Appendix for Cross-Validation for Longitudinal Datasets with Unstable Correlations

Meera Krishnamoorthy
meerak@umich.edu
University of Michigan
Division of Computer Science and Engineering
Ann Arbor, Michigan, USA

Michael Sjoding
msjoding@umich.edu
University of Michigan
Department of Internal Medicine
Ann Arbor, Michigan, USA

Jenna Wiens
wiensj@umich.edu
University of Michigan
Division of Computer Science and Engineering
Ann Arbor, Michigan, USA

CCS Concepts

• **Applied computing → Health informatics**; • **Computing methodologies → Cross-validation**.

Keywords

cross-validation, applied machine learning, model evaluation, model selection, healthcare, health informatics

ACM Reference Format:
Meera Krishnamoorthy, Michael Sjoding, and Jenna Wiens. 2025. Appendix for Cross-Validation for Longitudinal Datasets with Unstable Correlations. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/XXXXXX.XXXXXX

A Theoretical Analysis Appendix with One Unstable Feature

A.1 K-Fold Random CV Derivations

A.1.1 Stable Model Weight. For each validation fold k, we derive that the stable model weight (calculated using ordinary least squares for simple linear regression) is

$$\frac{\mathbb{E}_{t,i\in\tilde{T}(k)}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i\in\tilde{T}(k)}\left(\left(x_0^{(t,i)}\right)^2\right)} = \frac{\mathbb{E}_{t,i}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right)}$$

It follows from our data generation process that for all t, i

$$\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right) = \mathbb{E}_{t,i}\left(x_0^{(t,i)}\right)^2 + \mathrm{Var}\left(x_0^{(t,i)}\right) = 1$$

Thus

$$\begin{aligned}
&\frac{\mathbb{E}_{t,i}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right)} = \mathbb{E}_{t,i}\left(x_0^{(t,i)}y^{(t,i)}\right)\\
&= \mathbb{E}_{t,i}\left(x_0^{(t,i)}\left(ax_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)\\
&= \mathbb{E}_{t,i}\left(a\left(x_0^{(t,i)}\right)^2\right) + \mathbb{E}_{t,i}\left(bx_0^{(t,i)}x_1^{(t,i)}p_u^{(t,i)}\right)\\
&= a\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right) + b\mathbb{E}_{t,i}\left(x_0^{(t,i)}\right)\mathbb{E}_{t,i}\left(x_1^{(t,i)}\right)\mathbb{E}_{t,i}\left(p_u^{(t,i)}\right)\\
&= a
\end{aligned}$$

A.1.2 Expected MSE of Stable Model on All Validation Folds Derivation. For each validation fold k, we derive that the expected MSE of the stable model is

$$\begin{aligned}
&\mathbb{E}_{t,i}\left(\left(ax_0^{(t,i)} - ax_0^{(t,i)} - bp_u^{(t,i)}x_1^{(t,i)}\right)^2\right)\\
&= \mathbb{E}_{t,i}\left(\left(-bp_u^{(t,i)}x_1^{(t,i)}\right)^2\right)\\
&= b^2\mathbb{E}_{t,i}\left(\left(p_u^{(t,i)}\right)^2\right)\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)\\
&= b^2(A_u^2 + V_u)
\end{aligned}$$

A.1.3 Unstable Model Weight Derivation. For each validation fold k, we derive that the unstable model weight (calculated using ordinary least squares for simple linear regression) is

$$\frac{\mathbb{E}_{t,i\in\tilde{T}(k)}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i\in\tilde{T}(k)}\left(\left(x_1^{(t,i)}\right)^2\right)} = \frac{\mathbb{E}_{t,i}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)}$$

It follows from our data generation process that for all t, i

$$\mathbb{E}\left(\left(x_1^{(t,i)}\right)^2\right) = \mathbb{E}\left(x_1^{(t,i)}\right)^2 + \mathrm{Var}\left(x_1^{(t,i)}\right) = 1$$

KDD '25, August 3–7, 2025, Toronto, ON, Canada.

ACM ISBN 979-8-4007-1454-2/25/08
https://doi.org/10.1145/XXXXXX.XXXXXX

Thus

$$\frac{\mathbb{E}_{t,i}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)} = \mathbb{E}_{t,i}\left(x_1^{(t,i)}y^{(t,i)}\right)$$
$$= \mathbb{E}_{t,i}\left(x_1^{(t,i)}\left(ax_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)$$
$$= \mathbb{E}_{t,i}\left(ax_0^{(t,i)}x_1^{(t,i)}\right) + \mathbb{E}_{t,i}\left(b\left(x_1^{(t,i)}\right)^2 p_u^{(t,i)}\right)$$
$$= a\mathbb{E}_{t,i}\left(x_0^{(t,i)}\right)\mathbb{E}_{t,i}\left(x_1^{(t,i)}\right) + b\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)\mathbb{E}_{t,i}\left(p_u^{(t,i)}\right)$$
$$= bA_u$$

A.1.4 Expected MSE of Unstable Model on All Validation Folds Derivation.

$$\mathbb{E}_{t,i}\left(\left(A_u bx_1^{(t,i)} - ax_0^{(t,i)} - bp_u^{(t,i)}x_1^{(t,i)}\right)^2\right)$$
$$= \mathbb{E}_{t,i}\left(\left(b\left(A_u - p_u^{(t,i)}\right)x_1^{(t,i)} - ax_0^{(t,i)}\right)^2\right)$$
$$= b^2\mathbb{E}_{t,i}\left(\left(A_u - p_u^{(t,i)}\right)^2\left(x_1^{(t,i)}\right)^2\right) - 2ab\mathbb{E}_{t,i}\left(x_0^{(t,i)}\left(A_u - p_u^{(t,i)}\right)x_1(t,i)\right)$$
$$+ a^2\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right)$$
$$= b^2\mathbb{E}_{t,i}\left(\left(A_u - p_u^{(t,i)}\right)^2\right)\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)$$
$$- 2ab\mathbb{E}_{t,i}\left(x_0^{(t,i)}\right)\mathbb{E}_{t,i}\left(A_u - p_u^{(t,i)}\right)\mathbb{E}_{t,i}\left(x_1^{(t,i)}\right)$$
$$+ a^2\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right)$$
$$= b^2\mathbb{E}_{t,i}\left(A_u^2 - 2A_u p_u^{(t,i)} + \left(p_u^{(t,i)}\right)^2\right) + a^2$$
$$= b^2\left(A_u^2 - 2A_u\mathbb{E}_{t,i}\left(p_u^{(t,i)}\right) + \mathbb{E}_{t,i}\left(\left(p_u^{(t,i)}\right)^2\right)\right) + a^2$$
$$= b^2\left(A_u^2 - 2A_u^2 + A_u^2 + V_u\right) + a^2$$
$$= b^2 V_u + a^2$$

A.2 K-Fold Block CV Derivations

A.2.1 Stable Model Weight. For each validation fold k, we derive that the stable model weight (calculated using ordinary least squares for simple linear regression) is

$$\frac{\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_0^{(t,i)}\right)^2\right)}$$

It follows from our data generation process that for all t, i

$$\mathbb{E}\left(\left(x_0^{(t,i)}\right)^2\right) = \mathbb{E}\left(x_0^{(t,i)}\right)^2 + \mathrm{Var}\left(x_0^{(t,i)}\right) = 1$$

Thus

$$\frac{\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_0^{(t,i)}\right)^2\right)} = \mathbb{E}_{t\neq k}\left(x_0^{(t,i)}y^{(t,i)}\right)$$
$$= \mathbb{E}_{t\neq k}\left(x_0^{(t,i)}\left(ax_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)$$
$$= \mathbb{E}_{t\neq k}\left(a\left(x_0^{(t,i)}\right)^2\right) + \mathbb{E}_{t\neq k}\left(bx_0^{(t,i)}x_1^{(t,i)}p_u^{(t,i)}\right)$$
$$= a\mathbb{E}_{t\neq k}\left(\left(x_0^{(t,i)}\right)^2\right)$$
$$+ b\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(p_u^{(t,i)}\right)$$
$$= a$$

A.2.2 Expected MSE of Stable Model on on All Validation Folds Derivation. Same as the expected MSE of stable model on all validation folds estimated by K=fold random CV: $b^2(A_u^2 + V_u)$

A.2.3 Unstable Model Weight Derivation. For each validation fold k, we derive that the unstable model weight (calculated using ordinary least squares for simple linear regression) is

$$\frac{\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_1^{(t,i)}\right)^2\right)}$$

It follows from our data generation process that for all t, i

$$\mathbb{E}\left(\left(x_1^{(t,i)}\right)^2\right) = \mathbb{E}\left(x_1^{(t,i)}\right)^2 + \mathrm{Var}\left(x_1^{(t,i)}\right) = 1$$

Thus

$$\frac{\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_1^{(t,i)}\right)^2\right)} = \mathbb{E}_{t\neq k}\left(x_1^{(t,i)}y^{(t,i)}\right)$$
$$= \mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\left(ax_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)$$
$$= \mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\left(ax_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)$$
$$= \mathbb{E}_{t\neq k}\left(ax_0^{(t,i)}x_1^{(t,i)}|t\neq k\right) + \mathbb{E}_{t\neq k}\left(b\left(x_1^{(t,i)}\right)^2 p_u^{(t,i)}\right)$$
$$= a\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\right)$$
$$+ b\mathbb{E}_{t\neq k}\left(\left(x_1^{(t,i)}\right)^2\right)\mathbb{E}_{t\neq k}\left(p_u^{(t,i)}\right)$$
$$= b\mathbb{E}_{t\neq k}\left(p_u^{(t,i)}\right)$$
$$= b\frac{KA_u - p_u^{(k)}}{K-1}$$

A.2.4 Expected MSE of Unstable Model on Validation Fold k Derivation.

$$
\begin{aligned}
&\mathbb{E}_i\left(\left(b\frac{KA_u - p_u^{(k)}}{K-1}x_1^{(k,i)} - ax_0^{(k,i)} - bp_u^{(k,i)}x_1^{(k,i)}\right)^2\right)\\
&= \mathbb{E}_i\left(\left(\left(b\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)\right)x_1^{(k,i)} - ax_0^{(k,i)}\right)^2\right)\\
&= \mathbb{E}_i\left(\left(\left(b\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)\right)x_1^{(k,i)}\right)^2\right)\\
&\quad ab\mathbb{E}_i\left(x_0^{(t,i)}\right)\mathbb{E}_i\left(x_1^{(t,i)}\right)\mathbb{E}_i\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)\\
&\quad + a^2\mathbb{E}_i\left(\left(x_0^{(k,i)}\right)^2\right)\\
&= b^2\mathbb{E}_i\left(\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)^2\right) + a^2\\
&= \frac{b^2}{(K-1)^2}\mathbb{E}_i\left(\left(KA_u - p_u^{(k)} - (K-1)p_u^{(k,i)}\right)^2\right) + a^2\\
&= \frac{b^2}{(K-1)^2}\mathbb{E}_i\left(K^2A_u^2 - 2KA_up_u^{(k)} - 2KA_u(K-1)p_u^{(k,i)} + \left(p_u^{(k)}\right)^2\right)\\
&\quad + \frac{b^2}{K-1}\mathbb{E}_i\left(2p_u^{(k)}(K-1)p_u^{(k,i)} + \left((K-1)p_u^{(k,i)}\right)^2\right) + a^2\\
&= \frac{b^2}{(K-1)^2}\left(K^2A_u^2 - 2KA_up_u^{(k)} - 2KA_u(K-1)p_u^{(k)} + \left(p_u^{(k)}\right)^2\right)\\
&\quad + \frac{b^2}{K-1}\left(2(K-1)\left(p_u^{(k)}\right)^2 + \left((K-1)p_u^{(k)}\right)^2\right) + a^2\\
&= \frac{b^2}{(K-1)^2}\left(K^2A_u^2 - p_u^{(k)}\left(2KA_u + 2KA_u(K-1)\right)\right)\\
&\quad + \frac{b^2}{K-1}\left(\left(p_u^{(k)}\right)^2\left(1 + 2(K-1) + (K-1)^2\right)\right) + a^2\\
&= \frac{b^2}{(K-1)^2}\left(K^2A_u^2 - 2K^2A_up_u^{(k)}\right)\\
&\quad + \frac{b^2}{K-1}K^2\left(p_u^{(k)}\right)^2 + a^2\\
&= \frac{b^2K^2}{(K-1)^2}\left(A_u^2 - 2A_up_u^{(k)} + \left(p_u^{(k)}\right)^2\right) + a^2
\end{aligned}
$$

A.2.5 Expected MSE of Unstable Model on All Validation Folds.

$$
\begin{aligned}
&\mathbb{E}_t\left(\frac{b^2K^2}{(K-1)^2}\left(A_u^2 - 2A_up_u^{(t)} + \left(p_u^{(t)}\right)^2\right) + a^2\right)\\
&= \frac{b^2K^2}{(K-1)^2}\left(A_u^2 - 2A_u^2 + A_u^2 + V_u^2\right) + a^2\\
&= \frac{b^2K^2}{(K-1)^2}V_u + a^2
\end{aligned}
$$

A.2.6 When will the unstable hyperparameter be selected?

$$
\begin{aligned}
\frac{b^2K^2}{(K-1)^2}V_u + a^2 &< b^2(A_u^2 + V_u)\\
b^2V_u\left(\frac{K^2 - K^2 + 2K - 1}{(K-1)^2}\right) + a^2 &< b^2A_u^2\\
b^2V_u\left(\frac{2K-1}{(K-1)^2}\right) &< b^2A_u^2 - a^2\\
V_u &< \frac{(K-1)^2}{2K-1}\left(A_u^2 - \frac{a^2}{b^2}\right)
\end{aligned}
$$

B Theoretical Analysis Appendix with Two Unstable Features

In this setting, given constants A_s and V_s such that $0 < A_s < 1$ and $0 < V_s < V_u$, and random variables

$$
\begin{aligned}
p_s^{(t)} &\sim \text{Beta}(\mu = A_s, \sigma^2 = V_s)\\
p_s^{(t,i)} &\sim \text{Bernoulli}\left(p_s^{(t)}\right)
\end{aligned}
$$

We define the relationship between $y^{(t,i)}$ and $\mathbf{x}^{(t,i)}$ to be

$$
y^{(t,i)} = ax_0^{(t,i)}p_s^{(t,i)} + bx_1^{(t,i)}p_u^{(t,i)}
$$

B.1 K-Fold Random CV Derivations

B.1.1 Stable Model Weight.

$$
\begin{aligned}
&\frac{\mathbb{E}_{t,i}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right)} = \mathbb{E}_{t,i}\left(x_0^{(t,i)}y^{(t,i)}\right)\\
&= \mathbb{E}_{t,i}\left(x_0^{(t,i)}\left(ap_s^{(t,i)}x_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)\\
&= \mathbb{E}_{t,i}\left(ap_s^{(t,i)}\left(x_0^{(t,i)}\right)^2\right) + \mathbb{E}_{t,i}\left(bx_0^{(t,i)}x_1^{(t,i)}p_u^{(t,i)}\right)\\
&= a\mathbb{E}_{t,i}\left(p_s^{(t,i)}\right)\mathbb{E}_{t,i}\left(\left(x_0^{(t,i)}\right)^2\right)\\
&\quad + b\mathbb{E}_{t,i}\left(x_0^{(t,i)}\right)\mathbb{E}_{t,i}\left(x_1^{(t,i)}\right)\mathbb{E}_{t,i}\left(p_u^{(t,i)}\right)\\
&= aA_s
\end{aligned}
$$

B.1.2 Expected MSE of Stable Model on All Validation Folds Derivation.

$$
\begin{aligned}
&\mathbb{E}_{t,i}\left(\left(aA_s x_0^{(t,i)} - ap_s^{(t,i)}x_0^{(t,i)} - bp_u^{(t,i)}x_1^{(t,i)}\right)^2\right)\\
=&\mathbb{E}_{t,i}\left(\left(a\left(A_s - p_s^{(t,i)}\right)x_0^{(t,i)} - bp_u^{(t,i)}x_1^{(t,i)}\right)^2\right)\\
=&\mathbb{E}_{t,i}\left(a^2\left(A_s - p_s^{(t,i)}\right)^2\left(x_0^{(t,i)}\right)^2 + 2ab\left(A_s - p_s^{(t,i)}\right)p_u^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\right.\\
&\left.+b^2\left(p_u^{(t,i)}\right)^2\left(x_1^{(t,i)}\right)^2\right)\\
=&\mathbb{E}_{t,i}\left(a^2\left(A_s - p_s^{(t,i)}\right)^2\left(x_0^{(t,i)}\right)^2\right)\\
&+\mathbb{E}_{t,i}\left(2ab\left(A_s - p_s^{(t,i)}\right)p_u^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\right)\\
&+\mathbb{E}_{t,i}\left(b^2\left(p_u^{(t,i)}\right)^2\left(x_1^{(t,i)}\right)^2\right)\\
=&a^2\mathbb{E}_{t,i}\left(A_s^2 - 2A_s p_s^{(t,i)} + \left(p_s^{(t,i)}\right)^2\right) + b^2(A_u^2 + V_u)\\
=&a^2\left(A_s^2 - 2A_s^2 + A_s^2 + V_s\right) + b^2(A_u^2 + V_u)\\
=&a^2 V_s + b^2(A_u^2 + V_u)
\end{aligned}
$$

B.1.3 Unstable Model Weight Derivation.

$$
\begin{aligned}
&\frac{\mathbb{E}_{t,i}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)} = \mathbb{E}_{t,i}\left(x_1^{(t,i)}y^{(t,i)}\right)\\
&= \mathbb{E}_{t,i}\left(x_1^{(t,i)}\left(ap_s^{(t,i)}x_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)\\
&= \mathbb{E}_{t,i}\left(ap_s^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\right) + \mathbb{E}_{t,i}\left(b\left(x_1^{(t,i)}\right)^2 p_u^{(t,i)}\right)\\
&= a\mathbb{E}_{t,i}\left(p_s^{(t,i)}\right)\mathbb{E}_{t,i}\left(x_0^{(t,i)}\right)\mathbb{E}_{t,i}\left(x_1^{(t,i)}\right) + b\mathbb{E}_{t,i}\left(\left(x_1^{(t,i)}\right)^2\right)\mathbb{E}_{t,i}\left(p_u^{(t,i)}\right)\\
&= bA_u
\end{aligned}
$$

B.1.4 Expected MSE of Unstable Model on All Validation Folds Derivation.

$$
\begin{aligned}
&\mathbb{E}_{t,i}\left(\left(bA_u x_1^{(t,i)} - ap_s^{(t,i)}x_0^{(t,i)} - bp_u^{(t,i)}x_1^{(t,i)}\right)^2\right)\\
=&\mathbb{E}_{t,i}\left(\left(b\left(A_u - p_u^{(t,i)}\right)x_1^{(t,i)} - ap_s^{(t,i)}x_0^{(t,i)}\right)^2\right)\\
=&\mathbb{E}_{t,i}\left(b^2\left(A_u - p_u^{(t,i)}\right)^2\left(x_1^{(t,i)}\right)^2 + 2ab\left(A_u - p_u^{(t,i)}\right)p_s^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\right.\\
&\left.+a^2\left(p_s^{(t,i)}\right)^2\left(x_0^{(t,i)}\right)^2\right)\\
=&\mathbb{E}_{t,i}\left(b^2\left(A_u - p_u^{(t,i)}\right)^2\left(x_1^{(t,i)}\right)^2\right)\\
&+\mathbb{E}_{t,i}\left(2ab\left(A_u - p_u^{(t,i)}\right)p_s^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\right)\\
&+\mathbb{E}_{t,i}\left(a^2\left(p_s^{(t,i)}\right)^2\left(x_0^{(t,i)}\right)^2\right)\\
=&b^2\mathbb{E}_{t,i}\left(A_u^2 - 2A_u p_u^{(t,i)} + \left(p_u^{(t,i)}\right)^2\right) + a^2(A_s^2 + V_s)\\
=&b^2\left(A_u^2 - 2A_u^2 + A_u^2 + V_u\right) + a^2(A_s^2 + V_s)\\
=&b^2 V_u + a^2(A_s^2 + V_s)
\end{aligned}
$$

B.2 K-Fold Block CV Derivations

B.2.1 Stable Model Weight. For each validation fold k, we derive that the stable model weight (calculated using ordinary least squares for simple linear regression) is

$$
\frac{\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_0^{(t,i)}\right)^2\right)}
$$

It follows from our data generation process that for all t, i

$$
\mathbb{E}\left(\left(x_0^{(t,i)}\right)^2\right) = \mathbb{E}\left(x_0^{(t,i)}\right)^2 + \mathrm{Var}\left(x_0^{(t,i)}\right) = 1
$$

Thus

$$
\begin{aligned}
&\frac{\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_0^{(t,i)}\right)^2\right)} = \mathbb{E}_{t\neq k}\left(x_0^{(t,i)}y^{(t,i)}\right)\\
&= \mathbb{E}_{t\neq k}\left(x_0^{(t,i)}\left(ap_s^{(t,i)}x_0^{(t,i)} + bp_u^{(t,i)}x_1^{(t,i)}\right)\right)\\
&= \mathbb{E}_{t\neq k}\left(ap_s^{(t,i)}\left(x_0^{(t,i)}\right)^2\right) + \mathbb{E}_{t\neq k}\left(bx_0^{(t,i)}x_1^{(t,i)}p_u^{(t,i)}\right)\\
&= a\mathbb{E}_{t\neq k}\left(p_s^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(\left(x_0^{(t,i)}\right)^2\right)\\
&+ b\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(p_u^{(t,i)}\right)\\
&= a\frac{KA_s - p_s^{(k)}}{K-1}
\end{aligned}
$$

B.2.2 Expected MSE of Stable Model on All Validation Folds.

$$
\begin{aligned}
&\mathbb{E}_{k,i}\left(\left(a\frac{KA_s-p_s^{(k)}}{K-1}x_0^{(k,i)}-ap_s^{(k,i)}x_0^{(k,i)}-bp_u^{(k,i)}x_1^{(k,i)}\right)^2\right)\\
=&\mathbb{E}_{k,i}\left(\left(a\left(\frac{KA_s-p_s^{(k)}}{K-1}-p_s^{(k,i)}\right)x_0^{(k,i)}-bp_u^{(k,i)}x_1^{(k,i)}\right)^2\right)\\
=&\mathbb{E}_{k,i}\left(a^2\left(\frac{KA_s-p_s^{(k)}}{K-1}-p_s^{(k,i)}\right)^2\left(x_0^{(k,i)}\right)^2\right.\\
&-2ab\left(\frac{KA_s-p_s^{(k)}}{K-1}-p_s^{(k,i)}\right)p_u^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\\
&\left.+\ b^2\left(p_u^{(k,i)}\right)^2\left(x_1^{(k,i)}\right)^2\right)\\
=&\mathbb{E}_{k,i}\left(a^2\left(\frac{KA_s-p_s^{(k)}}{K-1}-p_s^{(k,i)}\right)^2\left(x_0^{(k,i)}\right)^2\right)\\
&-\mathbb{E}_{k,i}\left(2ab\left(\frac{KA_s-p_s^{(k)}}{K-1}-p_s^{(k,i)}\right)p_u^{(t,i)}x_0^{(t,i)}x_1^{(t,i)}\right)\\
&+\mathbb{E}_{k,i}\left(b^2\left(p_u^{(k,i)}\right)^2\left(x_1^{(k,i)}\right)^2\right)\\
=&\frac{a^2}{K-1}\mathbb{E}_{k,i}\left(\left(KA_s-p_s^{(k)}-(K-1)p_s^{(k,i)}\right)^2\right)+b^2(A_u^2+V_u)\\
=&\frac{a^2}{(K-1)^2}\left(\mathbb{E}_{k,i}\left((KA_s)^2\right)+\mathbb{E}_{k,i}\left(\left(p_s^{(k)}\right)^2\right)+\mathbb{E}_{k,i}\left((K-1)^2\left(p_s^{(k,i)}\right)^2\right)\right.\\
&\left.-\mathbb{E}_{k,i}\left(2KA_sp_s^{(k)}\right)-\mathbb{E}_{k,i}\left(2KA_s(K-1)p_s^{(k,i)}\right)+\mathbb{E}_{k,i}\left(2(K-1)p_s^{(k)}p_s^{(k,i)}\right)\right)\\
&+b^2(A_u^2+V_u)\\
=&\frac{a^2}{(K-1)^2}\left(K^2A_s^2+A_s^2+V_s+(K-1)^2(A_s^2+V_s)\right.\\
&\left.-2KA_s^2-2K(K-1)A_s^2+2(K-1)(A_s^2+V_s)\right)\\
&+b^2(A_u^2+V_u)\\
=&\frac{a^2}{(K-1)^2}\left(A_s^2(K^2-2K-2K(K-1))+(A_s^2+V_s)(1+(K-1)^2+2(K-1))\right)\\
&+b^2(A_u^2+V_u)\\
=&\frac{a^2}{(K-1)^2}\left(A_s^2(-K^2)+(A_s^2+V_s)(K^2)\right)+b^2(A_u^2+V_u)\\
=&\frac{a^2K^2V_s}{(K-1)^2}+b^2(A_u^2+V_u)
\end{aligned}
$$

B.2.3 Unstable Model Weight Derivation. For each validation fold k, we derive that the unstable model weight (calculated using ordinary least squares for simple linear regression) is

$$
\frac{\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_1^{(t,i)}\right)^2\right)}
$$

It follows from our data generation process that for all t, i

$$
\mathbb{E}\left(\left(x_1^{(t,i)}\right)^2\right)=\mathbb{E}\left(x_1^{(t,i)}\right)^2+\mathrm{Var}\left(x_1^{(t,i)}\right)=1
$$

Thus

$$
\begin{aligned}
&\frac{\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}y^{(t,i)}\right)}{\mathbb{E}_{t\neq k}\left(\left(x_1^{(t,i)}\right)^2\right)}=\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}y^{(t,i)}\right)\\
&=\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\left(ap_s^{(t,i)}x_0^{(t,i)}+bp_u^{(t,i)}x_1^{(t,i)}\right)\right)\\
&=\mathbb{E}_{t\neq k}\left(bp_u^{(t,i)}\left(x_1^{(t,i)}\right)^2\right)+\mathbb{E}_{t\neq k}\left(ax_0^{(t,i)}x_1^{(t,i)}p_s^{(t,i)}\right)\\
&=b\mathbb{E}_{t\neq k}\left(p_u^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(\left(x_1^{(t,i)}\right)^2\right)\\
&+a\mathbb{E}_{t\neq k}\left(x_0^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(x_1^{(t,i)}\right)\mathbb{E}_{t\neq k}\left(p_s^{(t,i)}\right)\\
&=b\frac{KA_u-p_u^{(k)}}{K-1}
\end{aligned}
$$

B.2.4 Expected MSE of Unstable Model on All Validation Folds.

$$
\begin{aligned}
&\mathbb{E}_{k,i}\left(\left(b\frac{KA_u - p_u^{(k)}}{K-1}x_1^{(k,i)} - ap_s^{(k,i)}x_0^{(k,i)} - bp_u^{(k,i)}x_1^{(k,i)}\right)^2\right)\\
=&\mathbb{E}_{k,i}\left(\left(b\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)x_1^{(k,i)} - ap_s^{(k,i)}x_0^{(k,i)}\right)^2\right)\\
=&\mathbb{E}_{k,i}\left(b^2\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)^2\left(x_1^{(k,i)}\right)^2\right.\\
&+2ab\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)x_1^{(k,i)}p_s^{(k,i)}x_0^{(k,i)}\\
&\left.+a^2\left(p_s^{(k,i)}\right)^2\left(x_0^{(k,i)}\right)^2\right)\\
=&\mathbb{E}_{k,i}\left(b^2\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)^2\left(x_1^{(k,i)}\right)^2\right)\\
&+\mathbb{E}_{k,i}\left(2ab\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)x_1^{(k,i)}p_s^{(k,i)}x_0^{(k,i)}\right)\\
&+\mathbb{E}_{k,i}\left(a^2\left(p_s^{(k,i)}\right)^2\left(x_0^{(k,i)}\right)^2\right)\\
=&b^2\mathbb{E}_{k,i}\left(\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)^2\right)\mathbb{E}_{k,i}\left(\left(x_1^{(k,i)}\right)^2\right)\\
&+a^2\mathbb{E}_{k,i}\left(\left(p_s^{(k,i)}\right)^2\right)\mathbb{E}_{k,i}\left(\left(x_0^{(k,i)}\right)^2\right)\\
=&b^2\mathbb{E}_{k,i}\left(\left(\frac{KA_u - p_u^{(k)}}{K-1} - p_u^{(k,i)}\right)^2\right)\\
&+a^2\left(A_s^2 + V_s\right)\\
=&\frac{b^2K^2V_u}{(K-1)^2} + a^2\left(A_s^2 + V_s\right)
\end{aligned}
$$

B.3 Proposed Approach Derivations

Our proposed approach will select the stable hyperparameter when

$$|R_s - B_s| < |R_u - B_u|$$

$$
\begin{aligned}
R_s &= a^2V_s + b^2(A_u^2 + V_u)\\
B_s &= b^2(A_u^2 + V_u) + \frac{a^2K^2V_s}{(K-1)^2}\\
|B_s - R_s| &= \frac{a^2(2K-1)}{(K-1)^2}V_s\\
R_u &= a^2(A_s^2 + V_s) + b^2V_u\\
B_u &= a^2(A_s^2 + V_s) + \frac{b^2K^2V_u}{(K-1)^2}\\
|B_u - R_s| &= \frac{b^2(2K-1)}{(K-1)^2}V_u
\end{aligned}
$$

$|B_s - R_s| < |B_u - R_u|$ when

$$a^2V_s < b^2V_u$$

C Implementation Details Appendix

Algorithm 1 outlines our implementation of our proposed approach.

Algorithm 1 Proposed Approach Implementation.

Inputs: dataset D, set of sets of hyperparameters $\{M_j\}_{j=1}^m$; functions l_1, l_2; function types (either loss or accuracy) l_1Type, l_2Type

Parameters: threshold T
Definitions: K-fold block CV's estimate of model performance given dataset D, performance measurement function l, and set of hyperparameters M_j: $B(D, M_j, l)$; K-fold random CV's estimate of model performance given dataset D, performance measurement function l, and set of hyperparameters M_j: $R(D, M_j, l)$

```
CandidateH = []
for j ∈ {1, ..., m} do
    if l1Type = Loss and B(D, Mj, l1) < T then
        CandidateH.append(j)
    end if
    if l1Type = Accuracy and B(D, Mj, l1) > T then
        CandidateH.append(j)
    end if
end for
FinalH, FinalP = −1, 10000
for j ∈ CandidateH do
    if |B(D, Mj, l2) − R(D, Mj, l2)| < FinalP then
        FinalH = j
        FinalP = |B(D, Mj, l2) − R(D, Mj, l2)|
    end if
    if |B(D, Mj, l2) − R(D, Mj, l2)| = FinalP then
        if l2type = Loss and B(D, Mj, l2) < B(M_FinalH, l2) then
            FinalH = j
            FinalP = |B(D, Mj, l2) − R(D, Mj, l2)|
        end if
        if l2type = Accuracy and B(D, Mj, l2) > B(M_FinalH, l2)
        then
            FinalH = j
            FinalP = |B(D, Mj, l2) − R(D, Mj, l2)|
        end if
    end if
end for

return FinalH
```

D Experimental Setup Appendix

D.1 Proposed Approach Details

For all experiments, our proposed approach removes hyperparameters from consideration whose corresponding models have an MSE estimated by K-fold block CV of above 100 (for our synthetic data tasks) or an AUROC estimated by K-fold block CV of below 0.5 (for our semi-synthetic and real data tasks). Furthermore, if multiple hyperparameters result in the same estimate of model performance via our proposed approach, we select the hyperparameter whose corresponding model has the lowest MSE (for our synthetic data task) or negative log loss (for our semi-synthetic and real data tasks) estimated by K-fold block CV.

D.2 Software Details

All experiments were run on a distribution of Ubuntu 20.04.6 LTS with Python 3.8.10. For all experiments, we use scikit-learn 1.1.1 [?], scipy 1.10.1 (1.8.1) [?], numpy 1.22.4 [?], and pandas 1.4.2 [?]. Matplotlib 3.5.2 [?] and seaborn 0.13.2 [?] were used to generate figures.

D.3 Synthetic Data Task

For all datasets, we fix $K+1=10$ and $N=1000$. For each dataset, we generate a value for b between 0 and 10. We then generate a value for a between 0 and one of the following values:

$$\left\{\frac{b\sqrt{A_u(2-A_u)}}{10}, \frac{b\sqrt{A_u(2-A_u)}}{5}, \frac{b\sqrt{A_u(2-A_u)}}{2}, \frac{b\sqrt{A_u(2-A_u)}}{1.1}\right\}$$

, thus varying the maximum possible value for a, which we call $\max(a)$. We define $\max(a)$ such that the stable model has better performance than the unstable model when $p_u^{(t,i)}=1$, but worse performance than the unstable model when $p_u^{(t,i)}=0$.

Claim 1. *When $a^2 < b^2A_u(2-A_u)$ the stable model will have a lower mean squared error than the unstable model when $p_u^{(t,i)}=1$; and the stable model will have a higher MSE than the unstable model when $p_u^{(t,i)}=1$.*

Proof 1. *Trained on all training data, the stable model weight a and the unstable model weight is $A_u b$. When the input data is of the form $y^{(t,i)}=ax_0^{(t,i)}$ (i.e., $p_u^{(t,i)}=0$), the stable model's expected MSE is 0, but the unstable model's expected MSE is $a^2+A_u^2b^2$. Given that $a>0$, $a^2+A_u^2b^2$ is always greater than 0.*

When the input data is of the form $y^{(t,i)}=ax_0^{(t,i)}+bx_1^{(t,i)}$ (i.e., $p_u^{(t,i)}=1$), the stable model's expected MSE is b^2. The unstable model's expected MSE is $a^2+(-1+A_u)^2b^2$. $a^2+(-1+A_u)^2b^2<b^2$ when $a^2<b^2A_u(2-A_u)$.

Furthermore, for each dataset, either A_u or V_u is fixed, where $A_u \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$ and $V_u \in \{0.010, 0.575, 0.105, 0.152, 0.240\}$. We generate 20,000 datasets for all A_u and $\max(a)$ combinations, and 20,000 datasets for all V_u and $\max(a)$ combinations,

We selected these values of $\max(a), A_u$, and V_u such that our synthetic datasets contain a large variety of values of A_u, V_u, $\frac{a}{b}$ and $\frac{(A_u^2b^2-a^2)(K-1)^2}{b^2(2K-1)}$, and that we can use these datasets to examine the performance of all approaches when $A_u<\frac{a}{b}$ (the condition under which K-fold random CV selects the stable hyperparameter); $A_u \geq \frac{a}{b}$ (the condition under which K-fold random CV selects the unstable hyperparameter); $V_u < \frac{(K-1)^2}{2K-1}\left(A_u^2-\frac{a^2}{b^2}\right)$, the condition under which K-fold block CV selects the unstable hyperparameter; and $V_u \geq \frac{(A_u^2b^2-a^2)(K-1)^2}{b^2(2K-1)}$, the condition under which K-fold block CV selects the stable hyperparameter.

D.4 Semi Synthetic Data Task

Following past work, our inclusion criteria are patients whose length of stay is at least 24 hours; we use the first 24 hours of each patient's stay to predict whether their length of stay will exceed 3 days; and the data resolution for time series data is 2 hours [?]. MIMIC-IV places each sample into a three-year time range. Following past work, we approximate the year of each sample by taking the midpoint of its time range [?].

For each of the semi-synthetic datasets, the number of time periods within the dataset, $K+1$, is 4. Data corresponding to time $1 \leq t \leq K+1$ are from years $2008+3(t-1)$ to $2011+3(t-1)$. Data from $t \leq K$ is used for training models and data from $t=K+1$ is used for testing. For each of the semi-synthetic datasets, we fix N in the training set to be 100,000.

Each dataset differs in the average and variance of the correlation between a and y. To vary the average and variance of this correlation, in each training time period, we vary the degree to which we upsample patients whose $y=a$ and downsample patients whose $y \neq a$. Specifically, for each of the training time periods, we sample $M \in \{0, 5000, 10000, 23000, 34000, 50000\}$ data where $y \neq a$ and $100000-M$ data where $y=a$. We do this for all combinations of M for all training time periods except the ones where M is equal for all training time periods. For all semi-synthetic data tasks, we use the same test set that we resample such that $N=100,000$ and there are an equal number of data where $y=a$ and $y \neq a$. The correlation between y and a in the test set is -0.048.

Following past work in developing models to predict length of stay on MIMIC-IV data, the stable model uses the following features: all diagnoses and the labs/vitals from the list provided by Wang et al [? ?]. The features are pre-processed using the pre-processing pipeline in Gupta et al[?]. We define the unstable model to take in the same features as the stable model as well as the feature a. The random seed used to train the stable and unstable models on all tasks was set to be 0.

D.5 Real Data Task

The Surveillance, Epidemiology, and End Results (SEER) Incidence Database contains multiple features (patient demographics, tumor sites, courses of treatment, etc). and the survival outcomes of 47% of cancer patients in the U.S. from 1975 to 2018 [?]. We consider 300 real data tasks. Each task differs in the training data, the random seed used to train all models and split the folds for K-fold random CV, and the test set. We consider multiple random seeds after observing that the performance of K-fold random CV changes significantly depending on the random seed.

Specifically, we consider 6 different training sets, each that contain four years of data, following past work that used this method of selecting training data from the SEER database [?]: data from 1990-1994, 1991-1995, 1992-1996, 1993-1997, 1994-1998, 1995-1999. For each training set, the number of time periods in the training set $K=4$. We consider 5 random seeds (0-4). We consider 10 test sets, each of which contain data from a single year ranging from 2000-2009.

We define our cohorts and the total set of features available to each model uses in the same way as past work [?], including the first entry for each patient ID with the cancer site of the lung or bronchus; and excluding all individuals with invalid survival months and unknown causes of death. The total set of features, which we take directly from past work, is below.

The name of all features used to train all models for our real data task are in our Github repository.

Each model selection method is selecting among models with the following L1 regularization parameter values

$$\left(C \in \{10^5, 10^4, 10^3, 10^2, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}\right)$$

E Overall Results Appendix

The output of our proposed approach—the instability signal, measured as the absolute difference between block and random CV losses ($|B - R|$)—shows a clear decreasing trend across settings. We observe the largest discrepancy in the synthetic tasks (median $|B - R| = 0.59$), a smaller one in the semi-synthetic tasks (0.05), and the smallest in the real-world setting (1.6×10^{-5}). This pattern suggests that temporal variability—and thus the potential for unstable feature-outcome relationships—is highest in synthetic settings and lowest in real clinical data.

Received 10 February 2025; accepted 15 May 2025