

Group-Sparse Smoothing for Longitudinal Models with Time-Varying Coefficients

TV-Select: research gap, structural selection, and post-selection
inference

Yu Lu¹, Tianni Zhang¹, Yuyao Wang¹, Mengfei Ran¹

¹ Academy of Life and Natural Sciences, Xi'an Jiaotong-Liverpool University

ECONOMETRICS AND STATISTICS Conference Presentation



Xi'an Jiaotong-Liverpool University

西交利物浦大學



Talk Roadmap

- ① Motivation and Research Gap
- ② Methodology
- ③ Theory and Proof Ideas
- ④ Empirical Evidence
- ⑤ Conclusion

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Motivation and Research Gap

Research Question: Three Types of Effects

For each covariate, is its effect zero, constant, or genuinely time varying?

Zero effect

$$\beta_k(t) \equiv 0$$

Irrelevant predictor

Constant effect

$$\beta_k(t) \equiv \mu_k \neq 0$$

Stable association

Varying effect

$$\beta_k(t) = \mu_k + g_k(t)$$

Temporal association

- Making all effects dynamic inflates variance and can produce spurious time patterns.
- The target is **structural identification**, not prediction alone.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

What Has the Literature Already Solved?

Literature	Strength	Remaining gap
VC models	Smooth effects	Which effects truly vary?
Group penalties	Select blocks	Constant is not zero
P-splines	Control roughness	Identifiable function norms
Clustered inference	Cluster-aware SEs	Inference after selection
Examples: Hastie–Tibshirani (1993); Wang et al. (2008); Antoniadis et al. (2012).		

Missing: an integrated three-class selection, normalized smoothing, and cluster-robust inference framework.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Research Gap I: The Selection Target Matters

Gap 1: Nonzero does not mean time varying

Selecting β_k as a whole conflates nonzero constant and changing effects.

Gap 2: Redundant spline coordinates

Raw centered B-splines have a null direction, obscuring coefficient-norm interpretation.

$$\beta_k(t) = \underbrace{\mu_k}_{\text{constant mean}} + \underbrace{g_k(t)}_{\text{temporal deviation}}, \quad \int_0^1 g_k(t) dt = 0.$$

Penalize the deviation g_k in full-rank L^2 -normalized coordinates.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



Research Gap II: Selection Is Not Inference

Why one penalized fit is insufficient

- Group shrinkage biases the curves.
- Zero and constant effects still need separation.
- Repeated visits create within-subject dependence.

Missing link

- Select the changing effects.
- Refit for accurate smooth curves.
- Classify zero vs. constant.
- Infer with cluster-robust SEs.

TV-Select uses separate estimators for selection, curve estimation, classification, and inference.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



Contributions: A Connected Pipeline

1. **Identify:** decompose each effect into mean + centered deviation.
2. **Normalize:** remove the spline null direction, so group norm has an L^2 meaning.
3. **Select and smooth:** combine group sparsity and a roughness penalty.
4. **Refit and infer:** obtain three-class recovery and oracle-equivalent cluster-robust inference under stated assumptions.

Decomposition lies at its core: breaking down complex entities into fundamental components.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Methodology

Model and Three-Class Identification

For subject $i = 1, \dots, N$ at visit $j = 1, \dots, n_i$:

$$y_{ij} = \beta_0 + \sum_{k=1}^p x_{ijk} \beta_k(t_{ij}) + \varepsilon_{ij}, \quad t_{ij} \in [0, 1].$$

Decompose $\beta_k(t) = \mu_k + g_k(t)$, with $\int_0^1 g_k(t) dt = 0$.

$$\begin{aligned} S_{\text{zero}} &= \{k : \mu_k = 0, g_k \equiv 0\}, \\ S_{\text{const}} &= \{k : \mu_k \neq 0, g_k \equiv 0\}, \\ S_{\text{vary}} &= \{k : g_k \not\equiv 0\}. \end{aligned}$$

Subjects are independent sampling units; repeated observations within a subject may be dependent.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Effective Spline Basis: Why Normalization Matters

Start from q B-splines $B(t)$ and center them:

$$\tilde{B}(t) = B(t) - \int_0^1 B(u) du.$$

The centered Gram matrix has a null direction. Keep only positive eigenvalues:

$$G_B = \int_0^1 \tilde{B}(t) \tilde{B}(t)^\top dt = U_e \Lambda_e U_e^\top, \quad C(t) = \Lambda_e^{-1/2} U_e^\top \tilde{B}(t).$$

Central identity

$$\int_0^1 \{C(t)^\top a\}^2 dt = \|a\|_2^2 \implies \|g_k\|_{L^2} = \|\theta_k\|_2 \quad (g_k = C^\top \theta_k).$$

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

The Penalized Selection Criterion

Let Z_k have rows $x_{ijk}C(t_{ij})^\top$, $\bar{X} = [\mathbf{1}, X]$ and $\gamma = (\beta_0, \mu^\top)^\top$.

$$\hat{\Theta} = \arg \min_{\gamma, \{\theta_k\}} \left\{ \frac{1}{2n} \left\| y - \bar{X}\gamma - \sum_{k=1}^p Z_k \theta_k \right\|_2^2 + \lambda_1 \sum_{k=1}^p \|\theta_k\|_2 + \lambda_2 \sum_{k=1}^p \theta_k^\top \Omega \theta_k \right\}.$$

λ_1 : group penalty

Entire varying blocks become zero.

λ_2 : roughness penalty

Selected curves stay smooth.

$$\hat{S}_{\text{vary}} = \{k : \|\hat{\theta}_k\|_2 > 0\}.$$

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Optimization: A Short Algorithmic Summary

The objective is convex. Use cyclic block proximal-gradient updates.

1. Given $\{\theta_k\}$, update unpenalized γ by least squares.
2. For each k , form its partial residual $r^{(k)}$.
3. Gradient step for the smooth part ℓ_k and block threshold:

$$v_k = \theta_k - \alpha_k \nabla \ell_k(\theta_k), \quad \theta_k \leftarrow \left(1 - \frac{\alpha_k \lambda_1}{\|v_k\|_2}\right)_+ v_k.$$

4. Choose $\alpha_k \leq L_k^{-1}$, where $L_k = \lambda_{\max}(Z_k^\top Z_k/n + 2\lambda_2 \Omega)$.

Theoretical convergence follows from descent, vanishing block increments, and KKT stationarity.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Four Estimators, Four Separate Roles

Stage	Penalty	Purpose
Selection	Group + roughness	Select varying blocks $\hat{S}_{\text{vary}}$
Smooth refit	Roughness only	Curves and prediction
Classification	None; all X	Separate zero vs. constant
Inference	None; selected model	Constant contrasts + cluster SEs

The four estimators $\hat{\Theta}$, F , R , and $PS = DB$ target different statistical tasks.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Theory and Proof Ideas

Theorem Map: Assumptions to Guarantees

Core assumptions

- Independent subjects; clustered score concentration.
- Spline smoothness $r > 2$.
- Full-rank $\bar{X}$, compatibility, stable inverse.
- Selection: irrepresentability and beta-min.

Main results

- **Thm 1:** estimation / function error.
- **Thm 2:** varying-set recovery.
- **Cor 1–2:** refit rate / full classification.
- **Thm 3:** cluster-robust oracle inference.

Tuning must satisfy suitable rates; the paper does not prove that EBIC automatically finds such rates.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Proof 1: Estimation via the Basic Inequality

Step 1: Profile out γ using $M_{\bar{X}} = I - P_{\bar{X}}$, and set $\tilde{Z} = M_{\bar{X}}Z$. **Step 2:** Compare $Q(\hat{\theta})$ with $Q(\theta_0^*)$, with $\Delta = \hat{\theta} - \theta_0^*$. Cluster-score concentration and controlled approximation / smoothing bias imply

$$\frac{\|\tilde{Z}\Delta\|_2^2}{2n} + \frac{\lambda_1}{2} \|\Delta_{S^c}\|_{2,1} \leq \frac{3\lambda_1}{2} \|\Delta_S\|_{2,1}.$$

Step 3: The group cone plus compatibility controls active errors.

$$\frac{\|\tilde{Z}\Delta\|_2^2}{n} = O_p(s_v \lambda_1^2), \quad \max_{k \in S} \|\hat{g}_k - g_{0k}\|_{L^2} = O_p(\lambda_1 + q_e^{-r}), \quad (s_v = O(1)).$$

Proof logic: stochastic bound $\rightarrow$ cone $\rightarrow$ restricted curvature $\rightarrow$ coefficient and function rates.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Proof 2: Exact Recovery (Primal–Dual Witness)

1. **Restricted solution:** solve on the true varying set S , setting $\tilde{\theta}_{S^c} = 0$.
2. **Avoid false negatives:** active inverse control gives

$$\left\| \tilde{\theta}_S - \theta_{0S}^* \right\|_{2,\infty} \lesssim C_G \lambda_1;$$

beta-min keeps all active blocks nonzero.

3. **Avoid false positives:** irrepresentability yields strict inactive KKT inequalities,

$$\max_{k \notin S} \|R_k\|_2 < \lambda_1.$$

4. The restricted candidate satisfies the *full* KKT system: $\Pr(\hat{S}_{\text{vary}} = S) \rightarrow 1$.

Two distinct conditions: beta-min controls missed signals; dual feasibility excludes spurious dynamics.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Proof 3: Refitting and Three-Class Recovery

Corollary 1: smooth refit

Given $\widehat{S} = S$, remove group shrinkage; retain roughness regularization:

$$\left\| \widehat{\alpha}^F - \alpha_0^* \right\|_2^2 = O_p \left(\frac{d_F}{n} + s_v q_e^{-2r} + s_v \delta_{F,n}^2 \right), \quad d_F = p + 1 + s_v q_e.$$

Corollary 2: zero vs. constant

Full-rank unpenalized classification refit gives

$$\left\| \widehat{\mu}^R - \mu_0 \right\|_\infty = O_p(r_{\mu,n}), \quad r_{\mu,n} = \sqrt{\log p/n} + q_e^{-r}.$$

Pick $r_{\mu,n} = o(\tau_N)$ and $\min_{k \in S_{\text{const}}} |\mu_{0k}| > 2\tau_N$.

Then true zeros lie below the threshold and true nonzero constants lie above it.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Proof 4: Cluster-Robust Oracle Inference

Let $X_c = X_{S_{\text{const}}}$, $W = [\mathbf{1}, X_S, Z_S]$ and $\tilde{X}_c = M_W X_c$.

$$\hat{\mu}_c^{DB} = \hat{\mu}_c^{PS} = A_{c,n}^{-1} \frac{\tilde{X}_c^\top M_W y}{n}, \quad A_{c,n} = \tilde{X}_c^\top \tilde{X}_c / n.$$

1. **FWL:** $\tilde{X}_c^\top W = 0$ cancels nuisance starting values exactly.
2. **Linear expansion:** estimation error = projected cluster score + approximation bias.
3. **Cluster CLT:** apply a CLT to independent subject scores $\tilde{X}_{c,i}^\top \varepsilon_i$.

$$\sqrt{n} a^\top (\hat{\mu}_c - \mu_{0c}) \Rightarrow N\left(0, a^\top \Sigma_c^{-1} \Gamma_c \Sigma_c^{-1} a\right).$$

Root- n inference needs stronger undersmoothing, e.g. $\sqrt{n} s_v q_e^{-r} \rightarrow 0$.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



Theory: What It Does Not Cover

- All p constant means are **unpenalized**; $\bar{X}$ must have full column rank.
- Effective size must obey $p + s_v q_e = o(n)$; for time-invariant baseline covariates, $p < N$ is needed.
- Exact selection uses **irrepresentability and beta-min**; weak effects are harder to recover.
- Cluster inference covers **fixed-dimensional constant contrasts**, not simultaneous curve bands.

This is not a proof of unrestricted $p \gg n$ selection or of valid EBIC tuning.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Empirical Evidence



Simulation Design: Stress the Structure

- $(N, p) = (100, 50)$ and $(200, 100)$; eight visits per subject.
- Six varying, six nonzero constant, and remaining zero effects.
- Cubic B-splines: $q = 8$ raw, $q_e = 7$ effective coordinates; 200 replications.
- Scenarios A–F vary covariate correlation, serial errors, heavy tails, visit-level covariates, and signal strength.
- Competitors: Group-Lasso, Marginal-VC, and VC-Ridge.

Key outcomes: exact varying-set recovery, false positives, curve error, and held-out prediction.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Simulation Results: Structural Precision

Exact recovery of S_{vary} , $N = 200$, $p = 100$ (paper Table 2)

Scenario	TV-Select	Group-Lasso	Marginal-VC
A	1.000	0.720	0.995
B	0.915	0.470	0.005
C	1.000	0.400	0.365
D	0.925	0.670	0.065
E	0.990	0.815	0.155
F (weak)	0.625	0.680	0.035

- A–E: TV-Select has stronger exact varying-set recovery.
- F: Group-Lasso does better on exact recovery, but TV-Select has lower FPR (0.0002 vs. 0.0027).

Fewer spurious varying effects; weak-signal detection is harder.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Sleep-EDF: Parsimony Without Prediction Loss

153 recordings from 78 subjects; nine physiological candidates for time variation.

Method	RMSE	Roughness	# varying
TV-Select	0.7617	43.65	7.12
VC-Ridge	0.7596	49.62	9.00
Group-Lasso	0.7605	90.73	8.52
Marginal-VC	0.7628	147.34	6.00

- Similar held-out RMSE; TV-Select has the lowest curve roughness.
- Full-data selected varying effects: alpha, sigma, beta, EOG line length, and respiratory line length.

Smoother, more interpretable curves; predictive accuracy remains similar.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Conclusion



Take-Home Message

1. **Research gap:** classical tools did not jointly address three-class identification, meaningful spline groups, refitting, and clustered post-selection inference.
2. **Method and theory:** centered normalized splines + group/roughness penalties lead to a clear proof chain:

basic inequality $\rightarrow$ support recovery $\rightarrow$ refit $\rightarrow$ cluster CLT.
3. **Evidence:** parsimonious temporal effects and smoother curves, with competitive prediction.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion



Remaining Research Opportunities

- Penalize μ too, to allow **ultra-high-dimensional** baseline covariates.
- Prove guarantees for the **data-driven EBIC** tuning procedure.
- Develop **covariance-aware** selection for more efficient longitudinal fitting.
- Extend inference from fixed constants to **simultaneous confidence bands** for varying curves.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Selected References I

-  Lu, Y., Zhang, T., Wang, Y. and Ran, M. (2026). Group-Sparse Smoothing for Longitudinal Models with Time-Varying Coefficients. arXiv:2603.07656v2.
-  Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *JRSS-B*.
-  Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*.
-  Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *JRSS-B*.
-  Wang, L., Li, H. and Huang, J. (2008). Variable selection in nonparametric varying-coefficient models. *JASA*.
-  Antoniadis, A. et al. (2012). Variable selection in varying-coefficient models using P-splines. *JCGS*.
-  Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*.

TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Thank You



TV-Select

Lu et al.

Motivation
and Research
Gap

Methodology

Theory and
Proof Ideas

Empirical
Evidence

Conclusion

Thank you very much!

Any questions?