Economics Bulletin

Volume 46, Issue 3

Time-varying betas for Bitcoin: A generalized cross-validation approach

Peter H. Wang
St. Mary's College of Maryland

Bluford Putnam
St. Mary's College of Maryland

Abstract

We estimate a time-varying market beta for Bitcoin using a data-driven, non-parametric kernel regression framework that models how market participants weight new information relative to older information through a time-decay process. The research highlights the boundary issues inherent in the standard Akaike Information Criterion (AIC) approach for bandwidth selection of time-series kernel smoothers and applies Generalized Cross-Validation (GCV) across four distinct decay families (Exponential, Gaussian, Rectangular, and Triangular). We document the optimal kernel bandwidth that minimizes the residual sum of squares subject to the trace penalty provided by GCV. Despite the differences in optimal bandwidths for each decay family, they converge to a common effective historical lookback, confirming the rationale of the approach. Empirically, the optimized estimates reveal profound shifts in Bitcoin's market risk over time, capturing its transition from a zero-beta speculative asset pre-2020 to a highly correlated, near-unity beta asset post-pandemic.

Citation: Peter H. Wang and Bluford Putnam, (2026) "Time-varying betas for Bitcoin: A generalized cross-validation approach", *Economics Bulletin*, Volume 46, Issue 3, pages 1066-1081
Contact: Peter H. Wang - hwang@smcm.edu, Bluford Putnam - bluputnam@gmail.com.
Submitted: April 17, 2026. **Published:** September 30, 2026.

1. Introduction

As Bitcoin matures from a fringe speculative asset into a staple of institutional portfolios, accurately quantifying its systematic market risk has become a critical objective for financial economists. The standard Capital Asset Pricing Model (CAPM), developed by Sharpe (1964) and Lintner (1965), relies on a static estimate of beta. However, assuming a constant covariance structure is fundamentally flawed for highly volatile, rapidly evolving digital assets. Bitcoin has undergone multiple structural shifts, driven by regulatory changes, macroeconomic tightening, and institutional adoption, necessitating a time-varying parameter (TVP) approach.

Traditional TVP models in empirical finance often rely on arbitrary rolling windows (e.g., a fixed 60-day or 252-day OLS regression). This approach is suboptimal: hard thresholds give equal weight to an observation from 59 days ago while discarding one from 61 days ago, creating arbitrary data cutoffs and an artificial discontinuity in the beta estimates. Parametric alternatives such as the Kalman filter and DCC-GARCH models impose structural restrictions on how beta evolves that may be especially ill-suited to an asset prone to sudden regime breaks.

To resolve this, we employ a non-parametric, weighted least squares regression to dynamically track Bitcoin's market risk. Rather than relying on arbitrary cutoffs or subjective researcher inputs, we implement a fully data-driven bandwidth selection process using Generalized Cross-Validation (GCV). This paper makes three contributions to the literature. First, it demonstrates analytically why standard AIC-based bandwidth selection breaks down when applied to time-series kernel smoothers, specifically, that the log-likelihood penalty cannot overcome the singularity that arises as the bandwidth approaches zero, and shows how GCV resolves this pathology. Second, it establishes an empirical regularity: despite producing substantially different optimal bandwidth values in normalized terms across four kernel families, all four kernels converge to an effective historical lookback of approximately 69 to 84 trading days, providing cross-kernel evidence that GCV is capturing a genuine data feature rather than a methodological artifact. Third, it provides a clean, algorithmically derived characterization of Bitcoin's evolving market beta from 2014 to 2026, documenting the transition from a near-zero pre-pandemic beta to a persistently positive post-pandemic beta close to unity.

2. Literature Review

The theoretical case for time-varying betas rests on the observation that investment opportunities and macroeconomic risk change over time, so that a rational asset's risk profile should evolve accordingly. Ferson and Harvey (1991, 1993) were among the first to formally model risk premiums as functions of conditioning information, and Jagannathan and Wang (1996) developed the conditional CAPM, allowing both betas and the market risk premium to vary with business-cycle instruments. Lettau and Ludvigson (2001) further showed that a scaled conditional CAPM explains the value premium that the static model cannot. However, Lewellen and Nagel (2006) argued that the degree of time variation implied by standard conditional CAPM formulations is far too small to rescue the model from well-known anomalies, providing a compelling motivation for more agnostic, model-free approaches.

Three broad empirical approaches dominate the TVP beta literature. Rolling-window OLS is prevalent in practice, but it weights all in-window observations equally and discards all out-of-

window information, producing artificial discontinuities. State-space models with the Kalman filter (Harvey, 1989) update estimates optimally under an assumed state-transition equation, but their performance degrades when the state-transition law is misspecified, a particular concern for assets subject to sudden regime breaks. Dynamic Conditional Correlation (DCC) GARCH models (Engle, 2002) allow pairwise correlations and, therefore, betas to vary dynamically but impose parametric restrictions on how correlations evolve, thereby incurring substantial estimation burdens. For a non-normal asset like Bitcoin, these restrictions may be especially consequential.

The theoretical foundations for non-parametric kernel regression are developed in Härdle (1990) and Fan and Gijbels (1996), who showed that locally weighted least squares estimators yield consistent estimates of unknown regression functions without imposing global parametric structure. Cai (2007) established the asymptotic properties of kernel-smoothed TVP models with serially correlated errors, a setting directly relevant to financial time series, thereby providing formal justification for the approach taken here.

The central challenge in kernel smoothing is bandwidth selection. The AIC (Akaike, 1974) and its finite-sample correction, AICc (Hurvich, Simonoff, and Tsai, 1998), are widely used but exhibit boundary pathologies in this setting as we discuss in Section 4. Craven and Wahba (1978) introduced GCV as a computationally tractable approximation to LOOCV for spline smoothing; Golub, Heath, and Wahba (1979) derived the foundational GCV formulation, which we adapt to the growing-window setting as described in Section 5; and Li (1987) established GCV's asymptotic optimality under mild regularity conditions.

Bitcoin's relationship to traditional financial markets has evolved substantially. Early studies, like Yermack (2015), Dyhrberg (2016), Baur, Hong, and Lee (2018), and Corbet et al. (2018), documented near-zero correlations with equity indices and characterized Bitcoin primarily as a speculative asset, suggesting it may offer diversification benefits for investors with short investment horizons, prior to 2018. Putnam and Wang (2025) showed that Bitcoin’s correlation with the S&P 500 strengthened materially from late 2017 to early 2018, coinciding with the introduction of CME Bitcoin futures, and then shifted persistently toward positive market-beta status after the pandemic. Liu and Tsyvinski (2021) and Liu, Tsyvinski, and Wu (2022) provided rigorous cross-sectional asset pricing analyses showing that cryptocurrency returns are largely unexplained by standard equity risk factors over early sample periods but exhibit increasingly factor-like behavior as institutional participation deepens. This body of evidence strongly suggests that Bitcoin's market beta is neither constant nor zero but follows a complex time path, making a data-driven, non-parametric approach to beta estimation especially valuable. A more detailed discussion of the full range of macroeconomic and asset-specific drivers of shifts in Bitcoin's beta is provided in Putnam and Wang (2025).

3. Time-Varying Parameter Estimation via Kernel Smoothing

To capture the evolving relationship between Bitcoin returns (y_t) and the broader equity market (proxied by the S&P 500, x_t), we estimate a time-varying beta using a weighted least squares (WLS) regression. For a given target day t, we minimize the locally weighted sum of squared residuals:

$$\min_{\alpha_t,\beta_t} \sum_{s=1}^{t} w_{t,s}(y_s - \alpha_t - \beta_t x_s)^2 \dots (1)$$

We explore four kernel decay functions: Exponential, Gaussian (Normal), Rectangular, and Triangular (Linear). To facilitate comparison across kernels, we normalize the time lag t−s into a common bandwidth parameter h ∈ [0, 1] by defining the normalized lag as $x_{t-s} = (t-s)/(252h)$. This parameterization is designed so that h → 1 approximates a long-horizon estimator and h → 0 places maximum weight on recent observations. Note that for the exponential and Gaussian kernels, h → 1 sets the half-life or inflection point at 252 days, approximating a form of long-memory decay, while for the rectangular and triangular kernels, h = 1 corresponds to an exact 252-day rolling window. The kernels are defined as:

$$\text{Exponential: } w_{t,s} = 0.5^{x_{t-s}} \quad \dots (2)$$

$$\text{Gaussian: } \quad w_{t,s} = e^{-\frac{x_{t-s}^2}{2}} \quad \dots (3)$$

For the rectangular and triangular kernels, weights are set to zero beyond the 252-day window ($x_{t-s} > 1$):

$$\text{Rectangular: } w_{t,s} = \begin{cases} 1, & x_{t-s} \le 1 \\ 0, & x_{t-s} > 1 \end{cases} \dots (4)$$

$$\text{Triangular: } w_{t,s} = \begin{cases} 1 - x_{t-s}, & x_{t-s} \le 1 \\ 0 \qquad\quad , & x_{t-s} > 1 \end{cases} \dots (5)$$

A maximum normalization of 252 trading days broadly aligns with common practice in empirical finance, and in unreported robustness checks, the GCV-optimal bandwidth lies well within the interior of the search grid for all four kernels, confirming that this maximum is not a binding constraint on the results.

4. Generalized Cross-Validation (GCV) Approach

The central challenge in kernel smoothing is choosing the bandwidth h that balances historical stability against rapid adaptation to structural breaks. Standard practice often relies on the Akaike Information Criterion (AIC). The foundational result of statistical learning theory (Efron, 2004) states that expected out-of-sample error equals in-sample error plus an optimism penalty (ω): $E[Err_{out}] = Err_{in} + \omega$. Efron (1986) has shown that the expected "optimism" is strictly equal to the sum of the covariances between the observed data points (y_t) and their corresponding predicted values ($\hat{y}_t$), scaled by sample size T. In other words, $\omega = \frac{2}{T}\sum_{t=1}^{T} Cov(y_t, \hat{y}_t)$. Because kernel weighting is a linear smoother, predictions $\hat{y} = Hy$, where the hat matrix is $H = X(X'WX)^{-1}X'W$.

It can be shown that this optimism simplifies to $(2\sigma^2/T)tr(H)$, and under the working assumption that the residuals have variance σ^2 (invoked here for illustration, not a maintained parametric assumption), we have[1]:

$$AIC = 2tr(H) + Tln(\frac{RSS}{T}) \dots \quad (6)$$

However, applying AIC to time-series kernel smoothers produces a boundary pathology: as h → 0, the effective degrees of freedom *tr(H)* approach the sample size T, the model perfectly

[1] See, for example, Burnham and Anderson (2002), or Hastie, Tibshirani and Friedman (2009) for a detailed proof of this expression.

interpolates recent observations, and RSS → 0. Because AIC relies on the natural logarithm of the variance, an RSS → 0 produces an unbounded negative log-likelihood that the linear trace penalty cannot overcome, causing AIC to spuriously select the minimum bandwidth regardless of the true data-generating process. This failure motivates a distribution-free alternative that does not rely on asymptotic likelihood approximations.

The gold standard for estimating expected prediction error is Leave-One-Out Cross-Validation (LOOCV), which computes:

$$LOOCV = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_{-t})^2 \ldots (7)$$

Using the Sherman-Morrison-Woodbury identity, this is equivalent to the in-sample residuals scaled by the hat-matrix leverage:

$$LOOCV = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t - \hat{y}_t}{1 - H_{tt}}\right)^2 \ldots (8)$$

Financial time series, however, are highly non-normal. Extreme market events exhibit near-unit leverage ($H_{tt} \approx 1$), which drives the denominator of equation (8) toward zero for those observations, causing a single outlier day to dominate bandwidth selection. Golub, Heath, and Wahba (1979) resolved this by replacing the individual, highly volatile H_{tt} values with the average leverage across all observations, the rotation-invariant quantity tr(H)/T, so that no single high-leverage observation can distort the criterion. Substituting and factoring yields the GCV objective:

$$GCV = \frac{RSS/T}{\left(1 - \frac{tr(H)}{T}\right)^2} \ldots (9)$$

Intuitively, as $h \to 0$ and the model overfits, $tr(H) \to T$ and the denominator collapses to zero, sending $GCV \to +\infty$. This mathematical barrier prevents over-discounting and forces the algorithm to select an interior optimal bandwidth purely from the data. Li (1987) established that GCV is asymptotically optimal in the sense of minimizing expected out-of-sample prediction error under mild regularity conditions, without requiring any distributional assumptions on the errors. In the standard smoother setting, H is a single fixed operator and H_{tt} denotes its t-th diagonal element. Section 5 describes how this criterion is adapted to the time-varying parameter setting, where a separate hat matrix $H^{(t)}$ is associated with each target day t.

5. Estimation Strategy

Since our estimation problem is to find a global optimal bandwidth for a "growing-window" of weighted least squares regressions, we need to adapt the GCV approach described in Section 4. We implement the following estimation strategy, which is also a computationally efficient algorithm that avoids explicit construction of full hat matrices. Explicitly, for each target day t, the scalar leverage $H_{tt}^{(t)}$ (for the t-th hat matrix $H^{(t)}$) is computed analytically from the weighted moment sums, which are accumulated as a running global trace. This reduces the computational complexity from $O(T^3)$ matrix operations to $O(T^2)$ per bandwidth value.

Specifically, we conduct a grid search of 100 values of h ∈ [0.01, 1] in increments of 0.01[2]. For each value of h and each day t, we compute the squared residual $RSS_t = (y_t - \hat{y}_t)^2$ and the diagonal leverage $H_{tt}^{(t)}$, defined as[3]:

$$H_{tt}^{(t)} = \frac{S_2 - (2 \times S_1 \times x_t) + (S_0 \times x_t^2)}{S_0 \times S_2 - S_1^2} \ldots (10)$$

where $S_0 = \sum_{j=1}^{T} w_{tj}$, $S_1 = \sum_{j=1}^{T} w_{tj} x_j$, $S_2 = \sum_{j=1}^{T} w_{tj} x_j^2$. Here $H_{tt}^{(t)}$ denotes the leverage of the terminal observation in the separate WLS regression run at the target day t, that is, the $[t, t]$ diagonal element of $H^{(t)} = X_t (X_t' W_t X_t)^{-1} X_t' W_t$, where $X_t$and W_t denote the design matrix and kernel weight matrix for the regression ending at day t. This generalizes the fixed-smoother $H_{tt}^{(t)}$ of equation (8) to the sequence of growing-window regressions. The RSS_t and $H_{tt}^{(t)}$ are then aggregated across all t to form the global GCV score[4] per the GCV equation:

$$GCV(h) = \frac{RSS/T}{\left(1 - \frac{1}{T}\sum_t H_{tt}^{(t)}\right)^2} \ldots (11)$$

The optimal h is selected as the grid value minimizing GCV across all 100 candidates. A 252-day burn-in period is employed so that the first beta estimate has sufficient observations to yield a meaningful residual and reliable error bands.

Obtaining the point estimate $\hat{\beta}_t$ at each target day t via weighted least squares, we next discuss how to construct uncertainty bands around it. A natural approach is to use the standard errors reported by a WLS regression, but this is inappropriate here. WLS standard errors are valid only when the weights are precision weights, that is, when $w_{t,s} \propto 1/\mathrm{var}(\varepsilon_s)$, so that observations with higher weight genuinely have lower error variance. Our kernel weights are temporal distance weights: they reflect how far the observation s lies in the past relative to the target day t, not the

---

[2] We start the algorithm at 0.01 to avoid the issue of division by zero for the normalization.

[3] See Appendix A for the derivation of $H_{tt}^{(t)}$ from the WLS hat matrix.

[4] This criterion differs from the textbook Golub–Heath–Wahba (1979) GCV in that the hat matrix is not a single fixed operator but a sequence of time-indexed WLS hat matrices $\{H^{(t)}\}$, one per target day, each accumulating observation from 1 to t. The appropriate bandwidth selection objective is the average of T individual leave-one-out prediction errors, one for each target day. For each individual regression at time t, the kernel weights $w_{t,s}$ are deterministic functions of the time lag and do not depend on y_t, so the Sherman–Morrison–Woodbury identity holds exactly: $y_t - \hat{y}_{-t}^{(t)} = (y_t - \hat{y}_t)/(1 - H_{tt}^{(t)})$. The resulting criterion is therefore a LOOCV-based objective, with $H_{tt}^{(t)}$ computed in closed form via equation (10) for each t. The single approximation enters when the individual leverage values are replaced by their cross-time average $\bar{H} = \frac{1}{T}\sum_t H_{tt}^{(t)}$: this is the GCV step, justified by the observation that $H_{tt}^{(t)}$ stabilizes to a smooth function of h alone once t is large relative to the effective bandwidth, the analog of Li's (1987) nil-trace condition for a sequence of regressions. Li (1987) establishes that GCV and LOOCV are asymptotically equivalent under the nil-trace condition for fixed linear smoothers with i.i.d. errors; Xia and Li (2002) establish asymptotic optimality of LOOCV for local polynomial estimators under α-mixing dependence, the setting directly relevant to financial time series. Andrews (1991) shows that with heteroskedastic errors, LOOCV retains asymptotic optimality under general conditions while GCV requires an additional eigenvalue-spread condition; since the present criterion computes $H_{tt}^{(t)}$ exactly and only approximates at the averaging step, it is closer in spirit to LOOCV than to classical GCV and inherits the stronger optimality guarantees of the former. The asymptotic theory for the underlying local linear TVP estimator with serially correlated errors is established in Cai (2007). A formal proof of the asymptotic properties of this adapted GCV criterion is beyond the scope of this paper, but is left for future research.

variance of the return error at s. Under the correct model, in which the errors ε_s are homoscedastic with $Var(\varepsilon_s) = \sigma^2$, the true variance of $\hat{\beta}_t$ from the kernel WLS estimator is not $\sigma^2(X'WX)^{-1}$ but the sandwich form:

$$Var(\hat{\beta}_t) = (X'WX)^{-1}(X'W^2X)(X'WX)^{-1} \dots (12)$$

where W^2 denotes the diagonal matrix of squared kernel weights[5]. Because W^2 concentrates mass even more heavily on the most recent observations than W does, equation (12) produces larger, and correct standard errors relative to the misspecified WLS formula, with the discrepancy largest in the early sample when only a small number of observations carry meaningful weight.

To compute equation (12) without constructing any matrices explicitly, we define three additional weighted moment sums analogous to S_0, S_1, and S_2 but built from the squared weights $w_{t,j}^2$: $S_0^w = \sum_{j=1}^{t} w_{t,j}^2$, $S_1^w = \sum_{j=1}^{t} w_{t,j}^2 x_j$, $S_2^w = \sum_{j=1}^{t} w_{t,j}^2 x_j^2$. Using the same 2×2 analytical inverse as in the derivation of $H_{tt}^{(t)}$ (see Appendix B), the (2,2) element of the sandwich, which corresponds to the slope $\hat{\beta}_t$, reduces to the closed-form scalar:

$$Var(\hat{\beta}_t) = \hat{\sigma}_t{}^2 \times \frac{(S_1^2 S_0^w - 2S_0 S_1 S_1^w + S_0^2 S_2^w)}{(S_0 S_2 - S_1^2)^2} \dots (13)$$

where $\hat{\sigma}_t{}^2 = (\frac{1}{t-2}) \sum_{j=1}^{t} (y_j - \hat{\alpha}_t - \hat{\beta}_t x_j)^2$ is the unweighted residual variance estimated with t − 2 degrees of freedom, consistent with the homoscedastic error assumption. The standard error of $\hat{\beta}_t$ is $SE(\hat{\beta}_t) = \sqrt{Var(\hat{\beta}_t)}$ and we report the 95% confidence bands constructed as $\hat{\beta}_t \pm 1.96 \times SE(\hat{\beta}_t)$.

6. Data

We gather daily adjusted closing prices for the S&P 500 (via the SPY ETF) and Bitcoin spot prices (BTC-USD) from Yahoo Finance, covering the period from September 18, 2014, to April 14, 2026. We use daily simple percentage changes as the unit of observation[6]. Since Bitcoin trades on weekends while the SPY ETF does not, we address this timing mismatch by retaining only the trading days when U.S. equity markets are open, excluding Saturdays, Sundays, and U.S. holidays.

7. Results

Figures 1 through 4 plot the time-varying beta coefficients for each kernel, together with 95% confidence interval bands constructed using the sandwich standard errors described in Section 5. Two broad regimes are visually apparent across all four kernel choices: a pre-pandemic period in which the beta is close to zero (save for a sharp spike around early 2018 associated with the CME

[5] Note that, for the rectangular kernel, equation (12) simplifies to the standard estimator $Var(\hat{\beta}_t) = \sigma^2(X'WX)^{-1}$.

[6] Note that the choice of simple percentage changes over log returns is standard practice. Although since Bitcoin is prone to fat-tailed daily moves, simple and log returns can diverge substantially for specific intervals, we leave robustness checks using log returns for future research.

Bitcoin futures launch), and a post-pandemic period in which the beta stabilizes above zero and frequently approaches a value of one. Formal tests for regime switches are left for future research.

Note that the simple linear regression beta coefficient of the entire period is 0.8034 (with a standard error of 0.0618). If we break the sample into pre- and post-pandemic periods, the coefficient before the pandemic is 0.1 (with a standard error of 0.1277), and after the pandemic, it is 1.0815 (with a standard error of 0.0646). This illustrates the substantial changes in Bitcoin's market beta since 2014 and further motivates our data-driven approach to time-varying beta estimation to capture Bitcoin's current market risk.

Figure 1: TVP of Bitcoin, Exponential Kernel

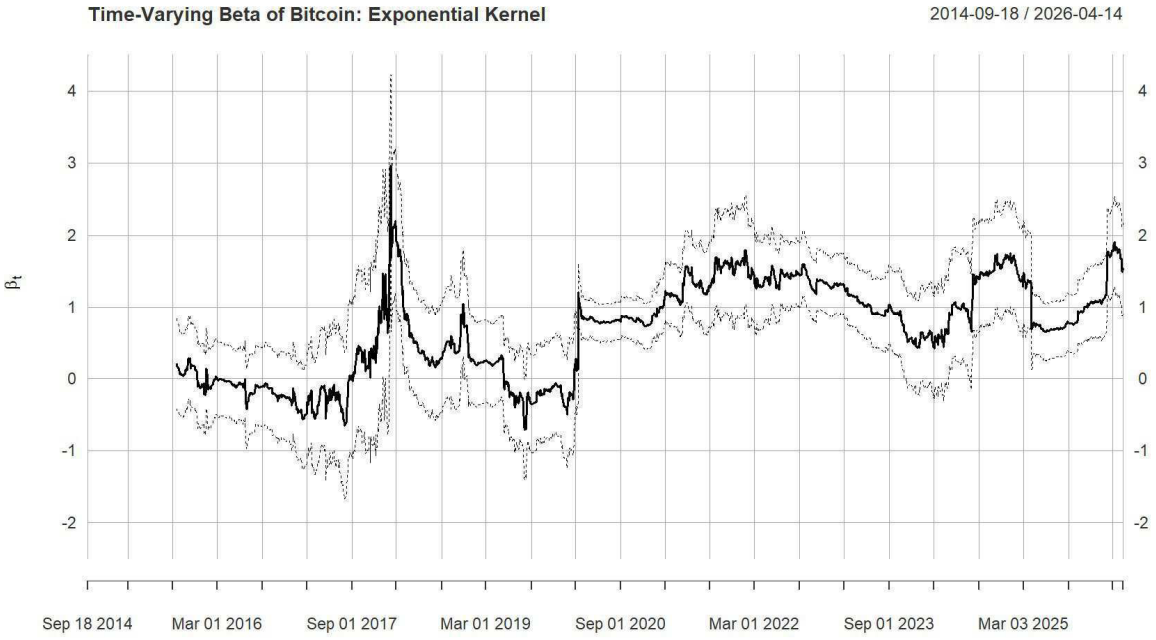

Figure 2: TVP of Bitcoin, Gaussian (Normal) Kernel

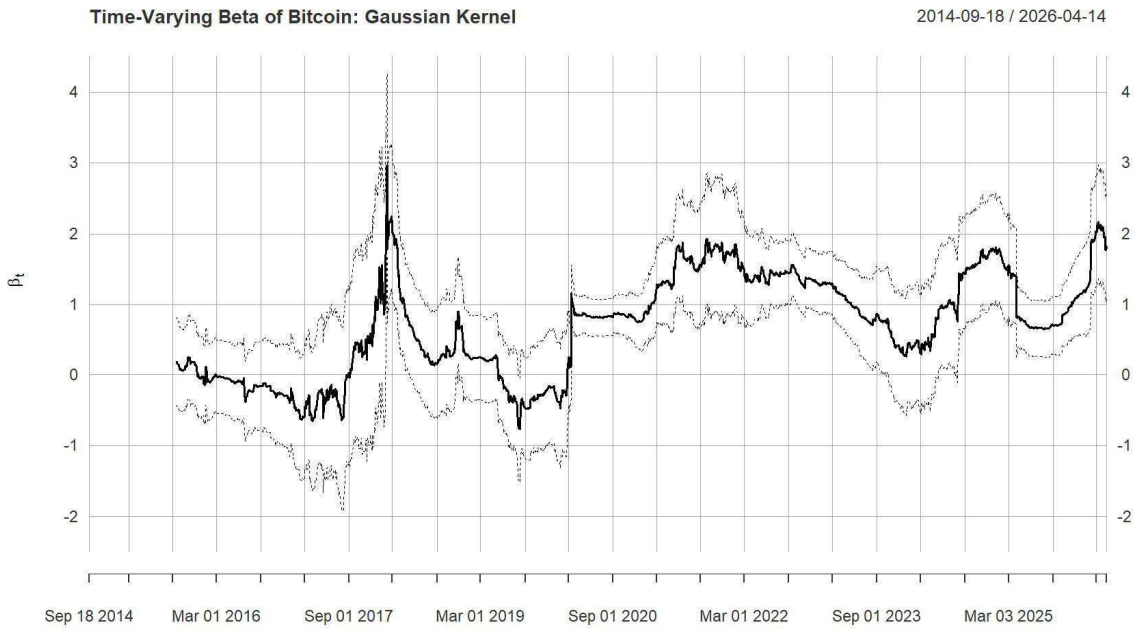

Figure 3: TVP of Bitcoin, Rectangular Kernel

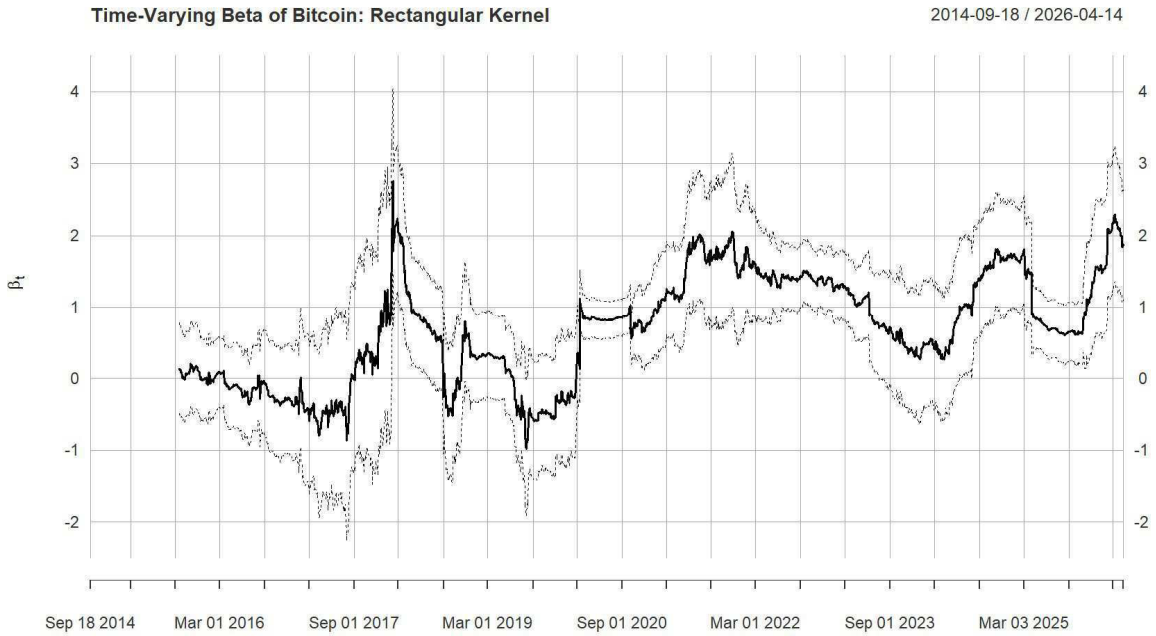

Figure 4: TVP of Bitcoin, Triangular (Linear) Kernel

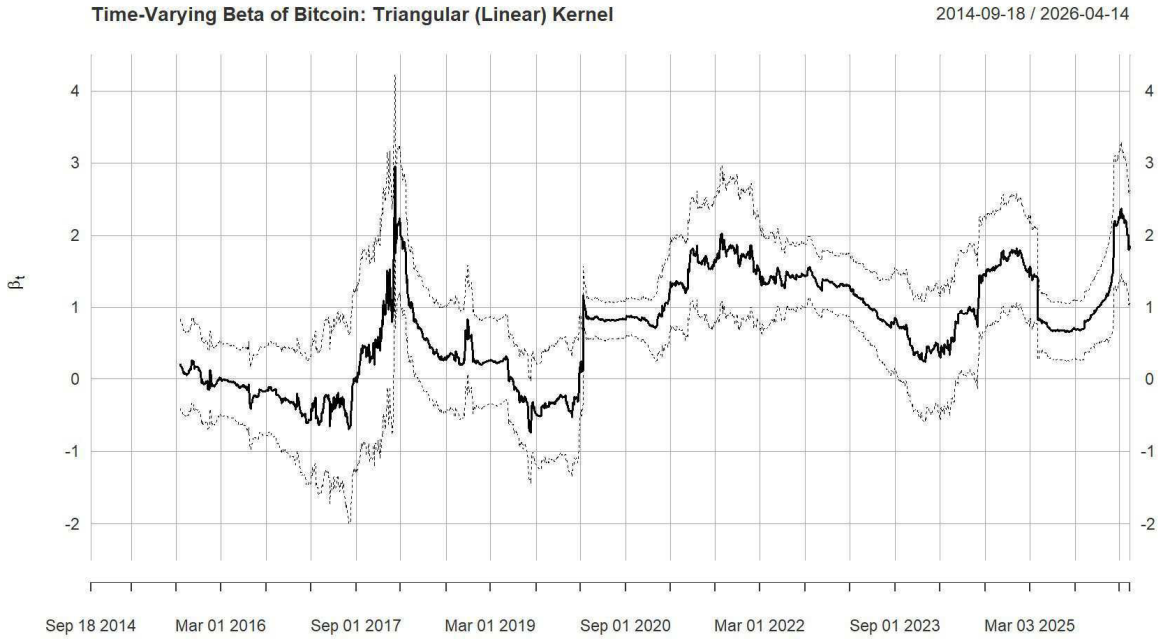

Table 1 reports the GCV-optimal bandwidth h and approximate lookback for each kernel.

Table 1. GCV-Optimal Bandwidth by Kernel

Kernel	Optimal h	Approx. Days[1]
Exponential	0.23	58
Gaussian	0.34	86
Rectangular	0.57	144
Triangular	0.82	207

[1] Interpretation differs by kernel: for exponential and Gaussian kernels, h parameterizes the half-life and inflection point, respectively, so weights remain positive (though diminishing) beyond 252 days; for rectangular and triangular kernels, the window is exact, with weights equal to zero beyond the reported number of days.

Despite the wide range of optimal h values across kernels (0.23 to 0.82), the implied historical centers of mass are notably similar. The rectangular kernel places its center of mass at 144/2 = 72 days; the triangular at 207/3 = 69 days; the Gaussian at $86 \times \sqrt{(2/\pi)} \approx 69$ days; and the exponential at $58/\ln(2) \approx 84$ days. All four converge to some "historical depth", or effective historical lookback of roughly 69–84 trading days, providing cross-kernel corroboration that GCV is recovering a genuine property of the data-generating process rather than an artifact of the chosen kernel form.

One technical concern is worth addressing regarding the implementation of the GCV-optimal kernel bandwidth selection. In our growing window estimations (from 252 days to about 2900 days), intuitively, the leverage $H_{tt}^{(t)}$ calculations would be larger for early regressions and could therefore dominate the average. In Table 2, we provide $H_{tt}^{(t)}$ values for each kernel at the first day,

the 250th day, and the terminal day, as well as the mean for the full sample and the mean excluding the first 250 $H_{tt}^{(t)}$ observations.

Table 2. Leverage Value Robustness Check

Kernel	First $H_{tt}^{(t)}$	250th-day $H_{tt}^{(t)}$	Last $H_{tt}^{(t)}$	Average	Late Average[1]
Exponential	0.0375	0.0120	0.0233	0.0231	0.0233
Gaussian	0.0305	0.0082	0.0153	0.0159	0.0158
Rectangular	0.0244	0.0053	0.0088	0.0097	0.0096
Triangular	0.0217	0.0041	0.0060	0.0071	0.0069

[1] Late average refers to the average of $H_{tt}^{(t)}$ excluding the first 250 days of the leverage observations.

It is true that the first $H_{tt}^{(t)}$ observations are larger than the last across all kernels. However, as an illustrative example, for the exponential kernel, with an optimal bandwidth of 0.23 and an approximate half-life of 58 days, 3 to 4 half-lives should stabilize the $H_{tt}^{(t)}$ since the effective sample carrying meaningful weight at day t should also become stable. From column 3, this is shown to be the case across all kernels after about 250 days. Additionally, if we compare the overall average (column 4) versus the "late average" (column 5), that is, the average excluding the first 250 $H_{tt}^{(t)}$ observations, the values are nearly identical, reducing concern that early observations disproportionately influence the bandwidth selection.

8. Discussion and Conclusions

Our research highlights the critical importance of studying how market participants process new information relative to older information. The GCV framework balances two opposing forces: a bandwidth that is too wide ($h \to 1$) smooths out the time-varying parameter entirely, producing a large RSS but near-constant beta; a bandwidth that is too narrow ($h \to 0$) tracks idiosyncratic daily noise, reducing RSS to near-zero but causing the model to overfit. The GCV score, by inflating the apparent fit loss when leverage is high, so that narrow bandwidths that nearly interpolate the data are penalized through a collapsing denominator rather than an additive term, locates an interior optimum that balances fit against excessive parameter flexibility.

The time-varying beta series reveals several distinct and economically interpretable episodes. The first is the sharp beta spike in early 2018, coinciding with the December 2017 launch of Bitcoin futures on the CME, which institutional participants used to short Bitcoin, temporarily increasing its co-movement with broader risk assets. The most consequential structural break occurred in March 2020, when the COVID-19 pandemic produced a synchronized global asset sell-off. Critically, Bitcoin's beta remained persistently elevated post-pandemic, rarely returning to near-zero for extended periods, consistent with the growing presence of institutional investors who treat Bitcoin as a risk asset within diversified portfolios. Bitcoin's beta rose sharply again following a weak U.S. job report in August 2024, reflecting its increased sensitivity to U.S. macroeconomic risk. More recently, Bitcoin's beta experienced a distinct downward shift in early 2025 following the "Liberation Day" tariff announcements, then a sharp upward spike, approaching a coefficient near 2, following the 2026 US-Iran conflict and the associated surge in global oil prices.

There are various macroeconomic and asset-specific reasons for these regime shifts, several of which are discussed in Putnam and Wang (2025); a full treatment is beyond the scope of the present paper. We also note that a natural next step is to conduct formal empirical tests for structural breaks in the time-varying parameter series, for example, using Bai and Perron's (1998) test applied to the estimated beta trajectory; we leave this for future research.

From a methodological standpoint, this paper demonstrates that standard AIC-based bandwidth selection is unsuitable for time-series kernel smoothers due to a log-likelihood singularity at narrow bandwidths, and that GCV, by virtue of its distribution-free construction and the asymptotic negligibility of individual leverage deviations around their cross-time average, resolves this pathology and delivers reliable interior optima even for highly non-normal financial series. The convergence of GCV-selected centers of mass across four structurally distinct kernel families provides an internal validation of the approach. Together, these results offer practitioners a methodologically defensible, computationally efficient tool for dynamic risk monitoring of cryptocurrency portfolios.

Finally, this paper adapts the GCV bandwidth selection criterion to a "growing-window" time-varying parameter regression in which a separate WLS regression is run at each target date t, accumulating all observations from the start of the sample through t. The standard GCV formulation assumes a single fixed hat matrix; here, we instead average the terminal leverage values $H_{tt}^{(t)}$ across all target dates to form a global penalty term. Because each $H_{tt}^{(t)}$ is computed exactly via the Sherman–Morrison–Woodbury identity, rather than approximated, this criterion is closer in spirit to leave-one-out cross-validation than to classical GCV, providing an intuitive basis for its asymptotic validity. A formal proof of asymptotic optimality for this criterion in the growing-window setting, however, remains an open question and is left for future research.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

Andrews, D. W. K. (1991). Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors. Journal of Econometrics, 47(2–3), 359–377.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica, 66(1), 47–78.

Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions and Money, 54, 177–189.

Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.). Springer-Verlag.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics, 136(1), 163–188.

Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. Economics Letters, 165, 28–34.

Craven, P., & Wahba, G. (1978). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik, 31(4), 377–403.

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar – a GARCH volatility analysis. Finance Research Letters, 16, 85–92.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? Journal of the American Statistical Association, 81(394), 461–470.

Efron, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. Journal of the American Statistical Association, 99(467), 619–632.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate GARCH models. Journal of Business & Economic Statistics, 20(3), 339–350.

Fan, J., & Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman & Hall.

Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. Journal of Political Economy, 99(2), 385–415.

Ferson, W. E., & Harvey, C. R. (1993). The risk and predictability of international equity returns. Review of Financial Studies, 6(3), 527–566.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21(2), 215–223.

Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press.

Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

Hurvich, C. M., Simonoff, J. S., & Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society: Series B, 60(2), 271–293.

Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. Journal of Finance, 51(1), 3–53.

Lettau, M., & Ludvigson, S. (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy, 109(6), 1238–1287.

Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics, 82(2), 289–314.

Li, K. C. (1987). Asymptotic optimality for C_n, CL, cross-validation and generalized cross-validation: discrete index set. Annals of Statistics, 15(3), 958–975.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, 47(1), 13–37.

Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. Review of Financial Studies, 34(6), 2689–2727.

Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common risk factors in cryptocurrency. Journal of Finance, 77(2), 1133–1177.

Putnam, B., & Wang, P. H. (2025). How the Pandemic Triggered a Sea Change in the Way Bitcoin Traded. Commodity Insights Digest. 3(2), 74-81.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, 19(3), 425–442.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.

Xia, Y., & Li, W.K. (2002). Asymptotic behavior of bandwidth selected by the cross-validation method for local polynomial fitting. *Journal of Multivariate Analysis*, 83(2), 265–287.

Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In D. Lee Kuo Chuen (Ed.), Handbook of Digital Currency (pp. 31–43). Elsevier.

Appendix A: Derivation of the Diagonal Hat-Matrix Element $H_{tt}^{(t)}$

For the WLS regression with weight matrix W, the hat matrix is $H^{(t)} = X(X'WX)^{-1}X'W$. Defining the weighted moment sums $S_0 = \sum_{j=1}^{T} w_{tj}$, $S_1 = \sum_{j=1}^{T} w_{tj}x_j$, $S_2 = \sum_{j=1}^{T} w_{tj}x_j^2$, the 2×2 matrix X'WX becomes:

$$X'WX = \begin{bmatrix} S_0 & S_1 \\ S_1 & S_2 \end{bmatrix}$$

Inverting this matrix, we get:

$$(X'WX)^{-1} = \frac{1}{S_0S_2 - S_1^2}\begin{bmatrix} S_2 & -S_1 \\ -S_1 & S_0 \end{bmatrix}$$

and multiplying to extract the diagonal element for day t yields:

$$H_{tt}^{(t)} = \frac{S_2 - (2 \times S_1 \times x_t) + (S_0 \times x_t^2)}{S_0 \times S_2 - S_1^2}$$

Note that this formula requires the weight of day t on itself (w_{tt}) to equal one. By construction, the normalized lag at zero distance is $x_{t-t} = 0/(252h) = 0$ for all $h > 0$. Substituting into each kernel: the exponential gives $0.5^0 = 1$; the Gaussian gives $\exp(0) = 1$; the rectangular gives 1 by design; and the triangular gives 1–0 = 1. Thus, $w_{tt} = 1$ holds exactly for all four kernels, confirming that equation (10) is fully general across all kernel families studied in this paper.

Appendix B: Derivation of the Sandwich Variance Estimator

In this appendix we will show the derivation of the variance of the $\hat{\beta}_t$ using the kernel weighted least squares regression. First, the true model at target day t is:

$$y_s = \alpha_t + \beta_t x_s + \varepsilon_s$$

With $s = 1, \dots, t$. Since $\mathrm{Var}(\varepsilon_s) = \sigma^2$ for all s (i.e., assuming homoscedastic errors). In matrix form we can write it as: $\mathrm{y} = \mathrm{X}\theta_t + \varepsilon$, hence $\mathrm{Var}(\varepsilon) = \sigma^2\mathrm{I}$.

The WLS estimator with kernel weight matrix $\mathrm{W} = \mathrm{diag}(w_{t,1}, \dots, w_{t,t})$ is:

$$\hat{\theta}_t = (\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{W}\mathrm{y}$$

Substitute $\mathrm{y} = \mathrm{X}\theta_t + \varepsilon$ in to the equation above yields:

$$\hat{\theta}_t - \theta_t = (\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{W}\varepsilon$$

Taking the variance:

$$\mathrm{Var}(\hat{\theta}_t) = (\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}\mathrm{X}'\mathrm{W} \cdot \mathrm{Var}(\varepsilon) \cdot \mathrm{W}\mathrm{X}(\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}$$

Lastly, substitute $\mathrm{Var}(\varepsilon) = \sigma^2\mathrm{I}$:

$$\mathrm{Var}(\hat{\theta}_t) = \sigma^2(\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}(\mathrm{X}'\mathrm{W}^2\mathrm{X})(\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}$$

Which is equation (12). Note that the W^2 term arises because $\mathrm{W} \cdot \sigma^2\mathrm{I} \cdot \mathrm{W} = \sigma^2\mathrm{W}^2$, not $\sigma^2\mathrm{W}^{-1}$ as standard WLS assumes. The standard WLS formula $\sigma^2(\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}$ would only be correct if $\mathrm{Var}(\varepsilon) = \sigma^2\mathrm{W}^{-1}$, meaning observations with higher kernel weight genuinely have lower error variance, which is exactly the precision-weight assumption that fails with the time weighted kernel regression, since kernel weights $w_{t,s}$ parameterize temporal distance, not inverse error variance. This is a specific, well-studied instance of the general sandwich principle established in White (1980). Additionally, we should emphasize that this sandwich form is a specific non-stochastic variant, in which the discrepancy stems purely from temporal filtering and is independent of the error variance status.

Now, we need to extract the [2,2] element — corresponding to $\hat{\beta}_t$— from the full sandwich matrix $\sigma^2(\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}(\mathrm{X}'\mathrm{W}^2\mathrm{X})(\mathrm{X}'\mathrm{W}\mathrm{X})^{-1}$. Using the weighted moment sums already defined in the paper:

$$\mathrm{X}'\mathrm{W}\mathrm{X} = \begin{pmatrix} S_0 & S_1 \\ S_1 & S_2 \end{pmatrix}, \mathrm{X}'\mathrm{W}^2\mathrm{X} = \begin{pmatrix} S_0^w & S_1^w \\ S_1^w & S_2^w \end{pmatrix}$$

The inverse of $\mathrm{X}'\mathrm{W}\mathrm{X}$ is:

$$(\mathrm{X}'\mathrm{WX})^{-1} = \frac{1}{S_0 S_2 - S_1^2}\begin{pmatrix} S_2 & -S_1 \\ -S_1 & S_0 \end{pmatrix}$$

For any conformable matrices A and B, the [2,2] element of ABA equals $\mathrm{r}_2'\mathrm{B}\mathrm{r}_2$, where r_2 is the second row of A written as a column vector. From the inverse above:

$$\mathrm{r}_2 = \frac{1}{S_0 S_2 - S_1^2}\begin{pmatrix} -S_1 \\ S_0 \end{pmatrix}$$

Next, we compute $\mathrm{X}'\mathrm{W}^2\mathrm{X} \cdot \mathrm{r}_2$, which gives us:

$$\begin{pmatrix} S_0^w & S_1^w \\ S_1^w & S_2^w \end{pmatrix}\frac{1}{S_0 S_2 - S_1^2}\begin{pmatrix} -S_1 \\ S_0 \end{pmatrix} = \frac{1}{S_0 S_2 - S_1^2}\begin{pmatrix} -S_0^w S_1 + S_1^w S_0 \\ -S_1^w S_1 + S_2^w S_0 \end{pmatrix}$$

Pre-multiply by r_2', we get:

$$\mathrm{r}_2' \cdot (\mathrm{X}'\mathrm{W}^2\mathrm{X} \cdot \mathrm{r}_2) = \frac{1}{(S_0 S_2 - S_1^2)^2}(-S_1, S_0)\begin{pmatrix} -S_0^w S_1 + S_1^w S_0 \\ -S_1^w S_1 + S_2^w S_0 \end{pmatrix}$$

Expanding term by term:

$$\begin{aligned} &= \frac{1}{(S_0 S_2 - S_1^2)^2}[(-S_1)(-S_0^w S_1 + S_1^w S_0) + S_0(-S_1^w S_1 + S_2^w S_0)] \\ &= \frac{S_1^2 S_0^w - S_0 S_1 S_1^w - S_0 S_1 S_1^w + S_0^2 S_2^w}{(S_0 S_2 - S_1^2)^2} \\ &= \frac{S_1^2 S_0^w - 2S_0 S_1 S_1^w + S_0^2 S_2^w}{(S_0 S_2 - S_1^2)^2} \end{aligned}$$

Multiply through by σ^2:

$$\mathrm{Var}(\hat{\beta}_t) = \sigma^2 \cdot \frac{S_1^2 S_0^w - 2S_0 S_1 S_1^w + S_0^2 S_2^w}{(S_0 S_2 - S_1^2)^2}$$

Lastly, replacing σ^2 with its estimator $\hat{\sigma}_t^2 = \frac{1}{t-2}\sum_{s=1}^{t}(y_s - \hat{\alpha}_t - \hat{\beta}_t x_s)^2$ gives equation (13).