B. Technical appendix to "Efficient estimation of autoregression parameters and innovation distributions for semiparametric integer-valued AR(p) models"

This technical appendix contains the proofs of the results in "Efficient estimation of autoregression parameters and innovation distributions for semiparametric integer-valued AR(p) models". Proofs for the following results are gathered in this note: Lemma A.1, Theorem 2.1, Lemma A.2, and Lemma A.3. Let us briefly comment on these results and their proofs. Lemma A.1 contains auxiliary results, which are needed to prove the other results. Some parts are already known from the literature; the new parts are established by exploiting the V-uniform-ergodicity of an INAR process. Theorem 2.1 shows that the NPMLE is consistent. After a compactification of the parameter space, this consistency follows by Wald's method. Lemma A.2 is essential in establishing the limiting distribution of the NPMLE. The proof of this lemma is rather complicated and long. Finally, Lemma A.3 contains a LAN-result for parametric submodels of the semiparametric INAR(p) model. Apart from negligibility of the initial value, this result follows from the main theorem in Drost et al. (2006b). The negligibility is proved using Lemma A.1.D.

For notational convenience, δ_x denotes the Dirac measure concentrated in x. Also recall the notation $Z_t = (X_t, X_{t-1}, \ldots, X_{t-p})^T$ and $Y_t = (X_{t-1}, \ldots, X_{t-p})^T$.

B.1. Proof of Lemma A.1

Proof of (A): For the existence of the stationary distribution see Dion et al. (1995), Latour (1998), or Drost et al. (2007). The existence of moments follows from Drost et al. (2007).
Proof of (B): Notice first that $\nu_{\theta,G} \otimes P^{\theta,G}$ is the stationary distribution of Z, and that Z is an irreducible, aperiodic Markov chain on $\mathcal{Z} = \text{support}(\nu_{\theta,G} \otimes P^{\theta,G})$. Let Q^n denote the n-step transition-operator of Z (we drop the superscript θ, G). From well-known results on mixing-numbers for Markov chains (see, for example, Doukhan (1994) pages 87-89) it follows that it is sufficient to prove that there exists a function $A : \mathbb{Z}_+^{p+1} \to (0, \infty)$ such that $\int A \,\mathrm{d}(\nu_{\theta,G} \otimes P^{\theta,G}) < \infty$ and

$$\|Q^n(z, \cdot) - \nu_{\theta,G} \otimes P^{\theta,G}\|_{\mathrm{TV}} \le A(z)\rho^n, \quad z \in \mathcal{Z}, \tag{26}$$

for some $0 < \rho < 1$, where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variational norm of a signed measure. Recall (Meyn and Tweedie (1994) Chapter 16) that for Markov transition-probabilities P_1 and P_2 and a function $1 \le V < \infty$ the V-norm distance between P_1 and P_2 is defined by $|||P_1 - P_2|||_V = \sup_{z\in\mathcal{Z}} \|P_1(z, \cdot) - P_2(z, \cdot)\|_V / V(z)$, where, for a signed measured μ, $\|\mu\|_V = \sup_{f:\,|f|\le V} |\int f \,\mathrm{d}\mu|$. Introduce $V : \mathbb{Z}_+^{p+1} \to [1, \infty)$ by $V(z) = 1 + \sum_{i=1}^p a_i z_i$ with $a_i = \theta_i + \cdots + \theta_p$. Then it is straightforward to check (see also Drost et al. (2007)) that the following drift condition holds. There exists a constant $\beta > 0$ such that for all $z \in \mathcal{Z}$, except for some finite set, we have $\mathbb{E}_{\theta,G}[V(Z_t) \mid Z_{t-1} = z] - V(z) \le -\beta V(z)$. We conclude from Meyn and Tweedie (1994) Theorem 16.01 that there exist constants $\rho < 1$ and $\tilde{C} < \infty$ such that for all $n \in \mathbb{Z}_+$ $|||Q^n - \nu_{\theta,G} \otimes P^{\theta,G}|||_V \le \tilde{C}\rho^n$, i.e. Z is V-uniformly ergodic. Since $\mathbb{E}_{\nu_{\theta,G},\theta,G} V(Z_0) < \infty$ (by Lemma A.1A) and $V \ge 1$ (26) immediately follows, which concludes the proof of (A).
Proof of (C): this follows from Doukhan et al. (1995) Theorem 1 and Application 4. Take $r = 3/2$, notice that, using Markov's inequality and $\mathbb{E}_{\nu_{\theta,G},\theta,G} X_0^3 < \infty$ (by Lemma A.1A), the envelope belongs to $\Lambda_3(P) = \Lambda_{x\sqrt{x}}(P)$. Next, take $b > 3/2$ such that $b \ge 3/a$, and note

that there exists $C>0$ such that $n^{-b}\geq C\rho^n$ for all $n\geq 1$.
Proof of (D): Notice first that $\nu_{\theta,G}$ is the stationary distribution of Y, and that Y is an irreducible, aperiodic Markov chain on $\mathbb{Z}_+^p$. Let $Q^{\theta,G}$ denote the transition-probabilities of Y and Q^n denotes the n-step transition-operator of Y (we drop the superscripts θ,G for the n step operator, since we only consider this operator at (θ,G)). Following the proof of (B) it follows that the Markov chain Y on $\mathbb{Z}_+^p$ is V-uniformly ergodic for $V(Y_t)=1+\sum_{i=1}^p a_iX_{t-i}$, $a_i=\theta_i+\cdots+\theta_p$, i.e. there exist constants $C>0$ and $0<\rho<1$ such that $|||Q^n-\nu_{\theta,G}|||_V\leq C\rho^n$ for all $n\in\mathbb{Z}_+$. Since Y is uniformly ergodic in the norm $|||\cdot|||_V$, an application of Kartashov (1985) Theorem B (it is easy to see that $|||\cdot|||_V$ satisfies the conditions) yields that Y is strongly stable in this norm: each transition-probability Q' in some neighborhood of Q has a unique stationary measure $\nu(Q')$ and $|||Q'_n-Q|||_V\to 0$ implies $\|\nu(Q'_n)-\nu_{\theta,G}\|_V\to 0$. This yields (D). $\square$

B.2. Proof of consistency Theorem 2.1

Let $(\hat\theta_n,\hat G_n)$ be a nonparametric maximum likelihood estimator of (θ,G). It suffices to prove $\hat\theta_n\xrightarrow{p}\theta_0$ and $\hat g_n(k)\xrightarrow{p}g_0(k)$ for all $k\in\mathbb{Z}_+$. We prove this pointwise convergence by an application of Wald's consistency proof. To that end, we first compactify the parameter space, starting with $\mathcal{G}$.

Introduce $\bar{\mathcal{G}}$: the class of all probability distributions on $\mathbb{Z}_+\cup\{\infty\}$. Identify each $G\in\bar{\mathcal{G}}$ with the sequence $(g(k))_{k\in\mathbb{Z}_+}$. Notice that this correspondence is 1-to-1, since $g(\infty)=1-\sum_{k=0}^\infty g(k)$. As a result, $\bar{\mathcal{G}}$ is a subset of $[0,1]^{\mathbb{Z}_+}$ equipped with the norm $\|a\|=\sum_{k=0}^\infty 2^{-k}|a(k)|$, that is, we endow $[0,1]^{\mathbb{Z}_+}$ with the product topology. Notice that a sequence in $[0,1]^{\mathbb{Z}_+}$ converges if and only if all coordinates, which are sequences in $[0,1]$, converge. Using Helly's lemma (see, for example, Van der Vaart (2000) Lemma 1.5) it is an easy exercise to show that $\bar{\mathcal{G}}$ is a compact subset of $[0,1]^{\mathbb{Z}_+}$. For $G\in\bar{\mathcal{G}}$ define $P_{x,\infty}^{\theta,G}=1-\sum_{j\in\mathbb{Z}_+}P_{x,j}^{\theta,G}=g(\infty)$ for $x\in\mathbb{Z}_+^p$ and $P_{x,\infty}^{\theta,G}=1$ if $\max_{i=1}^p x_i=\infty$.

Now, consider the parameter θ as well. Define $\bar E=[0,1]^p\times\bar{\mathcal{G}}$, and equip $\bar E$ with the "sum-distance" $d((\theta,G),(\theta',G'))=|\theta-\theta'|+\|(g(k))_{k\in\mathbb{Z}_+}-(g'(k))_{k\in\mathbb{Z}_+}\|$. $\bar E$ is the product of two compact spaces and, hence, itself compact.

Define

$$m^{\theta,G}(x_{-p},\ldots,x_0)=\log P^{\theta,G}_{(x_{-1},\ldots,x_{-p}),x_0},$$

and the (random) function $M_n:\bar E\to[-\infty,\infty)$ by

$$M_n(\theta,G)=\frac{1}{n}\sum_{t=0}^n m^{\theta,G}(X_{t-p},\ldots,X_t).$$

From an appropriate law of large number for Markov chains, we find that M_n converges in probability to a (nonrandom) function $M:\bar E\to[-\infty,\infty)$ defined by

$$M(\theta,G)=\mathbb{E}_{\nu_{\theta_0,G_0},\theta_0,G_0}m^{\theta,G}(X_{-p},\ldots,X_0).$$

Note that, by Lemma A.1A, the stationary distribution ν_{θ_0,G_0} indeed exists.

The following holds.

(A) For fixed $x_{-p}, \dots, x_0 \in \mathbb{Z}_+$, the map $\bar{E} \ni (\theta, G) \mapsto m^{\theta,G}(x_{-p}, \dots, x_0)$ is continuous. This is easy to see, since there appear only a finite number of $g(j)$'s in $P^{\theta,G}_{(x_{-1},\dots,x_{-p}),x_0}$.

(B) For all $x_{-p}, \dots, x_0 \in \mathbb{Z}_+$ we have $m^{\theta,G}(x_{-p}, \dots, x_0) \le \log(1) = 0$.

(C) The map $\bar{E} \ni (\theta, G) \mapsto M(\theta, G)$ has a unique maximum at (θ_0, G_0). Since we have the identification $P^{\theta,G}_{(X_{-1},\dots,X_{-p}),X_0} = P^{\theta_0,G_0}_{(X_{-1},\dots,X_{-p}),X_0}$ $\mathbb{P}_{\nu_{\theta_0,G_0},\theta_0,G_0}$-a.s. $\implies (\theta, G) = (\theta_0, G_0)$, this easily follows using the following well-known argument (use $\log x \le 2(\sqrt{x} - 1)$ for $x \ge 0$):

$$\begin{aligned} M(\theta, G) - M(\theta_0, G_0) &\le 2\mathbb{E}_{\nu_{\theta_0,G_0},\theta_0,G_0}\left(\sqrt{\frac{P^{\theta,G}_{(X_{-1},\dots,X_{-p}),X_0}}{P^{\theta_0,G_0}_{(X_{-1},\dots,X_{-p}),X_0}}} - 1\right) \\ &= 2\sum_{y\in\mathbb{Z}^p_+} \nu_{\theta_0,G_0}\{y\} \sum_{x_0=0}^{\infty} \sqrt{P^{\theta,G}_{y,x_0} P^{\theta_0,G_0}_{y,x_0}} - 2 \\ &\le -\sum_{y\in\mathbb{Z}^p_+} \nu_{\theta_0,G_0}\{y\} \sum_{x_0=0}^{\infty} \left(\sqrt{P^{\theta,G}_{y,x_0}} - \sqrt{P^{\theta_0,G_0}_{y,x_0}}\right)^2 \le 0. \end{aligned}$$

(D) $M_n(\hat{\theta}_n, \hat{G}_n) \ge M_n(\theta_0, G_0)$, since $(\hat{\theta}_n, \hat{G}_n)$ maximizes the likelihood.

Hence all conditions to Wald's consistency theorem hold (see, for example, the proof of Theorem 5.14 in Van der Vaart (2000), where the law of large numbers for the i.i.d. case has to be replaced by the result above. thus we obtain $d((\hat{\theta}_n, \hat{G}_n), (\theta_0, G_0)) \xrightarrow{p} 0$, which immediately yields $\hat{\theta}_n \xrightarrow{p} \theta_0$ and, for all $k \in \mathbb{Z}_+$, $\hat{g}_n(k) \xrightarrow{p} g_0(k)$. □

B.3. Proof of Lemma A.2

Throughout ν_0 is shorthand for ν_{θ_0,G_0}. If no confusion can arise, sub- and superscripts are sometimes dropped for notational convenience.

B.3.1. Proof of (L1)

To enhance readability the proof is decomposed in three steps. In the first step we show that $\dot{\Psi}$ is indeed linear and continuous. And in the second and third step we prove the Fréchet-differentiability of Ψ_1 and Ψ_2 respectively.

Step 1:

The linearity of $\dot{\Psi}$ is obvious. For the continuity, note that it suffices to prove that both $\dot{\Psi}_1$ and $\dot{\Psi}_2$ are continuous. We consider $\dot{\Psi}_1$ which is the sum of $\dot{\Psi}_{11}$ and $\dot{\Psi}_{12}$; the continuity of $\dot{\Psi}_2$ proceeds in the same way. Of course, $\dot{\Psi}_{11}$ is continuous. So the only thing left is to show that $\dot{\Psi}_{12}$ is continuous. It is easy to see that, here $\dot{\ell}_{\theta,i}$ refers to the ith coordinate of the p-vector $\dot{\ell}_\theta$,

$$\left|\dot{\ell}_{\theta,i}(x_{-p}, \dots, x_0; \theta, G)\right| \le \frac{x_{-i}}{\theta_i(1 - \theta_i)}, \tag{27}$$

which yields, using that ε_0 and X_{-i} are independent,

$$\left|\mathbb{E}_{\nu_0,\theta_0}\left[\dot{\ell}_{\theta,i}(X_{-p}, \dots, X_0; \theta, G) \mid \varepsilon_0\right]\right| \le \frac{\mathbb{E}_{\nu_0} X_{-i}}{\theta_i(1 - \theta_i)}.$$

Thus the map

$$\mathbb{Z}_+ \ni e \mapsto \left|\mathbb{E}_{\nu_0,\theta_0}\left[\dot{\ell}_\theta(x_{-p},\ldots,x_0;\theta,G) \mid \varepsilon_0 = e\right]\right|$$

is bounded, say by C. This yields, for $H, G \in \operatorname{lin}\mathcal{G}$,

$$\begin{aligned}|\dot{\Psi}_{12}(G-H)| &= \left|\int \mathbb{E}_{\nu_0,\theta_0}\left[\dot{\ell}_\theta(X_{-p},\ldots,X_0;\theta_0,G_0) \mid \varepsilon_0 = e\right] \mathrm{d}(H-G)(e)\right| \\ &\le C\sum_{e=0}^{\infty}|h(e)-g(e)| = C\|H-G\|_1,\end{aligned}$$

which yields the continuity of $\dot{\Psi}_{12}$.

Step 2:
Rewrite,

$$\begin{aligned}\Psi_1(\theta,G) - \Psi_1(\theta_0,G_0) - \dot{\Psi}_{11}(\theta-\theta_0) - \dot{\Psi}_{12}(G-G_0) &= \Psi_1(\theta,G) - \Psi_1(\theta_0,G) - \dot{\Psi}_{11}(\theta-\theta_0) \\ &\quad + \Psi_1(\theta_0,G) - \Psi_1(\theta_0,G_0) - \dot{\Psi}_{12}(G-G_0).\end{aligned}$$

Let θ_n be a sequence in $[0,1]^p$ converging to θ_0 and G_n a sequence in $\mathcal{G}$ converging to G_0. In Step 2a we show that

$$\frac{\left|\Psi_1(\theta_n,G_n) - \Psi_1(\theta_0,G_n) - \dot{\Psi}_{11}(\theta_n-\theta_0)\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1} \to 0, \tag{28}$$

and in Step 2b we show that

$$\frac{\left|\Psi_1(\theta_0,G_n) - \Psi_1(\theta_0,G_0) - \dot{\Psi}_{12}(G_n-G_0)\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1} \to 0, \tag{29}$$

which will conclude the proof of Step 2.
Step 2a:
First we recall from Drost et al. (2006b) that the usual information-identity holds, i.e.

$$I_\theta(\theta_0,G_0) = \mathbb{E}_{\nu_0,\theta_0,G_0}\dot{\ell}_\theta\dot{\ell}_\theta^T(X_{-p},\ldots,X_0;\theta_0,G_0) = -\mathbb{E}_{\nu_0,\theta_0,G_0}\frac{\partial}{\partial\theta^T}\dot{\ell}_\theta(X_{-p},\ldots,X_0;\theta_0,G_0).$$

From the mean-value theorem we obtain, for $i = 1,\ldots,p$,

$$\begin{aligned}&\dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\theta,G) - \dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\theta_0,G) \\ &\qquad = \frac{\partial}{\partial\theta^T}\dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\tilde{\theta}_i(\theta,G),G)(\theta-\theta_0),\end{aligned}$$

where $\tilde{\theta}_i(\theta,G) = \tilde{\theta}_i(X_{-p},\ldots,X_0;\theta,G,\theta_0)$ is a point on the line segment between θ and θ_0. Let $J(X_{-p},\ldots,X_0;\theta,G)$ be the $p\times p$ random matrix given by

$$J(X_{-p},\ldots,X_0;\theta,G) = \begin{pmatrix}\frac{\partial}{\partial\theta^T}\dot{\ell}_{\theta,1}(X_{-p},\ldots,X_0;\tilde{\theta}_1(\theta,G),G) \\ \vdots \\ \frac{\partial}{\partial\theta^T}\dot{\ell}_{\theta,p}(X_{-p},\ldots,X_0;\tilde{\theta}_p(\theta,G),G)\end{pmatrix}.$$

It is easy to see, since we only have to deal with a finite number of $g(k)$'s, that we have for fixed $x_{-p}, \ldots, x_0$, $J(x_{-p}, \ldots, x_0; \theta_n, G_n) \to (\partial/\partial\theta^T)\dot{\ell}_\theta(X_{-p}, \ldots, X_0; \theta_0, G_0)$. From Drost et al. (2006b) we have,

$$\left|\frac{\partial}{\partial\theta_j}\dot{\ell}_{\theta,i}(x_{-p}, \ldots, x_0; \theta, G)\right| \leq \frac{3}{2\theta_i(1-\theta_i)\theta_j(1-\theta_j)}(X_{-i}^2 + X_{-j}^2),$$

which is $\mathbb{P}_{\nu_0,\theta_0,G_0}$-integrable. Thus, using dominated convergence, we obtain

$$\begin{aligned}
&\frac{\left|\Psi_1(\theta_n, G_n) - \Psi_1(\theta_0, G_n) - \dot{\Psi}_{11}(\theta_n - \theta_0)\right|}{|\theta_n - \theta_0|} \\
&\qquad \leq \frac{\mathbb{E}_{\nu_0,\theta_0,G_0}\left|\left(I_\theta(\theta_0, G_0) + J(Z_0; \theta_n, G_n)\right)(\theta_n - \theta_0)\right|}{|\theta_n - \theta_0|} \to 0,
\end{aligned}$$

which yields (28).
Step 2b:
We have, using that $\mathbb{E}_{\nu_0,\theta_0,G}\left[\cdot \mid \varepsilon_0\right]$ does not depend on G,

$$\begin{aligned}
&\Psi_1(\theta_0, G) - \Psi_1(\theta_0, G_0) - \dot{\Psi}_{12}(G - G_0) \\
&\quad = \mathbb{E}_{\nu_0,\theta_0,G_0}\dot{\ell}_\theta(X_{-p}, \ldots, X_0; \theta_0, G) + \mathbb{E}_G\mathbb{E}_{\nu_0,\theta_0}\left[\dot{\ell}_\theta(X_{-p}, \ldots, X_0; \theta_0, G_0) \mid \varepsilon_0\right] \\
&\quad = \mathbb{E}_{\nu_0,\theta_0,G_0}\dot{\ell}_\theta(X_{-p}, \ldots, X_0; \theta_0, G) + \mathbb{E}_{\nu_0,\theta_0,G}\dot{\ell}_\theta(X_{-p}, \ldots, X_0; \theta_0, G_0) \\
&\quad = \mathbb{E}_{\nu_0} f(X_{-p}, \ldots, X_{-1}; G),
\end{aligned}$$

where (using that $\mathbb{E}_{\nu_0,\theta_0,H}\left[\dot{\ell}_\theta(X_{-p}, \ldots, X_0; \theta_0; H) \mid X_{-1}, \ldots, X_{-p}\right] = 0$ for $H \in \mathcal{G}$)

$$\begin{aligned}
&f(X_{-p}, \ldots, X_{-1}; G) \\
&\quad = \sum_{x_0=0}^{\infty}\left(P_{Y_0,x_0}^{\theta_0,G} - P_{Y_0,x_0}^{\theta_0,G_0}\right)\left(\dot{\ell}_\theta(Y_0, x_0; \theta_0, G_0) - \dot{\ell}_\theta(Y_0, x_0; \theta_0, G)\right) \\
&\quad = \sum_{x_0=0}^{\infty}\sum_{k=0}^{\infty}(g(k) - g_0(k))\left(*_{i=1}^p \mathrm{Bin}_{X_{-i},\theta_{0,i}}\right)\{x_0 - k\}\left(\dot{\ell}_\theta(Y_0, x_0; \theta_0, G_0) - \dot{\ell}_\theta(Y_0, x_0; \theta_0, G)\right) \\
&\quad = \sum_{k=0}^{\infty}(g(k) - g_0(k))\sum_{x_0=0}^{\infty}\left(*_{i=1}^p \mathrm{Bin}_{X_{-i},\theta_{0,i}}\right)\{x_0 - k\}\left(\dot{\ell}_\theta(Y_0, x_0; \theta_0, G_0) - \dot{\ell}_\theta(Y_0, x_0; \theta_0, G)\right).
\end{aligned}$$

From this we obtain the bound,

$$|f(X_{-1}, \ldots, X_{-p}; G)| \leq \|G - G_0\|_1 \sum_{x_0=0}^{X_{-1}+\cdots+X_{-p}}\left|\dot{\ell}_\theta(Y_0, x_0; \theta_0, G_0) - \dot{\ell}_\theta(Y_0, x_0; \theta_0, G)\right|.$$

Since G_n is a sequence in $\mathcal{G}$ converging to G_0, we obtain, for fixed $x_{-p}, \ldots, x_{-1}$,

$$\sum_{x_0=0}^{x_{-1}+\cdots+x_{-p}}\left|\dot{\ell}_\theta(x_{-p}, \ldots, x_{-1}, x_0; \theta_0, G_0) - \dot{\ell}_\theta(x_{-p}, \ldots, x_{-1}, x_0; \theta_0, G_n)\right| \to 0.$$

Furthermore, using (27),

$$\sum_{x_0=0}^{x_{-1}+\cdots+x_{-p}} \left|\dot{\ell}_\theta(X_{-p},\ldots,X_{-1},x_0;\theta_0,G_0) - \dot{\ell}_\theta(X_{-p},\ldots,X_{-1},x_0;\theta_0,G_n)\right|$$
$$\leq 2\sum_{j=1}^{p} \frac{X_{-j}}{\theta_{0,j}(1-\theta_{0,j})}.$$

Thus $f(X_{-p},\ldots,X_{-1};G_n)/\|G_n - G_0\|_1$ converges $\mathbb{P}_{\nu_0}$-a.s. to 0, and is dominated by a $\mathbb{P}_{\nu_0}$-integrable function. An application of the dominated convergence theorem yields (29).

Step 3:
Rewrite,

$$\Psi_2(\theta,G) - \Psi_2(\theta_0,G_0) - \dot{\Psi}_{21}(\theta-\theta_0) - \dot{\Psi}_{22}(G-G_0) = \Psi_2(\theta,G) - \Psi_2(\theta_0,G) - \dot{\Psi}_{21}(\theta-\theta_0)$$
$$+ \Psi_2(\theta_0,G) - \Psi_2(\theta_0,G_0) - \dot{\Psi}_{22}(G-G_0).$$

Let θ_n be a sequence in $[0,1]^p$ converging to θ_0 and G_n a sequence in $\mathcal{G}$ converging to G_0. We will verify that

$$\frac{\sup_{h\in\mathcal{H}_1}\left|\Psi_2(\theta_n,G_n)h - \Psi_2(\theta_0,G_n)h - \dot{\Psi}_{21}(\theta_n-\theta_0)h\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1} \to 0, \tag{30}$$

and,

$$\frac{\sup_{h\in\mathcal{H}_1}\left|\Psi_2(\theta_0,G_n)h - \Psi_2(\theta_0,G_0)h - \dot{\Psi}_{22}(G_n-G_0)h\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1} \to 0, \tag{31}$$

which will conclude the proof.
Step 3a:
First note that

$$\Psi_2(\theta_n,G_n)h - \Psi_2(\theta_0,G_n)h - \dot{\Psi}_{21}(\theta_n-\theta_0)h$$
$$= \mathbb{E}_{\nu_0,\theta_0,G_0}\left(A_{\theta_n,G_n}h(Z_0) - A_{\theta_0,G_n}h(Z_0) + A_{\theta_0,G_0}h(Z_0)\dot{\ell}_\theta^T(Z_0;\theta_0,G_0)(\theta_n-\theta_0)\right).$$

It is straightforward to check that, for $i=1,\ldots,p$,

$$\frac{\partial}{\partial\theta_i}A_{\theta,G}h(X_{-p},\ldots,X_0) = \mathbb{E}_{\theta,G}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\theta_i\circ X_{-i}) \mid X_0,\ldots,X_{-p}\right]$$
$$- A_{\theta,G}h(X_{-p},\ldots,X_0)\dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\theta,G),$$

and for $i,j=1,\ldots,p$,

$$\frac{\partial^2}{\partial\theta_j\partial\theta_i}A_{\theta,G}h(X_{-p},\ldots,X_0) = \mathbb{E}_{\theta,G}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\theta_i\circ X_{-i})\dot{s}_{X_{-j},\theta_j}(\theta_j\circ X_{-j})|X_0,\ldots,X_{-p}\right]$$
$$- \mathbb{E}_{\theta,G}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\theta_i\circ X_{-i})|X_0,\ldots,X_{-p}\right]\dot{\ell}_{\theta,j}(X_{-p},\ldots,X_0;\theta,G)$$

$$
\begin{aligned}
&- A_{\theta,G}h(X_{-p},\ldots,X_0)\ddot{\ell}_{\theta,ij}(X_{-p},\ldots,X_0;\theta,G)\\
&- \dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\theta,G)\frac{\partial}{\partial\theta_j}A_{\theta,G}h(X_{-p},\ldots,X_0)\\
&+ 1\{i=j\}\mathbb{E}_{\theta,G}\left[h(\varepsilon_0)\ddot{s}_{X_{-i},\theta_i}(\theta_i\circ X_{-i}) \mid X_0,\ldots,X_{-p}\right],
\end{aligned}
$$

where $\ddot{s}_{n,\alpha}(k) = (\partial/\partial\alpha)\dot{s}_{n,\alpha}(k)$. Now it is easy, but a bit tedious, to see that there exists a constant $C_\theta > 0$, which is bounded in θ in a neighborhood of θ_0 and not depending on h, such that, for $i,j = 1,\ldots,p$,

$$
\left|\frac{\partial}{\partial\theta_i}A_{\theta,G}h(X_{-p},\ldots,X_0)\right| + \left|\frac{\partial^2}{\partial\theta_j\partial\theta_i}A_{\theta,G}h(X_{-p},\ldots,X_0)\right| \leq C_\theta(X_{-i}^2 + X_{-j}^2). \tag{32}
$$

A second order Taylor expansion in θ yields

$$
\begin{aligned}
&A_{\theta_n,G_n}h(Z_0) - A_{\theta_0,G_n}h(Z_0) + A_{\theta_0,G_n}h(Z_0)\dot{\ell}_\theta^T(Z_0;\theta_0,G_n)(\theta_n-\theta_0)\\
&\quad= \sum_{i=1}^{p}(\theta_{n,i}-\theta_{0,i})\mathbb{E}_{\theta_0,G_n}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_{0,i}}(\theta_i\circ X_{-i}) \mid X_0,\ldots,X_{-p}\right]\\
&\qquad+ \frac{1}{2}(\theta_n-\theta_0)^T\frac{\partial^2}{\partial\theta\partial\theta^T}A_{\tilde{\theta}_n,G_n}h(X_{-p},\ldots,X_0)(\theta_n-\theta_0),
\end{aligned}
$$

where $\tilde{\theta}_n$ is a random point on the line segment between θ_0 and θ_n (also depending on h, Z_0, and G_n). Using (32) it easily follows, using dominated convergence, that

$$
\sup_{h\in\mathcal{H}_1}\frac{\left|\mathbb{E}_{\nu_0,\theta_0,G_0}(\theta_n-\theta_0)^T\frac{\partial^2}{\partial\theta\partial\theta^T}A_{\tilde{\theta}_n,G_n}h(X_{-p},\ldots,X_0)(\theta_n-\theta_0)\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1} \to 0.
$$

Hence we obtain (30) once we show that

$$
\begin{aligned}
&\sup_{h\in\mathcal{H}_1}\frac{\left|\sum_{i=1}^{p}(\theta_{n,i}-\theta_{0,i})\mathbb{E}_{\nu_0,\theta_0,G_0}\mathbb{E}_{\theta_0,G_n}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_{0,i}}(\theta_i\circ X_{-i}) \mid X_0,\ldots,X_{-p}\right]\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1}\\
&\qquad\to 0,
\end{aligned}
\tag{33}
$$

and,

$$
\begin{aligned}
&\sup_{h\in\mathcal{H}_1}\frac{\left|\mathbb{E}_{\nu_0,\theta_0,G_0}\left(A_{\theta_0,G_n}h(Z_0)\dot{\ell}_\theta^T(Z_0;\theta_0,G_n)(\theta_n-\theta_0) - A_{\theta_0,G_0}h(Z_0)\dot{\ell}_\theta^T(Z_0;\theta_0,G_0)(\theta_n-\theta_0)\right)\right|}{|\theta_n-\theta_0| + \|G_n-G_0\|_1}\\
&\qquad\to 0
\end{aligned}
\tag{34}
$$

both hold. It is easy to see that we have, for $i = 1,\ldots,p$,

$$
\begin{aligned}
&\left|\mathbb{E}_{\theta_0,G_n}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_{0,i}}(\theta_i\circ X_{-i}) \mid Z_0\right] - \mathbb{E}_{\theta_0,G_0}\left[h(\varepsilon_0)\dot{s}_{X_{-i},\theta_{0,i}}(\theta_i\circ X_{-i}) \mid Z_0\right]\right|\\
&\quad\leq \left|\frac{P^{\theta_0,G_0}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\theta_0,G_n}_{(X_{-1},\ldots,X_{-p}),X_0}} - 1\right|\frac{X_{-i}}{\theta_{0,i}(1-\theta_{0,i})}\\
&\qquad+ \frac{\sum_{e=0}^{X_0}\sum_{k=0}^{X_{-i}}|g_n(e)-g_0(e)|\left(*_{j\neq i}\mathrm{Bin}_{X_{-j},\theta_{0,j}}\right)\{X_0-k-e\}\dot{s}_{X_{-i},\theta_{0,i}}(k)\,\mathrm{b}_{X_{-i},\theta_{0,i}}(k)}{P^{\theta_0,G_n}_{(X_{-1},\ldots,X_{-p}),X_0}},
\end{aligned}
$$

which for fixed $X_{-p},\dots,X_0$ converges to 0. Note that the left-hand-side of this display is bounded by the ν_0-integrable variable $2X_{-i}/(\theta_{0,i}(1-\theta_{0,i}))$. Since $\mathbb{E}_{\nu_0,\theta_0,G_0} h(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\theta_i \circ X_{-i}) = 0$, by independence of ε_0 and $\theta_i \circ X_{-i} - \theta_{0,i}X_{-i}$, Display (33) easily follows using dominated convergence. In a similar fashion we obtain (34).
Step 3b:
Note first that we have

$$\begin{aligned}\Psi_2(\theta_0,G_n)h &- \Psi_2(\theta_0,G_0)h - \dot{\Psi}_{22}(G_n-G_0)h \\ &= \mathbb{E}_{\nu_0,\theta_0,G_0} A_{\theta_0,G_n}h(Z_0) - \int h\,\mathrm{d}G_n + \mathbb{E}_{\nu_0,\theta_0,G_n} A_{\theta_0,G_0}h(Z_0) - \int h\,\mathrm{d}G_0.\end{aligned}$$

It now follows that we have

$$\Psi_2(\theta_0,G_n)h - \Psi_2(\theta_0,G_0)h - \dot{\Psi}_{22}(G_n-G_0)h = \mathbb{E}_{\nu_0} f^h(X_{-p},\dots,X_{-1};G_n),$$

where

$$f^h(X_{-p},\dots,X_{-1};G_n) = \sum_{x_0=0}^{\infty}\left(P_{Y_0,x_0}^{\theta_0,G_n} - P_{Y_0,x_0}^{\theta_0,G_0}\right)\left(A_{\theta_0,G_0}h(Y_0,x_0) - A_{\theta_0,G_n}h(Y_0,x_0)\right).$$

Proceeding as in Step 2b we obtain the bound

$$|f^h(X_{-p},\dots,X_{-1};G_n)| \le \|G_n-G_0\|_1 \sum_{x_0=0}^{X_{-p}+\cdots+X_{-1}} \left|A_{\theta_0,G_0}h(Y_0,x_0) - A_{\theta_0,G_n}h(Y_0,x_0)\right|.$$

Using that, for $x_0 \in \{0,\dots,X_{-p}+\cdots+X_{-1}\}$,

$$\begin{aligned}\sup_{h\in\mathcal{H}_1} &\left|A_{\theta_0,G_n}h(X_{-p},\dots,X_{-1},x_0) - A_{\theta_0,G_0}h(X_{-p},\dots,X_{-1},x_0)\right| \\ &\le \left|\frac{P^{\theta_0,G_0}_{(X_{-p},\dots,X_{-1}),x_0}}{P^{\theta_0,G_n}_{(X_{-p},\dots,X_{-1}),x_0}} - 1\right| \left|A_{\theta_0,G_0}h(X_{-p},\dots,X_{-1},x_0)\right| \\ &\quad + \frac{\sum_{e=0}^{x_0}|g_n(e)-g_0(e)|\left(*_{i=1}^p \mathrm{Bin}_{X_{-i},\theta_{0,i}}\right)\{x_0-e\}}{P^{\theta_0,G_n}_{(X_{-p},\dots,X_{-1}),x_0}},\end{aligned}$$

we see that for fixed $(X_{-p},\dots,X_{-1})$ $\sup_{h\in\mathcal{H}_1}|f^h(X_{-p},\dots,X_{-1};G_n)|/\|G_n-G_0\|_1 \to 0$. Since $2(X_{-p}+\cdots+X_{-1})$ is an ν_0-integrable envelope for $\sup_{h\in\mathcal{H}_1}|f^h(X_{-p},\dots,X_{-1};G_n)|/\|G_n-G_0\|_1$, an application of dominated convergence yields (31).

B.3.2. *Proof of (L2)*

First we prove (L2) for the case $\mathrm{support}(G_0) = \mathbb{Z}_+$. To enhance readability we decompose the proof into the following steps.

(1) In this step we show that we can rewrite some parts of the derivative $\dot{\Psi}$ as follows,

$$\dot{\Psi}_{12}(G-G_0) = -\int A_0^*\dot{\ell}_\theta(e)\,\mathrm{d}(G-G_0)(e), \tag{35}$$

$$\dot{\Psi}_{22}(G-G_0)h = -\int A_0^* A_0 h(e)\,\mathrm{d}(G-G_0)(e), \quad h \in \mathcal{H}_1, \tag{36}$$

where A_0^* is the L_2-adjoint of $A_0 = A_{\theta_0,G_0}$. This representation allows us to invoke results from Hilbert space theory.

(2) This step shows that to prove that $\dot{\Psi}$ has a continuous inverse, it suffices to prove that a certain operator from $\ell^\infty(\mathbb{Z}_+)$ into itself is onto and continuously invertible.

(3) This step shows that the operator from Step 2 is indeed onto and continuously invertible.

Step 1:
Let $[\varepsilon]$ denote $\{f(\varepsilon_0) \mid f : \mathbb{Z}_+ \to \mathbb{R}, \mathbb{E}_{G_0} f^2(\varepsilon_0) < \infty\}$ equipped with the $\mathrm{L}_2(G_0)$ norm and let $[X]$ denote $\{f(X_{-p}, \ldots, X_0) \mid f : \mathbb{Z}_+^{p+1} \to \mathbb{R}, \mathbb{E}_{\nu_0,\theta_0,G_0} f^2(X_{-p}, \ldots, X_0) < \infty\}$ equipped with the $\mathrm{L}_2(\nu_0 \otimes P^{\theta_0,G_0})$ norm. It is not hard to see that both these spaces are, in fact, Hilbert spaces (that these spaces are already in their "a.s.-equivalence class form", follows from support$(G_0) = \mathbb{Z}_+$). We view upon A_0 as an operator from $[\varepsilon]$ into $[X]$. From the definition it is easy to see that A_0 is linear and continuous. Since A_0 is a continuous linear map between two Hilbert spaces, it has an adjoint map $A_0^* : [X] \to [\varepsilon]$ (which is a continuous linear map that satisfies and is uniquely determined by the equations $< A_0^* h_2, h_1 >_{[\varepsilon]} = < h_2, A_0 h_1 >_{[X]}$ for $h_1 \in [\varepsilon]$, $h_2 \in [X]$) given by

$$A_0^* f = A_0^* f(\varepsilon_0) = \mathbb{E}_{\nu_0,\theta_0}[f(X_{-p}, \ldots, X_0) \mid \varepsilon_0].$$

Now, invoking the definitions of $\dot{\Psi}_{12}$ and $\dot{\Psi}_{22}$, (35) and (36) are immediate.

Step 2:
To prove that $\dot{\Psi}$ is continuously invertible, it suffices to prove that $\dot{\Psi}_{11} : \mathbb{R}^p \to \mathbb{R}^p$ and $\dot{V} = \dot{\Psi}_{22} - \dot{\Psi}_{21}\dot{\Psi}_{11}^{-1}\dot{\Psi}_{12} : \mathrm{lin}\,\mathcal{G} \to \ell^\infty(\mathcal{H}_1)$ are both continuously invertible. The invertibility of $\dot{\Psi}_{11}$ is immediate, since the $p \times p$ Fisher information-matrix $I_{\theta_0} = \mathbb{E}_{\nu_0,\theta_0,G_0} \dot{\ell}_\theta \dot{\ell}_\theta^T(Z_0; \theta_0, G_0)$ is invertible (see Drost et al. (2006b)). To prove that $\dot{V}$ is continuously invertible is much harder. In this step, we will give an easier sufficient condition which is proved to hold true in Step 3. Introduce the operator $C : \mathcal{H}_1 \to [\varepsilon]$ by

$$Ch(e) = -\left[\mathbb{E}_{\nu_0,\theta_0,G_0} A_0 h(X_{-p}, \ldots, X_0)\dot{\ell}_\theta^T(X_{-p}, \ldots, X_0; \theta_0, G_0)\right] I_{\theta_0}^{-1}(A_0^*(\dot{\ell}_\theta(\cdot; \theta_0, G_0)))(e),$$

for $e \in \mathbb{Z}_+$, where $A_0^*(\dot{\ell}_\theta(\cdot; \theta_0, G_0)) = (A_0^*(\dot{\ell}_{\theta,1}(\cdot; \theta_0, G_0)), \ldots, A_0^*(\dot{\ell}_{\theta,p}(\cdot; \theta_0, G_0)))^T \in [\varepsilon]^p$. Then $\dot{V}$ can be rewritten as

$$\dot{V}(G-G_0)h = -\int \left(A_0^* A_0 h + Ch\right)(e)\,\mathrm{d}(G-G_0)(e), \quad h \in \mathcal{H}_1.$$

The mapping $\dot{V} : \mathrm{lin}\,\mathcal{G} \to \ell^\infty(\mathcal{H}_1)$ has a continuous inverse on its range if and only if there exists $\epsilon > 0$ such that

$$\|\dot{V}(G-G_0)\| = \sup_{h \in \mathcal{H}_1} |\dot{V}(G-G_0)h| \geq \epsilon \|G-G_0\|_1, \quad \text{for all } G \in \mathrm{lin}\,\mathcal{G}.$$

Notice that we have, since $(e \mapsto \mathrm{sgn}(g(e) - g_0(e))) \in \mathcal{H}_1$,

$$\|G-G_0\|_1 = \sum_{e=0}^{\infty} |g(e) - g_0(e)| \leq \sup_{h \in \mathcal{H}_1} \left| \int h\,\mathrm{d}(G-G_0) \right|.$$

Hence it suffices to prove that there exists $\epsilon > 0$ such that, for all $G \in \operatorname{lin} \mathcal{G}$,

$$\begin{aligned}\|\dot{V}(G - G_0)\| &= \sup_{h \in \mathcal{H}_1} |\dot{V}(G-G_0)h| = \sup_{h\in\mathcal{H}_1} \left| \int (A_0^* A + C) h \,\mathrm{d}(G - G_0) \right| \\ &\geq \epsilon \sup_{h\in\mathcal{H}_1} |\int h \,\mathrm{d}(G-G_0)|.\end{aligned}$$

Of course, a sufficient condition for this is given by $\epsilon\mathcal{H}_1 \subset \{(A_0^* A_0 + C)h \mid h \in \mathcal{H}_1\}$, which in turn holds if $B = A_0^* A_0 + C : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is onto and continuously invertible. To see this, first note that $\epsilon\mathcal{H}_1 \subset \{(A_0^* A_0 + C)h \mid h \in \mathcal{H}_1\}$ is equivalent to $\epsilon B^{-1}\mathcal{H}_1 \subset \mathcal{H}_1$. Since $\mathcal{H}_1$ is the unit-ball of $\ell^\infty(\mathbb{Z}_+)$ it thus suffices to show that there exists $\epsilon > 0$ such that $\|B^{-1}h\|_\infty \leq \epsilon^{-1}$ for all $h \in \mathcal{H}_1$. Since B^{-1} is continuous, there exists $\epsilon > 0$ such that $\|Bf\|_\infty \geq \epsilon\|f\|_\infty$ for all $f \in \ell^\infty(\mathbb{Z}_+)$. Taking $h \in \mathcal{H}_1$ and $f = B^{-1}h$ (which is possible, because B is onto), we indeed arrive at $\|B^{-1}h\|_\infty = \|f\|_\infty \leq \epsilon^{-1}\|Bf\|_\infty = \epsilon^{-1}\|h\|_\infty \leq \epsilon^{-1}$. Thus $\dot{\Psi}$ is continuously invertible if we prove that $A_0^* A_0 + C : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is onto and continuously invertible. This concludes Step 2.

Step 3:
In this step we prove that $B = A_0^* A_0 + C : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is onto and continuously invertible, which will conclude the proof of (L2). Notice that $C : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is a compact operator, since it has finite dimensional range. From functional analysis (see, for example, Van der Vaart (2000) Lemma 25.93), it is known that (all operators are defined on and take values in a common Banach space) the sum of a compact operator and a continuous operator, which is onto and has a continuous inverse, is continuously invertible and onto if the sum operator is 1-to-1. Thus it suffices to prove that $A_0^* A_0 : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is continuous, onto, and has a continuous inverse (Step 3a), and that B is one-to-one (Step 3b).
Step 3a:
The continuity of $A_0^* A_0$ is immediate,

$$\begin{aligned}\sup_{e\in\mathbb{Z}_+} |A_0^* A_0 h(e) - A_0^* A_0 h'(e)| &= \sup_{e\in\mathbb{Z}_+} |\mathbb{E}_{\nu_0,\theta_0} \left[\mathbb{E}_{\theta_0,G_0}\left[h(\varepsilon_0) - h'(\varepsilon_0) \mid X_0, \ldots, X_{-p}\right] \mid \varepsilon_0 = e\right]| \\ &\leq \sup_{e\in\mathbb{Z}_+} |h(e) - h'(e)|.\end{aligned}$$

Next we show that to prove that $A_0^* A_0 : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is onto and continuously invertible, it suffices to prove that $A_0^* A_0 : [\varepsilon] \to [\varepsilon]$ is onto and continuously invertible. If we already know that $A_0^* A_0 : [\varepsilon] \to [\varepsilon]$ is invertible, then $A_0^* A_0 : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is also invertible (since there are no "a.s.-problems" if support$(G_0) = \mathbb{Z}_+$). If $h \in \ell^\infty(\mathbb{Z}_+)$ it is clear that $A_0^* A_0 h \in \ell^\infty(\mathbb{Z}_+)$ Suppose next that $A_0^* A_0 h \in \ell^\infty(\mathbb{Z}_+)$. Since

$$\begin{aligned}A_0^* A_0 h(e) &= \sum_{y\in\mathbb{Z}_+^p} \sum_{x_0=0}^{\infty} \nu_0\{y\} \left(*_{i=1}^p \mathrm{Bin}_{y_i,\theta_{0,i}}\right)\{x_0 - e\} \mathbb{E}_{\theta_0,G_0}[h(\varepsilon_0) \mid Y_0 = y] \\ &\geq \nu_0\{0,\ldots,0\} h(e),\end{aligned}$$

this implies $h \in \ell^\infty(\mathbb{Z}_+)$. Thus, since $A_0^* A_0 : [\varepsilon] \to [\varepsilon]$ is onto and $\ell^\infty(\mathbb{Z}_+) \subset [\varepsilon]$, $A_0^* A_0 : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is indeed onto. Thus $A_0^* A_0 : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is a linear continuous operator, whose range is a Banach space, we conclude, from Banach's theorem, that $A_0^* A_0 :$

$\ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is continuously invertible. Hence, the proof of Step 3a is complete once we show that $A_0^* A_0 : [\varepsilon] \to [\varepsilon]$ is onto and continuously invertible. First we show that $A_0 : [\varepsilon] \to R_2(A_0) \subset \mathrm{L}_2(\nu_0 \otimes P^{\theta_0,G_0})$ ($R_2(A_0)$ is the range of A_0, where we use the "subscript 2" to stress that we working in L_2) is one-to-one, i.e. that the null space of A_0 is trivial. Let $h : \mathbb{Z}_+ \to \mathbb{R}$ such that $\mathbb{E}_{G_0} h^2(\varepsilon_0) < \infty$ and

$$0 = \mathbb{E}_{\theta_0,G_0}[h(\varepsilon_0) \mid X_0, \ldots, X_{-p}] \quad \mathbb{P}_{\nu_0,\theta_0,G_0} - \text{a.s.}$$

Since $\mathrm{support}(G_0) = \mathbb{Z}_+$, we can drop the "a.s." and we obtain

$$0 = \mathbb{E}_{\theta_0,G_0}[h(\varepsilon_0) \mid X_0 = e, X_{-1} = 0, \ldots, X_{-p} = 0] = h(e) \quad \forall e \in \mathbb{Z}_+$$

We see that $h(\varepsilon_0) = 0$ and hence A_0 is invertible, with inverse

$$(A_0^{-1} f)(\varepsilon_0) = f(0, \ldots, 0, \varepsilon_0).$$

Of course this is a linear operator. Moreover it is continuous since (remember that $P^{\theta_0,G_0}_{(0,\ldots,0),x_0} = g_0(x_0)$)

$$\begin{aligned}
\mathbb{E}_{G_0}\left(A_0^{-1} f(\varepsilon_0) - A_0^{-1} f'(\varepsilon_0)\right)^2 &= \mathbb{E}_{G_0}\left(f(0, \ldots, 0, \varepsilon_0) - f'(0, \ldots, 0, \varepsilon_0)\right)^2 \\
&\le \frac{1}{\nu_0\{0, \ldots, 0\}} \mathbb{E}_{\nu_0,\theta_0,G_0}\left(f(X_{-p}, \ldots, X_0) - f'(X_{-p}, \ldots, X_0)\right)^2.
\end{aligned}$$

Since $A_0 : [\varepsilon] \to R_2(A_0)$ is linear, continuous, one-to-one, and has a continuous inverse, we conclude from Banach's theorem that $R_2(A_0)$ is a closed subspace of $\mathrm{L}_2(\nu_0 \otimes P^{\theta_0,G_0})$. Since A_0 is one-to-one, and $R_2(A_0)$ is closed we conclude that the operator $A_0^* A_0 : [\varepsilon] \to [\varepsilon]$ is one-to-one, onto and has a continuous inverse (fact from Hilbert-space theory). This concludes Step 3a.

Step 3b:

In this step we show that $B : \ell^\infty(\mathbb{Z}_+) \to \ell^\infty(\mathbb{Z}_+)$ is one-to-one. This essentially follows from the proof of Lemma 25.92 in Van der Vaart (2000). For completeness we repeat the arguments, where we circumvent the need to consider the efficient information matrix for θ. Let $h \in \ell^\infty(\mathbb{Z}_+)$, with $Bh = 0$. We have to prove that $h = 0$. Introduce $\mathbb{R}^p \ni a = -I_{\theta_0}^{-1} \mathbb{E}_{\nu_0,\theta_0,G_0} A_0 h(Z_0) \dot{\ell}_\theta(Z_0; \theta_0, G_0)$, and notice that $Ch = a^T A_0^* \dot{\ell}_\theta(\cdot; \theta_0, G_0)$. Let $S = a^T \dot{\ell}_\theta(Z_0; \theta_0, G_0) + A_0 h(Z_0) - \int h \, \mathrm{d}G_0$. First we show that for $a \neq 0$ we have $\mathbb{E}_{\nu_0,\theta_0,G_0} S^2 > 0$. Suppose that $S = 0$ $\mathbb{P}_{\nu_0,\theta_0,G_0}$-a.s. Then conditioning on $X_{-p} = \cdots = X_{-1} = 0$ yields $h(e) - \int h \, \mathrm{d}G_0 = 0$ for all e. And we obtain, since I_{θ_0} is positive definite (Drost et al. (2006b) Theorem 3.1), $\mathbb{E}_{\nu_0,\theta_0,G_0} S^2 = a^T I_{\theta_0} a > 0$ for $a \neq 0$, which contradicts $\mathbb{E}_{\nu_0,\theta_0,G_0} S^2 = 0$. Conclude that we have, for $a \neq 0$,

$$0 < \mathbb{E}_{\nu_0,\theta_0,G_0} S^2 = \mathbb{E}_{\nu_0,\theta_0,G_0}\left(A_0 h(Z_0) - \int h \, \mathrm{d}G_0\right)^2 - a^T I_{\theta_0} a.$$

On the other hand $Bh = 0$, yields

$$\begin{aligned}
0 &= \mathbb{E}_{\nu_0,\theta_0,G_0} h(\varepsilon_0) Bh(\varepsilon_0) = \mathbb{E}_{\nu_0,\theta_0,G_0}(A_0 h(Z_0))^2 + a^T \mathbb{E}_{\nu_0,\theta_0,G_0} A_0 h(Z_0) \dot{\ell}_\theta(Z_0; \theta_0, G_0) \\
&\ge \mathbb{E}_{\nu_0,\theta_0,G_0}\left(A_0 h(Z_0) - \int h \, \mathrm{d}G_0\right)^2 - a^T I_{\theta_0} a.
\end{aligned}$$

From the previous two displays we conclude $a = 0$, which by definition of a and C yields $Ch = 0$. Hence $A_0^* A_0 h = 0$, which, by Step 3a, yields $h = 0$. This concludes the proof.

So we have proved (L2) for the case support$(G_0) = \mathbb{Z}_+$. The proof for the general case uses exactly the same arguments, if we replace in the arguments where "a.s." plays a role $\mathbb{Z}_+$ by support(G_0). Recall that we always have, by assumption, $g_0(0) > 0$.

B.3.3. Proof of (L3)

The weak-convergence of $\sqrt{n}\left(\Psi_{n1} - \Psi_1^{\theta_0,G_0}\right)(\theta_0, G_0)$ follows from Lemma A.1C, since we are dealing with a finite function class and $|\dot{\ell}_{\theta,i}(Z_0;\theta_0,G_0)| \le X_{-i}(\theta_{0,i}(1-\theta_{0,i}))^{-1}$, $i = 1,\ldots,p$. Hence, due to the form of $\sqrt{n}\left(\Psi_n - \Psi^{\theta_0,G_0}\right)(\theta_0,G_0)$, it suffices to prove that $\sqrt{n}\left(\Psi_{n2} - \Psi_2^{\theta_0,G_0}\right)(\theta_0,G_0)$ weakly converges, under $\mathbb{P}_{\nu_0,\theta_0,G_0}$, in $\ell^\infty(\mathcal{H}_1)$ to a tight Gaussian process. This can be reexpressed as the weak convergence of the empirical process $\{\mathbb{Z}_n f \mid f \in \mathcal{F}\}$, where $\mathcal{F} = \{\mathbb{Z}_+^{p+1} \ni (x_{-p},\ldots,x_0) \mapsto A_0 h(x_{-p},\ldots,x_0) \mid h \in \mathcal{H}_1\}$. We use Lemma A.1B to verify this. Let $\delta > 0$. Take $M_\delta = \lceil (8(p+1)\mathbb{E}_{\nu_0,\theta_0,G_0} X_0^{p+2})^{1/(p+2)}\delta^{-2/(p+2)}\rceil$. By Markov's inequality we have

$$\mathbb{P}_{\nu_0,\theta_0,G_0}\{\max_{i=0,\ldots,p} X_{-i} \ge M_\delta\} \le \frac{\delta^2}{8}.$$

Next, form a grid of cubes with sides of length $\epsilon_\delta = \delta/2\sqrt{2}$ over $[-1,1]^{\{0,\ldots,M_\delta-1\}^{p+1}}$. This yields $N_\delta \le \lceil 2/\epsilon_\delta \rceil^{M_\delta^{p+1}}$ points. Each point yields a mapping $f : \{0,\ldots,M_\delta - 1\}^{p+1} \to [-1,1]$. We label these as $f_1,\ldots,f_{N_\delta}$. Since for $h \in \mathcal{H}_1$ we have $|A_0 h| \le 1$, there exists $i \in \{1,\ldots,N_\delta\}$ such that $f_i(x_{-p},\ldots,x_0) - \delta/2\sqrt{2} \le A_0 h(x_{-p},\ldots,x_0) \le f_i(x_{-p},\ldots,x_0) + \delta/2\sqrt{2}$ for $x_{-p},\ldots,x_0 \le M_\delta - 1$. Next we introduce mappings f_i^L, f_i^U, $i = 1,\ldots,N_\delta$, from $\mathbb{Z}_+^{p+1}$ into $[-1,1]$ by $f_i^L = -1 \vee (f_i - \delta/2\sqrt{2})$ if $\max\{x_{-p},\ldots,x_0\} \le M_\delta - 1$, $f_i^L = -1$ for $\max\{x_{-p},\ldots,x_0\} \ge M_\delta$, and $f_i^U = 1 \wedge (f_i + \delta/2\sqrt{2})$ if $\max\{x_{-p},\ldots,x_0\} \le M_\delta - 1$ and $f_i^U = 1$ if $\max\{x_{-p},\ldots,x_0\} \ge M_\delta$. Conclude that for $h \in \mathcal{H}_1$ there exists $i \in \{1,\ldots,N_\delta\}$ such that $f_i^L \le A_0 h \le f_i^U$. So the brackets $\left[f_i^L, f_i^U\right]$, $i = 1,\ldots,N_\delta$, cover $\mathcal{F}$ and satisfy

$$\mathbb{E}_{\nu_0,\theta_0,G_0}(f_i^U - f_i^L)^2 \le \left(\frac{\delta}{\sqrt{2}}\right)^2 + 4\mathbb{P}_{\nu_0,\theta_0,G_0}\{\max_{i=0,\ldots,p} X_{-i} \ge M_\delta\} \le \delta^2.$$

Conclude that $N_{[\,]}(\delta,\mathcal{F}) \le N_\delta$. Using $\log(x) \le m(x^{1/m} - 1)$ for $x > 0$, $m \in \mathbb{N}$, it easily follows that we can find $a > 0$ such that $\int_0^1 x^{-a}\left(\log N_{[\,]}(x,\mathcal{F})\right)^{1/2} \mathrm{d}x < \infty$. Since the envelope of $\mathcal{F}$ is bounded by 2, an application of Lemma A.1C concludes the proof.

B.3.4. Proof of (L4)

In step A we prove

$$\sqrt{n}\left(\Psi_{n2} - \Psi_2^{\theta_0,G_0}\right)(\hat{\theta}_n,\hat{G}_n) - \sqrt{n}\left(\Psi_{n2} - \Psi_2^{\theta_0,G_0}\right)(\theta_0,G_0) = o(1;\mathbb{P}_{\nu_0,\theta_0,G_0}), \tag{37}$$

and in step B we prove

$$\sqrt{n}\left(\Psi_{n1} - \Psi_1^{\theta_0,G_0}\right)(\hat{\theta}_n,\hat{G}_n) - \sqrt{n}\left(\Psi_{n1} - \Psi_1^{\theta_0,G_0}\right)(\theta_0,G_0) = o(1;\mathbb{P}_{\nu_0,\theta_0,G_0}), \tag{38}$$

which will conclude the proof. Introduce for $\delta > 0$ $B_0(\delta) = \{(\theta, G) \in \Theta \times \mathcal{G} \mid |\theta - \theta_0| + \|G - G_0\|_1 \leq \delta\}$.

Step A: If we prove that there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \sup_{h\in\mathcal{H}_1} \mathbb{E}_{\nu_0,\theta_0,G_0} \left(A_{\theta_n,G_n} h(X_{-p},\ldots,X_0) - A_{\theta_0,G_0} h(X_{-p},\ldots,X_0)\right)^2 = 0,$$

for all sequences (θ_n, G_n) in $\Theta \times \mathcal{G}$ converging to (θ_0, G_0), and that the empirical process $\{\mathbb{Z}_n f \mid f \in \mathcal{F}^\delta\}$ with $\mathcal{F}^\delta$ given by

$$\mathcal{F}^\delta = \{(x_{-p},\ldots,x_0) \mapsto A_{\theta,G} h(x_{-p},\ldots,x_0) - A_{\theta_0,G_0} h(x_{-p},\ldots,x_0) \mid h \in \mathcal{H}_1,\ (\theta,G) \in B_0(\delta)\},$$

weakly converges to a tight Gaussian process, then (37) follows from (the proof of) Lemma 3.3.5 in Van der Vaart and Wellner (1993). Since

$$\sup_{h\in\mathcal{H}_1} |A_{\theta_n,G_n} h(X_{-p},\ldots,X_0) - A_{\theta_0,G_0} h(X_{-p},\ldots,X_0)| \leq 2,$$

and since, for fixed $X_{-p},\ldots,X_0$,

$$\sup_{h\in\mathcal{H}_1} |A_{\theta_n,G_n} h(X_{-p},\ldots,X_0) - A_{\theta_0,G_0} h(X_{-p},\ldots,X_0)| \leq \left| \frac{P^{\theta_0,G_0}_{Y_0,X_0}}{P^{\theta_n,G_n}_{Y_0,X_0}} - 1 \right| + \frac{\|G_n - G_0\|_1}{P^{\theta_n,G_n}_{Y_0,X_0}} \to 0,$$

the first condition easily follows by an application of the dominated convergence theorem. That the process $\{\mathbb{Z}_n f \mid f \in \mathcal{F}^\delta\}$ weakly converges to a tight Gaussian process follows by the same arguments as in the proof of (L3).

Step B: We consider the first coordinate. The others proceed in exactly the same way. If we prove that there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \mathbb{E}_{\nu_0,\theta_0,G_0} \left(\dot{\ell}_{\theta,1}(X_{-p},\ldots,X_0;\theta_n,G_n) - \dot{\ell}_{\theta,1}(X_{-p},\ldots,X_0;\theta_0,G_0)\right)^2 = 0,$$

for all sequences (θ_n, G_n) in $\Theta \times \mathcal{G}$ converging to (θ_0, G_0), and that the empirical process $\{\mathbb{Z}_n f \mid f \in \mathcal{F}^\delta\}$ with $\mathcal{F}^\delta$ given by

$$\mathcal{F}^\delta = \left\{(x_{-p},\ldots,x_0) \mapsto \dot{\ell}_{\theta,1}(x_{-p},\ldots,x_0;\theta,G) - \dot{\ell}_{\theta,1}(x_{-p},\ldots,x_0;\theta_0,G_0) \mid (\theta,G) \in B_0(\delta)\right\},$$

converges weakly to a tight Gaussian process, then (38) follows from Lemma 3.3.5 in Van der Vaart and Wellner (1993). Choose $\delta > 0$ such that for all θ in the ball we have $(\theta_i(1-\theta_i))^{-1} \leq C$ for certain $C > 0$ and all $i = 1,\ldots,p$. The first condition easily follows using dominated convergence (use $4CX_{-1}^2$ as dominating function). We use Lemma A.1C to verify the second condition. Let $\eta > 0$. Take $M_\eta = \lceil \alpha^{1/(p+4)} \eta^{-2/(p+2)} \rceil$, where the constant α is given by $\alpha = (p+1)\left(8C^2 \mathbb{E}_{\nu_0,\theta_0,G_0} X_0^{p+4}\right)^{(p+4)/(p+2)}$. By Markov's inequality we have

$$\mathbb{P}_{\nu_0,\theta_0,G_0}\{\max_{i=0,\ldots,p} X_{-i} \geq M_\eta\} \leq \frac{\mathbb{E}_{\nu_0,\theta_0,G_0} X_0^{p+4}}{\left(8C^2 \mathbb{E}_{\nu_0,\theta_0,G_0} X_0^{p+4}\right)^{(p+4)/(p+2)}} \eta^{2\frac{p+4}{p+2}},$$

and using Hölder's inequality we now obtain

$$\begin{aligned}
&\mathbb{E}_{\nu_0,\theta_0,G_0} X_{-1}^2 1\{\max_{i=0,\dots,p} X_{-i} \geq M_\eta\} \\
&\quad \leq \left(\mathbb{E}_{\nu_0,\theta_0,G_0} X_{-1}^{p+4}\right)^{2/(p+4)} \left(\mathbb{P}_{\nu_0,\theta_0,G_0}\{\max_{i=0,\dots,p} X_{-i} \geq M_\eta\}\right)^{(p+2)/(p+4)} \leq \frac{\eta^2}{8C^2}. \quad (39)
\end{aligned}$$

Notice that for all $(\theta, G) \in B_0(\delta)$ we have

$$|\dot{\ell}_{\theta,1}(x_{-p},\dots,x_0;\theta,G) - \dot{\ell}_{\theta,1}(x_{-p},\dots,x_0;\theta_0,G_0)| \leq 2Cx_{-1}.$$

Next, form a grid of cubes with sides of length $\epsilon_\eta = \eta/2\sqrt{2}$ over $[-2CM_\eta, 2CM_\eta]^{\{0,\dots,M_\eta-1\}^{p+1}}$. This yields $N_\eta \leq \lceil 4CM_\eta/\epsilon_\eta \rceil^{M_\eta^{p+1}}$ points. Each point yields a mapping $f : \{0,\dots,M_\eta - 1\}^{p+1} \to [-2CM_\eta, 2CM_\eta]$. We label these as $f_1,\dots,f_{N_\eta}$. So, for $(\theta,G) \in B_0(\delta)$, there exists $i \in \{1,\dots,N_\eta\}$ such that, for $x_{-p},\dots,x_0 \leq M_\eta - 1$,

$$\begin{aligned}
f_i(x_{-p},\dots,x_0) - \frac{\eta}{2\sqrt{2}} &\leq \dot{\ell}_{\theta,1}(x_{-p},\dots,x_0;\theta,G) - \dot{\ell}_{\theta,1}(x_{-p},\dots,x_0;\theta_0,G_0) \\
&\leq f_i(x_{-p},\dots,x_0) + \frac{\eta}{2\sqrt{2}}.
\end{aligned}$$

Next we introduce mappings f_i^L, f_i^U, $i = 1,\dots,N_\eta$, from $\mathbb{Z}_+^{p+1}$ into $\mathbb{R}$ by $f_i^L = -2CM_\eta \vee (f_i - \eta/2\sqrt{2})$ if $\max\{x_{-p},\dots,x_0\} \leq M_\eta - 1$ and $f_i^L = -2Cx_{-1}$ if $\max\{x_{-p},\dots,x_0\} \geq M_\eta$, and $f_i^U = 2CM_\eta \wedge (f_i + \eta/2\sqrt{2})$ if $\max\{x_{-p},\dots,x_0\} \leq M_\eta - 1$ and $f_i^U = 2Cx_{-1}$ if $\max\{x_{-p},\dots,x_0\} \geq M_\eta$. Conclude that for $(\theta,G) \in B_0(\delta)$ there exists $i \in \{1,\dots,N_\eta\}$ such that $f_i^L \leq \dot{\ell}_{\theta,1}(\theta,G) - \dot{\ell}_{\theta,1}(\theta_0,G_0) \leq f_i^U$. So the brackets $\left[f_i^L, f_i^U\right]$, $i = 1,\dots,N_\eta$, cover $\mathcal{F}^\delta$ and satisfy, by (39),

$$\mathbb{E}_{\nu_0,\theta_0,G_0}(f_i^U - f_i^L)^2 \leq \left(\frac{\eta}{\sqrt{2}}\right)^2 + 4C^2 \mathbb{E}_{\nu_0,\theta_0,G_0} X_{-1}^2 1\{\max_{i=0,\dots,p} X_{-i} \geq M_\eta\} \leq \eta^2.$$

Conclude that $N_{[\,]}(\eta, \mathcal{F}^\delta) \leq N_\eta$. Using $\log(x) \leq m(x^{1/m} - 1)$ for $x > 0$, $m \in \mathbb{N}$, it easily follows that we can find $a > 0$ such that $\int_0^1 x^{-a} \left(\log N_{[\,]}(x, \mathcal{F}^\delta)\right)^{1/2} \mathrm{d}x < \infty$. Since the envelope of $\mathcal{F}^\delta$ is bounded by the integrable variable $2CX_{-1}$, an application of Lemma A.1C concludes the proof. □

B.4. *Proof of LAN Theorem A.3*

By an application of the main theorem of Drost et al. (2006b) the lemma is proved once we prove that $\nu_n\{X_{-p},\dots,X_{-1}\} - \nu_{\theta,G}\{X_{-p},\dots,X_{-1}\} \xrightarrow{p} 0$, under $\mathbb{P}_{\nu_{\theta,G},\theta,G}$. By Lemma A.1D this follows if we show (recall that $Y_t = (X_{t-1},\dots,X_{t-p})^T$)

$$\lim_{n\to\infty} \sup_{y \in \mathbb{Z}_+^p} \frac{\sup_{f:\,|f|\leq V} \left|\mathbb{E}_{\delta_y,\theta_n,G_n} f(Y_1) - \mathbb{E}_{\delta_y,\theta,G} f(Y_1)\right|}{V(y)} = 0, \qquad (40)$$

where $V(y) = 1 + \sum_{i=1}^p c_i y_i$, $c_i = \theta_i + \dots, \theta_p$ for $i = 1,\dots,p$. Straightforward computations yield

$$\mathbb{E}_{\delta_y,\theta_n,G_n} f(Y_1) - \mathbb{E}_{\delta_y,\theta,G} f(Y_1) = \frac{1}{\sqrt{n}} \mathbb{E}_{\delta_y,\theta_n,G}(h(\varepsilon_0) - \mathbb{E}_G h(\varepsilon_0)) f(Y_1)$$

$$+\int_0^{\frac{1}{\sqrt{n}}}\sum_{i=1}^p a_i \mathbb{E}_{\delta_y,\theta+\tau a,G} f(Y_1)\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\,\mathrm{d}\tau.$$

We have, for a constant $C>0$, the bound

$$\begin{aligned}\sup_{f:\,|f|\le V}\left|\mathbb{E}_{\delta_y,\theta_n,G}(h(\varepsilon_0)-\mathbb{E}_G h(\varepsilon_0))f(Y_1)\right| &\le 2\|h\|_\infty\left(1+\sum_{i=2}^p c_i y_{i-1}+\mu_G+\left(\theta+\frac{a}{\sqrt{n}}\right)^T y\right)\\ &\le CV(y).\end{aligned}$$

Next let $i\in\{1,\dots,p\}$. Of course the supremum in

$$\sup_{f:\,|f|\le V}\left|\mathbb{E}_{\delta_y,\theta+\tau a,G} f(Y_1)\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\right|$$

is taken for $f=V1_A-V1_{A^c}$, where $A=\{\dot{s}_{X_{-i},\theta_i}(\theta\circ X_{-i})>0\}$. Consequently, in the first equality we exploit $\mathbb{E}_{\delta_y,\theta+\tau a,G}\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})=0$,

$$\begin{aligned}&\sup_{f:\,|f|\le V}\left|\mathbb{E}_{\delta_y,\theta+\tau a,G} f(Y_1)\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\right|\\ &\quad=\sup_{f:\,|f|\le V}\left|\mathbb{E}_{\delta_y,\theta+\tau a,G}(f(Y_1)-\mathbb{E}_{\delta_y,\theta+\tau a,G}f(Y_1))\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\right|\\ &\quad=\mathbb{E}_{\delta_y,\theta+\tau a,G}1_A(V(Y_1)-\mathbb{E}_{\delta_y,\theta+\tau a,G}V(Y_1))\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\\ &\qquad-\mathbb{E}_{\delta_y,\theta+\tau a,G}1_{A^c}(V(Y_1)-\mathbb{E}_{\delta_y,\theta+\tau a,G}V(Y_1))\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\end{aligned}$$

(fill in V and use $\mathbb{E}_{\delta_y,\theta+\tau a,G}\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})=0$)

$$\begin{aligned}&=c_1\mathbb{E}_{\delta_y,\theta+\tau a,G}1_A(X_0-\mathbb{E}_{\delta_y,\theta+\tau a,G}X_0)\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\\ &\quad-c_1\mathbb{E}_{\delta_y,\theta+\tau a,G}1_{A^c}(X_0-\mathbb{E}_{\delta_y,\theta+\tau a,G}X_0)\dot{s}_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})\\ &\le c_1\sqrt{\mathbb{E}_{\delta_y,\theta+\tau a,G}(X_0-\mathbb{E}_{\delta_y,\theta+\tau a,G}X_0)^2}\sqrt{\mathbb{E}_{\delta_y,\theta+\tau a,G}\dot{s}^2_{X_{-i},\theta_i+\tau a_i}(\theta_i\circ X_{-i})}\\ &=c_1\sqrt{\sigma_G^2+\sum_{j=1}^p(\theta_j+ta_j)y_j}\sqrt{\theta_i(1-\theta_i)y_i}\\ &\le CV(y),\end{aligned}$$

for a constant $C>0$. A combination of the previous four displays easily yields (40). □