**S1**

**Preliminaries**: The indicator function $\mathbf{1}_{\{A\}}$ takes the value one when the requirement $\mathbf{A}$ holds and zero otherwise. For a vector $x \subset \mathbb{R}^N$, the $L_P$-norm is defined as $\|\mathbf{x}\|_p = [\sum_{i=1}^{N}|x_i|^p]^{\frac{1}{p}}$, with $\|\mathbf{x}\|_\infty = \max_i|x_i|$ representing a vital limiting case. The cardinality of $\mathbf{x}$ is represented by $|\,x\,|$. However, the number of its non-zero components is given by $M(\mathbf{x}) = \sum_{i=1}^{N}\mathbf{1}\{x_i \neq 0\}$. The Orlicz norm is defined as $\|\cdot\|_\psi = \inf\{c > 0: \mathbb{E}[\psi(\frac{|\cdot|}{c})] \leq 1\}$, where $\psi(\cdot)$: $\mathbb{R}^+ \to \mathbb{R}^+$ denotes a convex and monotonically increasing function satisfying $\psi(0) = 0$ and $\psi(x) \to \infty$ as $x \to \infty$. Several matrix norms are employed throughout the analysis. Regarding to a matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, the matrix norms induced by the vector $L_P$-norms are defined by $\|\mathbf{A}\|_p = \sup_{x\subset\mathbb{R}^N}(\frac{\|Ax\|_p}{\|x\|_p})$. Notable special cases include $\|\mathbf{A}\|_1 = \max_{1\leq j\leq N}\sum_{i=1}^{M}|a_{i,j}|$, the spectral norm $\|\mathbf{A}\|_2 = [\lambda_{max}(\mathbf{A'A})]^{\frac{1}{2}}$, where $\lambda_{max}(\cdot)$ denotes the largest eigenvalue, and $\|A\|_\infty = \max_{1\leq i\leq M}\sum_{j=1}^{N}|a_{i,j}|$. For convenience, define $\|\mathbf{A}\|_{\vdash} = \max\{\|\mathbf{A}\|_1, \|\mathbf{A}\|_\infty\}$. Let $S \subseteq \{1, \dots, N\}$ be an index set with cardinality $|\,\mathbf{S}\,|$. The vector $\mathbf{x}_S$ denotes the $|\,S\,|$-dimensional subvector of $\mathbf{x}$ containing the elements indexed by $S$, while $\mathbf{A}_S$represents the $\mathbf{M}\times|\,\mathbf{S}\,|$submatrix of $\mathbf{A}$formed by the corresponding columns. Furthermore, define $\mathcal{D}_A(k) = \{a_{i,j}: |i-j| = k\}$. as the collection of coefficients located on the $k-$th (pairs of) diagonals of the matrix $\mathbf{A}$, and $\mathcal{B}_h(\mathbf{A}) = (a_{i,j}\,\mathbf{1}_{\{|i-j|\leq h\}})$, as the banded representation of $\mathbf{A}$ with bandwidth $h$.

**S2**

**Algorithm S1 SPLASH($\alpha$, $\lambda$) implementation (MATLAB):** Algorithm S1 outlines the computational procedure for implementing the SPLASH $(\alpha,\lambda)$estimator in high-dimensional KRW exchange-rate systems. Step 1 (Sample autocovariance estimation). Compute the $(N\times N)$ sample autocovariance matrices $\widehat{\mathbf{\Sigma}_0} = \frac{\sum_{t=2}^{T}\mathbf{x}_t\mathbf{x}_t'}{T}$ and $\widehat{\mathbf{\Sigma}_1} = \frac{\sum_{t=2}^{T}\mathbf{x}_t\mathbf{x}_{t-1}'}{T}$. Step 2 (Bandwidth selection). Determine the bandwidth parameter $h$ using a bootstrap-based procedure as described in Gao et al. (2016, p. 7), adapted to the temporal dependence structure of exchange-rate returns. Step 3 (Banded covariance construction). Using the selected bandwidth from Step 2, compute the banded estimators $\mathfrak{B}_h(\widehat{\mathbf{\Sigma}_0})$ and $\mathfrak{B}_h(\widehat{\mathbf{\Sigma}_1})$. Construct the quantities $\widehat{\boldsymbol{\sigma}_h}$, $\widehat{\mathbf{V}_h^{(d)}}$, and the index set $\mathcal{G}$ as defined in Section 3.2.1. Step 4 (Hyperparameter grid specification). Specify a grid of candidate values for the tuning parameters $\alpha$ and $\lambda$. In practical applications, a representative grid for $\alpha$ may be chosen as $\alpha_{grid} = \{0,0.25,0.5,0.75,1\}$, while the regularization parameter $\lambda$ is selected from a logarithmically spaced grid $\{\lambda_1, \dots, \lambda_K\}$ between $\lambda_{max}$ and $\lambda_{min} = \eta\lambda_{max}$ for some small $\eta > 0$, where $\lambda_{max} = \max\{\max_{g\in\mathcal{G}}\frac{\left\|\widehat{\mathbf{V}_{h,g}^{(d)}}\widehat{\boldsymbol{\sigma}_h}\right\|_2}{T\sqrt{|g|}}, \max_{1\leq i\leq N_c}\frac{\left\|\widehat{\mathbf{V}_{h,1}^{(d)}}\widehat{\boldsymbol{\sigma}_h}\right\|_2}{T}\}$. Step 5 (Optimization and model selection). For each candidate pair $(\alpha,\lambda)$, compute the SPLASH solution $\hat{\mathbf{c}}(\alpha,\lambda)$ by minimizing the objective function in (8) using a sparse group lasso (SGL) algorithm. The optimal hyperparameters are selected via time-series cross-validation that preserves the temporal ordering of the KRW exchange-rate data (Hyndman and Athanasopoulos, 2018).

**S3**

**Auxiliary Result 1:** The proofs of the **Auxiliary Results** reported in this appendix are provided in the supplementary material. **Auxiliary Result** 1 defines the quantities $\mathbf{N_c} = |\mathbf{c}|$, $S = \{j: c_j \neq 0\}$, $\mathcal{G}_S = \{g\in\mathcal{G}: \mathbf{c_g} \neq \mathbf{0}\}$, $\mathcal{G}_S^c = \{g\in\mathcal{G}: \mathbf{c_g} = \mathbf{0}\}$, and $\overline{\omega_\alpha} = \max\{(1-\alpha)\sum_{g\in\mathcal{G}_S}\sqrt{|g|}, \alpha\sqrt{|S|}\}$. For any $\mathbf{x}\in\mathbb{R}^{N_c}$, **Auxiliary Results** 1 also defines $\mathbf{x}\in\mathfrak{L}_{N_c}(\mathcal{G},S) := \{\mathbf{x}\in\mathbb{R}^{N_c}: \boldsymbol{\mathcal{P}}_{\alpha,S^c} \leq 3\boldsymbol{\mathcal{P}}_{\alpha,S}(\mathbf{x})\}$, where $\boldsymbol{\mathcal{P}}_{\alpha,S^c}(\mathbf{x}) = (1-\alpha)\mathbf{\Sigma}_{g\in\mathcal{G}_S}\sqrt{|g|}\left\|\mathbf{x}_g\right\|_2 + \alpha\|\mathbf{x}_{S^c}\|_1$, and $\boldsymbol{\mathcal{P}}_{\alpha,S}(\mathbf{x}) = (1-\alpha)\mathbf{\Sigma}_{g\in\mathcal{G}_S}\sqrt{|g|}\left\|\mathbf{x}_g\right\|_2 + \alpha\|\mathbf{\Delta}_S\|_1$. Then, under **Condition** 4, it holds that $\min_{\mathbf{x}\in\mathfrak{L}(\mathcal{G}_S,S)}\frac{\overline{\omega_\alpha}\|\mathbf{V}^{(d)}\mathbf{x}\|_2}{\boldsymbol{\mathcal{P}}_{\alpha,S}(\mathbf{x})} \geq \frac{\phi_0}{2}$. This result establishes a compatibility requirement between the design matrix $V^{(d)}$ and the structured sparsity pattern induced by **Condition** 3 and 4. In the

context of KRW-centered exchange-rate systems, the **Auxiliary Result** ensures that economically meaningful groups of coefficients—corresponding to admissible diagonals of $\mathbf{A}$ and $\mathbf{B}$—are sufficiently identifiable when restricted to the relevant sparse subspace.

**Auxiliary Result 2:** We define the event $\mathcal{V}(x) := \{\left\|\widehat{\mathbf{V}}_h^{(d)}\mathbf{V}\right\|_2 \le x\}$ under **Condition** 4, it holds on the event $\mathcal{V}(\frac{\phi_0}{4})$ that $\min_{\mathbf{x}\in\mathcal{Q}(\mathcal{G}_S,S)} \frac{\overline{\omega_\alpha}\left\|\widehat{\mathbf{V}^{(d)}}\mathbf{x}\right\|_2}{\mathcal{P}_{\alpha,S}(\mathbf{x})} \ge \frac{\phi_0}{4}$. This result shows that the restricted eigenvalue property established at the population level remains valid for the banded sample estimator $\widehat{\mathbf{V}}_h^{(d)}$ with high probability. In the context of high-dimensional KRW exchange-rate systems, the **Auxiliary Result** guarantees that the structured sparsity induced by economically localized cross-currency interactions is preserved when replacing population quantities with their sample-based counterparts. This stability property is essential for deriving finite-sample guarantees for the SPLASH estimator.

**Auxiliary Result 3**: We define the $N(p+1)$ -dimensional vector $\boldsymbol{\xi}_t = (\mathbf{ex}_t', \mathbf{ex}_{t-1}', \ldots, \mathbf{ex}_{t-p}')'$. For each $j = 1, \ldots, p$, the elements of the sample autocovariance matrix $\widehat{\boldsymbol{\Sigma}}_j = \frac{\sum_{t=p+1}^{T}\xi_{t,i}\xi_{t,j}}{T}$ admit the representation after a suitable choice of the indices $(i, j)$. (a) If **Conditions** 1 and 2(b1) (polynomial tail behaviour) are satisfied, then for some $0 < \delta < 1$, $\mathbb{P}(\left|\sum_{t=p+1}^{T}\xi_{t,i}\xi_{t,j} - \mathbb{E}(\xi_{t,i}\xi_{t,j})\right| > T\epsilon) \le [b_1\mathbf{T}^{\frac{1-\delta}{3}} + \frac{b_3}{\epsilon}]\exp(-\frac{\mathbf{T}^{\frac{1-\delta}{3}}}{2b_1^2}) + \frac{b_2}{\epsilon^2\mathbf{T}^{\frac{\delta(d-1)}{2}}}$. (b) If **Conditions** 1 and 2 (subexponential tail behaviour) are satisfied, then $\mathbb{P}(\left|\sum_{t=p+1}^{T}\xi_{t,i}\xi_{t,j} - \mathbb{E}(\xi_{t,i}\xi_{t,j})\right| > T\epsilon) \le [\frac{k_1}{\epsilon} + \frac{2(T\epsilon^2)^{\frac{1}{7}}}{k_2}]\exp(-\frac{(T\epsilon^2)^{\frac{1}{7}}}{k_3})$. **Auxiliary Result** 3 establishes probabilistic bounds for the deviation of sample autocovariance matrices from their population counterparts under alternative tail requirements on the innovation process. In high-dimensional KRW exchange-rate systems, these bounds ensure that sample-based estimates of cross-currency dependence converge sufficiently fast to their population analogues, thereby supporting the error bounds derived for banded autocovariance estimation and the subsequent theoretical guarantees of the SPLASH estimator. Explicit expressions for the constants $b_1$, $b_2$, $b_3$, $k_1$, $k_2$, and $k_3$ are provided in the proof.

**S4**

**Proof of Proposition 1:** The proof proceeds by establishing a sequence of intermediate results, reported in parts (a)–(e). These results characterize the approximation and bandedness properties of the matrices arising in the reduced-form representation of the KRW exchange-rate system. **Proposition** 1 then follows by combining these components. (a) Consider the matrix $\hat{\mathbf{C}}_s = (\sum_{j=0}^{s-1}\mathbf{A}^j)\mathbf{B} =: \widehat{\mathbb{D}}_s\mathbf{B}$. Owing to the banded structure imposed on the contemporaneous interaction matrix $\mathbf{A}$ and the lagged matrix $\mathbf{B}$, $\hat{\mathbf{C}}_s$ inherits a finite interaction range. In particular, $\hat{\mathbf{C}}_s$ has a maximum bandwidth of $(s+1)(k_0-1)+1$. Moreover, the deviation between $\hat{\mathbf{C}}_s$ and the reduced-form coefficient matrix $\mathbf{C}$ can be controlled in operator norm, yielding $\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash} \le \delta_C\delta_A^s$. This bound reflects the fact that higher-order contemporaneous spillovers decay geometrically in the KRW exchange-rate network as the interaction order $s$ increases. (b) Define $\boldsymbol{\Sigma}_0^{r,s} = \sum_{j=0}^{r}\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s\boldsymbol{\Sigma}_\epsilon\widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j$, with $\hat{\mathbf{C}}_s$ as in part (a). The matrix $\boldsymbol{\Sigma}_0^{r,s}$ is banded, with bandwidth no larger than $2(rs+r+s)(k_0-1)+2\ell_0+1$. In addition, its approximation error relative to the population autocovariance matrix $\boldsymbol{\Sigma}_0$satisfies $\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_{\vdash} \le \frac{8C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{C_\epsilon\delta_C^{2(r+1)}}{(1-\delta_A)^2(1-\delta_C^2)}$, whenever $s$ is sufficiently large such that $\delta_C(1+\delta_A^s) < 1$. This result formalizes the approximate bandedness of the contemporaneous covariance structure implied by economically localized interactions among KRW bilateral exchange rates. (c) Let $\boldsymbol{\Sigma}_1^{r,s} = \sum_{j=0}^{r}\hat{\mathbf{C}}_s\boldsymbol{\Sigma}_0^{r,s}$, where $\hat{\mathbf{C}}_s$ and $\boldsymbol{\Sigma}_0^{r,s}$ are defined as above. The matrix $\boldsymbol{\Sigma}_1^{r,s}$ is also banded, with bandwidth no larger than $2(rs+2r+3s+1)(k_0-1)+2\ell_0+1$. Furthermore, its deviation from the population lag-one autocovariance matrix $\boldsymbol{\Sigma}_1$ is bounded by $\|\boldsymbol{\Sigma}_1^{r,s} - \boldsymbol{\Sigma}_1\|_{\vdash} \le \frac{9C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{C_\epsilon\delta_C^{2(r+1)}}{(1-\delta_A)^2(1-\delta_C^2)}$, provided that $s$ is large enough to ensure $\delta_C(1+\delta_A^s) <$

1. (d) Let $h_1 \geq 2(rs + r + s)(k_0 - 1) + 2\ell_0 + 1$. Consider the banded estimator $\mathfrak{B}_{h_1}(\widehat{\boldsymbol{\Sigma}_0})$ of the sample contemporaneous autocovariance matrix. Under this choice of bandwidth, the deviation between the banded estimator and the population matrix $\boldsymbol{\Sigma}_0$ admits the bound $\left\|\mathfrak{B}_{h_1}(\widehat{\boldsymbol{\Sigma}_0}) - \boldsymbol{\Sigma}_0\right\|_2 \leq \frac{16C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{2C_\epsilon\delta_C^{2(r+1)}}{(1-\delta_A)^2(1-\delta_C^2)}$, with probability at least $1 - \boldsymbol{\mathcal{P}}_1(\varepsilon, N, T)$ in the case of polynomial tail decay, or $1 - \boldsymbol{\mathcal{P}}_2(\varepsilon, N, T)$ in the case of exponential tail decay. This result shows that, once the bandwidth exceeds the effective interaction range implied by the KRW exchange-rate network, banding the sample autocovariance matrix preserves consistency while controlling the estimation error in operator norm. (e) Let $h_2 \geq (2rs + 2r + 3s)(k_0 - 1) + 2\ell_0 + 1$. For the lag-one autocovariance matrix, the corresponding banded estimator $\mathfrak{B}_{h_2}(\widehat{\boldsymbol{\Sigma}_1})$ satisfies $\left\|\mathfrak{B}_{h_2}(\widehat{\boldsymbol{\Sigma}_1}) - \boldsymbol{\Sigma}_1\right\|_2 \leq \frac{18C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{2C_\epsilon\delta_C^{2(r+1)}}{(1-\delta_A)^2(1-\delta_C^2)}$, with probability at least $1 - \boldsymbol{\mathcal{P}}_1(\varepsilon, N, T)$under polynomial tail behaviour, or $1 - \boldsymbol{\mathcal{P}}_2(\varepsilon, N, T)$ under exponential tail behaviour. This bound establishes that the banded estimator of the lagged autocovariance matrix achieves the same order of approximation accuracy as in part (d), provided that the bandwidth is chosen sufficiently large relative to the admissible interaction range of the KRW exchange-rate system. Taken together, parts (a)-(e) confirm that both the reduced-form coefficients and the relevant autocovariance matrices associated with KRW bilateral exchange rates can be approximated by banded objects with explicitly controlled errors and high probability. These results complete the proof of **Proposition** 1 and provide the theoretical foundation for employing banded covariance estimation in high-dimensional KRW exchange-rate analysis.

Closed-form representations of the constants $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, together with the restriction $0 \leq \delta_C < 1$, are established in the subsequent derivations. (a) The argument relies on foundational results for banded vector autoregressive processes developed by Guo et al. (2016), which are particularly relevant for high-dimensional exchange-rate systems. Recall that the reduced-form matrix satisfies $\mathbf{C} = \mathbf{DB}$, where $\mathbf{D} = (\mathbf{I}_N - \mathbf{A})^{-1}$. Its truncated series approximation, $\widehat{\mathbf{D}}_s = \sum_{j=0}^{s-1}\mathbf{A}^j$, possesses a bandwidth bounded above by $s(k_0 - 1) + 1$. Moreover, the discrepancy between $\widehat{\mathbf{D}}_s$ and $\mathbf{D}$ can be bounded in operator norm as $\left\|\widehat{\mathbf{D}}_s - \mathbf{D}\right\|_\vdash = \left\|\sum_{j=0}^{s-1}\mathbf{A}^j - (\mathbf{I}_N - \mathbf{A})^{-1}\right\|_\vdash = \left\|-\sum_{j=s}^{\infty}\mathbf{A}^j\right\|_\vdash \leq \sum_{j=s}^{\infty}\|\mathbf{A}\|_\vdash^j \leq \frac{\delta_A^s}{1-\delta_A}$. The matrix product $\hat{\mathbf{C}}_s = \widehat{\mathbf{D}}_s\mathbf{B}$ therefore exhibits a maximal bandwidth equal to $(s + 1)(k_0 - 1) + 1$, reflecting the propagation of dynamic interactions across KRW exchange-rate components. Given that $\|\mathbf{B}\|_\vdash \leq \delta_B$and that $\frac{\delta_B}{1-\delta_A} =: \delta_C < 1$ holds under **Condition** 3, it follows that $\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_\vdash = \left\|(\widehat{\mathbf{D}}_s - \mathbf{D})\mathbf{B}\right\|_\vdash \leq \left\|\widehat{\mathbf{D}}_s - \mathbf{D}\right\|_\vdash\|\mathbf{B}\|_\vdash \leq \delta_C\delta_A^s$.

(b) Building on the implication stated in footnote 7, it follows that the bandwidth of $\hat{\mathbf{C}}_s^r$ is bounded above by $r[(s + 1)(k_0 - 1) + 1] - (r - 1) = r(s + 1)(k_0 - 1) + 1$. Consequently, the bandwidth of $\boldsymbol{\Sigma}_0^{r,s}$ does not exceed $2[r(s + 1)(k_0 - 1) + 1] + 2s(k_0 - 1) + 1 + (2\ell_0 + 1) - 4 = 2(rs + r + s)(k_0 - 1) + 2\ell_0 + 1$. The next step establishes an upper bound for $\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_1$. Noting that $\boldsymbol{\Sigma}_0 = \sum_{j=0}^{\infty}\hat{\mathbf{C}}_s^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j$, one obtains $\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_1 \leq \left\|\sum_{j=0}^{r}[\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s\boldsymbol{\Sigma}_\epsilon\widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j - \hat{\mathbf{C}}_s^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j]\right\|_\vdash + \left\|\sum_{j=r+1}^{\infty}\hat{\mathbf{C}}_s^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j\right\|_\vdash$ (D-1), where the first term captures approximation error induced by truncation and banding, while the second term reflects tail decay in the dynamic propagation of KRW exchange-rate shocks. The leading term on the right-hand side of (D-1) can be further decomposed as $\sum_{j=0}^{r}[\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s\boldsymbol{\Sigma}_\epsilon\widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j - \hat{\mathbf{C}}_s^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j] = \sum_{j=0}^{r}\{(\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D})\boldsymbol{\Sigma}_\epsilon\widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j - \mathbf{C}^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon[\mathbf{D}'(\mathbf{C}')^j - \widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_S')^j]\} = \sum_{j=0}^{r}\{(\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D})\boldsymbol{\Sigma}_\epsilon[\widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_S')^j - \mathbf{D}'(\mathbf{C}')^j]\} + \sum_{j=0}^{r}\{(\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D})\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j - \mathbf{C}^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon[\mathbf{D}'(\mathbf{C}')^j - \widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j]\}$ . As a result, the following bound is obtained: $\left\|\sum_{j=0}^{r}[\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s\boldsymbol{\Sigma}_\epsilon\widehat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j - \hat{\mathbf{C}}_s^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j]\right\|_\vdash \leq \mathbf{c}_\epsilon\sum_{j=0}^{r}\left\|\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D}\right\|_\vdash^2 + 2\mathbf{C}\sum_{j=0}^{r}\left\|\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D}\right\|_\vdash\left\|\mathbf{C}^j\mathbf{D}\right\|_\vdash$ (D-2), where the constant $\mathbf{c}_\epsilon$ depends on the innovation covariance structure relevant for KRW exchange-rate disturbances. Finally, observe that the approximation error satisfies $\sum_{j=0}^{r}\left\|\hat{\mathbf{C}}_s^j\widehat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D}\right\|_\vdash \leq \left\|\hat{\mathbf{C}}_s^j - \mathbf{C}^j\right\|_\vdash\left\|\widehat{\mathbf{D}}_s\right\|_\vdash + \left\|\mathbf{C}^j\right\|_\vdash\left\|\mathbf{D} - \widehat{\mathbf{D}}_s\right\|_\vdash$ (D-3).

which separates the contribution of dynamic truncation from that of static inverse approximation. This decomposition is particularly suitable for analysing error propagation in high-dimensional KRW exchange-rate systems, where both contemporaneous and lagged interactions must be controlled simultaneously. The analysis proceeds by bounding each component appearing in (B.3). To this end, expanding the relevant matrix powers yields $\left\|\hat{\mathbf{C}}_s^j - \mathbf{C}^j\right\|_{\vdash} = \left\|(\hat{\mathbf{C}}_s - \mathbf{C} + \mathbf{C})^j - \mathbf{C}^j\right\|_{\vdash} = \left\|[(\hat{\mathbf{C}}_s - \mathbf{C})^j + (\hat{\mathbf{C}}_s - \mathbf{C})^{j-1}\mathbf{C} + (\hat{\mathbf{C}}_s - \mathbf{C})^{j-2}\mathbf{C}(\hat{\mathbf{C}}_s - \mathbf{C}) + \ldots + \mathbf{C}(\hat{\mathbf{C}}_s - \mathbf{C})^{j-1} \ldots + \mathbf{C}^j] - \mathbf{C}^j\right\|_{\vdash} \le \sum_{k=1}^{j}\binom{j}{k}\|\mathbf{C}^j\|_{\vdash}^{j-k}\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash}^{k} = \sum_{k=0}^{j-1}\binom{j}{k+1}\|\mathbf{C}^j\|_{\vdash}^{j-1-k}\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash}^{k+1} = \left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash}\sum_{k=0}^{j-1}\frac{j}{k+1}\binom{j-1}{k}\|\mathbf{C}^j\|_{\vdash}^{j-1-k}\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash}^{k} \le \delta_C\delta_A^s j[\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash} + \|\mathbf{C}\|_{\vdash}]^{j-1} \le \delta_C\delta_A^s j[\delta_C(1+\delta_A^s)]^{j-1} = \delta_C^j j\delta_A^s(1+\delta_A^s)^{j-1}$, which follows from the uniform bounds imposed on $\|\mathbf{C}\|_{\vdash}$ and $\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash}$ in the KRW exchange-rate system. Next, it is noted that $\left\|\mathbf{D} - \hat{\mathbf{D}}_s\right\|_{\vdash} = \left\|\sum_{j=s}^{\infty}\mathbf{A}^j\right\|_{\vdash} \le \frac{\delta_A^s}{1-\delta_A}$, $\left\|\hat{\mathbf{D}}_s\right\|_{\vdash} \le \frac{1}{1-\delta_A}$ and $\left\|\mathbf{C}^j\right\|_{\vdash} \le \delta_C^j$, so that inequality (D-3) can be combined with these bounds to yield $\left\|\hat{\mathbf{C}}_s^j\hat{\mathbf{D}}_s - \mathbf{C}^j\mathbf{D}\right\|_{\vdash} \le \frac{\delta_A^s j\delta_C^j(1+\delta_A^s)^j}{(1+\delta_A^s)(1-\delta_A)} + \frac{\delta_A^s\delta_C^j}{1-\delta_A} \le \frac{2\delta_A^s j\delta_C^j(1+\delta_A^s)^j}{1-\delta_A}$ (D-4). This bound quantifies the combined effect of truncation and dynamic propagation in the approximation of high-dimensional KRW exchange-rate interactions. Substituting the bound in (D-4) into inequality (D-2) yields $\left\|\sum_{j=0}^{r}[\hat{\mathbf{C}}_s^j\hat{\mathbf{D}}_s\boldsymbol{\Sigma}_\epsilon\hat{\mathbf{D}}_s'(\hat{\mathbf{C}}_s')^j - \mathbf{C}^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j]\right\|_{\vdash} \le \frac{4C_\epsilon\delta_A^{2s}}{(1-\delta_A)^2}\sum_{j=0}^{r} j^2\,\delta_C^{2s}(1+\delta_A^s)^{2j} + \frac{4C_\epsilon\delta_A^s}{(1-\delta_A)^2}\sum_{j=0}^{r} j\,\delta_C^{2s}(1+\delta_A^s)^j \le \frac{4C_\epsilon\delta_A^{2s}(1+\delta_A^s)^2}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{4C_\epsilon\delta_A^s(1+\delta_A^s)}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)]^2} \le \frac{8C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2}$ (D-5), where the two summation terms reflect, respectively, second-order and first-order accumulation of approximation errors along the dynamic propagation paths of the KRW exchange-rate system. The above inequalities rely on the condition that the truncation level s is sufficiently large so that $\delta_C(1+\delta_A^s) < 1$, standard results for geometric series can be applied, namely $\sum_{j=0}^{\infty} jz^j = \frac{z}{(1-z)^2}$ and $\sum_{j=0}^{\infty} j^2z^j = \frac{z+1}{(1-z)^3}$ for $|z| < 1$. Attention is next turned to the second term on the right-hand side of (D-1). This term satisfies $\left\|\sum_{j=0}^{\infty}\mathbf{C}^j\mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}'(\mathbf{C}')^j\right\|_{\vdash} \le \|\boldsymbol{\Sigma}_\epsilon\|_{\vdash}\|\mathbf{D}\|_{\vdash}^2\sum_{j=r+1}^{\infty}\|\mathbf{C}\|_{\vdash}^{2j} \le \frac{C_\epsilon}{(1-\delta_A)^2}\sum_{j=r+1}^{\infty}\delta_C^{2j} = \frac{C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)}$ (D-6). Combining bounds (D-5) and (D-6) with inequality (D-1) yields the final estimate $\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_{\vdash} \le \frac{8C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)}$, which establishes the intermediate bound stated in part (b).

(c) By construction, the first-lag autocovariance matrix satisfies $\boldsymbol{\Sigma}_1 = (\mathbf{I}_N - \mathbf{A})^{-1}\mathbf{B}\boldsymbol{\Sigma}_0 = \mathbf{C}\boldsymbol{\Sigma}_0$, which implies that the approximation error for the lag-one covariance matrix can be decomposed as $\|\boldsymbol{\Sigma}_1^{r,s} - \boldsymbol{\Sigma}_1\|_{\vdash} = \left\|\hat{\mathbf{C}}_S(\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0) + (\hat{\mathbf{C}}_s - \mathbf{C})\boldsymbol{\Sigma}_0\right\|_{\vdash} \le \left\|\hat{\mathbf{C}}_S\right\|_{\vdash}\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_{\vdash} + \left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash}\boldsymbol{\Sigma}_0 \le \frac{8C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)} + \frac{C_\epsilon\delta_C\delta_A^s}{(1-\delta_A)^2(1-\delta_C^2)} \le \frac{9C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)}$ (D-7). The above inequalities make use of the intermediate result in part (a), which establishes that $\left\|\hat{\mathbf{C}}_s - \mathbf{C}\right\|_{\vdash} \le \delta_C\delta_A^s$. Moreover, since the truncation parameter s is chosen sufficiently large, it follows that $\left\|\hat{\mathbf{C}}_S\right\|_{\vdash} \le \delta_C(1+\delta_A^s) < 1$. Together, these requirements ensure that the approximation error of the lag-one covariance matrix in the KRW exchange-rate system remains uniformly controlled, thereby completing the argument for part (c). (d) Consider the banded covariance operator $\mathfrak{B}_{h_1}(\cdot)$. The deviation between the banded estimator applied to the sample covariance and the population covariance matrix satisfies

$\left\|\mathfrak{B}_{h_1}(\hat{\boldsymbol{\Sigma}}_0) - \boldsymbol{\Sigma}_0\right\|_{\vdash} \le \left\|\mathfrak{B}_{h_1}(\hat{\boldsymbol{\Sigma}}_0) - \mathfrak{B}_{h_1}\boldsymbol{\Sigma}_0\right\|_{\vdash} + \left\|\mathfrak{B}_{h_1}(\boldsymbol{\Sigma}_0) - \mathfrak{B}_{h_1}\boldsymbol{\Sigma}_0^{r,s}\right\|_{\vdash} + \left\|\mathfrak{B}_{h_1}(\boldsymbol{\Sigma}_0^{r,s}) - \boldsymbol{\Sigma}_0\right\|_{\vdash} = \left\|\mathfrak{B}_{h_1}(\hat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)\right\|_{\vdash} + \left\|\mathfrak{B}_{h_1}(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0^{r,s})\right\|_{\vdash} + \|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_{\vdash} \le \left\|\mathfrak{B}_{h_1}(\hat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)\right\|_{\vdash} + 2\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_{\vdash}$ (D-8). This inequality follows from the fact that $\boldsymbol{\Sigma}_0^{r,s}$ is itself banded, so the application of the banding operator cannot increase the operator-norm distance between $\boldsymbol{\Sigma}_0^{r,s}$ and $\boldsymbol{\Sigma}_0$. The two terms on the right-hand side of (D-8) are therefore examined separately. By symmetry of the covariance matrices, it holds that $\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\sigma}_0)\right\|_{\vdash} = \left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\sigma}_0)\right\|_1$.

Letting $\xi_t$ be defined as in **Auxiliary Result** 3, the following probability bound is obtained: $\mathbb{P}(\left\|\mathfrak{B}_{h_1}(\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)\right\|_{\vdash} \le x) = \mathbb{P}[\max_{i\le j\le N} \sum_{i=1\wedge(j-h_1)}^{NV(j-h_1)} \left|\sum_{t=2}^{T} \xi_{it}\xi_{jt} - \mathbb{E}(\xi_{it}\xi_{jt})\right| \le Tx] \ge 1 - \sum_{j=1}^{N} \mathbb{P}[\sum_{i=1\wedge(j-h_1)}^{NV(j-h_1)} \left|\sum_{t=2}^{T} \xi_{it}\xi_{jt} - \mathbb{E}(\xi_{it}\xi_{jt})\right| \ge Tx] \ge 1 - \sum_{j=1}^{N} \sum_{i=1\wedge(j-h_1)}^{NV(j-h_1)} \mathbb{P}\left|\sum_{t=2}^{T} \xi_{it}\xi_{jt} - \mathbb{E}(\xi_{it}\xi_{jt})\right| \ge \frac{Tx}{2h_1+1}]$ (D-9). Applying **Auxiliary Result** 3 yields the explicit probability bounds $\ge 1 - (2h_1+1)N[b_1 T^{\frac{1-\delta}{3}} + \frac{(2h_1+1)b_3}{x}] \exp[-\frac{T^{\frac{1-\delta}{3}}}{2b_1^2} + \frac{b_2(2h_1+1)^d}{x^d T^{\frac{\delta(d-1)}{2}}}]$ for polynomial tails and $\ge 1 - (2h_1+1)N[\frac{k_1(2h_1+1)}{x} + \frac{2}{k_2}[\frac{Tx^2}{(2h_1+1)^2}]^{\frac{1}{7}}] \exp[-\frac{1}{k_3}[\frac{Tx^2}{(2h_1+1)^2}]^{\frac{1}{7}}]$ for exponential tails. The probability bounds in (D-9) coincide with those defined as $1 - P_1(x, N, T)$ and $1 - P_2(x, N, T)$in **Proposition** 1. Consequently, if the event $\left\|\mathfrak{B}_{h_1}(\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)\right\|_{\vdash} \le \epsilon$ holds, then combining this result with the intermediate bound established in part (b) and applying inequality (D-8) yields $\left\|\mathfrak{B}_{h_1}(\widehat{\boldsymbol{\Sigma}}_0) - \boldsymbol{\Sigma}_0\right\|_{\vdash} \le \epsilon + 2\|\boldsymbol{\Sigma}_0^{r,s} - \boldsymbol{\Sigma}_0\|_{\vdash} \le \epsilon + \frac{16C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{2C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)}$. This establishes the desired bound for the banded covariance estimator in the KRW exchange-rate system and completes the argument for part (d). (e) Proceeding in the same manner as in part (d), an analogous bound is obtained for the banded estimator of the lag-one covariance matrix. Specifically, $\left\|\mathfrak{B}_{h_2}(\widehat{\boldsymbol{\Sigma}}_1) - \boldsymbol{\Sigma}_1)\right\|_{\vdash} \le \epsilon + 2\|\boldsymbol{\Sigma}_1^{r,s} - \boldsymbol{\Sigma}_1\|_{\vdash} \le \epsilon + \frac{16C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{2C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)}$, provided that the intermediate result in part (c) is applied and the event $\left\|\mathfrak{B}_{h_2}(\widehat{\boldsymbol{\Sigma}}_1) - \boldsymbol{\Sigma}_1)\right\|_{\vdash} \le \epsilon$ occurs. The probability of this event equals $1 - P_1(x, N, T)$ under polynomial tail behaviour and $1 - P_2(x, N, T)$ under exponential tail behaviour. All intermediate bounds are now combined to recover the statement of the proposition. Parts (d) and (e) are jointly applicable since the effective bandwidth satisfies $h = \max(h_1, h_2) = h_2 = (2rs + 2r + 3s + 1)(k_0 - 1) + 2\iota_0 + 1 \le [2r(s+1) + 3(s+1)](k_0 - 1) + 2\iota_0 + 1 \le (s+1)(2r+3)(k_0 - 1) + 2\iota_0 + 1$ (D-10). Using the intermediate results from parts (d) and (e), it follows that $\left\|\widehat{\mathbf{V}}_1 - \mathbf{V})\right\|_{\vdash} \le \left\|\mathfrak{B}_{h_2}(\widehat{\boldsymbol{\Sigma}}_1) - \boldsymbol{\Sigma}_1)\right\|_{\vdash} + \left\|\mathfrak{B}_{h_1}(\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)\right\|_{\vdash} \le 2\epsilon + \frac{34C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} + \frac{4C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)}$, with probabilities $1 - P_1(x, N, T)$ and $1 - P_2(x, N, T)$ under polynomial and exponential tail decay, respectively. It remains to select s and r such that $\delta_C(1+\delta_A^s) < 1$ and $\frac{34C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} = \epsilon$, $\frac{4C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)} = \epsilon$. First, the requirement $\delta_C(1+\delta_A^s) < 1$ can be ensured by choosing $s > \log\frac{\delta_C}{(K-\delta_C)\delta_A}$. Under this choice, $\frac{34C_\epsilon\delta_A^s}{(1-\delta_A)^2[1-\delta_C^2(1+\delta_A^s)^2]^2} = \frac{34C_\epsilon\delta_A^s}{(1-\delta_A)^2(1-K)^2} = \epsilon \Leftarrow s = \frac{\log\frac{34C_\epsilon}{(1-\delta_A)^2(1-K)^2\epsilon}}{|\log\delta_A|}$ (D-11). Similarly, the second equality yields $\frac{4C_\epsilon\delta_C^{2(1+r)}}{(1-\delta_A)^2(1-\delta_C^2)} = \epsilon \Leftarrow r = \frac{\log\frac{4C_\epsilon}{(1-\delta_A)^2(1-\delta_C^2)\epsilon}}{2|\log\delta_C|} - \frac{1}{2}$ (D-12). Then, we set $C_1 = \frac{34C_\epsilon}{(1-\delta_A)^2(1-K^2)^2}$ and $C_2 = \frac{4C_\epsilon}{(1-\delta_A)^2(1-\delta_C^2)}$. Substituting (D-11)–(D-12) into (D-10) shows that, for some $K \in (\delta_C, 1)$, $h = \{\max[\frac{\log\frac{K-\delta_C}{\delta_C}}{|\log\delta_A|}, \frac{\log\frac{C_1}{\epsilon}}{|\log\delta_A|}] + 1\}[\frac{\log\frac{C_1}{\epsilon}}{|\log\delta_C|} + 2](k_0 - 1) + 2\iota_0 + 1$, and consequently, $\left\|\widehat{\mathbf{V}}_1 - \mathbf{V}\right\|_{\vdash} \le 4\epsilon$ with probability at least $1 - 2P_1(x, N, T)$ under polynomial tails and $1 - 2P_2(x, N, T)$ under exponential tails.

Proof of **Proposition 2**.

The argument is based on fundamental properties of dual norms associated with composite penalties used in high-dimensional KRW exchange-rate systems. Recall that the penalty function is defined as $P_a(\mathbf{c}) = \alpha\sum_{g\in\mathcal{G}}\sqrt{|g|}\left\|\mathbf{c}_g\right\|_2 + (1-\alpha)\|\mathbf{c}\|_1$. Using standard properties of the $\|\cdot\|_1$ and $\|\cdot\|_2$ norms, it follows directly that $P_a(\cdot)$ satisfies the axioms of a norm for any $\alpha \in [0,1]$. For an arbitrary norm $\|\cdot\|^*$, its dual norm $\|c\|^*$ is defined by $\|\mathbf{c}\|^* = \sup_{\boldsymbol{x}\neq\mathbf{0}} \frac{|\mathbf{c}`\mathbf{x}|}{\|\mathbf{x}\|}$. The corresponding dual-norm inequality implies that $\mathbf{c}`\mathbf{x} \le \|\mathbf{c}\|^*\|\mathbf{x}\|$ (D-13) for all conformable vectors c and x. For the penalty $P_a(\mathbf{c})$, the associated dual norm

$P_\alpha^*(\mathbf{c})$ can be bounded as $P_\alpha^*(\mathbf{c}) = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{c}`\mathbf{x}|}{P_a(\mathbf{x})} \sup_{\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{c}`\mathbf{x}|}{\alpha\sum_{g\in\mathcal{G}}\sqrt{|g|}\|\mathbf{x}_g\|_2+(1-\alpha)\|\mathbf{x}\|_1} \leq \frac{|\mathbf{c}`\mathbf{x}|}{\alpha\sum_{g\in\mathcal{G}}\sqrt{|g|}\|\mathbf{x}_g\|_2} + (1-\alpha)\sup_{\mathbf{x}\neq\mathbf{0}}\frac{|\mathbf{c}`\mathbf{x}|}{\|\mathbf{x}\|_1} \leq \alpha\max_{g\in\mathcal{G}}\frac{\|\mathbf{c}_g\|_2}{\sqrt{|g|}} + (1-\alpha)\|\mathbf{c}\|_\infty$ (D-14). The first inequality exploits the convexity of the mapping $f(x) = x^{-1}$, while the second inequality follows from standard duality relations, namely $\|\mathbf{c}\|_1^* = \|\mathbf{c}\|_\infty$, together with the identity $\sup_{\mathbf{x}\neq\mathbf{0}}\frac{|\mathbf{c}`\mathbf{x}|}{\sum_{g\in\mathcal{G}}\sqrt{|g|}\|\mathbf{x}_g\|_2} = \sup_{\mathbf{x}\neq\mathbf{0},\sum_{g\in\mathcal{G}}\sqrt{|g|}\|\mathbf{x}_g\|_2=1}\left|\sum_{g\in\mathcal{G}}\mathbf{c}_g`\mathbf{x}_g\right| \leq \sup_{\mathbf{x}\neq\mathbf{0},\sum_{g\in\mathcal{G}}\sqrt{|g|}\|\mathbf{x}_g\|_2=1}\sum_{g\in\mathcal{G}}\left\|\frac{\mathbf{c}_g}{\sqrt{|g|}}\right\|_2 \leq \max_{g\in\mathcal{G}}\frac{\|\mathbf{c}_g\|_2}{\sqrt{|g|}}$. which completes the derivation of the bound for the dual norm $P_\alpha^*(\mathbf{c})$ in the high-dimensional KRW exchange-rate setting. The formal derivation now proceeds. Recall that $\widehat{\boldsymbol{\sigma}}_h = \mathrm{vec}[\mathfrak{B}_h(\widehat{\boldsymbol{\Sigma}}_1)`]$, $\widehat{\mathbf{V}}_h = [\mathfrak{B}_h(\widehat{\boldsymbol{\Sigma}}_1)`\ \mathfrak{B}_h(\widehat{\boldsymbol{\Sigma}}_0)]$, $\mathbf{V} = [\boldsymbol{\Sigma}_1`\ \boldsymbol{\Sigma}_0]$, and $\mathbf{C} = [\mathbf{A}\ \mathbf{B}]$. Invoking standard properties of the vectorization operator, it follows that $\widehat{\boldsymbol{\sigma}}_h - \widehat{\mathbf{V}}_h^{(d)}\mathbf{c} = \mathrm{vec}[\mathfrak{B}_h(\widehat{\boldsymbol{\Sigma}}_1)` - \widehat{\mathbf{V}}_h\mathbf{C}`] = \mathrm{vec}\{[\mathfrak{B}_h(\widehat{\boldsymbol{\Sigma}}_1) - \boldsymbol{\Sigma}_1]` - [\widehat{\mathbf{V}}_h - \mathbf{V}]\mathbf{C}` + [\boldsymbol{\Sigma}_1` - \mathbf{V}\mathbf{C}`]\}$. The last term vanishes by construction of the population Yule – Walker relation. Consequently, $\widehat{\boldsymbol{\sigma}}_h - \widehat{\mathbf{V}}_h^{(d)}\mathbf{c} = \mathrm{vec}[\mathfrak{B}_h(\widehat{\boldsymbol{\Sigma}}_1)` - \boldsymbol{\Sigma}_1`] - [\widehat{\mathbf{V}}_h^{(d)} - \mathbf{V}^{(d)}]\mathbf{c} = \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Delta}}_{\mathbf{V}}\mathbf{c}$ (D-15). Invoking relation (D-15), the following decomposition is obtained: $\left\|\widehat{\boldsymbol{\sigma}}_h - \widehat{\mathbf{V}}_h^{(d)}\mathbf{c}\right\|_2^2 = \left\|[\widehat{\boldsymbol{\sigma}}_h - \widehat{\mathbf{V}}_h^{(d)}\mathbf{c}] - \widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 = \left\|\widehat{\boldsymbol{\sigma}}_h - \widehat{\mathbf{V}}_h^{(d)}\mathbf{c}\right\|_2^2 + \left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 - 2(\hat{\mathbf{c}}-\mathbf{c})`\widehat{\mathbf{V}}_h^{(d)}`(\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Delta}}_{\mathbf{V}}\mathbf{c})$. Recalling the penalized objective function $\mathcal{L}_\alpha(\mathbf{c}:\lambda) = \left\|\widehat{\boldsymbol{\sigma}}_h - \widehat{\mathbf{V}}_h^{(d)}\mathbf{c}\right\|_2^2 + \lambda P_a(\mathbf{c})$, and noting that the construction of the estimator guarantees $\mathcal{L}_\alpha(\hat{\mathbf{c}}:\lambda) \leq \mathcal{L}_\alpha(\mathbf{c}:\lambda)$, the following inequality is obtained: $\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + \lambda P_a(\hat{\mathbf{c}}) \leq 2(\hat{\mathbf{c}}-\mathbf{c})`\widehat{\mathbf{V}}_h^{(d)}`(\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Delta}}_{\mathbf{V}}\mathbf{c}) + \lambda P_a(\mathbf{c}) \leq 2P_a(\hat{\mathbf{c}}-\mathbf{c})P_a^*[\widehat{\mathbf{V}}_h^{(d)}`(\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Delta}}_{\mathbf{V}}\mathbf{c})] + \lambda P_a(\mathbf{c}) \leq P_a(\hat{\mathbf{c}}-\mathbf{c})\{2P_a^*[\widehat{\mathbf{V}}_h^{(d)}`\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}}] + 2P_a^*[\widehat{\mathbf{V}}_h^{(d)}`\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}}]\mathbf{c}\} + \lambda P_a(\mathbf{c})$ (D-16). The preceding bounds rely on the dual-norm inequality given in (B.13), together with the triangle inequality for dual norms, which are invoked in the second and third inequalities, respectively. To facilitate subsequent probabilistic control of the error terms arising in the KRW exchange-rate system, introduce the following events: $\mathcal{H}_1(x) = \{2P_a^*[\widehat{\mathbf{V}}_h^{(d)}`\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}}] \leq x\}$ and $\mathcal{H}_2(x) = \{2P_a^*[\widehat{\mathbf{V}}_h^{(d)}`\widehat{\boldsymbol{\Delta}}_{\mathbf{V}}\mathbf{c}] \leq x\}$ (D-17). On the event $\mathcal{H}_1(\frac{\lambda}{4}) \cap \mathcal{H}_2(\frac{\lambda}{4})$, inequality (D-16) can be rescaled by a factor of two, which leads to the bound $2\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + 2P_a(\hat{\mathbf{c}}) \leq \lambda P_a(\hat{\mathbf{c}}-\mathbf{c}) + 2\lambda P_a(\mathbf{c})$ (D-18). The next step examines the structure of the penalty terms $P_a(\hat{\mathbf{c}})$ and $P_a(\hat{\mathbf{c}}-\mathbf{c})$. Applying the reverse triangle inequality yields $P_a(\hat{\mathbf{c}}) = \alpha\sum_{g\in\mathcal{G}}\sqrt{|g|}\left\|\hat{\mathbf{c}}_g\right\|_2 + (1-\alpha)\|\hat{\mathbf{c}}\|_1 \geq \alpha\sum_{g\in\mathcal{G}_S}\sqrt{|g|}\left[\left\|\mathbf{c}_g\right\|_2 - \left\|\hat{\mathbf{c}}_g - \mathbf{c}_g\right\|_2\right] + \alpha\sum_{g\in\mathcal{G}_{S^c}}\sqrt{|g|}\left[\left\|\hat{\mathbf{c}}_g\right\|_2 + (1-\alpha)[\|\mathbf{c}_S\|_1 - \|\hat{\mathbf{c}}_S - \mathbf{c}_S\|_1 + \left\|\hat{\mathbf{c}}_{S^c}\right\|_1] = P_{a,S}(\mathbf{c}) + P_{a,S^c}(\hat{\mathbf{c}}-\mathbf{c}) - P_{a,S}(\hat{\mathbf{c}}-\mathbf{c})$ (D-19). Here, the index sets $\mathcal{G}_S$ and $\mathcal{G}_{S^c}$ are defined as in **Auxiliary Result** 1. A direct decomposition further gives $P_a(\hat{\mathbf{c}}-\mathbf{c}) = \alpha\sum_{g\in\mathcal{G}}\sqrt{|g|}\left\|\hat{\mathbf{c}}_g - \mathbf{c}_g\right\|_2 + (1-\alpha)\|\hat{\mathbf{c}}_S - \mathbf{c}_S\|_1 + \alpha\sum_{g\in\mathcal{G}_{S^c}}\sqrt{|g|}\left[\left\|\hat{\mathbf{c}}_g\right\|_2 + (1-\alpha)\|\hat{\mathbf{c}}_S\|_1 = P_{a,S}(\hat{\mathbf{c}}-\mathbf{c}) + P_{a,S^c}(\hat{\mathbf{c}}-\mathbf{c})$. Combining inequalities (D-18)–(D-20) leads directly to $2\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + 2\lambda P_{a,S^c}(\hat{\mathbf{c}}-\mathbf{c}) \leq 3\lambda P_{a,S}(\hat{\mathbf{c}}-\mathbf{c})$ (D-21), which implies that the estimation error vector $\hat{\mathbf{c}}-\mathbf{c}$ satisfies the cone-type restriction and therefore belongs to the set $\ell_{N_c}(\mathcal{G},S)$ defined in **Auxiliary Result** 1.

In conjunction with **Auxiliary Result** 2, it is required that the event $\mathcal{H}_1(\frac{\lambda}{4}) \cap \mathcal{H}_2(\frac{\lambda}{4}) \cap v(\frac{\phi_0}{2})$ holds. Under this requirement, one obtains $2\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + 2\lambda P_a(\hat{\mathbf{c}}-\mathbf{c}) = 2\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + 2\lambda P_{a,S^c}(\hat{\mathbf{c}}-\mathbf{c}) + \lambda P_{a,S}(\hat{\mathbf{c}}-\mathbf{c}) \leq 4\lambda P_{a,S}(\hat{\mathbf{c}}-\mathbf{c}) \leq 16\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2(\frac{\widehat{\omega}_\alpha\lambda}{\phi_0}) \leq \left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + \frac{64\widehat{\omega}_\alpha^2\lambda^2}{\phi_0^2}$ (D-22), where step $4\lambda P_{a,S}(\hat{\mathbf{c}}-\mathbf{c})$ follows from inequality (D-21), step $\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2$ is implied by $P_{a,S}(\hat{\mathbf{c}}-\mathbf{c}) \leq \frac{4\widehat{\omega}_0\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2}{\phi_0}$ for $\hat{\mathbf{c}}-\mathbf{c} \in \ell_{N_c}(\mathcal{G},S)$ as established in **Auxiliary Result** 2, and step $\left\|\widehat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}}-\mathbf{c})\right\|_2^2 + \frac{64\widehat{\omega}_\alpha^2\lambda^2}{\phi_0^2}$ applies the elementary inequality $16\mu\upsilon \leq \mu^2 + 64\nu^2$, which is equivalent to $(\mu - 8\nu)^2 \geq 0$. A direct rearrangement of inequality (D-22) yields the desired bound stated in **Proposition** 2. It remains to

establish a lower bound for the probability of the event $\mathcal{H}_1(\frac{\lambda}{4}) \cap \mathcal{H}_2(\frac{\lambda}{4}) \cap \mathrm{v}(\frac{\phi_0}{2})$. To this end, the following elementary inequality is employed to control the individual components: $\mathbb{P}[\mathcal{H}_1(\frac{\lambda}{4}) \cap \mathcal{H}_2(\frac{\lambda}{4}) \cap \mathrm{v}(\frac{\phi_0}{2})] \geq 1 - \mathbb{P}[\mathcal{H}_1(\frac{\lambda}{4})^c - \mathbb{P}[\mathcal{H}_2(\frac{\lambda}{4})^c] - \mathbb{P}[\mathrm{v}(\frac{\phi_0}{2})^2]$ . Consider first the probability associated with $\mathcal{H}_1(\frac{\lambda}{4})$. By definition, $\mathbb{P}[\mathcal{H}_1(\frac{\lambda}{4})^c] = \mathbb{P}\{2\mathrm{P}_\alpha^*[\widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} \grave{} \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}}] > \frac{\lambda}{4}\} \leq \mathbb{P}(\left\| \widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} \grave{} \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} \right\|_\infty > \frac{\lambda}{4})$. The last inequality follows directly from the bound implied by (D-14), which yields $\mathrm{P}_\mathrm{a}(\mathbf{c}) \leq \alpha \max_{\mathrm{g} \in \mathcal{G}} \frac{\|\mathbf{c}_\mathrm{g}\|_2}{\sqrt{|\mathrm{g}|}} + (1-\alpha)\|\mathbf{c}\|_\infty \leq \alpha \max_{\mathrm{g} \in \mathcal{G}} \frac{\sqrt{|\mathrm{g}|}\|\mathbf{c}_\mathrm{g}\|_2}{\sqrt{|\mathrm{g}|}} + (1-\alpha)\|\mathbf{c}\|_\infty = \|\mathbf{c}\|_\infty$ (D-23) for any conformable vector $\mathbf{c}$. Next, observe that $\mathbb{P}(\left\| [\widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} \grave{} \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}}] \right\|_\infty > \frac{\lambda}{8}) = \mathbb{P}(\left\| \{[\widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} - \mathbf{V}^{(\mathrm{d})}] + \mathbf{V}^{(\mathrm{d})}\} \grave{} \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} \right\|_\infty > \frac{\lambda}{8}) \leq \mathbb{P}(\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_1 \left\| \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} \right\|_\infty + \mathbf{C}_\mathbf{V} \left\| \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} \right\|_\infty > \frac{\lambda}{8}) \leq \mathbb{P}(\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash^2 + \mathbf{C}_\mathbf{V} \left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda}{8}) \leq \mathbb{P}(\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda^{\frac{1}{2}}}{4}) + \mathbb{P}(\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda}{16\mathbf{C}_\mathbf{V}})$ (D-24), where the block-diagonal structure of $\widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} - \mathbf{V}^{(\mathrm{d})}$ and $\mathbf{V}^{(\mathrm{d})}$ is exploited to obtain $\left\| \widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} - \mathbf{V}^{(\mathrm{d})} \right\|_1 \max_{1 \leq \mathrm{i} \leq \mathrm{N}} \left\| \widehat{\mathbf{V}}_{\mathrm{i,h}} - \mathbf{V}_\mathrm{i} \right\|_1 \leq \left\| \widehat{\mathbf{V}}_{\mathrm{i,h}} - \mathbf{V} \right\|_1$, together with the bound $\left\| \mathbf{V}^{(\mathrm{d})} \right\|_1 \leq \|\mathbf{V}\|_1 \leq \mathbf{C}_\mathbf{V}$, as explicitly assumed in **Proposition** 2. Probability bounds for the final right-hand-side terms in (D-24) follow directly from **Proposition** 1. Second, consider the complement event associated with $\mathcal{H}_2(\frac{\lambda}{4})$. Its probability can be bounded as $\mathbb{P}[\mathcal{H}_2(\frac{\lambda}{4})^c] = \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} \grave{} \widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}} \mathbf{c} \right\|_\infty > \frac{\lambda}{8}\} \leq \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} \grave{} [\widehat{\mathbf{V}}_{\mathrm{h}}^{(\mathrm{d})} - \mathbf{V}^{(\mathrm{d})}] \right\|_\infty > \frac{\lambda}{8\|\mathbf{c}\|_\infty}\} \leq \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_1 \left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\infty + \mathbf{C}_\mathbf{V} \left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\infty > \frac{\lambda}{8\|\mathbf{c}\|_\infty}\} \leq \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda^{\frac{1}{2}}}{4}\} + \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda}{16\mathbf{C}_\mathbf{V}}\}$ (D-25). The final inequality follows from an application of the union bound together with the fact that $\|\mathbf{c}\|_\infty < 1$, which is guaranteed by **Condition** 1. Consequently, the complement events $\mathcal{H}_1(\frac{\lambda}{4})^c$ and $\mathcal{H}_2(\frac{\lambda}{4})^c$ are governed by identical probability bounds. Finally, consider the complement event associated with $\mathcal{v}(\frac{\phi_0}{2})^c$ . By definition, $\mathbb{P}\{\mathcal{v}(\frac{\phi_0}{2})^c\} = \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_1 > \frac{\phi_0}{2}\} \leq \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\phi_0}{2}\}$. Combining this bound with the preceding probability controls for $\mathcal{H}_1(\frac{\lambda}{4})$ and $\mathcal{H}_2(\frac{\lambda}{4})$ yields $\mathbb{P}\{\mathcal{H}_1(\frac{\lambda}{4}) \cap \mathcal{H}_2(\frac{\lambda}{4}) \cap \mathcal{v}(\frac{\phi_0}{2})\} \geq 1 - 2\mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda^{\frac{1}{2}}}{4}\} - 2\mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\lambda}{16\mathbf{C}_\mathbf{V}}\} - \mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > \frac{\phi_0}{2}\} \geq 1 - 5\mathbb{P}\{\left\| \widehat{\mathbf{V}}_{\mathrm{h}} - \mathbf{V} \right\|_\vdash > 6\mathrm{f}(\lambda, \phi_0)\}$ (D-26), where $\mathrm{f}(\lambda, \phi_0) = \min(\frac{\lambda^{\frac{1}{2}}}{24}, \frac{\lambda}{96\mathbf{C}_\mathbf{V}}, \frac{\phi_0}{12})$. The argument is completed by evaluating the final probability term in (D-26) using the concentration bounds established in **Proposition** 1.

Proof of **Theorem** 1: The argument begins by identifying requirements under which the event supporting the performance bound in **Proposition** 2 occurs with probability converging to one. Under **Condition** 2, together with the remaining conditions imposed in **Proposition** 2, this probability is characterized by $1 - \mathcal{P}_1[\mathrm{f}(\lambda, \phi_0), \mathrm{N}, \mathrm{T}]$. Recalling from **Proposition** 1, the function $\mathcal{P}_1(\cdot)$ admits the representation $\mathcal{P}_1[\mathrm{f}(\lambda, \phi_0), \mathrm{N}, \mathrm{T}] = \{2h[\mathrm{f}(\lambda, \phi_0)] + 1\}\mathrm{N}\{\mathrm{b}_1 \mathrm{T}^{\frac{1-\delta}{3}} + \frac{\{2h[\mathrm{f}(\lambda,\phi_0)]+1\}\mathrm{b}_3}{\mathrm{f}(\lambda,\phi_0)}\} \exp[-\frac{\mathrm{T}^{\frac{1-\delta}{3}}}{2\mathrm{b}_1^2}] + \{2h[\mathrm{f}(\lambda, \phi_0)] + 1\}\mathrm{N}\frac{\mathrm{b}_2\{2h[\mathrm{f}(\lambda,\phi_0)]+1\}^\mathrm{d}}{\mathrm{f}(\lambda,\phi_0)^\mathrm{d}\mathrm{T}^{\frac{\delta(\mathrm{d}-1)}{2}}}$ (D-27) for some $0 < \delta < 1$, where $\mathrm{f}(\lambda, \phi_0) = \min(\frac{\lambda^{\frac{1}{2}}}{24}, \frac{\lambda}{96\mathbf{C}_\mathbf{V}}, \frac{\phi_0}{12})$. Suppose that the regularization parameter satisfies $\lambda = 0(\mathrm{T}^{-\mathrm{q}_\lambda})$ for some $\mathrm{q}_\lambda > 0$. It then follows directly that $\mathrm{f}(\lambda, \phi_0) = \min(\frac{\lambda^{\frac{1}{2}}}{24}, \frac{\lambda}{96\mathbf{C}_\mathbf{V}}, \frac{\phi_0}{12}) = 0(\mathrm{T}^{-\mathrm{q}_\lambda})$ (D-28), and recalling the definition of the bandwidth function $\mathrm{h}[\mathrm{f}(\lambda, \phi_0)]$ in (4), $\mathrm{h}[\mathrm{f}(\lambda, \phi_0)] = 0[\log(\mathrm{T})^2 \mathrm{T}^{\mathrm{q}_\lambda}]$ (D-29). Since both expressions in (D-28) and (D-29) grow at polynomial rates in T, it follows that the first term on the right-hand side of (D-27) converges to zero at an exponential rate as $\mathrm{T} \to \infty$ for any $\delta < 1$. The remaining term in (D-27) likewise vanishes under the stated growth requirements, which establishes that the probability associated with the event in **Proposition** 2 converges to one. This completes the proof of **Theorem** 1. The second term on the right-hand side, by contrast, vanishes at most at a polynomial rate. In particular, it

satisfies $\mathcal{P}_1[f(\lambda,\phi_0),N,T] = O\{\frac{\{2h[f(\lambda,\phi_0)]+1\}N\{h[f(\lambda,\phi_0)]+1\}^d}{\lambda^d T^{\frac{\delta(d-1)}{2}}}\} = O\{\log(T)^{2(d-1)}\, T^{q_N+(d+1)q_K-dq_\lambda=-\frac{\delta(d-1)}{2}}\}$ (D-30). It follows from (D-30) that $\mathcal{P}_1[f(\lambda,\phi_0),N,T] \to 0$ as $T \to \infty$ provided that $q_N + (d+1)q_K - dq_\lambda = -\frac{\delta(d-1)}{2} < 0$, which is equivalent to $q_\lambda < \frac{\delta(d-1)}{2d} - \frac{(d+1)q_K}{d} - \frac{q_N}{d}$. An analogous argument applies to $\mathcal{P}_2[f(\lambda,\phi_0),N,T]$. Specifically, observe that $\mathcal{P}_2[f(\lambda,\phi_0),N,T] = \{2h[f(\lambda,\phi_0)]+1\}N\{\frac{k_1\{2h[f(\lambda,\phi_0)]+1\}}{f(\lambda,\phi_0)}\} + \frac{2}{k_2}\{\frac{Tf(\lambda,\phi_0)^2}{[2h+1]^2}\}^{\frac{1}{7}} \exp\{-\frac{1}{k_3}\{\frac{Tf(\lambda,\phi_0)^2}{[2h[f(\lambda,\phi_0)]+1]^2}\}^{\frac{1}{7}}\}$ (D-31), which converges to zero at an exponential rate in $T$ whenever $\frac{Tf(\lambda,\phi_0)^2}{[2h[f(\lambda,\phi_0)]+1]^2}$ diverges at a polynomial rate. Using relations (D-28) and (D-29), this ratio satisfies $\frac{Tf(\lambda,\phi_0)^2}{[2h[f(\lambda,\phi_0)]+1]^2} = O\{\log(T)^{-4}\, T^{1-2q_\lambda-2q_K}$, which diverges provided that $1 - 2q_\lambda - 2q_K > 0$, or equivalently, $q_\lambda < \frac{1}{2} - q_K$. These restrictions jointly establish requirements $4\lambda P_{a,S}(\hat{\mathbf{c}} - \mathbf{c})$ and $16\left\|\hat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}} - \mathbf{c})\right\|_2 (\frac{\hat{\omega}_\alpha \lambda}{\phi_0})$ stated in **Theorem** 1. The final step characterizes the asymptotic order of the performance bound derived in **Proposition** 2. Observe that $\sum_{g\in\mathcal{G}_S}\sqrt{|g|} \leq |\mathcal{G}_S| \max_{g\in\mathcal{G}_S}\sqrt{|g|} = O(T^{q_g+\frac{q_N}{2}})$. This implies that the effective penalty constant satisfies $\hat{\omega} = O[(1-\alpha)T^{q_g+\frac{q_N}{2}} + \alpha T^{\frac{q_S}{2}}]$. Applying **Proposition** 2 then yields $\left\|\hat{\mathbf{V}}_h^{(d)}(\hat{\mathbf{c}} - \mathbf{c})\right\|_2^2 \leq \frac{4\hat{\omega}_\alpha^2\lambda^2}{\phi_0^2} = O[(1-\alpha)T^{2q_g+2q_N-2q_\lambda} + \alpha T^{q_s-2q_\lambda}]$. These bounds hold on an event whose probability is given by $1 - \mathcal{P}_1[f(\lambda,\phi_0),N,T]$ or $1 - \mathcal{P}_2[f(\lambda,\phi_0),N,T]$, depending on whether **Condition** 2 applies. Since it has already been established that both $\mathcal{P}_1[f(\lambda,\phi_0),N,T] \to 0$ and $\mathcal{P}_2[f(\lambda,\phi_0),N,T] \to 0$ under the requirements stated in **Theorem** 1, the argument is complete.

Proof of **Proposition** 3: Let $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1`,\ldots,\boldsymbol{\beta}_K`)`$, define $\mathbf{U}^{(d)} = [\mathbf{W}_1^{*(d)},\ldots,\mathbf{W}_K^{*(d)}]$, and set $\mathbf{Q}^{(d)} = [\mathbf{V}^{*(d)}\ \mathbf{U}^{(d)}]$, together with their corresponding sample counterparts. The residual term appearing in the $L_2$ component of the SPLASHX objective function can be written as $\hat{\boldsymbol{\sigma}}^* = \mathbf{V}^{*(d)}\mathbf{c} - \sum_{k=1}^K \mathbf{W}_k^{*(d)}\boldsymbol{\beta}_K = \hat{\boldsymbol{\sigma}}^*[\mathbf{V}^{*(d)}\ \mathbf{U}^{(d)}]\begin{bmatrix}\mathbf{c}\\ \boldsymbol{\beta}^*\end{bmatrix} = \boldsymbol{\sigma}^* - [\mathbf{V}^{*(d)}\ \mathbf{U}^{(d)}]\mathbf{q} + [\hat{\boldsymbol{\sigma}}^* - \boldsymbol{\sigma}^*] - [\hat{\mathbf{V}}^{*(d)} - \mathbf{V}^{*(d)}]\mathbf{c} - [\hat{\mathbf{U}}^{(d)} - \mathbf{U}^{(d)}]\boldsymbol{\beta}^* = \hat{\boldsymbol{\Delta}}_1 - \hat{\boldsymbol{\Delta}}_2\mathbf{c} - \hat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*$ (D-32). Invoking the notation introduced above together with identity (D-32), it follows that $\left\|\boldsymbol{\sigma}^* - \mathbf{V}^{*(d)}\hat{\mathbf{c}} - \sum_{k=1}^K \mathbf{W}_k^{*(d)}\hat{\boldsymbol{\beta}}_K\right\|_2^2 = \left\|\boldsymbol{\sigma}^* - \hat{\mathbf{Q}}^{(d)}\hat{\mathbf{q}}\right\|_2^2 = \left\|\{\boldsymbol{\sigma}^* - \hat{\mathbf{Q}}^{(d)}\mathbf{q}\} - \{\hat{\mathbf{Q}}^{(d)}(\hat{\mathbf{q}} - \mathbf{q})\}\right\|_2^2 = \left\|\boldsymbol{\sigma}^* - \hat{\mathbf{Q}}^{(d)}\mathbf{q}\right\|_2^2 + \left\|\hat{\mathbf{Q}}^{(d)}(\hat{\mathbf{q}} - \mathbf{q})\right\|_2^2 - 2(\hat{\mathbf{q}} - \mathbf{q})`\hat{\mathbf{Q}}^{(d)}`(\hat{\boldsymbol{\Delta}}_1 - \hat{\boldsymbol{\Delta}}_2\mathbf{c} - \hat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*)$ (D-33). The penalty structure is next modified to explicitly account for coefficients associated with the exogenous variables. Let the index set $g_k$ be defined such that $\mathbf{q}_{g_k} = \boldsymbol{\beta}_k$ for $k = 1,\ldots,K$, and $\mathcal{G}^* = G \cup \bigcup_{k=1}^K g_k$. Since $|\boldsymbol{\beta}_k| = |g_k| = N$, the combined penalty can be written as $\mathbf{P}_a(\mathbf{c}) = \sum_{k=1}^K (1-\alpha)\sqrt{N}\|\boldsymbol{\beta}_k\|_2 + \alpha\|\boldsymbol{\beta}_k\|_1 = (1-\alpha)\sum_{g\in\mathcal{G}}^*\sqrt{|g|}\left\|\mathbf{q}_g\right\|_2 + \|\mathbf{q}\|_1 =: \hat{\mathbf{P}}_a(\mathbf{q})$. Using this augmented norm $\hat{\mathbf{P}}_a(\mathbf{q})$, the SPLASHX objective function admits the compact representation $\mathcal{L}_\alpha^*(\hat{\mathbf{q}};\lambda) = \left\|\boldsymbol{\sigma}^* - \hat{\mathbf{Q}}^{(d)}\mathbf{q}\right\|_2^2 + \lambda\hat{\mathbf{P}}_a(\mathbf{q})$. Moreover, the dual norm $\hat{\mathbf{P}}_\alpha^*(\mathbf{q})$ satisfies $\hat{\mathbf{P}}_\alpha^*(\mathbf{q}) \leq \|\mathbf{q}\|_\infty$, as implied by the argument in (D-23). Since $\mathcal{L}_\alpha^*(\hat{\mathbf{q}};\lambda) \leq \mathcal{L}_\alpha^*(\mathbf{q};\lambda)$ by construction, it follows that $\left\|\hat{\mathbf{Q}}^{(d)}(\hat{\mathbf{q}} - \mathbf{q})\right\|_2^2 + \lambda\hat{\mathbf{P}}_a(\mathbf{q}) \leq 2(\hat{\mathbf{q}} - \mathbf{q})`\hat{\mathbf{Q}}^{(d)}`(\hat{\boldsymbol{\Delta}}_1 - \hat{\boldsymbol{\Delta}}_2\mathbf{c} - \hat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*) + \hat{\mathbf{P}}_a(\mathbf{q}) \leq \hat{\mathbf{P}}_a(\hat{\mathbf{q}} - \mathbf{q})\{2\hat{\mathbf{P}}_\alpha^*(\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_1) + 2\hat{\mathbf{P}}_\alpha^*(\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_2\mathbf{c}) + 2\hat{\mathbf{P}}_\alpha^*(\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*)\} + \lambda\hat{\mathbf{P}}_a(\mathbf{q}) \leq \hat{\mathbf{P}}_a(\hat{\mathbf{q}} - \mathbf{q})\{2\left\|\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_1\right\|_\infty + 2\left\|\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_2\mathbf{c}\right\|_\infty + 2\left\|\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*\right\|_\infty + \lambda\hat{\mathbf{P}}_a(\mathbf{q})\}$ (D-34). To facilitate probabilistic control of the three stochastic components appearing in the bound, introduce the following events: $\mathcal{H}_1^*(x) = \{2\left\|\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_1\right\|_\infty \leq x\}$, $\mathcal{H}_2^*(x) = \{2\left\|\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_2\mathbf{c}\right\|_\infty \leq x\}$, and $\mathcal{H}_3^*(x) = \{2\left\|\hat{\mathbf{Q}}^{(d)}`\hat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*\right\|_\infty \leq x\}$. Rescaling inequality (D-34) by a factor of two and restricting attention to the event $\mathcal{H}_1^*(\frac{\lambda}{6}) \cap \mathcal{H}_2^*(\frac{\lambda}{6}) \cap \mathcal{H}_3^*(\frac{\lambda}{6})$, one obtains $2\left\|\hat{\mathbf{Q}}^{(d)}(\hat{\mathbf{q}} - \mathbf{q})\right\|_2^2 + 2\lambda\hat{\mathbf{P}}_a(\hat{\mathbf{q}}) \leq \lambda\hat{\mathbf{P}}_a(\hat{\mathbf{q}} - \mathbf{q}) + 2\lambda\hat{\mathbf{P}}_a(\mathbf{q})$ (D-35). The penalty term $\hat{\mathbf{P}}_a(\hat{\mathbf{q}})$ on the left-hand side of (D-35) is handled first. Proceeding as in (D-19) of the main text, it follows that $\hat{\mathbf{P}}_a(\hat{\mathbf{q}}) \geq \hat{\mathbf{P}}_{a,S^*}(\mathbf{q}) + \hat{\mathbf{P}}_{a,S^{*c}}(\hat{\mathbf{q}} - \mathbf{q}) - \hat{\mathbf{P}}_{a,S^*}(\hat{\mathbf{q}} - \mathbf{q})$, where $S^*$ denotes the index set corresponding to all nonzero components of $\mathbf{q}$. For the term $\hat{\mathbf{P}}_a(\hat{\mathbf{q}} - \mathbf{q})$ appearing on the right-hand side of (D-35), re-

lation (D-20) implies that $\widehat{\mathbf{P}}_{a}(\widehat{\mathbf{q}}-\mathbf{q})=\widehat{\mathbf{P}}_{a,S^*}(\widehat{\mathbf{q}}-\mathbf{q})+\widehat{\mathbf{P}}_{a,S^{*c}}(\widehat{\mathbf{q}}-\mathbf{q})$. Combining these expressions yields $2\left\|\widehat{\mathbf{Q}}^{(d)}(\widehat{\mathbf{q}}-\mathbf{q})\right\|_2^2+2\lambda\widehat{\mathbf{P}}_{a,S^{*c}}(\widehat{\mathbf{q}}-\mathbf{q})\leq 3\lambda\widehat{\mathbf{P}}_{a,S^*}(\widehat{\mathbf{q}}-\mathbf{q})$ (D-36). Accordingly, letting $N^*=N_c+NK$, the estimation error satisfies $(\widehat{\mathbf{q}}-\mathbf{q})\in\ell_{N^*}^*(\mathcal{G},S):=\{\boldsymbol{\Delta}\in\mathbb{R}^{N^*}:\widehat{\mathbf{P}}_{a,S^*}(\boldsymbol{\Delta})\leq 3\lambda\widehat{\mathbf{P}}_{a,S^*}(\boldsymbol{\Delta})\}$.

**Auxiliary Result** 1 continues to apply with $\widehat{\omega}_\alpha^*$ defined as in **Proposition** 3. To invoke **Auxiliary Result** 2, introduce the event $\mathscr{v}^*(x)=\{\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_2\leq x\}$, and assume that $\mathscr{v}^*(\frac{\phi_0^*}{2})$ holds. Repeating arguments identical to those leading to inequality (D-22) establishes the core bound stated in **Proposition** 3. To determine the probability with which this inequality is satisfied, it suffices to control the probability of the joint event $\mathcal{H}_1^*(\frac{\lambda}{6})\cap\mathcal{H}_2^*(\frac{\lambda}{6})\cap\mathcal{H}_3^*(\frac{\lambda}{6})\cap\mathscr{v}^*(\frac{\phi_0^*}{2})$. This probability is bounded from below by $1-\mathbb{P}[\mathcal{H}_1^*(\frac{\lambda}{6})^c]-\mathbb{P}[\mathcal{H}_2^*(\frac{\lambda}{6})^c]-\mathbb{P}[\mathcal{H}_3^*(\frac{\lambda}{6})^c]-\mathscr{v}^*(\frac{\phi_0^*}{2})$ (D-37), Each probability appearing in (D-37) can be reduced to tail bounds involving $\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_F$. Proceeding as in (D-24)-(D-25), one obtains $\mathbb{P}[\mathcal{H}_1^*(\frac{\lambda}{6})^c]=\mathbb{P}[\left\|\widehat{\mathbf{Q}}^{(d)\backprime}\widehat{\boldsymbol{\Delta}}_1\right\|_\infty>\frac{\lambda}{12}]\leq\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash^2+C_Q\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda}{12}]\leq\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda^{\frac{1}{2}}}{\sqrt{24}}]+\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda}{24C_Q}]$ (D-38), and similarly, $\mathbb{P}[\mathcal{H}_2^*(\frac{\lambda}{6})^c]=\mathbb{P}[\left\|\widehat{\mathbf{Q}}^{(d)\backprime}\widehat{\boldsymbol{\Delta}}_2\mathbf{c}\right\|_\infty>\frac{\lambda}{12}]\leq\mathbb{P}[\left\|\widehat{\mathbf{Q}}^{(d)\backprime}(\widehat{\mathbf{Q}}^{(d)}-\mathbf{Q}^{(d)})\right\|_\infty>\frac{\lambda}{12}]\leq\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda^{\frac{1}{2}}}{\sqrt{24}}]+\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda}{24C_Q}]$ (D-39), while for the third event, $\mathbb{P}[\mathcal{H}_3^*(\frac{\lambda}{6})^c]=\mathbb{P}[\left\|\widehat{\mathbf{Q}}^{(d)\backprime}\widehat{\boldsymbol{\Delta}}_3\boldsymbol{\beta}^*\right\|_\infty>\frac{\lambda}{12}]\leq\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda^{\frac{1}{2}}}{\sqrt{24}C_\beta}]+\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\lambda}{24C_QC_\beta}$ (D-40). These bounds show that all complement events appearing in (D-37) are controlled by deviations of $\widehat{\mathbf{Q}}$ from its population counterpart. Combining the above inequalities with the concentration results for $\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_F$ yields the desired probability statement, thereby completing the proof of **Proposition** 3. For inequalities (D-39) and (D-40), **Conditions** 1* ensures that $\|\mathbf{c}\|_1\leq 1$ and $\|\boldsymbol{\beta}^*\|_1\leq C_\beta$, respectively. Moreover, $\mathbb{P}[\mathscr{v}^*(\frac{\phi_0^*}{2})^c]=\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_2>\frac{\phi_0^*}{2}]\leq\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>\frac{\phi_0^*}{2}]$. Combining these bounds, inequality (D-37) can be rewritten as $\mathbb{P}[\mathcal{H}_1^*(\frac{\lambda}{6})\cap\mathcal{H}_2^*(\frac{\lambda}{6})\cap\mathcal{H}_3^*(\frac{\lambda}{6})\cap\mathscr{v}^*(\frac{\phi_0^*}{2})]\geq 1-7\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash>6f^*(\lambda,\phi_0^*]$. There into $f^*(\lambda,\phi_0^*]=\min(\frac{\lambda^{\frac{1}{2}}}{12\sqrt{6}},\frac{\lambda}{144C_Q},\frac{\lambda^{\frac{1}{2}}}{144C_QC_\beta},\frac{\phi_0^*}{12})$. The remaining task is to derive a lower bound for $\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_F>x]$. Rather, an upper bound for $\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_F\leq x]$ is obtained as shown: $\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\vdash\leq x]=\mathbb{P}[\max(\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_1,\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_\infty)\leq x]\geq\mathbb{P}[(K+2)N\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_{max}\leq x]=1-\mathbb{P}[\left\|\widehat{\mathbf{Q}}-\mathbf{Q}\right\|_{max}>\frac{x}{(K+2)N}]\geq 1-\mathbb{P}[\cup_{1\leq i\leq(K+2)N,1\leq j\leq(K+2)N}\left|[\widehat{\mathbf{Q}}-\mathbf{Q}]_{ij}\right|>\frac{x}{(K+2)N}]\geq 1-\sum_{i=1}^{(K+1)N}\cdot\sum_{j=1}^{(K+2)N}\mathbb{P}\left[\left|\sum_{t=2}^{T}\xi_{it}\xi_{jt}\right|>\frac{x}{(K+2)N}\right]\geq 1-(K+1)(K+2)N\{[b_1T^{\frac{1-\delta}{3}}+\frac{(k+2)Nb_3}{x}]\exp[-\frac{T^{\frac{1-\delta}{3}}}{2b_1^2}+\frac{b_2(K+2)^dN^d}{x^dT^{\frac{\delta(d-1)}{2}}}]$ for polynomial tails or $1-(K+1)(K+2)N^2\{[\frac{k_1(k+2)Nb_3}{x}+\frac{2}{k_2}[\frac{Tx^2}{(K+2)^2N^2}]^{\frac{1}{7}}]\exp[-\frac{1}{k_3}\{\frac{Tx^2}{(K+2)^2N^2}\}^{\frac{1}{7}}]$ for exponential tails. Here, $\xi_{it}$ represents an individual entry drawn from the autocovariance structure introduced in **Auxiliary Result** 3. Consistent with the characterization provided in **Proposition** 3, the probabilities appearing on the right-hand side reduce to $1-\mathcal{P}_1^*(x,N,T)$ when the underlying distribution exhibits polynomially decaying tails, and to $1-\mathcal{P}_2^*(x,N,T)$ when exponential tail decay prevails.