similarities and some key differences. As will be discussed, both methods can be viewed as "minimum chi-squared" methods and hence share the same linear algebraic structure of their first order analyses, although the regularity conditions underlying each are different in important ways. Since both methods involve moment-based estimation, it is natural to expect that issues relevant to GMM estimation are also relevant to II. In this paper, we investigate the extent to which is the case focusing on three particular topics that have received attention within the GMM framework: moment selection, identification failure and inference in misspecified models.

From this discussion, it emerges that concerns have been raised about the consequences of identification failure in certain applications of both GMM and II. However, while this topic has received considerable attention in the context of GMM, there is very little guidance available for models estimated via II. In

[1] See http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/press.html.

[2] See The Royal Swedish Academy of Sciences (2013b), p.24.

[3] 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).

this paper, we consider the case where first order local identification fails but the second order local identification holds. Sargan (1983) demonstrates how this situation can arise in models that are nonlinear in parameters. More recently, Madsen (2009) and Dovonon and Renault (2009, 2013) show this scenario can arise in panel data models and factor models respectively. Although this situation has been recognized to arise in models of interest, there are no general results available on the limiting distribution of either GMM or II estimators in this case. In this paper, we fill this gap in the literature. We present the limiting distribution of both the GMM estimator under second order identification and also the II estimator in cases where the auxiliary model is 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 outline of the paper is as follows. Section 2 compares and contrasts the GMM and II methods. Section 3 reviews the similarities and differences in the way the issues of moment selection, identification failure and inference in misspecified models have been approached in the GMM and II frameworks. Section 4 considers the behaviour of GMM and II under second order identification, and Section 5 concludes.

2 First order asymptotics of GMM and II: similarities and differences

In this section, we explore the similarities and the differences of the basic GMM and II inference frameworks based on first order asymptotics. As will be seen, the similarities stem from both being essentially "minimum chi-squared" methods. 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 $\theta_0 \in \Theta \subset R^p$. For some of the discussion only a subset of the parameters may be of primary interest, and so we write $\theta = (\phi', \psi')'$ where $\phi \in \Phi \subset R^{p_\phi}$ and $\psi \in \Psi \subset R^{p_\psi}$. 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:[4]

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

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

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

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

where

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

$\{x_t\}_{t=1}^T$ 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;[5] in model estimated via quasi-maximum likelihood then the moment condition is the quasi-score.[6]

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 $h_T^{(i)}(\theta) = h\left(\{v_t^{(i)}(\theta)\}_{t=1}^T\right)$. Assume $dim(h_T) = \ell > p$. The II estimator of θ_0 is:[7]

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

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

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

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

[7]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).

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,[8] *e.g.* Gallant and Tauchen (1996), Garcia, Renault, and Veredas (2011); estimated moments from the auxiliary model, such as impulse response functions in DSGE models, *e.g.* see Christiano, Eichenbaum, and Evans (2005).

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 as we now demonstrate.

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, $\tilde{m}_T(\phi_0)' V_m^{-1} \tilde{m}_T(\phi_0) \xrightarrow{d} \chi^2_\ell$, 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:[9]

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

where

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

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

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

[9] See Ferguson (1958).

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)* $Q_T(\phi) \xrightarrow{p} Q(\phi) = m(\phi)'Wm(\phi)$ *uniformly in* ϕ*; (iii)* $Q(\phi_0) < Q(\phi)\ \forall\, \phi \neq \phi_0$, $\phi \in \Phi$.

Assumption 2(iii) serves as an 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 $N_\epsilon = \{\phi;\, \|\phi - \phi_0\| < \epsilon\}$.

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 2-3 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: $V_\phi = \{M(\phi_0)'V_m^{-1}M(\phi_0)\}^{-1}$.[10] 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).

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

Other useful properties of the estimator also stem from the first order conditions. Since $M_T(\phi)$ is $n \times p_\phi$, it follows that (9) involves calculating $\hat{\phi}_{MC}$ as the value of ϕ that sets the p linear combinations of $m_T(.)$ to zero. Thus the estimator is, in effect, based on the information that

$$M(\phi_0)'Wm(\phi) \;=\; 0. \tag{10}$$

Our starting point, Assumption 1, implies that $m(\phi_0) = 0$. However, $m(\phi_0) = 0$ and (10) are only equivalent if $n = p_\phi$ (and Assumptions 2(i) and 3(iii) hold); if $n > p_\phi$ then $m(\phi_0) = 0$ implies (10) but the reverse is not true. Therefore, if $n > p_\phi$ then the estimation affects a decomposition of the original information, $m(\phi_0) = 0$ into two parts: the part used in estimation, which can be characterized as[11]

$$PV_m^{-1/2}m(\phi_0) \;=\; 0, \tag{11}$$

where $P = N(N'N)^{-1}N'$ and $N = W^{1/2}M(\phi_0)$; and the part unused in estimation,

$$(I \,-\, P)V_m^{-1/2}m(\phi_0) \;=\; 0. \tag{12}$$

While unused in estimation, the restrictions in (12) form a basis for a model diagnostic test for if $m(\phi_0) = 0$ then so too must (12). Such a test can be conveniently constructed by noting that

$$T^{1/2}V_m^{-1/2}m_T(\hat{\phi}_{MC}) \;\simeq\; (I_n \,-\, P)V_m^{-1/2}\tilde{m}_T(\phi_0), \tag{13}$$

which leads to test statistic

$$\xi_T \;=\; \tilde{m}_T(\hat{\phi}_{MC})'\hat{V}_m^{-1}\tilde{m}_T(\hat{\phi}_{MC}). \tag{14}$$

Under $H_0 : m(\phi_0) = 0$, $\xi_T \xrightarrow{d} \chi^2_{n-p_\phi}$, where $\hat{V}_m \xrightarrow{p} V_m$.

Discussion:
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

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

and also propose certain other model specifications tests. Following Sowell (1996),[12] the decomposition in (11)-(12) is referred to as being in terms of

[11] This decomposition applies the results due to Sowell (1996) for GMM to minimum chi-squared; see below for further discussion.

[12] Hansen (1982) characterized the overidentifying restrictions in nonlinear models, generalizing Sargan's (1958) linear model analysis and adopting his terminology; Sowell (1996) characterizes the decomposition in the form presented here.

identifying and overidentifying restrictions respectively. In GMM, this decomposition is of the population moment condition: $E[g(X, \phi_0)] = 0$. Ghysels and Guay (2004) extend Sowell's (1996) analysis to II where the decomposition involves $k(h_*, \theta_0) = 0$. As a result, the statistic in (14) is commonly referred to as the overidentifying restrictions test in both GMM and II.

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 in two important ways. First, Hansen's (1982) framework includes the restriction that X is strictly stationary and ergodic but the simulated series in II, $\{x_t^{(i)}(\theta)\}$, do not satisfy these conditions because the initial conditions are typically not drawn from the stationary distribution rendering the simulated series locally nonstationary. Second, Hansen (1982) establishes the uniform convergence of the sample GMM minimand to its population analog (Assumption 2(ii) above) using certain first moment continuity assumptions that are not satisfied in II because the simulated process depends on the unknown parameters. This has led to the development of alternative analyses for II estimators using geometric ergodicity or near-epoch dependence on mixing processes, and Lipschitz conditions, see Duffie and Singleton (1993) or Ghysels and Guay (2003, 2004).

3 GMM and II Inference

GMM has been widely applied in empirical econometric analysis and the diversity of these applications has helped to inspire the development of a broad array of inference techniques based on GMM estimators. Since, by its nature, II can only be applied in a more restrictive set of circumstances, the inference framework for II has been less well developed. However, since both methods are moment-based and have a common underlying structure, it is natural to expect that issues relevant to GMM estimation to be relevant to II as well. In this section, we investigate the extent to which this is the case focusing on three topics that have received particular attention within the GMM framework: moment selection, identification failure and inference in misspecified models. We note that on each of these topics the literature on GMM is far larger than the corresponding treatment for II. As a result, we concentrate on the parts of the GMM literature that are most relevant to II. In view of this, each sub-section below begins with a brief summary of the relevant GMM literature on each topic followed by a discussion of the extent to which these methods are relevant to and have been explored for II. A more complete review of the GMM literature on these topics can be found in Hall (2015).

3.1 Moment selection

GMM:
GMM works for any choice of $g(\cdot)$ that satisfies the assumptions mentioned above. While this flexibility can be seen as strength of the method, it leaves

open the question of which moments to employ for any given application. For in most cases, there is a potentially infinite candidate set of moment conditions upon which to base the estimation. In seeking answers, this question has been split in two parts: What is the optimal choice out of a candidate set consisting of only valid moment conditions? - Which moments are valid in a candidate set consisting of potentially valid and invalid moment conditions? We consider each in turn.

Consider first the case where the candidate set consists entirely of valid moment conditions. In terms of first order asymptotic properties, the only difference between estimators based on different moment conditions is in the variance of the limiting distribution. Therefore, from this perspective, the optimal choice is the score function associated with the true distribution of the data. However, Maximum Likelihood (ML) is infeasible in many economic models. As noted in the Introduction, this is most often because the distribution of the data is unknown but sometimes this is because the distribution is known but the likelihood function is intractable. In the latter cases, it can still be possible to achieve the efficiency of ML via GMM. This will happen if the true score function lies in the space spanned by the moment conditions; for example see Singleton (2001).[13]

However, more often than not, the distribution is not part of the specification. In many such situations, GMM is based on a moment condition derived from the orthogonality of a function of the data and ϕ_0, $u_t(\phi_0)$, and a vector of instruments z_t, and so the only difference in possible moments is in the choice of possible instruments. This scenario - often referred to as "generalized instrumental variables" (GIV)[14] - has received a lot of attention within the GMM literature but is less relevant to II. We, therefore, only provide a brief summary here designed to provide an indication of the approaches taken rather than specific details. Hansen (1985) characterizes the optimal choice of instruments and the associated efficiency bound.[15] However, recalling the partial information nature of GMM, this characterization may involve assumptions about aspects of the data generation process that are not specified as part of the underlying economic model. While non-parametric methods of calculating the optimal instrument are available for i.i.d. data,[16] their extension to time series is complicated by the dependence of the optimal instrument on aspects of the dynamic structure of the data that are often not part of the economic model. For this reason and others, attention has focussed on instrument selection based on some data-based criterion. Examples of such criterion include: the estimated second order mean squared error of the estimators;[17] Lasso techniques.[18]

More relevant to exploring connections between GMM and II are methods

[13] Also see discussion in Carrasco and Florens (2000).
[14] See Hansen and Singleton (1982).
[15] Also see Hayashi and Sims (1983), Hansen, Heaton, and Ogaki (1988), Heaton and Ogaki (1991), Anatolyev (2003) and West, Wong, and Anatolyev (2009).
[16] See Newey (1990).
[17] For example see Donald and Newey (2000) and Carrasco (2012).
[18] See Belloni, Chen, Chernozhukov, and Hansen (2012).

that can handle more general functional forms of $g(\cdot)$. Hall, Inoue, Jana, and Shin (2007) propose a Relevant Moment Selection Criterion (RMSC) that is designed to exclude any redundant moment conditions. As defined by Breusch, Qian, Schmidt, and Wyhowski (1999), a sub-set of the moment conditions used in the estimation are redundant if their inclusion/exclusion does not affect the first order asymptotic properties of the GMM estimator. However, while there may be no first order effects, the inclusion of redundant moment conditions can lead to a deterioration of the accuracy of first order asymptotic to the finite sample of GMM based inference techniques.[19] Although RMSC is essentially based on first order asymptotic arguments, Hall, Inoue, Jana, and Shin (2007) report simulation evidence that its use to eliminate redundant moments from the candidate set can yield an estimator with better finite sample properties than the estimator based on all moment conditions in the entire candidate set.

We now consider methods for choosing which moments are valid. Andrews (1999) considers both sequential testing and information criterion methods based on the overidentifying restrictions test. In this case, the objective is to uncover the maximal number of valid moment conditions from the candidate set, and Andrews (1999) delineates conditions under which this happens with probability one.[20] As pointed out by Hall and Peixe (2003), a weakness of this criterion is that it leads to the inclusion of valid moments irrespective of whether their inclusion is informative about θ_0. Thus, the chosen moment condition set may contain some moments that are redundant. Hall, Inoue, Jana, and Shin (2007) show that the sequential use of Andrews's (1999) Moment Selection Criterion (MSC) and RMSC leads to the selected moments are both valid and contain no redundancies with probability one. Liao (2013) proposes a Lasso based method for selecting valid moment conditions. As with Andrews's (1999) MSC, this approach includes all valid moments irrespective of their information content. Cheng and Liao (2013) propose a modification to the criterion to ensure the Lasso method does not include any redundant moments.

II:

Although comparatively unusual in the GMM context, II involves, by its very nature, scenarios in which the distribution of the data is specified. Therefore, it is possible to choose the moments to achieve asymptotic efficiency. As pointed out by Gallant and Tauchen (1996), it is natural in this context to make the auxiliary model a QML that is chosen to be as close as possible to the true (intractable) likelihood. The ideal situation is if the simulator model is *smoothly embedded* within the auxiliary model that is, when the joint probability density functions of the simulator evaluated at θ_0 equals that under the auxiliary model evaluated at $h_0 = b(\theta_0)$, where we have replaced the $*$ subscript h by 0 to emphasize that it represents the true value of the parameters in the auxiliary model as that model is now a valid alternative representation of the data generation process. In this case, II is as efficient as MLE as $s \to \infty$. Even if this ideal is not

[19] See Hall and Peixe (2003) and Hall, Inoue, Jana, and Shin (2007).

[20] See Andrews and Lu (2001) extend Andrews's (1999) method to select parameters as well.

attainable, careful choice of the auxiliary model can yield II estimators that are close to the asymptotic efficiency of MLE. For example, Garcia, Renault, and Veredas (2011) consider II estimation of the parameters of stable distributions using the skewed-t distribution as auxiliary model, and find II comes close to achieving the Cramer-Rao lower bound for this model.

If a suitable choice of QML is not known *a priori* then Gallant and Tauchen (1996) argue for using a flexible functional form that is known to be able to approximate the true likelihood arbitrarily well as the sample size increases; they refer to such a distribution as a "general purpose score generator". In this case, if $h(\cdot)$ is allowed to expand so that it nests the score of the true distribution then the resulting II estimator is as efficient asymptotically as maximum likelihood. To illustrate, suppose it is desired to estimate a stochastic volatility model, the likelihood of which is intractable. Then Gallant and Tauchen (1996) suggest setting $h(\cdot)$ equal to the score from a semi-nonparametric (SNP) density function whose lead term is the probability density function of a Gaussian ARCH model.[21] If the order of expansion inherent in the SNP increases with the sample size then the resulting II estimator will be (almost) efficient.[22] For obvious reasons, Gallant and Tauchen (1996) termed this version of II: Efficient Method of Moments. Expressing this notion in terms of the quasi-scores instead of QMLE - as Gallant and Tauchen (1996) do - then the argument is essentially the same as noted in the GMM context that is, asymptotic efficiency is achieved if the moments in the quasi score span the true score function.

The foregoing discussion deals with the case where the auxiliary model is a quasi-likelihood. Moment selection may also be an issue in other settings too, and has received some attention in the context of DSGE estimation based impulse response matching. In this setting, it is customary to use a relatively large number of impulse responses. Using the direct analogy to the impact of redundant moments in GMM, Hall, Inoue, Nason, and Rossi (2012) observe that the inclusion of (multiple) redundant impulse responses can lead to a severe deterioration in the quality of the first order asymptotic distribution theory as an approximation to the finite sample behaviour of the resulting estimator. They propose the Relevant Impulse Response Selection Criterion (RIRSC), modeled on RMSC in GMM, and that can be used to screen the candidate set of impulse responses to exclude those that provide no information. Hall, Inoue, Nason, and Rossi (2012) demonstrate that its use can improve the the quality of the first order asymptotic theory to the post-selection estimator.

The issue of selecting which moments are valid has also received attention. At first sight, this might seem strange as the maintained assumption so far is that the simulator is correct data generation process. However, Dridi, Guay, and Renault (2007) argue that DSGE models are inherently misspecified as a general representation of the economy. However, while we may not believe all of it, there may be parts of the model that permit consistent estimation of certain parameters that are of primary interest. So in this setting, if we partition the

[21] SNP densities are introduced in Gallant and Nychka (1987).

[22] The use of simulation introduces a multiplication factor of $(1 + 1/s)$ in the large sample variance, but this can be made arbitrarily close to 1 by making s large.

parameter vector again into $\theta' = (\phi', \psi')$, where ϕ represent the parameters of interest, ψ are "nuisance parameters",[23] then this setting is described by the existence of a binding function $b_1(\,\cdot\,)$ such that,

$$k_1(h_{*,1}, \phi_0, \psi_*) \;=\; h_{*,1} \;-\; b_1(\phi_0, \psi_*) \tag{15}$$

where $h_{1,*}$ is a subset of $h_*(\cdot)$, ϕ_0 represents the true value of the parameters of interest and ψ_* is the pseudo-true value of the nuisance parameters. This scenario is outside the original II framework, and so Dridi, Guay, and Renault (2007) extend it by both defining *partial* II (PII) estimators to cover this situation and presenting their first order asymptotic theory. To implement PII to obtain consistent estimators of the parameters of interest, it is necessary to identify the valid relations. Dridi, Guay, and Renault (2007) propose a sequential testing strategy based on a variant of the overidentifying restrictions test. Hall, Inoue, Nason, and Rossi (2012) propose a Valid Impulse Response Selection Criterion (VIRSC) that is an adaptation of Andrews's (1999) MSC to this setting.

As the above discussion suggests, the treatment of moment selection in II depends on the setting. If the auxiliary model is a quasi-score then moment selection is handled through using knowledge of the simulator to make a judicious choice of QML or through the use of a SNP-based QML. In these cases, there seems little scope for using the kind of methods developed within the GMM literature. However, in settings such as estimation of DSGE based on impulse response functions, the parallels between GMM and II moment selection seem far stronger. To date, moment selection strategies based on the overidentifying restrictions test and RMSC have been applied in this setting. In principle, there seems no reason why other approaches developed for special cases of GMM - such as selection based on minimizing second order mean square error - could not be applied in this context as well.

We conclude this section by considering the wider question of whether to use GMM or II in models where both are feasible, which is, in a sense, an issue of moment selection. In terms of first order asymptotics, the choice is one of efficiency. As noted above, II has the potential for (near) asymptotic efficiency with appropriate choice of auxiliary model. In contrast, GMM estimation can be based on moment conditions implied by the distribution such as, either the polynomial moments (mean, variance, skewness *etc.*) or the characteristic function (if feasible). Extant evidence suggests II dominates: for example see Garcia, Renault, and Veredas (2011) in the context of estimation of parameters of the stable distribution, or compare the results of simulation studies of GMM and EMM estimation of the stochastic volatility model in, respectively, Andersen and Sørensen (1996) and Andersen, Chung, and Sørensen (1999).

3.2 Identification failure

GMM:

[23] Some of which are estimated along with ϕ and others calibrated.

Asymptotic normality of the GMM estimator (Proposition 2 above) is predicated on first order local identification (Assumption 3(iii) above). It has been realized that in certain circumstances of interest this assumption fails or is close to doing so with the result that Proposition 2 either does not hold or provides a poor guide to behaviour in the sample sizes relevant to certain types of applications. The literature on this topic is voluminous and so here we confine our attention to describing the main ways in which violations of Assumption 3(iii) have been modeled and the consequences for asymptotic behaviour of GMM estimators.[24] As with moment selection, a large part of the GMM literature has focussed on the GIV case, and we note that in the case where this is applied to a linear model then identification (Assumption 2(iii)) and first order local identification are the same. In nonlinear models, they differ and this is something we return to below. Below, we use GIV-L to denote GIV applied to linear models.

Identification failure can either be complete or partial: if complete then ϕ_0 is unidentified; if partial then ϕ_0 is unidentified but certain linear combinations of ϕ_0 are identified. For GIV-L, Phillips (1989) shows $\hat{\phi}_{GMM}$ converges to a random limit and so consistency is lost, but if identification is partial then Choi and Phillips (1992) show that the identified linear combinations of ϕ_0 can be consistently estimated but have a non-standard asymptotic distribution.

In nonlinear models, ϕ_0 can be identified even if first order identification fails. Dovonon and Renault (2013) demonstrate that first order identification fails but second order, and hence global, identification holds in certain models of interest in asset pricing. In this case, the Jacobian is null and so the usual analysis behind the the first order asymptotic distribution does not apply. Dovonon and Renault (2013) show that in this case, the overidentifying restrictions test converges to a random variable whose distribution is bounded from below by $\chi^2_{q-p_\phi}$ and from above by χ^2_q, meaning the use of the standard first order asymptotic distribution leads to over-sized tests. We return to the implications of second order identification for the GMM estimator in Section 4.

The first order identification also does not hold if ϕ_0 is on the boundary of Φ. In this case, $\hat{\phi}_T$ cannot be characterized via the standard the first order conditions in (10) asymptotically and so the conventional first order analysis cannot be applied to deduce the limiting distribution in Proposition 2. Andrews (2002) demonstrates that the limiting distribution of $T^{1/2}(\hat{\phi}_T - \phi_0)$ is non-standard but can be simulated. Further, he shows that even if only a subvector of ϕ_0 is on the boundary, the limiting distributions of all elements of $T^{1/2}(\hat{\phi}_T - \phi_0)$ are non-standard unless a certain block-diagonality condition holds.

In the cases described above first order identification failure is exact. This case has received relatively little attention in the GMM literature to date.[25] Instead, driven by some high profile empirical examples, attention has focused on the case where first order identification is technically satisfied but in some

[24] A more comprehensive review is contained in Hall (2015).

[25] Arellano, Hansen, and Sentana (2012) propose methods for both testing for exact identification failure and also learning about the dimensions in which identification fails based on the overidentifying restrictions test statistic.

sense close to being violated. It is this scenario that is covered by the concepts of "weak" and "nearly-weak" identification.

Staiger and Stock (1997) introduced the concept of weak identification in GIV-L, Stock and Wright (2000) refined the concept and extended the analysis to nonlinear models. The key technical restriction behind weak identification is that the Jacobian is full rank - and so first order locally identified - for finite T but is converging to a rank deficient matrix at rate $T^{-1/2}$ so that first order local identification fails in the limit. Under weak identification, Stock and Wright (2000) demonstrate the standard first order asymptotic framework for GIV does not go through, with the limiting behaviour qualitatively the same as derived by Phillips (1989) and Choi and Phillips (1992) for GIV-L with exact identification failure. These analyses therefore indicate that even if identification holds but is close to failure then the standard first order asymptotic theory may provide a poor approximation to finite sample behaviour. This clearly raises a problem for a practitioner who is concerned the parameters of his/her model may be weakly identified. Two solutions suggest themselves: first, to pre-test the quality of the identification of candidate moments; second, to base inference on procedures that are robust to the quality of the identification. The first approach has been explored in the context of the linear model estimated by IV - for which the Jacobian only depends on the relationship between the endogenous regressors and instruments - with a number of different statistics being proposed.[26] The second approach is based on finding statistics that can be inverted to construct confidence sets for ψ_0 irrespective of the quality of the identification. Well known statistics of this type include the Anderson-Rubin (AR) statistic[27], the K- statistic (Kleibergen, 2002, 2006) and the conditional likelihood ratio (CLR) statistic (Moreira, 2003; Kleibergen, 2005). While such confidence sets have the attractive feature of being robust to failures of identification, the computational burden associated with their calculation increases with p_ϕ and makes this approach infeasible for large p_ϕ. This burden can be reduced if only a subset of the parameters are of primary interest; see *inter alia* Dufour and Taamouti (2005), Kleibergen and Mavroeidis (2009) and Chaudhuri and Zivot (2011). The key difference between these approaches and the standard Wald-type confidence intervals - "estimator" $\pm$ 2 $\times$ "standard error" - is that the confidence sets based on these identification-robust statistics can be infinite and non-contiguous whereas the Wald intervals are of finite length and contiguous by construction. Thus if ϕ_0 is unidentified, the identification-robust confidence set can be infinite demonstrating nothing has been learned about ϕ_0 from the model, whereas the Wald interval implies spuriously that something has been learnt.[28]

The weak identification framework is designed to approximate situations in which the information content of moments, while non-zero, is sufficiently low to

[26] For example see Cragg and Donald (1993), Hall, Rudebusch, and Wilcox (1996), Shea (1997) and Stock, Wright, and Yogo (2002).

[27] See Anderson and Rubin (1949), Dufour (1997) and Staiger and Stock (1997).

[28] If ϕ_0 is not first order identified then the Wald intervals are invalid; see Dufour (1997) for further discussion.

undermine standard first order asymptotic inferences. For this end, the choice of $T^{-1/2}$ as the rate of decay of Jacobian is critical. Hahn and Kuersteiner (2002) considered the the limiting behaviour of the Two Stage Least Squares (2SLS) estimator when the rate of decay is slower, a scenario they refer to as nearly-weak identification. They show that in this case consistency is restored and many conventional GMM statistics have the same properties as under standard first order asymptotics. The difference is that compared to the standard case, convergence to the limiting properties is slower.[29]

Both weak and nearly-weak identification are technical devices designed to understand how the estimator behaves in the case where first order identification is technically satisfied but in some sense close to being violated. Taken together, the derived results indicate the "proximity" to first order identification failure is key. Weak and nearly-weak identification yield different large sample theories that provide approximations that are appropriate in different circumstances. To our knowledge, it remains an open question as to how to decide which is more appropriate in nonlinear models in any given circumstance.

If the moments are less informative, one solution is to increase their number as the sample size increases. In linear models, Chao and Swanson (2005) establish conditions under which various estimators, including 2SLS, are consistent as the number of instruments increases with T. Han and Phillips (2006) consider the extension of this framework to nonlinear models estimated via GMM with a constant weighting matrix ($W_T = W$). They demonstrate that many different types of asymptotic behaviour of such estimators - including consistency and asymptotic normality - are possible depending on the rate of growth of information about θ_0 as the number of moments (q_T) increases. Collectively these results indicate the key feature here is the rate at which the information in the moment conditions increases as the set of moments is expanded. However, extant evidence suggests that GMM is dominated in such scenarios by other moment-based estimators, such as Generalized Empirical Likelihood; see Bekker (1994), Chao and Swanson (2005), Hansen, Hausman, and Newey (2008), and Newey and Windmeijer (2009).

II:

While all the original presentations of SMM-II-EMM note the need for first order local identification in order to justify the first order asymptotic framework discussed above, the consequences of its violation has received far less attention in this context than in GMM. This can be explained in the II-EMM case by the focus on QMLE of the auxiliary model, and the freedom of the researcher to choose an auxiliary model that is convenient for the situation in hand. However, in other settings, the assumption of first order local identification may be more tenuous.

Canova and Sala (2009) have raised concerns about the nature of the identification in the context of DSGE models estimated by matching impulse response functions. Here $h(\cdot)$ is a composite mapping, $h(\cdot) = r(\alpha(\theta))$ where

[29] Antoine and Renault (2009) and Caner (2010) extended these results to nonlinear models.

- $r(\cdot)$ is the mapping from the impulse responses to the parameters of the auxiliary model (the VAR), termed the "moment mapping" by Canova and Sala (2009);
- $\alpha(\cdot)$ is the mapping from the auxiliary model parameters to the parameters of the DSGE, termed the "solution mapping".

Canova and Sala (2009) observe that solution mapping is typically highly nonlinear, and so it can be that θ_0 is only partially identified. Furthermore, weak identification can also be present. Within this setting, the Jacobian of $h(\cdot)$, H say, equals RA where R is the Jacobian of the moment mapping and A, the Jacobian of the solution mapping. By analogy to GMM, the problems associated with weak identification will be present if H is rank deficient or close to being so. Clearly, as pointed out by Canova and Sala (2009), this can happen here due to rank deficiency in either R or A. As in GMM this can be difficult to assess *a priori*, especially as these derivatives may have to be obtained via simulation. Canova and Sala (2009) recommend examining the eigenvalues (or some function thereof) of $R'R$ and $A'A$ to investigate whether either (or both) of these two mappings may be a source of identification problems.

If the identification problems stem from the auxiliary model then one solution is the use of Constrained Indirect estimation (Calzolari, Fiorentini, and Sentana, 2004). This method extends the original Indirect Estimation framework (described in Section 2) by allowing restrictions on the parameter space of the auxiliary model. If these problem areas of the parameter space in the auxiliary model can be deduced *a priori* then it makes sense to avoid them as the auxiliary model is just a target for the simulator. In such settings, the potential problems caused by identification are simply side-stepped by appropriately restricting the parameter space over which the auxiliary model is estimated.

However, in more general settings, this solution may not be available because the identification problems either apply to the whole parameter space of the auxiliary or to the simulator. For these cases, it is, in principle, possible to follow the approaches taken in the GMM literature to develop a companion theory of II. However, to our knowledge, there are no formal treatments of II when first order local identification fails or is close to failing. We return to this topic in Section 4.

3.3 Misspecified models

GMM:
The first order asymptotic theory in Section 2 is predicated on the assumption that the population moment condition is correct. While this is typically the assumption under which inference is performed, there are circumstances when the underlying model is acknowledged to be misspecified and that as a result the population moment condition is invalid. A leading example is in asset pricing. Hansen and Jaganathan (1991) demonstrate the mean and standard deviation of the stochastic discount factor (SDF) that prices a set of assets must fall in admissible region that can be estimated nonparameterically. It has been found

that in many cases parametric forms of the SDF do not attain this admissible region, and thus are misspecified. Hansen and Jaganathan (1997) propose a measure of the size of the pricing error (made with a misspecifed SDF), known thereafter as the "Hansen-Jaganathan (HJ) distance". Hansen, Heaton, and Luttmer (1995) present methods for testing hypotheses about the HJ distance; Kan and Robotti (2009) present methods for testing which of two models has a smaller HJ distance.[30] Within this context, it may also be of interest to perform inference about the parameters of the proxy SDF. Hall and Inoue (2003) develop an asymptotic distribution theory for GMM estimators in overidentified models.[31] They show the GMM estimator converges to the "pseudo-true value", θ_*, but that the rate of convergence and θ_* itself depends on the weighting matrix. They further show that, in certain leading cases, $T^{1/2}(\hat{\theta}_T - \theta_*)$ converges to a mean zero normal distribution with a variance that is different from V_ϕ in Proposition 2 but can be consistently estimated.

II:
As mentioned in Section 3.1, the idea of inference within misspecified models has arisen in discussion of the interpretation of DSGE models. Dridi, Guay, and Renault (2007) have argued in this context that the simulator cannot be considered the true data generation process. However, there might be parts of the model that can be used to consistently estimate the parameters of interest, as described in (15). They propose the PII for estimation in this case. Here the parameters of interest are consistently estimated, and the misspecified nature of the model manifests itself through the interpretation of ψ_* as a pseudo-true value, and in the variance of the first order asymptotic distribution of PII estimator of ϕ_0 as a distinction now needs to be made in the limits of certain matrices evaluated at the actual and simulated data.

The above scenario is possible in DSGE models given the underlying models cover different aspects of the economy. In simpler settings, misspecification may be anticipated to impact on all the parameters of interest. This gives rise to the question of how the parameters should be interpreted in this setting. Given the similarities between II and GMM, it is not surprising that the qualitative features of GMM analysis in misspecified models carries over to II. Specifically, the II estimator converges in probability to a pseudo-true value defined as the argmin of the population analog to the minimand in (5)- a value that depends then on the weighting matrix; see Aguirre-Torres and Toribio (2004) for EMM, and Oh and Patton (2013) for SMM estimation of copula-based multivariate models.[32] Aguirre-Torres and Toribio (2004) also present a limiting distribu-

[30] See also *inter alia* Gospodinov, Kan, and Robotti (2013).

[31] Maasoumi and Phillips (1982) present the large sample behaviour of IV in misspecified linear models.

[32] Oh and Patton (2013) explore the use of SMM to estimate copula-based multivariate models in which the parameters are estimated by matching certain "pure" dependence measures, such as Spearmans's rank correlation, that are unaffected by the marginal distributions of the data. They establish conditions for the consistency and asymptotic normality of the estimators of the parameters indexing the copula.

tion for EMM under misspecification. Their analysis highlights an important difference between GMM and EMM: in GMM, misspecification inevitably implies the moment on which estimation is based cannot be set to zero; in EMM, misspecification of the simulator is compatible with both $k(h_*, \theta_0) = 0$ and $k(h_*, \theta_0) \neq 0$. The intuition is best understood by thinking in terms of matching the score of the auxiliary model. For a finite order SNP, the score involves a finite set of moments and there is always the potential that these can be matched by the misspecified simulator. This is important because it is the non-zero mean of the moment evaluated at the pseudo-true value that complicates the asymptotic behavior of the GMM estimator. Thus, Aguirre-Torres and Toribio (2004) show that the limiting distribution of EMM is asymptotically normal but its variance depends critically on whether or not $k(h_*, \theta_0)$ equals zero. They note, however, that as the order of the SNP increases, it is the case with $k(h_*, \theta_0) \neq 0$ that must ultimately apply because as the order of the SNP increases then so do the number of moments involved in the score of the auxiliary model, and a misspecified distribution can not match all the moments of the data.

Therefore, the parallels between the treatments of misspecification in GMM and II depend on the context of the II estimation.

4 GMM and II asymptotic behaviour under first order local identification failure

In this section, we consider the moment condition model (1) and study the asymptotic behaviour of the GMM estimator when the standard local identification condition (Assumption 3(iii)) fails. We also derive the asymptotic distribution of the II estimator when the auxiliary model is a moment condition model that has such a local identification issue.

4.1 Asymptotic distribution of the GMM estimator

While the global identification condition ensures consistency of the GMM estimator, its the asymptotic distribution depends on how sharply the moment function $m(\phi) \equiv Eg(X, \phi)$ moves away from 0 in the neighborhood of ϕ_0. Standard results are derived under the so-called first order local identification condition, i.e. $M(\phi_0)$ has rank p_ϕ. But in nonlinear models, global identification is possible without first order local identification as highlighted by the example in Section A of Appendix. Sargan (1983) has studied the IV estimator in this context whereas Dovonon and Renault (2009, 2013) have recently studied the GMM overidentification test when the moment condition moment is rank deficient at the true parameter value. In globally identified models, local identification can be ensured by higher order derivatives of the moment function $m(\phi)$. The second order local identification condition is introduced by Dovonon and Renault (2009) 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 $\frac{\partial m'}{\partial \phi}(\phi_0)$ and all v in the null space of $\frac{\partial m}{\partial \phi'}(\phi_0)$, we have:

$$\left(\frac{\partial m}{\partial \phi'}(\phi_0)u + \left(v' \frac{\partial^2 m_k}{\partial \phi \partial \phi'}(\phi_0) v \right)_{1 \le k \le q} = 0 \right) \Rightarrow (u = v = 0).$$

Without requiring that the Jacobian matrix $M(\phi_0)$ has 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.

We will 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. 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:

$$Rank \begin{pmatrix} \frac{\partial m}{\partial \phi^{1'}}(\phi_0) & \frac{\partial^2 m}{\partial \phi_{p_\phi}^2}(\phi_0) \end{pmatrix} = p_\phi,$$

where ϕ is partitioned into $(\phi^{1'}, \phi_{p_\phi})'$. 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 $G = \frac{\partial^2 m}{\partial \phi_{p_\phi}^2}(\phi_0)$, 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) $\sqrt{T} \begin{pmatrix} m_T(\phi_0) \\ \frac{\partial m_T}{\partial \phi_{p_\phi}}(\phi_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbb{Z}_0 \\ \mathbb{Z}_1 \end{pmatrix}$.

(iii) $W_T - W = o_P(T^{-1/4})$, $\frac{\partial m_T}{\partial \phi^{1'}}(\phi_0) - D = O_P(T^{-1/2})$, $\frac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_0) - G = O_P(T^{-1/2})$ *and* $\frac{\partial^2 m_T}{\partial \phi^{1'} \partial \phi_{p_\phi}}(\lambda_0) - G_{1p_\phi} = o_P(1)$, *with* $G_{1p_\phi} = \frac{\partial^2 m}{\partial \phi^{1'} \partial \phi_{p_\phi}}(\phi_0)$.

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) are 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 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 of $\begin{pmatrix} W^{1/2}D & W^{1/2}G \end{pmatrix}$:

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

Let

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

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,

$$\sqrt{T}\begin{pmatrix} \hat{\phi}^1 - \phi_0^1 \\ (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} H\mathbb{Z}_0 + HG\mathbb{V}/2 \\ \mathbb{V} \end{pmatrix},$$

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 $m_T'(\hat{\phi})W_T m_T(\hat{\phi})$ which are collected into $K_T(\phi)$ as given by (39) in 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:

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

showing that the minimum is reached at $(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})$ having opposite sign to R_{1T}. See (40) 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 $H = q = 1$ (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.*

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 example in Section A of Appendix. 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: $\lambda = R\eta$ and consider the model:

$$E\left(g(X, R\eta)\right) = 0. \tag{17}$$

By the chain rule, it is not hard to see that Model (17) identifies $\eta_0 = R^{-1}\phi_0$ with local identification properties matching Assumption 4. More precisely, we have:

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

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:

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

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

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

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,

$$B_T \hat{R}^{-1}(\hat{\phi} - \phi_0) = B_T R^{-1}(\hat{\phi} - \phi_0) + \epsilon_T \tag{18}$$

$\epsilon_T = -B_T \hat{R}^{-1}(\hat{R} - R) R^{-1}(\hat{\phi} - \phi_0)$. 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 $B_T \hat{R}^{-1}(\hat{\phi} - \phi_0)$ is asymptotically distributed as $\tilde{\mathbb{X}}$

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:

$$\epsilon_T = -A\sqrt{T}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi})^2 + o_P(1), \tag{19}$$

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, $B_T \hat{R}^{-1}(\hat{\phi} - \phi_0)$ is asymptotically distributed as $\tilde{\mathbb{X}} - A(\tilde{\mathbb{X}}_{p_\phi})^2$. It is worth mentioning that this change of joint asymptotic distribution of $\hat{\phi} - \phi_0$ does not affect the marginal distribution of its slowest converging linear combinations.

Our example in Section A of Appendix falls in *Case 2* since possible choices of R depend on the true value of the parameter of interest. We can actually choose:

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2\delta\Omega_2 \end{pmatrix}, \qquad \hat{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2\hat{\delta}\hat{\Omega}_2 \end{pmatrix},$$

with $\hat{\Omega}_2 = \hat{h}_2 - \hat{h}_3/\hat{\delta}$.

4.2 Asymptotic distribution of II with under-identified auxiliary

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:

$$E[g(x,h)] = 0, \tag{20}$$

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 (20) using the data and simulated series providing the sequences h_T and $h_T^{(i)}(\theta)$, $i = 1, \ldots, s$ that are the auxiliary features used to estimate the parameter of interest θ by the quadratic optimization (5).

We assume that (20) 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 continue 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:

$$\hat{\theta}_{II} - \theta_0 = -\left(M_{IT}(\hat{\theta}_{II})'\Omega_T M_{IT}(\dot{\theta}_T) \right)^{-1} M_{IT}(\hat{\theta}_{II})'\Omega_T m_{IT}(\theta_0).$$

That is:

$$\hat{\theta}_{II} - \theta_0 = \dot{F}_T \left(h_T - \frac{1}{s}\sum_{i=1}^{s} h_T^{(i)}(\theta_0) \right), \tag{21}$$

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 $i = 1, \ldots, s$ 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[33], 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 $\mathbb{X}_0, \mathbb{X}_1, \ldots \mathbb{X}_s$ 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 complicate than in the standard case. Pre-multiplying (21) by $T^{1/4}$, we have:

$$T^{1/4}(\hat{\theta}_{II} - \theta_0) = \dot{F}_{T,\bullet\ell} T^{1/4} m_{IT,\ell}(\theta_0) + o_P(1) = F_{\bullet\ell} T^{1/4} m_{IT,\ell}(\theta_0) + o_P(1), \quad (22)$$

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 $\mathbb{X}$.

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 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 other 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 imposed 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 not 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:

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

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

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:

$$\hat{\theta}_{II} - \theta_0 = \hat{F}_T \left(m_{IT}(\theta_0) - \frac{1}{2} \left((\hat{\theta}_{II} - \theta_0)' \Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0) \right)_{1 \leq k \leq \ell} \right), \quad (23)$$

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 $\hat{S}_T$ 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 $\hat{S}_T^1$ be the $(p-1) \times p$ submatrix of the first $(p-1)$ rows of $\hat{S}_T$. The last remark in this section gives how the matrix $\hat{S}_T$ can be determined as a continuous function of the last column of $\hat{F}_T$. By definition, $\hat{S}_T^1 \hat{F}_T m_{IT}(\theta_0)$ does not depend on the slow converging component, $m_{IT,\ell}(\theta_0)$, of $m_{IT}(\theta_0)$. We therefore have:

$$\begin{aligned} & \sqrt{T} \hat{S}_T^1 \left(\hat{\theta}_{II} - \theta_0 \right) \\ = \quad & \hat{S}_T^1 \hat{F}_T B_T \left(m_{IT}(\theta_0) - \tfrac{1}{2} \left((\hat{\theta}_{II} - \theta_0)' \Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II} - \theta_0) \right)_{1 \leq k \leq \ell} \right). \end{aligned} \quad (24)$$

By combining (22) and (24) and letting S be the probability limit of $\hat{S}_T$ 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:*

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

where S^1 is the submatrix 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 $B_{IT} \hat{S}_T (\hat{\theta}_{II} - \theta_0)$ 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 $\mathbb{Y}$ which is based on that of $\mathbb{X}$ which is described in the previous section.

Remark 3. *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 (21), 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 (23), 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 S_{RT} be row-wise, the orthonormal basis obtained by completing the last column of $\hat{F}_T\hat{R}$ according to Remark 4 below,and S_R its probability limit, we have that:

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

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 4. *Let us now describe a procedure that can be used to determine the matrix of orthogonal directions S_T 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 $\Re^p$, 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.

5 Concluding remarks

In this paper, we make two contributions. Firstly, we explore the connections between Generalized Method of Moments and Indirect Inference, and secondly, 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.

Our examination of the connections between GMM and II reveal some interesting similarities and differences between the methods. Although GMM and II are implemented in different ways, both can be viewed as "minimum chi-squared" methods and hence share the same linear algebraic structure of their first order analyses, although the regularity conditions underlying each are different in important ways. Since both methods involve moment-based estimation, it is natural to expect that issues relevant to GMM estimation are also relevant to II. In this paper, we investigate the extent to which this is the case, focusing on three particular topics: moment selection, identification failure and inference in misspecified models. We find that extent of the influence of GMM analyses of these topics for II depends on the context of the II estimation.

In applications such as estimation of stochastic volatility models, it is natural to choose the auxiliary model to be a quasi-likelihood based on a distribution with similar properties to the simulator. In such cases, there seems little need for GMM-type moment selection methods or GMM-type weak identification robust inference procedures. However, in applications such as estimation of DSGE models, the auxiliary model consists of impulse response functions and GMM-type methods of moment selection seem more relevant. Furthermore, DSGE models are often highly nonlinear and so it is quite possible that parameters may be only partially or weakly or second-order identified. All these scenarios for identification failures, near or exact, have been explored in the GMM literature to some degree but, to our knowledge, these approaches have previously not been extended to II.

Within the GMM framework, misspecification implies the population moment condition is invalid, and as a result the estimator is inconsistent and standard first order asymptotic theory does not apply. Within II, misspecification implies the simulator is not the true model and, again, the consequences of this depend on the context. In applications such as the stochastic volatility model, the consequences for II are similar to those for GMM. In applications to

DSGE, the focus has been on finding parts of the model that permit consistent estimation of certain parameters of interest. This approach is termed Partial II (PII), and the first order asymptotic distribution is similar to that of II once allowance is made for the fact that other aspects of the model are wrong.

Our second contribution is to present the limiting distribution of both the GMM estimator under second order identification and also the II estimator in cases where the auxiliary model is 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. 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 therefore be interesting to explore ways to generate confidence sets based on these estimators that are robust to the rank deficiency issue. 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

Aguirre-Torres, V., and Toribio, M. D. (2004). 'Efficient Method of Moments in misspecified i.i.d. models', *Econometric Theory*, 20: 513–534.

Anatolyev, S. (2003). 'The form of the optimal nonlinear instrument for multiperiod conditional moment restrictions', *Econometric Theory*, 19: 602609.

Andersen, T. G., Chung, H.-J., and Sørensen, B. (1999). 'Efficient method of moments of a stochastic volatility model: a Monte Carlo study', *Journal of Econometrics*, 91: 61–87.

Andersen, T. G., and Sørensen, B. (1996). 'GMM estimation of a stochastic volatility model: a Monte Carlo study', *Journal of Econometrics*, 14: 328–352.

Anderson, T. W., and Rubin, H. (1949). 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *Annals of Mathematical Statistics*, 20: 46–63.

Andrews, D. W. K. (1999). 'Consistent moment selection procedures for Generalized Method of Moments estimation', *Econometrica*, 67: 543–564.

——— (2002). 'Generalized Method of Moments estimation when a parameter is on a boundary', *Journal of Business and Economic Statistics*, 20: 530–544.

Andrews, D. W. K., and Lu, B. (2001). 'Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models', *Journal of Econometrics*, 101: 123–164.

Antoine, B., and Renault, E. (2009). 'Efficient GMM with nearly-weak instruments', *The Econometrics Journal*, 12: S135–S171.

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

Barankin, E., and Gurland, J. (1951). 'On asymptotically nor,al efficient estimators: I', *University of California Publications in Statistics*, 1: 86–130.

Bekker, P. A. (1994). 'Alternative approximations to the distributions of instrumental variables estimators', *Econometrica*, 63: 657–681.

Belloni, D., Chen, D., Chernozhukov, V., and Hansen, C. (2012). 'Sparse models and methods for optimal instruments with an application to eminent domain', *Econometrica*, 80: 2369–2430.

Breusch, T., Qian, H., Schmidt, P., and Wyhowski, D. (1999). 'Redundancy of moment conditions', *Journal of Econometrics*, 91: 89–111.

Calzolari, G., Fiorentini, G., and Sentana, E. (2004). 'Constrained indirect estimation', *Review of Economic Studies*, 71: 945–973.

Caner, M. (2010). 'Testing, estimation in GMM and CUE with nearly-weak identification', *Econometric Reviews*, 29: 330–363.

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

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

Carrasco, M., and Florens, J.-P. (2000). 'Generalization of GMM to a continuum of moment conditions', *Econometric Theory*, 16: 797–834.

Chao, J., and Swanson, N. (2005). 'Consistent estimation with alarge number of weak instruments', *Econometrica*, 73: 1673–1692.

Chaudhuri, S., and Zivot, E. (2011). 'A new method of projection-based inference in GMM with weakly identified nuisance parameters', *Journal of Econometrics*, 164: 239–251.

Cheng, X., and Liao, Z. (2013). 'Select the valid and relevant moments: an information-based LASSO for GMM with many moments', Discussion paper, Department of Economics, University of Pennsylvania, Philadeplhia PA, USA.

Choi, I., and Phillips, P. C. B. (1992). 'Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations', *Journal of Econometrics*, 51: 113–150.

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.

Cragg, J. G., and Donald, S. G. (1993). 'Testing identifiability and specification in instrumental variables', *Econometric Theory*, 9: 222–240.

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.

Donald, S. G., and Newey, W. K. (2000). 'A jackknife interpretation of the continuous updating estimator', *Economics Letters*, 67: 239–243.

Dovonon, P. (2013). 'Conditionally heteroskedastic factor models with skewness and leverage effects', *Journal of Applied Econometrics*, 28: 1110–1137.

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

Dridi, R., Guay, A., and Renault, E. (2007). 'Indirect inference and calibration of dynamic stochastic general equilibrium models', *Journal of Econometrics*, 136: 397–430.

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., and Taamouti, M. (2005). 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica*, 73: 1351–1365.

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.

Gallant, A. R., and Nychka, D. W. (1987). 'Semi-nonparametric maximum likelihood estimation', *Econometrica*, 55: 363–390.

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.

——— (2004). 'Testing for structural change in the presence of auxiliary models', *Econometric Theory*, 20: 1168–1202.

Gospodinov, N., Kan, R., and Robotti, C. (2013). 'Chi-squared tests for evaluation and comparison of asset pricing models', *Journal of Econometrics*, 173: 108–125.

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

Hahn, J., and Kuersteiner, G. (2002). 'Discontinuities of weak instruments limiting distributions', *Economics Letters*, 75: 325–331.

Hall, A. R. (2015). 'Econometricians have their moments: GMM at 32', Discussion paper, Department of Economics, University of Manchester, Manchester UK.

Hall, A. R., and Inoue, A. (2003). 'The large sample behaviour of the Generalized Method of Moments estimator in misspecified models', *Journal of Econometrics*, 114: 361–394.

Hall, A. R., Inoue, A., Jana, K., and Shin, C. (2007). 'Information in Generalized Method of Moments Estimation and Entropy Based Moment Selection', *Journal of Econometrics*, 138: 488–512.

Hall, A. R., Inoue, A., Nason, J. M., and Rossi, B. (2012). 'Information criteria for impulse response function matching estimation in DSGE models', *Journal of Econometrics*, 170: 499–518.

Hall, A. R., and Peixe, F. P. M. (2003). 'A consistent method for the selection of relevant instruments', *Econometric Reviews*, 22: 269–288.

Hall, A. R., Rudebusch, G., and Wilcox, D. (1996). 'Judging instrument relevance in instrumental variables estimation', *International Economic Review*, 37: 283–298.

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

Han, C., and Phillips, P. C. B. (2006). 'GMM with many moment conditions', *Econometrica*, 74: 147–192.

Hansen, C., Hausman, J., and Newey, W. K. (2008). 'Estimation with many instrumental variables', *Journal of Business and Economic Statistics*, 26: 398–422.

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

——— (1985). 'A method of calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators', *Journal of Econometrics*, 30: 203–238.

Hansen, L. P., Heaton, J., and Luttmer, E. G. J. (1995). 'Econometric evaluation of asset pricing models', *The Review of Financial Studies*, 8: 237–274.

Hansen, L. P., Heaton, J., and Ogaki, M. (1988). 'Efficiency bounds implied by multi–period conditional moment restrictions', *Journal of the American Statistical Association*, 83: 863–871.

Hansen, L. P., and Jaganathan, R. (1991). 'Implications of security market data for models of dynamic economies', *Journal of Political Economy*, 99: 225–262.

——— (1997). 'assessing specification errors in stochastic discount factor models', *Journal of Finance*, 52: 557–590.

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

Hayashi, F., and Sims, C. (1983). 'Nearly efficient estimation of time series models with predetermined, but not exogenous instruments', *Econometrica*, 51: 783–798.

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

Heaton, J., and Ogaki, M. (1991). 'Efficiency bound calculations for a time series model with conditional heteroscedasticity', *Economic Letters*, 35: 167–171.

Kan, R., and Robotti, C. (2009). 'Model comparison using the Hansen-Jaganathan distance', *The Review of Financial Studies*, 22: 3449–3490.

Kleibergen, F. (2002). 'Pivotal statistics for testing structural parameters in instrumental variables regression', *Econometrica*, 70: 1781–1803.

——— (2005). 'Testing parameters in GMM without assuming that they are ideintified', *Econometrica*, 73: 1103–1124.

Kleibergen, F., and Mavroeidis, S. (2009). 'Weak instrument robust tests in GMM and the new Keynesian Phillips curve', *Journal of Business and Economic Statistics*, 27: 293–310.

Liao, Z. (2013). 'Adaptive GMM shrinkage estimation with consistent moments election', *Econometric Theory*, 29: 1–48.

Maasoumi, E., and Phillips, P. C. B. (1982). 'On the behaviour of inconsistent instrumental variable estimators', *Journal of Econometrics*, 19: 183–201.

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.

Moreira, M. J. (2003). 'A conditional likelihood ratio test for structural models', *Econometrica*, 71: 1027–1048.

Newey, W. K. (1990). 'Efficient instrumental variables estimation of nonlinear models', *Econometrica*, 58: 809–838.

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 Windmeijer, F. (2009). 'Generalized Method of Moments with many weak moment conditions', *Econometrica*, 77: 687–719.

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.

Oh, D. H., and Patton, A. J. (2013). 'Simulated Mtheod of Moments estimation for copula-based multivariate models', *Journal of the American Statistical Association*, 108: 689–700.

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.

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

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

Sargan, J. D. (1958). 'The estimation of economic relationships using instrumental variables', *Econometrica*, 26: 393–415.

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

Shea, J. (1997). 'Instrument relevance in multivariate linear models', *Review of Economic Studies*, 79: 348–352.

Singleton, K. J. (2001). 'Estimation of affine asset pricing models using the empirical characteristic function', *Journal of Econometrics*, 102: 111–141.

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

Sowell, F. (1996). ‘Optimal tests of parameter variation in the Generalized Method of Moments framework’, *Econometrica*, 64: 1085–1108.

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

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

Stock, J. H., Wright, J. H., and Yogo, M. (2002). ‘A survey of weak instruments and weak identification in generalized method of moments’, *Journal of Business and Economic Statistics*, 20: 518–529.

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

——— (2013b). *Prizes in Economic Sciences 2013, Scientific Background*. http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/advanced-economicsciences2013.pdf.

West, K. D., Wong, K.-F. J., and Anatolyev, S. (2009). ‘Instrumental variables estimation of heteroskedastic linear models using all lags of instruments’, *Econometric Reviews*, 28: 441467.

A Example: Indirect estimation of the conditionally heteroskedastic factor model

In this example, we consider the conditionally heteroskedastic factor CHF) model that is specified in terms of parameter θ to be estimated. We derive a natural auxiliary model implied by this CHF model. The auxiliary model is a moment condition model with parameter vector h that can be expressed in terms of θ ($h \equiv h(\theta)$) allowing for indirect inference on θ. We show that the auxiliary model is first order locally underidentified and indirect inference on θ can be performed via our asymptotic theory in Section 4.

Consider the conditionally heteroskedastic factor model of two asset returns:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} f_t + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \tag{25}$$

with

$$\begin{gathered} E\left((f_t, u_t)|\mathfrak{F}_{t-1}\right) = 0,\ Var\left(f_t|\mathfrak{F}_{t-1}\right) = \sigma^2_{t-1}, \\ Var((u_{1t}, u_{2t})'|\mathfrak{F}_{t-1}) = \mathrm{Diag}(\Omega_1, \Omega_2),\ Cov(f_t, u_t|\mathfrak{F}_{t-1}) = 0. \end{gathered} \tag{26}$$

In this model, f_t is the latent common GARCH factor, u_t is the vector of idiosyncratic shocks and σ^2_{t-1} 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. The conditions $\gamma_1 > 0$ and $Var(f_t) = 1$ are added for identification purpose.

This model has been introduced by Diebold and Nerlove (1989) and further studied by Fiorentini, Sentana, and Shephard (2004) and Doz and Renault (2006). It is sometimes assumed that $(f_t, u_t)'$ is conditionally normally distributed.

We are interested in estimating the parameter vector $\theta \equiv (\gamma_1, \gamma_2, \Omega_1, \Omega_2)'$.

Estimation: In the literature, this model has been estimated by:

- *Kalman filter* and other *simulation methods* (Fiorentini, Sentana, and Shephard, 2004)): They specify an AR(1) dynamics for σ^2_t along with some distributional assumption for f_t and u_t. They are then able to write some state-space representation for the model that can be optimally estimated when the assumed distribution is correct.
- *GMM*: Moment conditions are derived that identify all parameters up to one (say, γ_1) that is given a 'reasonable' value. (See Doz and Renault, 2006; Dovonon, 2013). An extra benefit from doing this sort of calibration is that $Var(u_t|\mathfrak{F}_{t-1})$ can be identified even if it is nondiagonal.)

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 appropriate instrument z_{t-1} from $\mathfrak{F}_{t-1}$, (e.g lagged square returns), we have:

$$E\left\{\begin{pmatrix} 1 \\ z_{t-1} \end{pmatrix} \left[(y_{1t} - \delta y_{2t})^2 - c\right]\right\} = 0.$$

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

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

The auxiliary model is defined as:

$$\begin{array}{rcl} E\left\{\begin{pmatrix} 1 \\ z_{t-1} \end{pmatrix}\left[(y_{1t} - \delta y_{2t})^2 - c\right]\right\} & = & 0 \\ Ey_{1t}^2 & = & b_1 \\ Ey_{2t}^2 & = & b_2 \\ Ey_{1t}y_{2t} & = & b_3. \end{array} \tag{27}$$

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

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

at the true parameter value is:

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

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:

$$\left(2\delta\Omega_2\begin{pmatrix} 1 \\ Ez_{t-1} \end{pmatrix} \quad -\begin{pmatrix} 1 \\ Ez_{t-1} \end{pmatrix}\right)$$

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

$$\begin{pmatrix} 0 & 0 & 0 & 2\delta\Omega_2 & -1 \\ 0 & 0 & 0 & 2\delta\Omega_2 Ez_{t-1} & -Ez_{t-1} \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}$$

which is of rank 4 instead of 5.

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 ϕ_0^1 yields:

$$m_T(\hat{\phi}^1, \hat{\phi}_{p_\phi}) = m_T(\phi_0^1, \hat{\phi}_{p_\phi}) + \frac{\partial m_T}{\partial \phi^{1\prime}}(\bar{\phi}^1, \hat{\phi}_{p_\phi})(\hat{\phi}^1 - \phi_0^1),$$

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:

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

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 $\hat{\phi}_{p_\phi} - \phi_{0,p_\phi} = o_P(1)$, we have:

$$\begin{aligned} m_T(\hat{\phi}) &= m_T(\phi_0) + \tfrac{\partial m_T}{\partial \phi^{1\prime}}(\bar{\phi}^1, \hat{\phi}_{p_\phi})(\hat{\phi}^1 - \phi_0^1) \\ &\quad + \tfrac{1}{2}\tfrac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_0^1, \bar{\phi}_{p_\phi})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(T^{-1/2}). \end{aligned} \tag{28}$$

Let us define $\bar{D} = \frac{\partial m_T}{\partial \phi^{1\prime}}(\bar{\phi}^1, \hat{\phi}_{p_\phi})$ and $\bar{G} = \frac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_0^1, \bar{\phi}_{p_\phi})$. Pre-multiplying (28) by $\bar{D}' W_T$, we get

$$\begin{aligned} \hat{\phi}^1 - \phi_0^1 &= \left(\bar{D}' W_T \bar{D}\right)^{-1} \bar{D}' W_T \left(m_T(\hat{\phi}) - m_T(\phi_0)\right) \\ &\quad - \tfrac{1}{2}\left(\bar{D}' W_T \bar{D}\right)^{-1} \bar{D}' W_T \bar{G}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 + o_P(T^{-1/2}). \end{aligned} \tag{29}$$

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 (28), 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,

$$\begin{aligned} & m_T'(\hat{\phi}) W_T m_T(\hat{\phi}) \\ = \; & m_T'(\phi_0) W_T m_T(\phi_0) + \tfrac{1}{4}\bar{G}' W_T^{1/2} \bar{M}_d W_T^{1/2} \bar{G}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 \\ & + (\hat{\phi}_{p_\phi} - \phi_{0,1})^2 O_P(T^{-1/2}) + O_P(T^{-1}) \end{aligned} \tag{30}$$

The orders of magnitude in (30) 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 $m'_T(\hat\phi)W_T m_T(\hat\phi) \le m'_T(\phi_0)W_T m_T(\phi_0)$ (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 right hand side of (30) is less or equal to $m'_T(\phi_0)W_T m_T(\phi_0)$ and this gives:

$$\begin{aligned}&\tfrac{1}{4}G'W^{1/2}M_d W^{1/2}GT(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^4 + o_P(1)T(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^4\\ \le\quad & O_P(1)+\sqrt{T}(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2 O_P(1)\end{aligned} \tag{31}$$

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 (31) 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 $\hat\phi^1-\phi^1_0 = O_P(T^{-1/2})$ from (29).

(b) From (a) and (28), we have

$$m_T(\hat\phi) = m_T(\phi_0) + D(\hat\phi^1-\phi^1_0) + \frac{1}{2}G(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2 + o_P(T^{-1/2}).$$

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

$$(D'+o_P(1))W\left(\sqrt{T}m_T(\phi_0) + D\sqrt{T}(\hat\phi^1-\phi^1_0) + \frac{1}{2}G\sqrt{T}(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2 + o_P(1)\right) = 0.$$

This gives:

$$\sqrt{T}(\hat\phi^1-\phi^1_0) = -(D'WD)^{-1}D'W\left(\sqrt{T}m_T(\phi_0) + \frac{1}{2}G\sqrt{T}(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2\right) + o_P(1). \tag{32}$$

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

$$\begin{aligned}&\left(G'T^{1/4}(\hat\phi_{p_\phi}-\phi_{0,p_\phi}) + o_P(1)\right)\\ \times\quad & W\left(\sqrt{T}m_T(\phi_0) + D\sqrt{T}(\hat\phi^1-\phi^1_0) + \tfrac{1}{2}G\sqrt{T}(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2 + o_P(1)\right) = 0.\end{aligned} \tag{33}$$

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 (32) into (33), we get:

$$\begin{aligned}&T^{1/4}(\hat\phi_{p_\phi}-\phi_{0,p_\phi})\\ \times\quad & \left(G'W^{1/2}M_d W^{1/2}\sqrt{T}m_T(\phi_0) + \tfrac{1}{2}G'W^{1/2}M_d W^{1/2}G\sqrt{T}(\hat\phi_{p_\phi}-\phi_{0,p_\phi})^2\right) = o_P(1).\end{aligned} \tag{34}$$

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 (34), $(\mathbb{Z}_0,\mathbb{V}_0)$ satisfies:

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

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 (32) 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:

$$\begin{aligned}
& m_T(\hat{\phi}) \\
= \quad & m_T(\phi_0) + \frac{\partial m_T}{\partial \phi^{1\prime}}(\phi_0)(\hat{\phi}^1 - \phi_0^1) + \frac{\partial m_T}{\partial \phi_p}(\phi_0)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \frac{1}{2}\frac{\partial^2 m_T}{\partial \phi_{p_\phi}^2}(\phi_0)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
& + \frac{\partial^2 m_T}{\partial \phi_{p_\phi}\partial \phi^{1\prime}}(\phi_0)(\hat{\phi}^1 - \phi_0^1)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \frac{1}{6}\frac{\partial^3 m_T}{\partial \phi_{p_\phi}^3}(\dot{\phi})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + O_P(T^{-1}),
\end{aligned}$$

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

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

Hence,

$$\begin{aligned}
m_T(\hat{\phi}) \quad \equiv \quad & Z_{0T} + D(\hat{\phi}^1 - \phi_0^1) + Z_{1T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \frac{1}{2}G(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
& + G_{1p_\phi}(\hat{\phi}^1 - \phi_0^1)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \frac{1}{6}L(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + o_P(T^{-3/4}).
\end{aligned} \tag{35}$$

The first order condition for the $\hat{\phi}$ in the direction of ϕ^1 is:

$$0 = \frac{\partial m_T'}{\partial \phi^1}(\hat{\phi})W_Tm_T(\hat{\phi}) = \Big(D + G_{1p_\phi}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})\Big)' Wm_T(\hat{\phi}) + o_P(T^{-3/4}) \tag{36}$$

Plugging (35) into (36) and solving this in $(\hat{\phi}^1 - \phi_0^1)$ from the linear term and plugging back the outcome into the quadratic terms, we obtain:

$$\begin{aligned}
\hat{\phi}^1 - \phi_0^1 &= H\Big(Z_{0T} + (Z_{1T} + G_{1p_\phi} H Z_{0T})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{2}G(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
&\quad + \left(\tfrac{1}{2}G_{1p_\phi} HG + \tfrac{1}{6}L\right)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3\Big) \\
&\quad + H_1\Big((Z_{0T} + DHZ_{0T})(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{2}(DHG + G)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3\Big) \\
&\quad + o_P(T^{-3/4}) \\
&= H\Big(Z_{0T} + \tfrac{1}{2}G(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2\Big) \\
&\quad + \left(HZ_{1T} + HG_{1p_\phi} HZ_{0T} + H_1 Z_{0T} + H_1 DHZ_{0T}\right)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) \\
&\quad + \tfrac{1}{2}\left(H(G_{1p_\phi} HG + \tfrac{L}{3}) + H_1(DHG + G)\right)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + o_P(T^{-3/4}).
\end{aligned}$$

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

$$(\hat{\phi}^1 - \phi_0^1) = A_1 + B_1(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + C_1(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + o_P(T^{-3/4}) \tag{37}$$

Using (35), we have:

$$\begin{aligned}
& m'_T(\hat{\phi})W_T m_T(\hat{\phi}) = m'_T(\hat{\phi})W m_T(\hat{\phi}) + o_P(T^{-5/4}) \\
= \;& Z'_{0T}WZ_{0T} + (\hat{\phi}^1 - \phi_0^1)'D'WD(\hat{\phi}^1 - \phi_0^1) + \tfrac{1}{4}G'WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 \\
& + 2Z'_{0T}WD(\hat{\phi}^1 - \phi_0^1) + 2Z'_{0T}WZ_{1T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + Z'_{0T}WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
& + 2Z'_{0T}WG_{1p_\phi}(\hat{\phi}^1 - \phi_0^1)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{3}Z'_{0T}WL(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 \\
& + 2(\hat{\phi}^1 - \phi_0^1)'D'WZ_{1T}(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + (\hat{\phi}^1 - \phi_0^1)'D'WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \\
& + 2(\hat{\phi}^1 - \phi_0^1)'D'WG_{1p_\phi}(\hat{\phi}^1 - \phi_0^1)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}) + \tfrac{1}{3}(\hat{\phi}^1 - \phi_0^1)'D'WL(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 \\
& + Z'_{1T}WG(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 + G'WG_{1p_\phi}(\hat{\phi}^1 - \phi_0^1)(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^3 \\
& + \tfrac{1}{6}G'WL(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^5 + o_P(T^{-5/4}).
\end{aligned} \tag{38}$$

Replacing $\hat{\phi}^1 - \phi_0^1$ by its expression from (37) into (38), 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,

$$K_T(\phi_{p_\phi}) = Z'_{0T}W^{1/2}M_dW^{1/2}Z_{0T} + Z'_{0T}W^{1/2}M_dW^{1/2}G(\phi_{p_\phi} - \phi_{0,p_\phi})^2 + \tfrac{1}{4}G'W^{1/2}M_dW^{1/2}G(\phi_{p_\phi} - \phi_{0,p_\phi})^4. \quad (39)$$

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:

$$\begin{aligned}
&R_T = (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})\times \\
&\Big\{2A'_1D'WDB_1 + 2Z'_{0T}WDB_1 + 2Z'_{0T}WZ_{1T} \\
&+2Z'_{0T}WG_{1p_\phi}A_1 + 2A'_1D'WZ_{1T} + 2A'_1D'WG_{1p_\phi}A_1 \\
&+(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2\Big(2A'_1D'WDC_1 + 2Z'_{0T}WDC_1 + \tfrac{1}{3}Z'_{0T}WL + B'_1D'WG \\
&+\tfrac{1}{3}A'_1D'WL + Z'_{1T}WG + G'WG_{1p_\phi}A_1\Big) \\
&+(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4\left(C'_1D'WG + \tfrac{1}{6}G'WL\right)\Big\}
\end{aligned}$$

$$\begin{aligned}
&R_T = (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})\times \\
&\Big\{2Z'_{0T}H'D'WDB_1 + 2Z'_{0T}WDB_1 + 2Z'_{0T}WZ_{1T} + 2Z'_{0T}WG_{1p_\phi}HZ_{0T} \\
&+2Z'_{0T}H'D'WZ_{1T} + 2Z'_{0T}H'D'WG_{1p_\phi}HZ_{0T} \\
&+(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2\quad\Big(2Z'_{0T}H'D'WDC_1 + 2Z'_{0T}WDC_1 + \tfrac{1}{3}Z'_{0T}WL + B'_1D'WG \\
&+\tfrac{1}{3}Z'_{0T}H'D'WL + Z'_{1T}WG + G'WG_{1p_\phi}HZ_{0T} + G'H'D'WDB_1 \\
&+G'H'D'WZ_{1T} + Z'_{0T}WG_{1p_\phi}HG + Z'_{0T}H'D'WG_{1p_\phi}HG + G'H'D'WG_{1p_\phi}HZ_{0T}\Big) \\
&+(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4\quad\Big(C'_1D'WG + \tfrac{1}{6}G'WL + \tfrac{1}{2}G'H'D'WG_{1p_\phi}HG + G'H'D'WDC_1 \\
&+\tfrac{1}{6}G'H'D'WL + \tfrac{1}{2}G'WG_{1p_\phi}HG\Big)\Big\} \equiv (\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})\times 2R_{1T}.
\end{aligned}$$

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

$$
\begin{aligned}
2R_{1T} = \quad & 2Z_{0T}' W^{1/2} M_d W^{1/2} Z_{1T} + 2Z_{0T}' W^{1/2} M_d W^{1/2} G_{1p_\phi} H Z_{0T} \\
& +(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^2 \Big(\tfrac{1}{3} Z_{0T}' W^{1/2} M_d W^{1/2} L + Z_{1T}' W^{1/2} M_d W^{1/2} G \\
& + G' W^{1/2} M_d W^{1/2} G_{1p_\phi} H Z_{0T} + Z_{0T}' W^{1/2} M_d W^{1/2} G_{1p_\phi} H G \Big) \\
& +(\hat{\phi}_{p_\phi} - \phi_{0,p_\phi})^4 \Big(\tfrac{1}{6} G' W^{1/2} M_d W^{1/2} L + \tfrac{1}{2} G' W^{1/2} M_d W^{1/2} G_{1p_\phi} H G \Big).
\end{aligned} \tag{40}
$$

We can check that the GMM estimator $\hat{\phi}_{p_\phi}$ as given by the first order condition (34) 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

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

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 $m_T'(\hat{\phi}) W_T m_T(\hat{\phi})$. 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 (40) and scaling by T, we can see, using the continuous mapping theorem, that TR_{1T} converges in distribution towards $\mathbb{R}_1$:

$$
\begin{aligned}
\mathbb{R}_1 \quad = \quad & \mathbb{Z}_0' W^{1/2} M_{dg} W^{1/2} (\mathbb{Z}_1 + G_{1p_\phi} H \mathbb{Z}_0) \\
& + \Big(\mathbb{Z}_0' W^{1/2} (M_d - M_{dg}) W^{1/2} \mathbb{Z}_0 G' - G' W^{1/2} M_d W^{1/2} \mathbb{Z}_0 \mathbb{Z}_0' \Big) \\
& \times W^{1/2} M_d W^{1/2} \Big(\tfrac{1}{3} L + G_{1p_\phi} H G \Big) / \sigma_G,
\end{aligned} \tag{41}
$$

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

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 $\Big(\sqrt{T}(\hat{\phi}^1 - \phi_0^1),\ T^{1/4} |\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}|,\ (-1)^{B_T} \Big) = O_P(1)$, any subsequence of the left hand side has a further subsequence that converges in distribution. Using (b), such subsequence satisfies:

$$
\Big(\sqrt{T}(\hat{\phi}^1 - \phi_0^1),\ T^{1/4} |\hat{\phi}_{p_\phi} - \phi_{0,p_\phi}|,\ (-1)^{B_T} \Big) \xrightarrow{d} \Big(H\mathbb{Z}_0 + HG\mathbb{V}/2, \sqrt{\mathbb{V}}, (-1)^{\mathbb{B}} \Big).
$$

(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 $(\lambda_1, \lambda_2) \in \Re \times \Re$,

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

Let $x \in \Re$ 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 $T \to \infty$,

$$\begin{aligned} &P(\lambda_2 Y_T \leq x - \lambda_1, X_T < 0) \to P(\lambda_2 Y \leq x - \lambda_1, X < 0) \qquad and \\ &P(\lambda_2 Y_T \leq x + \lambda_1, X_T \geq 0) \to P(\lambda_2 Y \leq x + \lambda_1, X \geq 0) \end{aligned} \tag{42}$$

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 (42). 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 Theorem 2: We have:

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

From (24), we have

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

with $z_T=B_T\left((\hat{\theta}_{II}-\theta_0)'\Delta_{IT,k}(\dot{\theta}_T)(\hat{\theta}_{II}-\theta_0)\right)_{1\le k\le \ell}$. 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 (22), 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 $1\le k\le \ell-1$,

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

and

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

Thus,

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

Since the last column of $\hat{S}_T^1\hat{F}_T$ is nil, we can write:

$$\begin{aligned}&\sqrt{T}\hat{S}_T^1\left(\hat{\theta}_{II}-\theta_0\right)\\ =\;&\hat{S}_T^1\hat{F}_T\left(B_T m_{IT}(\theta_0)-\tfrac{1}{2}\left(F'_{\bullet\ell}\Delta_{I,k}(\theta_0)F_{\bullet\ell}\right)_{1\le k\le \ell}\left(T^{1/4}m_{IT,\ell}(\theta_0)\right)^2\right)+o_P(1).\end{aligned} \tag{43}$$

Using again (22), we have

$$T^{1/4}\hat{S}_{T,p}(\hat{\theta}_{II}-\theta_0)=\hat{S}_{T,p}F_{\bullet\ell}T^{1/4}m_{IT,\ell}(\theta_0)+o_P(1). \tag{44}$$

By the continuous mapping theorem, $\hat{S}_T^1\hat{F}_T$ converges in probability towards S^1F with nil last column and $\hat{S}_{T,p}$ converges in probability towards $S_{p\bullet}$. Since $B_T m_{IT}(\theta_0)$ converges in distribution towards $\mathbb{Y}$, we can deduce from (43) and (44) that:

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

□

Proof of Equation (19)**:** Since $\hat{\phi} - \phi_0 = R(\hat{\eta} - \eta_0)$, we have

$$T^{1/4}(\hat{\phi} - \phi_0) = R_{\bullet p_\phi} T^{1/4}(\hat{\eta}_{p_\phi} - \eta_{0,p_\phi}). \tag{45}$$

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

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

where the last equality uses (45) 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. □