Daniel Pellatt, 2026

# Supplementary Model Note: Kim et al. Arbitrage Portfolios and Nonlinear Alpha-Step Extensions

## 1. Kim et al. model and residualized-alpha interpretation

Within a rolling estimation window, let $N$ be the number of stocks, $T$ the number of months, $L$ the number of characteristics, and $K$ the number of latent factors. Let

$$R = [R_1, \dots, R_T] \in \mathbb{R}^{N \times T}$$

be the matrix of excess returns, where $R_t \in \mathbb{R}^N$ is the cross-section of returns in month $t$. Let $X \in \mathbb{R}^{N \times L}$ be the characteristic matrix used in the Kim-style rolling window and $F = [f_1, \dots, f_T]' \in \mathbb{R}^{T \times K}$ the latent factor realizations. Kim, Korajczyk, and Neuhierl (2021) write the local return model as

$$R = (G_\alpha(X) + \Gamma_\alpha)\mathbf{1}_T' + \big(G_\beta(X) + \Gamma_\beta\big)F' + E.$$

Here $G_\alpha(X) \in \mathbb{R}^N$ is the characteristic-driven mispricing function, $G_\beta(X) \in \mathbb{R}^{N \times K}$ is the characteristic-driven factor-loading function, and $\Gamma_\alpha$ and $\Gamma_\beta$ are components not explained by the chosen characteristics. $\mathbf{1}_T$ denotes the $(T \times 1)$ vector of ones.

Define the rolling-window average return vector

$$\bar{R} = \frac{1}{T} R \mathbf{1}_T \in \mathbb{R}^N.$$

After estimating $G_\beta(X)$, define the estimated factor residual-maker

$$\widehat{M}_\beta = I_N - \hat{G}_\beta(X)(\hat{G}_\beta(X)'\hat{G}_\beta(X))^{-1}\hat{G}_\beta(X)'.$$

Kim et al.'s constrained alpha step can be written as

$$\hat{\theta} = (X'X)^{-1}X'\widehat{M}_\beta \bar{R},$$

so the fitted linear mispricing function proposed in the paper is

$$\hat{G}_\alpha^{\text{lin}}(X) = X\hat{\theta}.$$

Thus, the baseline alpha step, that proposed by Kim et al., is equivalent to a linear regression of factor-residualized average returns on characteristics (equivalent to regressing $\widehat{M}_\beta \bar{R}$ on $X$). The residualization removes the component of average returns explained by the estimated characteristic-driven factor-loading space before attributing the remaining cross-sectional pattern to the mispricing function.

## 2. Motivation for nonlinear extensions

This interpretation motivates a direct nonlinear extension. Rather than changing the Kim et al. factor-loading step, I keep the estimated factor residual-maker $\widehat{M}_\beta$ and replace only the linear alpha map with XGBoost. This is an empirical extension, not a new proof of Kim et al.'s asymptotic arbitrage result.

**ML extension 1: mean target.** The mean-target model uses the same averaged target structure as the Kim alpha step:

$$\hat{y}^{\text{mean}} = \widehat{M}_\beta \bar{R}.$$

It fits a nonlinear map $m_{\text{mean}}$ by solving the supervised-learning problem

$$\hat{m}_{\text{mean}} \approx \arg\min_{m \in \mathcal{M}_{\text{XGB}}} \sum_{i=1}^{N} \big(\hat{y}_i^{\text{mean}} - m(x_i)\big)^2.$$

At the portfolio-formation date, with current characteristics $X_f$, the model produces

$$\hat{g}_f^{\text{ML}} = \hat{m}_{\text{mean}}(X_f).$$

I then apply a final beta-space projection before forming weights:

$$\hat{g}_f^{\mathrm{ML,resid}} = \widehat{M}_{\beta,f}^{(q)} \hat{g}_f^{\mathrm{ML}}.$$

Here $q$ denotes the selected projection mode. In the window mode following what is visible in Kim et al's repo, $\widehat{M}_{\beta,f}^{(q)}$ uses the same estimated beta-loading space from the Kim-style rolling-window estimator (i.e. $\widehat{M}_{\beta}$). In the updated mode, the estimated beta-loading relation is applied to the current formation characteristics $X_f$ (known at rebalance time), analogous to Kim et al.'s use of updated characteristics in portfolio formation. Validation determines which projection mode is used.

**ML extension 2: panel target.** The panel-target model uses the month-level analogue of the same residualized target. For each month $t$ in the estimation window,

$$\hat{y}_t^{\mathrm{panel}} = \widehat{M}_{\beta} R_t.$$

The model is trained on stacked observations,

$$\{\left(x_{i,t-1}, \hat{y}_{i,t}^{\mathrm{panel}}\right) : i = 1, \dots, N,\ t = 1, \dots, T\},$$

and then predicts the formation-date signal using $X_f$. The resulting signal is also projected against the selected estimated beta-loading space before weights are formed. The panel-target version uses more within-window information, but it is also more experimental than the mean-target version because it trains on noisier month-level residualized returns rather than the averaged target closest to Kim et al.'s Lemma 6 expression.

In the implementation, XGBoost labels may be standardized for numerical stability. Predictions are mapped back into return units before projection and portfolio formation, so this standardization is a training-scale normalization rather than a change in the economic target.

## 3. Interpretation

The ML estimators are motivated by two facts. First, Kim et al.'s linear alpha step is exactly a regression of $\widehat{M}_{\beta}\bar{R}$ on $X$ (see Appendix D for detail). Second, Kim et al. allow for nonlinear characteristic functions in principle and show that updating both the characteristic values and the characteristic-to-mispricing relation matters empirically. This is also consistent with the broader motivation in Freyberger, Neuhierl, and Weber (2020), who emphasize that nonlinearities can matter in characteristic-return modeling.

The important limitation is that these ML versions do not inherit Kim et al.'s Theorem 2 or Theorem 3. They should be read as empirical nonlinear replacements for the alpha-regression step, evaluated by out-of-sample performance, factor regressions, exposure, turnover, and implementation-aware cost diagnostics.

Daniel Pellatt, 2026

# Appendix: Mathematical Details

## A. Normalized convergence convention

For vectors $a_N \in \mathbb{R}^N$, write

$$a_N = o_{p,N}(1)$$

when

$$\frac{1}{N} \| a_N \|^2 \xrightarrow{p} 0.$$

For matrices $A_N, B_N \in \mathbb{R}^{N\times m}$ with fixed $m$, Kim et al.'s large-$N$ convergence convention is

$$A_N \xrightarrow{p} B_N \quad \text{if} \quad \frac{1}{N}(A_N - B_N)'(A_N - B_N) \xrightarrow{p} 0_{m\times m}.$$

This is the convergence notion used below.

## B. The population factor residual-maker

For a single month $t$, write the model as

$$R_t = G_\alpha(X) + G_\beta(X) f_t + u_t,$$

where

$$u_t = \Gamma_\alpha + \Gamma_\beta f_t + e_t.$$

Define

$$P_\beta = G_\beta(X)(G_\beta(X)'G_\beta(X))^{-1}G_\beta(X)',$$

and

$$M_\beta = I_N - P_\beta.$$

Since $M_\beta G_\beta(X) = 0$,

$$M_\beta R_t = M_\beta G_\alpha(X) + M_\beta u_t.$$

Equivalently,

$$M_\beta R_t = G_\alpha(X) + u_t - P_\beta(G_\alpha(X) + u_t).$$

Let

$$q_t = G_\alpha(X) + u_t.$$

Under the same cross-sectional orthogonality assumptions used in Kim et al. - in particular Assumptions 2 and 3 -

$$\frac{1}{N} G_\beta(X)' q_t \xrightarrow{p} 0.$$

Also,

$$\frac{1}{N} G_\beta(X)' G_\beta(X) \to I_K.$$

Therefore,

$$\frac{1}{N} \| P_\beta q_t \|^2 = \left(\frac{1}{N} G_\beta(X)' q_t\right)' \left(\frac{1}{N} G_\beta(X)' G_\beta(X)\right)^{-1} \left(\frac{1}{N} G_\beta(X)' q_t\right) \xrightarrow{p} 0.$$

Hence,

$$M_\beta R_t = G_\alpha(X) + u_t + o_{p,N}(1).$$

This is the population residualized-target interpretation: after removing the characteristic-driven factor-loading space, the remaining cross-sectional return target is the mispricing function plus noise.

Daniel Pellatt, 2026

## C. Replacing $M_\beta$ by $\widehat{M}_\beta$

Let

$$\hat{P}_\beta = \hat{G}_\beta(X)(\hat{G}_\beta(X)'\hat{G}_\beta(X))^{-1}\hat{G}_\beta(X)',$$

and

$$\widehat{M}_\beta = I_N - \hat{P}_\beta.$$

Kim et al.'s Theorem 1 gives

$$\frac{1}{N}(\hat{G}_\beta - G_\beta)'(\hat{G}_\beta - G_\beta) \xrightarrow{p} 0_{K\times K}$$

in normalized cross-sectional distance. Let $\Delta_N = \hat{G}_\beta - G_\beta$, and let $\Delta_{N,k}$ denote its $k$th column. For a vector $y_N$ satisfying $N^{-1}y_N'y_N = O_p(1)$, define

$$a_N = \frac{1}{N}G_\beta' y_N, \qquad \hat{a}_N = \frac{1}{N}\hat{G}_\beta' y_N,$$

and

$$B_N = \frac{1}{N}G_\beta' G_\beta, \qquad \hat{B}_N = \frac{1}{N}\hat{G}_\beta'\hat{G}_\beta.$$

Then

$$\hat{a}_N - a_N = \frac{1}{N}(\hat{G}_\beta - G_\beta)' y_N = \frac{1}{N}\Delta_N' y_N.$$

For each component $k$,

$$\left|\frac{1}{N}\Delta_{N,k}' y_N\right| \le \left(\frac{1}{N}\Delta_{N,k}'\Delta_{N,k}\right)^{1/2}\left(\frac{1}{N}y_N'y_N\right)^{1/2} = o_p(1)$$

by Cauchy-Schwarz, since the first factor is $o_p(1)$ from Kim et al.'s Theorem 1 and the second is $O_p(1)$ by assumption. Hence

$$\hat{a}_N - a_N = o_p(1).$$

Likewise,

$$\hat{B}_N - B_N = \frac{1}{N}\Delta_N' G_\beta + \frac{1}{N}G_\beta'\Delta_N + \frac{1}{N}\Delta_N'\Delta_N.$$

For each $(i,j)$ element,

$$\left|\frac{1}{N}\Delta_{N,i}' G_{\beta,j}\right| \le \left(\frac{1}{N}\Delta_{N,i}'\Delta_{N,i}\right)^{1/2}\left(\frac{1}{N}G_{\beta,j}'G_{\beta,j}\right)^{1/2} = o_p(1),$$

since $B_N \to I_K$ implies $N^{-1}G_{\beta,j}'G_{\beta,j} = O_p(1)$. The same argument applies to $N^{-1}G_\beta'\Delta_N$, and also

$$\left|\frac{1}{N}\Delta_{N,i}'\Delta_{N,j}\right| \le \left(\frac{1}{N}\Delta_{N,i}'\Delta_{N,i}\right)^{1/2}\left(\frac{1}{N}\Delta_{N,j}'\Delta_{N,j}\right)^{1/2} = o_p(1).$$

Therefore,

$$\hat{B}_N - B_N = o_p(1),$$

and since $B_N \to I_K$, continuity of matrix inversion gives

$$\hat{B}_N^{-1} - B_N^{-1} = o_p(1).$$

Now write the population and estimated beta-space projections as

$$P_\beta y_N = G_\beta B_N^{-1} a_N, \qquad \hat{P}_\beta y_N = \hat{G}_\beta \hat{B}_N^{-1}\hat{a}_N.$$

Then

$$(\hat{P}_\beta - P_\beta)y_N = (\hat{G}_\beta - G_\beta)\hat{B}_N^{-1}\hat{a}_N + G_\beta(\hat{B}_N^{-1} - B_N^{-1})\hat{a}_N + G_\beta B_N^{-1}(\hat{a}_N - a_N).$$

Hence,

$$\frac{1}{\sqrt{N}} \| \left(\hat{P}_\beta - P_\beta\right) y_N \| \leq \frac{1}{\sqrt{N}} \| \hat{G}_\beta - G_\beta \|_F \| \hat{B}_N^{-1} \| \| \hat{a}_N \| + \frac{1}{\sqrt{N}} \| G_\beta \|_F \| \hat{B}_N^{-1} - B_N^{-1} \| \| \hat{a}_N \| + \frac{1}{\sqrt{N}} \| G_\beta \|_F \| B_N^{-1} \| \| \hat{a}_N - a_N \| = o_p(1),$$

because $N^{-1/2} \| \hat{G}_\beta - G_\beta \|_F = o_p(1)$, $N^{-1/2} \| G_\beta \|_F = O_p(1)$, $\hat{a}_N = O_p(1)$, and the inverse and cross-product differences are $o_p(1)$. Therefore,

$$\frac{1}{N} \| \left(\hat{P}_\beta - P_\beta\right) y_N \|^2 \xrightarrow{p} 0,$$

so

$$\widehat{M}_\beta y_N = M_\beta y_N + o_{p,N}(1).$$

Applying the result in Appendix B, taking $y_N = R_t$ gives

$$\widehat{M}_\beta R_t = G_\alpha(X) + u_t + o_{p,N}(1).$$

Similarly, taking $y_N = \bar{R}$ gives

$$\widehat{M}_\beta \bar{R} = G_\alpha(X) + \bar{u} + o_{p,N}(1),$$

where

$$\bar{u} = \frac{1}{T} \sum_{t=1}^{T} u_t.$$

This is the formal bridge between the Kim et al. residual-maker and the ML targets.

## D. Lemma 6 and the mean-target extension

Kim et al.'s Theorem 2 estimates the linear mispricing function by solving

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} (\bar{R} - X\theta)'(\bar{R} - X\theta) \quad \text{subject to} \quad \hat{G}_\beta(X)'X\theta = 0_K.$$

Kim et al.'s Lemma 6 gives the closed-form solution

$$\hat{\theta} = (X'X)^{-1}X'\bar{R} - (X'X)^{-1}X'\hat{G}_\beta(X)(\hat{G}_\beta(X)'\hat{G}_\beta(X))^{-1}\hat{G}_\beta(X)'\bar{R}.$$

Factoring the right-hand side,

$$\hat{\theta} = (X'X)^{-1}X'\left[I_N - \hat{G}_\beta(X)(\hat{G}_\beta(X)'\hat{G}_\beta(X))^{-1}\hat{G}_\beta(X)'\right]\bar{R}.$$

Therefore,

$$\hat{\theta} = (X'X)^{-1}X'\widehat{M}_\beta \bar{R}.$$

The fitted baseline mispricing function is

$$\hat{G}_\alpha^{\text{lin}}(X) = X\hat{\theta} = X(X'X)^{-1}X'\widehat{M}_\beta \bar{R}.$$

Thus, Kim et al.'s linear alpha step is exactly the projection of the residualized average-return target $\widehat{M}_\beta \bar{R}$ onto the span of $X$. The mean-target ML extension replaces this linear projection with a nonlinear regression map while keeping the same target.

## E. Relationship between mean and panel targets

Because $\widehat{M}_\beta$ is fixed within a given rolling estimation window,

$$\widehat{M}_\beta \bar{R} = \widehat{M}_\beta \left(\frac{1}{T} \sum_{t=1}^{T} R_t\right) = \frac{1}{T} \sum_{t=1}^{T} \widehat{M}_\beta R_t.$$

So the mean target is the time average of the month-level residualized targets.

The panel-target extension uses

$$\hat{y}_t^{\text{panel}} = \widehat{M}_\beta R_t$$

directly. By the result above,

$$\hat{y}_t^{\text{panel}} = G_\alpha(X) + u_t + o_{p,N}(1).$$

The mean-target version averages these noisy month-level targets before fitting the nonlinear map:

$$\hat{y}^{\text{mean}} = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t^{\text{panel}}.$$

This makes the mean-target extension closest to Kim et al.'s baseline alpha step. The panel-target extension exposes the learner to more observations and can use within-window variation in lagged characteristics, but it also trains on noisier month-level residualized targets.

## References

Kim, Soohun, Robert A. Korajczyk, and Andreas Neuhierl. 2021. "Arbitrage Portfolios." *Review of Financial Studies* 34(6): 2813-2856.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020. "Dissecting Characteristics Nonparametrically." *Review of Financial Studies* 33(5): 2326-2377.