Measuring Economic Growth with Many Signals
Supplemental Appendix

1 Proofs

1.1 Proof of Proposition 1

Proof. Given the model, Assumption 1 implies that we then are able to estimate the γ_k as follows:

$$\begin{aligned}
\gamma_2 &= \frac{cov(x_2, x_3)}{cov(x_1, x_3)} = \frac{\gamma_2 \gamma_3 var(x^*)}{\gamma_3 var(x^*)} \\
\sigma^2_{x^*} &= cov(x_1, x_2)/\gamma_2 = \frac{cov(x_1, x_2) cov(x_1, x_3)}{cov(x_2, x_3)} \\
\gamma_k &= \frac{cov(x_k, x_1)}{\sigma^2_{x^*}} = \frac{\gamma_k var(x^*)}{\sigma^2_{x^*}}, \quad k = 3, ..., K.
\end{aligned}$$

The normalization condition in Assumption 2 implies that the means of the errors can be estimated as follows:

$$\begin{aligned}
E[x^*] &= E[x_1] \\
E[\varepsilon] &= E[x] - \gamma E[x^*] \\
&= E[x] - \gamma E[x_1]
\end{aligned}$$

The linear combination of the measurements to approximate the latent x^* is defined as

$$\begin{aligned}
\hat{x}^*(\lambda) &\equiv \lambda_0 + (\lambda_1, \lambda_2, ..., \lambda_K)(x_1, x_2, ..., x_K)^T \\
&\equiv \lambda_0 + \lambda_1^T x.
\end{aligned}$$

We pick λ to minimize the mean squared error as follows:

$$\lambda^{\text{BLUE}} = \arg\min_{\lambda} E[\lambda_0 + \lambda_1^T x - x^*]^2. \tag{13}$$

The first order condition with respect to λ_0 follows:

$$E[\lambda_0 + \lambda_1^T x - x^*] = 0,$$

which implies

$$\lambda_0 \;=\; -\lambda_1^T E[x] + E[x^*], \tag{14}$$

and

$$\begin{aligned}
\hat{x}^*(\lambda) &= \lambda_0 + \lambda_1^T x \\
&= -\lambda_1^T E[x] + E[x^*] + \lambda_1^T(\gamma x^* + \varepsilon) \\
&= -\lambda_1^T E[\gamma x^* + \varepsilon] + E[x^*] + \lambda_1^T(\gamma x^* + \varepsilon) \\
&= E[x^*] + \lambda_1^T(\gamma x^* + \varepsilon - E[\gamma x^* + \varepsilon]) \\
&= E[x^*] + \lambda_1^T\gamma(x^* - E[x^*]) + \lambda_1^T(\varepsilon - E[\varepsilon]) \\
&= x^* + (\lambda_1^T\gamma - 1)(x^* - E[x^*]) + \lambda_1^T(\varepsilon - E[\varepsilon]).
\end{aligned}$$

The mean squared error can be simplified as follows:

$$\begin{aligned}
E[\hat{x}^*(\lambda) - x^*]^2 &= (\lambda_1^T\gamma - 1)^2\sigma_{x^*}^2 + \lambda_1^T V(\varepsilon)\lambda_1 \\
&= (\bar{\lambda}_1^T\mathbf{1} - 1)^2\sigma_{x^*}^2 + \bar{\lambda}_1^T\Gamma^{-1}V(\varepsilon)\Gamma^{-1}\bar{\lambda}_1 \\
&\equiv (\bar{\lambda}_1^T\mathbf{1} - 1)^2\sigma_{x^*}^2 + \bar{\lambda}_1^T\bar{V}\bar{\lambda}_1
\end{aligned}$$

where $\Gamma = diag\{\gamma_1, \gamma_2, ..., \gamma_K\}$ and $\bar{\lambda}_1 = \Gamma\lambda_1$.

Part (i): MSE-optimal (bias-allowing) case. We minimize

$$E[\hat{x}^*(\lambda) - x^*]^2 = (\lambda_1^T\gamma - 1)^2\sigma_{x^*}^2 + \lambda_1^T V\lambda_1$$

over λ_1 without any constraint. The first-order condition is

$$\frac{\partial}{\partial\lambda_1}\left[(\lambda_1^T\gamma - 1)^2\sigma_{x^*}^2 + \lambda_1^T V\lambda_1\right] \;=\; 2(\lambda_1^T\gamma - 1)\sigma_{x^*}^2\gamma + 2V\lambda_1 \;=\; 0.$$

Letting $c := 1 - \lambda_1^T\gamma$, this rearranges to $V\lambda_1 = c\sigma_{x^*}^2\gamma$, so

$$\lambda_1^{\text{MSE}} \;=\; c\,\sigma_{x^*}^2\,V^{-1}\gamma.$$

Premultiplying by γ^T gives $\gamma^T\lambda_1^{\text{MSE}} = c\,\sigma_{x^*}^2\gamma^T V^{-1}\gamma = 1 - c$, which yields

$$c \;=\; \frac{1}{1 + \sigma_{x^*}^2\gamma^T V^{-1}\gamma}, \qquad \lambda_1^{\text{MSE}} \;=\; \frac{\sigma_{x^*}^2\,V^{-1}\gamma}{1 + \sigma_{x^*}^2\gamma^T V^{-1}\gamma}.$$

Substituting into $\lambda_0 = E[x_1] - \lambda_1^T E[x]$ gives the predictor

$$\hat{x}^*(\lambda^{\text{MSE}}) \;=\; \frac{\sigma_{x^*}^2}{1 + \sigma_{x^*}^2\gamma^T V^{-1}\gamma}\,\gamma^T V^{-1}(x - E[x]) + E[x_1].$$

The minimum mean squared error follows by substituting λ_1^{MSE} back into the decomposition. With $\lambda_1^T\gamma - 1 = -c$ and $\lambda_1^T V \lambda_1 = c^2\sigma_{x^*}^4\gamma^T V^{-1}\gamma$,

$$\begin{aligned} E[\hat{x}^*(\lambda^{\text{MSE}}) - x^*]^2 &= c^2\sigma_{x^*}^2 + c^2\sigma_{x^*}^4\gamma^T V^{-1}\gamma \;=\; c^2\sigma_{x^*}^2(1 + \sigma_{x^*}^2\gamma^T V^{-1}\gamma) \\ &= \frac{\sigma_{x^*}^2}{1 + \sigma_{x^*}^2\gamma^T V^{-1}\gamma}. \end{aligned}$$

Part (ii): BLUE (unbiased) case. If we additionally require the linear approximation to have conditional mean equal to the latent variable, i.e., $E[\hat{x}^*(\lambda)|x^*] = x^*$ whenever the errors are conditional mean independent of x^*, then we impose $\bar{\lambda}_1^T\mathbf{1} = 1$. Define

$$\Lambda := \{\lambda : \lambda_1^T\gamma = 1; \lambda_0 = -\lambda_1^T E[x] + E[x_1]\}.$$

We pick λ or $\bar{\lambda}$ to minimize the mean squared errors as follows:

$$\lambda^{\text{BLUE}} = \arg\min_{\lambda\in\Lambda} E[\hat{x}^*(\lambda) - x^*]^2 \tag{15}$$

or

$$\bar{\lambda}_1^{\text{BLUE}} = \arg\min_{\bar{\lambda}_1^T\mathbf{1}=1} \bar{\lambda}_1^T\bar{V}\bar{\lambda}_1 \tag{16}$$

The first-order condition of this minimization problem with the Lagrange multiplier δ

$$\begin{aligned} \frac{\partial}{\partial\lambda_1}[\bar{\lambda}_1^T\bar{V}\bar{\lambda}_1 + \delta(1 - \mathbf{1}^T\bar{\lambda}_1)] &= 2\bar{V}\bar{\lambda}_1 - \delta\mathbf{1} = 0 \\ \bar{\lambda}_1 &= \bar{V}^{-1}\times\mathbf{1}\times\delta/2 \end{aligned}$$

With $\bar{\lambda}_1^T\mathbf{1} = 1$, we have

$$\begin{aligned} \delta &= \frac{2}{\mathbf{1}^T\times\bar{V}^{-1}\times\mathbf{1}} \\ \bar{\lambda}_1 &= \frac{\bar{V}^{-1}\times\mathbf{1}}{\mathbf{1}^T\times\bar{V}^{-1}\times\mathbf{1}} \\ \lambda_1 = \Gamma^{-1}\bar{\lambda}_1 &= \frac{1}{\gamma^T V^{-1}\gamma}V^{-1}\gamma \end{aligned}$$

We may then solve for λ_0 as follows

$$\begin{aligned} \lambda_0 &= -\lambda_1^T E[x] + E[x^*] \\ &= -\lambda_1^T E[x] + E[x_1] \end{aligned}$$

That means

$$\begin{aligned}\hat{x}^*(\lambda^{\text{BLUE}}) &= \lambda_0 + \lambda_1^T x \\ &= -\lambda_1^T E[x] + E[x_1] + \lambda_1^T x \\ &= \frac{1}{\gamma^T V^{-1}\gamma}\gamma^T V^{-1}(x - E[x]) + E[x_1]\end{aligned}$$

Notice that the variance-covariance matrix V

$$V(\varepsilon) = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 & ... & 0 \\ 0 & \sigma_2^2 & 0 & \sigma_{24} & ... & \sigma_{2K} \\ 0 & 0 & \sigma_3^2 & \sigma_{34} & ... & \sigma_{3K} \\ 0 & \sigma_{42} & \sigma_{43} & \sigma_4^2 & ... & \sigma_{4K} \\ \vdots & \vdots & \vdots & \vdots & ... & \vdots \\ 0 & \sigma_{K2} & \sigma_{K3} & \sigma_{K4} & ... & \sigma_{K=K}^2 \end{pmatrix}$$

can be directly estimated as follows:

$$\begin{aligned}\sigma_{x^*}^2 &= \frac{cov(x_1, x_2)cov(x_1, x_3)}{cov(x_2, x_3)} \\ \sigma_{jk} &= cov(x_j, x_k) - \gamma_j\gamma_k\sigma_{x^*}^2 \\ \sigma_k^2 &= \sigma_{kk}\end{aligned}$$

■

1.2 Proof of Proposition 2

Proof. We have $V = U^T DU$ where $U^T U = I$ and $D := diag\{\varsigma_1^2, \varsigma_2^2, ..., \varsigma_K^2\}$ is a diagonal matrix containing all the eigenvalues of V with $\varsigma_1^2 \geq \varsigma_2^2 \geq ... \geq \varsigma_K^2 > 0$. Based on the results in Proposition 1, the minimum mean squared error can written as

$$\begin{aligned}E[\hat{x}^*(\lambda^{\text{BLUE}}) - x^*]^2 &= E[\lambda_1^T(\varepsilon - E[\varepsilon])]^2 \\ &= \lambda_1^T V \lambda_1 \\ &= \left(\frac{1}{\gamma^T V^{-1}\gamma}\right)^2 (V^{-1}\gamma)^T V V^{-1}\gamma \\ &= \left(\gamma^T V^{-1}\gamma\right)^{-1}\end{aligned}$$

With a diagonal $V = U^T DU$ and $K \to \infty$, we have

$$\begin{aligned} E[\hat{x}^*(\lambda^{\text{BLUE}}) - x^*]^2 &= \left(\gamma^T V^{-1}\gamma\right)^{-1} \\ &= \left(\gamma^T U^T D^{-1} U\gamma\right)^{-1} \\ &\leq \left(\frac{1}{(\varsigma_K^2)^{-1}\gamma^T U^T U\gamma}\right) \\ &= \left(\frac{1}{(\varsigma_K^2)^{-1}\gamma^T\gamma}\right) \\ &= \left(\frac{\varsigma_K^2}{\sum_{k=1}^K \gamma_K^2}\right) \\ &= O\left(\frac{K^{-\beta}}{\sum_{k=1}^K k^{-2\alpha}}\right) \end{aligned}$$

The results then follow. ■

1.3 Proof of Proposition 3

Proof. The BLUE linear approximation is

$$\begin{aligned} \hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) &= \frac{1}{\hat{\gamma}^T\hat{V}^{-1}\hat{\gamma}}\hat{\gamma}^T\hat{V}^{-1}(x - \bar{x}) + \bar{x}_1 \\ &\equiv \sum_{k=0}^K \hat{\lambda}_{k,N} \times x_k \end{aligned}$$

where $x_0 \equiv 1$. Our goal is to show the mean square convergence of $\hat{x}_n^*(\lambda)$, i.e.,

$$\lim_{N\to\infty} E[\hat{x}^*(\hat{\lambda}) - \hat{x}^*(\lambda)]^2 = 0.$$

We consider a plug-in estimator as follows:

$$\begin{aligned} \hat{\lambda}_{k,N} &\equiv G_k(\bar{x}_1, ..., \bar{x}_K, \frac{\sum_{i=1}^N x_{1,i}x_{1,i}}{N}, ..., \frac{\sum_{i=1}^N x_{1,i}x_{K,i}}{N}, ..., \frac{\sum_{i=1}^N x_{K,i}x_{K,i}}{N}) \\ &\equiv G_k(\hat{\mu}) \end{aligned}$$

where $\hat{\mu}$ is a $(\frac{K\times(K+3)}{2})$ by 1 vector of all the sample averages $\frac{1}{N}\sum_{i=1}^N x_{j,i}x_{k,i}$ for $j \geq k$ and $j, k = 0, 1, ..., K$. Let $\mu = E\hat{\mu}$. Notice that $\lambda_k = G_k(\mu)$. Similarly, we may define $\lambda_0 = G_0(\mu)$. Although G_k are complicated, it is a known analytic function over the real line.

We define the remaining term in a Taylor expansion as follows:

$$e_{k,N} \equiv \hat{\lambda}_{k,N}(\hat{\mu}) - \lambda_k - \frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu} - \mu).$$

We have

$$
\begin{aligned}
E|\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) - \hat{x}^*(\lambda^{\text{BLUE}})|^2 &= E\left|\sum_{k=0}^{K}[(\hat{\lambda}_{k,N} - \lambda_k)x_k\right|^2 \\
&\leq E\left[\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu) + e_{k,N}\right|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\
&\leq 2E\left[\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu)\right|^2 + \sum_{k=0}^{K}|e_{k,N}|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\
&\leq 2E\left[\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu)\right|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\
&\quad +2E\left[\left(\sum_{k=0}^{K}|e_{k,N}|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \qquad (17)
\end{aligned}
$$

The first term in Equation (17) is

$$
\begin{aligned}
& E\left[\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu)\right|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\
\leq & \left[E\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu)\right|^2\right)^2 E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2} \\
= & \left[E\left(\sum_{k=0}^{K}\left|\sum_{j=1}^{\frac{K(K+3)}{2}}\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu}_j-\mu_j)\right|^2\right)^2 E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2} \\
\leq & \left[E\left(\sum_{k=0}^{K}\frac{K(K+3)}{2}\sum_{j=1}^{\frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu}_j-\mu_j)\right|^2\right)^2 E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2} \\
\leq & \frac{K(K+3)}{2}\left[\frac{K(K+1)(K+3)}{2}E\left(\sum_{k=0}^{K}\sum_{j=1}^{\frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu}_j-\mu_j)\right|^4\right)E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2} \\
\leq & \frac{K(K+3)}{2}\left[\left(\frac{K(K+1)(K+3)}{2}\right)^2 E\left(\max_{0\leq k\leq K;1\leq j\leq\frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2}
\end{aligned}
$$

and

$$
\begin{aligned}
&\leq \frac{K(K+3)}{2}\frac{K(K+1)(K+3)}{2}\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2}\\
&\leq O(K^5)\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)(K+1)E\left(\sum_{k=0}^{K}|x_k|^4\right)\right]^{1/2}\\
&\leq O(K^5)\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)(K+1)^2E\left(\max_{0\leq k\leq K}|x_k|^4\right)\right]^{1/2}\\
&\leq O(K^5)(K+1)\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)E\left(\max_{0\leq k\leq K}|x_k|^4\right)\right]^{1/2}\\
&\leq O(K^6)\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)E\left(\max_{0\leq k\leq K}|x_k|^4\right)\right]^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq O(K^6)\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)(K+1)\left(\max_{0\leq k\leq K}E|x_k|^4\right)\right]^{1/2}\\
&\leq O(K^{6.5})\left[E\left(\max_{0\leq k\leq K;1\leq j\leq \frac{K(K+3)}{2}}\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right)\left(\max_{0\leq k\leq K}E|x_k|^4\right)\right]^{1/2}
\end{aligned}
$$

The second last step uses the fact that $\max_i |Z_i| \leq \sum_i |Z_i|$.Assumption 5 guarantees $\max_{0\leq k\leq K} E|x_k|^4 < \infty$. We consider the first-order term

$$\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j) \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{\partial G_k(\mu)}{\partial \mu_j}(x_{k,i}x_{l,i}-E(x_{k,i}x_{l,i}))$$

where μ_j corresponds to $E(x_{k,i}x_{l,i})$ in vector μ. Notice that the expression above has a zero mean and G_k is a real analytic function. Assumption 5 implies for each k

$$E\left|\frac{\partial G_k(\mu)}{\partial \mu_j}(x_{k,i}x_{l,i}-E(x_{k,i}x_{l,i}))\right|^4 \leq \frac{4!}{2}M^2.$$

holds for a finite M. Bernstein's inequality (Corollary 14.1 in Bühlmann and Van De Geer (2011)

) then implies that

$$\begin{aligned} & E\left(\max_{0\le k\le K; 1\le j\le \frac{K(K+3)}{2}} \left|\frac{\partial G_k(\mu)}{\partial \mu_j}(\hat{\mu_j}-\mu_j)\right|^4\right) \\ = & \left(\sqrt{\frac{2\ln(K(K+1)(K+3))}{N}}+\frac{M\ln(K(K+1)(K+3))}{N}\right)^4 \\ = & O\left(\frac{(\ln K)^2}{N^2}\right) \end{aligned}$$

Thus the first term in Equation (17) is bounded as follows:

$$\begin{aligned} & E\left[\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu)\right|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\ = & O(K^{6.5})O\left(\frac{\ln K}{N}\right) \\ = & O\left(\frac{K^{6.5}\ln K}{N}\right) \end{aligned}$$

We then consider the second term in Equation (17):

$$\begin{aligned} & E|\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}})-\hat{x}^*(\lambda^{\text{BLUE}})|^2 \\ \le & 2E\left[\left(\sum_{k=0}^{K}\left|\frac{\partial G_k(\mu)}{\partial \mu^T}(\hat{\mu}-\mu)\right|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\ & +2E\left[\left(\sum_{k=0}^{K}|e_{k,N}|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\ = & O\left(\frac{K^{6.5}\ln K}{N}\right)+2E\left[\left(\sum_{k=0}^{K}|e_{k,N}|^2\right)\left(\sum_{k=0}^{K}|x_k|^2\right)\right] \\ \le & O\left(\frac{K^{6.5}\ln K}{N}\right)+2\left[E\left(\sum_{k=0}^{K}|e_{k,N}|^2\right)^2 E\left(\sum_{k=0}^{K}|x_k|^2\right)^2\right]^{1/2} \\ \le & O\left(\frac{K^{6.5}\ln K}{N}\right)+2\left[(K+1)E\left(\sum_{k=0}^{K}|e_{k,N}|^4\right)(K+1)E\left(\sum_{k=0}^{K}|x_k|^4\right)\right]^{1/2} \\ \le & O\left(\frac{K^{6.5}\ln K}{N}\right)+2\left[(K+1)^2E\left(\max_{0\le k\le K}|e_{k,N}|^4\right)(K+1)^2E\left(\max_{0\le k\le K}|x_k|^4\right)\right]^{1/2} \\ = & O\left(\frac{K^{6.5}\ln K}{N}\right)+O(K^2)\left[E\left(\max_{0\le k\le K}|e_{k,N}|^4\right)E\left(\max_{0\le k\le K}|x_k|^4\right)\right]^{1/2} \end{aligned}$$

Again, we use

$$E\left(\max_{0\le k\le K}|e_{k,N}|^4\right) \le (K+1)\left(\max_{0\le k\le K}E|e_{k,N}|^4\right)$$

$$E\left(\max_{0\le k\le K}|x_k|^4\right) \le (K+1)\left(\max_{0\le k\le K}E|x_k|^4\right)$$

Therefore,

$$\begin{aligned}
& E|\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) - \hat{x}^*(\lambda^{\text{BLUE}})|^2. \\
\le\ & O\left(\frac{K^{6.5}\ln K}{N}\right) + O(K^2)\left[(K+1)\left(\max_{0\le k\le K}(E|e_{k,N}|^4)^{1/4}\right)^4 (K+1)\left(\max_{0\le k\le K}E|x_k|^4\right)\right]^{1/2} \\
\le\ & O\left(\frac{K^{6.5}\ln K}{N}\right) + O(K^3)\left(\max_{0\le k\le K}(E|e_{k,N}|^4)^{1/4}\right)^2\left(\max_{0\le k\le K}E|x_k|^4\right)^{1/2}
\end{aligned}$$

We have shown $\max_{0\le k\le K}E|x_k|^4 < \infty$. The second term is negligible when it converges faster than the leading term so that

$$\max_{0\le k\le K}(E|e_{k,N}|^4)^{1/4} = o\left(\sqrt{\frac{K^{6.5-3}\ln K}{N}}\right)$$

which is implied by Assumption 5.

Finally, we have

$$E|\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) - \hat{x}^*(\lambda^{\text{BLUE}})|^2 = O\left(\frac{K^{6.5}\ln K}{N}\right).$$

That leads to

$$\begin{aligned}
E[\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) - x^*]^2 &= E[\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) - \hat{x}^*(\lambda^{\text{BLUE}}) + \hat{x}^*(\lambda^{\text{BLUE}}) - x^*]^2 \\
&\le 2E[\hat{x}^*(\hat{\lambda}_N^{\text{BLUE}}) - \hat{x}^*(\lambda^{\text{BLUE}})]^2 + 2E[\hat{x}^*(\lambda^{\text{BLUE}}) - x^*]^2 \\
&= O\left(\frac{K^{6.5}\ln K}{N}\right) + O\left(\frac{K^{-\beta}}{\sum_{k=1}^K k^{-2\alpha}}\right)
\end{aligned}$$

■

1.4 Proof of Proposition 4

Proof. Let us denote $\sigma^2_{*|23} := \text{Var}(x^* \mid x_2, x_3) = E\big[(x^* - E[x^* \mid x_2, x_3])^2\big]$ for the smallest mean squared error in predicting x^* from (x_2, x_3) alone, and $\sigma_1^2 := \text{Var}(\varepsilon_1)$. Recall $x_1 = x^* + \varepsilon_1$, and that by Corollary 1 each generated measurement $x_a = h(x_2, x_3)$ satisfies $\text{cov}(\varepsilon_1, x_a) = 0$.

First, we show the MSE-optimal limit equals the oracle bound. By Corollary 1, under Assumptions 1, 2 and equation (5) the augmented system $\{x_1, x_2, x_3, h_1, \ldots, h_{K_d}\}$ satisfies Proposition 1 for every d. The MSE-optimal combination is the L^2 projection of x^* onto the closed linear span $\mathcal{H}_d := \overline{\text{span}}\{1, x_1, x_2, x_3, h_1, \ldots, h_{K_d}\}$, so

$$E\big[\hat{x}^*(\lambda_{K_d}^{\text{MSE}}) - x^*\big]^2 = \big\|x^* - P_{\mathcal{H}_d} x^*\big\|^2,$$

where $P_{\mathcal{H}_d}$ is orthogonal projection and $\|\cdot\|^2 = E[(\cdot)^2]$. The subspaces are nested and increasing; since polynomials are dense in $L^2(x_2, x_3)$ and the sieve contains no powers of x_1, their closure is

$$\mathcal{H}_\infty = \big\{b\, x_1 + g(x_2, x_3) : b \in \mathbb{R},\ g \in L^2(x_2, x_3)\big\}.$$

Orthogonal projections onto an increasing sequence of closed subspaces converge strongly to the projection onto the closure of their union, so $\|x^* - P_{\mathcal{H}_d} x^*\|^2 \downarrow \|x^* - P_{\mathcal{H}_\infty} x^*\|^2$. To evaluate the limit, minimize $E\big[(x^* - b\, x_1 - g(x_2, x_3))^2\big]$ over $b \in \mathbb{R}$ and $g \in L^2(x_2, x_3)$. Substituting $x_1 = x^* + \varepsilon_1$ and using equation (5), which makes ε_1 mean-independent of (x^*, x_2, x_3) and hence orthogonal to $(1-b)x^* - g(x_2, x_3)$, the objective separates:

$$E\big[(x^* - b\, x_1 - g)^2\big] = E\big[\big((1-b)x^* - g(x_2, x_3)\big)^2\big] + b^2\sigma_1^2.$$

For fixed b the inner minimizer is $g = (1-b)\, E[x^* \mid x_2, x_3]$, leaving $(1-b)^2\sigma^2_{*|23} + b^2\sigma_1^2$; minimizing over b gives

$$\big\|x^* - P_{\mathcal{H}_\infty} x^*\big\|^2 = \Big(\tfrac{1}{\sigma_1^2} + \tfrac{1}{\sigma^2_{*|23}}\Big)^{-1},$$

attained by $\hat{x}^*_\infty = b^\star x_1 + (1 - b^\star)E[x^* \mid x_2, x_3]$ with $b^\star = \sigma^2_{*|23}/(\sigma^2_{*|23} + \sigma_1^2)$. Under the affine restriction (7), $E[x^* \mid x_1, x_2, x_3]$ has the form $b\, x_1 + g(x_2, x_3)$ and therefore coincides with $\hat{x}^*_\infty$; hence $\|x^* - P_{\mathcal{H}_\infty} x^*\|^2 = \text{Var}(x^* \mid x_1, x_2, x_3) = O$, which establishes (8).

Second, we show that the sufficient condition implies (7). Under the stated Gaussianity, conditional on (x_2, x_3) we have a Gaussian location model with prior $x^* \sim \mathcal{N}\big(E[x^* \mid x_2, x_3], \sigma^2_{*|23}\big)$ and one observation $x_1 = x^* + \varepsilon_1$, $\varepsilon_1 \sim \mathcal{N}(0, \sigma_1^2)$ independent. The posterior is Gaussian with mean $b^\star x_1 + (1 - b^\star)E[x^* \mid x_2, x_3]$—affine in x_1—and constant variance $\big(1/\sigma_1^2 + 1/\sigma^2_{*|23}\big)^{-1}$. Thus (7) holds with $\beta = b^\star$, and $\text{Var}(x^* \mid x_1, x_2, x_3) = \big(1/\sigma_1^2 + 1/\sigma^2_{*|23}\big)^{-1} = O$.

Lastly, the BLUE limit derives from the relationship in equation (4). Applying it to the degree-d system and writing $P_d := \gamma_{K_d}^T V_{K_d}^{-1} \gamma_{K_d}$ for its precision, (4) reads

$$\frac{1}{E[\hat{x}^*(\lambda_{K_d}^{\text{MSE}}) - x^*]^2} = \frac{1}{\sigma^2_{x^*}} + P_d = \frac{1}{\sigma^2_{x^*}} + \frac{1}{E[\hat{x}^*(\lambda_{K_d}^{\text{BLUE}}) - x^*]^2},$$

so that

$$\frac{1}{E[\hat{x}^*(\lambda_{K_d}^{\text{BLUE}}) - x^*]^2} = \frac{1}{E[\hat{x}^*(\lambda_{K_d}^{\text{MSE}}) - x^*]^2} - \frac{1}{\sigma^2_{x^*}}.$$

Letting $d \to \infty$ and substituting the limit (8),

$$\lim_{d\to\infty} \frac{1}{E[\hat{x}^*(\lambda_{K_d}^{\mathrm{BLUE}}) - x^*]^2} = \frac{1}{O} - \frac{1}{\sigma_{x^*}^2} = \frac{\sigma_{x^*}^2 - O}{\sigma_{x^*}^2\, O},$$

which is positive because $O < \sigma_{x^*}^2$. Inverting gives $\lim_{d\to\infty} E[\hat{x}^*(\lambda_{K_d}^{\mathrm{BLUE}}) - x^*]^2 = \sigma_{x^*}^2 O/(\sigma_{x^*}^2 - O)$, and this exceeds O since $\sigma_{x^*}^2/(\sigma_{x^*}^2 - O) > 1$. This proves (9). ■

References

Bühlmann, P., and S. Van De Geer (2011): *Statistics for High-dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media.