

Koch, Matthias

**Conference Paper**

## The Estimation of Spatial Autoregressive Models with Missing data of the Dependent Variables

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Koch, Matthias (2010) : The Estimation of Spatial Autoregressive Models with Missing data of the Dependent Variables, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118803>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The Estimation of Spatial Autoregressive Models with Missing data of the Dependent Variables

Matthias Koch  
Matthias.Koch@wu.ac.at

## Abstract

This paper focuses on several estimation methods for spatial autoregressive (SAR)- models in case of missing observations in the dependent variable. First, we show with an example and then in general, how missing observations can change the model and thus resulting in the failure of the "available" estimation methods. To estimate the SAR- model with missings we propose different estimation methods, such as GMM, NLS and OLS. Some of the estimators are based on a model approximation. A Monte Carlo Simulation is conducted to compare the different estimation methods in their diverse numerical and sample size aspects.

## 1 Introduction

...Finance, market values of sold properties to unsold properties. (To turn a blind eye on this problem will result in biased estimators)...

...Our focus is on spatially dependent models where we treat unobserved market transactions as missing data in the dependent variable. Le Sage 2004 provided a framework for this problem via Maximum Likelihood Estimations. The main problem with Maximum Likelihood Estimations is that one must assume the correct Distribution (Outliers are different handled weather you use a t- distribution or a Normal distribution)...

...Another advantage is that one can derive unbiased estimators which are less computational expensive than estimators based on maximum likelihood....

## 2 Spatial Dependence and Missing data

In this section we will focus how on the one hand missing data effect a spatial autoregressive (SAR) model and on the other how one can derive

estimators without assuming the specific distribution of the error term. In the latter we will examine different details of the estimators...

## 2.1 Subsection

First we will start with an example to illustrate the effect of missing data of the dependent variable on the estimation of a spatial autoregressive process. We will assume that the spatial dependence is represented by an one forward one behind neighboring pattern. That means we will have an almost complete sparse matrix  $\mathbf{W}$  except that we will have ones in the upper and lower diagonal. Therefore our data generating process is represented by (1).

$$\mathbf{Y} = \rho_0 \mathbf{W} \mathbf{Y} + \mathbf{X} \beta_0 + \boldsymbol{\varepsilon} \text{ where } \varepsilon_i \sim i.i.d.(0, \sigma_0^2) \quad (1)$$

where  $\mathbf{X}$  is a  $n \times k$  dimensional matrix of exogenous variables. Assume now that the third entry of  $\mathbf{Y}$  is not observed.  $\bar{\mathbf{Y}}$  will in general denote the observed data, or in that special case:  $\bar{\mathbf{Y}} = (y_1, y_2, y_4, \dots, y_n)'$ . In the appendix it is shown that the data generating process will now be represented by (2)<sup>1</sup>

$$\bar{\mathbf{Y}} = \rho_0 \mathbf{W}_1 \bar{\mathbf{Y}} + \mathbf{X}_1 \beta_0 + \frac{\mathbf{X}_3 \beta_0}{1 - \rho_0^2} + \rho_0 \frac{\mathbf{W}_2 \bar{\mathbf{Y}} + \mathbf{X}_2 \beta_0}{1 - \rho_0^2} + \rho_0^2 \frac{\mathbf{W}_3 \bar{\mathbf{Y}}}{1 - \rho_0^2} + \bar{\boldsymbol{\varepsilon}} \quad (2)$$

Observe that we no longer have a linear Model and that the  $\bar{\varepsilon}_i$  are no longer independent and identically distributed. It is intuitively clear that by simply ignoring missing data in the dependent variable is like substituting  $\bar{\mathbf{Y}} = \rho_0 \bar{\mathbf{W}} \bar{\mathbf{Y}} + \bar{\mathbf{X}} \beta_0 + \bar{\boldsymbol{\varepsilon}}$  for (2), where  $\bar{\mathbf{X}} = (\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_4, \dots, \mathbf{x}'_n)'$ ,  $\bar{\varepsilon}_i \sim i.i.d.(0, \sigma_0^2)$  and  $\bar{\mathbf{W}} = \begin{pmatrix} \mathbf{W}_{2,2} & \mathbf{0}_{2,n-3} \\ \mathbf{0}_{n-3,2} & \mathbf{W}_{n-3,n-3} \end{pmatrix}$ . Therefore, if  $\rho_0 \neq 0$  one is not surprised that the ignoring of missing dependent variables causes a biased estimation, since one is no longer estimating the true data generating process.

If the  $\mathbf{W}$ -matrix is sparse like in this case, one might use the following approach: Classify all the  $y_i$  of the data generating process into one of the following three sets: "{Missing}", "{Border}" and "{Inside}". The "{Missing}"-set is selfexplaining. The "{Border}"- set contains all the observed  $y_i$  that have a missing observation as neighbour and the "{Inside}" set contains the remaining  $y_i$  that are not elements of the sets Missing or Border. Let  $\bar{\mathbf{Y}}_I := y_i \in \{\text{Inside}\}$  and  $\bar{\mathbf{Y}}_B := y_i \in \{\text{Border}\}$  then one can rewrite the data generating process for  $\bar{\mathbf{Y}}_I$  with

<sup>1</sup>For the definition of  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$  and  $\bar{\boldsymbol{\varepsilon}}$  see the Appendix

the following equation:  $\bar{\mathbf{Y}}_I = \rho_0 (\mathbf{W}_{I,I} \mathbf{W}_{I,B}) \begin{pmatrix} \bar{\mathbf{Y}}_I \\ \bar{\mathbf{Y}}_B \end{pmatrix} + \mathbf{X}_I \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_I$ , where  $\mathbf{W}_{I,I}$ ,  $\mathbf{W}_{I,B}$  represent the neighbouring pattern of the set  $\{\text{Inside}\}$  and between  $\{\text{Inside}\}$  and  $\{\text{Border}\}$ . The  $\mathbf{X}_I$  and  $\boldsymbol{\varepsilon}_I$  contain the  $\mathbf{x}_i$  and  $\varepsilon_i$  of the corresponding  $y_i \in \{\text{Inside}\}$ . This approach is only practical if a small part of the data set is unobserved or the missing data is spatially clustered. The estimations are always based on the sample size  $|\{\text{Inside}\}|$ . In this paper we assume that one observes  $\mathbf{x}_i$  for every  $y_i$  and as result our estimations are based on the sample size  $|\{\text{Inside}\}| + |\{\text{Border}\}|$ .

### 2.1.1 Formalization of missing data

As we have seen before solving the data generating process with missing observations can be complicated it seems feasible to try a more formal approach. First some notation:

Let  $N$  denote the sample size of the unobserved data.  $n$  denotes the sample size of the observed dependent variables. We assume that  $n$  is a function of  $N$  and that  $\lim_{N \rightarrow \infty} n(N) = \infty$  in order to derive asymptotic properties of our estimators. That means the observed data sample approaches infinity if the unobserved data sample approaches infinity. Furthermore we assume that our true data generating process is sorted in a way that the first  $n$  observations are observed and the other  $N - n$  represent the unobserved.

We now can define the observation matrix  $\mathbf{S}_n$  which extracts the observed  $y$  from the vector  $\mathbf{Y}$ :  $\mathbf{S}_n := (\mathbf{I}_{n \times n} \mathbf{0}_{n \times (N-n)})$ . This definition yields that  $\bar{\mathbf{Y}}_n = \mathbf{S}_n \mathbf{Y}_N$ . Therefore, the true data generating process for  $\bar{\mathbf{Y}}$  is:

$$\bar{\mathbf{Y}}_n = \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \mathbf{X}_N \boldsymbol{\beta}_0 + \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon}_N \quad (3)$$

where  $\varepsilon_i \sim i.i.d.(0, \sigma_0^2)$ . One can see that (3) is for  $y_i$  a nonlinear function in  $\rho_0$  and therefore one can no longer use directly linear estimation methods to find estimators for  $\theta_0 := (\boldsymbol{\beta}'_0, \rho_0)'$ . The next section derives estimators based on the generalized method of moments, nonlinear least squares and ordinary least squares. Of course it is possible to estimate (3) with Maximum Likelihood if one for example assumes that the  $\varepsilon_i$  are independent normal distributed<sup>2</sup>. One has to point out, that the only way to estimate (3) for huge data sets ( $N > 5000$ ) is to

---

<sup>2</sup>One simple method is to maximize the log likelihood:  $\max_{\boldsymbol{\beta}, \rho, \sigma^2} \ln(\mathbf{L}(\boldsymbol{\beta}, \rho, \sigma^2))$  where  $\mathbf{L}(\boldsymbol{\beta}, \rho, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\boldsymbol{\Omega}_{ML}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\bar{\mathbf{Y}} - \boldsymbol{\Upsilon})' \boldsymbol{\Omega}_{ML}^{-1} (\bar{\mathbf{Y}} - \boldsymbol{\Upsilon})\right)$ ,  $\boldsymbol{\Omega}_{ML} = \mathbf{S}(\mathbf{I} - \rho \mathbf{W})^{-1} (\mathbf{I} - \rho \mathbf{W}')^{-1} \mathbf{S}' \sigma^2$  and  $\boldsymbol{\Upsilon} = \mathbf{S}(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta}$

This likelihood has some numerical difficulties. To programm the Likelihood Method efficiently see LeSage 2004

use approximations for  $(\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1}$ . We will use a finite Neumann series to find approximations for  $(\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1}$ .

## 2.2 Estimation Methods

### 2.2.1 Model Assumptions<sup>3</sup>

1. Let  $\theta_0 = (\beta_0' \rho_0)'$   $\in \Theta$ , where  $\Theta$  is a compact contiguous subspace of  $\mathbb{R}^k \times (-1, 1)$ . We call  $\Theta$  the parameter space of our SAR- model.
2. For every  $N : \|\mathbf{W}_N\|_1 = 1$ , where  $\|\mathbf{W}_N\|_1 = \max_{j \in \{1, \dots, N\}} \left\{ \sum_{i=1}^N |w_{j,i}| \right\}$  and  $w_{j,i} \in \mathbf{W}_N$ . (The weigh matrix  $\mathbf{W}_N$  is maximum absolute row sum normalized.<sup>4</sup>)
3. Let  $E[\mathbf{X}'_N \boldsymbol{\epsilon}_N] = 0$ . The sequence  $\mathbf{x}_i \in \mathbb{R}^{1 \times k}$  is an identically and independently distributed random vector with finite mean.
4.  $rank(\mathbf{X}) = k$  and for every  $i \in \{1, \dots, \max(m(n), z)\} : rank(\mathbf{W}_N^i \mathbf{X}_N) = k$

### 2.2.2 Model Approximation

To derive different estimators for (3), namely two different GMM- Estimators (GMM-e and GMM-a), two NLS estimators (NLS-e and NLS-a) and one estimator based on OLS we use for the GMM-a, NLS-a and OLS estimators a model approximation. Like noted before, if one wants to estimate (3) for huge data sets at some point an approximation for  $(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}$  will be needed. In that light we suggest to use the Neumannseries not in the estimator itself but use an approximation for the data generating process like (4).

$$\bar{\mathbf{Y}}_n \approx \tilde{\mathbf{Y}}_n = \mathbf{S}_n \sum_{k=0}^{m(n)} \rho_0^k \mathbf{W}_N^k \mathbf{X}_N \beta_0 + \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \boldsymbol{\epsilon}_N \quad (4)$$

where  $m(n)$  is  $\mathbb{N} \rightarrow \mathbb{N} : n \mapsto m(n)$  and  $\lim_{n \rightarrow \infty} m(n) = \infty$ . The property of  $m(n)$  that  $\lim_{n \rightarrow \infty} m(n) = \infty$  is necessary to derive the asymptotic

<sup>3</sup>For more detatiled Assumptions of the data generating process see Appendix

<sup>4</sup>The main idea is to have a prameterspace where  $(\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1}$  is bounded in row and coulounm sums... see Prucha 200X page X.

For the simplicity of the proofs we also assume that for every  $N : \|\mathbf{W}_N\|_\infty = 1$ . This may seem strong but all proofs will also work as long as  $\|\mathbf{W}_N\|_\infty < \infty$ .

distribution of the different estimators. In other words we need  $m(n)$  to grow as the sample size grows. One can show that the approximation has an maximum error  $\delta(\rho_0)$  for every  $y_i$  of:  $|\delta(\rho_0)| \leq kx_m\beta_m \frac{|\rho_0|^{m(n)+1}}{1-|\rho_0|}$  where  $\delta := \max_{1 \leq i \leq n} \{|y_i - \tilde{y}_i|\}$ ,  $x_m = \max_{1 \leq i \leq n, 1 \leq j \leq k} (|x_{i,j}|)$ ,  $\beta_m = \max_{1 \leq j \leq k} (|(\beta_0)_j|)$ . Since  $|\rho_0| < 1$  is assumed, the model error will always decrease exponentially with  $m(n)$  since,  $\delta(\rho_0) \leq c \exp(\ln(|\rho_0|)(m(n) + 1))$  where  $c = \frac{kx_m\beta_m}{1-|\rho_0|} \in \mathbb{R}$  and  $\ln(|\rho_0|) < 0$ . Therefore, an adequate high  $m(n)$  leads to negligible small errors  $\delta(\rho_0)$ .

In practice we define a numerical maximum relative error (condition,  $eps = 10^{-2}$  in our MC- simulation) for the model approximation like it is common in numerical mathematics. Then we guess  $\rho_m$  as a upper limit for  $\rho_0$ . For example we say that  $\rho_0 < \rho_m = 0.5$ . With  $\rho_m$  and  $eps$  it is possible to find an  $m(n)$  that will lead to the postulated model accuracy. The model approximation leads to an estimator  $\hat{\rho}$  and if  $\hat{\rho}$  is smaller than  $\rho_m$  our guess for  $\rho_m$  was good, otherwise we have to estimate the model again and take a higher value for  $\rho_m$  like  $\rho_m = .75$ .

### 2.2.3 GMM- Estimation

Now we derive two estimators which are based on the GMM- method. One Estimator (GMM-e) will use the true data generating process (3) and one that will use the approximation stated in (4) (GMM-a). Both estimators are based on the following moment condition  $E[\mathbf{g}] = \mathbf{0}$ , where

$$\mathbf{g} = (g'_0, g'_1, \dots, g'_z)' \text{ where } g_i = \frac{1}{n} (\mathbf{S}_n \mathbf{W}_N^i \mathbf{X}_N)' \overline{\boldsymbol{\varepsilon}}_n, z \geq 1, z < m(n) \quad (5)$$

and  $\overline{\boldsymbol{\varepsilon}}_n = \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon}$ . The GMM- condition fulfills  $E[g_i] = 0$  for every  $1 \leq i \leq z$ . Since  $z \geq 1$  we have to estimate  $(k+1)$  parameters with  $(z+1)k$  linear independent moment conditions. Therefore, our estimator is overidentified. This means that we need a 2-step procedure. In the first step we minimize  $\mathbf{g}'\mathbf{g}$  to get an estimator for  $\theta_0$ . Regardless which model we are using (the true or approximate), we have to minimize  $\mathbf{g}'\mathbf{A}\mathbf{g}$  in a second step, where  $\mathbf{A}$  is a positive definite matrix. In order to get an asymptotic efficient estimators for  $\theta_0$  one might use  $\mathbf{A} = \boldsymbol{\Omega}_{0,gmm}^{-1}$ , since  $\boldsymbol{\Omega}_{0,gmm} = Var(\mathbf{g})$  for the true model.

$$\boldsymbol{\Omega}_{0,gmm} = \frac{\sigma_0^2}{n^2} \begin{pmatrix} \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{X}}_n & \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{W}} \overline{\mathbf{X}}_n & \dots & \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{W}}^z \overline{\mathbf{X}}_n \\ \overline{\mathbf{W}} \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{X}}_n & \overline{\mathbf{W}} \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{W}} \overline{\mathbf{X}}_n & \dots & \overline{\mathbf{W}} \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{W}}^z \overline{\mathbf{X}}_n \\ \vdots & & & \vdots \\ \overline{\mathbf{W}}^z \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{X}}_n & \dots & \dots & \overline{\mathbf{W}}^z \overline{\mathbf{X}}_n' \boldsymbol{\Sigma}_{0,n} \overline{\mathbf{W}}^z \overline{\mathbf{X}}_n \end{pmatrix} \quad (6)$$

where  $\Sigma_{0,n} = \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} (\mathbf{I}_N - \rho_0 \mathbf{W}'_N)^{-1} \mathbf{S}'_n$ ,  $\overline{\mathbf{W}^i \mathbf{X}_n} = \mathbf{S}_n \mathbf{W}_N^i \mathbf{X}_N$  and  $0 \leq i \leq z$ . As usual when using GMM- methods one will use the estimator from step1 to calculate  $\mathbf{\Omega}_{0,gmm}$ .

There are two reasons why one can use  $\mathbf{\Omega}_{0,gmm}^{-1}$  as weighting matrix for the moment conditions of the GMM-a estimator. First, since the efficiency proof for GMM-estimators only holds asymptotically it doesn't matter whether the exact or the approximate model was used due to the assumption that both models are asymptotically equivalent. Second, we noted that the we will use a  $m(n)$  so high, that the approximation error is negligible small. Therefore, the difference between  $Var(\mathbf{g})$  for the approximate and the true model should be negligible small.

**GMM- exact- estimator** The Appendix shows that if one uses for  $\overline{\boldsymbol{\varepsilon}}_n = \overline{\boldsymbol{\varepsilon}}_n(\beta, \rho) = \overline{\mathbf{Y}}_n - \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \beta$ <sup>5</sup> which is based on the true model stated in (3), the first minimization step yields a consistent estimator for  $\theta_0$ .

**step 1:**

$$\hat{\theta}_1 = \arg \min_{\theta \in \Theta} \mathbf{g}(\theta)'_1 \mathbf{g}(\theta)_1 \quad (7)$$

**step 2:** Now one uses the estimator  $\hat{\theta}_1$  from (7) to calculate  $\widehat{\mathbf{\Omega}}_{gmm}$  and minimize (8)

$$\hat{\theta}_2 = \arg \min_{\theta \in \Theta} \mathbf{g}(\theta)'_1 \widehat{\mathbf{\Omega}}_{gmm}^{-1} \mathbf{g}(\theta)_1 \quad (8)$$

The appendix shows that the estimators obtained by (8) have the following asymptotic distribution:

$$n^{-1/2} \left( \hat{\theta}_2 - \theta_0 \right) \sim N \left( 0, (\mathbf{G}'_0 \mathbf{\Omega}_{0,gmm} \mathbf{G}_0)^{-1} \right) \text{ where } \mathbf{G}_0 = \frac{\partial \mathbf{g}_1}{\partial \theta'} \Big|_{\theta=\theta_0} \quad (9)$$

$$\text{with } \frac{\partial g_{1,i}}{\partial \theta'} \Big|_{\theta=\theta_0} = \frac{-1}{n} \left( \begin{array}{c} \overline{\mathbf{W}^i \mathbf{X}'_n} \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \mathbf{X}_N \\ \overline{\mathbf{W}^i \mathbf{X}'_n} \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-2} \mathbf{W}_N \mathbf{X}_N \beta_0 \end{array} \right).$$

---

<sup>5</sup>If one would use instead of  $\overline{\boldsymbol{\varepsilon}}_n(\beta, \rho) = \overline{\mathbf{Y}}_n - \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \beta$  the smaller "Inner"-  $\overline{\boldsymbol{\varepsilon}}_n(\beta, \rho)$ - vector  $\boldsymbol{\varepsilon}_I(\beta, \rho) = \overline{\mathbf{Y}}_I - \rho (\mathbf{W}_{I,I} \mathbf{W}_{I,B}) \begin{pmatrix} \overline{\mathbf{Y}}_I \\ \overline{\mathbf{Y}}_B \end{pmatrix} - \mathbf{X}_I \beta$ , then one could solve the minimization problem analytically and would derive an instrumental variable estimator.

Futhermore if one assumes that  $\mathbf{S}_n = \mathbf{I}_n$  (the case of no missings) then one can write instead of the highly nonlinear  $\overline{\boldsymbol{\varepsilon}}_n(\beta, \rho) = \overline{\mathbf{Y}}_n - (\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} \mathbf{X}_n \beta$  a linear equivalent  $\overline{\boldsymbol{\varepsilon}}_n(\beta, \rho) = \mathbf{Y}_n - \rho \mathbf{W}_n \mathbf{Y}_n - \mathbf{X}_n \beta$  solve the GMM- minimization problem analytically and get the instrumental variable estimator proposed by Kelejian, Prucha [1998].

**GMM- approx- estimator:** The Appendix shows that if one uses for  $\overline{\varepsilon}_n(\beta, \rho) = \overline{\mathbf{Y}}_n - \mathbf{S}_n \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^k \mathbf{X}_N \beta$  the approximation model proposed in (4) the first minimization step yields an consistent estimator for  $\theta$ .

**step 1:**

$$\hat{\theta}_1 = \arg \min_{\theta \in \Theta} \mathbf{g}(\theta)'_2 \mathbf{g}(\theta)_2 \quad (10)$$

Now one uses the estimator  $\hat{\theta}_1$  from (10) to estimate  $\widehat{\Omega}_{gmm}$  and minimize (11)

$$\hat{\theta}_2 = \arg \min_{\theta \in \Theta} \mathbf{g}'_2 \widehat{\Omega}_{gmm}^{-1} \mathbf{g}_2 \quad (11)$$

The appendix shows that the estimator obtained by (11) has the following asymptotic distribution:

$$n^{-1/2} (\hat{\theta}_2 - \theta_0) \sim N \left( 0, (\mathbf{G}'_0 \mathbf{\Omega}_{0,gmm} \mathbf{G}_0)^{-1} \right) \text{ where } \mathbf{G}_0 = \frac{\partial \mathbf{g}_2}{\partial \theta'} \Big|_{\theta=\theta_0} \quad (12)$$

$$\frac{\partial g_{1,i}}{\partial \theta'} \Big|_{\theta=\theta_0} = \frac{-1}{n} \overline{\mathbf{W}^i \mathbf{X}_n}' \mathbf{S}_n \left( \left( \sum_{j=0}^{m(n)} \rho_0^j \mathbf{W}_N^j \mathbf{X}_N \right)', \left( \sum_{j=1}^{m(n)} j \rho_0^{j-1} \mathbf{W}_N^j \mathbf{X}_N \beta_0 \right)' \right)'$$

One should note that as  $N$  approaches infinity  $\frac{\partial \mathbf{g}_2}{\partial \theta'} = \frac{\partial \mathbf{g}_1}{\partial \theta'}$  and therefore both estimators have the same asymptotic distribution. Additionally observe that the asymptotic distribution of the approximate estimator (GMM-a) is essentially the same as the asymptotic distribution of the exact estimator (GMM-e) if one uses in the exact estimator the finite Neumannseries instead of  $(\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1}$ .

### 2.3 NLS- estimation

An other possibility to derive an estimator for the models (3) and (4)

is to use the nonlinear least squares method. This method, like the GMM- Estimator needs no assumption about the actual distribution of the error term. Similar to the GMM- Estimator we will derive two different estimators, one based on (3) and one on (4). Like in the GMM case we call them NLS-e and NLS-a. Both estimators are based on the following minimization problem:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \overline{\varepsilon}_n' \overline{\varepsilon}_n \text{ where } \overline{\varepsilon}_n \text{ depends on the chosen model} \quad (13)$$



### 2.3.1 NLS- exact- estimator

The appendix shows that the minimization of (13) where  $\overline{\varepsilon}_n(\beta, \rho) = \overline{\mathbf{Y}}_n - \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \beta$  yields a consistent estimator for  $\theta_0 = (\beta_0' \rho_0)$ . It is also shown that this estimator has the following asymptotic distribution:

$$n^{-1/2} (\widehat{\theta} - \theta_0) \xrightarrow{D} N(0, \mathbf{A}_0^{-1} \mathbf{B}_0 \mathbf{A}_0^{-1}) \quad (14)$$

with  $\mathbf{B}_0 = \mathbf{D}_0 \mathbf{\Lambda}_0 \mathbf{D}_0'$ ,  $\mathbf{\Lambda}_0 = \sigma_0^2 \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{G}_N'(\rho_0) \mathbf{S}_n'$ ,

$$\mathbf{D}_0 = \frac{-1}{n} \begin{pmatrix} \mathbf{X}_N' \mathbf{G}_N(\rho_0)' \mathbf{S}_n' \\ \beta_0' \mathbf{X}_N' \overline{\mathbf{G}}_N(\rho_0)' \mathbf{S}_n' \end{pmatrix}, \quad \mathbf{A}_0 = n \mathbf{D}_0 \mathbf{D}_0'$$

where  $\overline{\mathbf{G}}_N(\rho) = \mathbf{G}_N(\rho) \mathbf{W}_N \mathbf{G}_N(\rho)$  and  $\mathbf{G}_N(\rho) = (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}$ . In practice one will use the estimators from (13) for  $\rho_0$  and  $\beta_0$ .

### 2.3.2 NLS- approximate- estimator

The appendix shows that the minimization of (13) where  $\overline{\varepsilon}_n(\beta, \rho) =$

$\overline{\mathbf{Y}}_n - \mathbf{S}_n \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^k \mathbf{X}_N \beta$  yields a consistent estimator. It is also shown, that the NLS- approximate estimator has the following asymptotic distribution:

$$n^{-1/2} (\widehat{\theta} - \theta_0) \xrightarrow{D} N(0, \mathbf{A}_0^{-1} \mathbf{B}_0 \mathbf{A}_0^{-1}) \quad (15)$$

with  $\mathbf{B}_0 = \mathbf{D}_0 \mathbf{\Lambda}_0 \mathbf{D}_0'$ ,  $\mathbf{\Lambda}_0 = \sigma_0^2 \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{G}_N'(\rho_0) \mathbf{S}_n'$ ,

$$\mathbf{D}_0 = \frac{-1}{n} \begin{pmatrix} \mathbf{X}_N' \sum_{k=0}^{m(n)} \rho_0^k \mathbf{W}_N^k \mathbf{S}_n' \\ \beta_0' \mathbf{X}_N' \sum_{k=1}^{m(n)} k \rho_0^{k-1} \mathbf{W}_N^k \mathbf{S}_n' \end{pmatrix}, \quad \mathbf{A}_0 = n \mathbf{D}_0 \mathbf{D}_0'$$

It is also possible to interpret (15) as the numerical approximation of (14).

## 2.4 OLS- estimation

The OLS- Estimator is based on the approximation model (4). But instead of making an nonlinear approach the model is linearized in the following way:

$$\bar{\mathbf{Y}}_n = \mathbf{Z}_n \eta_0 + \boldsymbol{\nu}_n \quad (16)$$

where  $\mathbf{Z}_n = \mathbf{S}_n[\mathbf{W}_N^0 \mathbf{X}_N, \mathbf{W}_N^1 \mathbf{X}_N, \mathbf{W}_N^2 \mathbf{X}_N, \dots, \mathbf{W}_N^{m(n)} \mathbf{X}_N]$ ,  $\eta_0 = (\beta_0', \rho_0 \beta_0', \dots, \rho_0^{m(n)} \beta_0')'$  and  $\boldsymbol{\nu}_n = \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon}$ . The estimator is based on the GMM- condition  $\mathbf{g}_{OLS,N} = n(N)^{-1} \mathbf{Z}'_n \boldsymbol{\nu}_n$ . Thus resulting in the case of an exact identified estimator. The appendix shows that  $\hat{\eta} = (\mathbf{Z}'_n \mathbf{Z}_n)^{-1} \mathbf{Z}'_n \bar{\mathbf{Y}}_n$  is a consistent estimator for  $\eta_0$ . Our next concern is to find a Transformation  $T_1 : \mathbb{R}^{(m(n)+1)k} \rightarrow \mathbb{R}$  so that:  $E[T_1(\eta_0)] = \rho_0$ . With this Transformation we are able to use the asymptotic distribution of  $n^{-1/2} (\hat{\eta} - \eta_0)$ :

$$n^{-1/2} (\hat{\eta} - \eta_0) \sim N(0, \mathbf{R}_n^{-1} \mathbf{V}_{0,N} \mathbf{R}_n^{-1}) \quad (17)$$

where  $\mathbf{R}_n = \frac{1}{n} \mathbf{Z}'_n \mathbf{Z}_n$  and  $\mathbf{V}_{0,N} = \frac{\sigma_0^2}{n} \mathbf{Z}'_n \mathbf{S}_{n(N)} \mathbf{G}_N(\rho_0) \mathbf{G}'_N(\rho_0) \mathbf{S}'_{n(N)} \mathbf{Z}_n$ . We suggest to consider the following  $(k+1)$  continuous differentiable Transformations  $T_{2,i} : \mathbb{R}^{2(m(n)+1)} \rightarrow \mathbb{R}$  so that  $E[T_{2,i}(\eta_0, c_{i,1}, \dots, c_{i,m(n)+1})] = \theta_{0,i+1}$  where  $\theta_{0,i+1}$  denotes the  $i$ -th element of  $\theta_0$ . In addition we need the  $c_{i,j}$  to be chosen so that the minimization of the variance of the transformation is achieved:

$$c_{i,j} = \arg \min_{c_{i,j}} V[T_{2,i}(\hat{\eta}, c_{i,1}, \dots, c_{i,m(n)+1})] \quad (18)$$

The main issue is that due to the linearization of (4) one has to make an additional transformation to find estimators for  $\theta_0$ . Obviously there are an infinite number of possible transformations to do that. Therefore, we suggest to find one transformation that is at least the most efficient one in a class of transformations.

One possibility for the transformation  $T_1$  would be:

$$T_1(\eta) = \frac{1}{km} \sum_{i=1}^m d^6(\gamma_i) d(\gamma_{i-1})^{-1}$$

where  $\gamma_i = (\eta_{k(i-1)+1}, \eta_{k(i-1)+2}, \dots, \eta_{ki})'$ . One possibility for the transformation  $T_2$

---

<sup>6</sup> $d()$  denotes  $diag()$

$$T_{2,0}(\widehat{\eta}, c_{1,1}, \dots, c_{1,m(n)k}) = \sum_{i=1}^m d(\alpha_i)d(\gamma_i)d(\gamma_{i-1})^{-1}$$

$$T_{2,j}(\widehat{\eta}, c_{j,1}, \dots, c_{j,m(n)k}) = c_{i,1}\eta_j + \sum_{i=2}^m c_{j,i} (\eta_{ik+1})^{1-i} (\eta_{(i-1)k+1})^i$$

for  $j \in \{1, 2, \dots, k\}$  where  $\alpha_i = (c_{1,k(i-1)+1}, c_{1,k(i-1)+2}, \dots, c_{1,ki})'$  and  $\sum_{i=1}^m \sum_{j=1}^k \alpha_{i,j} = 1$ .  $T_{2,0}$  is the transformation for  $\rho$  and  $T_{2,j}$  for  $\beta_j$ . The transformations  $T_2$  have the following Variance:

$$Var(T_{2,k}(\widehat{\eta})) = \left( \frac{\partial T_{2,k}(\eta)}{\partial \eta} \Big|_{\eta=\widehat{\eta}} \right)' n \mathbf{R}_n^{-1} \mathbf{V}_{0,N} \mathbf{R}_n^{-1} \frac{\partial T_{2,k}(\eta)}{\partial \eta} \Big|_{\eta=\widehat{\eta}}$$

$$\frac{\partial T_{2,1}(\gamma)}{\partial \gamma_i} = \begin{cases} -d(\alpha_i)d(\gamma_i)d(\gamma_{i-1})^{-2} & \text{if } i = 1 \\ d(\alpha_i)d(\gamma_{i-1})^{-1} - d(\alpha_{i+1})d(\gamma_{i+1})d(\gamma_i)^{-2} & \text{if } i \in \{2, 3, \dots, m-1\} \\ d(\alpha_i)d(\gamma_{i-1})^{-1} & \text{if } i = m \end{cases}$$

$$\frac{\partial T_{2,j}(\eta)}{\partial \eta} = \left\{ \right.$$

One can see that  $T_{2,0} = T_1$  if  $c_{1,j} = c_{1,1}$  for every  $j$  and  $\sum_{i=1}^{mk} c_{1,i} = 1$ .

### 3 Estimator properties

#### 3.1 Efficiency of the proposed estimators

The proposed GMM, NLS and OLS- estimators are all mean distance estimators. Since the GMM- Estimator is the only one where the moment conditions are weighed in a way to reduce the classical "sandwich" form of the variance matrix of  $n^{1/2} (\widehat{\theta}_N - \theta_0)$ , the used GMM- estimator is the most efficient one in his class of GMM- estimators. That is the main reason why we regard the GMM- estimator as the most preferable method for small to medium size samples.

Obviously some of the NLS and OLS estimators' inefficiency lies in the fact that in (4)  $\overline{\mathbf{Y}}_n$  has the variance  $\sigma_0^2 \mathbf{\Sigma}_0 = \sigma_0^2 \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} (\mathbf{I}_N - \rho_0 \mathbf{W}'_N)^{-1} \mathbf{S}'_n$ . Therefore, it might seem plausible to multiply the model with  $\widehat{\mathbf{\Sigma}}^{-1/2}$ ,

since OLS and NLS produce almost consistent estimators for  $\rho_0$ . If we multiply (3) with  $\Sigma_0^{-1/2}$  we get:

$$\Sigma_0^{-1/2}\bar{\mathbf{Y}}_n = \Sigma_0^{-1/2}\mathbf{S}_n(\mathbf{I}_N - \rho_0\mathbf{W}_N)^{-1}\mathbf{X}_N\beta_0 + \boldsymbol{\varepsilon}_N \quad (19)$$

Green (page 207) shows that  $\Sigma_0^{-1/2}\bar{\mathbf{Y}}_n$ , like  $\boldsymbol{\varepsilon}_N$  has now a variance of

$\sigma_0^2\mathbf{I}_n$ . But the matrix  $\Sigma_0^{-1/2}\mathbf{S}_n(\mathbf{I}_N - \rho_0\mathbf{W}_N)^{-1}$  induces two problems:

First, let us consider two additional assumptions in the theoretical framework: As  $n$  approaches  $\infty$  we will no longer have any missing observations and that our weigh matrix is symmetric. This would cause  $\lim_{n \rightarrow \infty} \Sigma_0^{-1/2}\mathbf{S}_n(\mathbf{I}_N - \rho_0\mathbf{W}_N)^{-1} = \lim_{n \rightarrow \infty} \mathbf{I}_n$ . Therefore (19) would be asymptotically reduced to  $\Sigma_0^{-1/2}\bar{\mathbf{Y}}_n = \mathbf{X}_n\beta_0 + \boldsymbol{\varepsilon}_n$ . In that case the parameter  $\rho$  is no longer identified<sup>7</sup>.

Secondly, this theoretical problem also accures if  $n$  is near  $N$ . In that case depending on the specified  $\mathbf{W}_N$  it is possible that  $\Sigma_0^{-1/2}\mathbf{S}_n(\mathbf{I}_N - \rho_0\mathbf{W}_N)^{-1}$  is near  $\mathbf{S}_n$ . As a result,  $\rho$  maybe badly identified by ().

A third problem is posed by the transformation in a numerical sense, since in most cases  $\Sigma_0$  will not be a sparse matrix and therefore even efficient algorithms may not be able to calculate  $\Sigma_0^{-1/2}$  in reasonable time if the observed data is huge ( $n > 5000$ ). An other obstacle is that porgrams like matlab don't have a sparse routine for  $\mathbf{X}^{1/2}$ . As a result, the time increases for calculating  $\mathbf{X}^{1/2}$  with  $\sim n^3$ .

In the following subsection we assume in addition<sup>8</sup>:

1.  $n$ ,  $N$  and  $\mathbf{W}_N$  take values so that  $\rho$  is still identified by equation (19).
2.  $\left\| \Sigma_0^{-1/2} \right\|_1 < \infty$  and  $\left\| \Sigma_0^{-1/2} \right\|_\infty < \infty$  for all sample sizes and all possible  $\rho$

### 3.1.1 Enhancing efficiency for the NLS Estimators

The minimization of (13) yielded consistent or almost consistent estimators for  $\rho_0$ . We use this estimation to calculate  $\tilde{\mathbf{Y}}_n := \hat{\Sigma}_n^{-1/2}\bar{\mathbf{Y}}_n$ . Further

<sup>7</sup>In that case where  $\rho$  is not identified by () it is still possible to do a simple regression of  $\mathbf{S}_n\mathbf{X}_N$  on  $\Sigma_0^{-1/2}\bar{\mathbf{Y}}_n$  in order to get efficient estimates for  $\beta$ . This is like doing a Corcane Orcut transformation on ().

<sup>8</sup>These assumptions are necessary so that all the proof- logic shown in the appendix still holds for consicentcy, almost consictnecy and asymptotic normality.

we define  $\tilde{\mathbf{S}}_n := \hat{\Sigma}_n^{-1/2} \mathbf{S}_n = \left( \hat{\Sigma}_n^{-1/2} \mathbf{0}_{n \times (N-n)} \right)$ . We are now faced with the following data generating process:

$$\tilde{\mathbf{Y}}_n = \tilde{\mathbf{S}}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \mathbf{X}_N \beta_0 + \varepsilon_n \quad (20)$$

$$\tilde{\mathbf{Y}}_n \approx \tilde{\mathbf{S}}_n \sum_{k=0}^{m(n)} \rho_0^k \mathbf{W}_N^k \mathbf{X}_N \beta_0 + \varepsilon_n \quad (21)$$

Note that the estimator for  $\theta_0$  will have the following asymptotic distribution, since for both estimators  $\mathbf{A}_0 = \sigma_0^2 \mathbf{I}_n$  and  $\mathbf{A}_0 = \mathbf{D}_0 \mathbf{D}_0'$ .

$$n^{-1/2} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{D} N \left( 0, \sigma_0^2 \mathbf{A}_0^{-1} \right) \quad (22)$$

$$\text{where } \mathbf{D}_0 = \frac{1}{n} \begin{cases} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N (\rho_0)' \tilde{\mathbf{S}}'_n \\ \beta'_0 \mathbf{X}'_N \overline{\mathbf{G}}_N (\rho_0)' \tilde{\mathbf{S}}'_n \end{pmatrix} & \text{for GMM-e} \\ \begin{pmatrix} \mathbf{X}'_N \sum_{k=0}^{m(n)} \rho_0^k \mathbf{W}_N^k \tilde{\mathbf{S}}'_n \\ \beta'_0 \mathbf{X}'_N \sum_{k=1}^{m(n)} k \rho_0^{k-1} \mathbf{W}_N^k \tilde{\mathbf{S}}'_n \end{pmatrix} & \text{for GMM-a} \end{cases}$$

Obviously the variance of (22) has smaller eigenvalues than that of (15) and (14) and hence more efficient.

### 3.1.2 Enhancing efficiency for the OLS Estimator

If we use the same notation as in the NLS- case we have to estimate the following process:

$$\tilde{\mathbf{Y}}_n = \mathbf{Z}_n \eta_0 + \varepsilon_n \quad (23)$$

where  $\tilde{\mathbf{Z}}_n = \tilde{\mathbf{S}}_n [\mathbf{W}_N^0 \mathbf{X}_N, \mathbf{W}_N^1 \mathbf{X}_N, \mathbf{W}_N^2 \mathbf{X}_N, \dots, \mathbf{W}_N^{m(n)} \mathbf{X}_N]$ . If we use the consistent estimator<sup>9</sup>  $\hat{\eta}_n = \left( \tilde{\mathbf{Z}}'_n \tilde{\mathbf{Z}}_n \right)^{-1} \tilde{\mathbf{Z}}'_n \tilde{\mathbf{Y}}_n$  the asymptotic distribution is

$$n^{-1/2} (\hat{\eta} - \eta_0) \sim N \left( \mathbf{0}, \frac{\sigma_0^2}{n} \left( \tilde{\mathbf{Z}}'_n \tilde{\mathbf{Z}}_n \right)^{-1} \right) \quad (24)$$

Obviously the variance of (24) has smaller eigenvalues than that of (17)

and hence is more efficient. Additionally, one only needs to do the second Transformation  $T_2$  in order to get estimations for  $\theta_0$ .

<sup>9</sup>this is basically a GLS-estimator (see Green page 207)

## 3.2 Numerical properties

In this section we focus on the different numerical properties of the estimators. The GMM-e and the NLS-e Estimator are obviously the most expensive in a numerical sense. Both have to minimize a criteria function over  $k + 1$  dimensions and in each optimization step the inverse of  $(\mathbf{I}_N - \rho \mathbf{W}_N)$  has to be calculated. In each optimization run calculations with matrices of the size of  $N$  need to be handed. Keep in mind that to calculate a gradient of the objective function one has to evaluate the Inverse at least at two points in each optimization step. Of course the GMM- method is more expensive then the NLS- estimation, since the GMM- procedure consists of two steps and the criteria function is more complicated. Since for exact algorithms the time to compute  $(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}$  rises with  $\sim N^3$  if  $\mathbf{W}_N$  isn't a sparse matrix, both estimators should only be used for small to medium sized  $N$  in these circumstances.

Since the GMM-a, NLS-a and OLS Estimator only use approximations there is no need to calculate  $(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}$ . Furthermore it is possible to calculate  $\mathbf{S}_n[\mathbf{W}_N^0 \mathbf{X}_N, \mathbf{W}_N^1 \mathbf{X}_N, \mathbf{W}_N^2 \mathbf{X}_N, \dots, \mathbf{W}_N^{m(n)} \mathbf{X}_N]$  only once at the beginning of the optimization and therefore in each method one only have to handle matrices with the size  $n$ .

GMM- a is next to the GMM-e and NLS-e the most numerical expensive, since it poses a  $k + 1$  dimensional optimization problem with a relatively complicated criteria function.

On the other hand NLS-a poses, in some sense, only a unidimensional problem: First restrict the maximization parameter only to  $\rho$ , calculate in each optimization step  $\mathbf{X}\mathbf{X} = \mathbf{S}_n[\rho^0 \overline{\mathbf{W}^0 \mathbf{X}}, \rho^1 \overline{\mathbf{W}^1 \mathbf{X}}, \rho^2 \overline{\mathbf{W}^2 \mathbf{X}}, \dots, \rho^m \overline{\mathbf{W}^m \mathbf{X}}]^{10}$ , regress  $\mathbf{X}\mathbf{X}$  on  $\overline{\mathbf{Y}}$  and use the estimated sum of squared residuals as criteria function.

The OLS- estimator is obviously numerical the cheapest. It only needs to perform one regression of  $\mathbf{Z}$  on  $\overline{\mathbf{Y}}$  which can be programmed very efficiently. The only optimization routine that is needed to get an efficient Transformation  $T_2$ , is a  $mk + k$  dimensional problem. The matrices handled during this optimization are only of the size  $mk + k$ . One must also keep in mind that in order to find consistent estimations for  $\theta_0$  it is not necessary to find a global minimum for the Variance.

Time of the different algorithms taken in MC:

Graph...

---

<sup>10</sup> where  $\overline{\mathbf{W}^i \mathbf{X}} = \mathbf{S}_n \mathbf{W}_N^i \mathbf{X}_N$

## 4 Monte-Carlo study

### 4.1 Basic Monte-Carlo Design

### 4.2 Spatial dependence - used $\mathbf{W}$ -matrices

### 4.3 Results

## 5 Appendix

### 5.1 Useful Lemmas

**Proof.**

**Lemma 1** *Due to Assumption (1) and (2) it follows:  $\|(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}\|_1 \leq a < \infty$*

**Proof.**  $\|(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}\|_1 = \left\| \sum_{k=0}^{\infty} \rho^k \mathbf{W}_N^k \right\|_1 \leq \sum_{k=0}^{\infty} |\rho|^k \|\mathbf{W}\|_N^k = \frac{1}{1-|\rho|}$  ■

■

**Lemma 2** *Let  $|\rho_0 + \Delta\rho| < 1$ . It follows that  $\mathbf{G}_N(\rho_0 + \Delta\rho) = \mathbf{G}_N(\rho_0) + \Delta\mathbf{G}_N(\rho_0, \Delta\rho)$  where  $\Delta\mathbf{G}_N(\rho_0, \Delta\rho) = \mathbf{0} \Rightarrow \Delta\rho = 0$  and  $\|\Delta_N \mathbf{G}(\rho_0, \Delta\rho)\|_1 \leq \Delta a < \infty$*

**Proof.**  $\mathbf{G}(\rho_0 + \Delta\rho) = (\mathbf{I}_N - (\rho_0 + \Delta\rho) \mathbf{W}_N)^{-1} = \sum_{k=0}^{\infty} (\rho_0 + \Delta\rho)^k \mathbf{W}_N^k =$   
 $\sum_{k=0}^{\infty} \mathbf{W}_N^k \sum_{i=0}^k \binom{k}{i} \rho_0^{k-i} \Delta\rho^i = \sum_{k=0}^{\infty} \mathbf{W}_N^k \left( \rho_0^k + \Delta\rho^k + \sum_{i=1}^{k-1} \binom{k}{i} \rho_0^{k-i} \Delta\rho^i \right) =$   
 $(\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} + \sum_{k=0}^{\infty} \mathbf{W}_N^k \left( \Delta\rho^k + \sum_{i=1}^{k-1} \binom{k}{i} \rho_0^{k-i} \Delta\rho^i \right) = \mathbf{G}_N(\rho_0) +$   
 $\Delta\mathbf{G}_N(\rho_0, \Delta\rho)$  ■

**Proof.**  $\sum_{k=0}^{\infty} \mathbf{W}_N^k \left( \Delta\rho^k + \sum_{i=1}^{k-1} \binom{k}{i} \rho_0^{k-i} \Delta\rho^i \right) = \mathbf{0} \Rightarrow \Delta\rho = 0$  trivial; ■

**Proof.**  $\left\| \sum_{k=0}^{\infty} \mathbf{W}_N^k \left( \Delta \rho^k + \sum_{i=1}^{k-1} \binom{k}{i} \rho_0^{k-i} \Delta \rho^i \right) \right\|_1 \leq$

$$\sum_{k=0}^{\infty} \left\| \mathbf{W}_N^k \right\|_1 \left| \left( \Delta \rho^k - \rho_0^k + \rho_0^k + \sum_{i=1}^{k-1} \binom{k}{i} \rho_0^{k-i} \Delta \rho^i \right) \right| \leq \sum_{k=0}^{\infty} \left| (\rho_0 + \Delta \rho)^k - \rho_0^k \right| \leq$$

$$\frac{1}{1-|\rho_0+\Delta\rho|} + \frac{1}{1-|\rho_0|} \leq \Delta a < \infty$$

**Lemma 3** Under the proposed assumptions one can write:

$$g_{2,i}(X_N, S_{n(N)}, W_N, \theta) = \mathbf{g}_{1,i}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta) +$$

$$\frac{1}{n} \overline{\mathbf{W}_N^i \mathbf{X}_n}' \mathbf{S}_{n(N)} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) \blacksquare$$

**Proof.**  $g_{2,i}(X_N, S_{n(N)}, W_N, \theta) =$

$$\frac{1}{n} \overline{\mathbf{W}_N^i \mathbf{X}_n}' \left( \overline{\mathbf{Y}_n} - \mathbf{S}_{n(N)} \sum_{j=0}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) \right)$$

$$+ \frac{1}{n} \overline{\mathbf{W}_N^i \mathbf{X}_n}' \mathbf{S}_{n(N)} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta)$$

$$= \mathbf{g}_{1,i}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta) +$$

$$\frac{1}{n} \overline{\mathbf{W}_N^i \mathbf{X}_n}' \mathbf{S}_{n(N)} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta)$$

**Lemma 4** For every  $i, j$ :  $E \left[ {}_j \widetilde{\mathbf{X}}_i \cdot \overrightarrow{\mathbf{X}} \right] = 0$  only if  $\overrightarrow{\mathbf{X}} = 0$  where  $\widetilde{\mathbf{X}}_i = (\mathbf{S}_{n(N)} \mathbf{W}_N^i \mathbf{X}_N)'$  and  $\overrightarrow{\mathbf{X}} = -\mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \Delta \beta - \Delta \mathbf{G}_N(\rho_0, \Delta \rho) \mathbf{X}_N (\beta_0 + \Delta \beta)$  (proof is not correct)

■

**Proof.**  $E \left[ {}_j \widetilde{\mathbf{X}}_i \cdot \overrightarrow{\mathbf{X}} \right] = 0$  is only possible if every (with  $j$  indicated) row vector of  $\widetilde{\mathbf{X}}_i$  is orthogonal with  $\overrightarrow{\mathbf{X}}$ . If  $\Delta \beta \neq 0$  and  $\Delta \rho \neq 0$  it follows that for at least one  $(i, j) \in \{\{0, 1, \dots, z\}, \{1, \dots, k\}\}$  that  $\langle {}_j \widetilde{\mathbf{X}}_i, \overrightarrow{\mathbf{X}} \rangle \neq 0$

$$\langle {}_j \widetilde{\mathbf{X}}_i, \overrightarrow{\mathbf{X}} \rangle = - \langle {}_j \widetilde{\mathbf{X}}_i, S_n G_N(\rho_0) X_N \Delta \beta \rangle - \langle {}_j \widetilde{\mathbf{X}}_i, S_n \Delta G_N(\rho_0, \Delta \rho) X_N \beta_0 \rangle$$

$$- \langle {}_j \widetilde{\mathbf{X}}_i, \Delta G_N(\rho_0, \Delta \rho) X_N \Delta \beta \rangle \neq 0 \text{ since}$$

$$\langle {}_j \widetilde{\mathbf{X}}_i, S_n G_N(\rho_0) X_N \Delta \beta \rangle = \langle {}_j (\mathbf{S}_{n(N)} \mathbf{W}_N^i \mathbf{X}_N), S_n \sum_{k=0}^{\infty} \rho_0^k W_N^k X_N \Delta \beta \rangle$$

$\neq 0$  for at least one  $(i, j) \in \{\{0, 1, \dots, z\}, \{1, \dots, k\}\}$  if  $\Delta \beta \neq 0$  and  $\Delta \beta \neq -\beta_0$

$$\langle {}_j \widetilde{\mathbf{X}}_i, \Delta G_N(\rho_0, \Delta \rho) X_N \beta_0 \rangle =$$



$\langle_j (\mathbf{S}_{n(N)} \mathbf{W}_N^i \mathbf{X}_N), S_n \sum_{k=0}^{\infty} W_N^k \left( \sum_{i=1}^{k-1} \binom{k}{i} \rho_0^{k-i} \Delta \rho^i \Delta \rho^k \right) X_N(\beta_0 + \Delta\beta) \rangle \neq 0$  for at least one  $(i, j) \in \{\{0, 1, \dots, z\}, \{1, \dots, k\}\}$  if  $\Delta\rho \neq 0$  ■

## 5.2 Appendix for chapter 2

### 5.2.1 Rewriting (1) for missing $y_3$

$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \beta + \boldsymbol{\varepsilon}$  where  $\varepsilon_i \sim i.i.d.(0, \sigma^2)$ . This process can be rewritten as:

$$y_i = \begin{cases} \rho y_2 + \mathbf{x}_1 \beta + \epsilon_1 & \text{if } i = 1 \\ \rho y_{i-1} + \rho y_{i+1} + \mathbf{x}_i \beta + \epsilon_i & \text{if } i \in \{2, 3, \dots, n-1\} \\ \rho y_{n-1} + \mathbf{x}_n \beta + \epsilon_n & \text{if } i = n \end{cases}$$

If  $y_3$  is not observed we have to substitute it in the expressions of  $y_4$  and  $y_2$ :

$$y_4 = \rho y_3 + \rho y_5 + \mathbf{x}_4 \beta + \epsilon_4 = \rho(\rho y_2 + \rho y_4 + \mathbf{x}_3 \beta + \epsilon_3) + \rho y_5 + \mathbf{x}_4 \beta + \epsilon_4 \\ \Rightarrow y_4 = \frac{1}{1-\rho^2} (\rho y_5 + \mathbf{x}_4 \beta + \rho \mathbf{x}_3 \beta + \rho^2 y_2 + \rho \epsilon_3 + \epsilon_4)$$

and for  $y_2$ :

$$y_2 = \frac{1}{1-\rho^2} (\rho y_1 + \mathbf{x}_2 \beta + \rho \mathbf{x}_3 \beta + \rho^2 y_4 + \rho \epsilon_3 + \epsilon_2)$$

If we rewrite this matrix notation:

$$\bar{\mathbf{Y}} = \rho \mathbf{W}_1 \bar{\mathbf{Y}} + \frac{1}{1-\rho^2} \mathbf{X}_3 \beta + \frac{\rho}{1-\rho^2} (\mathbf{W}_2 \bar{\mathbf{Y}} + \mathbf{X}_2 \beta) + \frac{\rho^2}{1-\rho^2} \mathbf{W}_3 \bar{\mathbf{Y}} + \mathbf{X}_1 \beta + \boldsymbol{\varepsilon} \text{ where}$$

$$\mathbf{W}_1 = \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} & \mathbf{0}_{3,n-2} \\ \mathbf{0}_{n-3,2} & \widetilde{\mathbf{W}}_{n-3,n-2} \end{pmatrix}, \quad \widetilde{\mathbf{W}}_{n-3,n-2} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & & : \\ 0 & 1 & \dots & & : \\ : & \dots & \dots & & 0 \\ 0 & \dots & 0 & 1 & 0 & 1 \end{pmatrix} \quad \mathbf{W}_2 =$$

$$\begin{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \mathbf{0}_{n-3,4} & \mathbf{0}_{n-3,n-4} \end{pmatrix} & \mathbf{0}_{3,n-4} \\ \mathbf{0}_{n-3,4} & \mathbf{0}_{n-3,n-4} \end{pmatrix}, \quad \mathbf{W}_3 = \begin{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ \mathbf{0}_{n-3,4} & \mathbf{0}_{n-3,n-4} \end{pmatrix} & \mathbf{0}_{3,n-4} \\ \mathbf{0}_{n-3,4} & \mathbf{0}_{n-3,n-4} \end{pmatrix}, \quad \mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{0}_{1,k} \\ \mathbf{0}_{1,k} \\ \mathbf{x}_5 \\ : \\ \mathbf{x}_n \end{pmatrix},$$

$$\mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{1,k} \\ \mathbf{x}_3 \\ \mathbf{x}_3 \\ \mathbf{0}_{n-3,k} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{1,k} \\ \mathbf{x}_2 \\ \mathbf{x}_4 \\ \mathbf{0}_{n-3,k} \end{pmatrix}, \quad \bar{\mathbf{Y}} = \begin{pmatrix} y_1 \\ y_2 \\ y_4 \\ : \\ y_n \end{pmatrix} \text{ and } \boldsymbol{\varepsilon} = \begin{pmatrix} \epsilon_1 \\ \frac{1}{1-\rho^2} \epsilon_2 + \frac{\rho}{1-\rho^2} \epsilon_3 \\ \frac{1}{1-\rho^2} \epsilon_4 + \frac{\rho}{1-\rho^2} \epsilon_3 \\ \epsilon_5 \\ : \\ \epsilon_n \end{pmatrix}$$

### 5.2.2 Upper bound for approximation error:

**Proof.**  $\delta(\rho_0) := \max_{1 \leq i \leq n} \{|y_i - \tilde{y}_i|\} =$

$$\begin{aligned}
& \left\| \mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \mathbf{X}_N \beta_0 + \bar{\boldsymbol{\varepsilon}}_N - \mathbf{S}_N \sum_{j=0}^{m(n)} \rho_0^j \mathbf{W}_N^j \mathbf{X}_N \beta_0 - \bar{\boldsymbol{\varepsilon}} \right\| = \\
& \left\| \mathbf{S}_N \sum_{j=m(n)+1}^{\infty} \rho_0^j \mathbf{W}_N^j \mathbf{X}_N \beta_0 \right\| \leq \sum_{j=m(n)+1}^{\infty} |\rho_0^j| \|\mathbf{X}_N\| \|\beta_0\| k = \\
& k x_m \beta_m \left( \frac{1}{1-|\rho|} - \frac{1-|\rho|^{m(n)+1}}{1-|\rho|} \right) = k x_m \beta_m \frac{|\rho|^{m(n)+1}}{1-|\rho|} \text{ where} \\
& x_m = \max_{1 \leq i \leq n, 1 \leq j \leq k} (|x_{i,j}|), \beta_m = \max_{1 \leq j \leq k} (|\beta_{0,j}|) \blacksquare
\end{aligned}$$

### 5.2.3 Consistency proof for the GMM- estimators and OLS-estimator:

Matyas 2007 shows on page 12-14 that if the following 3 GMM- conditions are fulfilled the GMM- estimator is consistent.

**GMM- property 1:** (i)  $E [\mathbf{g} (X_N, S_{n(N)}, W_N, \theta)]$  exists and is finite for all  $\theta \in \Theta$  and for all  $N$

(ii) There exists only one  $\theta_0 \in \Theta$  for all  $N$  such that  $E [\mathbf{g} (X_N, S_{n(N)}, W_N, \theta_0)] = 0$

**Proof (i):**  $\Leftarrow E [\sup_{\theta \in \Theta} \|\mathbf{g} (X_N, S_{n(N)}, W_N, \theta)\|] < \infty$  for all  $N$   
 $\Leftrightarrow \forall i \in \{0, 1, \dots, z\} : E [\sup_{\theta \in \Theta} \|\frac{1}{n} g_i (X_N, S_{n(N)}, W_N, \theta)\|] < \infty$  for all  $N$

a.) GMM- exakt:

**Proof.**  $\Leftrightarrow \forall i \in \{0, 1, \dots, z\} :$

$$\begin{aligned}
& E \left[ \sup_{\theta \in \Theta} \left\| \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} (\overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho_0 + \Delta\rho) \mathbf{X}_N (\beta_0 + \Delta\beta)) \right\| \right] \\
& =: E [\sup_{\theta \in \Theta} \|F_1\|] \text{ now considering Lemma 2 for} \\
& F_1: F_1 = \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} (\overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \beta_0) - \\
& \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} (\mathbf{S}_n \Delta \mathbf{G}_N(\rho_0, \Delta\rho) \mathbf{X}_N (\beta_0 + \Delta\beta) - \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \Delta\beta) \Leftrightarrow F_1 = \\
& \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} \overline{\mathbf{Y}_n} \mathbf{G}(\rho_0) \epsilon - \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} (\mathbf{S}_n \Delta \mathbf{G}_N(\rho_0, \Delta\rho) \mathbf{X}_N (\beta_0 + \Delta\beta) - \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \Delta\beta) \\
& \Rightarrow E [\sup_{\theta \in \Theta} \|F_1\|] = \\
& E \left[ \sup_{\theta \in \Theta} \left\| \frac{-1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} (\mathbf{S}_n \Delta \mathbf{G}_N(\rho_0, \Delta\rho) \mathbf{X}_N (\beta_0 + \Delta\beta) - \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \Delta\beta) \right\| \right] \leq \\
& E \left[ \sup_{\theta \in \Theta} \left\| \frac{-1}{n} \overline{\mathbf{W}^i \mathbf{X}'_n} (\mathbf{S}_n \Delta \mathbf{G}_N(\rho_0, \Delta\rho) \mathbf{X}_N (\beta_0 + \Delta\beta) - \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \Delta\beta) \right\| \right]
\end{aligned}$$

By using the Cauchy- Schwarz equality:  $E [\sup_{\theta \in \Theta} \|F_1\|] \leq$

$$\begin{aligned}
& E \left[ \sup_{\theta \in \Theta} \frac{k}{n} \left( \left\| \overline{\mathbf{W}^i \mathbf{X}'_n} \right\| \mathbf{1}_{n \times k} \right)' (\mathbf{1}_{n \times k} \|\mathbf{S}_{n(N)}\| \|\Delta \mathbf{G}_N(\rho_0, \Delta\rho)\| \|\mathbf{X}_N\| \|(\beta_0 + \Delta\beta)\|) \right] \\
& + E \left[ \sup_{\theta \in \Theta} \left\| \frac{k}{n} \left( \left\| \overline{\mathbf{W}^i \mathbf{X}'_n} \right\| \mathbf{1}_{n \times k} \right)' (\mathbf{1}_{n \times k} \|\mathbf{S}_{n(N)}\| \|\mathbf{G}_N(\rho_0)\| \|\mathbf{X}_N\| \|\Delta\beta\|) \right\| \right] \leq
\end{aligned}$$

$E [\sup_{\theta \in \Theta} kx_m \mathbf{1}_{k \times 1} \Delta a x_m \beta_m] + E [\sup_{\theta \in \Theta} kx_m \mathbf{1}_{k \times 1} a \mathbf{x}_m 2 \beta_m] < \infty \Rightarrow \forall i \in \{0, 1, \dots, z\} : E [\sup_{\theta \in \Theta} \left\| \frac{1}{n} g_{1,i} (X_N, S_{n(N)}, W_N, \theta) \right\|] < \infty$  for all  $N$  ■

b.) GMM- approx:

**Proof.**  $\Leftrightarrow \forall i \in \{0, 1, \dots, z\} :$

$$E \left[ \sup_{\theta \in \Theta} \left\| \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}_n}' \left( \overline{\mathbf{Y}_n} - \mathbf{S}_{n(N)} \sum_{j=0}^{m(n)} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) \right) \right\| \right] =:$$

$$E [\sup_{\theta \in \Theta} \|F_2\|]$$

$$F_2 = \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}_n}' \left( \overline{\mathbf{Y}_n} - \mathbf{S}_{n(N)} \sum_{j=0}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) \right) +$$

$$\frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}_n}' \mathbf{S} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}^j \mathbf{X} (\beta_0 + \Delta \beta)$$

$$\Rightarrow E [\sup_{\theta \in \Theta} \|F_2\|] \leq E [\sup_{\theta \in \Theta} \|F_1\|]$$

$$+ E \left[ \sup_{\theta \in \Theta} \left\| \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}_n}' \mathbf{S}_{n(N)} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) \right\| \right] =$$

$$\Leftrightarrow E [\sup_{\theta \in \Theta} \|F_2\|] \leq E [\sup_{\theta \in \Theta} \|F_1\|] + E \left[ \sup_{\theta \in \Theta} \left\| \frac{1}{n} x_m^2 \mathbf{1}_{k \times 1} \beta_m \frac{|\Delta \rho + \rho_0|^{m(n)+1}}{1 - |\Delta \rho + \rho_0|} \right\| \right] <$$

$\infty$  ■

c.) OLS:

**Proof.**  $E [\mathbf{g}_{OLS,N} (X_N, S_{n(N)}, W_N, \eta)] =$

$$E \left[ n^{-1} \left[ \overline{\mathbf{W}^0 \mathbf{X}_n}, \overline{\mathbf{W}^1 \mathbf{X}_n}, \dots, \overline{\mathbf{W}^m \mathbf{X}_n} \right]' \left( \overline{\mathbf{Y}_n} - \left[ \overline{\mathbf{W}^0 \mathbf{X}_n}, \overline{\mathbf{W}^1 \mathbf{X}_n}, \dots, \overline{\mathbf{W}^m \mathbf{X}_n} \right] \eta \right) \right] \leq$$

$$E \left[ n^{-1} x_m \mathbf{1}_{(m(n)+1) \cdot k \times n(N)} \left( y_m \mathbf{1}_{n(N) \times 1} - x_m \eta_m \mathbf{1}_{n(N) \times 1} \right) \right] = x_m (y_m - x_m \eta_m) \mathbf{1}_{(m(n)+1)k \times 1} <$$

$\infty$  where  $\eta_m = \max_{i,j} \{\rho_0^i \beta_{0,j}\} = \beta_m$  ■

**Proof (ii):**

a.) GMM- exact:

**Proof.**  $E [\mathbf{g}_1 (\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta)] = 0 \rightarrow \theta = \theta_0$

$$\Leftrightarrow \forall i \in \{0, 1, \dots, z\} : E [g_{1,i} (\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta)] = 0 \rightarrow \theta = \theta_0$$

$$\Leftrightarrow \forall i \in \{0, 1, \dots, z\} : E \left[ \frac{1}{n} \overline{\mathbf{W}^i \mathbf{X}_n}' \left( \overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho) \mathbf{X}_N \beta \right) \right] = 0 \rightarrow \theta = \theta_0$$

The use of Lemma 2, Lemma 4 and setting  $\beta = \beta_0 + \Delta \beta$  and  $\rho = \rho_0 + \Delta \rho$  shows that  $E [] = 0$  is only possible if  $E [\overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho) \mathbf{X}_N \beta] = 0$ .

Rewriting  $E [\overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho) \mathbf{X}_N \beta]$  to  $E [-\mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{X}_N \Delta \beta - \Delta \mathbf{G}_N(\rho_0, \Delta \rho) \mathbf{X}_N (\beta_0 + \Delta \beta)]$  one can see that  $E [\overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho) \mathbf{X}_N \beta] = 0 \Rightarrow \Delta \rho = 0, \Delta \beta = 0$  ■

b.) GMM- approximate:

**Proof.** The GMM- approximate estimator only satisfies this condition

approximately. The minimization has an error that gets exponentially smaller as  $m(n)$  increases. Using the same logic as in Lemma 4 one

has only to show weather:  $E \left[ \bar{\mathbf{Y}}_n - \mathbf{S}_n \sum_{j=0}^{m(n)} \rho^j \mathbf{W}_N^j \mathbf{X}_N \beta \right] = 0 \Rightarrow \theta = \theta_0$

Setting  $\rho = \rho_0 + \Delta\rho$  and  $\beta = \beta_0 + \Delta\beta$

$$\begin{aligned} & E \left[ \bar{\mathbf{Y}}_n - \mathbf{S}_n \sum_{j=0}^{m(n)} (\rho_0 + \Delta\rho)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta\beta) \right] = 0 \\ \Leftrightarrow & E \left[ \bar{\mathbf{Y}}_n - \mathbf{S}_n \left( \mathbf{G}_N(\rho_0 + \Delta\rho) - \sum_{j=m(n)+1}^{\infty} (\rho_0 + \Delta\rho)^j \mathbf{W}_N^j \right) \mathbf{X}_N (\beta_0 + \Delta\beta) \right] = \\ 0 & \\ \Leftrightarrow & E \left[ \bar{\mathbf{Y}}_n - \mathbf{S}_n \mathbf{G}_N(\rho_0 + \Delta\rho) \mathbf{X}_N (\beta_0 + \Delta\beta) \right] - \\ & E \left[ \underbrace{\sum_{j=m(n)+1}^{\infty} (\rho_0 + \Delta\rho)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta\beta)}_{=\delta(m(n))} \right] = 0 \end{aligned}$$

One can see that the estimation problem has an error to exact minimization problem of  $\delta(m(n))$  :

$$\delta(m(n)) = \sum_{j=m(n)+1}^{\infty} (\rho_0 + \Delta\rho)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta\beta)$$

This error must be smaller then

$$\|\delta(m(n))\| \leq \left( \frac{1}{1-|\rho_0+\Delta\rho|} - \frac{1-|\rho_0+\Delta\rho|^{m(n)+1}}{1-|\rho_0+\Delta\rho|} \right) x_m \beta_m$$

$$\|\delta(m(n))\| \leq \frac{|\rho_0+\Delta\rho|^{m(n)+1}}{1-|\rho_0+\Delta\rho|} x_m \beta_m \quad \blacksquare$$

c.) OLS- estimator:

**Proof.**  $E \left[ \mathbf{g}_{OLS,N} (X_N, S_{n(N)}, W_N, \eta) \right] = \mathbf{0} \Rightarrow \theta = \theta_0$

$$\Leftrightarrow E \left[ n(N)^{-1} \mathbf{Z}'_n (\bar{\mathbf{Y}}_n - \mathbf{Z}_n \eta) \right] = 0 \text{ let } \eta = \eta_0 + \Delta\eta$$

$$\Rightarrow E \left[ n(N)^{-1} \mathbf{Z}'_n (\bar{\mathbf{Y}}_n - \mathbf{Z}_n (\eta_0 + \Delta\eta)) \right] = 0$$

$$\Leftrightarrow E \left[ n(N)^{-1} \mathbf{Z}'_n \left( \bar{\mathbf{Y}}_n - \mathbf{S}_n \sum_{l=0}^{m(n)} \rho^l \mathbf{W}_N^l \mathbf{X}_N \beta_0 - \mathbf{S}_n \sum_{l=0}^{m(n)} \Delta\rho^l \mathbf{W}_N^l \mathbf{X}_N \Delta\beta \right) \right] =$$

$$0 \\ \Leftrightarrow E \left[ n(N)^{-1} \mathbf{Z}'_n \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon} \right] -$$

$$\begin{aligned}
& \mathbf{E} \left[ n(N)^{-1} \mathbf{Z}'_n \left( \mathbf{S}_n \sum_{l=m(n)+1}^{\infty} \rho_0^l \mathbf{W}_N^l \mathbf{X}_N \beta_0 - \mathbf{S}_n \sum_{l=0}^{m(n)} \Delta \rho^l \mathbf{W}_N^l \mathbf{X}_N \Delta \beta \right) \right] = \\
0 & \\
& \Leftrightarrow E \left[ n(N)^{-1} \mathbf{Z}'_n \mathbf{S}_n \left( \underbrace{\sum_{l=m(n)+1}^{\infty} \rho_0^l \mathbf{W}_N^l \mathbf{X}_N \beta_0}_{=:\delta(m(n))} - \mathbf{Z} \Delta \eta \right) \right] = \mathbf{0}
\end{aligned}$$

If it is true that  $\text{rank}(\mathbf{Z}') = m(n)k + k$  then the minimization problem is identified only at  $\Delta \eta = \mathbf{0}$  if and only if  $\delta(m(n)) = \mathbf{0}$ . Since  $\delta(m(n)) \leq \frac{|\rho_0|^{m(n)+1}}{1-|\rho_0|} x_m \beta_m \mathbf{1}_{(m(n)+1)k}$  is near  $\mathbf{0}$  if  $m(n)$  is high enough, the estimator fulfills almost this property. ■

**GMM- property 2:** (i)  $\Theta$  is compact

(ii)  $\mathbf{g}(X_N, S_{n(N)}, W_N, \theta) - E[\mathbf{g}(X_N, S_{n(N)}, W_N, \theta)] \xrightarrow{p} 0$  pointwise on  $\Theta$

(iii)  $\mathbf{g}(X_N, S_{n(N)}, W_N, \theta)$  is stochastically equicontinuous and  $E[\mathbf{g}(X_N, S_{n(N)}, W_N, \theta)]$  is equicontinuous.

(i) **Property 1 is fulfilled due assumption 1**

(ii) **Proof Property 2**  $\forall \epsilon > 0 : \lim_{N \rightarrow \infty} P(\|\mathbf{g}_{1,2}() - E[\mathbf{g}_{1,2}()]\| > \epsilon) \stackrel{?}{=} 0$

$0 \Leftrightarrow$

$\forall i \in \{0, 1, \dots, z\}, \epsilon > 0 : \lim_{N \rightarrow \infty} P(\|g_{1,2;i}() - E[g_{1,2;i}()]\| > \epsilon) \stackrel{?}{=} 0$

**GMM- exact:**

**Proof.**  $\lim_{N \rightarrow \infty} P\left(\left\|E[g_{1,i}] - E[g_{1,i}] - \frac{1}{n(N)} \overline{\mathbf{W}_N^i \mathbf{X}_n'} \mathbf{S}_{n(N)} \mathbf{G}_N(\rho_0) \epsilon\right\| > \epsilon\right) = \lim_{N \rightarrow \infty} P\left(\left\|E\left[-\frac{1}{n(N)} \overline{\mathbf{W}_N^i \mathbf{X}_n'} \mathbf{S}_{n(N)} \mathbf{G}_N(\rho_0) \cdot \epsilon\right]\right\| > \epsilon\right) = 0 \quad \blacksquare$

**GMM- approx:**

**Proof.**  $\lim_{N \rightarrow \infty} P(\|\mathbf{g}_2(X_N, S_{n(N)}, W_N, \theta) - E[\mathbf{g}_2(X_N, S_{n(N)}, W_N, \theta)]\| > \epsilon)$

Using Lemma (3)

$$\begin{aligned}
& \Leftrightarrow \lim_{N \rightarrow \infty} P \left( \left\| g_{1,i} + \frac{1}{n} \overline{\mathbf{W}_N^i \mathbf{X}_n'} \mathbf{S}_{n(N)} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) - \right. \right. \\
& \left. \left. E \left[ g_{1,i} + \frac{1}{n} \overline{\mathbf{W}_N^i \mathbf{X}_n'} \mathbf{S}_{n(N)} \sum_{j=m(n)+1}^{\infty} (\Delta \rho + \rho_0)^j \mathbf{W}_N^j \mathbf{X}_N (\beta_0 + \Delta \beta) \right] \right\| \right) > \epsilon
\end{aligned}$$

$$\Leftrightarrow \lim_{N \rightarrow \infty} P \left( \|g_{1,i}(X_N, S_{n(N)}, W_N, \theta) - E[g_{1,i}(X_N, S_{n(N)}, W_N, \theta)]\| > \epsilon \right) = 0 \quad \blacksquare$$

OLS- estimator:

**Proof.**  $\mathbf{g}_{OLS,N}(X_N, S_{n(N)}, W_N, (\eta_0 + \Delta\eta)) - E[\mathbf{g}_{OLS,N}(X_N, S_{n(N)}, W_N, (\eta_0 + \Delta\eta))] \xrightarrow{p} 0$

$$\begin{aligned} E[\mathbf{g}_{OLS,N}(X_N, S_{n(N)}, W_N, (\eta_0 + \Delta\eta))] &= E[n(N)^{-1} \mathbf{Z}'_n (\bar{\mathbf{Y}}_n - \mathbf{Z}_n (\eta_0 + \Delta\eta))] = \\ n(N)^{-1} \mathbf{Z}'_n \left( \mathbf{S}_n \sum_{l=m(n)+1}^{\infty} \rho_0^l \mathbf{W}_N^l \mathbf{X}_N \beta_0 - \mathbf{Z}_n \Delta\eta \right) \\ \Rightarrow \mathbf{g}_{OLS,N} - E[\mathbf{g}_{OLS,N}] &= n(N)^{-1} \mathbf{Z}'_n \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \boldsymbol{\epsilon} \xrightarrow{p} 0 \quad \blacksquare \end{aligned}$$

**Proof Assumption 3 Proof.** (iii) b.)  $E[\mathbf{g}(X_N, S_{n(N)}, W_N, \theta)] = 0$  is equicontinuous since the multiplication and addition of equicontinuous functions is equicontinuous.

(iii) a.)  $g(X_N, S_{n(N)}, W_N, \theta)$  is stochastically equicontinuous:

The sequence of stochastic functions  $\{\mathbf{g}(\epsilon, \theta)_N\}$  is said to be stochastically equicontinuous if there exists a set  $M \in \Omega$  where  $P(M) = 1$  and for every  $\gamma > 0$  there exists a  $\delta$  and such that for every  $\epsilon \in M$ :

$$\begin{aligned} \sup_{|\theta_1 - \theta_2| \leq \delta} |\mathbf{g}(\epsilon, \theta_1)_N - \mathbf{g}(\epsilon, \theta_2)_N| &\leq \gamma \\ \text{for all } N > N(\epsilon) \end{aligned}$$

...since for all  $\rho$  of the parameter space it holds that  $(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} = \sum_{k=0}^{\infty} \rho^k \cdot W_N^k$ , it follows directly that  $g(X_N, S_{n(N)}, W_N, \theta)$  is only a sum of polynoms in  $\rho$  multiplied with  $X\beta$ . Since every function of the sum is equicontinuous and the uniformly converging sum of equicontinuous functions is itself equicontinuous, it follows that  $g(X_N, S_{n(N)}, W_N, \theta)$  is equicontinuous... (actual proof is still work in progress)  $\blacksquare$

**GMM- property 3** There exists a non-random matrix sequence of positive definite matrices  $\bar{\boldsymbol{\Omega}}_N$  such that,  $\bar{\boldsymbol{\Omega}}_N - \hat{\boldsymbol{\Omega}}_N \xrightarrow{p} 0$ .

**Proof.** (i) This is obviously true for the first step for both estimators

since:  $\hat{\boldsymbol{\Omega}}_N = \mathbf{I}_N \quad \blacksquare$

**Proof.** (ii) The minimization of (10) and (7) are both yielding to consistent estimators for  $\rho_0$ . Therefore, one would use the inverse of (6) as weighing matrix:

Due to assumption (4) and the following rewriting of  $\hat{\boldsymbol{\Omega}}_N$  (for further details, see Green page 835,  $\hat{\boldsymbol{\Omega}}_N$  is positive definite:  $\mathbf{G}(\hat{\rho})$ )

$$\widehat{\Omega}_N = \frac{1}{n^2} \cdot \begin{pmatrix} \overline{\mathbf{X}}_n' \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \\ \overline{\mathbf{W}\mathbf{X}}_n' \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \\ \vdots \\ \overline{\mathbf{W}^z \mathbf{X}}_n' \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \end{pmatrix} \begin{pmatrix} \overline{\mathbf{X}}_n' \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \\ \overline{\mathbf{W}\mathbf{X}}_n' \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \\ \vdots \\ \overline{\mathbf{W}^z \mathbf{X}}_n' \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \end{pmatrix}'$$

Obviously  $\overline{\Omega}_N - \widehat{\Omega}_N \xrightarrow{p} 0$  holds since

$$\overline{\Omega}_N - \widehat{\Omega}_N = \frac{1}{n^2} \cdot \begin{pmatrix} \overline{\mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{X}}_n & \overline{\mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{W}\mathbf{X}}_n & \dots & \overline{\mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{W}^z \mathbf{X}}_n \\ \overline{\mathbf{W}\mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{X}}_n & \overline{\mathbf{W}\mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{W}\mathbf{X}}_n & \dots & \overline{\mathbf{W}\mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{W}^z \mathbf{X}}_n \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mathbf{W}^z \mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{X}}_n & \dots & \dots & \overline{\mathbf{W}^z \mathbf{X}}_n' \widetilde{\Omega}_S \overline{\mathbf{W}^z \mathbf{X}}_n \end{pmatrix}$$

where  $\widetilde{\Omega}_S = \mathbf{S}_n \sum_{k=0}^{\infty} \rho_0^k \mathbf{W}_N^k \sum_{k=0}^{\infty} \rho_0^k \mathbf{W}_N^k - \sum_{k=0}^{\infty} \widehat{\rho}^k \mathbf{W}_N^k \sum_{k=0}^{\infty} \widehat{\rho}^k \mathbf{W}_N^k \mathbf{S}_n' = \mathbf{S}_n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \rho_0^{k+l} \mathbf{W}_N^{k+l} - \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \widehat{\rho}^{k+l} \mathbf{W}_N^{k+l} \mathbf{S}_n' = \mathbf{S}_n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} (\rho_0^{k+l} - \widehat{\rho}^{k+l}) \mathbf{W}_N^{k+l} \mathbf{S}_n' \blacksquare$

#### 5.2.4 Asymptotic distributions of the GMM- estimators:

under the GMM- Assumptions the GMM- estimator  $\widehat{\theta}_n$  has the asymptotic distribution:

$$n^{-1/2} (\widehat{\theta}_n - \theta_0) \sim N(0, (\mathbf{G}_0' \Omega_{0,gmm} \mathbf{G}_0)^{-1}) \text{ where } \mathbf{G}_0 = \frac{\partial \mathbf{g}_1}{\partial \theta'} \Big|_{\theta=\theta_0}$$

(Matyas page 19, Theorem 1.2)

**GMM- property 4**  $\mathbf{g}(X_N, S_{n(N)}, W_N, \theta)$  is continuously differentiable with respect to  $\theta$  on  $\Theta$

GMM- exact:

**Proof.**  $\frac{\partial g_{1,i}}{\partial \theta} = \begin{pmatrix} \frac{\partial g_i}{\partial \beta} \\ \frac{\partial g_i}{\partial \rho} \end{pmatrix} = \frac{1}{n(N)} \begin{pmatrix} -\overline{\mathbf{W}^i \mathbf{X}}_n' \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \\ -\overline{\mathbf{W}^i \mathbf{X}}_n' \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-2} \mathbf{W}_N \mathbf{X}_N \beta \end{pmatrix} = \frac{1}{n(N)} \begin{pmatrix} -\overline{\mathbf{W}^i \mathbf{X}}_n' \mathbf{S}_n \sum_{k=0}^{\infty} \rho^k \mathbf{W}_N^k \mathbf{X}_N \\ -\overline{\mathbf{W}^i \mathbf{X}}_n' \mathbf{S}_n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \rho^{k+l} \mathbf{W}_N^{k+l+1} \mathbf{X}_N \beta \end{pmatrix}$  is continuous in  $\rho$  and  $\beta$ .  $\blacksquare$

GMM- approx:

**Proof.**  $\frac{\partial g_2}{\partial \theta} = \frac{-1}{n} \overline{\mathbf{W}^i \mathbf{X}}_n' \mathbf{S}_n \left( \left( \sum_{j=0}^{m(n)} \widehat{\rho}^j \mathbf{W}_N^j \mathbf{X}_N \right)', \left( \sum_{j=1}^{m(n)} j \widehat{\rho}^{j-1} \mathbf{W}_N^j \mathbf{X}_N \widehat{\beta} \right)' \right)'$

is continuous in  $\rho$  and  $\beta$ .  $\blacksquare$

**GMM- property 5** For any sequence  $\theta_N^*$  such that  $\theta_N^* \xrightarrow{p} \theta_0$ ,

$$\frac{1}{n} \sum_{k=1}^n \frac{\partial \mathbf{g}_k(X_k, S_{n(k)}, W_k, \theta)}{\partial \theta'} \Big|_{\theta=\theta_N^*} - \bar{F}_N \xrightarrow{p} 0$$

where  $\bar{F}_N$  is a sequence of  $q \times p$  matrices that do not depend on  $\theta$

**Proof.** Note that the proof of GMM- exact and GMM- approximate are

the same, since we have to show a probability limit and therefore the ap-

proximate model converges to the exact model:  $\frac{1}{n} \sum_{k=1}^n \frac{\partial \mathbf{g}_i(X_k, S_{n(k)}, W_k, \theta)}{\partial \theta'} \Big|_{\theta=\theta_N^*} =$

$$\begin{aligned} & \frac{1}{n(N)} \left( \begin{array}{c} -\overline{\mathbf{W}}_N^i \mathbf{X}_n' \mathbf{S}_n \sum_{k=0}^{\infty} (\rho_N^*)^k \mathbf{W}_N^k \mathbf{X}_N \\ \overline{\mathbf{W}}_N^i \mathbf{X}_n' \mathbf{S}_n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} (\rho_N^*)^{k+l} \mathbf{W}_N^{k+l+1} \mathbf{X}_N \beta_N^* \end{array} \right) \\ & \Rightarrow \frac{1}{n} \sum_{k=1}^n \frac{\partial \mathbf{g}_i(X_k, S_{n(k)}, W_k, \theta)}{\partial \theta'} \Big|_{\theta=\theta_N^*} - \bar{F}_N = \\ & \frac{1}{n(N)} \left( \begin{array}{c} -\overline{\mathbf{W}}_N^i \mathbf{X}_n' \mathbf{S}_n \sum_{k=0}^{\infty} (\rho_N^* - \rho_0)^k \mathbf{W}_N^k \mathbf{X}_N \\ (\mathbf{S}_n \overline{\mathbf{W}}_N^i \mathbf{X}_N)' \mathbf{S}_n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathbf{W}_N^{k+l+1} \mathbf{X}_N \left( (\rho_N^*)^{k+l} \beta_N^* - \rho_0^{k+l} \beta_0 \right) \end{array} \right) \xrightarrow{p} \end{aligned}$$

0, since  $\theta_N^* \xrightarrow{p} \theta_0$  ■

**OLS- estimator**

**Proof.**  $\frac{\partial \mathbf{g}_{OLS}(X_k, S_{n(k)}, W_k, \theta)}{\partial \eta'} \Big|_{\theta=\theta_N^*} = \frac{\partial}{\partial \eta'} n(N)^{-1} \mathbf{Z}_n' (\bar{\mathbf{Y}} - \mathbf{Z}_n \eta) - \bar{\mathbf{Z}}' \bar{\mathbf{Z}} \xrightarrow{p} 0$

where  $\bar{\mathbf{Z}}' \bar{\mathbf{Z}} = E [\lim_{n \rightarrow \infty} n(N)^{-1} \mathbf{Z}_n' \mathbf{Z}_n]$  ■

### 5.2.5 GMM- property 6

$\mathbf{g}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta)$  satisfies a central limit theorem, so that

$$\bar{\mathbf{V}}_N^{-1/2} \cdot \sqrt{n} \cdot \mathbf{g}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0) \xrightarrow{d} N(0, \mathbf{I}_{k+1})$$

where  $\bar{\mathbf{V}}_N = n \cdot \text{Var}(\mathbf{g}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0))$

**Proof.**

**Theorem 5** *Multivariate Lindberg- Feller Central Limit Theorem (Green page 912):*

If  $g(\mathbf{X}_N, \mathbf{S}_1, \mathbf{W}_N, \theta_0)_1, \dots, g(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0)_{n(N)}$  are random variables from a multivariate distribution with finite mean vector  $\mu$  and finite positive definite covariance Matrix  $\sigma^2 \bar{\mathbf{\Omega}}$ , then

$$\sqrt{n} (\bar{\mathbf{g}}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0) - \mu) \xrightarrow{d} N(0, \sigma^2 \bar{\mathbf{\Omega}})$$



where

$$\bar{\mathbf{g}}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0) = \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_N, \mathbf{S}_i, \mathbf{W}_N, \theta_0)_i$$

■

GMM- estimators:

**Proof.** i.)  $E[\mathbf{g}(\mathbf{X}_1, \mathbf{S}_{n(1)}, \mathbf{W}_1, \theta_0)] = 0$

ii.)  $\bar{\mathbf{\Omega}}$  is positive definite (see GMM\_Property 1.3)

iii.)  $\bar{\mathbf{\Omega}}$  has finite entries:

$$\|\sigma^2 \bar{\mathbf{\Omega}}\| \leq \frac{\sigma^2}{n^2} \left\| \begin{pmatrix} \bar{\mathbf{X}}_n' \|\mathbf{S}_n\| \|\mathbf{G}_N(\hat{\rho})\| \\ \overline{\mathbf{W}} \mathbf{X}_n' \|\mathbf{S}_n\| \|\mathbf{G}_N(\hat{\rho})\| \\ \vdots \\ \overline{\mathbf{W}}^z \mathbf{X}_n' \|\mathbf{S}_n\| \|\mathbf{G}_N(\hat{\rho})\| \end{pmatrix} \begin{pmatrix} \bar{\mathbf{X}}_n' \|\mathbf{S}_n\| \|\mathbf{G}_N(\hat{\rho})\| \\ \overline{\mathbf{W}} \mathbf{X}_n' \|\mathbf{S}_n\| \|\mathbf{G}_N(\hat{\rho})\| \\ \vdots \\ \overline{\mathbf{W}}^z \mathbf{X}_n' \|\mathbf{S}_n\| \|\mathbf{G}_N(\hat{\rho})\| \end{pmatrix}' \right\|$$

using Lemma 1

$$\|\sigma^2 \bar{\mathbf{\Omega}}\| \leq \frac{\sigma^2}{n^2} a^2 \left\| \begin{pmatrix} \bar{\mathbf{X}}_n' \\ \overline{\mathbf{W}} \mathbf{X}_n' \\ \vdots \\ \overline{\mathbf{W}}^z \mathbf{X}_n' \end{pmatrix} \begin{pmatrix} \bar{\mathbf{X}}_n' \\ \overline{\mathbf{W}} \mathbf{X}_n' \\ \vdots \\ \overline{\mathbf{W}}^z \mathbf{X}_n' \end{pmatrix}' \right\| < \infty \quad \blacksquare$$

OLS- estimator

**Proof.** (Ia):  $\bar{\mathbf{V}}_n = n \cdot \text{Var}(\mathbf{g}_{OLS}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0)) =$

$$E \left[ \frac{1}{n} \mathbf{Z}'_n \mathbf{S}_n \mathbf{G}_N(\rho_0) \epsilon_N \epsilon_N' \mathbf{G}'_N(\rho_0) \mathbf{S}'_{n(N)} \mathbf{Z}_n \right] = \frac{\sigma^2}{n} \mathbf{Z}'_n \mathbf{S}_n \mathbf{G}_N(\rho_0) \mathbf{G}'_N(\rho_0) \mathbf{S}'_n \mathbf{Z}_n$$

(Ib)  $\bar{\mathbf{V}}_n \dots$  is positive definite since  $\text{rank}(Z) = (m(n) + 1)k$

(Ic)  $\bar{\mathbf{V}}_n \dots$  has finite entries:  $\frac{1}{n} \mathbf{Z}'_n \mathbf{Z}_n \leq z_{\max}^2 \mathbf{1}_{m(n)k+k \times m(n)k+k}$

(II):  $\sqrt{n} \mathbf{g}(\mathbf{X}_N, \mathbf{S}_{n(N)}, \mathbf{W}_N, \theta_0) = \frac{1}{\sqrt{n}} \mathbf{Z}'_n \mathbf{S}_{n(N)} \mathbf{G}_N(\rho_0) \epsilon_N \xrightarrow{d} N(0, \bar{\mathbf{V}}_n)$

■

### 5.3 Consistency-proof for NLS- estimators:

If the following NLS- properties (NLS- properties 1-6) are fulfilled then the NLS- minimization yields consistent estimators (proof see Prucha Asymptotic script page 19). One must point out that in Prucha assumes that the  $y_i$  are i.i.d. but the proof logic also applies if the  $y_i$  have a dependence like  $(\mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon})_i$ . These conditions are written as properties for the function  $h_i(\cdot)$  where we define:  $\mathbf{v}_i = (\mathbf{S}_n (\mathbf{I}_N - \rho_0 \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon})_i$  where  $(\cdot)_i$  denotes the  $i$ th entry in this vector.

Notation for NLS- exact:

$$\begin{aligned}
Q_{1,n}(X_n, \beta, \rho) &= n^{-1} \sum_{i=1}^n q_1(x_i, \beta, \rho) = \\
&= \frac{1}{n} (\bar{\mathbf{Y}}_n - \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \beta)' (\bar{\mathbf{Y}}_n - \mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \beta) = \\
n^{-1} \sum_{i=1}^n (y_i - h_{1,i}(X, \rho, \beta))^2 &\Rightarrow h_{1,i}(X, \rho, \beta) = (\mathbf{S}_n (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \mathbf{X}_N \beta)_i
\end{aligned}$$

Notation for NLS- exact:

$$\begin{aligned}
Q_{1,n}(X_n, \beta, \rho) &= n^{-1} \sum_{i=1}^n q_2(x_i, \beta, \rho) = \\
&= \frac{1}{n} \left( \bar{\mathbf{Y}}_n - \mathbf{S}_n \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^k \mathbf{X}_N \beta \right)' \left( \bar{\mathbf{Y}}_n - \mathbf{S}_n \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^k \mathbf{X}_N \beta \right) = n^{-1} \sum_{i=1}^n (y_i - h_{2,i}(X, \rho, \beta))^2 \Rightarrow \\
h_{2,i}(X, \rho, \beta) &= \left( \mathbf{S}_n \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^k \mathbf{X}_N \beta \right)_i
\end{aligned}$$

### 5.3.1 NLS- property 1

$h_i(X, \rho, \beta)$  is a real valued function on  $\mathbb{R}^{N \times k} \times B$ ,  $B \subseteq \mathbb{R}^{k+1}$  where  $B$  is a Borel set

**Proof.** Both functions are real valued functions. Therefore NLS-1 property is fulfilled. ■

### 5.3.2 NLS- property 2

$B \subseteq \mathbb{R}^{k+1}$  is compact.

**Proof.** Fulfilled due to Assumption 1 ■

### 5.3.3 NLS- Property 3

$h_i(X, \cdot, \cdot)$  is Borel measurable for each  $[\beta', \rho]' \in B$

**Proof.** Both functions are continuous functions on auf  $[\beta', \rho]' \in \mathbb{R}^k \times$

$(-1, 1)$ . Therefore both are Borel- measurable. ■

### 5.3.4 NLS- Property 4

$\mathbf{z}_i = [\mathbf{y}_i, \mathbf{x}_i]$  with  $\mathbf{y}_i \in \mathbb{R}$  and  $\mathbf{x}_i \in \mathbb{R}^{n \times k}$  is a sequence of identically and independently distributed random vectors.

**Proof.** Fulfilled due to Assumption 3 ■

### 5.3.5 NLS- Property 5

$E[\mathbf{y}_i | \mathbf{x}_i] = h(\mathbf{x}_i, \rho_0, \beta_0)$  for  $\rho_0, \beta_0 \in B$ .

**Proof.** This obviously true for  $h_1$ , since

$$E[y_i|\mathbf{x}_i] = E[h_1(\rho_0, \beta_0, \mathbf{S}_i, \mathbf{W}_i, \mathbf{X}_i)_i + \mathbf{v}_i|\mathbf{x}_i] = E[h_1(\rho_0, \beta_0, \mathbf{S}_i, \mathbf{W}_i, \mathbf{X}_i)_i|\mathbf{x}_i] = E \left[ \left( \mathbf{S}_i \sum_{k=0}^{\infty} \rho_0^k \mathbf{W}_i^k \mathbf{X}_i \beta_0 \right)_i \right] = h_1(\rho_0, \beta_0, \mathbf{S}_i, \mathbf{W}_i, \mathbf{X}_i)_i$$

For  $h_2$  this is only true with an maximal error  $\delta(m(n)) \leq$ , since

$$E[y_i|\mathbf{x}_i] = E[h_2(\rho_0, \beta_0, \mathbf{S}_i, \mathbf{W}_i, \mathbf{X}_i)_i + \left( \mathbf{S}_i \sum_{k=m(n)+1}^{\infty} \rho_0^k \mathbf{W}_i^k \mathbf{X}_i \beta_0 \right)_i + \mathbf{v}_i|\mathbf{x}_i] = E[h_2(\rho_0, \beta_0, \mathbf{S}_i, \mathbf{W}_i, \mathbf{X}_i)_i|\mathbf{x}_i] + \delta(m(n), i) \quad \text{where } \delta(m(n), i) = \left( \mathbf{S}_i \sum_{k=m(n)+1}^{\infty} \rho_0^k \mathbf{W}_i^k \mathbf{X}_i \beta_0 \right)_i \Rightarrow \delta(m(n)) := \max_i \{|\delta(m(n), i)|\} \leq kx_m \beta_m \frac{|\rho_0|^{m(n)+1}}{1-|\rho_0|}$$

■

### 5.3.6 NLS- Property 6

$$E[y_i - h_{1,2}(\mathbf{x}_i, \rho_0, \beta_0)]^2 < \infty \text{ and } E[\sup_{[\beta', \rho]' \in B} (h_{1,2}(\mathbf{x}_i, \rho, \beta))^2] < \infty.$$

NLS- exact:

$$\mathbf{Proof.} \text{ (i) } E[y_i - h_1(\mathbf{x}_i, \rho_0, \beta_0)]^2 = E \left[ (\mathbf{S}_i (\mathbf{I}_i - \rho \mathbf{W}_i)^{-1})_i \varepsilon_i^2 \right] < a E \left[ (\varepsilon_i)_i^2 \right] =$$

$$a^2 \sigma^2 \quad \blacksquare$$

$$\mathbf{Proof.} \text{ (ii) } E[\sup_{[\beta', \rho]' \in B} (h_1(\mathbf{x}_i, \rho, \beta))^2] =$$

$$E \left[ \sup_{[\beta', \rho]' \in B} (\mathbf{S}_i (\mathbf{I}_i - \rho \mathbf{W}_i)^{-1} \mathbf{X}_i \beta)_i^2 \right] \leq (ax_m k \beta_m)^2 < \infty \quad \blacksquare$$

NLS- approx:

$$\mathbf{Proof.} \text{ (i) b.) } E[y_i - h_2(\mathbf{x}_i, \rho_0, \beta_0)]^2 =$$

$$E \left[ \left( - \left( \mathbf{S}_i \sum_{k=m(n)+1}^{\infty} \rho_0^k \mathbf{W}_i^k \mathbf{X}_i \beta_0 \right)_i + (\mathbf{S}_i (\mathbf{I}_i - \rho_0 \mathbf{W}_i)^{-1} \varepsilon_i)_i \right)^2 \right] \leq \frac{|\rho_0|^{m(n)+1}}{1-|\rho_0|} kx_{mi} \beta_0 + a^2 E[\varepsilon_i^2] < \infty \quad \blacksquare$$

$$\mathbf{Proof.} \text{ (ii) b.) } E[\sup_{[\beta', \rho]' \in B} (h_1(\mathbf{x}_i, \rho, \beta))^2] =$$

$$E \left[ \sup_{[\beta', \rho]' \in B} \left( h_1(\mathbf{x}_i, \rho, \beta) - \left( \mathbf{S}_i \sum_{k=m(n)+1}^{\infty} \rho^k \mathbf{W}_i^k \mathbf{X}_i \beta \right)_i \right)^2 \right] \leq (ax_m k \beta_m)^2 + 2ax_m^2 k^2 \beta_m^2 \frac{|\rho|^{m(n)+1}}{1-|\rho|} + x_m^2 k^2 \beta_m^2 \left( \frac{|\rho|^{m(n)+1}}{1-|\rho|} \right)^2 < \infty \quad \blacksquare$$

### 5.3.7 NLS- property 7

$$E[h_{1,2}(\mathbf{x}_i, \rho_0, \beta_0) - h_{1,2}(\mathbf{x}_i, \rho, \beta)]^2 > 0 \text{ for } \theta \neq \theta_0$$

NLS- Exact:

$$E[h_1(\mathbf{x}_i, \rho_0, \beta_0) - h_1(\mathbf{x}_i, \rho, \beta)]^2 = E \left[ \left( \mathbf{S}_i \sum_{k=0}^{\infty} (\rho_0^k \beta_0 - \rho^k \beta) \mathbf{W}_i^k \mathbf{X}_i \right)_i \right]^2 =$$

$0 \Rightarrow \theta = \theta_0$

NLS- approximate:

$$E[h_2(\mathbf{x}_i, \rho_0, \beta_0) - h_2(\mathbf{x}_i, \rho, \beta)]^2 = E \left[ \left( \mathbf{S}_i \sum_{k=0}^{m(n)} (\rho_0^k \beta_0 - \rho^k \beta) \mathbf{W}_i^k \mathbf{X}_i \right)_i \right]^2 =$$

$0 \Rightarrow \theta = \theta_0$

## 5.4 Asymptotic distribution- proof for NLS- estimators:

The following NLS- properties (1-7) have to be fulfilled in order that the following theorem holds (proof, see Prucha page 27; nonlinear econometric models)

$$\begin{aligned} n^{1/2} (\hat{\theta}_N - \theta_0) &\xrightarrow{D} \mathbf{A}_0^{-1} \mathbf{D}_0 \zeta \\ \text{if } \zeta &\sim N(0, \mathbf{\Lambda}_0), \text{ then} \\ n^{1/2} (\hat{\theta}_N - \theta_0) &\xrightarrow{D} N(0, \mathbf{A}_0^{-1} \mathbf{B}_0 \mathbf{A}_0^{-1}) \\ \text{with } \mathbf{B}_0 &= \mathbf{D}_0 \mathbf{\Lambda}_0 \mathbf{D}_0' \end{aligned}$$

### 5.4.1 NLS- property 8

The parameter space T and B are Compact subsets of  $\mathbb{R}^{p_\tau}$  and  $\mathbb{R}^{p_\theta}$ , respectively.

**Proof.** This is for both NLS estimators fulfilled due to Assumption 1 ■

### 5.4.2 NLS- property 9

$Q_N = Z^N \times T \times B \rightarrow \mathbb{R}$  where  $Q_N(z_1, \dots, z_N, \tau, \theta)$  is  $\mu$ - measurable for all  $(\tau, \theta) \in T \times B$  and  $Q_N(z_1, \dots, z_N, \tau, \theta)$  is a.s. twice continuous partially differentiable at every point  $(\tau, \theta)$  in the interior of  $T \times B$  (where exceptional null sets does not depend on  $(\tau, \theta)$ ).

NLS- exact:

**Proof.**  $Q_N(z_1, \dots, z_N, \tau_N, \theta_N) = \frac{1}{2n} \overline{\overline{\boldsymbol{\varepsilon}_n}}(\beta, \rho)' \overline{\overline{\boldsymbol{\varepsilon}_n}}(\beta, \rho)$  where  $\overline{\overline{\boldsymbol{\varepsilon}_n}}(\beta, \rho) = \overline{\mathbf{Y}_n} - \mathbf{S}_n \mathbf{G}_N(\rho) \mathbf{X}_N \beta$

$$\nabla_{\theta'} Q_N(z_1, \dots, z_N, \tau_N, \theta_N) = \frac{-1}{n} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\rho)' \mathbf{S}'_n \overline{\overline{\boldsymbol{\varepsilon}_n}}(\beta, \rho) \\ \beta' \mathbf{X}'_N \overline{\overline{\mathbf{G}_N}}(\rho)' \mathbf{S}'_n \overline{\overline{\boldsymbol{\varepsilon}_n}}(\beta, \rho) \end{pmatrix}$$

$Q_N$  is continuous differentiable at every point in  $(\tau, \theta)$

$$\nabla_{\theta\theta} Q_N(z_1, \dots, z_N, \tau_N, \theta_N) = \frac{1}{n} (\overline{\mathbf{A}} + \overline{\overline{\mathbf{A}}}),$$

$$\begin{aligned}\bar{\mathbf{A}} &= \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\rho)' \mathbf{S}'_n \\ \beta' \mathbf{X}'_N \bar{\mathbf{G}}_N(\rho)' \mathbf{S}'_n \end{pmatrix} (\mathbf{S}_n \mathbf{G}_N(\rho) \mathbf{X}_N \mathbf{S}_n \bar{\mathbf{G}}_N(\rho) \mathbf{X}_N \beta), \\ \bar{\bar{\mathbf{A}}} &= - \begin{pmatrix} \mathbf{0}_{k \times k} & \mathbf{X}'_N \bar{\mathbf{G}}_N(\rho) \mathbf{S}'_n \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho) \\ \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho)' \mathbf{S}_n \bar{\mathbf{G}}_N(\rho) \mathbf{X} & 2 \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho)' \mathbf{S}_n \bar{\mathbf{G}}_N(\rho) \mathbf{W}_N \mathbf{G}_N(\rho) \mathbf{X} \beta \end{pmatrix}, \\ Q_N &\text{ is twice continuous differentiable at every point in } (\tau, \theta) \text{ where} \\ \bar{\mathbf{G}}_N(\rho) &= \mathbf{G}_N(\rho) \mathbf{W}_N \mathbf{G}_N(\rho) \quad \blacksquare\end{aligned}$$

NLS- approx:

$$\text{Proof. } \nabla_{\theta'} Q_N(z_1, \dots, z_N, \tau_N, \theta_N) = \frac{-1}{n} \begin{pmatrix} \mathbf{X}'_N \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^{k'} \mathbf{S}'_n \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho) \\ \beta' \mathbf{X}'_N \sum_{k=1}^{m(n)} k \rho^{k-1} \mathbf{W}_N^{k'} \mathbf{S}'_n \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho)' \end{pmatrix}$$

$Q_N$  is continuous differentiable at every point in  $(\tau, \theta)$

$$\nabla_{\theta\theta} Q_N(z_1, \dots, z_N, \tau_N, \theta_N) = \frac{1}{n} \left( \tilde{\mathbf{A}} + \tilde{\tilde{\mathbf{A}}} \right),$$

$$\tilde{\mathbf{A}} = \begin{pmatrix} \sum_{k=0}^{m(n)} \rho^k \mathbf{H}' \\ \sum_{k=1}^{m(n)} k \rho^{k-1} \bar{\mathbf{H}}' \end{pmatrix} \begin{pmatrix} \sum_{k=0}^{m(n)} \rho^k \mathbf{H} & \sum_{k=1}^{m(n)} k \rho^{k-1} \bar{\mathbf{H}} \end{pmatrix},$$

$$\tilde{\tilde{\mathbf{A}}} = - \begin{pmatrix} \mathbf{0}_{k \times k} & \sum_{k=1}^{m(n)} k \rho^{k-1} \mathbf{H}' \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho) \\ \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho)' \sum_{k=1}^{m(n)} k \rho^{k-1} \mathbf{H} & \sum_{k=2}^{m(n)} k(k-1) \rho^k \bar{\mathbf{H}}' \bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho) \end{pmatrix}$$

$$\bar{\bar{\boldsymbol{\varepsilon}}}_n(\beta, \rho) = \bar{\mathbf{Y}}_n - \mathbf{S}_n \sum_{k=0}^{m(n)} \rho^k \mathbf{W}_N^k \mathbf{X}_N, \quad \mathbf{H} = \mathbf{S} \cdot \mathbf{W}^k \cdot \mathbf{X}, \quad \bar{\mathbf{H}} = \mathbf{H} \cdot \beta$$

$Q_N$  is twice continuous differentiable at every point in  $(\tau, \theta)$   $\blacksquare$

### 5.4.3 NLS- property 10

The estimators  $(\hat{\tau}_N, \hat{\theta}_N)$  take their values in  $T \times B$ , the true parameters  $(\tau_0, \theta_0)$  lie in the interior of  $T \times B$ ,

$$\begin{aligned}\hat{\theta}_N - \theta_0 &\text{ as } n \rightarrow \infty, \\ n^{1/2} \cdot (\hat{\tau}_N - \tau_0) &= O_p(1)\end{aligned}$$

**Proof.** Fulfilled due assumption (1)

NLS- Exact and NLS- approx fulfill:  $\hat{\theta}_N - \theta_0$  as  $n \rightarrow \infty$ , since  $\lim_{n \rightarrow \infty} m(n) = \infty$   $\blacksquare$

#### 5.4.4 NLS- property 11

The sequence  $\widehat{\theta}_N$  satisfies

$$n^{1/2} \cdot \nabla_{\theta'} \cdot Q_N(z_1, \dots, z_N, \widehat{\tau}_N, \widehat{\theta}_N) = o_p(1)$$

(I.e.  $\widehat{\theta}_N$  satisfies the normalized first order conditions up to an error of magnitude  $o_p(1)$ ).

**Proof.**  $p \lim_{n \rightarrow \infty} \frac{-1}{\sqrt{n}} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\widehat{\rho})' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \boldsymbol{\varepsilon}_N \\ \beta' \mathbf{X}'_N \overline{\mathbf{G}}_N(\widehat{\rho})' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \boldsymbol{\varepsilon}_N \end{pmatrix} \leq p \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \begin{pmatrix} a^2 \mathbf{X}'_N \boldsymbol{\varepsilon}_N \\ \beta_m a^2 \mathbf{X}'_N \boldsymbol{\varepsilon}_N \end{pmatrix} = o_p(1)$  ■

#### 5.4.5 NLS- property 12

For all sequences of random vectors  $(\widetilde{\tau}_N, \widetilde{\theta}_N)$  with  $\widetilde{\tau}_N \xrightarrow{p} \tau_0$  and  $\widetilde{\theta}_N \xrightarrow{p} \theta_0$  we have

$$\nabla_{\theta\theta} \cdot Q_N(z_1, \dots, z_N, \widehat{\tau}_N, \widehat{\theta}_N) \xrightarrow{p} A_0$$

as  $N \rightarrow \infty$ , where  $A_0$  is a real symmetric positive definite matrix:

**Proof.** Note that  $p \lim_{n \rightarrow \infty} \frac{1}{n} (\overline{\mathbf{A}} + \widetilde{\mathbf{A}}) = p \lim_{n \rightarrow \infty} \frac{1}{n} (\widetilde{\mathbf{A}} + \widetilde{\widetilde{\mathbf{A}}})$

$$\begin{aligned} & p \lim_{n \rightarrow \infty} \frac{1}{n} \overline{\mathbf{A}} \xrightarrow{p} \mathbf{0}_{k+1 \times k+1} \text{ for } \theta = \widetilde{\theta}_N, \text{ since } \overline{\boldsymbol{\varepsilon}}_n(\beta_0, \rho_0) = \mathbf{S}_n \mathbf{G}_N(\rho_0) \boldsymbol{\varepsilon} \text{ and} \\ & \widetilde{\theta}_N \xrightarrow{p} \theta_0 \\ & p \lim_{n \rightarrow \infty} \frac{1}{n} \overline{\mathbf{A}} - \frac{1}{n} \overline{\mathbf{A}}_0 = \\ & \frac{1}{n} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_{\Lambda N}(\widetilde{\rho}, \rho_0)' \mathbf{S}'_n \\ (\widetilde{\beta} - \beta_0)' \mathbf{X}'_N \overline{\mathbf{G}}'_{\Lambda N}(\widetilde{\rho}, \rho_0) \mathbf{S}'_n \end{pmatrix} (\mathbf{S}_n \mathbf{G}_{\Lambda N}(\widetilde{\rho}, \rho_0) \mathbf{X}_N \mathbf{S}_n \overline{\mathbf{G}}_{\Lambda N}(\widetilde{\rho}, \rho_0) \mathbf{X}_N \beta) \xrightarrow{p} \end{aligned}$$

$$\mathbf{0}_{k+1 \times k+1} \text{ where } \mathbf{G}_{\Lambda N}(\widetilde{\rho}, \rho_0) = \sum_{k=0}^{\infty} (\widetilde{\rho} - \rho_0)^k \mathbf{W}_N^k, \overline{\mathbf{G}}_{\Lambda N} = \sum_{k=1}^{\infty} k \cdot (\widetilde{\rho} - \rho_0)^{k-1} \mathbf{W}_N^k$$

since  $\widetilde{\theta}_N \xrightarrow{p} \theta_0$ .

$A_0$  is a real symmetric positive definite matrix since  $\mathbf{X}'_N \mathbf{G}_N(\rho_0)' \mathbf{S}'_n$  and  $\beta'_0 \mathbf{X}'_N \overline{\mathbf{G}}'_N(\rho_0) \mathbf{S}'_n$  are linear independent (for more details see Green page 835). ■

#### 5.4.6 NLS- property 13

For all sequences  $(\widetilde{\tau}_N, \widetilde{\theta}_N)$  as in NLS- property 12 we have

$$\nabla_{\theta\tau} \cdot Q_N(z_1, \dots, z_N, \widehat{\tau}_N, \widehat{\theta}_N) \xrightarrow{p} 0$$

**Proof.**  $p \lim_{N \rightarrow \infty} \frac{-1}{n} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\widehat{\rho})' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \boldsymbol{\varepsilon}_N \\ \beta' \mathbf{X}'_N \overline{\mathbf{G}}_N(\widehat{\rho})' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\widehat{\rho}) \boldsymbol{\varepsilon}_N \end{pmatrix} \xrightarrow{p} 0$  since  $\frac{1}{n} \|\mathbf{X}'_N \mathbf{G}_N(\widehat{\rho})' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\widehat{\rho})\| \leq x_m a^2 < \infty$  and  $\|\beta' \mathbf{X}'_N \overline{\mathbf{G}}_N(\widehat{\rho})' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\widehat{\rho})\| \leq \beta_m k x_m a^2$  ■

#### 5.4.7 NLS- property 14

There exists a real matrix  $D_0$  such that

$$-N^{1/2} \nabla_{\theta'} \cdot Q_N(z_1, \dots, z_N, \tau_0, \theta_0) \xrightarrow{p} D_0 \cdot \zeta_n + o_p(1)$$

where  $\zeta_n$  and  $\zeta$  are random vectors satisfying  $\zeta_n \xrightarrow{D} \zeta$

**Proof.** Obviously it is true that:  $p \lim_{N \rightarrow \infty} \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\rho_0)' \mathbf{S}'_n \overline{\boldsymbol{\varepsilon}}_n(\beta_0, \rho_0) \\ \beta' \mathbf{X}'_N \overline{\mathbf{G}}_N(\rho_0)' \mathbf{S}'_n \overline{\boldsymbol{\varepsilon}}_n(\beta_0, \rho_0) \end{pmatrix} - \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\rho_0)' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\rho_0) \boldsymbol{\varepsilon}_N \\ \beta' \mathbf{X}'_N \overline{\mathbf{G}}_N(\rho_0)' \mathbf{S}'_n \mathbf{S}_n \mathbf{G}_N(\rho_0) \boldsymbol{\varepsilon}_N \end{pmatrix} \xrightarrow{p} o_p(1)$ . Therefore,  $D_0 = \frac{-1}{n} \begin{pmatrix} \mathbf{X}'_N \mathbf{G}_N(\rho_0)' \mathbf{S}'_n \\ \beta' \mathbf{X}'_N \overline{\mathbf{G}}_N(\rho_0)' \mathbf{S}'_n \end{pmatrix}$  and  $n^{1/2} \zeta_n = n^{1/2} \mathbf{S}_n \mathbf{G}_N(\rho_0) \boldsymbol{\varepsilon}_N \xrightarrow{D} \zeta$  (????) ■