Broken Adaptive Ridge in Time Series: the TS-BAR

Gianluca Cubadda
Tor Vergata University of Rome

Luca Margaritella
Lund University

Orest Prifti
Tor Vergata University of Rome

February 3, 2026

1 Extended Abstract

The Lasso (Tibshirani, 1996) solves a convex relaxation of the ideal but intractable ℓ_0 penalised least squares. This convexification buys computational tractability, but at the cost of a global, linear penalty that shrinks all coefficients equally, regardless of their magnitude or local structure. Ever since its introduction, several adaptive alternatives have also come up in the literature, notably the adaptive Lasso of Zou (2006) and folded-concave penalties such as SCAD (Fan and Li, 2001). These methods replace the plain ℓ_1 penalty by a weighted version where the weights are data-dependent and chosen to downweight large preliminary coefficients. This adaptivity reduces shrinkage bias on strong signals and, under suitable conditions, delivers the so-called oracle property: correct support recovery and asymptotically unbiased estimation on the true support.[1]

Broken Adaptive Ridge (BAR) (Dai et al., 2018) belongs to the same broad family of *adaptive* regularisation methods, but with a different geometry. Instead of a weighted ℓ_1 penalty, BAR uses an adaptive quadratic penalty where the weights are updated iteratively via a reweighted ridge scheme. Each step solves a strictly convex ridge problem, and the resulting estimator inherits a ridge-type grouping effect: highly correlated regressors tend to receive similar coefficients. In contrast to the weighted ℓ_1 penalty of the adaptive Lasso, BAR penalises coefficients quadratically but with heterogeneous curvature: small coefficients face very steep curvature and are pushed to zero, whereas large coefficients face almost no shrinkage. This makes BAR a closer surrogate to the ideal ℓ_0 penalty, which only cares about whether a coefficient is zero or not and leaves large signals essentially unshrunk.

In high-dimensional VARs, where strong collinearity across lags and across variables is the rule rather than the exception, these features are particularly appealing. Adaptive Lasso, while reducing bias relative to Lasso, still relies on an ℓ_1 geometry. BAR, by contrast, stabilises the autoregressive structure: dominant own-lags are kept essentially unbiased, weak cross-lags are aggressively shrunk to zero, and highly correlated lagged variables are treated in a more grouped fashion.

In this paper we contribute by: (i) extending BAR to high- and ultra-high-dimensional time series frameworks, specifically vector autoregressive (VAR) processes under physical dependence. Like the adaptive Lasso and folded-concave penalties, we show that BAR attains high-dimensional estimation and prediction

[1] This, without the stringent *irrepresentable condition* required by the plain Lasso, namely that assumption for which variables inside and outside the true support need not be (strongly) correlated.

rates under a restricted eigenvalue condition and a beta-min condition, without requiring an irrepresentable condition. (ii) We work under a row-wise weak sparsity assumption on the VAR transition matrices, allowing for many small but nonzero coefficients, and derive non-asymptotic bounds for the BAR prediction error and ℓ_1/ℓ_2 estimation errors that accommodate both temporal dependence and sub-exponential growing dimension. (iii) We establish variable selection consistency (support recovery) for BAR in this dependent, high-dimensional setting: inactive lags are excluded with probability tending to one, while active lags are retained under a beta-min separation condition of the same order as the empirical process fluctuations. This mirrors oracle-type properties known for adaptive Lasso and folded-concave penalties, but is obtained here for a ridge-based, iteratively reweighted procedure in a time-series context. (iv) Through a comprehensive Monte Carlo study based on high-dimensional VAR designs with controlled block sparsity and cross-sectional correlation, we document that BAR systematically improves on Lasso and adaptive Lasso in terms of support recovery, shrinkage bias on dominant autoregressive coefficients, and structural recovery of the transition matrices, while achieving comparable or better prediction performance. (v) Finally, the BAR superior feature selection and forecasting performance are further confirmed by two empirical applications in macroeconomics and finance.

References

Dai, L., Chen, K., Sun, Z., Liu, Z., and Li, G. (2018). Broken adaptive ridge regression and its asymptotic properties. *Journal of multivariate analysis*, 168:334–351.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.