CONSISTENTLY DETERMINING THE NUMBER OF FACTORS IN MULTIVARIATE VOLATILITY MODELLING

Qiang Xia[1], Wangli Xu[1] and Lixing Zhu[2]

[1]Renmin University of China and [2]Hong Kong Baptist University

Supplementary Material

S1 Lemmas

Lemma 1. *When $d \to \infty$, under Conditions 5 and 7, we have*

$$\| \hat{\Omega} - \Omega \|_{max} = O_p(n^{-1/2}).$$

Proof. First, note that for two matrices, $\| MN \|_{max} \leq \| M \|_{max} \| N \|_{max}$, and $\| M \|_{max} = \| M^{\tau} \|_{max}$. Some elementary calculations lead to the following decomposition:

$$\begin{aligned}
& \| \hat{\Omega} - \Omega \|_{max} \\
\leq & \sum_{m=1}^{m_0} \| S_m S_m^{\tau} - [\frac{1}{d}\Sigma_Y(m)][\frac{1}{d}\Sigma_Y^{\tau}(m)] \|_{max} \} \\
\leq & \sum_{m=1}^{m_0} \| S_m - ES_m \|_{max}^2 \\
& + \sum_{m=1}^{m_0} \| ES_m - [\frac{1}{d}\Sigma_Y(m)] \|_{max}^2 \\
& +2 \sum_{m=1}^{m_0} \| S_m - ES_m \|_{max} \| ES_m - [\frac{1}{d}\Sigma_Y(m)] \|_{max} \\
& +2 \sum_{m=1}^{m_0} \| [\frac{1}{d}\Sigma_Y(m)] \|_{max} \| S_m - ES_m \|_{max} \\
& +2 \sum_{m=1}^{m_0} \| [\frac{1}{d}\Sigma_Y(m)] \|_{max} \| ES_m - [\frac{1}{d}\Sigma_Y(m)] \|_{max} \\
=: & K_1 + K_2 + K_3 + K_4 + K_5.
\end{aligned}$$

By the definition, we have

$$
\begin{aligned}
d\,S_m &= \frac{1}{(n-m)}\sum_{t=1}^{n-m}(Y_{t+m}-\bar{Y})(Y_t-\bar{Y})^\tau \\
&= \frac{1}{(n-m)}\sum_{t=1}^{n-m}Y_{t+m}Y_t^\tau - \frac{1}{(n-m)}\sum_{t=1}^{n-m}Y_t\bar{Y}^\tau \\
&\quad -\frac{1}{(n-m)}\sum_{t=1}^{n-m}\bar{Y}Y_t^\tau + \bar{Y}\bar{Y}^\tau. \qquad (S1.1)
\end{aligned}
$$

By Condition 5, an application of Markov inequality yields that all the elements of $\frac{1}{(n-m)}\sum_{t=1}^{n-m}Y_t-\bar{Y}$ and $\frac{1}{(n-m)}\sum_{t=1}^{n-m}Y_{t+m}-\bar{Y}$ uniformly have the order $O_p(1/n)$. Thus, without confusion, we write that $dS_m=\frac{1}{(n-m)}\sum_{t=1}^{n-m}Y_{t+m}Y_t^\tau - \bar{Y}\bar{Y}^\tau + O_p(1/n) := d\tilde{S}_m + O_p(1/n)$. In fact, to prove that $\sqrt{n}\parallel\{S_m-ES_m\}\parallel_{max}=O_p(1)$, which is equivalent to prove that $\sqrt{n}\parallel\{\tilde{S}_m-E\tilde{S}_m\}\parallel_{max}=O_p(1)$. Hence, we only prove that $\sqrt{n}\parallel\{\tilde{S}_m-E\tilde{S}_m\}\parallel_{max}=O_p(1)$.

Denote the (i,j)-th element of $(d\tilde{S}_m)$ and $\Sigma_Y(m)$ by $s_m(i,j)$ and $\sigma_Y(m)_{(i,j)}$ respectively. Note that $(d\cdot\tilde{S}_m)$ is a sample variance and $\Sigma_Y(m)$ is the corresponding population one.

By Condition 5 and Theorem 3 of Doukhan (1994), we have for $i=j$

$$
\begin{aligned}
&\max_{i,i} E(s_m(i,i)-Es_m(i,i))^2 \\
\le\ & 2\big\{\max_{i,i} Var\big(\frac{1}{n-m}\sum_{t=1}^{n-m} y_{t+m,i}y_{t,i}\big) + \max_{i,i} Var(\bar{y}_i^2)\big\} \\
=\ & 2\big\{\max_{i,i}\frac{1}{(n-m)^2}\sum_{1\le k,l\le n-m} Cov\big(y_{k,i}y_{k+m,i}, y_{l,i}y_{l+m,i}\big) + \max_{i,i}\frac{1}{n^4}\sum_{1\le h,j,k,l\le n} Cov\big(y_{h,i}y_{j,i}, y_{k,i}y_{l,i}\big)\big\} \\
\le\ & 2\big\{\frac{1}{(n-m)^2}\sum_{k,l}2\varphi^{1/2}(|k-l|)\max_{i,i,k,l}[E\mid y_{k,i}y_{k+m,i}\mid^2]^{1/2}[E\mid y_{l,i}y_{l+m,i}\mid^2]^{1/2} \\
& +\max_{i,i}\frac{12}{n^4}\sum_{1\le h\le j\le k\le l\le n} Cov\big(y_{h,i}y_{j,i}, y_{k,i}y_{l,i}\big) + \max_{i,i}\frac{12}{n^4}\sum_{1\le h\le k\le j\le l\le n} Cov\big(y_{h,i}y_{j,i}, y_{k,i}y_{l,i}\big)\big\} \\
\le\ & 2\big\{\frac{1}{(n-m)^2}\sum_{k,l}2\varphi^{1/2}(|k-l|)\max_{i,i,k,l}[E(y_{k,i})^4E(y_{k+m,i})^4]^{1/4}[E(y_{l,i})^4E(y_{l+m,i})^4]^{1/4} \\
& +\frac{12}{n^4}\sum_{1\le h\le j\le k\le l\le n}2\varphi^{1/2}(k-j)\max_{i,i,h,j,k,l}[E\mid y_{h,i}y_{j,i}\mid^2]^{1/2}[E\mid y_{k,i}y_{l,i}\mid^2]^{1/2} \\
& +\frac{12}{n^4}\sum_{1\le h\le k\le j\le l\le n}2\varphi^{1/2}(min\{k-i,j-k,l-j\})\max_{i,i,h,k,j,l}[E\mid y_{h,i}y_{j,i}\mid^2]^{1/2}[E\mid y_{k,i}y_{l,i}\mid^2]^{1/2}\big\}
\end{aligned}
$$

$$
\begin{aligned}
\leq\ & 2\{\frac{1}{(n-m)^2}\sum_{k,l}2\varphi^{1/2}(|k-l|)\max_{i,i,k,l}[E(y_{k,i})^4E(y_{k+m,i})^4]^{1/4}[E(y_{l,i})^4E(y_{l+m,i})^4]^{1/4} \\
& +\frac{12}{n^4}\sum_{1\leq h\leq j\leq k\leq l\leq n}2\varphi^{1/2}(k-j)\max_{i,i,h,j,k,l}[E(y_{h,i})^4E(y_{j,i})^4]^{1/4}[E(y_{k,i})^4E(y_{l,i})^4]^{1/4} \\
& +\frac{12}{n^4}\sum_{1\leq h\leq k\leq j\leq l\leq n}2\varphi^{1/2}(min\{k-i,j-k,l-j\})\max_{i,i,h,k,j,l}[E(y_{h,i})^4E(y_{j,i})^4]^{1/4}[E(y_{k,i})^4E(y_{l,i})^4]^{1/4}\} \\
=:\ & 4C\frac{1}{(n-m)^2}\sum_{k,l}|k-l|^{-1-c/2}+4C'\frac{1}{n^4}\sum_{1\leq h\leq j\leq k\leq l\leq n}(k-j)^{-1-c/2} \\
& +4C''\frac{1}{n^4}\sum_{1\leq h\leq k\leq j\leq l\leq n}(min\{k-i,j-k,l-j\})^{-1-c/2} \\
=\ & O(1/n)+O(1/n)+O(1/n)=O(1/n). \qquad \text{(S1.2)}
\end{aligned}
$$

For $i\neq j$, the absolute value of the corresponding element of $(d\cdot\tilde{S}_m)$ is controlled by those with $i=j$. Then, by Cauchy inequality, to prove $\sqrt{n}\parallel\{S_m-ES_m\}\parallel_{max}=O_p(1)$, we only need to prove that $E\parallel\{S_m-ES_m\}\parallel_{max}^2=O(1/n)$. Together with the above proof, we only need to prove that $E\parallel\{\tilde{S}_m-E\tilde{S}_m\}\parallel_{max}^2=O(1/n)$. By the definition, we have

$$
\begin{aligned}
& E\parallel\{\tilde{S}_m-E\tilde{S}_m\}\parallel_{max}^2 \\
=\ & E\frac{1}{d^2}\parallel\{d\tilde{S}_m-E(d\tilde{S}_m)\}\parallel_{max}^2\leq\frac{1}{d^2}tr\{E\Big([d\tilde{S}_m-E(d\tilde{S}_m)]^\tau[d\tilde{S}_m-E(d\tilde{S}_m)]\Big)\} \\
=\ & \frac{1}{d^2}\sum_{i,j=1}^{d}E[\tilde{s}_m(i,j)-E\tilde{s}_m(i,j)]^2=\frac{1}{d^2}\sum_{i,j=1}^{d}Var(\tilde{s}_m(i,j))=O(n^{-1}). \qquad \text{(S1.3)}
\end{aligned}
$$

By Condition 5 and Theorem 3 of Doukhan (1994), for all $i,j=1,...,d$ uniformly, we have that

$$|\sigma_Y(m)_{(i,j)}|=|Cov(Y_{t+m,i},Y_{t,j})|\leq 2\varphi^{1/2}(m)[E(Y_{t+m,i})^2]^{1/2}[E(Y_{t,j})^2]^{1/2}=O(1).$$

Hence,

$$\parallel\frac{1}{d}\Sigma_Y(m)\parallel_{max}=O(1). \qquad \text{(S1.4)}$$

By Condition 6, and together with above results, we can have $K_1=O(n^{-1}),K_2=O(n^{-1}),K_3=O_p(n^{-1})$, $K_4=O_p(n^{-1/2})$ and $K_5=O(n^{-1/2})$. Altogether, we have

$$\parallel\hat{\Omega}-\Omega\parallel_{max}=O_p(n^{-1/2}).$$

□

The following lemma is Theorem 8.1.10 in Golub and Van Loan (1996). For convenience in our proof, we cite it as a lemma. See the lemma 3 of Lam et al. (2011) also.

Lemma 2. *Suppose that A and $A+E$ are $d \times d$ are symmetric matrices and that*

$$Q = \underset{r}{[Q_1} \quad \underset{d-r}{Q_2]}$$

is an orthogonal matrix such that $ran(Q_1)$ is an invariant subspace for A. Partition the matrices $Q^\tau AQ$ and $Q^\tau EQ$ as follows:

$$Q^\tau AQ = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \qquad Q^\tau EQ = \begin{bmatrix} E_{11} & E_{21}^\tau \\ E_{21} & E_{22} \end{bmatrix}$$

If $sep(D1, D2) > 0$ and $\| E \|_{max} \leq \dfrac{sep(D1,D2)}{5}$, then there exists a matrix $P \in \mathbf{R}^{(d-r)\times r}$ with

$$\| P \|_{max} \leq \frac{4}{sep(D1,D2)} \| E_{21} \|_{max}$$

such that the columns of $\hat{Q}_1 = (Q_1 + Q_2 P)(I + P^\tau P)^{-1/2}$ define an orthonormal basis for a subspace that is invariant for $A+E$. Where $ran(Q_1) = \{y \in \mathbf{R}^r : y = Q_1 x$ for some $x \in \mathbf{R}^d\}$, and $sep(D1, D2) = \min\limits_{\lambda \in D_1, \mu \in D_2} | \lambda - \mu |$. □

S2 Proof of Theorems

In the proofs thereafter, $\sigma_j(M)$ is denoted to be the j-th singular value of the matrix M, and $\lambda_j(M)$ to be the j-th largest eigenvalue of M. Hence $\sigma_1(M) = \| M \|_{max}$.

Proof of Theorem 1. As the results are parallel to those in Lam et al. (2011) and Lam and Yao (2012), the proof will follow the spirit in their papers. Thus, we will give the proof briefly. As fixed d can be regarded as a special case with diverging d, we then directly go to prove the case of diverging d. The results (i), (ii), (iii) are about diverging d. Consider (i): proving the conclusion $\| \hat{\alpha}_i - \alpha_i \|_{max} \leq \| \hat{A} - A \|_{max} = O_p(n^{-1/2})$.

We assume that A is a half orthogonal matrix in model (2.1). Then $A^\tau \Omega A = \frac{1}{d^2} D$ where D is diagonal with the presentation:

$$D = \sum_{m=1}^{m_0} \left[\Sigma_X(m) A^\tau + \Sigma_{X\xi}(m)\right] \left[\Sigma_X(m) A^\tau + A\Sigma_{X\xi}(m)\right]^\tau$$

If B is an orthogonal complement of A, $\Omega B = 0$, and

$$\begin{bmatrix} A^\tau \\ B^\tau \end{bmatrix} \Omega \begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} \frac{1}{d^2} D & 0 \\ 0 & 0 \end{bmatrix}$$

with $sep(\frac{1}{d^2} D, 0) = \lambda_{min}(\frac{1}{d^2} D)$. By Conditions 2 and 3, we can easily obtain

$$\| \Sigma_{X\xi}(m) \|_{max} = O_p(d).$$

Hence, we have

$$\begin{aligned}
\lambda_{min}(D) &= \lambda_{min}\big(\sum_{m=1}^{m_0}\big[\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big]\big[\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big]^\tau\big)\\
&\geq \sum_{m=1}^{m_0}\lambda_{min}\big(\big[\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big]\big[\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big]^\tau\big)\\
&\geq \max_{1\leq m\leq m_0}\lambda_{min}\big(\big[\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big]\big[\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big]^\tau\big)\\
&= \max_{1\leq m\leq m_0}\big\{\sigma_r\big(\Sigma_X(m)A^\tau+\Sigma_{X\xi}(m)\big)\big\}^2 \qquad \text{(S2.1)}\\
&\geq \max_{1\leq m\leq m_0}\big\{\sigma_r\big(\Sigma_X(m)A^\tau\big)-\sigma_1\big(\Sigma_{X\xi}(m)\big)\big\}^2\\
&\asymp \max_{1\leq m\leq m_0}\big\{\sigma_r\big(\Sigma_X(m)A^\tau\big)\big\}^2\\
&\asymp d^2. \qquad \text{(S2.2)}
\end{aligned}$$

The second equality (3.5) is due to the relation between singular values and eigenvalues. Let $E_\Omega=\hat\Omega-\Omega$. Invoking Lemma 1, we have

$$\parallel E_\Omega \parallel_{max}=\parallel \hat\Omega-\Omega\parallel_{max}=O_p(n^{-1/2}). \qquad \text{(S2.3)}$$

Similar to Lam et al. (2011), applying Lemma 2, we have

$$\parallel \hat A-A\parallel_{max}=O_p(n^{-1/2}).$$

Hence,

$$\parallel \hat\alpha_i-\alpha_i\parallel_{max}\leq\parallel \hat A-A\parallel_{max}\leq O_p(n^{-1/2}). \qquad \text{(S2.4)}$$

(i) is proved. Deal with (ii) now: To prove that $\mid \hat\lambda_i-\lambda_i\mid=O_p(n^{-1/2})$ for $i=1,...,r$. First, we can have the decomposition:

$$\hat\lambda_i-\lambda_i=\hat\alpha_i^\tau\hat\Omega\hat\alpha_i-\alpha_i^\tau\Omega\alpha_i=L_1+L_2+L_3+L_4+L_5$$

where

$$\begin{aligned}
L_1 &= (\hat\alpha_i-\alpha_i)^\tau(\hat\Omega-\Omega)(\hat\alpha_i-\alpha_i), \quad i=1,...,r,\\
L_2 &= 2\alpha_i^\tau(\hat\Omega-\Omega)(\hat\alpha_i-\alpha_i), \quad i=1,...,r,\\
L_3 &= (\hat\alpha_i-\alpha_i)^\tau\Omega(\hat\alpha_i-\alpha_i), \quad i=1,...,r,\\
L_4 &= 2\alpha_i^\tau\Omega(\hat\alpha_i-\alpha_i), \quad i=1,...,r,\\
L_5 &= \alpha_i^\tau(\hat\Omega-\Omega)\alpha_i, \quad i=1,...,r.
\end{aligned}$$

From $(S2.1)$ and $(S2.3)$, similar to Lam and Yao (2012), it can be seen that $\mid L_1\mid\leq O_p(n^{-3/2})$, $\mid L_2\mid\leq O_p(n^{-1})$, and $\mid L_5\mid\leq O_p(n^{-1/2})$. Invoking $(S1.4)$ and $(S2.3)$, we obtain that $\mid L_3\mid\leq O_p(n^{-1})$ and $\mid L_4\mid\leq O_p(n^{-1/2})$. Altogether, we have $\mid\hat\lambda_i-\lambda_i\mid= O_p(n^{-1/2})$ for $i=1,...,r$ and the proof is done.

For (iii), we prove that $\mid \hat{\lambda}_i \mid = O_p(n^{-1})$ for $i = r+1, ..., d$. Also, we decompose $\hat{\lambda}_i$ to be:

$$\hat{\lambda}_i = \hat{\beta}_{i-r}^{\tau} \hat{\Omega} \hat{\beta}_{i-r} = M_1 + M_2 + M_3 + M_4$$

where

$$\begin{aligned}
M_1 &= (\hat{\beta}_{i-r} - \beta_{i-r})^{\tau} (\hat{\Omega} - \Omega)(\hat{\beta}_{i-r} - \beta_{i-r}), \quad i = r+1, ..., d, \\
M_2 &= 2(\hat{\beta}_{i-r} - \beta_{i-r})^{\tau} (\hat{\Omega} - \Omega)\beta_{i-r}, \quad i = r+1, ..., d, \\
M_3 &= (\hat{\beta}_{i-r} - \beta_{i-r})^{\tau} \Omega \hat{\beta}_{i-r} - \beta_{i-r}), \quad i = r+1, ..., d, \\
M_4 &= \beta_{i-r}^{\tau} (\hat{\Omega} - \Omega)\beta_{i-r}, \quad i = r+1, ..., d.
\end{aligned}$$

The same as the proof of ($S2.4$), or similar to Lam and Yao (2012), we also obtain

$$\| \hat{\beta}_{i-r} - \beta_{i-r} \|_{max} \leq \| \hat{B} - B \|_{max} \leq O_p(n^{-1/2}).$$

Then, ($S2.3$) and ($S2.4$) yield that similarly as the above for (ii), $\mid M_1 \mid \leq O_p(n^{-3/2})$, $\mid M_2 \mid \leq O_p(n^{-1})$ and

$$\begin{aligned}
\mid M_4 \mid &= \mid \beta_{i-r}^{\tau} (\hat{\Omega} - \Omega)\beta_{i-r} \mid = \mid \beta_{i-r}^{\tau} \Big\{ \sum_{m=1}^{m_0} \big[S_m S_m^{\tau} - [\frac{1}{d}\Sigma_Y(m)][\frac{1}{d}\Sigma_Y^{\tau}(m)] \big] \Big\} \beta_{i-r} \mid \\
&= \mid \beta_{i-r}^{\tau} \Big\{ \sum_{m=1}^{m_0} \big[S_m - [\frac{1}{d}\Sigma_Y(m)] \big] \big[S_m - [\frac{1}{d}\Sigma_Y(m)] \big]^{\tau} \Big\} \beta_{i-r} \mid \\
&\leq \sum_{m=1}^{m_0} \| \big[S_m - [\frac{1}{d}\Sigma_Y(m)] \big] \|_{max}^2 \\
&= O_p(n^{-1}).
\end{aligned}$$

By ($S1.4$) and ($S2.4$), we derive that $\mid M_3 \mid \leq O_p(n^{-1})$. Thus, $\mid \hat{\lambda}_i \mid = \hat{\lambda}_i = O_p(n^{-1})$ for $i = r+1, ..., d$. This completes the proof of (iii) and then the theorem. □

Proof of Theorem 2. Consistency of the ridge-type ratio estimate of (??). On one hand, in probability, for any $i < r$, as $\hat{\lambda}_i \to \lambda_i > c$ for a positive constant $c > 0$ and $\hat{\lambda}_{r+1} = O_p(n^{-1})$. Then for large n

$$\frac{\hat{\lambda}_{i+1} + log(n)/(10n)}{\hat{\lambda}_i + log(n)/(10n)} \approx \frac{\lambda_{i+1}}{\lambda_i} > \frac{log(n)}{10n\lambda_r} \approx \frac{\hat{\lambda}_{r+1} + log(n)/(10n)}{\hat{\lambda}_r + log(n)/(10n)}, \tag{S2.5}$$

in probability. On the other hand, for any j with $r < j \leq d$, as $\hat{\lambda}_j = O_p(n^{-1})$, $\log(n)/(10n)$ is a dominating term in $\hat{\lambda}_{j+1} + log(n)/(10n)$. Thus, for large n

$$\frac{\hat{\lambda}_{j+1} + log(n)/(10n)}{\hat{\lambda}_j + log(n)/(10n)} \to 1 > \frac{log(n)}{10n} \approx \frac{\hat{\lambda}_{r+1} + log(n)/(10n)}{\hat{\lambda}_r + log(n)/(10n)}$$

in probability uniformly over all $j > r$. Together with (S2.5), $\hat{r} \to r$ with probability one.

Consistency of the BIC-type estimate of (??). For any m, $\log(\hat{\lambda}_m+1)-\hat{\lambda}_m \leq 0$, because of (ii). For any $m > r$ as $\hat{\lambda}_m \to 0$ at the rate of $1/n$, then

$$\sum_{m=1}^{d}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\} = \sum_{m=1}^{r}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\} + \sum_{m=r+1}^{d}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\} \to b+O(d/n^2)$$

in probability for some $b<0$. For any k with $r>k$

$$G(r)-G(k) = \frac{n\sum_{m=k+1}^{r}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}}{2\sum_{m=1}^{d}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}} - log(n)\frac{r(r+1)-k(k+1)}{d}.$$

For large n, $\sum_{m=k+1}^{r}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}/\sum_{m=1}^{d}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\} > b_1 > 0$ for a positive constant b_1. Then, due to $log(n)/n \to 0$,

$$\frac{\sum_{m=k+1}^{r}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}}{\sum_{m=1}^{d}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}} > 2log(n)\frac{r(r+1)-k(k+1)}{dn} \tag{S2.6}$$

in probability uniformly over all $k<r$. In other words, $G(r)-G(k)>0$ in probability uniformly over all $k<r$. For k with $r<k$, and any $m>r$, $n\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\} = -n\{\hat{\lambda}_m^2+o_p(\hat{\lambda}_m^2)\} = O_p(1/n)$, and

$$G(r)-G(k) = -\frac{n\sum_{m=r+1}^{k}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}}{2\sum_{m=1}^{d}\{\log(\hat{\lambda}_m+1)-\hat{\lambda}_m\}} + log(n)\frac{k(k+1)-r(r+1)}{d}.$$

Thus, as $nlog(n)/d \to \infty$, $G(r)-G(k)>0$ in probability uniformly over all $k>r$

Together with (S2.6), $\hat{r} \to r$ in probability, which completes the proof. □