

Mean and Variance Regimes in Threshold VARs: Revisiting Regime-dependent Fiscal Multipliers

Valeria Gargiulo*

Universitat Pompeu Fabra

This version: October 13, 2026

[Click here for the latest version](#)

Abstract

This paper develops the estimation theory for threshold VARs in which regimes govern not only the conditional mean but also the conditional variance, with regime boundaries determined endogenously by the data. We introduce the decoupled threshold VAR, in which mean and variance regimes are driven by different observable state variables and separate thresholds, and show that the joint-switching model is its constrained special case. Existing threshold estimators identify regimes from the conditional mean alone; we show that this is inconsistent when the regime change is in the variance – a common feature of macroeconomic data – and inefficient when it is in both, and provide a likelihood-based estimator that uses both moments. We establish consistency, derive the nonstandard limiting distribution of the thresholds, and show that their super-consistent rate leaves inference on the regime parameters unaffected. We apply the framework to U.S. government spending multipliers, where wars and major fiscal events constitute a separate high-variance regime. Once these episodes are isolated, significant state dependence emerges even at the thresholds imposed in earlier work, and post-WWII multipliers are larger in expansions than in recessions, reconciling apparently conflicting findings in the literature.

*Email: valeria.gargiulo@upf.edu

1 Introduction

Nonlinearities in macroeconomic data are estimated by letting parameters differ across periods or regimes, and a central lesson of the last twenty years of this literature is that the parameters that change most are the second moments. A large body of evidence documents this. The Great Moderation was first of all a fall in the volatility of shocks; drifting-parameter models attribute most of the time variation in U.S. monetary dynamics to stochastic volatility rather than to the coefficients; and among regime-switching models of monetary policy, the best fit is obtained when only the variances switch. Thus, volatility is not only a nuisance parameter to be corrected for. It is an economic object with real effects of its own, worth measuring and dating.¹

Two things follow for estimation. Regimes in volatility deserve the same treatment as regimes in dynamics; and a model that lets the dynamics switch while ignoring the volatility will misattribute the one to the other, locating its regimes where the shocks are large rather than where the propagation changes. The available tools handle this differently. Time-varying-parameter models with stochastic volatility and Markov-switching models accommodate changes in the second moments, but treat the regime change as latent and, in their classic formulations, as exogenous in its timing: the model infers when the regime changed and the economist labels the result afterwards. Threshold models provide a parsimonious and interpretable alternative. They tie the regime change to an observable state crossing an estimated cutoff: the regimes carry an economic label from the start, the model is linear within each regime, and the only nonstandard object to estimate is the threshold itself. This has made them especially appealing for state-dependent macroeconomics, exploited for monetary policy by [Gargiulo et al. \(2026\)](#) and [Canova and Forero \(2024\)](#) among others. Their estimation theory, however, is silent on the second moments. This paper proposes a threshold VAR in which the variance has regimes of its own, dated like those in the dynamics by an observable state crossing an estimated threshold, and develops its estimation theory. The fiscal multiplier literature is the natural place to see why this matters. Whether government spending is more effective in bad times is a question addressed with regime-dependent models, and with few exceptions the regimes are imposed: the researcher chooses the state variable and the cutoff that separates expansions from recessions, as in [Auerbach and Gorod-](#)

¹[Cogley and Sargent \(2005\)](#) and [Primiceri \(2005\)](#) on drifting parameters with stochastic volatility; [Sims and Zha \(2006\)](#) on regime switching. On volatility as an economic object with real effects of its own, [Bloom \(2009\)](#) on uncertainty shocks and, closest to our application, [Fernández-Villaverde et al. \(2015\)](#) on fiscal volatility shocks.

nichenko (2012) and Ramey and Zubairy (2018). The identifying variation in this literature, moreover, often comes with wars and military build-ups, which are both episodes of unusual shock volatility and of unusual macroeconomic states – near-full employment in the mobilizations, deep slack in the Depression years that precede WWII. A regime structure that sees only the conditional mean has no way of telling the two apart, and will date its regimes by whichever is louder in the data. In this literature, that is the war. Fiscal policy is therefore the scenario in which volatility regimes are both most plausible and most dangerous to ignore.

This paper develops the estimation theory for threshold VARs in which regimes govern not only the conditional mean but also the conditional variance, with regime boundaries determined by the data. It extends threshold models along two dimensions: from univariate autoregressions to VARs, and from the first to the second moment. We introduce the *decoupled* threshold VAR, in which mean and variance regimes are driven by different observable state variables and separate thresholds, so that *propagation regimes* – how shocks transmit – and *shock regimes* – the environment in which the identifying shocks arrive – can be dated separately. The framework thus endogenizes not only the timing of a regime switch but also its source: whether what changes is how shocks propagate or how volatile they are. We then show that the simpler *joint* model, in which one state and one threshold govern both moments, is its constrained special case. We establish consistency, derive the non-standard limiting distributions of the thresholds, and show that their super-consistent rate leaves inference on the regime parameters unaffected. We then apply the framework to U.S. government spending multipliers in the data and framework of Ramey and Zubairy (2018), henceforth RZ2018, where wars and major fiscal events constitute a separate high-variance regime. Once these episodes are isolated, state dependence emerges even at the thresholds imposed in earlier work, and post-WWII multipliers are larger in the low-unemployment regime than in the high-unemployment regime, which reconciles apparently conflicting findings in the literature.

Our first contribution is econometric, and concerns the estimation of the threshold VAR (TVAR). Threshold models are normally estimated in two steps: first the threshold is estimated by selecting the value that best fits the data according to a criterion function; then, conditional on it, the rest of the parameters are obtained by estimating one linear model per regime. Under regularity conditions, the threshold estimator is super-consistent at the fast rate T , implying that its estimation uncertainty does not enter the asymptotic distribution of the remaining model parameters, thus justifying the two-step approach. The available

theory, however, has two limitations. First, it covers univariate autoregressions only (Tong and Lim, 1980; Chan, 1993; Hansen, 2000; Samia and Chan, 2011). Second, whatever the criterion in the first step, the threshold is identified through the conditional mean alone: the sum of squared residuals (SSR) of Chan (1993) and Hansen (2000) is a pure measure of mean fit, on which a regime change in the variance leaves no systematic imprint, and the likelihood criterion of Samia and Chan (2011) could in principle use second moments but its theory does not incorporate regime-dependent variances in the threshold step. In a threshold model estimated this way, the location of the regimes is chosen as if the variance never changed: a volatility regime is invisible to the threshold step, even when the covariance is allowed to differ in the second estimation step. The threshold step is blind to volatility.

This paper closes both gaps. For the decoupled and the joint model we provide a likelihood-based threshold estimator in the spirit of Samia and Chan (2011) and establish consistency, super-consistency at rate T of all thresholds, and standard $\sqrt{T}$ Gaussian inference on the regime parameters; a one-step feasible GLS estimator is asymptotically equivalent to the exact profile likelihood, which makes the decoupled model no harder to estimate than a linear VAR on each combination of mean and variance regime. The limiting distributions of the thresholds are of compound-Poisson-argmax form; in the decoupled model the two thresholds are asymptotically independent, and the variance threshold admits a near-pivotal plug-in interval. The results also supply the frequentist foundations for the Bayesian threshold estimation of Gargiulo et al. (2026). The two models are tied together formally: the decoupled estimator nests the joint one in estimation, and a likelihood-gap rule selects between them consistently (Proposition 1), so the researcher need not decide ex ante whether the mean and the variance switch together. A Monte Carlo study delineates when second moments matter, by comparing the likelihood estimator with the least-squares criterion in three cases: when only the mean switches the two are equivalent; when both moments switch the likelihood halves the threshold error at identical regime-parameter precision, as super-consistency implies; when only the variance switches least squares is inconsistent while the likelihood estimator retains the T -rate. A single threshold imposed on a decoupled process leaves the unmodeled dimension unlearned at any sample size, and the decoupled estimator recovers the joint model, and selects it, when the data are generated by it. Second moments are never a cost and sometimes the only source of identification.

Our second contribution is empirical: a reconciliation of the disagreement on state-dependent fiscal multipliers. Auerbach and Gorodnichenko (2012) find government spending multipliers larger in recessions than in expansions; RZ2018 find no evidence of larger multipliers in times

of slack and, under their preferred identification, no significant differences across regimes; [Alloza \(2022\)](#) finds multipliers larger in expansions and argues that the disagreement traces to how the recession regime is defined. In all three the regimes are imposed, so the natural question is whether the disagreement is due to regime misspecification, and the natural answer is to estimate the regimes. We do so in RZ2018’s data and framework in their *slack exercise*, where the state is lagged unemployment and the imposed threshold is 6.5%, on their three samples: the full historical sample, the sample excluding WWII, and the post-WWII sample. Since the original paper studies regime-dependent effects of the military news shock of [Ramey \(2011\)](#) using an instrumental variable approach, we follow [Plagborg-Møller and Wolf \(2021\)](#) and estimate an internal-proxy SVAR. The interest-rate exercise at the zero lower bound (ZLB) is a complement and is relegated to the Appendix.

Three results organize the application. First, a mean-only threshold estimator (SSR-based) essentially confirms RZ2018’s cutoff (6.95% against 6.5% on the full sample): the disagreement is not about where the line is drawn. The likelihood estimator departs from it, to 3.6%, only because it also sees the second moments, and what it sees is the shock environment: the volatility of fiscal innovations is two to three times higher in a small set of wartime quarters, and the regime selected by a single threshold is a shock regime in disguise, its indicator nearly collinear with the largest realizations of the instrument. What looked like a state is a shock. Second, the decoupled estimator separates the two. Using the absolute value of the lagged instrument as the variance state, we identify a *fiscal-event regime* of 33 quarters, composed of the World Wars, Korea, Vietnam, the Reagan build-up, the Gulf War, 2002, 2008–2013. The estimated variance threshold is the same in every sample and for both mean state variables, which says the shock environment is a feature of the data rather than of the sample. Outside these quarters, state dependence is in the data. At RZ2018’s own 6.5%, once the fiscal-event regime is isolated, multipliers on both historical samples are larger in the high-unemployment regime, with the difference significant at every horizon once WWII is excluded; isolating the fiscal-event regime also strengthens the weak first stage of RZ2018’s low-unemployment state, because the wartime quarters carry a contemporaneous news–spending relation of the opposite sign. At the estimated thresholds the ranking depends on the sample: larger in the high-unemployment regime when the pre-WWII extreme-slack episodes are included, larger in the low-unemployment regime after WWII – there, a point-estimate ranking.

Third, because the variance state is built from the instrument itself, one may object that the fiscal-event regime merely tracks the heteroskedasticity of the proxy equation rather than

a feature of the macroeconomy. A robustness exercise in the framework of [Auerbach and Gorodnichenko \(2012\)](#) answers the objection by removing the instrument from the system: identification is recursive, the news series enters only as a candidate variance state, and the variance threshold is selected by the second moments of the macroeconomic aggregates alone. It recovers the identical fiscal-event partition, quarter for quarter, over the common sample, and on post-WWII data the larger multipliers in good times – there, the high-growth regime – become a statistically significant difference under this second identification. The disagreement thus decomposes into three ingredients – the shock environment, the sample period, and the definition of the regimes – of which only the last is addressed by the existing debate. The relevant notion of a regime is two-dimensional: no choice of recession definition can absorb the shock-environment dimension, which, held fixed at RZ2018’s own regime definition, restores state dependence on the historical samples on its own. The choice of state variables remains the researcher’s, and the framework is what disciplines it: a candidate state is confronted with the location of its cutoff, with whether its regimes are in the dynamics or in the volatility, and with the stability of the partition it delivers. We justify ours on these grounds, and show in the appendix what happens with the natural alternative, government spending itself, which classifies quarters by the response to shocks rather than by their arrival and recovers neither the fiscal-event regime nor a stable partition.

Related literature. The paper contributes first to the econometrics of threshold time-series models, which originate with [Tong and Lim \(1980\)](#); the estimation theory (consistency, super-consistency at rate T and the limiting distributions) is developed for univariate autoregressions by [Chan \(1993\)](#) and [Hansen \(2000\)](#) under least squares, with [Tsay \(1998\)](#) for multivariate testing and [Hansen \(2011\)](#) for a survey. The closest paper is [Samia and Chan \(2011\)](#), who take the likelihood route: they establish consistency and the compound-Poisson-argmax limit of the maximum-likelihood threshold estimator in a univariate generalized threshold regression, and their argument is the template for ours. Their theory, however, does not cover regime-dependent variances in the threshold step, so that the threshold is still located by the conditional mean. We carry their approach to multivariate systems in which the covariance switches with the mean, where the second moment contributes to the location of the threshold and the jumps of the limit process carry both mean and variance differences, and to the decoupled case in which mean and variance regimes have separate states and thresholds, each estimated at rate T ; the double-threshold ARCH model of [Li and Li \(1996\)](#) is a univariate precursor of joint mean–variance thresholding.

Relative to the latent-regime traditions – Markov switching ([Hamilton, 1989](#); [Sims and Zha,](#)

2006), drifting parameters with stochastic volatility (Cogley and Sargent, 2005; Primiceri, 2005; Chan and Eisenstat, 2018; Justiniano and Primiceri, 2008), and the endogenous Markov chains of Kim et al. (2008) and Chang et al. (2017) – threshold models tie the regimes to observable states known at $t - 1$, endogenous to the economy yet predetermined for estimation, at the cost of requiring candidate state variables; we extend this observable-state logic to the variance dimension. Relative to smooth-transition specifications (Teräsvirta, 1994; Auerbach and Gorodnichenko, 2012) we estimate the location of the regime change rather than calibrating the shape of the transition, and relative to the state-dependent local-projection literature (Jordà, 2005; Ramey and Zubairy, 2018; Gonçalves et al., 2024) we supply the estimation theory for the regime structure it conditions on.

On the empirical side the paper speaks to the state dependence of fiscal multipliers. Building on linear benchmarks (Blanchard and Perotti, 2002; Ramey, 2011), Auerbach and Gorodnichenko (2012) and Auerbach and Gorodnichenko (2013) find multipliers larger in downturns; Owyang et al. (2013) and RZ2018 find little or no slack dependence on long historical samples with imposed thresholds; Fazzari et al. (2015) are the exception that estimates the threshold, on capacity utilization; Bernardini and Peersman (2018) condition on private debt; Ghassibe and Zanetti (2022) show that the source of the fluctuations underlying the state matters; and Alloza (2022) traces the disagreement to the definition of the recession regime. Rather than proposing another state variable or cutoff, we estimate the regime structure itself. Closest to us, Barnichon et al. (2022) locate the nonlinearity in the *sign* of the shock: our shock-environment dimension is the second-moment counterpart of their idea, and the two mechanisms can in principle coexist.

The rest of the paper is organized as follows. Section 2 presents the models, the estimators and the asymptotic theory; Section 3 the Monte Carlo evidence; Section 4 the application to RZ2018, with the robustness exercise in the framework of Auerbach and Gorodnichenko (2012) in Section 4.7; Section 5 concludes. Proofs and additional results are in the appendix.

2 Threshold VARs with Decoupled Mean and Variance Regimes

Threshold autoregressions, introduced by Tong (1977) and developed by Tong and Lim (1980) and Chan (1993), model nonlinearity through a small number of regimes selected by an observable, $\mathcal{F}_{t-1}$ -measurable state variable: a parsimonious and interpretable framework,

extended to vector systems by [Tsay \(1998\)](#) without, so far, an estimation theory for the threshold itself. Estimation proceeds in two steps: first, the threshold is estimated by optimizing a criterion over a grid of possible values, then the regime parameters are obtained conditionally on it. The choice of criterion in the first step matters: the sum of squared residuals (SSR) of [Chan \(1993\)](#) and [Hansen \(2000\)](#) identifies the threshold from breaks in the conditional mean alone; [Samia and Chan \(2011\)](#) replace it with the likelihood, which can in principle carry information from the entire conditional distribution. Regime-dependent variances – which carry threshold information of their own and which neither framework accommodates – are the starting point of this paper.

The economic motivation is the one that animates the applications of Section 4 and Appendix I. The literature on state-dependent fiscal multipliers disagrees on whether multipliers are larger in recessions ([Auerbach and Gorodnichenko, 2012](#)), in expansions ([Alloza, 2022](#)), or not at all ([Ramey and Zubairy, 2018](#)), and all three impose the regime structure – a threshold on a single macroeconomic state – rather than estimating it. Two distinct mechanisms plausibly generate regime dependence: the state of the economy, which governs how a shock propagates, and the shock environment, which governs how large and how volatile the identifying shocks are. There is no *ex ante* reason for these two forms of variation to be governed by the same observable: the level of economic slack may determine whether fiscal multipliers are large or small, while the size of the arriving shocks determines the volatility environment in which they are identified. Macroeconomic data display well-documented changes in volatility ([Cogley and Sargent, 2005](#); [Primiceri, 2005](#)), and among regime-switching specifications of U.S. monetary policy the best fit is obtained when only the variances switch ([Sims and Zha, 2006](#)); regimes in the conditional variance therefore need not coincide with regimes in the conditional mean, in state variable, threshold, or number. A threshold VAR in which mean regimes and variance regimes are *decoupled* – driven by their own state variables and thresholds, with regime-specific coefficient and covariance matrices – formalizes the two mechanisms separately and lets the data locate both.

This section develops the likelihood-based estimation theory for that model. The model in which both switch together at a common threshold – the specification of [Gargiulo et al. \(2026\)](#) – is the special case of Section 2.5, where we show that the restriction can be selected rather than imposed. The theory is organized in three layers, each with its own conditions: (i) geometric ergodicity and moments of the data-generating process; (ii) identification and consistency of the estimator; (iii) regularity local to the thresholds, delivering the convergence rates and limiting distributions. Sections 2.1–2.2 define the model and the estimator;

Section 2.3 states the conditions, grouped by layer, with remarks collected after each group; Section 2.4 states the results.

2.1 The Model

Let $y_t \in \mathbb{R}^n$ follow a threshold VAR of order p with k_μ mean regimes and k_σ variance regimes,

$$y_t = \sum_{i=1}^{k_\mu} \left(\Phi_{i,0} + \sum_{j=1}^p \Phi_{i,j} y_{t-j} \right) \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu\} + \sum_{\ell=1}^{k_\sigma} \Sigma_\ell^{1/2} \mathbb{I}\{q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma\} \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, I_n), \quad (1)$$

where q_t^μ and q_t^σ are scalar, $\mathcal{F}_{t-1}$ -measurable state variables observed at delays $d_\mu, d_\sigma \in \{1, \dots, p\}$, and the regimes are the intervals $\mathcal{R}_i^\mu = [\gamma_{i-1}^\mu, \gamma_i^\mu)$ and $\mathcal{R}_\ell^\sigma = [\gamma_{\ell-1}^\sigma, \gamma_\ell^\sigma)$ generated by the ordered threshold vectors $\gamma^\mu \in \mathbb{R}^{k_\mu-1}$ and $\gamma^\sigma \in \mathbb{R}^{k_\sigma-1}$, with the conventions $\gamma_0 = -\infty$ and $\gamma_k = +\infty$. At each t the observation belongs to the joint cell (i, ℓ) . We write $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]'$, $B_i = (\Phi_{i,0}, \Phi_{i,1}, \dots, \Phi_{i,p})$, and collect the parameters in $\theta = \{\gamma^\mu, \gamma^\sigma, \varphi_1, \dots, \varphi_{k_\mu}, \text{vech}(\Sigma_1), \dots, \text{vech}(\Sigma_{k_\sigma})\}$. When the state variables are coordinates of y_t , $q_t^\mu = e'_{s_\mu} y_t$ and $q_t^\sigma = e'_{s_\sigma} y_t$ (e_s the s -th unit vector), (1) is a self-exciting threshold VAR; the regime assignment is governed by the scalar states, not by the full lagged vector. We write θ_0 for the true parameter. Let $\pi_{i\ell} := P_\pi(q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu, q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma)$ be the stationary probability of joint cell (i, ℓ) under the invariant law π of Theorem 1, and $t_{i\ell} := |\{t : q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu, q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma\}|$ the corresponding sample count.

Both the variables and the delays may differ across the two regime structures. The leading cases are two different coordinates of y_t , the same coordinate at two different lags ($q_t^\mu = q_t^\sigma = e'_{s_\mu} y_t$ with $d_\mu \neq d_\sigma$), and an economic state for the mean regimes paired with a measure of shock size for the variance regimes, as in the applications of Section 4 and Appendix I. A shared state variable at a common delay is also admissible: with distinct threshold values the two regime structures form a staircase of feasible cells and the theory below applies (Remark B.8); with a common threshold value the two structures coincide and (1) reduces to the joint mean–variance model of Section 2.5, which we show is nested in estimation.

2.2 The Estimator

Under innovation gaussianity imposed in (1), y_t lies in joint cell (i, ℓ) with mean $(x'_t \otimes I_n) \text{vec}(B_i)$ and covariance Σ_ℓ . We estimate θ by maximizing the resulting likelihood, concentrating the regime parameters out in closed form for each candidate pair of thresholds.

The average log-likelihood is

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^{k_\mu} \sum_{\ell=1}^{k_\sigma} \left(-\frac{1}{2} \ln \det \Sigma_\ell - \frac{1}{2} \varepsilon'_{it} \Sigma_\ell^{-1} \varepsilon_{it} \right) \mathbb{I}(q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu) \mathbb{I}(q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma), \quad (2)$$

where $\varepsilon_{it} = y_t - (x'_t \otimes I_n) \text{vec}(B_i)$ and $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]'$. Concentrating out $\{B_i\}$ and $\{\Sigma_\ell\}$ yields the profile likelihood. Conditional on $(\gamma^\mu, \gamma^\sigma)$, the coefficients B_i are estimated at the mean-regime level using variance-regime-specific precision weights, while the covariance matrices Σ_ℓ are estimated at the variance-regime level by pooling residuals across all mean regimes within regime ℓ . The first-order condition for $\beta_i := \text{vec}(B_i)$ is

$$\sum_{\ell=1}^{k_\sigma} \sum_{t \in S_{i\ell}} (x_t \otimes \Sigma_\ell^{-1}) [y_t - (x'_t \otimes I_n) \beta_i] = 0, \quad (3)$$

where

$$S_{i\ell} := \left\{ t : q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu, q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma \right\}.$$

This is a GLS problem: within mean regime i , different observations receive different precision weights Σ_ℓ^{-1} according to their variance-regime membership. Solving (3) gives the feasible GLS estimator

$$\hat{\beta}_i^{GLS}(\gamma^\mu, \gamma^\sigma) = \left[\sum_{\ell=1}^{k_\sigma} \sum_{t \in S_{i\ell}} (x_t x'_t \otimes \hat{\Sigma}_\ell^{-1}) \right]^{-1} \left[\sum_{\ell=1}^{k_\sigma} \sum_{t \in S_{i\ell}} (x_t \otimes \hat{\Sigma}_\ell^{-1}) y_t \right], \quad (4)$$

where $\hat{\Sigma}_\ell$ is a first-step consistent estimate of Σ_ℓ (see Remark below). Define $\hat{\beta}_i^{GLS} := \text{vec}(\hat{B}_i^{GLS})$, $\hat{\beta}_i^{(0)} := \text{vec}(\hat{B}_i^{(0)})$. We use $\hat{\beta}_i$ inside the GLS normal equations and $\hat{B}_i$ for the corresponding matrix-valued coefficient blocks.

Similarly, the first-order condition for Σ_ℓ yields the variance-regime-specific residual covariance estimator, obtained by pooling residuals across all mean regimes within the same variance regime ℓ

$$\hat{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma) = \frac{1}{t_\ell} \sum_{i=1}^{k_\mu} \sum_{t \in S_{i\ell}} \hat{\varepsilon}_{it} \hat{\varepsilon}'_{it}, \quad t_\ell := \sum_{i=1}^{k_\mu} t_{i\ell}, \quad (5)$$

where $\hat{\varepsilon}_{it} = y_t - (x'_t \otimes I_n) \hat{\beta}_i^{GLS}$ are the GLS residuals.

Substituting $\{\hat{B}_i^{GLS}\}_{i=1}^{k_\mu}$ and $\{\hat{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma)\}_{\ell=1}^{k_\sigma}$ into the concentrated Gaussian log-likelihood²

² L_T^p as defined in (6) uses the feasible one-step GLS plug-ins. Consistency is established directly for this feasible criterion (Lemma T5-F and Step 4 of Appendix A.3); its first-order equivalence with the exact profile

yields

$$L_T^p(\gamma^\mu, \gamma^\sigma) = -\frac{1}{2} \sum_{\ell=1}^{k_\sigma} \frac{t_\ell}{T} \ln \det \widehat{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma) + \text{const}, \quad (6)$$

and the threshold estimator is

$$(\hat{\gamma}^\mu, \hat{\gamma}^\sigma) = \arg \max_{(\gamma^\mu, \gamma^\sigma) \in \Gamma^\mu \times \Gamma^\sigma} L_T^p(\gamma^\mu, \gamma^\sigma). \quad (7)$$

where $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma)$ denotes the *smallest maximizer*; the convention is discussed in Section 2.4.

Remark E.1 (Feasible profile-GLS implementation). Since the variance-regime covariance matrices $\{\Sigma_\ell\}_{\ell=1}^{k_\sigma}$ are unknown, the GLS estimator in (4) is implemented by a feasible two-step procedure. For any candidate pair $(\gamma^\mu, \gamma^\sigma) \in \Gamma^\mu \times \Gamma^\sigma$:

- Assign each observation to its mean regime i , variance regime ℓ , and joint cell $S_{i\ell} := \{t : q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu, q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma\}$.
- Estimate preliminary mean-regime coefficients $\{\widehat{B}_i^{(0)}(\gamma^\mu)\}_{i=1}^{k_\mu}$ by OLS within each mean regime, ignoring heteroskedasticity at this initial step. Let $\widehat{u}_{it} = y_t - (x'_t \otimes I_n) \widehat{\beta}_i^{(0)}$.
- Construct preliminary variance-regime covariance estimators by pooling these residuals across all mean regimes within variance regime ℓ :

$$\widehat{\Sigma}_\ell^{(0)}(\gamma^\mu, \gamma^\sigma) = \frac{1}{t_\ell} \sum_{i=1}^{k_\mu} \sum_{t \in S_{i\ell}} \widehat{u}_{it} \widehat{u}_{it}', \quad t_\ell := \sum_{i=1}^{k_\mu} t_{i\ell}.$$

- Using the weights $(\widehat{\Sigma}_\ell^{(0)})^{-1}$, compute the feasible GLS estimator $\widehat{B}_i^{GLS}(\gamma^\mu, \gamma^\sigma)$ from (4).
- $\widehat{\varepsilon}_{it} = y_t - (x'_t \otimes I_n) \widehat{\beta}_i^{GLS}$, and update the variance-regime covariance matrices by

$$\widehat{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma) = \frac{1}{t_\ell} \sum_{i=1}^{k_\mu} \sum_{t \in S_{i\ell}} \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{it}'.$$

- Substitute $\{\widehat{\beta}_i^{GLS}\}_{i=1}^{k_\mu}$ and $\{\widehat{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma)\}_{\ell=1}^{k_\sigma}$ into the concentrated Gaussian log-likelihood (6), and define the threshold estimator as

$$(\hat{\gamma}^\mu, \hat{\gamma}^\sigma) = \arg \max_{(\gamma^\mu, \gamma^\sigma) \in \Gamma^\mu \times \Gamma^\sigma} L_T^p(\gamma^\mu, \gamma^\sigma).$$

$\sup_\vartheta L_T(\gamma^\mu, \gamma^\sigma, \vartheta)$ at the $1/T$ localization scale is Appendix A.5, Steps 3–4.

In practice, further GLS iterations may be performed, but one covariance update is sufficient to define a coherent feasible estimator and keeps the asymptotic analysis transparent. The algorithm in the Remark describes the feasible computation of the profiled criterion $L_T^p(\gamma^\mu, \gamma^\sigma)$ for each candidate pair $(\gamma^\mu, \gamma^\sigma)$. The overall estimator nevertheless retains the standard two-step structure of threshold models: first, obtain $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma)$ by maximizing the feasible profile criterion in (6); second, conditional on these estimated thresholds, compute the corresponding regime-parameter estimators $\{\hat{B}_i\}_{i=1}^{k_\mu}$ and $\{\hat{\Sigma}_\ell\}_{\ell=1}^{k_\sigma}$ from (4)–(5) evaluated at $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma)$.

By Theorem 2 and Theorem 3(i), the preliminary covariance estimators used in the feasible GLS step are consistent, and one covariance update is asymptotically first-order equivalent to the infeasible GLS estimator. Appendix A.5 provides the corresponding score-expansion argument. This is the standard justification for one-step FGLS under heteroskedasticity (see, e.g., Hamilton, 1994, Chapter 7)³. On the diagonal of the threshold space treated in Section 2.5, where mean and variance regimes coincide, the profile is closed-form and the one-step and exact maximizers coincide. Throughout, ℓ_t denotes a per-observation contribution to the log-likelihood, $L_T = T^{-1} \sum_t \ell_t$ the sample criterion in all its arguments, and $\ell = \mathbb{E}_\pi[\ell_t]$ its population counterpart, π being the stationary law of Theorem 1; a superscript p marks the profiled versions L_T^p, ℓ^p , obtained by concentrating out the regime parameters; $\mathcal{L}_T$ denotes a normalized local process and $l := T L_T$ the unnormalized likelihood, which appears only in Lemma 7. All $O_p(\cdot)$ statements refer to this normalization.

2.3 Assumptions

We state the conditions in the two groups in which they are used: those guaranteeing geometric ergodicity and moments of the data-generating process, and those delivering identification, the convergence rates and the limiting distributions of the estimator. Remarks collected after each group discuss their role and their relation to the literature.

³The estimator used throughout is this one-step version: GLS at the pooled residual covariance, then a single update of each Σ_ℓ . Iterating to the Gaussian maximum likelihood conditional on the regimes raises the log-likelihood by 8 to 9 points at the estimated thresholds on the two historical samples of Section 4 and by less than 2 elsewhere, and moves the multipliers reported there by at most 0.1 (0.2 in one cell); the profile criterion for the thresholds is evaluated at the one-step estimates.

2.3.1 Geometric Ergodicity

To study the stability of (1) we represent it as a Markov chain on $\mathbb{R}^{np}$. We stack the state

$$X_t := \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} \in \mathbb{R}^{np}.$$

and assume that the threshold variables $q_{t-d_\mu}^\mu$ and $q_{t-d_\sigma}^\sigma$ are measurable functions of the stacked state. This covers the empirical specifications of interest, such as the lagged level and the lagged change of a component of y_t , provided the relevant lags are contained in X_t . Hence no enlargement of the state space is needed. For each mean regime $i = 1, \dots, k_\mu$ let A_i denote the usual $np \times np$ companion matrix built from the regime- i coefficient blocks, and let $b_i \in \mathbb{R}^{np}$ collect the regime intercept placed in the first block. Let H_ℓ denote the $np \times n$ noise-loading matrix in regime ℓ (H_ℓ has only the first $n \times n$ block nonzero and equal to $\Sigma_\ell^{1/2}$).

Write the one-step evolution in terms of the (random) active regime $i^*(X_t)$ ⁴ and $\ell^*(X_t)$:

$$X_{t+1} = A_{i^*(X_t)} X_t + b_{i^*(X_t)} + H_{\ell^*(X_t)} \varepsilon_{t+1} \quad (8)$$

where $i^*(X_t)$ selects the mean-regime via $q_{t-d_\mu}^\mu$ and $\ell^*(X_t)$ selects the variance-regime via $q_{t-d_\sigma}^\sigma$; when indexing by time we write $i(t) := i^*(X_{t-1})$ and $\ell(t) := \ell^*(X_{t-1})$, so that observation t lies in true cell $(i(t), \ell(t))$ and $y_t = B_{i(t)}^0 x_t + \Sigma_{\ell(t)}^{0,1/2} \varepsilon_t$ under the true parameter. This is a time-homogeneous Markov chain with kernel

$$P(x, B) = \mathbb{P}(A_{i^*(x)} x + b_{i^*(x)} + H_{\ell^*(x)} \varepsilon \in B), \quad x \in \mathbb{R}^{np}, \quad B \in \mathcal{B}_{np}.$$

where $\mathcal{B}_{np}$ is the Borel σ -field on $\mathbb{R}^{np}$ and P is the probability measure on the probability space on which X_0 and $\{\varepsilon_t\}_{t \geq 1}$ are defined. We need to find the conditions under which model (1) is geometrically ergodic.

Geometric ergodicity. The Markov chain $\{X_t\}$ is *geometrically ergodic* if it has a unique invariant probability measure π and there exist constants $M(x) < \infty$ and $\rho \in (0, 1)$ such

⁴When indexing by time we write $i(t) := i^*(X_{t-1})$ for the regime active at t ; note $i(t)$ is therefore $\mathcal{F}_{t-1}$ -measurable.

that for all x and integers $t \geq 0$

$$\|P^t(x, \cdot) - \pi(\cdot)\|_{\text{TV}} \leq M(x) \rho^t.$$

Equivalently, there exist a function $g \geq 1$ and constants $R < \infty$, $\rho \in (0, 1)$ with $\|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq Rg(x)\rho^t$; in particular geometric ergodicity implies geometric β -mixing and exponential forgetting of initial conditions.

To achieve that goal, we make the following assumptions on model (1):

A1 (Innovations.) $\{\varepsilon_t\}$ is i.i.d., independent of X_0 , with $E[\varepsilon_t] = 0$, $E[\varepsilon_t \varepsilon_t'] = I_n$, an everywhere positive and continuous Lebesgue density on $\mathbb{R}^n$, and $m_4 := E\|\varepsilon_t\|^4 < \infty$.

A2 (Markov structure.) The coefficient matrices $\Phi_{i,j}$ and intercepts $\Phi_{i,0}$ are finite constants; the regime-selection maps $i^*(\cdot)$ and $\ell^*(\cdot)$ are measurable, so that X_t is a time-homogeneous Markov chain.

A3 (Non-degeneracy.) Each Σ_ℓ is non-singular and $\sup_\ell \|\Sigma_\ell\| \leq C_\Sigma < \infty$.

A4 (Joint stability of the mean regimes.) There exist $P \succ 0$ and $\lambda_i \in (0, 1)$ such that

$$A_i' P A_i \preceq \lambda_i^2 P, \quad \forall i = 1, \dots, k_\mu.$$

Remark A.1 (Role of Gaussianity). The Gaussianity imposed in (1) implies all the conditions in **A1**; it is used only for the likelihood-based estimation, not for the ergodicity results of this subsection, which require only a positive continuous innovation density and finite fourth moments.

Remark A.2 (Common Lyapunov function). **A4** is stronger than individual stability $\rho(A_i) < 1$ for all i : individual stability is necessary but not sufficient for a common P . The condition requires all mean regimes to contract the same energy metric $g(x) = 1 + x' P x$, which is what prevents instability from switching; for given parameter values it can be verified by LMI optimization (Liberzon, 2003, Chapter 2). A sufficient primitive condition is that $\rho(A_i) < \bar{\lambda} < 1$ for all i and $\max_{i,j} \|A_i - A_j\| < \varepsilon$ for some $\varepsilon > 0$ depending only on $\bar{\lambda}$, the dimension np , and $\max_i \|A_i\|$ (through the conditioning of the Lyapunov matrix constructed from any single regime), and not on the sample size: regimes with sufficiently similar dynamics cannot diverge by switching.

Remark A.3 (Variance regimes and stability). **A4** restricts only the k_μ mean companion matrices. No Lyapunov condition involving the variance-regime matrices Σ_ℓ is required, since they enter (8) only through the bounded noise loading $H_{\ell^*(X_t)}$; in the proof of Theorem 1 they affect only the additive drift constant.

Remark A.4 (Relation to the literature). The argument combines three classical ingredients: ν -irreducibility and aperiodicity from the everywhere-positive innovation density, a Foster–Lyapunov drift condition verified for the norm-inducing function g (Meyn and Tweedie, 1993, Ch. 15), and the equivalence of drift and geometric ergodicity with moment bounds for the invariant law (Tong, 1990; Chen and Chen, 2000). The novelty is the control of fourth moments under arbitrary regime switching, obtained by lifting **A4** to the Kronecker-squared companion matrices (Appendix A.1).

2.3.2 Identification, Rates and Limit Laws

We now state the assumptions that extend Chan (1993) and Samia and Chan (2011) to the multivariate threshold VAR setting with regimes in the variance.

Throughout, $\Lambda_t := 1 + \|x_t\|^2 + \|y_t\|^2$ is the moment envelope; $x_t = (q_{t-d}, x_t^{(-)})$ splits the regressor vector into a threshold coordinate and the remaining components where a state is a coordinate of x_t ; $\mathbb{I}_{i,t}^\mu := \mathbb{I}\{|q_{t-d_\mu}^\mu - \gamma_i^{\mu,0}| \leq \Delta\}$ and $\mathbb{I}_{\ell,t}^\sigma$ are the band indicators; f_μ and f_σ are the stationary marginal densities of the two state variables.

B1 (Compact parameter space.) $\Theta = \Gamma^\mu \times \Gamma^\sigma \times \mathcal{B} \times \mathcal{S}$ is compact, with $\Gamma^\mu = \{\gamma^\mu : \underline{\gamma} \leq \gamma_1^\mu < \dots < \gamma_{k_\mu-1}^\mu \leq \bar{\gamma}\}$ and Γ^σ defined analogously, $\mathcal{B} = \prod_i \mathcal{B}_i$ compact and consistent with **A4**, and $\mathcal{S} = \prod_\ell \{\Sigma : \underline{\sigma}I \prec \Sigma \prec \bar{\sigma}I\}$ for $0 < \underline{\sigma} < \bar{\sigma} < \infty$.

B2 (Separation at the thresholds.) For each $i = 1, \dots, k_\mu - 1$: f_μ is continuous and positive at $\gamma_i^{\mu,0}$, and $P((B_{i+1}^0 - B_i^0)x_t \neq 0 \mid q_{t-d_\mu}^\mu = \gamma_i^{\mu,0}) > 0$. For each $\ell = 1, \dots, k_\sigma - 1$: f_σ is continuous and positive at $\gamma_\ell^{\sigma,0}$, and $\Sigma_\ell^0 \neq \Sigma_{\ell+1}^0$. Every joint cell (i, ℓ) with $\mathcal{R}_i^\mu \cap \mathcal{R}_\ell^\sigma \neq \emptyset$ has $\pi_{i\ell} > 0$.

B3 (Regimes are different.) $B_i^0 \neq B_{i'}^0$ for $i \neq i'$ and $\Sigma_\ell^0 \neq \Sigma_{\ell'}^0$ for $\ell \neq \ell'$.

B4 (Interior parameters.) The true parameter θ_0 is interior to Θ .

B5 (Local ρ -mixing.) There exists $\Delta > 0$ such that the process stacking the two flagged

localized observation vectors,

$$V_t^D := \left(\mathbb{I}\{|q_{t-d_\mu}^\mu - \gamma^{\mu,0}| \leq \Delta\}, x_t' \mathbb{I}\{|q_{t-d_\mu}^\mu - \gamma^{\mu,0}| \leq \Delta\}, y_t' \mathbb{I}\{|q_{t-d_\mu}^\mu - \gamma^{\mu,0}| \leq \Delta\}, \right. \\ \left. \mathbb{I}\{|q_{t-d_\sigma}^\sigma - \gamma^{\sigma,0}| \leq \Delta\}, x_t' \mathbb{I}\{|q_{t-d_\sigma}^\sigma - \gamma^{\sigma,0}| \leq \Delta\}, y_t' \mathbb{I}\{|q_{t-d_\sigma}^\sigma - \gamma^{\sigma,0}| \leq \Delta\} \right)',$$

is ρ -mixing with summable mixing coefficients. Every localized summand appearing in the proofs is a fixed measurable function of V_t^D and inherits its mixing coefficients.

B6 (Conditional moments near the thresholds.) There exist $\tau > 0$, $M < \infty$ with $E[\Lambda_t^2 \mid q_{t-d_\mu}^\mu = v] \leq M$ for a.e. v within τ of each $\gamma_i^{\mu,0}$ and $E[\Lambda_t^2 \mid q_{t-d_\sigma}^\sigma = v] \leq M$ for a.e. v within τ of each $\gamma_\ell^{\sigma,0}$.

B7 (Weak continuity at the mean thresholds.) The conditional distribution of the pair $(x_t^{(-)}, \mathbb{I}\{q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^{\sigma,0}\}_{\ell=1,\dots,k_\sigma})$ given $q_{t-d_\mu}^\mu = v$ is weakly continuous at $v = \gamma_i^{\mu,0}$ for each i , where $x_t^{(-)}$ collects the components of x_t other than $q_{t-d_\mu}^\mu$ (when the latter is a coordinate of x_t). In particular the conditional variance-regime weights

$$p_\ell(v) := P(q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^{\sigma,0} \mid q_{t-d_\mu}^\mu = v), \quad \ell = 1, \dots, k_\sigma,$$

are continuous at $v = \gamma_i^{\mu,0}$ for each i ; write $p_\ell := p_\ell(\gamma_i^{\mu,0})$, suppressing the threshold index i when no confusion arises.

B8 (Bounded joint density near the threshold pairs.) For each pair (i, ℓ) , the pair $(q_{t-d_\mu}^\mu, q_{t-d_\sigma}^\sigma)$ admits a joint density bounded in a neighborhood of $(\gamma_i^{\mu,0}, \gamma_\ell^{\sigma,0})$.

(F) (Uniform candidate feasibility.) $\inf_{\gamma^\mu \in \Gamma^\mu} \min_i P_\pi(q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu(\gamma^\mu)) > 0$ and $\inf_{\gamma^\sigma \in \Gamma^\sigma} \min_\ell P_\pi(q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma(\gamma^\sigma)) > 0$.

Remark B.1 (Separation). The mean-side clause of **B2** requires that adjacent mean regimes differ with positive probability *at* the threshold, not merely near it: it is the multivariate form of the discontinuity condition of [Chan \(1993, Condition 4\)](#) and of [Samia and Chan \(2011, Assumption 1\)](#), and it is what drives both the identification argument ([Appendix A.3](#)) and the strictly negative drift of the mean-threshold local process ([Theorem 3](#)). It is weaker in spirit than a jump in the conditional mean of y_t : what matters is that the two regimes be distinguishable at the threshold with positive probability. Band-level separation, which is all consistency needs, is automatic from **B3** and the everywhere-positive stationary density

(Lemma 3(iii)): when $B_i^0 \neq B_{i+1}^0$ the set $\{(B_{i+1}^0 - B_i^0)x_t = 0\}$ is Lebesgue-null. The exact-threshold clause is the substantive content, consumed by Theorem 3 where, with **B7**, it guarantees nondegenerate jump laws; separation that degraded at the threshold would collapse the uniform negative drift behind the T -rate. The variance-side clause needs no conditioning, as covariance differences separate the laws everywhere; it does not follow from **B3**, which only requires the variance regimes to be pairwise distinct. The cell clause is the population counterpart of every feasible cell being populated, so that the GLS normal equations and the pooled covariances are well defined; it is required only of structurally feasible cells (Remark B.8). Taken together, **B2–B3** ensure that distinct regimes are not observationally equivalent either at the threshold boundaries or away from them.

Remark B.2 (Parameter bounds). The feasible estimator's population plug-in values must lie in $\mathcal{B} \times \mathcal{S}$ for the profile to dominate the plug-in criterion (Lemma T5-F(ii), Appendix A.3); since these values form a bounded family, continuous on the compact threshold space, a finite enlargement of the bounds in **B1** always suffices.

Remark B.3 (Candidate feasibility). **(F)** concerns *marginal* regime probabilities under *candidate* thresholds; empty joint cells at candidate pairs are harmless, since the GLS normal equations and the pooled covariances aggregate over cells within a mean regime (respectively, a variance regime). With two regimes per dimension the candidate regimes are half-lines, so **(F)** reduces to the state marginals placing positive mass below $\underline{\gamma}$ and above $\bar{\gamma}$, automatic whenever the marginal state densities are positive on $\mathbb{R}$ (Lemma 3(iii) for coordinates of y_t). With more regimes it additionally requires a minimum regime width in $\Gamma^\mu, \Gamma^\sigma$: the population counterpart of the minimum-observations guard in the implementation and of the trimming condition standard in the structural-break literature (Bai and Perron, 1998).

Remark B.4 (Local mixing). A stationary process $\{W_t\}$ is ρ -mixing with coefficients $\rho(k) \rightarrow 0$ if $|\text{corr}(f, g)| \leq \rho(k)$ for all square-integrable f measurable with respect to $\sigma(W_t, t \leq j)$ and g measurable with respect to $\sigma(W_t, t \geq j+k)$ (Billingsley, 1999, §19); see Doukhan (1994) and Bradley (2005) for the relations among mixing conditions. **B5** is the multivariate analogue of Assumption 7 of Samia and Chan (2011): it controls the dependence of the observations local to the thresholds, where the limiting distributions are decided. Three points. First, **B5** is a genuine strengthening of what Theorem 1 delivers: geometric ergodicity implies geometric β -mixing, which does not imply ρ -mixing for non-reversible chains (Bradley, 2005). Second, the assumption is stated for the *generating* localized vector – the band indicator together with x_t and y_t – rather than for a summary of it, so that every summand appearing in the proofs ($\Lambda_t \mathbb{I}$, the jump variables $\xi \mathbb{I}$, the band indicators themselves) is a fixed measurable function of the

assumed process and inherits its mixing coefficients; this is the level at which the condition is used in [Samia and Chan \(2011\)](#), where the two formulations coincide because the regressors are lags of the threshold variable. Third, **B5** is dispensable when both threshold variables are coordinates of y_t at fixed lags: Lemma 4 (Appendix A.2) derives the two consequences of **B5** used in the proofs from geometric ergodicity and the bounded one-step conditional density of Lemma 3, at the cost of strengthening **B6** to the conditional-moment condition **B6*** stated there. We keep **B5** in the statements to preserve the parallel with [Samia and Chan \(2011\)](#); with general state transformations it remains the operative condition. Formally, Theorems 3 and 4 hold with **B5** replaced by **B6*** in this case; we do not state that variant separately. Beyond coordinates, **B5–B7** hold for fixed linear combinations of y_t 's components with a nonzero contemporaneous coefficient, finite moving averages of such combinations, and absolute values thereof, since each is, conditionally on the past, a nondegenerate Gaussian functional with density bounded by a multiple of $C_* = (2\pi\sigma_*)^{-1/2}$, the constant of Lemma 3 (Appendix A.2) (for absolute values this is the folded-Gaussian bound: if $W \mid \mathcal{F} \sim N(m, s^2)$ with $s^2 \geq \underline{s}^2$, the density of $|W|$ is at most $2(2\pi\underline{s}^2)^{-1/2}$).

Remark B.5 (Conditional moments). This is the analogue of Assumption 6 of [Samia and Chan \(2011\)](#). It does not follow from the unconditional moment bound $\int \|x\|^4 \pi(dx) < \infty$ of Theorem 1, because an unconditional bound does not control conditional moments on the (null) event $\{q_{t-d} = v\}$; it must be imposed. Its role: the likelihood jumps incurred by reassigning an observation across regimes (defined before Theorem 3) are quadratic in (x_t, ε_t) and hence dominated by $C \Lambda_t$ for a constant C depending only on the parameter bounds in **B1**. **B6** therefore guarantees that the jumps have finite conditional second moments near the thresholds, which feeds tightness and uniform integrability in the proof of Theorem 3. **B6** is imposed at both thresholds; its role is identical on the two sides.

Remark B.6 (Weak continuity, mean side only). This is the analogue of Assumption 8 of [Samia and Chan \(2011\)](#); see [Feller \(1971\)](#) for weak continuity. Because, for a coordinate state, $q_{t-d_\mu}^\mu$ is itself a coordinate of x_t , conditioning on $q_{t-d_\mu}^\mu = v$ fixes one coordinate of the regressor vector, and the assumption is accordingly phrased on the conditional law of the remaining coordinates. Its role is to make the one-sided limiting jump laws in Theorem 3 well-defined probability distributions: the law of the reassignment loss given $q_{t-d_\mu}^\mu = \gamma_i^{\mu,0}$ from the right (respectively from the left) is defined as the limit of the conditional law given $\gamma_i^{\mu,0} \leq q_{t-d_\mu}^\mu < \gamma_i^{\mu,0} + \delta$ (respectively $\gamma_i^{\mu,0} - \delta \leq q_{t-d_\mu}^\mu < \gamma_i^{\mu,0}$) as $\delta \downarrow 0$, and **B7** guarantees these limits exist. Note that the innovation ε_t requires no such condition: it is independent of $\mathcal{F}_{t-1}$, hence of $q_{t-d_\mu}^\mu$. No weak-continuity condition is needed at the variance threshold.

The variance-side jump variables (V2) below are measurable functions of ε_t alone, and ε_t is independent of $\mathcal{F}_{t-1}$, hence of $q_{t-d_\sigma}^\sigma$: the variance-side marks are exactly i.i.d. with a fully explicit law, whatever the conditional distribution of the state at the threshold.

Remark B.7 (The price of two thresholds). **B8** is the local-boundedness counterpart of the cell clause of **B2**: the latter asks that every feasible joint cell have positive probability, the former that the joint law of the pair of threshold variables place no excess mass near any threshold pair. Its role is a counting bound. With a bounded joint density, the probability that an observation lies within ε of *both* a mean and a variance threshold is $O(\varepsilon^2)$, so at the $1/T$ scale of the localization the expected number of observations reassigned in both dimensions at once is $O(1/T)$: the two local criterion processes are driven, with probability tending to one, by disjoint sets of boundary observations. This is what makes the limiting Poisson random measure factorize and the threshold estimators asymptotically independent in Theorem 3(ii). **B8** holds automatically when the two threshold variables are distinct coordinates of y_t at (possibly different) lags: conditioning first on the information preceding the earlier of the two dates and then sequentially, each coordinate has a one-step conditional density bounded by C_* (Lemma 3), so the pair's joint density is bounded by C_*^2 ; when the two dates coincide, the pair is conditionally jointly Gaussian with a nonsingular 2×2 covariance submatrix of some Σ_ℓ^0 (**A3**), with the same conclusion.

Remark B.8 (Shared state variables). **B8** is sufficient but not necessary: what the proofs use is the coincidence bound $P(|q_{t-d_\mu}^\mu - \gamma_i^{\mu,0}| \leq \varepsilon, |q_{t-d_\sigma}^\sigma - \gamma_\ell^{\sigma,0}| \leq \varepsilon) \leq C\varepsilon^2$. As stated, **B8** excludes any shared state variable, since the pair then concentrates on a line and admits no bounded joint density; but the bound holds trivially when the state is shared and the threshold values differ: with $\Delta := |\gamma_i^{\mu,0} - \gamma_\ell^{\sigma,0}| > 0$ the two bands are disjoint for $\varepsilon < \Delta/2$. In that case all conclusions of Theorems 2 and 3 apply with two modifications – the cell clause of **B2** is required only of the structurally feasible cells, one joint cell being empty when the state is shared, and the mean-side jump mixture of Remark 1 degenerates to the single variance regime containing the mean threshold. Only equal threshold variables *and* equal threshold values fall outside Theorem 3(ii): that configuration is the joint model, whose thresholds remain consistently estimable at rate T by the unrestricted estimator, with a different limit law (Section 2.5, Proposition 1).

Remark B.9 (Relation to the literature). The conditions extend [Chan \(1993\)](#) and [Samia and Chan \(2011\)](#) to a multivariate threshold VAR with regime-dependent covariances and two threshold mechanisms. [Chan \(1993\)](#) assumes a compact parameter space, stationarity, threshold identifiability, regime distinctness and an interior true parameter; [Hansen \(2000\)](#)

adds the local-to-threshold conditional-moment conditions that drive the rate; the Assumptions of Samia and Chan (2011) are the likelihood counterparts, with their Assumptions 5–8 corresponding to our **B5–B7** at a single threshold. Stationarity, ergodicity and the fourth moments needed for the uniform law of large numbers are not assumed but delivered by Theorem 1 under **A1–A4**. What has no counterpart in these papers is the second threshold mechanism: the cell clause of **B2**, **B8**, and **(F)** exist because two partitions of the sample must be identified and localized at once.

2.4 Results

We state the three results in the order of the layers: ergodicity, consistency, and the rates and limiting distributions. All proofs are in the Appendix.

Theorem 1. *Under assumptions **A1–A4**, the Markov chain $\{X_t\}$ is geometrically ergodic, with a unique invariant probability measure π satisfying $\int \|x\|^4 \pi(dx) < \infty$.*

The proof can be found in Appendix A.1. It follows Chen and Chen (2000) in verifying a geometric drift condition for a norm-inducing Lyapunov function (Remark A.4).

Theorem 2. *Under assumptions **A1–A4**, **B1–B4** and condition **(F)**, if the data $\{y_t\}_{t=1}^T$ are generated by (1) at the true parameter θ_0 , the feasible estimator of Section 2.2 satisfies*

$$\hat{\theta}_T \rightarrow_{\text{a.s.}} \theta_0 \quad \text{as } T \rightarrow \infty; \quad (9)$$

*in particular $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma) \rightarrow_{\text{a.s.}} (\gamma^{\mu,0}, \gamma^{\sigma,0})$.*⁵

The proof is in Appendix A.3; it establishes a uniform law of large numbers over the compact parameter space, uniqueness of the population maximizer by the identification argument, and consistency of the feasible maximizer through the uniform validity of the one-step feasible profile (Lemma T5-F), via the argmax theorem of Newey and McFadden (1994).

Limit objects. For the mean threshold $\gamma_i^{\mu,0}$ write $\delta_{i,t} := (B_{i+1}^0 - B_i^0)x_t$ and $\kappa_{i,t} := \delta'_{i,t} \Sigma_{\ell(t)}^{0,-1} \delta_{i,t}$, where $\ell(t) = \ell^*(X_{t-1})$ is the variance regime of observation t .⁶ The log-likelihood loss from evaluating an observation of mean regime $i+1$ under the law of regime i , holding

⁵In the two-regime baseline $k_\mu = k_\sigma = 2$, condition **(F)** and the parameter-bound requirement of Remark B.2 hold automatically, the latter after at most a finite enlargement of the bounds in **B1**.

⁶The index ℓ denotes a variance regime throughout; it carries a time argument, $\ell(t)$, when it denotes the variance regime of observation t .

its variance regime fixed, is

$$\xi_{i,t}^{\mu,+} = -\frac{1}{2}\kappa_{i,t} - \delta'_{i,t}\Sigma_{\ell(t)}^{0,-1/2}\varepsilon_t, \quad \xi_{i,t}^{\mu,+} \mid (x_t, \ell(t)) \sim \mathcal{N}(-\kappa_{i,t}/2, \kappa_{i,t}), \quad (\text{V1})$$

and symmetrically for $\xi_{i,t}^{\mu,-}$ with the regime labels interchanged. The loss from evaluating an observation of variance regime $\ell + 1$ under the law of regime ℓ , holding its mean regime fixed, depends on ε_t alone: with $\lambda_1^{(\ell)}, \dots, \lambda_n^{(\ell)}$ the eigenvalues of $\Sigma_\ell^{0,-1}\Sigma_{\ell+1}^0$ and $Z_1, \dots, Z_n$ i.i.d. standard normal,

$$\xi_\ell^{\sigma,+} \stackrel{d}{=} \frac{1}{2} \sum_{j=1}^n [\ln \lambda_j^{(\ell)} + (1 - \lambda_j^{(\ell)})Z_j^2], \quad \xi_\ell^{\sigma,-} \stackrel{d}{=} \frac{1}{2} \sum_{j=1}^n [-\ln \lambda_j^{(\ell)} + (1 - 1/\lambda_j^{(\ell)})Z_j^2], \quad (\text{V2})$$

with laws $\mu_\ell^{\sigma,\pm}$. The one-sided limit laws of the mean jumps are the mixtures over variance regimes

$$\mu_i^{\mu,\pm} := \sum_{\ell=1}^{k_\sigma} p_\ell \int \mathcal{N}(-\kappa/2, \kappa) \Big|_{\kappa=\delta'\Sigma_\ell^{0,-1}\delta, \delta=(B_{i+1}^0-B_i^0)x} dF_i^\pm(x), \quad (\text{V3})$$

where $F_i^\pm$ is the one-sided conditional law of x_t given $q_{t-d_\mu}^\mu = \gamma_i^{\mu,0}$ and p_ℓ the variance-regime weights, both well defined by **B6–B7**. These laws may carry an atom at zero, when $\delta_{i,t} = 0$ with positive conditional probability at the threshold; by **B2** its mass $p_0 < 1$, and the nonzero part is absolutely continuous. Zero jumps do not move the process, which therefore coincides with a compound-Poisson process of thinned rate $f_\mu(\gamma_i^{\mu,0})(1 - p_0)$ and continuous jump law; the almost-sure uniqueness of the maximizing interval then follows as in Remark 1 of [Samia and Chan \(2011\)](#). All jump means are strictly negative: each equals minus a Kullback–Leibler divergence between adjacent regime laws, positive by **B2–B3**.

For a rate $f > 0$ and mark laws (μ^+, μ^-) , let $J(h)$, $h \in \mathbb{R}$, be the two-sided compound Poisson process with $J(0) = 0$, independent Poisson arrivals of rate f on each half-line, marks drawn from μ^+ for $h > 0$ and from μ^- for $h < 0$. Since both sides drift to $-\infty$, J has an almost surely unique maximizing interval $[M_-, M_+]$; we write $M_- =: \operatorname{argmax}_-(J)$ for its left endpoint, the *smallest maximizer*, which is the convention of the estimator, following [Chan \(1993\)](#) for the least-squares threshold estimator.⁷ Define

$$M_i^\mu := \operatorname{argmax}_-(J_i^\mu), \quad M_\ell^\sigma := \operatorname{argmax}_-(J_\ell^\sigma),$$

⁷[Yu \(2012\)](#) shows, in the i.i.d. threshold-regression setting, that the midpoint of the maximizing interval has smaller higher-order risk; nothing in our first-order theory depends on the choice, but the limit random variable is the left endpoint M_- under the smallest-maximizer convention.

where J_i^μ has rate $f_\mu(\gamma_i^{\mu,0})$ and marks $\mu_i^{\mu,\pm}$, J_ℓ^σ has rate $f_\sigma(\gamma_\ell^{\sigma,0})$ and marks $\mu_\ell^{\sigma,\pm}$, and all processes $\{J_i^\mu\}_i, \{J_\ell^\sigma\}_\ell$ are mutually independent. For threshold i of the mean structure define the normalized local criterion process $\mathcal{L}_{T,i}^\mu(h) := T[L_T^p(\gamma^{\mu,0} + h e_i/T, \gamma^{\mu,0}) - L_T^p(\gamma^{\mu,0}, \gamma^{\mu,0})]$, $h \in \mathbb{R}$, with e_i the i -th unit vector, and $\mathcal{L}_{T,\ell}^\sigma$ analogously. Finally let

$$M_i^{GLS} := E_\pi[x_t x_t' \otimes \Sigma_{\ell(t)}^{0,-1} \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^{\mu,0}\}], \quad \pi_\ell^\sigma := P_\pi(q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^{\sigma,0}),$$

and let D_n^+ be the Moore–Penrose inverse of the duplication matrix.

Theorem 3. *Let the estimator be that of Section 2.2 and assume **A1–A4**, **B1–B8** and **(F)**. Then:*

(i) *(Super-consistency of the thresholds and $\sqrt{T}$ -consistency of the regime parameters.)*

$$\begin{aligned} T(\hat{\gamma}_i^\mu - \gamma_i^{\mu,0}) &= O_p(1) \quad \text{and} \quad T(\hat{\gamma}_\ell^\sigma - \gamma_\ell^{\sigma,0}) = O_p(1) \quad \text{for every } i, \ell \\ \sqrt{T}(\hat{B}_i - B_i^0) &= O_p(1) \quad \text{and} \quad \sqrt{T} \text{vech}(\hat{\Sigma}_\ell - \Sigma_\ell^0) = O_p(1) \end{aligned}$$

(ii) *(Thresholds.) Jointly, the local processes $(\mathcal{L}_{T,i}^\mu)_i, (\mathcal{L}_{T,\ell}^\sigma)_\ell$ converge weakly in $D(\mathbb{R})^{k_\mu+k_\sigma-2}$, product Skorokhod topology, to the independent processes $(J_i^\mu)_i, (J_\ell^\sigma)_\ell$, and*

$$\left(T(\hat{\gamma}_i^\mu - \gamma_i^{\mu,0})_{i=1}^{k_\mu-1}, T(\hat{\gamma}_\ell^\sigma - \gamma_\ell^{\sigma,0})_{\ell=1}^{k_\sigma-1} \right) \Rightarrow \left((M_i^\mu)_{i=1}^{k_\mu-1}, (M_\ell^\sigma)_{\ell=1}^{k_\sigma-1} \right),$$

with all $k_\mu + k_\sigma - 2$ limits mutually independent.

(iii) *(Regime parameters.) Jointly with (ii) and asymptotically independently of the threshold estimators,*

$$\sqrt{T} \text{vec}(\hat{B}_i - B_i^0) \Rightarrow \mathcal{N}(0, (M_i^{GLS})^{-1}), \quad \sqrt{T} \text{vech}(\hat{\Sigma}_\ell - \Sigma_\ell^0) \Rightarrow \mathcal{N}(0, \frac{2}{\pi_\ell^\sigma} D_n^+(\Sigma_\ell^0 \otimes \Sigma_\ell^0) D_n^{+'}),$$

with all Gaussian blocks – across mean regimes, across variance regimes, and between the B - and Σ -blocks – mutually independent, and the one-step feasible GLS estimator first-order equivalent to its infeasible counterpart.

The proof is in Appendix A.5 (part (i), and parts (ii)–(iii)): the localization argument of part (i) extends the single-threshold argument to the box of candidate pairs, the feasible-GLS layer being handled by the profile-stability Lemma 7; parts (ii)–(iii) establish the convergence of the local criterion processes to the compound-Poisson limits, their asymptotic independence,

and the argmax continuous-mapping step, treating one mean and one variance threshold explicitly and extending to general k_μ, k_σ in the closing paragraph of Appendix A.5.

Remark 1 (The mean–variance asymmetry of the limit). The two threshold limits are not symmetric. The mean-threshold jumps are Gaussian conditionally on $(x_t, \ell(t))$, with precision $\kappa_{i,t} = \delta'_{i,t} \Sigma_{\ell(t)}^{0,-1} \delta_{i,t}$ that depends on the variance regime in force, so the limit law $\mu_i^{\mu,\pm}$ is a mixture over variance regimes with the weights p_ℓ of **B7**: the precision with which a mean threshold is located depends on which variance regime prevails near it. The variance-threshold jumps, by contrast, are free of x_t and homogeneous across mean regimes. Hence the limit law of $\hat{\gamma}_i^\mu$ depends on the interaction of the two regime structures, while that of $\hat{\gamma}_\ell^\sigma$ does not; Table 3 displays the finite-sample counterpart, with the mean threshold estimated more precisely than the variance one in every design. Both thresholds are super-consistent, converging at rate T rather than the parametric $\sqrt{T}$, because the likelihood is discontinuous in each threshold: moving a threshold by h/T reassigns $O_p(1)$ observations, each paying a strictly positive expected loss. Unlike in Chan (1993) and Hansen (2000), where the rate is driven by a jump in the conditional mean alone, here it follows from a jump in the entire conditional law, so that a shift in the covariance matrices alone suffices (Section 2.5, Remark 8); the rate would collapse to $\sqrt{T}$ only if the regimes joined continuously at the threshold, the case of Chan and Tsay (1998).

Remark 2 (Why the two thresholds are asymptotically independent). The proof shows that the probability that any observation lies simultaneously within $O(1/T)$ of both thresholds is $O(1/T)$ under **B8**: with probability approaching one, the two local processes are driven by *disjoint* sets of boundary observations, and the limiting marked point process on the four half-lines is a Poisson random measure, whose restrictions to disjoint carriers are independent. On the diagonal $q^\mu \equiv q^\sigma$ with $\gamma_i^{\mu,0} = \gamma_\ell^{\sigma,0}$ for some pair (i, ℓ) , the two boundary bands coincide, the double-crossing count is $O_p(1)$ rather than $o_p(1)$, and the factorization fails, which is exactly why that case is excluded and assigned to the joint model. With a shared state but distinct thresholds, the bands are separated by $|\gamma_i^{\mu,0} - \gamma_\ell^{\sigma,0}| > 0$ and disjoint deterministically for all large T , so the coincidence bound holds trivially.

Remark 3 (GLS efficiency). By Theorem 3(iii), the asymptotic distribution of

$$\sqrt{T} \text{vec}(\hat{B}_i - B_i^0)$$

is centered Gaussian with variance-covariance matrix

$$(M_i^{GLS})^{-1}, \quad M_i^{GLS} = E_\pi \left[(x_t x_t' \otimes (\Sigma_{\ell(t)}^0)^{-1}) \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu\} \right].$$

This is weakly more efficient than the OLS-based estimator in the sense that

$$(M_i^{GLS})^{-1} \preceq (M_i^{OLS})^{-1} V_i^{OLS} (M_i^{OLS})^{-1}$$

in the positive-semidefinite order, where $M_i^{OLS} := E_\pi [x_t x_t' \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^{\mu,0}\}] \otimes I_n$ and $V_i^{OLS} := E_\pi [x_t x_t' \otimes \Sigma_{\ell(t)}^0 \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^{\mu,0}\}]$ is the middle matrix of the OLS sandwich.

Remark 4 (Inference). The limit laws of the mean thresholds, M_i^μ , are not asymptotically pivotal: they depend on $f_\mu(\gamma_i^{\mu,0})$ and on the mixture jump laws $\mu_i^{\mu,\pm}$, hence on the conditional law of x_t at the threshold. Two consequences follow. First, the nonparametric bootstrap is inconsistent for the thresholds under fixed-threshold-effect asymptotics (Yu, 2014); bootstrap bands for objects computed *conditionally* on the threshold estimates – regime-specific coefficients, impulse responses, multipliers – remain justified by the super-consistency of Theorem 3(i). Second, the valid routes for the thresholds themselves are subsampling (Gonzalo and Wolf, 2005), whose validity requires only a nondegenerate limit law at a known rate, both delivered by Theorem 3, and the simulation plug-in of Li and Ling (2012), which for a mean threshold requires estimating the conditional law of x_t at the threshold together with $f_\mu(\gamma_i^{\mu,0})$. For the variance thresholds the plug-in becomes fully parametric (Remark 5).

Remark 5 (Near-pivotal inference for the variance thresholds). Because the variance-jump laws $\mu_\ell^{\sigma,\pm}$ depend on $(\Sigma_\ell^0, \Sigma_{\ell+1}^0)$ only through the eigenvalues $\lambda_j^{(\ell)}$, the limit law of $T(\hat{\gamma}_\ell^\sigma - \gamma_\ell^{\sigma,0})$ is completely determined by $(f_\sigma(\gamma_\ell^{\sigma,0}), \Sigma_\ell^0, \Sigma_{\ell+1}^0)$: there is no dependence on unknown conditional distributions of the state at the threshold. Confidence sets for $\gamma_\ell^{\sigma,0}$ can therefore be constructed by a parametric plug-in: estimate $\lambda_j^{(\ell)}$ from $(\hat{\Sigma}_\ell, \hat{\Sigma}_{\ell+1})$ and $f_\sigma(\gamma_\ell^{\sigma,0})$ by a univariate kernel estimate at $\hat{\gamma}_\ell^\sigma$, simulate the compound-Poisson argmax, and read off quantiles. No analogous shortcut exists for the mean thresholds, or for the joint threshold of Section 2.5, whose jump laws involve the conditional law of x_t at the threshold; there the routes of Remark 4 remain the recommended ones. The present plug-in is the fully parametric special case of the simulation approach, made feasible by the fact that the variance-jump law involves no conditional feature of the state.

Remark 6 (Bayesian estimation). The results above provide frequentist foundations for Bayesian estimation of both models. Gargiulo et al. (2026) estimate the joint model of Sec-

tion 2.5 (with $k = 3$) by combining a flat prior on the threshold over a fine grid with smooth conjugate priors on the regime parameters, maximizing the resulting profile posterior over the grid and conditioning slope inference on the threshold estimate. Our theory delivers the two properties that procedure relies on. First, since the profiled parameters move by $O_p(T^{-1})$ as the threshold sweeps an $O(T^{-1})$ -neighborhood (Lemma 7, Appendix A.4), any smooth prior contributes a penalty whose local increments are $O_p(T^{-1})$ against the $O_p(1)$ likelihood jumps: the profile-posterior mode is asymptotically equivalent to the maximum likelihood estimator, and Theorem 4 applies to it unchanged. Second, the asymptotic independence and rate separation of Theorem 4(iv) formalize the equivalence, asserted in Gargiulo et al. (2026), between conditioning the posterior of the regime parameters on the threshold estimate and on the threshold’s entire posterior. The decoupled model admits the same treatment with two independent flat priors on $(\gamma^\mu, \gamma^\sigma)$: since the local likelihood asymptotically factorizes across the two threshold blocks (Theorem 3; Remark 2), the joint posterior of the normalized thresholds converges to the product of the two marginal limit posteriors no matter the prior’s dependence structure: independent priors are asymptotically innocuous, marginal inference is asymptotically equivalent to joint, and, on the variance side, the limit posterior is a functional of the parametric jump law (V2) underlying Remark 5. We note that, with flat priors, posterior-mode estimation of the thresholds is maximum likelihood in substance; a decision-theoretic Bayesian treatment would instead use the posterior mean, which in non-regular models of this type is asymptotically efficient while the maximum likelihood estimator is not (Hirano and Porter, 2003; Chan and Kutoyants, 2012).

Remark 7 (Relation to the literature). Theorem 3 is the fixed-threshold-effect (compound Poisson) limit theory in the tradition of Chan (1993) and Samia and Chan (2011), extended to a multivariate Gaussian likelihood with regime-specific covariance matrices. Because the likelihood is maximized, the jump laws have negative mean and the walks drift to $-\infty$, the mirror image of the minimized SSR case in Chan (1993), where the jumps have positive mean and the walks drift to $+\infty$. Hansen (2000) obtains a different limit – the argmax of a two-sided Brownian motion with drift – under an asymptotic framework in which the threshold effect shrinks with the sample size. His framework yields asymptotically pivotal likelihood-ratio inference at the cost of the shrinking-effect device, while ours retains the fixed effect implied by Assumption B2 (Remark 4). Chan and Tsay (1998) show that if the regression function is instead continuous at the threshold, the rate drops to $\sqrt{T}$ with a Gaussian limit; the discontinuity guaranteed by the mean clause of B2 is thus essential for both the T -rate of Theorem 3 and the compound-Poisson form of Theorem 3.

2.5 Special Case: The Joint Model

The restriction. Let $q^\mu \equiv q^\sigma =: q$ with a common delay d , $k_\mu = k_\sigma =: k$, and $\gamma^\mu = \gamma^\sigma =: \gamma$. Every observation's mean regime then coincides with its variance regime, the regimes $\mathcal{R}_i = [\gamma_{i-1}, \gamma_i)$ are shared, and model (1) reduces to the *joint mean-variance* threshold VAR

$$y_t = \sum_{i=1}^k \left(\Phi_{i,0} + \sum_{j=1}^p \Phi_{i,j} y_{t-j} + \Sigma_i^{1/2} \varepsilon_t \right) \mathbb{I}(q_{t-d} \in \mathcal{R}_i), \quad (10)$$

with parameter $\theta = \{\gamma, \text{vec}(B_1)', \dots, \text{vec}(B_k)', \text{vech}(\Sigma_1)', \dots, \text{vech}(\Sigma_k)'\}$ and average log-likelihood

$$\ell_T(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^k \left(-\frac{1}{2} \ln \det \Sigma_i - \frac{1}{2} \varepsilon_{it}' \Sigma_i^{-1} \varepsilon_{it} \right) \mathbb{I}(q_{t-d} \in \mathcal{R}_i), \quad \varepsilon_{it} = y_t - (x_t' \otimes I_n) \text{vec}(B_i). \quad (11)$$

The profile is closed-form: with a common threshold each mean regime contains exactly one variance regime, so the GLS weights in (4) are constant within each mean regime and the feasible GLS estimator reduces to per-regime OLS, while per-regime sample covariances are the exact conditional maximum-likelihood estimators of the Σ_i ; the feasible and exact profiles of Section 2.2 therefore coincide, and the decoupled estimator nests the joint one exactly, at the level of both the criterion and the regime-parameter estimators, and the estimator is

$$\hat{\gamma} = \arg \max_{\gamma \in \Gamma} \ell_T(\gamma, \hat{B}(\gamma), \hat{\Sigma}(\gamma)), \quad \hat{\theta}_T = (\hat{\gamma}, \hat{B}(\hat{\gamma}), \hat{\Sigma}(\hat{\gamma})), \quad (12)$$

computed by grid search over Γ and the smallest maximizer in case of ties. We call it the *joint estimator*: it is the estimator of Gargiulo et al. (2026), and its threshold is identified from breaks in the conditional means *and* covariances jointly.

Nesting in estimation. The restriction need not be imposed: it can be selected.

Proposition 1 (Nesting). *Suppose the data are generated by (10) with $B_i^0 \neq B_{i+1}^0$ and $\Sigma_i^0 \neq \Sigma_{i+1}^0$ for all i , and let $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma)$ be the unrestricted estimator of Section 2.2 with the shared state q . Under **A1–A4**, **B1–B7** and **(F)**: (i) $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma) \rightarrow (\gamma^0, \gamma^0)$ a.s., with $T(\hat{\gamma}^\mu - \gamma^0) = O_p(1)$ and $T(\hat{\gamma}^\sigma - \gamma^0) = O_p(1)$; (ii) the regime-parameter estimates are $\sqrt{T}$ -asymptotically equivalent to those of the joint estimator; (iii) with $\hat{L}_T^F$ and $\hat{L}_T^J$ the maximized unrestricted and restricted criteria, $T(\hat{L}_T^F - \hat{L}_T^J) = O_p(1)$, whereas under distinct true thresholds on the shared state the same statistic diverges at rate T ; hence the rule “select the restriction if $T(\hat{L}_T^F - \hat{L}_T^J) < c_T$ ” is consistent for any $c_T \rightarrow \infty$ with $c_T = o(T)$; the simulations of*

Section 3.3 recommend $c_T = 2 \ln T$.

The proof is in Appendix A.3.3; it reuses the identification argument of Lemma 6 on the staircase of feasible cells, the localization of Theorem 3(i) with the three misclassification costs (mean only, variance only, both), and the $O_p(1)$ -versus-order- T gap for the selection step, the threshold analogue of the break-number selection logic of Bai and Perron (1998). Theorem 3(ii) does not apply on the diagonal: the two localizations share their boundary observations and **B8** fails, so the limit law of the unrestricted estimator there is a coupled two-dimensional argmax whose restriction to the diagonal is the law in Theorem 4 below. We do not pursue it, because the selection step directs the practitioner to the restricted estimator whenever the restriction holds. The recommended protocol is therefore: estimate unrestricted; apply the selection rule; if the restriction is selected, impose it and use Theorem 4 for inference.

Theory of the joint estimator. On the diagonal, **A4** applies to all k companion matrices; the cell clause of **B2** reduces to the positivity of each regime; **B5–B7** localize a single band around each γ_i^0 ; and **B8** is vacuous. Because the mean and variance jumps of a misclassified observation now occur together, the loss from evaluating an observation of regime $i+1$ under the law of regime i is the joint quantity

$$\xi_{i,t}^+ := m_i(y_t) - m_{i+1}(y_t) = \frac{1}{2} \ln \frac{\det \Sigma_{i+1}^0}{\det \Sigma_i^0} - \frac{1}{2} (\delta_{i,t} + \Sigma_{i+1}^{0,1/2} \varepsilon_t)' \Sigma_i^{0,-1} (\delta_{i,t} + \Sigma_{i+1}^{0,1/2} \varepsilon_t) + \frac{1}{2} \varepsilon_t' \varepsilon_t, \quad (\text{J1})$$

where m_i is the per-observation regime- i Gaussian log-likelihood at the true parameters and the expression for $\xi_{i,t}^-$ interchanges the regime labels; so that, conditionally on x_t , $\xi_{i,t}^+$ is an affine-plus-quadratic functional of ε_t with

$$E[\xi_{i,t}^+ | x_t] = -\text{KL}(\mathcal{N}(B_{i+1}^0 x_t, \Sigma_{i+1}^0) \parallel \mathcal{N}(B_i^0 x_t, \Sigma_i^0)) \leq 0, \quad (\text{J2})$$

where for multivariate Gaussians

$$\text{KL} = \frac{1}{2} \left[\text{tr}(\Sigma_i^{0,-1} \Sigma_{i+1}^0) - n - \ln \det(\Sigma_i^{0,-1} \Sigma_{i+1}^0) + \delta_{i,t}' \Sigma_i^{0,-1} \delta_{i,t} \right]. \quad (\text{J3})$$

Under **B2–B3** the divergence is strictly positive with positive probability at the threshold, so all jump means are strictly negative. With one-sided limit laws $\mu_i^\pm$ (well defined by **B6–B7**) and compound-Poisson processes J_i of rate $f(\gamma_i^0)$, f the stationary density of q_{t-d} , write $M_i := \arg\max_-(J_i)$, $S_i := E_\pi[x_t x_t' \mathbb{I}\{q_{t-d} \in \mathcal{R}_i^0\}]$ and $\pi_i := P_\pi(q_{t-d} \in \mathcal{R}_i^0)$. No counterpart of **(F)** is required, the profile being exact.

Theorem 4 (The joint model). *Consider model (10) and the joint estimator (12), with all assumptions restricted to the diagonal as described above. (i) Under **A1–A4**, $\{X_t\}$ is geometrically ergodic with finite fourth moments. (ii) Under **A1–A4** and **B1–B4**, $\hat{\theta}_T \rightarrow \theta_0$ a.s. (iii) Under **A1–A4** and **B1–B7**, $T(\hat{\gamma}_i - \gamma_i^0) = O_p(1)$ for each i , $\sqrt{T}(\hat{B}_i - B_i^0) = O_p(1)$ and $\sqrt{T} \text{vech}(\hat{\Sigma}_i - \Sigma_i^0) = O_p(1)$. (iv) Under the assumptions of (iii), jointly, the local processes converge weakly to the independent processes $(J_i)_{i=1}^{k-1}$, $(T(\hat{\gamma}_i - \gamma_i^0))_{i=1}^{k-1} \Rightarrow (M_i)_{i=1}^{k-1}$, and, asymptotically independently of them, $\sqrt{T} \text{vec}(\hat{B}_i - B_i^0) \Rightarrow \mathcal{N}(0, S_i^{-1} \otimes \Sigma_i^0)$ and $\sqrt{T} \text{vech}(\hat{\Sigma}_i - \Sigma_i^0) \Rightarrow \mathcal{N}(0, \frac{2}{\pi_i} D_n^+(\Sigma_i^0 \otimes \Sigma_i^0) D_n^{+'})$, with all Gaussian blocks mutually independent.*

Proofs are in Appendices A.1, A.3 and A.4; they are the single-threshold specializations of the arguments for Theorems 1–3, with the quartic drift of **A4** over all k matrices and the joint jump law in place of the separate mean and variance jumps.

Remark 8 (Identification through the second moment). The drift of each jump law is minus a conditional Kullback–Leibler divergence, (J2). Setting $\delta_{i,t} = 0$ in (J3) – identical conditional means – the divergence does not vanish; it collapses to the pure-covariance form

$$\text{KL} = \frac{1}{2} \sum_{j=1}^n (\lambda_j - \ln \lambda_j - 1), \quad \lambda_j := \text{eig}(\Sigma_i^{0,-1} \Sigma_{i+1}^0),$$

with each summand nonnegative and zero if and only if $\lambda_j = 1$; the sum is zero if and only if $\Sigma_i^0 = \Sigma_{i+1}^0$. Whenever the regime covariance matrices differ, the local process therefore has strictly negative drift and the threshold localizes at rate T with no contribution from the conditional mean. This is the population counterpart of the Monte Carlo design DGP3 in Section 3.2, where the SSR estimator of Chan (1993), whose local process is driven by $\delta_{i,t}$ alone, is inconsistent while the likelihood estimator retains the T -rate.

Remark 9 (Inference for the joint threshold). The limit law M_i is not pivotal, and no parametric shortcut of the kind available for variance thresholds exists, because the joint jump law depends on the conditional law of x_t at the threshold through $\delta_{i,t}$; the routes of Remark 4 apply.

3 Monte Carlo Evidence

This section provides Monte Carlo evidence in support of the theoretical results in Section 2. They show what the theory implies for practice: when second-moment information changes a threshold estimate, what is lost by imposing a single regime structure on a decoupled

process, and how the decoupled estimator behaves when the data are generated by the joint model. Then, they verify the asymptotic results themselves, by comparing finite-sample distributions with the limit laws of Theorems 3 and 4. Section 3.1 describes the designs, all of which are instances of (1); Sections 3.2–3.4 report the comparisons; Section 3.5 verifies the limiting distributions. Appendix B collects the full statistics behind every table.

3.1 Monte Carlo Design

All designs are bivariate ($n = 2$) first-order ($p = 1$) threshold VARs with two regimes per threshold, generated for $t = 1, \dots, T$ as

$$y_t = \Phi_{i(t),0} + \Phi_{i(t),1} y_{t-1} + \Sigma_{\ell(t)}^{1/2} \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, I_2), \quad \Phi_{1,1} = \Phi_{2,1} = 0.5 I_2, \quad (13)$$

where $i(t) = 1 + \mathbb{I}\{q_{t-1}^\mu \geq \gamma^{\mu,0}\}$ and $\ell(t) = 1 + \mathbb{I}\{q_{t-1}^\sigma \geq \gamma^{\sigma,0}\}$ are the mean and variance regimes of (1) with $d_\mu = d_\sigma = 1$, so that only the intercepts and the innovation covariances switch.⁸ Here $B_i = (\Phi_{i,0}, \Phi_{i,1})$ is 2×3 with $x_t = (1, y_{t-1}')'$. Each path starts at $y_0 = 0$ and discards 50 burn-in observations. Sample sizes are $T \in \{200, 1000\}$ with $S = 5,000$ replications per design unless stated otherwise, each replication with its own random stream.⁹ Performance is summarized by the root mean squared error (RMSE) of the thresholds about their true values and by the RMSE of $\hat{B}_i$ and $\hat{\Sigma}_\ell$ in Frobenius norm, averaged over regimes; classification accuracies and rate ratios for every design are in Appendix B.

Joint designs (DGP1–DGP4). A single self-exciting state drives both regimes, $q_t^\mu = q_t^\sigma = e_1' y_t$, with $\gamma^{\mu,0} = \gamma^{\sigma,0} = \gamma^0 = 0$, so that (13) is the joint model (10):

$$y_t = \begin{cases} \Phi_{1,0} + 0.5 y_{t-1} + \Sigma_1^{1/2} \varepsilon_t, & y_{1,t-1} < 0, \\ \Phi_{2,0} + 0.5 y_{t-1} + \Sigma_2^{1/2} \varepsilon_t, & y_{1,t-1} \geq 0, \end{cases} \quad (14)$$

⁸Holding $\Phi_{i,1}$ common across regimes satisfies the stability condition A4 with a trivial common Lyapunov matrix and isolates the threshold-identification problem from cross-regime stability. The implementation assigns the measure-zero event $q = \gamma$ to regime 1; with continuous states this is immaterial.

⁹Threshold candidates are 100 equally spaced points on the observed range of each state, followed by a local refinement (51 points for one threshold, 21×21 for a pair). The refinement matters at $T = 1000$, where the resolvable spacing of order T^{-1} falls below the coarse-grid resolution and the rate could not otherwise be detected. A candidate is feasible if every mean regime contains at least $nK + 1 = 7$ (with $K := np + 1$ regressors per equation) and every variance regime at least $n(n + 1)/2 + 1 = 4$ observations.

	$\Phi_{1,0}$	$\Phi_{2,0}$	Σ_1	Σ_2	Regime change in
DGP1	$(0, 0)'$	$(0, 0)'$	I_2	I_2	linear (no regime)
DGP2	$(1, 1)'$	$(-1, -1)'$	I_2	I_2	mean only
DGP3	$(0, 0)'$	$(0, 0)'$	I_2	$4 I_2$	variance only
DGP4	$(1, 1)'$	$(-1, -1)'$	I_2	$4 I_2$	mean and variance

DGP1 has no regime and serves as a check on the estimators' behavior when the threshold is unidentified; DGP2 is the mean-shift setting for which the SSR estimator is designed; DGP3 isolates a break in the second moments; DGP4 combines both.

Decoupled designs (DGP5–DGP6). Parameters as in DGP4, but the mean and variance regimes are driven by different states, with $\gamma^{\mu,0} = \gamma^{\sigma,0} = 0$:

$$\text{DGP5: } q_t^\mu = e_1' y_t, \quad q_t^\sigma = e_2' y_t; \quad \text{DGP6: } q_t^\mu = u_t, \quad q_t^\sigma = v_t, \quad (15)$$

with $u_t, v_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$, mutually independent and independent of $\{\varepsilon_t\}$. In DGP5 the two states are components of the system and weakly correlated through its dynamics; in DGP6 neither state carries information about the other, so each threshold must be identified without cross-leakage.

Limiting-distribution designs (DGP7–DGP8). Parameters as above except $\Sigma_2 = \begin{pmatrix} 4 & 0.8 \\ 0.8 & 2 \end{pmatrix}$, non-proportional to Σ_1 so that the variance-side theory is exercised in full-matrix form, and exogenous standard-normal states so that the Poisson rates and regime weights of Theorems 3 and 4 are known in closed form: DGP7 is the joint model on $q_t = u_t$ with $\gamma^0 = 0$; DGP8 is the decoupled model with $q_t^\mu = u_t$, $\gamma^{\mu,0} = 0$, and $q_t^\sigma = v_t$, $\gamma^{\sigma,0} = 0.3$. Here $S = 5,000$ at $T = 200$ and $S = 2,000$ at $T = 1000$, and the thresholds are computed by exact maximization over observed state values (the smallest maximizer, as in the theorems) rather than on the grid.¹⁰

3.2 Second-Moment Information and the Joint Estimator

Table 1 compares the SSR estimator with the likelihood-based joint estimator of Section 2.5 on DGP1–DGP4.

¹⁰The two-pass grid of Subsection 3.1 resolves the *rate* but, like any fixed grid, returns an interior point of the maximizing piece of the step-function criterion, displacing $T(\hat{\gamma} - \gamma^0)$ by up to one fine-grid spacing – roughly one unit at $T = 1000$, the same order as the object under study here. Exact maximization over observed state values removes the displacement and implements the smallest-maximizer convention of the theorems, whose limit is M_- .

Table 1: RMSE of the joint estimators on DGP1–DGP4

DGP	Estimator	Threshold $\hat{\gamma}$			$\hat{B}_i$		$\hat{\Sigma}_i$	
		$T = 200$	$T = 1000$	T -rate ratio	$T = 200$	$T = 1000$	$T = 200$	$T = 1000$
DGP1	SSR	–	–	–	1.758	2.259	0.360	0.263
	joint	–	–	–	1.908	2.590	0.495	0.426
DGP2	SSR	0.025	0.004	5.50	0.334	0.147	0.246	0.109
	joint	0.025	0.004	5.62	0.334	0.147	0.247	0.109
DGP3	SSR	2.004	3.204	0.63	2.902	4.138	1.756	1.629
	joint	0.080	0.013	6.01	0.454	0.191	0.760	0.337
DGP4	SSR	0.081	0.010	8.42	0.469	0.197	0.721	0.315
	joint	0.039	0.006	6.00	0.451	0.197	0.712	0.315

Notes: $S = 5,000$ replications, $\gamma^0 = 0$. SSR: least-squares threshold estimator of [Chan \(1993\)](#); joint: likelihood-based estimator of Section 2.5. Threshold RMSE about γ^0 ; ratio = RMSE($T=200$)/RMSE($T=1000$), equal to 5 under the rate T . Parameter RMSE: root of the mean over replications of the squared Frobenius error, averaged over regimes; $\hat{B}_i$ and $\hat{\Sigma}_i$ are the per-regime OLS and residual-covariance estimates at the estimated threshold. DGP1 has no true threshold.

The headline finding is the DGP3 row. In a DGP where regime identification must come entirely from the second moment, the SSR estimator does not converge: its threshold RMSE increases from 2.00 at $T = 200$ to 3.20 at $T = 1000$, and the empirical T-rate ratio of 0.63 is well below unity. The likelihood-based estimator, by contrast, achieves an RMSE of 0.013 at $T = 1000$ with a T-rate ratio of 6.01, consistent with the super-consistency claim of Theorem 4. The corresponding classification accuracy is 99.85% for the likelihood estimator versus 65.83% for SSR (Table A-3). Practical implication: when the candidate threshold variable plausibly drives volatility but not the conditional mean, the SSR-based threshold estimator is systematically misleading at any sample size; the likelihood-based estimator handles either source of identification.

The parameter columns add what the threshold columns cannot show. On DGP3 the inconsistency of the SSR threshold propagates to the regime parameters: its RMSEs for B and Σ do not fall with T and behave like those of DGP1, where no regime exists. On DGP2, where identification is carried by the conditional mean, the two estimators are indistinguishable in every column, so incorporating the covariance information costs nothing on the classical mean-shift problem. On DGP4 both estimators converge at the rate T , and the likelihood estimator halves the threshold RMSE (0.039 against 0.081 at $T = 200$); the parameter RMSEs are nevertheless identical across the two estimators. This is Theorem 4(iii)–(iv) at work: because the threshold is super-consistent, its precision has no first-order effect on the $\sqrt{T}$ -asymptotics of $\hat{B}_i$ and $\hat{\Sigma}_i$, so a better threshold estimator buys sharper regimes, not sharper regime parameters. The halving itself is the identification gain of Remark 8 made

visible: the likelihood criterion accumulates the mean break and the variance break at the same location, so its curvature around γ^0 is the sum of two contributions where the SSR criterion has one, and the threshold is located correspondingly more sharply.¹¹

DGP1 deserves a comment, because its parameter RMSEs are large and do not fall with T even though no regime exists. With a *fixed* partition this could not happen: the two sub-samples would come from the same process and per-regime OLS would estimate the common parameters consistently (on the same samples, a split at the sample median of the state gives RMSEs that fall at exactly the $\sqrt{T}$ rate, see Table A-2). The estimated partition is not fixed, however. The threshold estimator selects the split at which the two sub-sample fits differ most, so under no threshold effect the regime estimates are the extreme of many spurious comparisons rather than noisy versions of the truth: the estimated gap between the two regimes' intercepts is 1.7–2.3, against a true value of zero and a median-split gap that shrinks at the $\sqrt{T}$ rate, and it rises with T . Because the spurious gain is largest when one regime is small, the selected split increasingly isolates a tail of the state's support – outside the central 80% of its distribution in 26–41% of replications at $T = 200$ and 49–63% at $T = 1000$, with the smaller regime's median share falling to 5–10% and its minimum pinned at the feasibility bound. The equal-weight RMSE of Table 1 gives that small regime half the weight, which is why it rises; weighted by observation share it falls with T , though more slowly than $\sqrt{T}$.¹² The practical lesson is the classical one: a test for the existence of a threshold (Hansen, 1996) should precede its estimation. We take the existence of regimes as given throughout, as the questions of the applications presuppose.

3.3 Nesting the Joint Model

Proposition 1 states that, run with shared state and threshold, the decoupled estimator recovers the joint model: both thresholds converge to γ^0 at rate T , the regime parameters are asymptotically those of the joint estimator, and the statistic $T(\hat{L}_T^F - \hat{L}_T^J)$ is $O_p(1)$ under the restriction while diverging at rate T when the two thresholds differ, so that comparing it with any $c_T \rightarrow \infty$, $c_T = o(T)$ selects consistently between the two models. Table 2 tests each clause: the decoupled estimator is run with $q_t^\mu = q_t^\sigma = e_1' y_t$ on DGP1–DGP4

¹¹The SSR rate ratio of 8.42 on DGP4 reflects a finite-sample bias, from the misallocation of within-regime second-moment information by the mean-only objective, that vanishes faster than $1/T$; the asymptotic rate is still T (MAD, bias and standard deviations for all designs are in Table A-1).

¹²The likelihood estimator, which also searches over spurious covariance differences, lands in the tails more often than SSR (41% against 26% at $T = 200$) and is affected slightly more; $\hat{\Sigma}_i$, which a mean-targeted split barely moves, is affected less than $\hat{B}_i$. The comparison with a fixed split cannot separate pure selection from the mechanical imprecision of fitting an intercept on a small tail sub-sample; both act in the same direction.

(the same samples as Table 1), and the selection rule is evaluated on DGP1–DGP6 with $c_T \in \{\ln T, 2 \ln T\}$.¹³

The rule needs a penalty sequence, and it compares the likelihood gain from freeing the second threshold with a cost c_T : too small a cost and the unrestricted model is chosen under the restriction whenever the free threshold fits noise; too large and the restriction is accepted when the thresholds genuinely differ. Proposition 1 shows that any $c_T \rightarrow \infty$ with $c_T = o(T)$ avoids both errors asymptotically, because the gain is $O_p(1)$ under the restriction and of order T otherwise – the logic behind information criteria for the number of structural breaks (Bai and Perron, 1998; Liu et al., 1997). A BIC-type penalty would charge $\frac{1}{2} \ln T$ per additional parameter. We take $\ln T$ as the baseline and $2 \ln T$ as the alternative because the gain from a free threshold is a supremum over candidate locations rather than a single χ_1^2 term, so it is systematically larger than a parameter count suggests (Hansen, 1996); the modified criterion of Liu et al. (1997) is heavier than BIC for the same reason.

DGP4, where both moments break, is the case covered by the proposition, and every implication holds: $\hat{\gamma}^\mu$ and $\hat{\gamma}^\sigma$ converge at rate T (ratios 5.2 and 5.7), the parameter RMSEs coincide with the joint estimator’s, and the restriction is selected in 99.2% of replications with $c_T = \ln T$ and in all of them with $c_T = 2 \ln T$. DGP2 and DGP3 show why the proposition requires both moments to break: the threshold attached to the moment that does not switch is unidentified and does not converge (RMSE 1.7–2.0 for $\hat{\gamma}^\sigma$ on DGP2, 2.0–2.9 for $\hat{\gamma}^\mu$ on DGP3), while the identified one converges at rate T . The selection rule nevertheless remains consistent there: the gap is $O_p(1)$, with quantiles that do not move between $T = 200$ and $T = 1000$, so the frequency of selecting the restriction rises only as fast as c_T – 0.80–0.95 with $c_T = \ln T$, 0.99–1.00 with $c_T = 2 \ln T$. On DGP5–DGP6 the restriction is rejected in every replication at either penalty: the gap exceeds 13.8 in every sample at $T = 200$ and grows linearly in T . We therefore recommend $c_T = 2 \ln T$; larger penalties begin to cost power at $T = 200$ ($3 \ln T$ selects the restriction in three of 5,000 DGP5 samples; Table A-8 reports the full penalty sensitivity).¹⁴

¹³With a shared state the off-diagonal joint cells are empty by construction – the boundary nature of the joint case that Proposition 1 and Theorem 4 address – so the estimator’s cell-count feasibility rule is relaxed for this experiment only. The unrestricted criterion is evaluated on the two-dimensional grid augmented with the restricted estimate $(\hat{\gamma}, \hat{\gamma})$, so that $\hat{L}_T^F \geq \hat{L}_T^J$ holds by construction; on DGP5–DGP6 the restricted criterion is the larger of the joint fits on q^μ and on q^σ .

¹⁴The penalty must exceed the BIC-type $\frac{1}{2} \ln T$ per parameter because the unrestricted fit maximizes over a free, possibly unidentified threshold, so the gap behaves like a sup-type statistic rather than a χ_1^2 ; this is also why DGP1, where both thresholds are unidentified, has the heaviest tail (0.87 and 0.95 at $2 \ln T$). When nothing breaks, either model is correct and the choice is immaterial.

Table 2: Decoupled estimator with a shared state on the joint designs: RMSE and selection rule

DGP	T	Thresholds (decoupled)		$\hat{B}_i$		$\hat{\Sigma}_\ell$		Restriction selected	
		$\hat{\gamma}^\mu$	$\hat{\gamma}^\sigma$	dec.	joint	dec.	joint	$c_T = \ln T$	$c_T = 2 \ln T$
DGP1	200	–	–	3.558	1.908	0.823	0.495	0.610	0.872
	1000	–	–	5.132	2.590	0.777	0.426	0.761	0.948
DGP2	200	0.024	1.712	0.339	0.334	0.795	0.247	0.813	0.990
	1000	0.005	1.959	0.148	0.147	0.637	0.109	0.946	0.998
DGP3	200	1.954	0.092	3.788	0.454	0.786	0.760	0.800	0.997
	1000	2.894	0.013	5.127	0.191	0.340	0.337	0.920	1.000
DGP4	200	0.050	0.078	0.464	0.451	0.720	0.712	0.992	1.000
	1000	0.010	0.014	0.198	0.197	0.315	0.315	0.999	1.000
DGP5	200	<i>distinct true thresholds (Table 3)</i>						0.000	0.000
	1000							0.000	0.000
DGP6	200							0.000	0.000
	1000							0.000	0.000

Notes: Decoupled estimator with $q_t^\mu = q_t^\sigma = e_1' y_t$ on DGP1–DGP4, next to the joint estimator on the same samples; $S = 5,000$. Threshold RMSE about $\gamma^0 = 0$ (undefined for DGP1). “Restriction selected” is the share of replications with $T(\hat{L}_T^F - \hat{L}_T^J) < c_T$, the statistic of Proposition 1(iii), with $\ln T = 5.30$ and 6.91 at $T = 200$ and 1000 . Monte Carlo standard errors of the frequencies are below 0.006 .

3.4 Validating the Decoupled Estimator

Table 3 evaluates the decoupled estimator on DGP5–DGP6 against three single-threshold alternatives: the joint estimator with q^μ as its state (correctly oriented along the mean dimension, misspecified along the variance dimension), the joint estimator with q^σ (the mirror image), and the SSR estimator on q^μ – the estimator a practitioner would use – with per-regime OLS for B_i and a single pooled $\hat{\Sigma}$. Full statistics are in Appendix B, Tables A-4–A-5.

Three patterns emerge. First, the decoupled estimator delivers the smallest RMSE in every column, including on each threshold dimension against the joint estimator that observes the correct state for that dimension (on DGP5 at $T = 200$, 0.040 against 0.092 for $\hat{\gamma}^\mu$). The mechanism is that the joint estimator’s threshold search is conducted against an inflated within-cell residual covariance: pooling observations from both true variance regimes within each q^μ -cell amplifies the noise against which the mean-shift signal is judged. The decoupled estimator obtains cleaner per-regime covariance estimates from the q^σ split, and a correspondingly sharper mean-threshold profile. Its rate ratios lie between 5.3 and 6.0 on all eight (DGP, dimension) pairs (Table A-4), consistent with Theorem 3(i), and all four parameter RMSEs decline at rate $\sqrt{T}$, as Theorem 3(iii) predicts. Second, the misspecified joint

estimators and the SSR estimator carry misspecification floors on the dimension they cannot see: the covariance of the joint estimator on q^μ – and the pooled $\hat{\Sigma}$ of SSR, which is essentially tied with it – converges to $(\Sigma_1 + \Sigma_2)/2 = 2.5 I_2$, at Frobenius distance $1.5\sqrt{2} \approx 2.12$ from either truth, against observed RMSEs of 2.13 at $T = 1000$; symmetrically, $\hat{B}$ of the joint estimator on q^σ settles at distance $\sqrt{2} \approx 1.41$ from truth, against 1.42. Third, on the mean dimension the SSR estimator is as good as the correctly oriented joint estimator (and slightly better at $T = 200$ on DGP5), but it estimates no variance threshold at all: when the variance regime is driven by a different state, mean-only estimation is not merely inefficient but silent on one of the two regime structures.

Table 3: RMSE of decoupled, joint and SSR estimators on DGP5–DGP6

DGP	Estimator	$T = 200$				$T = 1000$			
		$\hat{\gamma}^\mu$	$\hat{\gamma}^\sigma$	$\hat{B}_i$	$\hat{\Sigma}_\ell$	$\hat{\gamma}^\mu$	$\hat{\gamma}^\sigma$	$\hat{B}_i$	$\hat{\Sigma}_\ell$
DGP5	joint on q^μ	0.092	–	0.458	2.205	0.009	–	0.192	2.127
	joint on q^σ	–	0.138	1.616	1.167	–	0.022	1.545	0.883
	decoupled	0.040	0.103	0.379	0.828	0.008	0.018	0.162	0.363
	SSR on q^μ	0.051	–	0.447	2.165	0.008	–	0.192	2.127
DGP6	joint on q^μ	0.033	–	0.282	2.229	0.006	–	0.122	2.143
	joint on q^σ	–	0.094	1.459	2.139	–	0.013	1.422	2.029
	decoupled	0.027	0.052	0.228	0.832	0.005	0.009	0.097	0.363
	SSR on q^μ	0.032	–	0.282	2.176	0.006	–	0.122	2.132

Notes: $S = 5,000$; $\gamma^{\mu,0} = \gamma^{\sigma,0} = 0$. “Joint on q^μ ” (“on q^σ ”) imposes a single threshold on the mean (variance) state for both moments; “SSR on q^μ ” estimates a mean threshold by least squares with per-regime OLS for B_i and one pooled $\hat{\Sigma}$. $\hat{B}_i$ RMSE as in Table 1. $\hat{\Sigma}_\ell$ RMSE on a common definition for all rows: each observation is assigned the estimated covariance of the regime its estimator places it in, the squared error is the average over observations of $\|\hat{\Sigma}_{\ell(t)} - \Sigma_{\ell(t)}\|_F^2$ with $\ell(t)$ the true variance regime, and the RMSE is the root of its mean over replications.

3.5 Limiting Distributions

DGP7 and DGP8 check the asymptotic theory directly: the compound-Poisson argmax laws for the thresholds (Theorems 4(iv) and 3(ii)), the Gaussian limits and independence structure of the regime parameters (Theorems 4(iv) and 3(ii)–(iii)), and the plug-in inference for the variance threshold of Remark 5. The limit laws are simulated by drawing $R = 200,000$ realizations of the two-sided compound Poisson processes – arrival rates $\phi(\gamma^0)$ on each side, ϕ the standard normal density, jump laws (J1) for DGP7 and (V3), (V2) for the mean and variance sides of DGP8 – and recording the smallest maximizer M_- .¹⁵

¹⁵With system-internal threshold variables the same comparison requires estimating $f_\pi(\gamma^0)$ and the conditional regressor law from a long simulated path – the band-resampling construction of Li and Ling (2012);

Thresholds. For each of the three thresholds – $\hat{\gamma}$ in DGP7, $\hat{\gamma}^\mu$ and $\hat{\gamma}^\sigma$ in DGP8 – the finite-sample distribution of $T(\hat{\gamma} - \gamma^0)$ is indistinguishable from its limit at Monte Carlo precision. The Kolmogorov–Smirnov distances are 0.011, 0.009 and 0.031 at $T = 200$, against a 5% critical value of 0.019, and 0.024, 0.025 and 0.012 at $T = 1000$ ($S = 2,000$, critical value 0.030). Figure 1 overlays the distribution functions; Table A-6 reports quantiles.

Regime parameters and independence. Table A-7 checks the Gaussian limits and the independence structure. The empirical standard deviations of $\sqrt{T} \text{vec}(\hat{B}_1 - B_1)$ match the asymptotic values – $S_1^{-1} \otimes \Sigma_1$ for DGP7, the GLS information inverse for DGP8 – with ratios of 1.04 at $T = 200$ and 1.00 at $T = 1000$, uniformly across components. The largest correlation between a normalized threshold and any normalized slope falls to 0.03–0.05 at $T = 1000$; and the correlation between the two normalized thresholds of DGP8 is -0.007 at $T = 200$ and 0.013 at $T = 1000$. This is the distinctive implication of the four-half-line Poisson random measure in Theorem 3(ii).

Plug-in inference. Remark 5 proposes confidence intervals for $\gamma^{\sigma,0}$ by simulating the variance-side limit law. Two versions are evaluated: an oracle version, which simulates the law from the true Σ_1 and Σ_2 , and the feasible version a practitioner would use, which re-simulates it in each replication from that replication’s $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$. Empirical coverage at nominal 90% and 95% is 0.86 and 0.92 at $T = 200$ and 0.89 and 0.94 at $T = 1000$ – and it is the same for the two versions to within Monte Carlo error (Table A-7). Replacing the nuisance objects of the limit law by their estimates therefore costs nothing in finite samples: the super-consistency of the thresholds and the $\sqrt{T}$ -consistency of the covariances make the plug-in procedure of Remark 5 operational as stated.

4 Empirical Application: Data-driven Regimes in Ramey and Zubairy (2018)

We apply the estimators to the setting of RZ2018, who ask whether government spending multipliers are larger in times of slack or at the zero lower bound, using more than a century of U.S. data. Their answer is a null: under their preferred military-news identification and with thresholds imposed at 6.5% unemployment and a 0.25% T-bill rate, multipliers are 0.60 in the high-unemployment regime and 0.59 in the low-unemployment regime, with a p -value of 0.95 for the difference; differences appear only under Blanchard–Perotti identification, and

nothing in the theory changes. The exogenous design isolates the question of interest: whether the finite-sample distributions converge to the limits of Theorems 3 and 4.

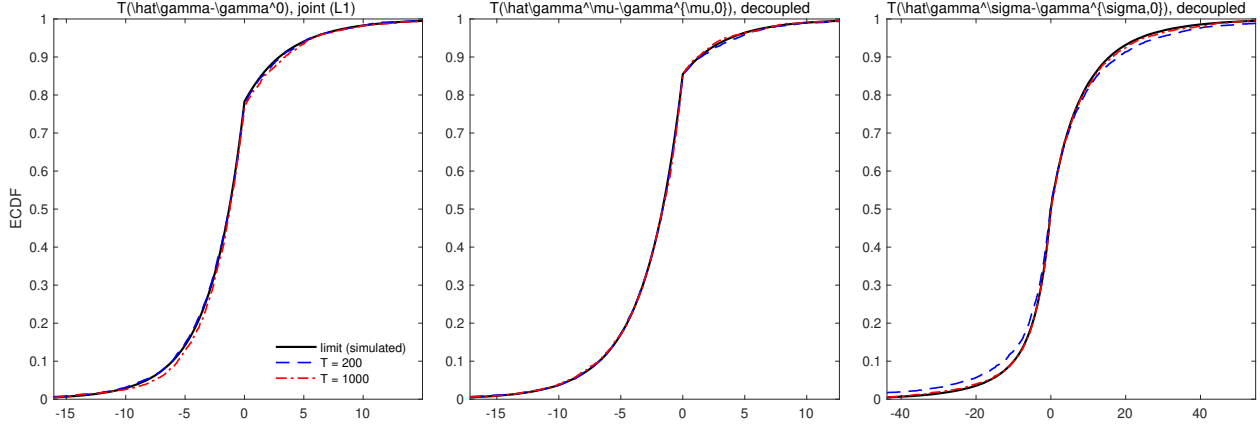


Figure 1: Empirical distribution functions of the normalized threshold estimators at $T = 200$ and $T = 1000$, overlaid on the simulated limiting distributions of Theorems 3 and 4 (solid). Left: joint model. Center and right: decoupled model, mean and variance thresholds. Note the wider scale and pronounced asymmetry of the variance-threshold law.

are driven by very low multipliers in the low-unemployment state.¹⁶ The question we ask is whether this conclusion survives when the regimes are estimated rather than imposed, and when the volatility of the identifying shocks is allowed to define a regime of its own.

4.1 Setup

The key variables are real output y_t and real government spending g_t , each normalized by a measure of potential GDP. RZ2018 employ a proxy identification scheme, where the instrument for government spending is the Ramey (2011) military news shock, normalized following the original specification. Since we differ from RZ2018 in that they use a local projection approach and the proposed estimator is built around VAR models¹⁷, we follow Plagborg-Møller and Wolf (2021) and set up the TVAR by including the proxy z_t as the first variable in the system $[z_t, y_t, g_t]'$, ensuring IRF equivalence between the internal-instrument TVAR and the LP-IV used by RZ2018.¹⁸ We set $p = 4$ lags throughout.

Two exercises are considered in the original paper:

- **Slack:** the state variable is the unemployment rate. RZ2018 impose a threshold of

¹⁶RZ2018, Table 1. At the zero lower bound they find differences on the sample excluding WWII under military-news identification, which we discuss in Appendix C.

¹⁷In a companion project, we show that the thresholds can be equivalently estimated in a regime-dependent local projection setup.

¹⁸Plagborg-Møller and Wolf (2021) show that local projections and VARs estimate the same impulse responses under the same identification. Ordering z_t first and using the Cholesky decomposition of the regime-specific covariance matrix recovers the proxy-SVAR IRF within each regime.

6.5% to capture slack in the economy.

- **ZLB**: the state variable is the interpolated three-month Treasury-bill rate. RZ2018 impose a threshold of 0.25%.

In both cases, regime assignment is based on the one-period lag of the state variable, so that it lies in the information set at the time the shock hits.

The slack exercise is estimated on three samples: (i) the full historical sample 1890Q1–2015Q4 ($T = 504$); (ii) the full sample excluding WWII rationing quarters 1941Q3–1945Q4 (labeled *Excl. WWII*); and (iii) the post-WWII sample 1947Q1–2015Q4 ($T = 276$, labeled *Post-WWII*). The ZLB exercise uses observations from 1915Q1 to 2015Q4 ($T = 404$), again with and without WWII. RZ2018’s headline result is a null: with the thresholds imposed at 6.5% unemployment and at a 0.25% T-bill rate, they find no significant differences in fiscal multipliers across regimes in either exercise. Our question is whether this null result is informative about fiscal transmission or is instead a consequence of regime misspecification: of the location of the imposed thresholds, of the single-state-variable regime structure, or of both. We therefore let the data choose the regimes. For each sample and state variable we first estimate the threshold in the joint mean-variance TVAR of Section 2.5, and we then re-estimate the model with the decoupled estimator of Section 2.2, retaining RZ2018’s economic state for the mean regimes while introducing the size of the identifying shock instrument as the variance-regime state. We estimate thresholds on all the subsamples considered in the original paper under both state variable specifications. The slack exercise is the focus of this section; the interest-rate exercise, in which the free estimates recover RZ2018’s 0.25% cut-off on the full sample once the shock environment is isolated, is reported in Appendix C and gives the same qualitative picture.

4.2 Joint Estimation and the Instrument Diagnostic

We begin with the joint estimator, in which a single threshold on the state variable governs both the dynamics and the innovation covariance – the model closest to the single-regime-structure practice of the literature. Table 4 reports the estimates next to two benchmarks: the thresholds RZ2018 impose, and the threshold a researcher would obtain with the standard least-squares (SSR) criterion, which uses the conditional mean only.

The comparison is instructive. On the full sample the SSR estimate for slack is 6.95% – essentially RZ2018’s 6.5% – and for the interest-rate exercise it is 0.21%, essentially their 0.25%: a mean-only estimator confirms the imposed thresholds on the full sample. The

Table 4: Joint threshold estimates: likelihood, least squares (SSR), and the imposed values

Sample	Joint likelihood		SSR (mean only)		Imposed (RZ2018)	
	$\hat{\gamma}$	Regime 1 obs. (%)	$\hat{\gamma}$	Regime 1 obs. (%)	γ	Regime 1 obs. (%)
<i>Panel A: Slack (unemployment rate, %; Regime 1 = low unemployment)</i>						
Full sample	3.56	57 (11.4)	6.95	343 (68.6)	6.50	319 (63.8)
Excl. WWII	3.64 [†]	49 (10.3)	4.20	105 (22.0)	6.50	300 (62.8)
Post-WWII	5.61	139 (51.1)	5.61	139 (51.1)	6.50	189 (69.5)
<i>Panel B: ZLB (T-bill rate, %; Regime 1 = low rate)</i>						
Full sample	3.48	223 (55.8)	0.21 [‡]	54 (13.5)	0.25	58 (14.5)
Excl. WWII	3.48	201 (53.2)	1.22	94 (24.9)	0.25	55 (14.6)

Notes: Thresholds estimated on a 300-point grid over the observed range of the state, with each regime required to contain more than $nK + n(n+1)/2 + 1 = 46$ ($K = np + 1$) observations. “Joint likelihood”: the estimator of Section 2.5, which uses both moments. “SSR”: the least-squares criterion of Chan (1993), which uses the conditional mean only, on the same grid, sample and trimming. Regime 1 is the below-threshold regime. The post-WWII ZLB exercise is infeasible (too few low-rate observations).

[†]Smallest grid point satisfying the minimum-size requirement (Regime 1 has 49 observations); the likelihood is still rising as the regime shrinks. [‡]Lowest feasible grid point.

likelihood estimator departs from them because it also sees the second moments. It places the slack threshold at 3.56%, isolating 57 quarters of wartime near-full employment and peacetime booms, and the interest-rate threshold at 3.48%, a broad low-rate environment rather than a zero lower bound; on the sample excluding WWII the slack estimate is driven to the smallest regime the trimming allows. On the post-WWII sample, where the wartime episodes are absent, the two criteria coincide at 5.61%. The departure of the joint estimate from the imposed threshold is therefore attributable entirely to variance information, which raises the question of what that information is.

Figure 2 answers it. It reports the residual standard deviation of each equation of the TVAR in the two regimes selected by the joint estimator on the full sample. The break is in government spending: the standard deviation of the spending-equation residual is 1.66 in the low-unemployment regime against 0.65 in the high-unemployment regime, a ratio of 2.6. Output’s is modestly larger (1.92 against 1.70), and the news equation’s own residual is not (5.55 against 5.35) – the largest news realizations of the sample, in 1941Q3–1942Q1, occur when unemployment is still above the threshold. The joint likelihood thus locates the threshold where the volatility of fiscal innovations changes, in the wartime spending surges, rather than where the propagation of shocks changes. The instrument’s own variation concentrates in the same quarters, with a lag (Figure 3: the low-unemployment regime holds 22% of the instrument’s absolute mass with 11% of the observations, and 20% of its top-decile realizations). The lag is the anticipation built into the instrument: the announcements arrive while

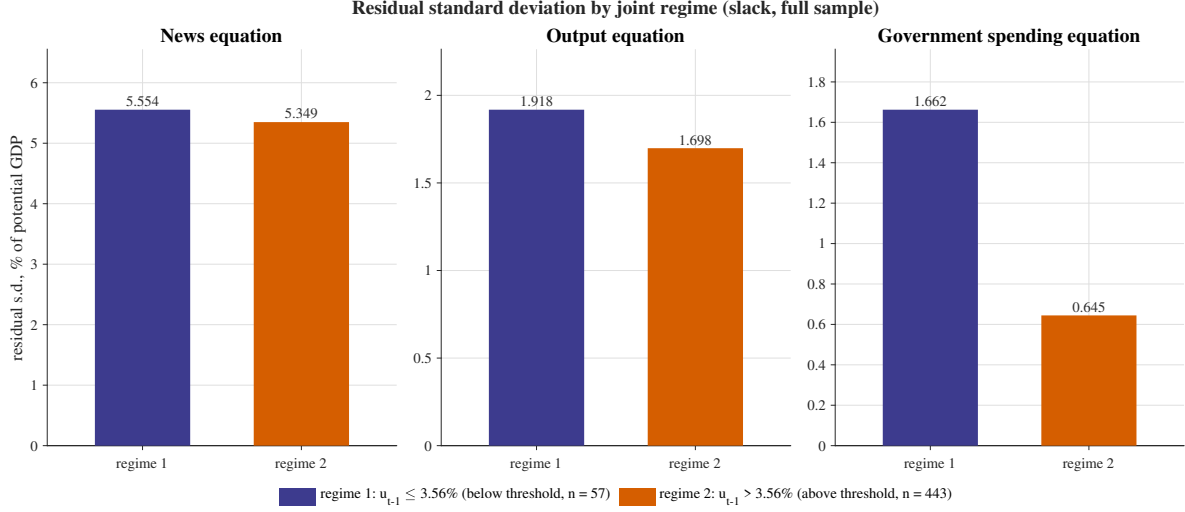


Figure 2: Residual standard deviation by joint regime: slack exercise, full sample ($\hat{\gamma} = 3.56\%$). Each panel reports the standard deviation of one equation’s residual in the low-unemployment regime ($u_{t-1} \leq 3.56\%$, 57 quarters) and in the high-unemployment regime (443 quarters), from the joint TVAR estimated at the stored threshold. Units: percent of potential GDP.

unemployment is still high, and the mobilization that follows them pushes unemployment below the threshold in the quarters in which the announced spending is executed, which is where the spending innovations are volatile. The regime indicator and the lagged instrument therefore carry much the same information, and the “state regime” is a proxy for a shock regime – specifically, for the wartime spending surge that the news announces. It is not a business-cycle regime: its 57 quarters pool wartime near-full employment with the peacetime booms of the 1900s and 1920s, wars and booms alike, not a recession–expansion partition.¹⁹ Under the SSR split of 6.95%, as under RZ2018’s own 6.5%, where 73% of the fiscal-event quarters fall in a regime with 64% of the observations, the concentration is proportionate to regime size. The pathology is therefore specific to the variance-driven split: a variance regime is being forced onto a single mean-regime partition. We therefore let the two dimensions have their own thresholds through the decoupled estimator.

For this reason we do not report multiplier estimates from the joint specification: under this degree of concentration, regime-specific multipliers would measure the particular wartime episodes rather than the economic states.

¹⁹Wartime near-full employment in 1943–46, 1951–53 and 1968–69; peacetime booms in 1902–07 and 1926–29.

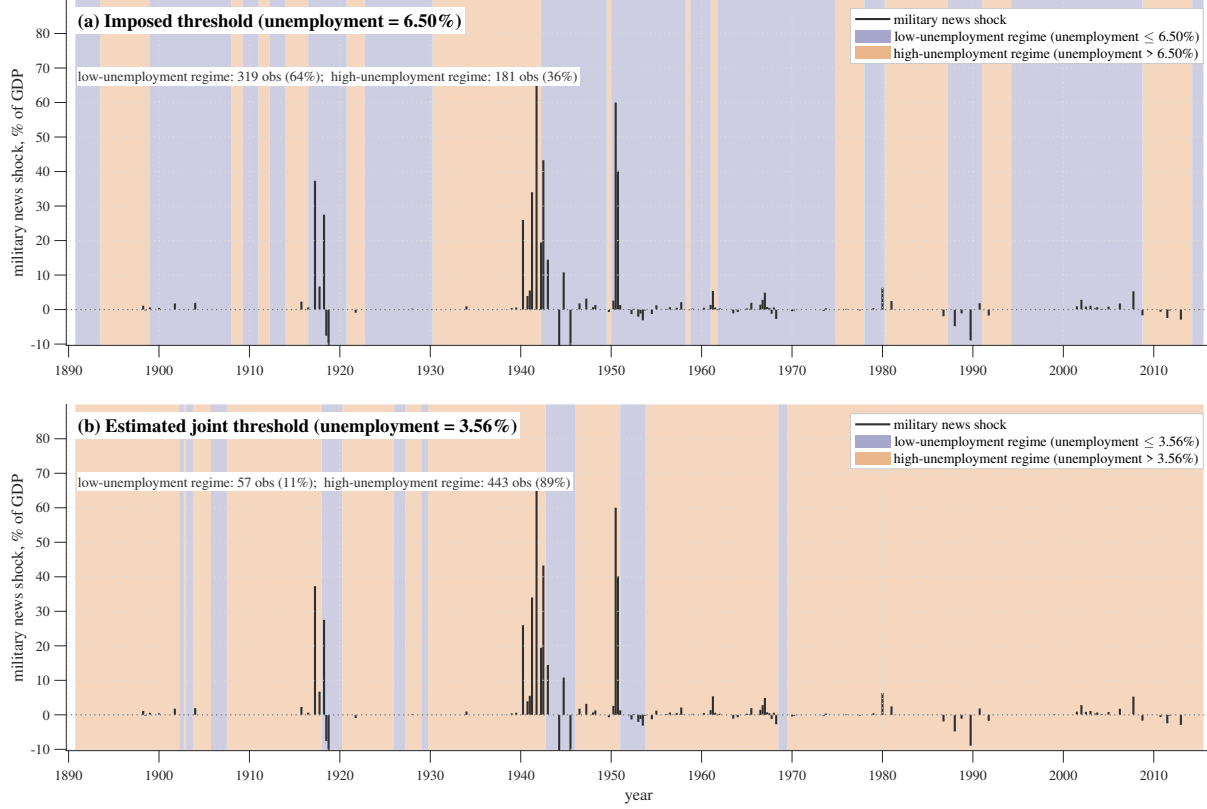


Figure 3: Unemployment regimes and the military news instrument, full sample. *Top*: the instrument colored by the imposed assignment at 6.5%; *bottom*: by the estimated joint assignment at 3.56%. The low-unemployment regime is shaded violet, the high-unemployment regime orange.

4.3 Decoupled Estimation: Separating the Economic State from the Shock Environment

The decoupled estimator keeps the unemployment rate as the mean state and introduces a second state for the variance regimes: the absolute value of the lagged instrument, $|z_{t-1}|$. The variance threshold is estimated jointly with the mean threshold; it is not imposed. The choice of the news series as the variance state, despite Figure 2 locating the variance break in the spending equation, lies in the interest of identifying a *shock* regime – the environment in which the identifying shocks arrive – and the news series is the identifying shock, whereas spending is the system’s *response* to it. A variance regime defined on spending, or on its innovation, would classify quarters by the consequence rather than the cause, conflating the two dimensions the model separates. Timing makes the point concrete: announcements lead spending by quarters or years, the premise of the news identification, so a spending-

based regime dates the environment after the identifying variation has entered the system. The absolute lagged news series isolates the arrival of large identified shocks, clustered into named episodes, and is known at $t - 1$ like the mean state. Both alternatives are nonetheless examined: the spending residual in Section 4.7, where it selects a dispersion split with no event structure. Lagged spending growth (Appendix H) behaves as the argument predicts: it isolates the build-ups and demobilizations of the three major wars together with pre-1940 peacetime spending volatility, misses the announcement quarters and every post-1968 fiscal event, and in two of the three samples produces boundary partitions whose cells are too thin for inference – indicating a state that classifies by consequence rather than cause.

Two features of the resulting partition matter for everything that follows. First, the variance threshold is remarkably stable, $\hat{\gamma}^\sigma \approx 0.024$ in every sample and for both state variables, and it isolates a small *fiscal-event regime* (VR2) of large military-news quarters – on the full sample 33 quarters comprising the two World Wars (1917–19, 1940–45), the Korean mobilization (1950–51, 1953Q4), the 1961 Berlin crisis, the Vietnam build-up (1967–68), the Reagan build-up (1980–81), the late Cold War and Gulf War announcements (1988, 1990), the post-9/11 build-up (2002), and the 2008–2013 fiscal responses. Everything else forms the *normal-shock regime* (VR1). Second, once that regime is isolated, the mean threshold moves to where the propagation break lies rather than where the volatility break lies (Table 5).

Table 5: Decoupled TVAR: threshold estimates and four-cell observation counts

Sample	Thresholds		Low unemployment		High unemployment		z share (%)
	$\hat{\gamma}^\mu$ (%)	$\hat{\gamma}^\sigma$	normal	fiscal-event [†]	normal	fiscal-event [†]	
Full sample	8.42	0.0243	382 (76.6)	28 (5.6)	84 (16.8)	5 (1.0)	36.8
Excl. WWII	3.44	0.0241	37 (7.8)	4 (0.8)	415 (87.0)	21 (4.4)	44.7
Post-WWII	4.23	0.0211	47 (17.3)	5 (1.8)	208 (76.5)	12 (4.4)	53.2
Imposed (RZ2018)	6.50	—					—

Notes: Slack exercise. $\hat{\gamma}^\mu$: mean threshold on the lagged unemployment rate, defining the low-unemployment ($u_{t-1} \leq \hat{\gamma}^\mu$, MR1) and high-unemployment (MR2) regimes. $\hat{\gamma}^\sigma$: variance threshold on $|z_{t-1}|$, the absolute lagged military news shock of Ramey (2011), defining the normal-shock (VR1) and fiscal-event (VR2) regimes. Cell entries: observations (% of the kept sample). “|z| share”: fraction of $\sum_t |z_t|$ in the fiscal-event regime. $\hat{B}_i$ by feasible GLS with variance-regime precision weights; $\hat{\Sigma}_\ell$ pooled within variance regime. [†]Too small for inference; not used below. The interest-rate exercise is in Appendix C.

The mean threshold now varies across samples, and the variation is informative. On the full sample it rises to 8.4%: the high-unemployment regime is an *extreme-slack* regime of 89 quarters, roughly three quarters of them pre-1947 (the depressions of the 1890s, 1914–15, 1921–22 and the Great Depression of 1930–41), joined by the deepest post-war reces-

sions (1975, 1982–84, 2009–12). On the post-WWII sample it falls to 4.2%, a conventional business-cycle cut-off below RZ2018’s 6.5%. On the sample excluding WWII it is 3.4%, isolating a small low-unemployment regime of 41 quarters; we read this estimate with caution, as it sits close to the trimming boundary, and the imposed-threshold exercise of Section 4.5 does not depend on it.²⁰

4.4 State-Dependent Multipliers in the Normal-Shock Regime

An important element of the analysis that follows is the identification of impulse responses in the decoupled TVAR. Under internal-proxy identification the $h = 0$ response is the first column of the Cholesky factor of $\hat{\Sigma}_\ell$. Since this depends only on the variance regime, the impact of a fiscal shock is the same across mean regimes within a variance regime, while its propagation – the response at $h \geq 1$ – is governed by the mean-regime coefficients $\hat{B}_i$, which differ across mean regimes but not across variance regimes. This separation motivates the terminology: the variance threshold identifies the impact channel, the mean threshold the propagation channel. We report cumulative multipliers, the ratio of the cumulative output response to the cumulative spending response, because any normalization of the shock rescales both responses together and leaves the ratio unchanged. What the ratio does not remove is the impact *direction*, the relative response of output and spending, which differs across variance regimes but is common across mean regimes. Accordingly, the relevant comparison for state-dependent transmission is across mean regimes within a fixed variance regime, where any difference at $h > 0$ is propagation; cells in different variance regimes are not compared.

Inference on the multipliers is by a fixed-design wild bootstrap. Residuals and fitted values come from the regime-specific regressions at the estimated thresholds; each pseudo-sample multiplies the residual vector of quarter t by a standard normal draw η_t , the same draw for the three equations, so that the cross-equation correlation and the regime-specific covariances of the residuals are preserved, and holds the regressors fixed (Gonçalves and Kilian, 2004). On each of $B = 499$ pseudo-samples the two thresholds are re-estimated on a coarser grid and the cells, coefficients, covariances and multipliers are recomputed; the bands are percentile intervals. The re-estimation of the thresholds inside the bootstrap is not the object of the exercise – Remark 4 explains why the bootstrap is not a valid route to the thresholds

²⁰The fiscal-event regime holds 37–53% of the instrument’s absolute mass with 6–7% of the observations, which is what isolating the large shocks is meant to achieve. The fiscal-event cells are reported in Appendix G for completeness only.

themselves – but it is harmless for the multipliers: by Theorem 3(i) the threshold draws vary at rate T , below the $\sqrt{T}$ scale of the bands, so the intervals are those of the regime parameters conditional on the regime structure, which is what we need. In the imposed-threshold exercise of Section 4.5 the threshold pair is held fixed and only the regime parameters are re-drawn. Appendix F checks the bands against a posterior computed under a flat prior and finds them of the same width in the well-populated cells.

Table 6 reports the cumulative multipliers in the two normal-shock cells, and Figure 4 plots them together with the point estimates obtained at the imposed 6.5% threshold, to be discussed in Section 4.5. We complement the cell-level bands with a direct test of the across-regime difference: within each bootstrap replication we compute $\Delta(h) = \hat{m}_{\text{high-}u}(h) - \hat{m}_{\text{low-}u}(h)$ and report the 90% percentile interval of its bootstrap distribution (the Δ rows). Because both cells are re-estimated from the same draw and share the normal-shock covariance $\hat{\Sigma}_1$, the two estimates are positively correlated and the interval for the difference is tighter than a comparison of the individual bands suggests; an interval excluding zero indicates a significant difference at the 10% level, paralleling the across-state tests of RZ2018 themselves.

The results show a sample-dependent direction. On the historical samples, multipliers are larger in the high-unemployment regime: on the full sample 1.24 against 0.44 at $h = 3$ and 0.98 against 0.45 at $h = 19$, the ranking of Auerbach and Gorodnichenko (2012), driven by the pre-1947 extreme-slack variation documented above. On the post-WWII sample the ranking reverses: 3.45 rising to 4.47 in the low-unemployment regime against 1.56 falling to 0.72 in the high-unemployment regime – larger multipliers in good times, the ranking of Alloza (2022). The post-WWII ranking is a point-estimate result in this exercise: the 90% interval for the difference includes zero at every horizon (Δ rows), reflecting the width of the low-unemployment bands in a 47-quarter cell. Its statistically significant counterpart is the robustness exercise of Section 4.7, where the same ranking is recovered under a different identification, with the difference significant through $h = 11$.

The contrast with RZ2018 is sharp. At their imposed 6.5% threshold and under their preferred identification, they find 0.60 in the high-unemployment regime and 0.59 in the low-unemployment regime, with a p -value of 0.95 for the difference. We find that the ranking of the two regimes is not fixed but depends on the sample – larger in the high-unemployment regime on the historical samples, larger in the low-unemployment regime after WWII – and, as the next subsection shows, that significant differences appear even at their own threshold once the fiscal-event regime is isolated.

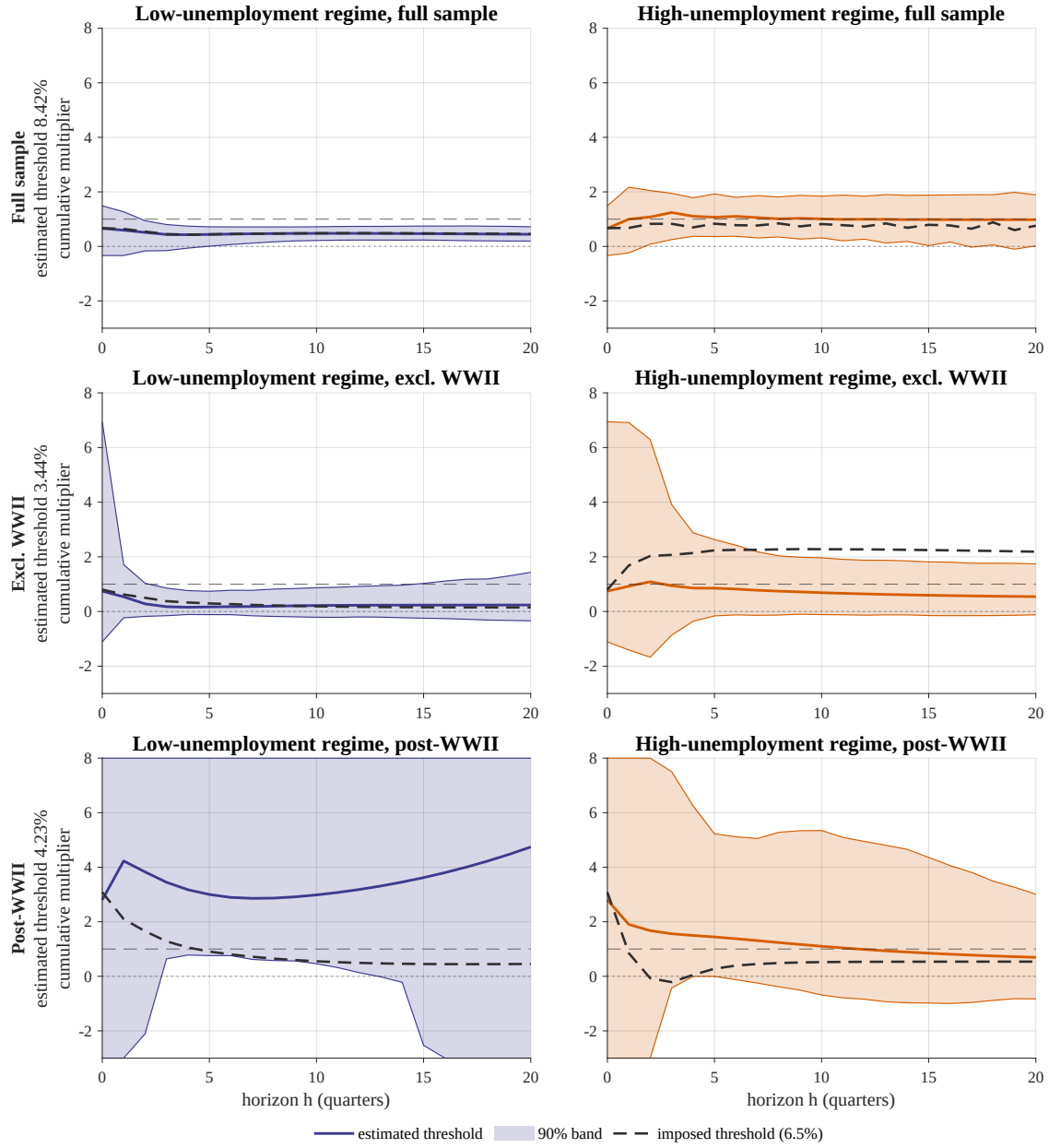


Figure 4: Cumulative multipliers in the normal-shock regime, slack exercise. Rows: full sample, excluding WWII, post-WWII; columns: low-unemployment and high-unemployment regimes at the estimated threshold, with 90% wild-bootstrap bands ($B = 499$). The dashed black line is the point estimate at the imposed 6.5% threshold (Section 4.5). Bands are truncated at the display range.

Table 6: Cumulative multipliers in the normal-shock regime at the estimated thresholds

Sample	Regime	Horizon h (quarters)			
		$h = 3$	$h = 7$	$h = 11$	$h = 19$
Full sample	low unemployment	0.44 [-0.15, 0.80]	0.46 [0.12, 0.71]	0.48 [0.23, 0.73]	0.45 [0.20, 0.74]
	high unemployment	1.24 [0.25, 1.95]	1.06 [0.31, 1.86]	0.99 [0.21, 1.88]	0.98 [-0.10, 1.98]
	Δ (high – low)	0.80 [0.01, 1.41]	0.59 [-0.14, 1.45]	0.51 [-0.33, 1.47]	0.53 [-0.69, 1.62]
Excl. WWII	low unemployment	0.18 [-0.15, 0.86]	0.18 [-0.16, 0.77]	0.23 [-0.21, 0.89]	0.24 [-0.33, 1.30]
	high unemployment	0.94 [-0.87, 3.91]	0.78 [-0.14, 2.19]	0.67 [-0.12, 1.91]	0.55 [-0.14, 1.76]
	Δ (high – low)	0.77 [-1.01, 3.35]	0.59 [-0.38, 1.76]	0.44 [-0.50, 1.65]	0.31 [-0.90, 1.61]
Post-WWII	low unemployment	3.45 [0.64, 11.24]	2.86 [0.62, 8.11]	3.08 [0.32, 10.71]	4.47 [-11.41, 25.88]
	high unemployment	1.56 [-0.43, 7.51]	1.31 [-0.25, 5.05]	1.04 [-0.79, 5.09]	0.72 [-0.82, 3.27]
	Δ (high – low)	-1.89 [-8.34, 3.86]	-1.55 [-6.11, 2.99]	-2.03 [-9.07, 4.14]	-3.75 [-25.59, 13.77]

Notes: Slack exercise, normal-shock regime ($|z_{t-1}| \leq \hat{\gamma}^\sigma$), mean regimes at the estimated $\hat{\gamma}^\mu$ of Table 5. Point estimates with 90% wild-bootstrap intervals below ($B = 499$). The multiplier at horizon h is $\sum_{s=0}^h \hat{\phi}_{Y,s} / \sum_{s=0}^h \hat{\phi}_{G,s}$, the ratio of cumulative responses of output and government spending to the identified news shock. The $h = 0$ multiplier is identical across mean regimes within the normal-shock regime (common $\hat{\Sigma}_1$, hence common impact vector). Δ rows: difference across mean regimes, with intervals computed from the same bootstrap draws as the cell bands, so the correlation between cells is accounted for; an interval excluding zero indicates a significant difference at the 10% level. Two cells call for caution at long horizons. In the post-WWII low-unemployment cell (47 quarters) the denominator changes sign in 9% of draws at $h = 19$, which is what widens that band; in the full-sample high-unemployment cell (84 quarters) the estimated within-regime dynamics are explosive (largest root 1.58), so its multipliers are informative at short horizons only (Appendix F).

4.5 State Dependence at the Imposed Thresholds

Two margins separate our estimates from RZ2018’s: the location of the mean threshold, and the isolation of the fiscal-event regime. To tell them apart we re-estimate the model with the mean threshold *imposed* at RZ2018’s 6.5% and only the variance threshold estimated (at the values of Table 5). Relative to RZ2018 this keeps their threshold and changes only the treatment of the shock environment; relative to our free estimates it changes only the location of the mean threshold. Inference is conditional on the fixed threshold pair, with only the regime parameters re-drawn in the wild bootstrap ($B = 499$).

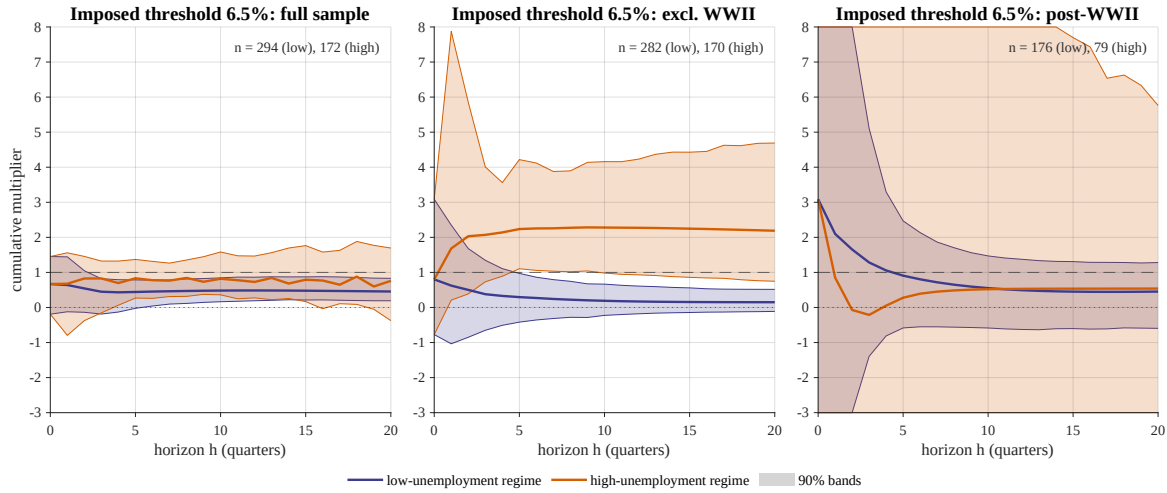


Figure 5: Cumulative multipliers in the normal-shock regime at the imposed 6.5% threshold, by sample. Each panel overlays the low-unemployment (violet) and high-unemployment (orange) regimes with 90% bands conditional on the fixed threshold pair.

Figure 5 and Table 7 give three answers, one per sample. On the sample excluding WWII, the two regimes separate cleanly at RZ2018’s own threshold: 2.07–2.28 in the high-unemployment regime against 0.15–0.38 in the low-unemployment regime, with the 90% interval for the difference excluding zero at every horizon ($\Delta = 1.7$ –2.1)²¹. On the full sample the ranking is the same and the gap smaller (0.83 against 0.45 at $h = 3$) and not significant; the three quarters of the U.S. entry into WWII, 1941Q3–1942Q1, are extreme leverage points inside the high-unemployment regime, and their inclusion is what separates the two historical samples (Ap-

²¹With the usual caveat for this cell. Once WWII is excluded, the high-unemployment quarters at 6.5% contain almost no news variation outside the fiscal-event regime (first-stage $F < 0.1$, Appendix E), exactly as in RZ2018’s own high-unemployment state on this sample. The impact vector of the shock is estimated from the whole normal-shock regime, but the propagation coefficients of the high-unemployment regime are estimated from little identifying variation, and we read the multiplier there with the same caution RZ2018 apply to theirs.

Table 7: Imposed-threshold exercise: normal-shock multipliers at RZ2018’s 6.5%

Sample	Regime	Horizon h (quarters)			
		$h = 3$	$h = 7$	$h = 11$	$h = 19$
Full sample	low unemployment	0.45 [-0.19, 0.82]	0.46 [0.10, 0.78]	0.48 [0.18, 0.87]	0.46 [0.19, 0.84]
	high unemployment	0.83 [-0.15, 1.32]	0.77 [0.31, 1.26]	0.78 [0.25, 1.47]	0.60 [-0.05, 1.77]
	Δ (high – low)	0.38 [-0.29, 0.76]	0.30 [-0.16, 0.81]	0.29 [-0.31, 0.99]	0.14 [-0.63, 1.31]
	low unemployment	0.38 [-0.65, 1.35]	0.24 [-0.31, 0.79]	0.18 [-0.20, 0.63]	0.15 [-0.12, 0.52]
	high unemployment	2.07 [0.73, 4.01]	2.26 [1.03, 3.88]	2.28 [0.95, 4.16]	2.20 [0.76, 4.68]
	Δ (high – low)	1.69 [0.77, 3.37]	2.01 [0.86, 3.56]	2.10 [0.75, 3.95]	2.05 [0.59, 4.53]
Excl. WWII	low unemployment	1.28 [-1.39, 5.10]	0.72 [-0.55, 1.87]	0.52 [-0.61, 1.41]	0.45 [-0.59, 1.27]
	high unemployment [§]	-0.21 [-23.51, 21.58]	0.45 [-13.41, 17.65]	0.53 [-8.62, 13.18]	0.54 [-3.85, 6.34]
	Δ (high – low)	-1.50 [-24.49, 20.93]	-0.27 [-14.38, 17.83]	0.01 [-9.14, 13.10]	0.09 [-4.34, 5.62]
Post-WWII	low unemployment	1.28 [-1.39, 5.10]	0.72 [-0.55, 1.87]	0.52 [-0.61, 1.41]	0.45 [-0.59, 1.27]
	high unemployment [§]	-0.21 [-23.51, 21.58]	0.45 [-13.41, 17.65]	0.53 [-8.62, 13.18]	0.54 [-3.85, 6.34]
	Δ (high – low)	-1.50 [-24.49, 20.93]	-0.27 [-14.38, 17.83]	0.01 [-9.14, 13.10]	0.09 [-4.34, 5.62]

Notes: Mean threshold imposed at 6.5%; variance threshold at the estimated value of Table 5; normal-shock regime only. 90% wild-bootstrap intervals conditional on the fixed threshold pair ($B = 499$, all draws valid). Normal-shock cell counts (low / high unemployment): 294 / 172 (full), 282 / 170 (excl. WWII), 176 / 79 (post-WWII). Δ rows as in Table 6. [§]The wide intervals reflect bootstrap draws in which the cumulative spending response, the denominator of the multiplier, changes sign: 23% of draws at $h = 3$ and 13–14% at $h \geq 7$ in this 79-quarter cell, against at most 1.6% in every other cell of the table. They are a feature of the ratio in a small cell with a weak first stage (Appendix E), not evidence of unstable dynamics, and the posterior bands of Appendix F reproduce them.

pendix D). On the post-WWII sample, no ranking emerges at 6.5%: the high-unemployment regime at that cut-off contains 79 normal-shock quarters, its bands are wide, and the difference is indistinguishable from zero; the larger multipliers in the low-unemployment regime of Table 6 appear only at the estimated 4.2%.

Overall, the exercise shows that imposing an exogenous threshold has first-order consequences for the conclusions, and that the consequences differ by sample. On the historical samples the imposed threshold is not the problem: holding it at 6.5%, the ranking appears on both samples as soon as the fiscal-event regime is isolated, and on the sample excluding WWII the difference between regimes is significant at every horizon, so the null result in RZ2018 there reflects the shock environment, not the regime definition. On the post-WWII

sample the opposite holds: at 6.5% no ranking emerges, and it is the estimated threshold that reveals the larger multipliers in the low-unemployment regime, so there the regime definition is what binds. Two margins, each binding in a different part of the sample.²²

The imposed-threshold exercise also answers a question about the instrument itself. RZ2018 document that the military news series is a weak instrument in their high-unemployment state and on the post-WWII sample. We find the same, and add that in the low-unemployment state part of the weakness is produced by the fiscal-event quarters. In those quarters the contemporaneous relation between the news and government spending runs the other way from the rest of the sample – the spending that a large announcement signals arrives only later – so a first stage that pools 24 war quarters with 294 ordinary ones is a cancellation. Restricting RZ2018’s low-unemployment regime to its normal-shock quarters raises the first-stage F -statistic on the news from 2.2 to 11.6 on the full sample and from 0.8 to 13.7 on the sample excluding WWII (Appendix E). The fiscal-event regime is thus where the news is both largest and least like the rest of the sample, a further reason to estimate it as a regime of its own.

4.6 Taking Stock

Two mechanisms account for the difference between our results and RZ2018’s. The first is the shock environment: the instrument’s variation concentrates in a few wartime quarters whose fiscal innovations are two to three times more volatile than the rest of the sample’s, and a single regime structure cannot represent them as anything but a distortion of the state regimes; once isolated, state dependence appears even at the imposed thresholds. The second is sample composition: the historical samples contain extreme-slack episodes – pre-1947 depressions and the deepest post-war recessions – that dominate the high-unemployment regime and deliver the [Auerbach and Gorodnichenko \(2012\)](#) ranking, whereas the post-WWII sample, on its own, delivers the opposite one. The regime definition of [Alloza \(2022\)](#) is the third ingredient, and it binds exactly where the first does not: after WWII.

These findings raise one question that this section cannot answer from within. The variance state is the absolute lagged instrument, and the instrument is a variable inside the VAR; a skeptic can attribute the fiscal-event regime to the heteroskedasticity of the proxy equation itself – an artifact of the identification approach – rather than to genuine atypical-variance

²²RZ2018’s own Blanchard–Perotti estimates on the full sample find significantly higher multipliers in the high-unemployment regime (0.68 against 0.30), the same direction we recover on the historical samples – visible, in their paper, only under the identification that does not inherit the instrument’s concentration.

episodes of the macroeconomy. The likelihood that selects $\hat{\gamma}^\sigma$ observes the news innovations directly, so the two readings cannot be separated here. The next subsection addresses this with a change of laboratory.

4.7 Robustness: Removing the Instrument from the System

We re-estimate the regime structure in the framework of [Auerbach and Gorodnichenko \(2012\)](#), henceforth AG2012²³, chosen because it removes the instrument from the system: identification is recursive following [Blanchard and Perotti \(2002\)](#) in a VAR of government spending, taxes and output over 1947Q1–2008Q4, the news series enters only as a candidate variance state and is absent from the VAR, and the mean state is a backward-looking version of AG2012’s growth state, known at $t - 1$ ²⁴. The likelihood that selects the variance threshold is therefore computed exclusively from the second moments of the macroeconomic aggregates. If the fiscal-event regime of Section 4.3 were a feature of the proxy, there would be no reason for this likelihood to select the same partition; among the 59 distinct partitions the news candidate induces on this sample, it could isolate the largest Korean War announcements only, sweep in dozens of moderate ones, or place its optimum anywhere else. Appendix I reports the full exercise; two results matter here.

First, the AG2012 likelihood selects a variance threshold of 2.04% of GDP, which corresponds to the RZ2018 estimate after the unit conversion between the two news series, and reproduces the fiscal-event partition of Section 4.3 *quarter for quarter* on the common 1949–2008 window (15 of 15 quarters). The second moments of the macroeconomic aggregates themselves break at the fiscal-event quarters: the shock regimes are a property of the macroeconomy, not of the instrument. Second, in the normal-shock regime the post-WWII ranking is the one found above – larger multipliers in the high-growth than in the low-growth regime, 1.11 rising to 3.12 against 0.18 to 0.31 (Table A-17) – and here the difference is statistically significant through $h = 11$ ($\Delta = 0.93, 1.69, 2.14$ with intervals excluding zero), under a different identification scheme, a different state variable and a different sample construction. Across

²³AG2012 estimate a smooth-transition VAR, but the framework is well suited to a hard threshold. In their footnote 10 they report that when the smoothness parameter of the logistic transition is estimated jointly with the regime parameters, its point estimates exceed 5–10 across specifications – which they describe as evidence that the model “is best described as a model switching regimes sharply at certain thresholds” – and they calibrate a smooth transition only because joint nonlinear estimation is imprecise in small samples. The T -rate super-consistency of the threshold (Theorems 3 and 4) resolves precisely that concern.

²⁴A predetermined state is also what [Alloza \(2022\)](#) argues for: AG2012’s centered moving average embeds output growth at $t + 1$ and $t + 2$, and he shows that regimes defined with information available at $t - 1$ reverse their ranking. Here the backward-looking state is required by Assumption B2, and the ranking we recover under recursive identification is his.

the two exercises, then, the shock-regime partition is identical and the direction of state dependence invariant to how “good times” are measured; the magnitudes differ by a factor of two to three, as the identifications and the mean partitions do.

Our relationship to [Alloza \(2022\)](#) is therefore one of confirmation on one margin and departure on another. We confirm that state-dependent conclusions hinge on how the regimes are defined, and we endogenize that definition. But the imposed-threshold exercise shows that the definitional margin is not exhaustive: holding the regime definition fixed at RZ2018’s, the shock-environment split alone restores state dependence on the historical samples. Two researchers using an identical definition of the regimes can reach opposite conclusions depending on whether the shock environment is modeled. The relevant notion of a regime is two-dimensional – a propagation regime defined on the macroeconomic state, and a shock regime defined on the environment in which the identifying shocks arrive – and the second dimension is invisible to any single-partition analysis, however the partition is defined.

5 Conclusion

Threshold models locate a cutoff in an observable state, and the theory for estimating it uses information about the regime-dependence of the conditional mean alone. This paper has developed the estimation theory for threshold VARs in which the conditional variance switches as well. The decoupled model gives mean and variance regimes their own states and thresholds; the joint model, in which one state and one threshold govern both moments, is its constrained special case, and a likelihood-gap rule selects it when it holds. In both, a likelihood-based estimator of the thresholds is consistent and super-consistent at rate T , with limiting distributions of compound-Poisson-argmax form, and the regime parameters retain standard $\sqrt{T}$ Gaussian inference. Estimating the regimes costs nothing asymptotically. The simulations say when the second moments matter: never when only the mean switches, where the likelihood and least-squares criteria are equivalent; for precision when both switch; and for consistency when only the variance switches, where a mean-based criterion has nothing to work with. A single threshold imposed on a process with separate mean and variance regimes misses the dimension it cannot see, at any sample size.

The application shows what that dimension is in a well-known dataset. In the fiscal multiplier literature the identifying variation arrives with wars, and wars are episodes of unusual shock volatility as much as of unusual macroeconomic states. A mean-based estimator confirms the threshold that [Ramey and Zubairy \(2018\)](#) impose. The likelihood estimator departs from

it only because it sees the second moments, and what it sees is the shock environment – a regime indicator nearly collinear with the largest realizations of the instrument – rather than a new state. The decoupled estimator separates the two. It isolates a fiscal-event regime of 33 quarters at a variance threshold that is the same in every sample and for both state variables, and a system that does not contain the instrument recovers the same quarters, one for one. Outside them, state dependence is in the data. On the historical samples multipliers are larger in the high-unemployment regime, at the estimated thresholds and at the one the earlier literature imposed – significantly so, at the imposed threshold, once WWII is excluded. After WWII they are larger in good times, significantly so under recursive identification. The disagreement in the literature decomposes into the shock environment, the sample period and the definition of the regimes, and only the last has been debated. A regime, in this literature, is two-dimensional.

The method does not choose the states; it disciplines the choice. A theory of state dependence suggests a variable and a direction, and the framework turns that statement into objects the data can confront: where the cutoff is, whether the regimes are in the dynamics, in the volatility, or in both, whether the partition survives a change of sample, and what the responses are once the regime structure is estimated rather than assumed. The application shows each of these at work. The unemployment state and the shock-environment state pass; government spending as a variance state does not, because it classifies quarters by their response rather than by what hits them, and the data say so. Two limits bound what the data can say at long horizons, and we have marked them where they apply: the instrument’s weakness in small cells, which we share with the literature that uses it, and the explosive within-regime dynamics of the high-unemployment regime on the century-long sample. Within those limits the lesson is the one the paper opened with. In macroeconomic data volatility has regimes of its own, and an estimator that cannot see them finds them anyway – in the wrong place.

References

- Alloza, Mario.** 2022. “Is fiscal policy more effective during recessions?” *International Economic Review* 63 (3): 1271–1292.
- An, Hong-Zhi, and Fu-Chun Huang.** 1996. “The geometrical ergodicity of nonlinear autoregressive models.” *Statistica Sinica* 6 (4): 943–956.
- Andrews, Donald WK.** 1988. “Laws of large numbers for dependent non-identically distributed random variables.” *Econometric theory* 4 (3): 458–467.
- Auerbach, Alan J, and Yuriy Gorodnichenko.** 2012. “Measuring the output responses to fiscal policy.” *American Economic Journal: Economic Policy* 4 (2): 1–27.
- Auerbach, Alan J, and Yuriy Gorodnichenko.** 2013. “Fiscal multipliers in recession and expansion.” In *Fiscal policy after the financial crisis*, 63–98, University of Chicago Press.
- Bai, Jushan, and Pierre Perron.** 1998. “Estimating and testing linear models with multiple structural changes.” *Econometrica* 47–78.
- Barnichon, Regis, Davide Debortoli, and Christian Matthes.** 2022. “Understanding the size of the government spending multiplier: It’s in the sign.” *The Review of Economic Studies* 89 (1): 87–117.
- Bernardini, Marco, and Gert Peersman.** 2018. “Private debt overhang and the government spending multiplier: Evidence for the United States.” *Journal of Applied Econometrics* 33 (4): 485–508.
- Billingsley, Patrick.** 1961. “The Lindeberg–Lévy theorem for martingales.” *Proceedings of the American Mathematical Society* 12 (5): 788–792.
- Billingsley, Patrick.** 1999. *Convergence of Probability Measures*. New York: Wiley, , 2nd edition.
- Blanchard, Olivier, and Roberto Perotti.** 2002. “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output.” *The Quarterly Journal of Economics* 117 (4): 1329–1368.

- Bloom, Nicholas.** 2009. “The impact of uncertainty shocks.” *econometrica* 77 (3): 623–685.
- Bradley, Richard C.** 2005. “Basic properties of strong mixing conditions. A survey and some open questions.” *Probability Surveys* 2 107–144.
- Canova, Fabio, and Fernando J Pérez Forero.** 2024. “Does the transmission of monetary policy shocks change when inflation is high?” Technical report, Banco Central de Reserva del Perú.
- Chan, Joshua CC, and Eric Eisenstat.** 2018. “Bayesian model comparison for time-varying parameter VARs with stochastic volatility.” *Journal of applied econometrics* 33 (4): 509–532.
- Chan, Kung S, and Howell Tong.** 1985. “On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations.” *Advances in applied probability* 17 (3): 666–678.
- Chan, Kung-Sik.** 1993. “Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model.” *Annals of Statistics* 21 520–533.
- Chan, Kung-Sik, and Ruey S. Tsay.** 1998. “Limiting properties of the least squares estimator of a continuous threshold autoregressive model.” *Biometrika* 85 (2): 413–426.
- Chan, Ngai Hang, and Yury A Kutoyants.** 2012. “On parameter estimation of threshold autoregressive models.” *Statistical inference for stochastic processes* 15 (1): 81–104.
- Chang, Yoosoon, Yongok Choi, and Joon Y Park.** 2017. “A new approach to model regime switching.” *Journal of Econometrics* 196 (1): 127–143.
- Chen, Min, and Gemai Chen.** 2000. “Geometric ergodicity of nonlinear autoregressive models with changing conditional variances.” *Canadian Journal of Statistics* 28 (3): 605–614.
- Cogley, Timothy, and Thomas J Sargent.** 2005. “Drifts and volatilities: monetary policies and outcomes in the post WWII US.” *Review of Economic dynamics* 8 (2): 262–302.

- Doukhan, Paul.** 1994. *Mixing: Properties and Examples*. Volume 85. of Lecture Notes in Statistics, New York: Springer-Verlag.
- Fazzari, Steven M, James Morley, and Irina Panovska.** 2015. “State-dependent effects of fiscal policy.” *Studies in Nonlinear Dynamics & Econometrics* 19 (3): 285–315.
- Feigin, Paul D, and Richard L Tweedie.** 1985. “Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments.” *Journal of time series analysis* 6 (1): 1–14.
- Feller, William.** 1971. *An Introduction to Probability Theory and Its Applications*. Volume II. New York: Wiley, , 2nd edition.
- Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Keith Kuester, and Juan Rubio-Ramírez.** 2015. “Fiscal volatility shocks and economic activity.” *American Economic Review* 105 (11): 3352–3384.
- Gargiulo, Valeria, Christian Matthes, and Katerina Petrova.** 2026. “Monetary policy across inflation regimes.” *European Economic Review* 105322.
- Ghassibe, Mishel, and Francesco Zanetti.** 2022. “State dependence of fiscal multipliers: the source of fluctuations matters.” *Journal of Monetary Economics* 132 1–23.
- Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento.** 2024. “State-dependent local projections.” *Journal of Econometrics* 244 (2): 105702.
- Gonçalves, Silvia, and Lutz Kilian.** 2004. “Bootstrapping autoregressions with conditional heteroskedasticity of unknown form.” *Journal of econometrics* 123 (1): 89–120.
- Gonzalo, Jesús, and Michael Wolf.** 2005. “Subsampling inference in threshold autoregressive models.” *Journal of Econometrics* 127 (2): 201–224.
- Hamilton, James D.** 1989. “A new approach to the economic analysis of nonstationary time series and the business cycle.” *Econometrica: Journal of the econometric society* 357–384.
- Hamilton, James D.** 1994. *Time series analysis*. Princeton university press.

- Hansen, Bruce E.** 1996. "Inference when a nuisance parameter is not identified under the null hypothesis." *Econometrica: Journal of the econometric society* 413–430.
- Hansen, Bruce E.** 2000. "Sample splitting and threshold estimation." *Econometrica* 68 (3): 575–603.
- Hansen, Bruce E.** 2011. "Threshold autoregression in economics." *Statistics and its Interface* 4 (2): 123–127.
- Hirano, Keisuke, and Jack R Porter.** 2003. "Asymptotic efficiency in parametric structural models with parameter-dependent support." *Econometrica* 71 (5): 1307–1338.
- Ibragimov, Ildar A., and Rafail Z. Has'minskii.** 1981. *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.
- Jordà, Òscar.** 2005. "Estimation and inference of impulse responses by local projections." *American economic review* 95 (1): 161–182.
- Justiniano, Alejandro, and Giorgio E Primiceri.** 2008. "The time-varying volatility of macroeconomic fluctuations." *American Economic Review* 98 (3): 604–641.
- Kallenberg, Olav.** 2002. *Foundations of Modern Probability*. New York: Springer, , 2nd edition.
- Kim, Chang-Jin, Jeremy Piger, and Richard Startz.** 2008. "Estimation of Markov regime-switching regression models with endogenous switching." *Journal of Econometrics* 143 (2): 263–273.
- Kurtz, Thomas G.** 1975. "Semigroups of conditioned shifts and approximation of Markov processes." *Annals of Probability* 3 (4): 618–642.
- Kushner, Harold J.** 1980. "A martingale method for the convergence of a sequence of processes to a jump-diffusion process." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 53 207–219.
- Kushner, Harold J.** 1984. *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*. Cambridge, MA: MIT Press.
- Li, CW, and Wai Keung Li.** 1996. "On a double-threshold autoregressive heteroscedastic time series model." *Journal of applied econometrics* 11 (3): 253–274.

- Li, Dong, and Shiqing Ling.** 2012. “On the least squares estimation of multiple-regime threshold autoregressive models.” *Journal of Econometrics* 167 (1): 240–253.
- Liberzon, D.** 2003. “Switching in Systems and Control, Lecture Notes.” Technical report, University of Illinois; see sections on common Lyapunov functions.
- Liu, Jian, Shiying Wu, and James V Zidek.** 1997. “On segmented multivariate regression.” *Statistica Sinica* 497–525.
- Meyn, S. P., and R. L. Tweedie.** 1993. *Markov Chains and Stochastic Stability*. Springer, , Available online; see Chapters 5, 15–16 for drift/petite set results.
- Montiel Olea, José Luis, and Carolin Pflueger.** 2013. “A Robust Test for Weak Instruments.” *Journal of Business & Economic Statistics* 31 (3): 358–369.
- Newey, Whitney K., and Daniel McFadden.** 1994. “Large Sample Estimation and Hypothesis Testing.” Chapter in the Handbook of Econometrics (or lecture notes), standard reference for M-estimators.
- Owyang, Michael T, Valerie A Ramey, and Sarah Zubairy.** 2013. “Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data.” *American Economic Review* 103 (3): 129–134.
- Plagborg-Møller, Mikkel, and Christian K Wolf.** 2021. “Local projections and VARs estimate the same impulse responses.” *Econometrica* 89 (2): 955–980.
- Primiceri, Giorgio E.** 2005. “Time varying structural vector autoregressions and monetary policy.” *The Review of economic studies* 72 (3): 821–852.
- Ramey, Valerie A.** 2011. “Identifying government spending shocks: It’s all in the timing.” *The quarterly journal of economics* 126 (1): 1–50.
- Ramey, Valerie A, and Sarah Zubairy.** 2018. “Government spending multipliers in good times and in bad: evidence from US historical data.” *Journal of political economy* 126 (2): 850–901.
- Samia, Noelle I, and Kung-Sik Chan.** 2011. “Maximum likelihood estimation of a generalized threshold stochastic regression model.” *Biometrika* 98 (2): 433–448.

- Sims, Christopher A, and Tao Zha.** 2006. “Were there regime switches in US monetary policy?” *American Economic Review* 96 (1): 54–81.
- Stroock, Daniel W., and S. R. Srinivasa Varadhan.** 1971. “Diffusion processes with boundary conditions.” *Communications on Pure and Applied Mathematics* 24 (2): 147–225.
- Teräsvirta, Timo.** 1994. “Specification, estimation, and evaluation of smooth transition autoregressive models.” *Journal of the American Statistical Association* 89 (425): 208–218.
- Tong, Howell.** 1977. “Some comments on the Canadian lynx data.” *Journal of the Royal Statistical Society Series A: Statistics in Society* 140 (4): 432–436.
- Tong, Howell.** 1990. *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- Tong, Howell, and Keng S Lim.** 1980. “Threshold autoregression, limit cycles and cyclical data.” *Journal of the Royal Statistical Society: Series B (Methodological)* 42 (3): 245–268.
- Tsay, Ruey S.** 1998. “Testing and modeling multivariate threshold models.” *Journal of the American Statistical Association* 93 (443): 1188–1202.
- Yu, Ping.** 2012. “Likelihood estimation and inference in threshold regression.” *Journal of Econometrics* 167 (1): 274–294.
- Yu, Ping.** 2014. “The bootstrap in threshold regression.” *Econometric Theory* 30 (3): 676–714.

Online Appendix for “Mean and Variance Regimes in Threshold VARs: Revisiting Regime-dependent Fiscal Multipliers”

A	Proofs	A-2
A.1	Geometric Ergodicity	A-2
A.2	Auxiliary Lemmas	A-6
A.3	Consistency	A-10
A.4	Rates and Limit Laws: the Single-threshold Case	A-17
A.5	Rates and Limit Laws: the Decoupled Case	A-28
B	Additional Tables from Monte Carlo Exercise	A-34
C	The Interest-Rate (ZLB) Exercise	A-36
D	Leverage Points and the Full Sample	A-39
E	Instrument Strength by Cell	A-42
F	Posterior Bands as a Check on the Bootstrap	A-45
G	Fiscal-Event Cells	A-46
H	Spending Growth as the Variance State	A-46
I	The Auerbach–Gorodnichenko Exercise	A-48

A Proofs

Proofs are organized by argument rather than by order of appearance in the text: the single-threshold arguments underlying Theorem 4 come first, because the proofs for the decoupled model (Theorems 1–3) extend their steps and refer to them.

A.1 Geometric Ergodicity

A.1.1 Geometric Ergodicity and Second Moments

Lemma 1. *Under assumptions A1 to A3, the Markov chain $(\mathbf{x}_t)$ defined by (8) is aperiodic and ν -irreducible with ν being the Lebesgue measure. Furthermore, a ν -non-null compact set is a small set.*

Proof. To establish that the Markov chain is aperiodic, ν -irreducible with small compact sets, we follow the approach of [Chan and Tong \(1985, Section 2\)](#) and [An and Huang \(1996, Lemma 5.2\)](#), adapted to the multivariate threshold VAR setting. The proof relies on verifying three fundamental properties:

- (i) ν -irreducibility. For $p > 1$ the one-step kernel is singular with respect to Lebesgue measure on $\mathbb{R}^{np}$: the noise enters only the first n coordinates of X_{t+1} , while the remaining $(p-1)n$ coordinates are deterministic shifts of X_t . Irreducibility is therefore established through the p -step kernel. Conditional on $X_t = x$, factor the joint law of $(y_{t+1}, \dots, y_{t+p})$ sequentially: for each j , conditionally on $(x, y_{t+1}, \dots, y_{t+j-1})$ the active regime at $t+j$ is determined (it is a measurable function of these variables), and y_{t+j} is Gaussian with nonsingular covariance Σ_i for the realized regime i (A1, A3), hence has an everywhere-positive conditional density on $\mathbb{R}^n$. The joint density of the block is the product of these conditional densities and is therefore everywhere positive on $\mathbb{R}^{np}$; since X_{t+p} is a fixed permutation of the block, the p -step transition kernel $P^p(x, \cdot)$ has an everywhere-positive density $q_p(x, \cdot)$ for every x . Consequently every set of positive Lebesgue measure is reached in p steps with positive probability from every starting point, and the chain is ν -irreducible with $\nu = \text{Leb}$.
- (ii) Aperiodicity. We must show the transition kernel has no cyclical structure. The same argument applied to $p+1$ steps shows that $q_{p+1}(x, \cdot)$ is also everywhere positive. A cycle of period $d \geq 2$ would require the sets visited at times p and $p+1$ from a common starting point to lie in distinct elements of the cyclic decomposition, contradicting the positivity of both densities on all of $\mathbb{R}^{np}$; since $\gcd(p, p+1) = 1$, the chain is aperiodic.
- (iii) Compact sets are small. The key structural fact is that the p -step density involves *no* integration over intermediate states: X_{t+p} consists exactly of the block $(y_{t+p}, \dots, y_{t+1})$, so for a

target point $z = (z'_p, \dots, z'_1)'$ the intermediate states χ_j (the stacked state after j steps) are deterministic functions of $(x, z_1, \dots, z_j)$, the realized regime path $i_1, \dots, i_p$ is the deterministic path $i_j = i^*(\chi_{j-1})$, and the p -step density is the finite product

$$q_p(x, z) = \prod_{j=1}^p \varphi_{\Sigma_{i_j}}(z_j - A_{i_j}^{(y)} \chi_{j-1} - b_{i_j}^{(y)}),$$

where φ_{Σ} is the $N(0, \Sigma)$ density on $\mathbb{R}^n$ and $A^{(y)}, b^{(y)}$ denote the first n rows of the companion objects. Now fix $C = \{x : \|x\| \leq R\}$ and any compact $K \subset \mathbb{R}^{np}$ of positive Lebesgue measure, and set $R_K := \sup_{z \in K} \|z\|$. Every intermediate state satisfies $\|\chi_{j-1}\| \leq R' := \max(R, R_K)$, since its coordinates are coordinates of x or of z . Hence every Gaussian argument in the product is bounded: $\|z_j - A_{i_j}^{(y)} \chi_{j-1} - b_{i_j}^{(y)}\| \leq R_K + \max_i(\|A_i\| R' + \|b_i\|) =: R''$, and

$$\delta_0 := \min_{1 \leq i \leq k} \inf_{\|w\| \leq R''} \varphi_{\Sigma_i}(w) > 0$$

because each φ_{Σ_i} is continuous and strictly positive (A3) and the minimum runs over finitely many regimes. Therefore $q_p(x, z) \geq \delta_0^p =: \delta > 0$ for all $x \in C$ and $z \in K$, uniformly over the (state-determined) regime path, and

$$P^p(x, B) = \int_B q_p(x, z) dz \geq \delta \text{Leb}(B \cap K) \quad \text{for all } x \in C, B \in \mathcal{B}_{np} :$$

C is small with $m = p$ and minorizing measure $\nu_p(\cdot) = \delta \text{Leb}(\cdot \cap K)$.

The general framework is [Meyn and Tweedie \(1993, Proposition 16.1.3\)](#): non-degenerate noise (A1, A3) with bounded coefficients (A2) ensures that the chain can be neither trapped nor periodic. Relative to the univariate state-dependent-volatility model of [Chen and Chen \(2000\)](#), the regime-specific constant volatility structure (A3) simplifies the verification, since non-degeneracy holds uniformly within each regime. $\square$

Theorem A1. *Under assumptions A1–A4 the Markov chain $\{X_t\}$ is geometrically ergodic. Moreover, the unique invariant probability measure π satisfies*

$$\int \|x\|^2 \pi(dx) < \infty,$$

so in particular the VAR process has finite second moment under stationarity.

In order to prove Theorem A1, we will check the conditions in the following lemma.

Lemma 2. *Suppose that $(\mathbf{x}_t)$ is an aperiodic, ν -irreducible Markov chain, where ν is the Lebesgue measure. If there exists a small set A , a non-negative measurable function g and constants $\rho \in (0, 1)$ and $b > 0$ such that:*

(i) g is bounded away from zero and infinity on A ;

(ii) $\sup_{\mathbf{x} \in A} \mathbb{E}\{g(\mathbf{x}_t) | \mathbf{x}_{t-1} = \mathbf{x}\} < \infty$;

(iii) $\mathbb{E}\{g(\mathbf{x}_t) | \mathbf{x}_{t-1} = \mathbf{x}\} < \rho g(\mathbf{x}) - b, \mathbf{x} \in A^c$;

then $(\mathbf{x}_t)$ is geometrically ergodic and

$$\int_{\mathbb{R}^{np}} g(\mathbf{x}) \pi(d\mathbf{x}) < \infty \quad (\text{A-1})$$

where π is an invariant probability measure on $\mathbb{R}^p$ with respect to the Markov chain $(\mathbf{x}_t)$.

Proof of Lemma 2. The geometric ergodicity follows from the drift criterion of [Meyn and Tweedie \(1993, Theorem 15.0.1 and Ch. 15–16\)](#) as presented in [Tong \(1990, Appendix 1\)](#), and (A-1) follows from [Feigin and Tweedie \(1985\)](#).

Proof of Theorem A1. By Lemma 1 the chain is ν -irreducible and aperiodic and compact sets are small, so it remains to exhibit a Lyapunov function satisfying (i)–(iii) of Lemma 2.

Step 1: Lyapunov function. Let $P \succ 0$ satisfy $A_i' P A_i \preceq \lambda_i^2 P$ for all i , with $\lambda := \max_i \lambda_i < 1$ (such a P exists under **A4**), and set $g(x) := 1 + x' P x$. Then $g \geq 1$, g is measurable, and $c_1(1 + \|x\|^2) \leq g(x) \leq c_2(1 + \|x\|^2)$ with $c_1 := \min(1, \lambda_{\min}(P))$ and $c_2 := \max(1, \lambda_{\max}(P))$; all drift and moment statements below transfer between g and the Euclidean norm through these constants.

Step 2: one-step drift. Fix $x \in \mathbb{R}^{np}$, let $i = i^*(x)$, and write $\bar{A} := \max_i \|A_i\|$, $\bar{b} := \max_i \|b_i\|$ (finite by **A2**). From $X_{t+1} = A_i x + b_i + H_i \varepsilon_{t+1}$, $E \varepsilon_{t+1} = 0$ and $E \varepsilon_{t+1} \varepsilon_{t+1}' = I$ (**A1**),

$$\begin{aligned} E[g(X_{t+1}) | X_t = x] &= 1 + (A_i x + b_i)' P (A_i x + b_i) + \text{tr}(H_i' P H_i) \\ &\leq 1 + \lambda^2 x' P x + 2\bar{A} \|P\| \bar{b} \|x\| + \lambda_{\max}(P) \bar{b}^2 + C_1 \\ &= \lambda^2 g(x) + 2\bar{A} \|P\| \bar{b} \|x\| + C_1', \end{aligned}$$

where $\text{tr}(H_i' P H_i) = \text{tr}(P_{11} \Sigma_i) \leq \lambda_{\max}(P) n C_\Sigma =: C_1$ by **A3**, and $C_1' := (1 - \lambda^2) + \lambda_{\max}(P) \bar{b}^2 + C_1$.

Step 3: conditions of Lemma 2. Fix $\rho \in (\lambda^2, 1)$ and $b > 0$. Since $g(x) \geq 1 + \lambda_{\min}(P) \|x\|^2$ grows quadratically while the cross term in Step 2 grows linearly, there is $R < \infty$ such that $(\rho - \lambda^2) \lambda_{\min}(P) R^2 \geq 2\bar{A} \|P\| \bar{b} R + C_1' + b$; for $\|x\| > R$ this gives $E[g(X_{t+1}) | X_t = x] \leq \rho g(x) - b$, i.e. condition (iii) of Lemma 2 on the complement of the compact set $A := \{x : \|x\| \leq R\}$. On A , which is small by Lemma 1, $1 \leq g(x) \leq 1 + \lambda_{\max}(P) R^2$ and $\sup_{x \in A} E[g(X_{t+1}) | X_t = x] \leq 1 + \lambda_{\max}(P) [(\bar{A} R + \bar{b})^2 + n C_\Sigma] < \infty$, which are conditions (i) and (ii). Lemma 2 yields geometric ergodicity and $\int g d\pi < \infty$, hence $\int \|x\|^2 \pi(dx) < \infty$ and $E\|y_t\|^2 < \infty$ under stationarity. $\square$

A.1.2 Proof of Theorem 4(i)

Proof. Geometric ergodicity is Theorem A1; the finite fourth moment follows by repeating its drift argument for a quartic Lyapunov function on the lifted system $z := x \otimes x \in \mathbb{R}^{(np)^2}$. Set $\tilde{A}_i := A_i \otimes A_i$. If $A'_i P A_i \preceq \lambda_i^2 P$ then, by the mixed-product property and because the Kronecker product of positive semidefinite matrices is positive semidefinite, $\tilde{A}'_i (P \otimes P) \tilde{A}_i = (A'_i P A_i) \otimes (A'_i P A_i) \preceq \lambda_i^4 (P \otimes P)$; hence $W := P \otimes P \succ 0$ satisfies the lifted inequality $\tilde{A}'_i W \tilde{A}_i \preceq \lambda^{(4)} W$ for all i , with $\lambda^{(4)} := \max_i \lambda_i^4 < 1$. Define

$$G(x) := 1 + z' W z, \quad z = x \otimes x,$$

so that $1 + \lambda_{\min}(W) \|x\|^4 \leq G(x) \leq 1 + \lambda_{\max}(W) \|x\|^4$, since $\|z\| = \|x\|^2$.

Step 1: one-step drift for G . Fix x , let $i = i^*(x)$, $y := A_i x + b_i$ and $\eta := H_i \varepsilon_{t+1}$, so that $X_{t+1} = y + \eta$ and

$$\begin{aligned} z_{t+1} &:= X_{t+1} \otimes X_{t+1} = \tilde{A}_i z + R_i(x) + \mathcal{N}, \\ R_i(x) &:= (A_i x) \otimes b_i + b_i \otimes (A_i x) + b_i \otimes b_i, \quad \mathcal{N} := y \otimes \eta + \eta \otimes y + \eta \otimes \eta. \end{aligned}$$

Expand $z'_{t+1} W z_{t+1}$ and take conditional expectations given $X_t = x$. Terms linear in ε_{t+1} vanish (A1); the remaining terms are bounded as follows, with $\bar{A} := \max_i \|A_i\|$, $\bar{b} := \max_i \|b_i\|$ (A2), $\|H_i\|_{\text{op}} \leq C_{\Sigma}^{1/2}$ (A3) and $E\|\eta\|^r \leq C_{\Sigma}^{r/2} E\|\varepsilon_t\|^r < \infty$ for $r \leq 4$ (A1):

- (a) the leading term, $z' \tilde{A}'_i W \tilde{A}_i z \leq \lambda^{(4)} z' W z$, by the lifted inequality;
- (b) the intercept terms, $2(\tilde{A}_i z)' W R_i(x) + R_i(x)' W R_i(x) \leq c_1 \|z\|^{3/2} + c_2 \|z\|$ for constants c_1, c_2 depending on $\|W\|, \bar{A}, \bar{b}$ only, since $\|R_i(x)\| \leq 2\bar{A}\bar{b}\|x\| + \bar{b}^2 \leq C(1 + \|z\|^{1/2})$;
- (c) the terms of second and third order in $\eta - E[(y \otimes \eta)' W (y \otimes \eta)], E[(\eta \otimes y)' W (\eta \otimes y)]$, the cross terms of $\tilde{A}_i z + R_i(x)$ with $\eta \otimes \eta$, and the cross terms of $y \otimes \eta, \eta \otimes y$ with $\eta \otimes \eta$ – are each bounded by $\|W\|$ times a product of $\|y\|$ or $\|\tilde{A}_i z + R_i(x)\|$ with $E\|\eta\|^2$ or $E\|\eta\|^3$; since $\|y\| \leq \bar{A}\|z\|^{1/2} + \bar{b}$ and $\|\tilde{A}_i z + R_i(x)\| \leq C(1 + \|z\|)$, their sum is at most $C_2(1 + \|z\|)$;
- (d) the purely quartic noise term, $E[(\eta \otimes \eta)' W (\eta \otimes \eta)] \leq \|W\| E\|\eta\|^4 \leq \|W\| C_{\Sigma}^2 m_4 =: C_4$.

Since $\|z\|^{3/2}$ and $\|z\|$ grow sub-quadratically, for any $\epsilon > 0$ there is $C_{\epsilon} < \infty$ with $c_1 \|z\|^{3/2} + (c_2 + C_2) \|z\| \leq \epsilon \lambda_{\min}(W) \|z\|^2 + C_{\epsilon} \leq \epsilon z' W z + C_{\epsilon}$ (Young's inequality). Choosing ϵ so that $\tilde{\lambda}^2 := \lambda^{(4)} + \epsilon < 1$ and collecting (a)–(d),

$$E[G(X_{t+1}) \mid X_t = x] \leq \tilde{\lambda}^2 G(x) + C, \quad C := C_{\epsilon} + C_2 + C_4 + 1 - \tilde{\lambda}^2 < \infty.$$

Step 2: conclusion. The display has the form of the drift inequality of Step 2 of the proof of Theorem A1, with (g, λ^2) replaced by $(G, \tilde{\lambda}^2)$ and no linear cross term, and G dominates

$\lambda_{\min}(W)\|x\|^4$. Step 3 of that proof therefore applies verbatim: for $\rho \in (\tilde{\lambda}^2, 1)$ and $b > 0$ choose R with $(\rho - \tilde{\lambda}^2)\lambda_{\min}(W)R^4 \geq C + b$, so that $E[G(X_{t+1}) \mid X_t = x] \leq \rho G(x) - b$ for $\|x\| > R$, while on the small set $A := \{x : \|x\| \leq R\}$ the function G lies between 1 and $1 + \lambda_{\max}(W)R^4$ and $\sup_{x \in A} E[G(X_{t+1}) \mid X_t = x] \leq 1 + \lambda_{\max}(W)(\bar{A}R + \bar{b})^4 + C_2(1 + R^2) + C_4 < \infty$. Lemma 2 gives geometric ergodicity with $\int G d\pi < \infty$, hence $\int \|x\|^4 \pi(dx) < \infty$ and $E\|y_t\|^4 < \infty$ under stationarity. $\square$

A.1.3 Proof of Theorem 1

Proof. The proofs of Theorems A1 and 4(i) apply with k replaced by k_μ and with the noise loading $H_{\ell^*(x)}$ in place of $H_{i^*(x)}$. The drift inequalities (Step 2 of the proof of Theorem A1 and Step 1 of the proof of Theorem 4(i)) involve the variance regime only through the noise constants: for any active pair (i, ℓ) , $\text{tr}(H'_\ell P H_\ell) = \text{tr}(P_{11} \Sigma_\ell) \leq \lambda_{\max}(P) n C_\Sigma = C_1$ and $E\|H_\ell \varepsilon_{t+1}\|^4 \leq C_\Sigma^2 m_4$, uniformly in ℓ by **A3**, while the contraction constants $\lambda, \lambda^{(4)}$ and the compact set $A = \{x : \|x\| \leq R\}$ are determined by the mean companion matrices $\{A_i\}_{i=1}^{k_\mu}$ alone (**A4**). Since $\ell^*(\cdot)$ is measurable (**A2**), Lemma 1 applies unchanged, and Lemma 2 gives the conclusion. The variance regimes thus affect only the additive constants in the drift inequalities, mirroring the structure of (1): the stability of $\{X_t\}$ is governed entirely by the mean regimes, while the variance regimes affect only the noise loading. $\square$

A.2 Auxiliary Lemmas

A.2.1 Lemma 3

A convenient feature of model (10) is that the density regularity that Chan (1993, Conditions 2–3) and Samia and Chan (2011, Assumption 3) must assume — boundedness of the marginal and pairwise densities of the threshold variable, and full support of the state — holds automatically under A1–A3 and B1. Define $\sigma_* := \min_{1 \leq \ell \leq k_\sigma} \lambda_{\min}(\Sigma_\ell^0)$. Since each Σ_i^0 is nonsingular (A3) and there are finitely many regimes, $\sigma_* > 0$; this is a property of the true parameter, not an additional assumption¹.

Lemma 3. *Under the Gaussian innovation assumption in model (1) (and its special case (10)), A1–A4 and B1, with π the stationary law of X_t from Theorem A1:*

- (i) *The conditional law of $q_t = e'_s y_t$ given $\mathcal{F}_{t-1}$ has a density bounded by $C_* := (2\pi\sigma_*)^{-1/2}$, uniformly over histories.*

¹Where statements uniform over the parameter space are needed, as in Lemma 7, the compactness of $\mathcal{S}$ in **B1** supplies the uniform counterpart.

- (ii) The stationary marginal density f_π of q_{t-d} exists, is everywhere positive, and is bounded by C_* ; and for every $t \neq s$ the pair (q_{t-d}, q_{s-d}) has a joint density bounded by C_*^2 , uniformly over (t, s) .
- (iii) π has an everywhere-positive Lebesgue density on $\mathbb{R}^{np}$. Consequently, for every nonzero matrix $C \in \mathbb{R}^{n \times (np+1)}$ whose restriction to the non-constant coordinates of x_t is nonzero, and every event $E \in \sigma(X_{t-1})$ with $P(E) > 0$, we have $P(Cx_t \neq 0 \mid E) > 0$.

Proof

- (i) Given $\mathcal{F}_{t-1}$, the active regime $i(t) = i^*(X_{t-1})$ is known (it is a measurable function of X_{t-1}), and y_t is Gaussian with covariance $\Sigma_{i(t)}$; hence $q_t \mid \mathcal{F}_{t-1} \sim N(e'_s B_{i(t)} x_t, e'_s \Sigma_{i(t)} e_s)$ with variance $e'_s \Sigma_{i(t)} e_s \geq \sigma_*$, and its density is bounded by C_* uniformly over histories.
- (ii) By the invariance equation, $f_\pi(u) = E_\pi[f_{q_t \mid \mathcal{F}_{t-1}}(u)] \leq C_*$, and $f_\pi(u) > 0$ everywhere because each conditional Gaussian density is everywhere positive. For $s < t$, iterating expectations,

$$f_{(q_{s-d}, q_{t-d})}(u, v) = E_\pi[f_{q_{s-d} \mid \mathcal{F}_{s-d-1}}(u) E[f_{q_{t-d} \mid \mathcal{F}_{t-d-1}}(v) \mid \mathcal{F}_{s-d}]] \leq C_*^2.$$

- (iii) By Lemma 1, the p -step kernel admits an everywhere-positive density $q_p(x, \cdot)$ on $\mathbb{R}^{np}$ for every x ; hence $\pi(B) = \int P^p(x, B) \pi(dx) > 0$ for every Borel B of positive Lebesgue measure, and, since $\pi = \pi P^p$ with P^p absolutely continuous, π admits the density $f_X(z) = \int q_p(x, z) \pi(dx)$, which is everywhere positive. For the last claim, decompose $Cx_t = c_0 + C_1 X_{t-1}$; if $C_1 \neq 0$, the set $\{c_0 + C_1 X_{t-1} = 0\}$ is a proper affine subset of $\mathbb{R}^{np}$, hence π -null and null under any conditioning on positive-probability $\sigma(X_{t-1})$ -events, while if $C_1 = 0$ and $c_0 \neq 0$ the event $\{Cx_t \neq 0\}$ is sure. $\square$

Remark. Lemma 3(ii) supplies, without assumption, the uniformly-bounded pairwise densities that Samia and Chan (2011) require in Assumption 3 (and invoke “for all i, j ” in their Theorem 3) and that drive the tightness bound in Step 3 below. Lemma 3(iii) is used in the identification proof of Appendix A.3 and in the remark following the mean clause of **B2**. In the decoupled model, parts (i)–(ii) apply verbatim to any threshold variable that is a coordinate of y_t at some lag; for general measurable transformations q^μ, q^σ of X_t , the corresponding regularity is maintained through the density clauses of **B2** and **B8**.

A.2.2 Lemma 4: discharging the mixing condition

Assumption **B5** is the one condition of the B-block that restricts *dependence* rather than moments or densities, and it is the hardest to verify: geometric ergodicity delivers β -mixing but not ρ -mixing (Bradley, 2005). Samia and Chan (2011) verify their Assumption 7 by hand in a univariate

example (condition (C7) of their online supplement), using two features special to the scalar case – an envelope depending only on the threshold variable and future noise, and a stationary density bounded below on the band – neither of which survives in the multivariate model. Lemma 4 takes a different route: the proofs never use ρ -mixing itself, only two of its consequences for functionals supported on the shrinking threshold band (the covariance bounds behind the Claim in Step 5 of the proof of Theorem 4(iii), and the dependence-removal ingredient (c) of Step 4 of the proof of Theorem 4(iv)), and for such functionals a *density bridge* substitutes for mixing: conditioning $d+1$ steps before the band time, the bounded one-step conditional density of the threshold variable (Lemma 3(i)) turns the functional’s conditional expectation into a bounded function of the state – bounded by C_* times the band width – which the weighted total-variation bound of Theorem A1 controls at a geometric rate. The price is a strengthening of **B6** from second to $(2+\delta)$ -th conditional moments of an extended envelope:

B6* There exist $\tau > 0$, $\delta > 0$ and $M < \infty$ such that, with $g(x) = 1 + x'Px$ the weight function of Theorem A1,

$$E\left[(\Lambda_t + g(X_t))^{2+\delta} \mid q_{t-d} = v\right] \leq M \quad \text{for a.e. } v \in [\gamma^0 - \tau, \gamma^0 + \tau].$$

The same bound is imposed around each threshold, mean and variance alike conditionally on $q_{t-d_\mu}^\mu = v$ for v near $\gamma^{\mu,0}$ and on $q_{t-d_\sigma}^\sigma = v$ for v near $\gamma^{\sigma,0}$. Both are conditional *moment* requirements of the same nature as **B6**, in place of a dependence requirement.

Lemma 4. *Let Assumptions A1–A4, B1 and B6* hold, and let $R < \infty$, $\rho \in (0, 1)$ be the constants of Theorem A1, so that $\|P^j(x, \cdot) - \pi\|_{TV} \leq Rg(x)\rho^j$ for all x and $j \geq 1$. Fix $\Delta \leq \tau$ and write $\mathbb{I}_t(I) := \mathbb{I}\{q_{t-d} \in I\}$ for intervals $I \subseteq [\gamma^0 - \Delta, \gamma^0 + \Delta]$. Then:*

(i) *(Covariance bounds.) There exist $H < \infty$ and a summable sequence $\{\bar{\rho}(j)\}_{j \geq 1}$ with $\bar{\rho}(j) \leq C\rho^{j\delta/(2+\delta)}$ such that, for every interval I as above, every pair $s < t$, and every pair of measurable functions $m_t = m(x_t, y_t)$, $m_s = m'(x_s, y_s)$ with $|m|, |m'| \leq C_m\Lambda$,*

$$|\text{Cov}(m_t \mathbb{I}_t(I), m_s \mathbb{I}_s(I))| \leq C_m^2 H \bar{\rho}(t-s) P_\pi(q_{t-d} \in I).$$

In particular, the variance bounds $\text{var}(\sum_{t \leq T} m_t \mathbb{I}_t(I)) \leq TH'P_\pi(q \in I)$ required in the Claim of Step 5 of the proof of Theorem 4(iii) hold.

(ii) *(Averaging bound.) Let $\varepsilon = 1/T$, let $I_\varepsilon := [\gamma^0, \gamma^0 + \kappa\varepsilon]$, let $J_t := \xi_t^+ \mathbb{I}_t(I_\varepsilon)$, and let F be bounded and measurable. Then for every $k \geq 1$, every $w \in \mathbb{R}$, and every sub- σ -field $\mathcal{G} \subseteq \mathcal{F}_t$, writing $h_k := F(w + J_{t+k}) - F(w)$,*

$$E|E[h_k \mid \mathcal{G}] - E[h_k]| \leq 4\|F\|_\infty C_* \kappa \varepsilon \left(\mathbb{I}\{k \leq d+1\} + R\rho^{k-d-1} E_\pi[g] \mathbb{I}\{k > d+1\} \right),$$

so that, summed over a window of $m_\varepsilon = \lceil T\tau_\varepsilon \rceil$ terms and normalized by τ_ε as in Step 4 of the proof of Theorem 4(iv), the total is $O(\varepsilon/\tau_\varepsilon) = O(T^{-2/3})$ for the window choice $\tau_\varepsilon = T^{-1/3}$ – the analogue of display (A49) of Samia and Chan (2011), with the mixing coefficient replaced by the geometric total-variation rate.

Proof. (The bridge.) For any random variable of the form $\eta_t = m(x_t, y_t) \mathbb{1}_t(I)$ with $|m| \leq L$ define $b_\eta(x) := E[\eta_t \mid X_{t-d-1} = x]$. Conditioning first on q_{t-d} ,

$$b_\eta(x) = \int_I E[m(x_t, y_t) \mid X_{t-d-1} = x, q_{t-d} = v] f_{q_{t-d} \mid X_{t-d-1}=x}(v) dv, \quad |b_\eta(x)| \leq L C_* |I|,$$

by Lemma 3(i). Thus b_η is *bounded*, and for any $j \geq 1$,

$$|P^j b_\eta(x) - \pi b_\eta| \leq \|b_\eta\|_\infty \min\{2, R g(x) \rho^j\} \leq L C_* |I| \min\{2, R g(x) \rho^j\}. \quad (\text{L7.1})$$

(i). Write $j := t - s$ and $\phi_t := m_t \mathbb{1}_t(I)$. For $j \leq d + 1$ the bound is immediate from Cauchy–Schwarz and B6*: $E[\phi_t^2] \leq C_m^2 \int_I E[\Lambda_t^2 \mid q = v] f_\pi(v) dv \leq C_m^2 M C_* |I|$, and similarly for ϕ_s , so $|\text{Cov}| \leq C |I|$, which is of the required form since there are finitely many such lags. For $j > d + 1$, truncate: $m_t = m_t^L + m_t^{>L}$ with $m_t^L := m_t \mathbb{1}_{\{\Lambda_t \leq L\}}$. By the Markov property and $\mathcal{F}_s$ -measurability of ϕ_s ,

$$\text{Cov}(m_t^L \mathbb{1}_t(I), \phi_s) = E[(\phi_s - E\phi_s) (P^{j-d-1} b_L(X_s) - \pi b_L)],$$

with b_L the bridge of $m_t^L \mathbb{1}_t(I)$. By (L7.1) and Cauchy–Schwarz with B6* ($E[\phi_s^2 g(X_s)^2]^{1/2} \leq C_m E[(\Lambda_s + g(X_s))^4]^{1/2} \leq C |I|^{1/2}$ when $\delta \geq 2$, and by Hölder with exponent $(2 + \delta)/2$ in general),

$$|\text{Cov}(m_t^L \mathbb{1}_t(I), \phi_s)| \leq C_m L C_* |I| R \rho^{j-d-1} E[(|\phi_s| + E|\phi_s|) (1 + g(X_s))] \leq C L \rho^{j-d-1} |I|.$$

For the tail, by Cauchy–Schwarz, B6* and Markov’s inequality,

$$|\text{Cov}(m_t^{>L} \mathbb{1}_t(I), \phi_s)| \leq \left(E[m_t^2 \mathbb{1}_{\{\Lambda_t > L\}} \mathbb{1}_t(I)] \right)^{1/2} (E[\phi_s^2])^{1/2} \leq C L^{-\delta/2} |I|.$$

Optimizing $L = \rho^{-2(j-d-1)/(2+\delta)}$ gives the stated $\bar{\rho}(j) \leq C \rho^{(j-d-1)\delta/(2+\delta)}$; summability yields the variance bound by the usual double sum. Finally $P_\pi(q \in I) \geq$ (a constant multiple of) $|I|$ on the band by the density clause of **B2** and Lemma 3(ii) after shrinking Δ , so $|I|$ may be replaced by $P_\pi(q \in I)$ throughout.

(ii). Here $|h_k| \leq 2\|F\|_\infty \mathbb{1}_{t+k}(I_\varepsilon)$, so the bridge $b(x) := E[h_k \mid X_{t+k-d-1} = x]$ satisfies $\|b\|_\infty \leq 2\|F\|_\infty C_* \kappa \varepsilon$ with *no truncation*. For $k > d + 1$, by the tower property, Jensen’s inequality ($\mathcal{G} \subseteq \mathcal{F}_t$),

the Markov property and (L7.1),

$$E |E[h_k | \mathcal{G}] - Eh_k| \leq E |P^{k-d-1}b(X_t) - \pi b| \leq 2\|F\|_\infty C_* \kappa \varepsilon R \rho^{k-d-1} E_\pi[g].$$

For $k \leq d+1$, bound both terms separately: $E|E[h_k | \mathcal{G}]| \leq E|h_k| \leq 2\|F\|_\infty P_\pi(q \in I_\varepsilon) \leq 2\|F\|_\infty C_* \kappa \varepsilon$, and likewise $|Eh_k|$. Summing over $k = 0, \dots, m_\varepsilon - 1$ and dividing by τ_ε gives $O(\kappa \varepsilon (d + 1 + \sum_k \rho^k) / \tau_\varepsilon) = O(\varepsilon / \tau_\varepsilon)$. $\square$

Remark (scope). Part (i) supplies the variance bounds invoked in the Claim of Step 5 of the proof of Theorem 4(iii) (with $m_t \in \{1, \Lambda_t, \xi_t^\pm\}$, all dominated by $C\Lambda_t$), and part (ii) supplies ingredient (c) of Step 4 of the proof of Theorem 4(iv), where the test functions have compact support and the increments h_k are of the stated band-supported form. Consequently, *Theorem 4(iii)–(iv) holds with Assumption B5 replaced by B6**. The same argument discharges **B5** for Theorem 3 whenever both threshold variables admit one-step conditional densities bounded uniformly over histories – automatic when q^μ and q^σ are coordinates of y at fixed lags, by Lemma 3(i) – with the bridge taken $d_\mu + 1$ (respectively $d_\sigma + 1$) steps before the relevant band time; for general measurable transformations of the state, **B5** is retained. Lemma 4 does *not* assert that the localized process is ρ -mixing – a property that does not follow from geometric ergodicity in general (Bradley, 2005) – only that the two consequences of ρ -mixing used in the proofs follow from Theorem A1 under the conditional moment condition B6*.

A.3 Consistency

A.3.1 Proof of Theorem 4(ii)

The proof follows the standard strategy for extremum estimators, verifying: (i) uniform convergence of the sample criterion, (ii) uniqueness of the population maximizer, and (iii) application of the argmax theorem (Newey and McFadden, 1994). Below we use the notation and assumptions of Section 2. We explicitly rely on the geometric ergodicity and moment results established in Theorem 4(i) and Appendix A.1.

Step 0: population log-likelihood. Define the population log-likelihood

$$\ell(\theta) = \mathbb{E}_\pi[\ell_t(\theta)]$$

where $\ell_t(\theta)$ is the time- t contribution:

$$\ell_t(\theta) = \sum_{i=1}^k \left[-\frac{1}{2} \ln \det \Sigma_i - \frac{1}{2} \varepsilon'_{it} \Sigma_i^{-1} \varepsilon_{it} \right] \mathbb{I}(q_{t-d} \in \mathcal{R}_i(\gamma))$$

By the density clause of **B2**, the marginal density $f_\pi(\gamma_i^0) > 0$ for all thresholds, which ensures that each regime has positive stationary probability:

$$\pi_i := P_\pi(q_{t-d} \in R_i(\gamma^0)) > 0 \quad \forall i = 1, \dots, k$$

Step 1: uniform LLN.

Lemma 5. Uniform law of large numbers (ULLN) *Assume **A1–A4** and **B1**, and in addition:*

(M1) *The stationary process $\{y_t\}$ is geometrically β -mixing (this follows from geometric ergodicity, Appendix A.1).*

(M2) $\mathbb{E}_\pi \|y_t\|^4 < \infty$ *(this follows from Theorem 4(i))*

Then

$$\sup_{\theta \in \Theta} |L_T(\theta) - \ell(\theta)| \xrightarrow{a.s.} 0 \quad (T \rightarrow \infty). \quad (\text{A-2})$$

Proof sketch For each fixed θ , ergodicity implies pointwise almost-sure convergence $L_T(\theta) \rightarrow \ell(\theta)$. Compactness of Θ , together with the fourth-moment bound and the uniform positive-definiteness restriction on Σ_i , yields an integrable envelope for $\ell_t(\theta)$. Uniformity over the smooth coordinates (B, Σ) follows from standard stochastic equicontinuity arguments, while uniformity over the threshold coordinate follows from the usual finite-grid approximation for threshold models together with a ULLN for dependent sequences; see Andrews (1988) and Newey and McFadden (1994).

Step 2: identification - unique maximizer.

Lemma 6. (Unique maximizer) *Under **A1–A4**, **B2–B4**, θ_0 uniquely maximizes $\ell(\theta)$:*

$$\theta_0 = \arg \max_{\theta \in \Theta} \ell(\theta) \quad (\text{A-3})$$

Proof of Lemma 6.

(i) **Information inequality.** Let $f(y_t | \mathcal{F}_{t-1}; \theta)$ be the conditional density implied by parameter θ . By Jensen's inequality (information inequality),

$$\ell(\theta) - \ell(\theta_0) = \mathbb{E}_\pi \left[\ln \frac{f(y_t | \mathcal{F}_{t-1}; \theta)}{f(y_t | \mathcal{F}_{t-1}; \theta_0)} \right] \leq 0,$$

with equality iff $f(\cdot; \theta) = f(\cdot; \theta_0)$ almost surely. Thus any θ that yields conditional densities different from the truth must have strictly smaller population log-likelihood.

(ii) **Parameter identification.** We now show that equality can hold only at $\theta = \theta_0$.

Case 1: $\gamma = \gamma^0$, $(B_i, \Sigma_i) \neq (B_i^0, \Sigma_i^0)$ for some regime i . On $\{q_{t-d} \in \mathcal{R}_i^0\}$, which has positive stationary probability by the density clause of **B2**, equality of the conditional densities requires $B_i x_t = B_i^0 x_t$ and $\Sigma_i = \Sigma_i^0$ for π -almost every state. If $B_i \neq B_i^0$, the set $\{(B_i - B_i^0)x_t = 0\}$ is contained in a lower-dimensional affine subspace, hence is Lebesgue-null and therefore π -null by Lemma 3(iii); if $B_i = B_i^0$ but $\Sigma_i \neq \Sigma_i^0$, the Gaussian laws differ everywhere. In either case the densities differ on a set of positive probability, so

$$\ell(\theta) < \ell(\theta_0)$$

Case 2: $\gamma \neq \gamma^0$. Superpose the two partitions: the candidate thresholds $\{\gamma_1, \dots, \gamma_{k-1}\}$ and the true thresholds $\{\gamma_1^0, \dots, \gamma_{k-1}^0\}$ jointly partition $\mathbb{R}$ into finitely many intervals, on each of which both the true regime index $\tau(v)$ and the candidate regime index $c(v)$ are constant. Since the stationary law of q_{t-d} has an everywhere-positive density (Lemma 3(iii)), every such interval of positive length has positive probability. Suppose the criterion equality $\ell(\theta) = \ell(\theta_0)$ held. Then on each positive-length interval I of the superposition, the candidate conditional law must coincide with the true one for π -almost every state with $q_{t-d} \in I$; by the argument of Case 1, this forces the *parameter* identity $(B_{c(I)}, \Sigma_{c(I)}) = (B_{\tau(I)}^0, \Sigma_{\tau(I)}^0)$ on every such interval. Now, since $\gamma \neq \gamma^0$ and both are ordered $(k-1)$ -vectors, some true threshold γ_j^0 is not a candidate threshold, and therefore lies in the *interior* of some candidate regime $\mathcal{R}_{c^*}(\gamma)$. Choose $\varepsilon > 0$ smaller than the distance from γ_j^0 to every other threshold of either partition; then $(\gamma_j^0 - \varepsilon, \gamma_j^0)$ and $(\gamma_j^0, \gamma_j^0 + \varepsilon)$ are two positive-length superposition intervals inside the same candidate regime c^* , with true indices j and $j+1$ respectively. The interval-wise parameter identities give $(B_{c^*}, \Sigma_{c^*}) = (B_j^0, \Sigma_j^0)$ and $(B_{c^*}, \Sigma_{c^*}) = (B_{j+1}^0, \Sigma_{j+1}^0)$ simultaneously, hence $(B_j^0, \Sigma_j^0) = (B_{j+1}^0, \Sigma_{j+1}^0)$, contradicting **B3**.

Combining the two cases shows that θ_0 is the unique maximizer of $\ell(\theta)$. (Unlike [Chan \(1993\)](#), which assumes constant variance across regimes, regime-specific Σ_i provides additional identification power; see Remark 8 and Section 3.2.)

Step 3: consistency via argmax theorem. By Lemma 5, $L_T \rightarrow \ell$ uniformly on Θ ; by Lemma 6, θ_0 is the unique maximizer of ℓ on the compact Θ (**B1**); Theorem 2.1 of [Newey and McFadden \(1994\)](#) gives $\hat{\theta}_T = \arg \max_{\theta \in \Theta} L_T(\theta) \rightarrow_{a.s.} \theta_0$. $\square$

A.3.2 Proof of Theorem 2

The proof parallels that of Theorem 4(ii), now applied to the product parameter space $\Gamma^\mu \times \Gamma^\sigma$, and to the profiled criterion of the decoupled model.

Lemma T5-F (uniform validity of the feasible profile). Write $L_T^F(\gamma^\mu, \gamma^\sigma)$ for the one-

step feasible criterion of the implementation Remark of Section 2.2 (the object denoted L_T^P there) reserving L_T^P in this appendix for the exact profile $\sup_{\vartheta} L_T(\gamma^\mu, \gamma^\sigma, \vartheta)$, with population counterparts $\tilde{\ell}$ (the population criterion evaluated at the population one-step maps) and ℓ^P . Assume **A1–A3**, **A4**, **B1–B4**, and:

$$\begin{aligned} (\mathbf{F}) \text{ (Uniform candidate feasibility.) } & \inf_{\gamma^\mu \in \Gamma^\mu} \min_{1 \leq i \leq k_\mu} P_\pi(q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu(\gamma^\mu)) > 0 \\ \text{and } & \inf_{\gamma^\sigma \in \Gamma^\sigma} \min_{1 \leq \ell \leq k_\sigma} P_\pi(q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma(\gamma^\sigma)) > 0. \end{aligned}$$

Then

- (i) $\sup_{(\gamma^\mu, \gamma^\sigma) \in \Gamma^\mu \times \Gamma^\sigma} |L_T^F - \tilde{\ell}| \rightarrow 0$ a.s.;
- (ii) $\tilde{\ell} \leq \ell^P$ on $\Gamma^\mu \times \Gamma^\sigma$, provided the parameter blocks $\mathcal{B}, \mathcal{S}$ of Θ contain the population pipeline values $\{\tilde{B}_i(\gamma^\mu, \gamma^\sigma), \tilde{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma)\}$; a finite enlargement of the bounds in **B1** always suffices, since these values form a bounded family, being continuous on the compact $\Gamma^\mu \times \Gamma^\sigma$; and
- (iii) $\tilde{\ell}(\gamma^{\mu,0}, \gamma^{\sigma,0}) = \ell^P(\gamma^{\mu,0}, \gamma^{\sigma,0}) = \ell(\theta_0)$.

Remark. Condition **(F)** concerns marginal regime probabilities under *candidate* thresholds; empty joint cells at candidate pairs are harmless, since the GLS normal equations and the pooled covariances aggregate over cells within a mean (respectively variance) regime. In the two-regime baseline the candidate regimes are half-lines and **(F)** reduces to the state marginals placing positive mass below $\underline{\gamma}$ and above $\bar{\gamma}$, automatic for coordinates of y_t by Lemma 3(iii); with more than two regimes per dimension it additionally requires a minimum regime width in $\Gamma^\mu, \Gamma^\sigma$ (thresholds separated by some $\epsilon > 0$), the population counterpart of the minimum-observations guard in the implementation.

Proof. (i) Every ingredient of the feasible pipeline is a function of finitely many cell sample moments of the form $T^{-1} \sum_t g(x_t, y_t) \mathbb{I}\{q_{t-d_\mu}^\mu \leq a\} \mathbb{I}\{q_{t-d_\sigma}^\sigma \leq b\}$ with g one of finitely many quadratic maps. For each fixed g the family indexed by (a, b) is monotone in each index with integrable envelope (Theorem 4), and its pointwise ergodic limits are continuous in (a, b) because the states have absolutely continuous laws (Lemma 3); uniform a.s. convergence over the compact $\Gamma^\mu \times \Gamma^\sigma$ follows by the monotonicity argument underlying Lemma 5. By **(F)**, the population mean-regime and variance-regime probabilities are bounded below by some $\underline{\pi} > 0$ uniformly over candidates. The population mean-regime second-moment matrices are uniformly nonsingular: for $a \neq 0$, $a' E_\pi[x_t x_t' \mathbb{I}\{q^\mu \in \mathcal{R}_i^\mu(\gamma^\mu)\}] a = E_\pi[(a' x_t)^2 \mathbb{I}\{\cdot\}]$ can vanish only if $a' x_t = 0$ a.s. on an event of probability at least $\underline{\pi}$, which is impossible for $a \neq 0$ because the stationary law of X_t has an everywhere-positive Lebesgue density (Lemma 3(iii)) and x_t contains the intercept; the map $\gamma^\mu \mapsto \lambda_{\min}(E_\pi[x_t x_t' \mathbb{I}\{\cdot\}])$ is continuous by dominated convergence (the states have atomless laws), hence bounded below on the compact Γ^μ . The population pooled residual covariances dominate the innovation covariances, $\tilde{\Sigma}_\ell(\gamma^\mu, \gamma^\sigma) \succeq \min_{\ell'} \lambda_{\min}(\Sigma_{\ell'}^0) I_n \succ 0$, uniformly. On this domain – regime

probabilities $\geq \underline{\pi}$, moment matrices with eigenvalues in a fixed compact of $(0, \infty)$ – the pipeline maps (cell OLS, pooled covariance, one-step GLS, concentrated criterion) are jointly Lipschitz-continuous functions of the cell moments; by the uniform convergence of those moments, the sample pipeline is, almost surely for all large T , well defined uniformly over $\Gamma^\mu \times \Gamma^\sigma$ (sample regime shares $\geq \underline{\pi}/2$, sample moment matrices within the compact eigenvalue domain), and uniform convergence is preserved under the composition. (ii) The plug-in evaluates ℓ at the population pipeline parameters, the profile at the supremum over $\mathcal{B} \times \mathcal{S}$; the inequality follows whenever the former lie in the latter, which is the stated qualification. Boundedness of the pipeline family holds because, by **(F)** and the uniform moment bounds of part (i), the maps $(\gamma^\mu, \gamma^\sigma) \mapsto (\tilde{B}_i, \tilde{\Sigma}_\ell)$ are continuous on the compact threshold space (dominated convergence, atomless state laws), hence have bounded range. (iii) At the true thresholds the cells are correctly specified, so the population within-mean-regime OLS estimand is B_i^0 regardless of the variance-regime heteroskedasticity, the pooled population residual covariance is Σ_ℓ^0 , the population GLS estimand is again B_i^0 , and the criterion value is $\ell(\theta_0)$. $\square$

Step 1. ULLN. It is convenient to work first with the full Gaussian criterion (2). For $\vartheta := (B, \Sigma) \in \mathcal{B} \times \mathcal{S}$ and $\theta := (\gamma^\mu, \gamma^\sigma, \vartheta) \in \Theta$, the per-observation contribution is

$$\ell_t(\theta) = \sum_{i=1}^{k_\mu} \sum_{\ell=1}^{k_\sigma} \left(-\frac{1}{2} \log \det \Sigma_\ell - \frac{1}{2} \varepsilon_{it}(B_i)' \Sigma_\ell^{-1} \varepsilon_{it}(B_i) \right) \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu(\gamma^\mu)\} \mathbb{I}\{q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^\sigma(\gamma^\sigma)\},$$

with $\varepsilon_{it}(B_i) = y_t - (x_t' \otimes I_n) \text{vec}(B_i)$, $L_T(\theta) = T^{-1} \sum_{t=1}^T \ell_t(\theta)$ and $\ell(\theta) = E_\pi[\ell_t(\theta)]$. We show

$$\sup_{\theta \in \Theta} |L_T(\theta) - \ell(\theta)| \xrightarrow{a.s.} 0. \quad (\text{A.23})$$

By **B1**, $\mathcal{B} \times \mathcal{S}$ is compact and $0 < \underline{\sigma} \leq \lambda_{\min}(\Sigma_\ell) \leq \lambda_{\max}(\Sigma_\ell) \leq \bar{\sigma} < \infty$ uniformly over $\Sigma_\ell \in \mathcal{S}_\ell$, so $|\log \det \Sigma_\ell|$ and $\|\Sigma_\ell^{-1}\|$ are uniformly bounded and $\sup_{B_i \in \mathcal{B}_i} \|\varepsilon_{it}(B_i)\| \leq C(1 + \|y_t\| + \|x_t\|)$; hence

$$\sup_{\theta \in \Theta} |\ell_t(\theta)| \leq M_t := C(1 + \|y_t\|^2 + \|x_t\|^2), \quad E_\pi M_t < \infty \quad (\text{A.24})$$

by Theorem 1 and the definition of x_t : $\{M_t\}$ is an integrable envelope. For the threshold coordinates, let $\Delta^\mu(\gamma^\mu, \tilde{\gamma}^\mu) := \bigcup_{j=1}^{k_\mu-1} [\gamma_j^\mu \wedge \tilde{\gamma}_j^\mu, \gamma_j^\mu \vee \tilde{\gamma}_j^\mu]$ and define $\Delta^\sigma(\gamma^\sigma, \tilde{\gamma}^\sigma)$ analogously. Two parameter points sharing the same ϑ induce different cell assignments only when $q_{t-d_\mu}^\mu \in \Delta^\mu$ or $q_{t-d_\sigma}^\sigma \in \Delta^\sigma$, so

$$|\ell_t(\theta) - \ell_t(\tilde{\theta})| \leq 2M_t \left(\mathbb{I}\{q_{t-d_\mu}^\mu \in \Delta^\mu(\gamma^\mu, \tilde{\gamma}^\mu)\} + \mathbb{I}\{q_{t-d_\sigma}^\sigma \in \Delta^\sigma(\gamma^\sigma, \tilde{\gamma}^\sigma)\} \right),$$

and by the density clauses of **B2** the stationary densities of $q_{t-d_\mu}^\mu$ and $q_{t-d_\sigma}^\sigma$ are bounded on the compact threshold spaces, whence $P_\pi(q_{t-d_\mu}^\mu \in \Delta^\mu) \leq C_\mu \|\gamma^\mu - \tilde{\gamma}^\mu\|$ and $P_\pi(q_{t-d_\sigma}^\sigma \in \Delta^\sigma) \leq C_\sigma \|\gamma^\sigma - \tilde{\gamma}^\sigma\|$. Together with (A.24) this gives stochastic equicontinuity in the threshold coordinates;

uniformity in the smooth coordinates (B, Σ) follows from compactness and continuity of $\ell_t(\theta)$ with the same envelope. Since $\{X_t\}$ is geometrically ergodic (Theorem 1), hence geometrically β -mixing, the ULLN for dependent processes applies (Andrews, 1988; Newey and McFadden, 1994), proving (A.23). Finally, with the exact profile criteria $L_T^P(\gamma^\mu, \gamma^\sigma) := \sup_{\vartheta \in \mathcal{B} \times \mathcal{S}} L_T(\gamma^\mu, \gamma^\sigma, \vartheta)$ and $\ell^P(\gamma^\mu, \gamma^\sigma) := \sup_{\vartheta \in \mathcal{B} \times \mathcal{S}} \ell(\gamma^\mu, \gamma^\sigma, \vartheta)$,

$$\sup_{(\gamma^\mu, \gamma^\sigma) \in \Gamma^\mu \times \Gamma^\sigma} |L_T^P(\gamma^\mu, \gamma^\sigma) - \ell^P(\gamma^\mu, \gamma^\sigma)| \leq \sup_{\theta \in \Theta} |L_T(\theta) - \ell(\theta)| \xrightarrow{a.s.} 0, \quad (\text{A.27})$$

which is the required ULLN for the profiled criterion.

Step 2. Unique maximizer. We show that θ_0 uniquely maximizes the full-criterion population objective $\ell(\theta) := E_\pi[\ell_t(\theta)]$ on Θ , and then pass to the profiled criterion. The argument follows the two-case superposition pattern of Lemma 6, applied to each threshold dimension.

- (i) **Information inequality.** For any $\theta \in \Theta$, letting $f(y_t \mid \mathcal{F}_{t-1}; \theta)$ denote the conditional Gaussian density implied by θ ,

$$\ell(\theta) - \ell(\theta_0) = \mathbb{E}_\pi \left[\ln \frac{f(y_t \mid \mathcal{F}_{t-1}; \theta)}{f(y_t \mid \mathcal{F}_{t-1}; \theta_0)} \right] \leq 0,$$

with equality if and only if $f(\cdot; \theta) = f(\cdot; \theta_0)$ almost surely under π .

- (ii) **Any $\theta \neq \theta_0$ produces a density mismatch on a set of positive stationary probability.**

If $\gamma^\mu \neq \gamma^{\mu,0}$, superpose the candidate and true mean partitions as in Case 2 of Lemma 6: some true mean threshold is interior to a candidate mean regime, so equality of the candidate and true conditional densities a.s. would force a single candidate matrix B_i to satisfy $B_i x_t = B_j^0 x_t$ and $B_i x_t = B_{j'}^0 x_t$ for distinct true regimes $j \neq j'$, each on a set of positive stationary probability (the density clause of **B2**); hence $(B_j^0 - B_{j'}^0) x_t = 0$ on a set of positive probability, which is impossible by **B3** and Lemma 3(iii) (the stationary law of X_t has an everywhere-positive density, so a nonzero linear functional of x_t vanishes only on a null set, the knife-edge intercept/threshold-coordinate case being excluded by the mean clause of **B2**). If $\gamma^\sigma \neq \gamma^{\sigma,0}$ (with γ^μ arbitrary), the same superposition on the variance side forces a single candidate Σ_ℓ to equal two distinct true matrices $\Sigma_m^0 \neq \Sigma_{m'}^0$ (**B3**) – impossible, and requiring no condition on x_t , since covariances are matched matrix-to-matrix. If both thresholds are correct but $(B_i, \Sigma_\ell) \neq (B_i^0, \Sigma_\ell^0)$ on some cell of positive probability (the cell clause of **B2**), the conditional Gaussian laws differ on that cell by Lemma 3(iii) (mean block) or directly (covariance block). In all cases $\ell(\theta) < \ell(\theta_0)$.

- (iii) **Moving to the profile.** $\ell(\gamma^\mu, \gamma^\sigma, \vartheta)$ is continuous in $\vartheta = (B, \Sigma)$ on the compact $\mathcal{B} \times \mathcal{S}$ (dominated convergence with the envelope (A.24)), so the supremum defining $\ell^P(\gamma^\mu, \gamma^\sigma) =$

$\sup_{\vartheta} \ell(\gamma^{\mu}, \gamma^{\sigma}, \vartheta)$ is attained. For $(\gamma^{\mu}, \gamma^{\sigma}) \neq (\gamma^{\mu,0}, \gamma^{\sigma,0})$, every attaining ϑ yields a parameter $\theta \neq \theta_0$, so by (i)–(ii), $\ell^{\mathbb{P}}(\gamma^{\mu}, \gamma^{\sigma}) < \ell(\theta_0) = \ell^{\mathbb{P}}(\gamma^{\mu,0}, \gamma^{\sigma,0})$. Hence $(\gamma^{\mu,0}, \gamma^{\sigma,0})$ is the unique maximizer of $\ell^{\mathbb{P}}$ on $\Gamma^{\mu} \times \Gamma^{\sigma}$.

Step 3. Consistency via the argmax theorem. By (A.27) the profiled criterion converges uniformly, and by Step 2 $(\gamma^{\mu,0}, \gamma^{\sigma,0})$ is the unique maximizer of $\ell^{\mathbb{P}}$ on the compact $\Gamma^{\mu} \times \Gamma^{\sigma}$; Theorem 2.1 of [Newey and McFadden \(1994\)](#) gives $(\hat{\gamma}^{\mu}, \hat{\gamma}^{\sigma}) = \arg \max_{\Gamma^{\mu} \times \Gamma^{\sigma}} L_T^{\mathbb{P}} \xrightarrow{\text{a.s.}} (\gamma^{\mu,0}, \gamma^{\sigma,0})$.

Step 4: the feasible estimator. Steps 1–3 establish consistency of the exact-profile maximizer. The estimator of Section 2.2 maximizes the feasible criterion L_T^{F} ; denote it $(\hat{\gamma}_{\text{F}}^{\mu}, \hat{\gamma}_{\text{F}}^{\sigma})$ within this step. Fix $\delta > 0$. By Step 2 and the continuity of $\ell^{\mathbb{P}}$ on the compact $\Gamma^{\mu} \times \Gamma^{\sigma}$ (dominated convergence, using that the states have atomless laws), there is $c = c(\delta) > 0$ with $\ell^{\mathbb{P}}(\gamma^{\mu}, \gamma^{\sigma}) \leq \ell^{\mathbb{P}}(\gamma^{\mu,0}, \gamma^{\sigma,0}) - c$ whenever $\|(\gamma^{\mu}, \gamma^{\sigma}) - (\gamma^{\mu,0}, \gamma^{\sigma,0})\| \geq \delta$. Almost surely, for all large T , Lemma T5-F(i) gives

$$\tilde{\ell}(\hat{\gamma}_{\text{F}}^{\mu}, \hat{\gamma}_{\text{F}}^{\sigma}) \geq L_T^{\text{F}}(\hat{\gamma}_{\text{F}}^{\mu}, \hat{\gamma}_{\text{F}}^{\sigma}) - c/3 \geq L_T^{\text{F}}(\gamma^{\mu,0}, \gamma^{\sigma,0}) - c/3 \geq \tilde{\ell}(\gamma^{\mu,0}, \gamma^{\sigma,0}) - 2c/3 = \ell^{\mathbb{P}}(\gamma^{\mu,0}, \gamma^{\sigma,0}) - 2c/3,$$

using the definition of the feasible maximizer in the second inequality and Lemma T5-F(iii) in the equality. Since $\tilde{\ell} \leq \ell^{\mathbb{P}}$ everywhere (Lemma T5-F(ii), with the bounds of **B1** enlarged if necessary so that the part-(ii) qualification holds), the feasible maximizer satisfies $\ell^{\mathbb{P}}(\hat{\gamma}_{\text{F}}^{\mu}, \hat{\gamma}_{\text{F}}^{\sigma}) \geq \ell^{\mathbb{P}}(\gamma^{\mu,0}, \gamma^{\sigma,0}) - 2c/3$ and therefore lies within δ of the truth for all large T . As $\delta > 0$ was arbitrary, $(\hat{\gamma}_{\text{F}}^{\mu}, \hat{\gamma}_{\text{F}}^{\sigma}) \rightarrow (\gamma^{\mu,0}, \gamma^{\sigma,0})$ a.s.; this is the estimator to which the statement of Theorem 2 refers. Consistency of the regime parameters then follows as in Steps 1–3, evaluating the (uniformly convergent) feasible parameter maps at the consistent thresholds. $\square$

A.3.3 Proof of Proposition 1

Throughout, $q^{\mu} \equiv q^{\sigma} =: q$ with a common delay, and the data follow (10) with k regimes, true threshold vector γ^0 , and $B_i^0 \neq B_{i+1}^0$, $\Sigma_i^0 \neq \Sigma_{i+1}^0$ for all i . With a shared state, a candidate pair $(\gamma^{\mu}, \gamma^{\sigma})$ generates the superposed partition of $\mathbb{R}$ by both threshold vectors; the feasible joint cells are the cells of that partition (a staircase), and the cell clause of **B2** is imposed on those cells only, as in Remark B.8. Condition **(F)** is unaffected, since it concerns marginal regime probabilities.

Step 1 (consistency). The uniform convergence of Lemma T5-F(i) and the population inequality (ii) hold verbatim: nothing in their proofs uses the distinctness of the two state variables. For (iii) and the identification step, note that the true pair (γ^0, γ^0) lies on the diagonal of the candidate space, where the two partitions coincide and the population pipeline returns (B_i^0, Σ_i^0) . Uniqueness of the population maximizer follows from the superposition argument of Lemma 6 applied to the common refinement of the candidate pair and γ^0 : if $\gamma^{\mu} \neq \gamma^0$, some true threshold is interior to a candidate

mean regime and one candidate coefficient matrix must equal two distinct true ones; if $\gamma^\sigma \neq \gamma^0$, some true threshold is interior to a candidate variance regime and one candidate covariance must equal two distinct true ones; either contradicts **B3**. The squeeze of Step 4 in the proof of Theorem 2 then gives $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma) \rightarrow (\gamma^0, \gamma^0)$ a.s.

Step 2 (rates). Localize both candidates at scale $1/T$ around γ^0 : $(\gamma^\mu, \gamma^\sigma) = (\gamma^0 + h_1/T, \gamma^0 + h_2/T)$. A boundary observation at position v (in $1/T$ units from γ^0) is reassigned in the mean only, in the variance only, or in both, according to the position of v relative to h_1 and h_2 ; its expected loss is, respectively, the mean-only divergence $\frac{1}{2}E[\kappa | \cdot] > 0$ (the mean clause of **B2** with **B7**), the variance-only divergence $\frac{1}{2}\sum_j(\lambda_j - \ln \lambda_j - 1) > 0$ (**B3**), or the joint divergence $(J3) > 0$. All three being strictly positive, the uniform negative drift of the local criterion over the box $\{(h_1, h_2)\}$ holds exactly as in Step 5 of the proof of Theorem 4(iii), with the single band counting of that proof in place of the two disjoint bands of Theorem 3(i); hence $T(\hat{\gamma}^\mu - \gamma^0) = O_p(1)$ and $T(\hat{\gamma}^\sigma - \gamma^0) = O_p(1)$. The number of misclassified observations is therefore $O_p(1)$, and the feasible-GLS regime-parameter estimates differ from the joint estimator's by $o_p(T^{-1/2})$: this is (ii).

Step 3 (selection). Both maximized criteria are evaluated within $O_p(1/T)$ of (γ^0, γ^0) by Steps 1–2, so $T(\hat{L}_T^F - \hat{L}_T^J)$ is the difference between the maximum of the local criterion over the box and its maximum over the diagonal $h_1 = h_2$; both are finite sums of the $O_p(1)$ boundary losses, hence the difference is $O_p(1)$. Under a shared-state truth with distinct thresholds $\gamma^{\mu,0} \neq \gamma^{\sigma,0}$, the restricted estimator misclassifies a positive fraction of the sample whatever its threshold, so $\hat{L}_T^J$ is bounded away from $\hat{L}_T^F$ by the identification gap of Theorem 2 and the difference is of order T . Consistency of the rule “select the restriction if $T(\hat{L}_T^F - \hat{L}_T^J) < c_T$ ” for $c_T \rightarrow \infty$, $c_T = o(T)$, follows; this is the threshold analogue of the break-number selection argument of [Bai and Perron \(1998\)](#). $\square$

A.4 Rates and Limit Laws: the Single-threshold Case

A.4.1 Lemma 7

Lemma 7 is the analogue, for our two models, of Lemmas 5 and 6 of [Samia and Chan \(2011\)](#) (and of displays (4.11)–(4.12) in [Chan, 1993](#)). Part (i) says the fixed-threshold estimators are insensitive, at the $o_p(T^{-1/2})$ scale, to $O(1/T)$ perturbations of the threshold; part (ii) says the normalized profile-likelihood increments may be computed as if the regime parameters were known. Part (i) is the converse direction of the argument already used in Step 6 of Appendix A.4 (there: threshold error does not disturb the parameter CLT; here: parameter estimation does not disturb the threshold local process), and both parts feed the proofs of Theorems 4(iv) and 3(ii)–(iii).

For the joint model, write $\varphi := (\text{vec}(B_1)', \text{vec}(B_2)', \text{vech}(\Sigma_1)', \text{vech}(\Sigma_2)')'$, let $\hat{\varphi}(\gamma)$ collect the fixed- γ maximum likelihood estimators, and let $l(\varphi, \gamma) := T L_T(\varphi, \gamma)$ denote the unnormalized log-likelihood, so that $T L_T^P(\gamma) = l(\hat{\varphi}(\gamma), \gamma)$. For the decoupled model, $\hat{\varphi}(\gamma^\mu, \gamma^\sigma)$ collects the feasible

profile-GLS estimators of Section 2.2 and the statements hold with the interval $|\gamma - \gamma^0| \leq K/T$ replaced by the box $\|(\gamma^\mu, \gamma^\sigma) - (\gamma^{\mu,0}, \gamma^{\sigma,0})\|_\infty \leq K/T$.

Lemma 7. *Let the assumptions of Theorem 4(iv) (respectively Theorem 3(ii)–(iii)) hold. Then for every $K > 0$:*

(i)

$$\sup_{|\gamma - \gamma^0| \leq K/T} \|\hat{\varphi}(\gamma) - \hat{\varphi}(\gamma^0)\| = O_p(T^{-1}) = o_p(T^{-1/2});$$

(ii) *with $\mathcal{L}_T(\kappa)$ the normalized profile increment of Theorem 4(iv) and $\mathcal{L}_T^0(\kappa) := l(\varphi_0, \gamma^0 + \kappa/T) - l(\varphi_0, \gamma^0)$ the increment at the true regime parameters,*

$$\sup_{|\kappa| \leq K} |\mathcal{L}_T(\kappa) - \mathcal{L}_T^0(\kappa)| = o_p(1).$$

The decoupled analogues hold with (κ) replaced by (h_μ, h_σ) ranging over $[-K, K]^2$.

Proof.

(i) Consider the joint model. The fixed- γ Gaussian MLE is regime-by-regime OLS plus the residual covariance. With $\mathbb{I}_{i,\gamma}(t) := \mathbb{I}\{q_{t-d} \in \mathcal{R}_i(\gamma)\}$,

$$\begin{aligned} \hat{S}_i(\gamma) &:= \sum_t x_t x_t' \mathbb{I}_{i,\gamma}(t), \quad \hat{h}_i(\gamma) := \sum_t x_t y_t' \mathbb{I}_{i,\gamma}(t), \quad \hat{B}_i(\gamma)' = \hat{S}_i(\gamma)^{-1} \hat{h}_i(\gamma), \\ \hat{\Sigma}_i(\gamma) &= T_i(\gamma)^{-1} \sum_t (y_t - \hat{B}_i(\gamma) x_t)(y_t - \hat{B}_i(\gamma) x_t)' \mathbb{I}_{i,\gamma}(t), \quad T_i(\gamma) := \sum_t \mathbb{I}_{i,\gamma}(t). \end{aligned}$$

Over $|\gamma - \gamma^0| \leq K/T$ the indicators $\mathbb{I}_{i,\gamma}(t)$ and $\mathbb{I}_{i,\gamma^0}(t)$ can differ only for observations in the band $\mathcal{B}_T := [\gamma^0 - K/T, \gamma^0 + K/T]$. By Lemma 3(ii), the stationary density of q_{t-d} is bounded by C_* , so $P_\pi(q_{t-d} \in \mathcal{B}_T) \leq C_* |\mathcal{B}_T| = 2KC_*/T$ and hence $E N_T \leq T \cdot C_* \cdot 2K/T = 2KC_*$, where $N_T := \#\{t : q_{t-d} \in \mathcal{B}_T\}$, so $N_T = O_p(1)$ by Markov's inequality; and by B6 each band observation contributes moment terms $\|x_t x_t'\|, \|x_t y_t'\| \leq \Lambda_t$ with $E[\Lambda_t \mathbb{I}\{q_{t-d} \in \mathcal{B}_T\}] \leq M^{1/2} C_* |\mathcal{B}_T| T \leq 2KC_* M^{1/2}$. Since the indicator differences are monotone in γ on each side of γ^0 , the supremum over the interval of the moment perturbations is dominated by the total over the band, so $\sup_{|\gamma - \gamma^0| \leq K/T} \|\hat{S}_i(\gamma) - \hat{S}_i(\gamma^0)\| = O_p(1)$, $\sup_{|\gamma - \gamma^0| \leq K/T} \|\hat{h}_i(\gamma) - \hat{h}_i(\gamma^0)\| = O_p(1)$,

and similarly $|T_i(\gamma) - T_i(\gamma^0)| \leq N_T = O_p(1)$. By the ergodic theorem, $T^{-1} \hat{S}_i(\gamma^0) \rightarrow_{a.s.} S_i = E_\pi[x_t x_t' \mathbb{I}\{q \in \mathcal{R}_i^0\}]$, which is positive definite: $\pi_i > 0$ by the density clauses of **B2** and x_t is not supported on any proper affine subspace conditionally on $\{q_{t-d} \in \mathcal{R}_i^0\}$, by Lemma 3(iii).

On the event $\{\lambda_{\min}(T^{-1}\hat{S}_i(\gamma^0)) \geq \lambda_{\min}(S_i)/2\}$, whose probability tends to one, the map $(\hat{S}, \hat{h}) \mapsto \hat{S}^{-1}\hat{h}$ is Lipschitz in a neighborhood, and the $O_p(1)$ unnormalized perturbations translate into $\sup_{|\gamma - \gamma^0| \leq K/T} \|\hat{B}_i(\gamma) - \hat{B}_i(\gamma^0)\| = O_p(T^{-1})$.

The same two-line argument, applied to the residual cross-product sums (whose perturbation combines $O_p(1)$ band terms and the just-established $O_p(T^{-1})$ coefficient perturbation multiplied through $O_p(T)$ sums, i.e. an $O_p(1)$ total) and to $T_i(\gamma)$, gives $\sup \|\hat{\Sigma}_i(\gamma) - \hat{\Sigma}_i(\gamma^0)\| = O_p(T^{-1})$.

For the decoupled model, the feasible profile-GLS estimator of Section 2.2 is the composition of three maps of the cellwise moments $\hat{S}_{i\ell}, \hat{h}_{i\ell}, t_{i\ell}$ — preliminary OLS, pooled preliminary covariance, feasible GLS — each smooth with nonsingular Jacobian limits ($E_\pi[x_t x'_t \mathbb{I}_{i\ell}] \succ 0$ by B2joint and Lemma 3(iii); $\Sigma_\ell^0 \succ 0$ by B1). An $O(1/T)$ move of either threshold perturbs each unnormalized cell moment by $O_p(1)$, by the same band-counting argument applied to the μ - and σ -bands separately (B6 and Lemma 3(ii)/B8). Chaining the Lipschitz maps yields the same $O_p(T^{-1})$ bound, uniformly over the box.

- (ii) Since $\hat{\varphi}(\gamma)$ maximizes $l(\cdot, \gamma)$, the profiling error $l(\hat{\varphi}(\gamma), \gamma) - l(\varphi_0, \gamma)$ has no linear term and is purely quadratic in the displacement $a_\gamma := \hat{\varphi}(\gamma) - \varphi_0$. Each such quadratic form is $O_p(1)$ — too large to discard — but the lemma concerns their *difference* across γ_κ and γ^0 , where the leading terms cancel and three channels remain: (a) the movement of the estimate, $O_p(T^{-1})$ by part (i); (b) the Hessian change from reallocating $O_p(1)$ band observations; (c) the Hessian change along the moved evaluation path. These are bounded in turn. Fix κ and write $\gamma_\kappa := \gamma^0 + \kappa/T$, $a_\gamma := \hat{\varphi}(\gamma) - \varphi_0$, and let $H(\varphi, \gamma) := -\nabla_\varphi^2 l(\varphi, \gamma)$. For T large, $\hat{\varphi}(\gamma)$ lies in the interior of the compact $\mathcal{B} \times \mathcal{S}$ with probability approaching one (B1, B4, Theorem 4(ii)), so $\nabla_\varphi l(\hat{\varphi}(\gamma), \gamma) = 0$ and a second-order Taylor expansion with integral remainder gives

$$l(\hat{\varphi}(\gamma), \gamma) - l(\varphi_0, \gamma) = \frac{1}{2} a'_\gamma \bar{H}_\gamma a_\gamma, \quad \bar{H}_\gamma := \int_0^1 2(1-u) H(\varphi_0 + u a_\gamma, \gamma) du.$$

Hence

$$\mathcal{L}_T(\kappa) - \mathcal{L}_T^0(\kappa) = \frac{1}{2} a'_{\gamma_\kappa} \bar{H}_{\gamma_\kappa} a_{\gamma_\kappa} - \frac{1}{2} a'_{\gamma^0} \bar{H}_{\gamma^0} a_{\gamma^0}. \quad (\text{A.32})$$

Now bound the difference of the two quadratic forms. Write $a_{\gamma_\kappa} = a_{\gamma^0} + r_\kappa$ with $\sup_{|\kappa| \leq K} \|r_\kappa\| = O_p(T^{-1})$ by part (i), and $a_{\gamma^0} = O_p(T^{-1/2})$ by standard maximum-likelihood asymptotics for the fixed-threshold model at γ^0 : the score at (φ_0, γ^0) is a stationary ergodic martingale difference with finite second moments (Theorem 1), the normalized Hessian converges to the negative-definite information matrix (the ULLN of Lemma 5 together with Lemma 3(iii)), and

a standard sphere argument (cf. [Samia and Chan, 2011](#), proof of Lemma 5, displays (A29)–(A37) therein) yields the rate. Note this is a fixed- γ^0 statement, logically prior to Theorem 3. Decompose $\bar{H}_{\gamma_\kappa} - \bar{H}_{\gamma^0} = D_\kappa^{ind} + D_\kappa^{sm}$, where D_κ^{ind} collects the terms generated by indicator reallocation and D_κ^{sm} the effect of the changed evaluation path in φ . Each Hessian summand is dominated by $C\Lambda_t$ on the compact parameter set, so $\sup_{|\kappa| \leq K} \|D_\kappa^{ind}\| \leq C \sum_t \Lambda_t \mathbb{I}\{q_{t-d} \in \mathcal{B}_T\} = O_p(1)$ as in part (i); and the Gaussian log-likelihood has third φ -derivatives dominated by an integrable envelope on $\mathcal{B} \times \mathcal{S}$, so $\varphi \mapsto H(\varphi, \gamma^0)$ is Lipschitz with random constant $O_p(T)$ on the relevant neighborhood, giving $\sup_{|\kappa| \leq K} \|D_\kappa^{sm}\| = O_p(T) \cdot O_p(T^{-1/2}) = O_p(T^{1/2})$. Expanding (A.32),

$$2|\mathcal{L}_T(\kappa) - \mathcal{L}_T^0(\kappa)| \leq 2\|a_{\gamma^0}\| \|\bar{H}_{\gamma^0}\| \|r_\kappa\| + \|r_\kappa\|^2 \|\bar{H}_{\gamma^0}\| + (\|a_{\gamma^0}\| + \|r_\kappa\|)^2 (\|D_\kappa^{ind}\| + \|D_\kappa^{sm}\|),$$

and with $\|\bar{H}_{\gamma^0}\| = O_p(T)$ (ULLN on the compact set) the three terms are, respectively, $O_p(T^{-1/2}) \cdot O_p(T) \cdot O_p(T^{-1}) = O_p(T^{-1/2})$, $O_p(T^{-2})O_p(T) = O_p(T^{-1})$, and $O_p(T^{-1})(O_p(1) + O_p(T^{1/2})) = O_p(T^{-1/2})$ — all $o_p(1)$, uniformly over $|\kappa| \leq K$ since every bound above was uniform. The decoupled version is identical over the box, with the feasible-GLS profile in place of the ML profile and the one-step equivalence of Appendix A.5 (score expansion, Steps 3–4 there) guaranteeing that the feasible and infeasible profiled criteria differ by $o_p(1)$ at this scale; there, the role of a_{γ^0} is played by the known-thresholds infeasible GLS estimator at $(\gamma^{\mu,0}, \gamma^{\sigma,0})$, whose $\sqrt{T}$ -rate follows from the same fixed-threshold argument, and the feasible–infeasible gap is $o_p(T^{-1/2})$ by the score expansion of Appendix A.5, Steps 3–4. $\square$

A.4.2 Proof of Theorem 4(iii)

Throughout, $\hat{\gamma}$ denotes the smallest maximizer (Section 2.5), and by Theorem 4(ii) all statements may be, and are, restricted to an arbitrarily small neighborhood of θ_0 , exactly as in Proposition 1 of [Chan \(1993\)](#) and Theorem 2 of [Samia and Chan \(2011\)](#). The proof invokes Lemma 3(ii) and, as an alternative to **B5**, Lemma 4(i) of Appendix A.2.

Part (i) - super consistency of the threshold

Steps 1–4: localization and reduction. For $h \in \mathbb{R}$ define the rescaled local process $\mathcal{L}_T(h) := T\{L_T^p(\gamma^0 + h/T) - L_T^p(\gamma^0)\}$, with L_T^p the profile criterion of Section 2.5. Perturbing γ^0 to $\gamma^0 + h/T$ reassigns only observations with q_{t-d} in a band of width $|h|/T$ at γ^0 ; their number $N_T(h/T) := \sum_{t=1}^T \mathbb{I}\{\gamma^0 \leq q_{t-d} < \gamma^0 + h/T\}$ has expectation $T \int_{\gamma^0}^{\gamma^0 + h/T} f_\pi(u) du \rightarrow f_\pi(\gamma^0) h$ by the density clause of **B2** and Lemma 3(ii), hence is $O_p(1)$ by Markov’s inequality (and in fact Poisson in the limit, Theorem 4(iv)): at the $1/T$ scale the band contains finitely many observations, which is what motivates the T -rescaling. Each reassigned observation is evaluated under the adjacent regime’s

conditional Gaussian law, which differs from the true one with positive conditional probability at γ^0 (the mean clause of **B2** and **B3**) and in a neighborhood of it (**B7**), so its expected contribution is minus a Kullback–Leibler divergence between distinct Gaussian laws, strictly negative. Observations outside the band enter only through the profiled estimates $(\hat{B}(\gamma), \hat{\Sigma}(\gamma))$: by Lemma 7(ii) the normalized profile increment coincides, uniformly over $|h| \leq K$, with the increment at the true regime parameters up to $o_p(1)$, because the profiled estimates satisfy the inner first-order conditions and move by only $O_p(T^{-1})$ across the band (Lemma 7(i)). The argmax of $\mathcal{L}_T$ is therefore $O_p(1)$ once the switched-observation sum is shown to be uniformly negative outside a K/T -neighborhood, uniformly over the regime parameters; this is Step 5.

Step 5: Conclusion for the threshold rate. By Theorem 4(ii), with probability approaching one $\hat{\theta}_T$ lies in $\Omega_1(\bar{\Delta}) := \{\theta : \|B_i - B_i^0\| \leq \bar{\Delta}, \|\Sigma_i - \Sigma_i^0\| \leq \bar{\Delta}, i = 1, 2; |\gamma - \gamma^0| \leq \Delta\}$ for any fixed $\bar{\Delta}, \Delta > 0$. We show: for every $\varepsilon > 0$ there exist $K > 0$ and small $\bar{\Delta}, \Delta$ such that, with probability greater than $1 - \varepsilon$,

$$\theta \in \Omega_1(\bar{\Delta}) \text{ and } K/T < \gamma - \gamma^0 \leq \Delta \implies L_T(\varphi, \gamma) - L_T(\varphi, \gamma^0) < 0,$$

with the mirror statement for $\gamma < \gamma^0$; since $L_T^p(\hat{\gamma}) \geq L_T^p(\gamma^0) \geq L_T(\hat{\varphi}(\hat{\gamma}), \gamma^0)$, this forces $|\hat{\gamma} - \gamma^0| \leq K/T$ with probability greater than $1 - \varepsilon$, i.e. $T(\hat{\gamma} - \gamma^0) = O_p(1)$.

Fix $\varphi \in \Omega_1(\bar{\Delta})$ and $r := \gamma - \gamma^0 \in (K/T, \Delta]$. For fixed φ , the two log-likelihoods $L_T(\varphi, \gamma^0 + r)$ and $L_T(\varphi, \gamma^0)$ assign every observation to the same regime except those with $q_{t-d} \in [\gamma^0, \gamma^0 + r)$, whose contribution switches from the regime-2 to the regime-1 conditional Gaussian log-density; all other terms cancel. Writing $m_i(y_t; \varphi)$ for the regime- i log-density at candidate parameters and $M_t(\varphi) := m_1(y_t; \varphi) - m_2(y_t; \varphi)$,

$$T[L_T(\varphi, \gamma^0 + r) - L_T(\varphi, \gamma^0)] = \sum_{t=1}^T M_t(\varphi) \mathbb{I}\{\gamma^0 \leq q_{t-d} < \gamma^0 + r\} = \sum_{t=1}^T [\xi_t^+ + \{M_t(\varphi) - M_t(\varphi_0)\}] \mathbb{I}_t(r),$$

where $\mathbb{I}_t(r) := \mathbb{I}\{\gamma^0 \leq q_{t-d} < \gamma^0 + r\}$ and, on the compact $\mathcal{B} \times \mathcal{S}$, $|M_t(\varphi) - M_t(\varphi_0)| \leq C \Lambda_t \bar{\Delta}$ by the Gaussian log-density's Lipschitz property in φ with envelope Λ_t . Let $Q(r) := P_\pi(\gamma^0 \leq q_{t-d} < \gamma^0 + r)$. Dividing by $TQ(r)$,

$$\frac{T[L_T(\varphi, \gamma^0 + r) - L_T(\varphi, \gamma^0)]}{TQ(r)} \leq \frac{\sum_t \xi_t^+ \mathbb{I}_t(r)}{TQ(r)} + C\bar{\Delta} \frac{\sum_t \Lambda_t \mathbb{I}_t(r)}{TQ(r)}. \quad (\text{A.20})$$

Claim (uniform ratio control; the analogue of Claim I in Samia and Chan (2011)). For all $\varepsilon, \zeta > 0$ there exists $K > 0$ such that, for all T , with probability greater than $1 - \varepsilon$, uniformly over $r \in (K/T, \Delta]$:

$$(i) \quad \left| \sum_t \mathbb{I}_t(r) / (TQ(r)) - 1 \right| < \zeta;$$

- (ii) $|\sum_t \Lambda_t \mathbb{I}_t(r) - E \sum_t \Lambda_t \mathbb{I}_t(r)|/(TQ(r)) < \zeta$, with $E[\Lambda_t \mathbb{I}_t(r)] \leq M_\Lambda Q(r)$;
- (iii) $|\sum_t \xi_t^+ \mathbb{I}_t(r) - E \sum_t \xi_t^+ \mathbb{I}_t(r)|/(TQ(r)) < \zeta$.

Proof of the Claim. The ingredients are the analogues of displays (A22)–(A28) of [Samia and Chan \(2011, online appendix\)](#). First, after shrinking Δ if necessary, there exist $0 < m < \bar{M} < \infty$ with $mr \leq Q(r) \leq \bar{M}r$ for all $r \in (0, \Delta]$: the upper bound is Lemma 3(ii) ($f_\pi \leq C_*$), and the lower bound follows since f_π is positive (Lemma 3(ii)) and continuous at γ^0 (the density clause of **B2**), hence bounded away from zero on $[\gamma^0, \gamma^0 + \Delta]$ for small Δ . Second, $\text{var}(\mathbb{I}_t(r)) \leq Q(r)$, and by **B6** together with $|\xi_t^+| \leq C\Lambda_t$, $E[\Lambda_t^2 \mathbb{I}_t(r)] \leq MQ(r)$ and $E[(\xi_t^+)^2 \mathbb{I}_t(r)] \leq C^2 MQ(r)$, whence the same bounds hold for the variances. Third, all three summand processes are measurable functions of the localized pair in **B5**; by the covariance inequality for ρ -mixing sequences ([Doukhan, 1994](#), p. 9) and summability of the mixing coefficients, the variances of the *sums* are bounded by $THQ(r)$ for some $H < \infty$, uniformly in $r \in (0, \Delta]$ (in model (10), Lemma 4(i) delivers the same variance bounds directly from Theorem A1, without **B5**). The uniform statements (i)–(iii) then follow by Chebyshev bounds on the dyadic grid $r_j = b^j K/T$, $j = 0, 1, \dots$, combined with the monotonicity of $r \mapsto \sum_t(\cdot) \mathbb{I}_t(r)$ and $r \mapsto Q(r)$ to interpolate between grid points, exactly as in [Chan \(1993, p. 529, displays \(4.9a\)–\(4.9e\)\)](#); K large makes all grid bounds summable to less than ε . $\diamond$

Drift. By **B7** and **B6**, the map $v \mapsto E[\xi_t^+ | q_{t-d} = v] = -E[\text{KL}(N(B_2^0 x_t, \Sigma_2^0) \| N(B_1^0 x_t, \Sigma_1^0)) | q_{t-d} = v]$ is continuous on $[\gamma^0, \gamma^0 + \Delta]$ (shrinking Δ if necessary) and strictly negative at $v = \gamma^0$ by the mean clause of **B2** and **B3**; hence $E[\xi_t^+ | q_{t-d} = v] \leq -\chi < 0$ on the band, and $E \sum_t \xi_t^+ \mathbb{I}_t(r) \leq -\chi TQ(r)$.

Combining the Claim with the drift bound in (A.20): with probability greater than $1 - \varepsilon$, uniformly over $\varphi \in \Omega_1(\bar{\Delta})$ and $r \in (K/T, \Delta]$,

$$\frac{T[L_T(\varphi, \gamma^0 + r) - L_T(\varphi, \gamma^0)]}{TQ(r)} \leq -\chi(1 - \zeta) + \zeta + C\bar{\Delta}(M_\Lambda + \zeta) < 0$$

upon choosing ζ and $\bar{\Delta}$ small enough that $\chi(1 - \zeta) > \zeta + C\bar{\Delta}(M_\Lambda + \zeta)$. The case $\gamma < \gamma^0$ is symmetric with ξ_t^- in place of ξ_t^+ . This establishes the uniform negativity, and hence

$$T(\hat{\gamma} - \gamma^0) = O_p(1).$$

Part (ii) - $\sqrt{T}$ -consistency of regime parameters

Step 6: Negligible impact of threshold estimation on regime parameters. Let $s_{it}(\gamma)$ denote the regime- i score contribution for parameters (B_i, Σ_i) evaluated at threshold γ . The number of observations that switch regimes when γ moves from γ^0 to $\hat{\gamma}$ is, by the band count of Steps 1–4,

of order $O_p(1)$ because $|\hat{\gamma} - \gamma^0| = O_p(T^{-1})$. Hence

$$\sum_{t=1}^T \mathbb{I}\{q_{t-d} \text{ misclassified under } \hat{\gamma}\} = O_p(1).$$

Each individual score term satisfies $\mathbb{E}_\pi \|s_{it}(\gamma^0)\|^2 < \infty$ by the fourth-moment bound (Theorem 1). Therefore

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T s_{it}(\hat{\gamma}) \mathbb{I}\{q_{t-d} \text{ misclassified under } \hat{\gamma}\} = O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1).$$

That is, the score contribution from the small number of misclassified observations is asymptotically negligible at the $\sqrt{T}$ scale. This result mirrors Lemma 5 of [Samia and Chan \(2011\)](#) (see also [Hansen \(2000\)](#) Lemma 1), showing that the plug-in of the estimated threshold does not affect first-order asymptotics; it is the converse direction of Lemma 7(i) in Appendix A.4.

Step 7: $\sqrt{T}$ -rate of the regime-specific estimators. Given Step 6, the usual ML first-order expansion within each regime holds at $\hat{\gamma}$:

$$0 = \frac{1}{\sqrt{T}} \sum_{t=1}^T s_{it}(\gamma^0) + \mathcal{I}_i \sqrt{T} (\hat{\beta}_i - \beta_i^0) + o_p(1), \quad \mathcal{I}_i := -\mathbb{E}_\pi [\partial s_{it}(\gamma^0) / \partial \beta_i'],$$

so that $\sqrt{T}(\hat{\beta}_i - \beta_i^0) = \mathcal{I}_i^{-1} T^{-1/2} \sum_{t=1}^T s_{it}(\gamma^0) + o_p(1) = O_p(1)$ by the central limit theorem for geometrically ergodic sequences, and analogously for $\hat{\Sigma}_i$; the Gaussian limits themselves are derived in Step 8 of the proof of Theorem 4(iv). Because threshold estimation affects only $O_p(1)$ observations (Step 6), the rate is the same as with γ^0 known.

This completes the proof of Theorem 4(iii). $\square$

A.4.3 Proof of Theorem 4(iv)

The proof follows the architecture of Theorem 3 of [Samia and Chan \(2011\)](#), which in turn adapts Proposition 2 of [Chan \(1993\)](#); the weak-convergence machinery is the operator-convergence method of [Kushner \(1980, Theorem 1\)](#), implemented through the perturbed-test-function device of [Kurtz \(1975, p. 625\)](#), with uniqueness of the limiting martingale problem from [Stroock and Varadhan \(1971\)](#); [Kushner \(1984, Ch. 5\)](#), cited by [Chan \(1993\)](#), presents the same circle of ideas. We spell out the steps and the exact points at which each assumption enters, and cite the two papers for the parts of the argument that transfer verbatim.

Step 1: reduction to the true-parameter local process. By Lemma 7(ii), $\sup_{|\kappa| \leq K} |\mathcal{L}_T(\kappa) - \mathcal{L}_T^0(\kappa)| = o_p(1)$ for every K , so it suffices to establish the weak convergence in part (i) for $\mathcal{L}_T^0$. This is the analogue of passing from (4.10) to (4.12) in [Chan \(1993\)](#) and of Lemma 6 in [Samia and Chan](#)

(2011).

Step 2: exact marked-point-process representation. Throughout this proof we treat a single threshold and drop its index; for the two-sided process J of Theorem 4(iv) we write J_1 for its left branch ($h < 0$) and J_2 for its right branch ($h > 0$), following Chan (1993). At the true regime parameters, perturbing the threshold reassigns observations exactly and the increments telescope:

$$\begin{aligned}\mathcal{L}_T^0(\kappa) &= \sum_{t=1}^T \xi_t^+ \mathbb{I}\{\gamma^0 \leq q_{t-d} < \gamma^0 + \kappa/T\}, & \kappa \geq 0, \\ \mathcal{L}_T^0(\kappa) &= \sum_{t=1}^T \xi_t^- \mathbb{I}\{\gamma^0 + \kappa/T \leq q_{t-d} < \gamma^0\}, & \kappa < 0,\end{aligned}\tag{A.33}$$

with no remainder. Thus $\mathcal{L}_T^0$ is a marked point process: rescaled locations $T(q_{t-d} - \gamma^0)$ within $[0, \kappa)$ (resp. $[\kappa, 0)$), marks $\xi_t^\pm$.

Step 3: tightness. By Lemma 3(ii), for every interval $I \subseteq [\gamma^0 - \Delta, \gamma^0 + \Delta]$ and all $t \neq s$,

$$P(q_{t-d} \in I, q_{s-d} \in I) \leq C_*^2 |I|^2, \tag{A.34}$$

the analogue of display (4.13) in Chan (1993); together with the first-moment bound $E[|\xi_t^\pm| \mathbb{I}\{q_{t-d} \in I\}] \leq C E[\Lambda_t \mathbb{I}\{q_{t-d} \in I\}] \leq CM^{1/2} C_* |I|$ (B6 and Lemma 3(ii)), (A.34) yields tightness of the pair $(\mathcal{L}_T^0(\cdot), \mathcal{L}_T^0(\cdot))$ in $D[0, \infty)^2$, by the argument in the proof of Proposition 2 of Chan (1993) (following Lemma 5.2 of Ibragimov and Has'minskii, 1981).

Step 4: finite-dimensional convergence and identification of the limit. We verify the three ingredients of the direct-averaging/martingale-problem argument of Chan (1993, proof of Proposition 2, displays (4.14)–(4.23)), as adapted by Samia and Chan (2011, proof of Theorem 3):

- (a) Local band rates. Uniformly over histories, $P(\gamma^0 \leq q_{j-d} < \gamma^0 + z\varepsilon \mid \mathcal{F}_{j-d-1}) \leq C_* z\varepsilon$ by Lemma 3(i) — the analogue of Chan's (4.17) — and, in the averaged sense required by the martingale problem,

$$\lim_{j \rightarrow \infty} P(\gamma^0 \leq q_{j-d} < \gamma^0 + z\varepsilon) / \varepsilon = f_\pi(\gamma^0) z + o(\varepsilon)$$

by continuity and positivity of f_π at γ^0 (the density clause of B2 and Lemma 3(ii)) — the analogue of Chan's (4.18). The removal of the conditioning on the remote past in these displays is where the local ρ -mixing B5 does its work, exactly as Assumption 7 does in Samia and Chan (2011); under the alternative route flagged in the Remark to B5, the total-variation forgetting bound of our Theorem A1 (Chan's Condition 1 with $g(x) = 1 + x'Px$) substitutes.

- (b) Convergence of the conditional mark laws. The marks are $\xi_t^\pm = \Phi^\pm(x_t, \varepsilon_t)$ for continuous maps $\Phi^\pm$ (display (J1)). Since $\varepsilon_t \perp \mathcal{F}_{t-1}$, the conditional law of $(x_t^{(-)}, \varepsilon_t)$ given $\gamma^0 \leq q_{t-d} < \gamma^0 + \delta$ equals (conditional law of $x_t^{(-)}$ given the band) $\otimes N(0, I_n)$, and by B7 the first factor converges weakly, as $\delta \downarrow 0$, to the conditional law at $q_{t-d} = \gamma^0$. By the continuous mapping theorem the conditional law of ξ_t^+ given the band converges weakly to μ^+ , and B6 supplies the uniform integrability needed to pass conditional expectations of the truncated test functions to the limit — the analogue of Chan’s (4.19) and the role of Assumption 6 in Samia and Chan (2011).
- (c) Averaging and the martingale problem. With (a), (b), and **B5** in hand (in model (10), Lemma 4(ii) supplies the role of **B5** here directly, without the mixing condition), the operator-convergence argument of Kushner (1980, Theorem 1)– implemented via the perturbed test functions of Kurtz (1975, p. 625), and applied exactly as in Chan (1993) and in the proof of Theorem 3 of Samia and Chan (2011) – identifies every subsequential weak limit of $\{\mathcal{L}_T^0(\kappa), 0 \leq \kappa \leq K\}$ as the unique solution (Stroock and Varadhan, 1971) of the martingale problem with generator $Af(w) = f_\pi(\gamma^0) z \int \{f(w+u) - f(w)\} \mu^+(du)$, i.e. as the compound Poisson process J_2 with rate $f_\pi(\gamma^0)$ and jump law μ^+ ; the Cramér–Wold device extends this to the finite-dimensional distributions of the pair $(\mathcal{L}_T^0(-\cdot), \mathcal{L}_T^0(\cdot))$. The two components are driven by observations in the disjoint bands $[\gamma^0, \gamma^0 + K/T)$ and $[\gamma^0 - K/T, \gamma^0)$, so the bivariate martingale problem factorizes and the limit pair (J_1, J_2) has independent components; equivalently, the limiting marked point process on the two half-lines is a Poisson random measure, whose restrictions to disjoint carriers are independent (Kallenberg, 2002, Ch. 12). Combined with Step 3, this proves the process convergence of Theorem 4(iv).

The convergence just established can be recorded formally as a marked-point-process statement, which the proof of Theorem 3(ii)–(iii) invokes componentwise. The proof below uses only the conclusions of Steps 2–4.

Lemma MPP. *Fix $K > 0$. Define the random point measures on $E_K := (0, K] \times \mathbb{R}$,*

$$N_T^+ := \sum_{t=1}^T \delta_{(T(q_{t-d}-\gamma^0), \xi_t^+)} \mathbb{I}\{0 < T(q_{t-d}-\gamma^0) \leq K\}, \quad N_T^- := \sum_{t=1}^T \delta_{(T(\gamma^0-q_{t-d}), \xi_t^-)} \mathbb{I}\{0 < T(\gamma^0-q_{t-d}) \leq K\},$$

*regarded as random elements of the space $\mathcal{N}(E_K)$ of finite point measures on E_K , endowed with the topology of weak convergence of measures and the corresponding Borel σ -field. Under **A1–A4** and **B1–B7**,*

$$(N_T^+, N_T^-) \Rightarrow (N^+, N^-),$$

where N^+ and N^- are independent Poisson random measures on E_K with intensity measures $f_\pi(\gamma^0) \text{Leb}_{(0,K]} \otimes \mu^+$ and $f_\pi(\gamma^0) \text{Leb}_{(0,K]} \otimes \mu^-$ respectively. Equivalently, the associated jump pro-

cesses $\kappa \mapsto \int_{(0,\kappa] \times \mathbb{R}} u N_T^\pm(dv, du)$ converge jointly in $D[0, K]^2$ to the compound Poisson processes (J_2, J_1) of Theorem 4(iv) restricted to $[0, K]$.

Proof. By Steps 2–4, the pair of one-sided normalized criterion processes converges in $D[0, K]^2$ to the compound Poisson pair (J_2, J_1) . The same direct-averaging/martingale-problem computation applies identically to the enlarged pair (count, sum) on each side: the generator displays of Step 4 depend on the state only through the arrival rates and the mark laws (ingredients (a)–(c)) and the counting component is the special case of a mark identically equal to one, so no new verification is required and the four processes (counts and sums, both sides) converge jointly, the limit counts being Poisson with rate $f_\pi(\gamma^0)$ and the limit sums their $\mu^\pm$ -marked compounds. A path pair (count, sum) with finitely many jumps and distinct jump positions determines the marked point measure – positions are the count’s jump times, marks the sum’s increments there – and this correspondence is continuous, in the weak topology of $\mathcal{N}(E_K)$, at every configuration with finitely many points and distinct positions, an almost-sure property of the Poisson limit. The continuous mapping theorem then yields $(N_T^+, N_T^-) \Rightarrow (N^+, N^-)$ with the stated intensities, the independence of the two sides being that of the limiting pairs established in Step 4 (their points are carried by disjoint sets of observations, display (A.34)). $\square$

Step 5: drift. By (J2) and iterated expectations, the means of the jump laws are

$$\int u \mu^+(du) = -E\left[\text{KL}(N(B_2^0 x_t, \Sigma_2^0) \parallel N(B_1^0 x_t, \Sigma_1^0)) \mid q_{t-d} = \gamma^0\right] < 0,$$

strictly, because the mean clause of B2 gives positive conditional probability, at the threshold, that the adjacent conditional Gaussian laws differ (and B3 gives $\Sigma_1^0 \neq \Sigma_2^0$ when the means coincide), while B7 carries this conditional separation to the one-sided limit laws, and the Gaussian KL divergence (J3) is zero only when both the means and the covariances coincide; symmetrically for μ^- . Both means are finite by B6. The random walks of jump sums therefore tend to $-\infty$ almost surely by the strong law of large numbers, so $\mathcal{J}(\kappa) := J(\kappa) \rightarrow -\infty$ (we write $\mathcal{J}$ for the limit process J of Theorem 4(iv) in this appendix) as $|\kappa| \rightarrow \infty$ almost surely.

Step 6: almost-sure uniqueness of the argmax interval. By the limit-objects paragraph of Section 2.4, the nonzero parts of $\mu^\pm$ are absolutely continuous, so the running maxima of the thinned compound Poisson processes are almost surely tie-free, and the global maximum of $\mathcal{J}$ is attained on an almost-surely unique interval $[M_-, M_+]$; this is the multivariate-Gaussian analogue of Remark 1 in Samia and Chan (2011).

Step 7: argmax. Theorem 4(iii) gives $T(\hat{\gamma} - \gamma^0) = O_p(1)$: the maximizer of the profile likelihood over Γ lies, with probability approaching one, within $[\gamma^0 - K/T, \gamma^0 + K/T]$ for K large. On that event, $T(\hat{\gamma} - \gamma^0)$ is the smallest maximizer of $\kappa \mapsto \mathcal{L}_T(\kappa)$ over $[-K, K]$. Using the Skorohod representation to realize the convergence of part (i) almost surely, and the almost-sure uniqueness

from Step 6, the smallest maximizer of $\mathcal{L}_T$ converges to the left endpoint M_- of the maximizing interval of $\mathcal{J}$ — exactly the concluding argument in the proof of Theorem 2 of [Chan \(1993\)](#) and of Theorem 3 of [Samia and Chan \(2011\)](#). Hence $T(\hat{\gamma} - \gamma^0) \Rightarrow M_-$, which proves the argmax statement of Theorem 4(iv).

Step 8: regime parameters and asymptotic independence. By Lemma 7(i) and $T(\hat{\gamma} - \gamma^0) = O_p(1)$,

$$\hat{\varphi}(\hat{\gamma}) = \hat{\varphi}(\gamma^0) + o_p(T^{-1/2}), \quad (\text{A.35})$$

the analogue of display (4.24) in [Chan \(1993\)](#): the two-step estimator is first-order equivalent to the known-threshold ML estimator. It therefore suffices to derive the limit of $\hat{\varphi}(\gamma^0)$. The known-threshold scores, evaluated at θ_0 and written with the true innovations $\varepsilon_t^0 := y_t - B_i^0 x_t = \Sigma_i^{0,1/2} \varepsilon_t$ on regime i ,

$$s_t^{B_i} = (x_t \otimes I_n) \Sigma_i^{0,-1} \varepsilon_t^0 \mathbb{I}\{q_{t-d} \in \mathcal{R}_i^0\}, \quad s_t^{\Sigma_i} = \frac{1}{2} D_n' \text{vec} \left(\Sigma_i^{0,-1} (\varepsilon_t^0 \varepsilon_t^{0'} - \Sigma_i^0) \Sigma_i^{0,-1} \right) \mathbb{I}\{q_{t-d} \in \mathcal{R}_i^0\},$$

are stationary, ergodic martingale difference sequences with respect to $\{\mathcal{F}_t\}$ (the indicators and x_t are $\mathcal{F}_{t-1}$ -measurable; $\varepsilon_t \perp \mathcal{F}_{t-1}$) with finite second moments (Theorem 1). By the martingale-difference central limit theorem for stationary ergodic sequences ([Billingsley, 1961](#)) and the Cramér–Wold device, the stacked normalized score converges to a centered Gaussian vector with covariance equal to the information matrix, which is block diagonal: across regimes because the indicators are disjoint, and between the B - and Σ -blocks within a regime because the corresponding cross moments involve odd Gaussian moments of ε_t , which vanish. The standard ML expansion within each regime (as in Step 7 of the proof of Theorem 4(iii)) then delivers the stated Gaussian limits: the B_i -block information is $E_\pi[x_t x_t' \mathbb{I}_i] \otimes \Sigma_i^{0,-1} = S_i \otimes \Sigma_i^{0,-1}$, with inverse $S_i^{-1} \otimes \Sigma_i^0$; and $\sqrt{T_i} \text{vech}(\hat{\Sigma}_i - \Sigma_i^0) \Rightarrow N(0, 2D_n^+(\Sigma_i^0 \otimes \Sigma_i^0)D_n^{+'})$ with $T_i/T \rightarrow_{a.s.} \pi_i$ gives the stated Σ -block. Finally, the asymptotic independence of $T(\hat{\gamma} - \gamma^0)$ from the Gaussian block follows as in [Chan \(1993, proof of Theorem 2\)](#) and [Samia and Chan \(2011, Theorem 4\)](#): by (A.35) the Gaussian block is a functional of the full-sample score partial sums, while by Steps 1–7 the threshold limit is a functional of the $O_p(1)$ observations in the shrinking bands, whose contribution to the normalized score is $o_p(1)$ (Step 6 of the proof of Theorem 4(iii)); the joint limit therefore factorizes. $\square$

Proof of Theorem 4(iv) (sketch). For $k > 2$, the true thresholds are separated by fixed positive distances (B1), so for K fixed and T large the $k - 1$ boundary bands $[\gamma_j^0 - K/T, \gamma_j^0 + K/T)$ are disjoint, and only adjacent-regime pairs are involved in each. Steps 1–4 apply to the $2(k - 1)$ one-sided component processes simultaneously; the limiting marked point process on the $2(k - 1)$ half-lines is a Poisson random measure whose restrictions to the disjoint carriers are independent,

so the $k - 1$ two-sided processes $\mathcal{J}^{(j)}$ are mutually independent, each built from the adjacent pair $(j, j + 1)$ with rate $f_\pi(\gamma_j^0)$. Steps 5–8 apply coordinatewise; the joint argmax convergence follows from the continuous mapping theorem applied coordinatewise on the product space. The result parallels Theorem 2 of [Li and Ling \(2012\)](#) for the least-squares estimator of multiple-regime TAR models, where the same asymptotic independence of the multiple threshold estimators is established. $\square$

A.5 Rates and Limit Laws: the Decoupled Case

A.5.1 Proof of Theorem 3(i)

The proof extends that of Theorem 4(iii) to the product parameter space and to the feasible profile-GLS estimator.

Part (i): super-consistency of $\hat{\gamma}^\mu$ and $\hat{\gamma}^\sigma$. We prove the rate jointly. For $h_\mu \in \mathbb{R}^{k_\mu-1}$ and $h_\sigma \in \mathbb{R}^{k_\sigma-1}$ define the local process

$$\mathcal{L}_T(h_\mu, h_\sigma) := T \left[L_T^p(\gamma^{\mu,0} + h_\mu/T, \gamma^{\sigma,0} + h_\sigma/T) - L_T^p(\gamma^{\mu,0}, \gamma^{\sigma,0}) \right]; \quad (\text{A.28})$$

it suffices to show that its joint maximizer is $O_p(1)$.

Step 1: local counting argument. For fixed compact sets $K_\mu \subset \mathbb{R}^{k_\mu-1}$, $K_\sigma \subset \mathbb{R}^{k_\sigma-1}$, define the boundary bands $\Delta_T^\mu(h_\mu) := \bigcup_{j=1}^{k_\mu-1} [\gamma_j^{\mu,0} \wedge (\gamma_j^{\mu,0} + h_{\mu,j}/T), \gamma_j^{\mu,0} \vee (\gamma_j^{\mu,0} + h_{\mu,j}/T)]$ and, analogously, $\Delta_T^\sigma(h_\sigma)$. Only observations with $q_{t-d_\mu}^\mu \in \Delta_T^\mu(h_\mu)$ or $q_{t-d_\sigma}^\sigma \in \Delta_T^\sigma(h_\sigma)$ can change cell when the threshold pair is perturbed from $(\gamma^{\mu,0}, \gamma^{\sigma,0})$ to $(\gamma^{\mu,0} + h_\mu/T, \gamma^{\sigma,0} + h_\sigma/T)$, so the number of switched observations is at most

$$N_T(h_\mu, h_\sigma) := \sum_{t=1}^T \mathbb{I}\{q_{t-d_\mu}^\mu \in \Delta_T^\mu(h_\mu)\} + \sum_{t=1}^T \mathbb{I}\{q_{t-d_\sigma}^\sigma \in \Delta_T^\sigma(h_\sigma)\},$$

which is $O_p(1)$ uniformly over $(h_\mu, h_\sigma) \in K_\mu \times K_\sigma$: by the density clauses of **B2** each band has stationary probability $O(1/T)$, so $EN_T = O(1)$ and Markov's inequality applies.

Step 2: local likelihood loss. If $h_{\mu,j} \neq 0$, the observations in the j -th mean band are assigned to the wrong pair of adjacent mean regimes, whose conditional means differ at $\gamma_j^{\mu,0}$ (the mean clause of **B2**); if $h_{\sigma,\ell} \neq 0$, those in the ℓ -th variance band are evaluated under the wrong adjacent covariance matrices (the variance clause of **B2**). Each switched observation thus incurs a non-vanishing expected Kullback–Leibler loss, so for every fixed $(h_\mu, h_\sigma) \neq (0, 0)$ the local criterion is driven by the $O_p(1)$ switched observations of Step 1, each contributing a loss of the sign required by the classical localization argument. The uniform ratio control follows exactly as in the Claim of

Step 5 of the proof of Theorem 4(iii), applied to each of the $(k_\mu - 1) + (k_\sigma - 1)$ bands separately – the band sums are combined by the triangle inequality, so no cross-band covariances are needed – with **B5** and **B6** supplying the covariance and moment bounds; for threshold variables of the class described in the remark after **B5**, Lemma 4(i) supplies these bounds without **B5**.

Step 3: control of the remainder. Observations outside the boundary bands do not change cell and affect the criterion only through the re-estimated regime parameters. By Lemma 7 (decoupled clauses), the profiled estimates move by $O_p(T^{-1})$ uniformly over the box (part (i)) and, since the score vanishes at the profiled maximizer, the induced criterion change is quadratic in this movement against an $O_p(T)$ Hessian (part (ii)); hence these terms are $o_p(1)$ after multiplication by T , uniformly over (h_μ, h_σ) in compact sets.

Step 4: localization. By Steps 1–3 the local process admits the localization mechanism of Chan (1993), Hansen (2000) and Samia and Chan (2011), applied jointly to the product threshold parameter: for every $\varepsilon > 0$ there is $M < \infty$ such that $P(\sup_{\|(h_\mu, h_\sigma)\| > M} \mathcal{L}_T(h_\mu, h_\sigma) \geq \mathcal{L}_T(0, 0)) < \varepsilon$ for all sufficiently large T . Since $(\hat{\gamma}^\mu, \hat{\gamma}^\sigma)$ maximizes L_T^p over $\Gamma^\mu \times \Gamma^\sigma$, the usual argmax localization argument gives

$$T(\hat{\gamma}^\mu - \gamma^{\mu,0}) = O_p(1), \quad T(\hat{\gamma}^\sigma - \gamma^{\sigma,0}) = O_p(1). \quad (\text{A.31})$$

This proves part (i).

Part (ii): $\sqrt{T}$ -consistency of the regime parameters. Step 1. Preliminary covariances. Let $\hat{\Sigma}_\ell^{(0)}$ be the preliminary pooled estimator of Remark E.1 (Section 2.2). Evaluated at the true threshold pair, the ergodic theorem and the fourth-moment bound give $\hat{\Sigma}_\ell^{(0)} \rightarrow_{a.s.} \Sigma_\ell^0$ for $\ell = 1, \dots, k_\sigma$ and, since $\Sigma_\ell^0 \succ 0$ by **B1**, also $(\hat{\Sigma}_\ell^{(0)})^{-1} \rightarrow_{a.s.} (\Sigma_\ell^0)^{-1}$.

Step 2. Negligible misclassification. By Part (i) and the local counting argument of its Step 1, the number of observations misclassified in either threshold dimension is $O_p(1)$; since each score term has finite second moment (Theorem 1), their contribution to the normalized score is $O_p(T^{-1/2}) = o_p(1)$.

Step 3. GLS score expansion for the mean block. Let $s_{it}^{GLS} := (x_t \otimes \Sigma_{\ell(t)}^{-1}) \varepsilon_t \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu\}$, where $\ell(t) = \ell^*(X_{t-1})$ is evaluated at the true threshold. The feasible GLS first-order condition yields

$$0 = \frac{1}{\sqrt{T}} \sum_{t=1}^T s_{it}^{GLS}(B_i^0) - M_i^{GLS} \sqrt{T} (\hat{\beta}_i^{GLS} - \beta_i^0) + o_p(1),$$

with $M_i^{GLS} = E[(x_t x_t' \otimes \Sigma_{\ell(t)}^{-1}) \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu\}]$.

Step 4. FGLS plug-in error. Since $\max_{1 \leq \ell \leq k_\sigma} \|(\hat{\Sigma}_\ell^{(0)})^{-1} - (\Sigma_\ell^0)^{-1}\| = o_p(1)$, the feasible and infeasible GLS scores differ by $T^{-1/2} \sum_{t=1}^T (x_t \otimes [(\hat{\Sigma}_{\ell(t)}^{(0)})^{-1} - (\Sigma_{\ell(t)}^0)^{-1}]) \varepsilon_t \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^\mu\} = o_p(1)$, by Cauchy–Schwarz and the fourth-moment bound: one covariance update is first-order equivalent

to infeasible GLS.

Steps 5–6. Mean- and variance-block asymptotics. Solving the score expansion, $\sqrt{T} \text{vec}(\widehat{B}_i^{GLS} - B_i^0) = (M_i^{GLS})^{-1} T^{-1/2} \sum_{t=1}^T s_{it}^{GLS}(B_i^0) + o_p(1) = O_p(1)$, and a central limit theorem for geometrically ergodic processes yields the centered Gaussian limit stated in the theorem. For each variance regime ℓ , $\widehat{\Sigma}_\ell$ in (5) is a pooled sample covariance of residuals that differ from the true innovations only through $o_p(1)$ parameter-estimation error and $O_p(1)$ misclassified observations, so the usual sample-covariance expansion gives $\sqrt{T} \text{vech}(\widehat{\Sigma}_\ell - \Sigma_\ell^0) = O_p(1)$, $\ell = 1, \dots, k_\sigma$. This proves Part (ii). $\square$

A.5.2 Proof of Theorem 3(ii)-(iii)

The proof extends the proof of Theorem 4(iv) (Appendix A.4) to the product threshold parameter, with two new ingredients: the elimination of double-boundary observations (Step 2, using B8), which is what delivers the asymptotic independence of the two threshold estimators; and the observation that the variance-side marks are exact i.i.d. functionals of the innovations (Step 3), which simplifies the identification of the σ -side limit.

Step 1: reduction. By the decoupled version of Lemma 7(ii), uniformly over $(h_\mu, h_\sigma) \in [-K, K]^2$ the normalized feasible-GLS profile increment

$$\mathcal{L}_T(h_\mu, h_\sigma) := T \left[L_T^p(\gamma^{\mu,0} + h_\mu/T, \gamma^{\sigma,0} + h_\sigma/T) - L_T^p(\gamma^{\mu,0}, \gamma^{\sigma,0}) \right]$$

differs by $o_p(1)$ from its true-parameter counterpart $\mathcal{L}_T^0(h_\mu, h_\sigma)$, computed from the full Gaussian criterion ℓ_t of (2) in Appendix A.3 at $\varphi_0 = \{B_i^0, \Sigma_\ell^0\}$. (The one-covariance-update feasibility gap is $o_p(1)$ at this scale by the score expansion of Appendix A.5, Steps 3–4).

Step 2: exact decomposition and elimination of double switchers. At the true parameters, an observation contributes to $\mathcal{L}_T^0(h_\mu, h_\sigma)$ only if it changes cell assignment, i.e. only if $q_{t-d_\mu}^\mu$ lies in the μ -band $\mathcal{B}_T^\mu(h_\mu)$ of width $|h_\mu|/T$ at $\gamma^{\mu,0}$, or $q_{t-d_\sigma}^\sigma$ lies in the analogous σ -band $\mathcal{B}_T^\sigma(h_\sigma)$, or both. Splitting by which band is hit,

$$\mathcal{L}_T^0(h_\mu, h_\sigma) = \mathcal{L}_T^\mu(h_\mu) + \mathcal{L}_T^\sigma(h_\sigma) + R_T(h_\mu, h_\sigma), \quad (\text{A.36})$$

where $\mathcal{L}_T^\mu$ sums the mean jumps $\xi_t^{\mu,\pm}$ (display (V1), evaluated at the observation's true variance regime $\ell(t)$, which is unchanged for a μ -only switcher) over observations hitting only the μ -band; $\mathcal{L}_T^\sigma$ sums the variance jumps $\xi_t^{\sigma,\pm}$ (display (V2), with the observation's true mean, unchanged for a σ -only switcher) over observations hitting only the σ -band; and R_T collects the double switchers, whose individual contributions are bounded by $C\Lambda_t$. By B8,

$$P\left(\exists t \leq T : q_{t-d_\mu}^\mu \in \mathcal{B}_T^\mu(h_\mu), q_{t-d_\sigma}^\sigma \in \mathcal{B}_T^\sigma(h_\sigma) \text{ for some } |h_\mu|, |h_\sigma| \leq K\right) \leq T \cdot \sup f_{\mu\sigma} \cdot \frac{2K}{T} \cdot \frac{2K}{T} = O(T^{-1}),$$

so with probability approaching one there are no double switchers in the box and

$$\sup_{\|(h_\mu, h_\sigma)\|_\infty \leq K} |R_T| = 0$$

in particular $R_T = o_p(1)$ uniformly. This is the precise sense in which the two threshold problems decouple: membership in one band is an event of probability $O(1/T)$ per observation, which occurs a stable $O_p(1)$ number of times in T trials and drives the limit, whereas membership in both bands requires two coincidences for the same observation and, under **B8**, which rules out any singular concentration of (q^μ, q^σ) along a curve, has probability $O(1/T^2)$, so that not even one double event occurs in the limit. The assumption is sharp: on the diagonal $q^\mu \equiv q^\sigma$ (same variable, same delay) one band membership *implies* the other, the joint law is singular, the double-switcher count is $O_p(1)$ and the criterion no longer separates – that configuration is exactly the joint model of Section 2.5.

Step 3: marks. The σ -side marks $\xi_t^{\sigma, \pm}$ are, by (V2), measurable functions of ε_t and (Σ_1^0, Σ_2^0) alone; since $\{\varepsilon_t\}$ is i.i.d. and independent of the $(\mathcal{F}_{t-1}$ -measurable) band indicators, the σ -side marks are exactly i.i.d. with law $\mu^{\sigma, \pm}$, whatever the conditional distribution of the state at the threshold — no weak-continuity condition is needed, and ingredient (b) of Step 4 in Appendix A.4 is automatic on this side. The μ -side marks are $\xi_t^{\mu, \pm} = \Phi^{\mu, \pm}(x_t, \ell(t), \varepsilon_t)$ with $\Phi^{\mu, \pm}$ continuous in (x_t, ε_t) for each value of the discrete $\ell(t)$; by B7 the conditional law of $(x_t^{(-)}, \ell(t))$ given the μ -band converges weakly to its value at $q^\mu = \gamma^{\mu, 0}$, so the conditional mark law converges to the mixture $\mu^{\mu, \pm}$ of (V3), with uniform integrability from B6 — ingredient (b) on the μ side.

Step 4: joint weak convergence and independence. Ingredient (a) of Step 4 in Appendix A.4 holds at each threshold by the density clauses of **B2** (continuity and positivity of f_μ, f_σ) together with the conditional-density bounds (Lemma 3(i) when the state variables are coordinates of y_t ; maintained through the density clauses of **B2** and B6 in general); ingredient (c) holds by **B5**; for threshold variables of the class described in the remark after **B5**, Lemma 4(ii) applied at each band supplies it without **B5**. Applying Lemma MPP and the direct-averaging argument of Step 4 in Appendix A.4 componentwise to the four one-sided component processes simultaneously – left and right of each threshold – identifies the limit as a Poisson random measure on the disjoint union of four half-lines, with intensity $f_\mu(\gamma^{\mu, 0})$ on the two μ -components and $f_\sigma(\gamma^{\sigma, 0})$ on the two σ -components, and the respective mark laws from Step 3. The four driving observation sets are pairwise disjoint with probability approaching one (Step 2 for μ -vs- σ ; the left/right bands are disjoint by construction), so the four-component martingale problem factorizes and the limiting components are mutually independent (Kallenberg, 2002, Ch. 12). Together with

the tightness bound (A.34) – the pairwise band-occupancy bound of Step 3 in Appendix A.4, which by Lemma 3(i) holds verbatim for each of q^μ and q^σ , applied at each threshold – this gives the process convergence in part (ii), with $(\mathcal{J}^\mu, \mathcal{J}^\sigma)$ independent.

Step 5: drifts. On the μ side, using (V1), B1 ($\Sigma_\ell^{0,-1} \succeq \bar{\sigma}^{-1} I_n$) and the mean clause of **B2**,

$$\int u \mu^{\mu,+}(du) = -\frac{1}{2} E \left[\delta_t^{\mu} \Sigma_{\ell(t)}^{0,-1} \delta_t^{\mu} \mid q_{t-d_\mu}^\mu = \gamma^{\mu,0} \right] \leq -\frac{1}{2\bar{\sigma}} E \left[\|\delta_t^\mu\|^2 \mid q^\mu = \gamma^{\mu,0} \right] < 0,$$

strictly, since $P(\delta_t^\mu \neq 0 \mid q^\mu = \gamma^{\mu,0}) > 0$; symmetrically for $\mu^{\mu,-}$. On the σ side the drift is the deterministic constant $-\frac{1}{2} \sum_j (\lambda_j - \ln \lambda_j - 1) < 0$, strictly by the variance clause of **B2**. Finiteness from B6 (the σ -side mean is finite outright). Both pairs of random walks tend to $-\infty$ almost surely.

Step 6: uniqueness of the argmax intervals. The σ -jump laws are absolutely continuous with no atom: $\Sigma_1^0 \neq \Sigma_2^0$ implies $I_n - Q \neq 0$ (if $Q = I_n$ then $\Sigma_2^0 = \Sigma_1^0$), so at least one weight $1 - \lambda_j$ in (V2) is nonzero and $\xi^{\sigma,\pm}$ has a Lebesgue density. The μ -jump laws may place an atom at zero of mass $P(\delta_t^\mu = 0 \mid q^\mu = \gamma^{\mu,0}) < 1$, with absolutely continuous nonzero part (conditionally on $\{\delta_t^\mu \neq 0\}$ and $\ell(t)$, the jump is Gaussian with variance $\kappa_t > 0$); the thinning argument of the limit-objects paragraph of Section 2.4 applies. Hence each of $\mathcal{J}^\mu, \mathcal{J}^\sigma$ attains its global maximum on an almost-surely unique interval, $[M_-^\mu, M_+^\mu)$ and $[M_-^\sigma, M_+^\sigma)$.

Step 7: argmax. Theorem 3(i) localizes: $T(\hat{\gamma}^\mu - \gamma^{\mu,0})$ and $T(\hat{\gamma}^\sigma - \gamma^{\sigma,0})$ are $O_p(1)$, so with probability approaching one the smallest maximizer of the profile criterion over $\Gamma^\mu \times \Gamma^\sigma$ lies in the box $[-K, K]^2$ for K large. By (A.36) with $R_T = 0$ on an event of probability approaching one, the criterion is *additively separable* on the box, so its smallest maximizer is the pair of coordinatewise smallest maximizers of $\mathcal{L}_T^\mu$ and $\mathcal{L}_T^\sigma$. Applying the Skorohod representation and the argmax argument of Step 7 in Appendix A.4 to each coordinate, and using the independence from Step 4,

$$\left(T(\hat{\gamma}^\mu - \gamma^{\mu,0}), T(\hat{\gamma}^\sigma - \gamma^{\sigma,0}) \right) \Rightarrow (M_-^\mu, M_-^\sigma), \quad M_-^\mu \perp M_-^\sigma,$$

proving part (ii).

Step 8: part (iii). By the decoupled Lemma 7(i) and part (i), $\hat{\varphi}(\hat{\gamma}^\mu, \hat{\gamma}^\sigma) = \hat{\varphi}(\gamma^{\mu,0}, \gamma^{\sigma,0}) + o_p(T^{-1/2})$, and by Steps 3–4 of Appendix A.5 the feasible one-step GLS estimator at the true thresholds is first-order equivalent to the infeasible GLS estimator. The infeasible-GLS scores at θ_0 , with $\varepsilon_t^0 = \Sigma_{\ell(t)}^{0,1/2} \varepsilon_t$,

$$s_t^{B_i} = (x_t \otimes I_n) \Sigma_{\ell(t)}^{0,-1} \varepsilon_t^0 \mathbb{I}\{q_{t-d_\mu}^\mu \in \mathcal{R}_i^{\mu,0}\}, \quad s_t^{\Sigma_\ell} = \frac{1}{2} D_n' \text{vec} \left(\Sigma_\ell^{0,-1} (\varepsilon_t^0 \varepsilon_t^{0'} - \Sigma_\ell^0) \Sigma_\ell^{0,-1} \right) \mathbb{I}\{q_{t-d_\sigma}^\sigma \in \mathcal{R}_\ell^{\sigma,0}\},$$

are stationary ergodic martingale differences with finite second moments (Theorem 1), so the

martingale-difference CLT (Billingsley, 1961) with Cramér–Wold gives joint asymptotic normality; the information is block diagonal across mean regimes (disjoint μ -indicators), across variance regimes (disjoint σ -indicators), and between the B - and Σ -blocks (odd Gaussian moments vanish — note the B – Σ cross moments here mix indicators from *different* partitions, but the factor $E[\varepsilon_t \otimes (\varepsilon_t \varepsilon_t' - I)] = 0$ annihilates them regardless). The B_i -block information is M_i^{GLS} as defined in the theorem, giving $\text{avar}(M_i^{GLS})^{-1}$; the Σ_ℓ -block gives $\sqrt{T_\ell} \text{vech}(\hat{\Sigma}_\ell - \Sigma_\ell^0) \Rightarrow N(0, 2D_n^+(\Sigma_\ell^0 \otimes \Sigma_\ell^0)D_n^{+'})$ with $T_\ell/T \rightarrow \pi_\ell^\sigma$ (the pooling across mean regimes within variance regime ℓ is already reflected in T_ℓ), yielding the stated law. Asymptotic independence from the threshold pair follows exactly as in Step 8 of Appendix A.4: the Gaussian block is driven by full-sample score sums, the threshold pair by the $O_p(1)$ band observations, whose score contribution is $o_p(1)$ (Appendix A.5, Step 2). $\square$

B Additional Tables from Monte Carlo Exercise

This appendix collects the full statistics from the Monte Carlo experiments of Section 3.

Table A-1: Joint MC – full threshold statistics.

DGP	T	RMSE		MAD		$\hat{\gamma}$: bias / std	
		SSR	Joint	SSR	Joint	SSR	Joint
DGP2	200	0.025	0.025	0.017	0.017	−0.013 / 0.021	−0.013 / 0.022
	1000	0.004	0.004	0.003	0.003	−0.002 / 0.004	−0.002 / 0.004
DGP3	200	2.004	0.080	1.752	0.050	+1.728 / 1.015	−0.001 / 0.080
	1000	3.204	0.013	2.872	0.008	+2.865 / 1.435	+0.001 / 0.013
DGP4	200	0.081	0.039	0.035	0.025	+0.006 / 0.081	−0.012 / 0.037
	1000	0.010	0.006	0.006	0.004	+0.001 / 0.010	−0.001 / 0.006
T-rate ratio DGP2		SSR 5.50, Joint 5.62					
T-rate ratio DGP3		SSR 0.63, Joint 6.01					
T-rate ratio DGP4		SSR 8.42, Joint 6.00					

Notes. DGP1 omitted because the threshold is unidentified in the no-regime DGP, so RMSE and MAD are not meaningful quantities. T-rate ratio is $\text{RMSE}(T = 200)/\text{RMSE}(T = 1000)$; target is 5 for super-consistency at rate T .

Table A-2: DGP1 (no regime): parameter RMSE under estimated and fixed partitions

	T	Estim. split, equal weights		Estim. split, share weights		Median split	No split
		SSR	joint	SSR	joint		
RMSE $\hat{B}_i$	200	1.758	1.908	0.796	0.799	0.345	0.164
	1000	2.259	2.590	0.488	0.502	0.152	0.072
RMSE $\hat{\Sigma}_i$	200	0.360	0.495	0.247	0.291	0.245	0.174
	1000	0.263	0.426	0.111	0.135	0.110	0.078
$\ \hat{\Phi}_{1,0} - \hat{\Phi}_{2,0}\ $	200	1.704	1.849			0.299	
	1000	1.856	2.294			0.133	
$\hat{\gamma}$ outside central 80%	200	0.257	0.414				
	1000	0.489	0.633				
Median share, smaller regime	200	0.188	0.125				
	1000	0.104	0.049				

Notes: $S = 5,000$; true parameters common across the two sub-samples. “Estimated split”: the SSR or joint threshold estimate, with RMSE averaged over regimes with equal weights (as in Table 1) or with weights equal to each regime’s share of observations. “Median split”: regimes defined by the sample median of $y_{1,t-1}$. “No split”: the constant-parameter VAR. The intercept gap has true value zero. The smaller regime’s minimum share is $11/T$ at both T , the estimator’s feasibility bound.

Table A-3: Joint MC – classification accuracy

DGP	T	Classification accuracy	
		SSR	joint
DGP2	200	0.9987	0.9986
	1000	0.9997	0.9997
DGP3	200	0.7257	0.9904
	1000	0.6583	0.9985
DGP4	200	0.9943	0.9975
	1000	0.9991	0.9995

Notes: $S = 5,000$. Fraction of observations assigned to their true regime under $\hat{\gamma}$. DGP1 has no true regime, so accuracy is undefined; its parameter RMSEs are in Tables 1 and A-2. Parameter RMSEs for DGP2–DGP4 are in Table 1.

Table A-4: Decoupled MC – full threshold statistics.

DGP	T	Estimator	RMSE	MAD	bias / std	T-rate ratio
DGP5	200	joint q^μ	0.092	0.036	$-0.019 / 0.090$	see below
		joint q^σ	0.138	0.082	$+0.008 / 0.138$	
		decoupled $\hat{\gamma}^\mu$	0.040	0.026	$-0.012 / 0.038$	
		SSR q^μ	0.051	0.031	$-0.014 / 0.049$	
	1000	joint q^μ	0.009	0.005	$-0.002 / 0.008$	
		joint q^σ	0.022	0.014	$+0.002 / 0.022$	
		decoupled $\hat{\gamma}^\mu$	0.008	0.006	$-0.001 / 0.008$	
		SSR q^μ	0.008	0.005	$-0.002 / 0.008$	
DGP5 decoupled $\hat{\gamma}^\sigma$: $T = 200$ RMSE = 0.103, $T = 1000$ RMSE = 0.018.						
DGP5 T-rate ratios: joint q^μ 10.75, joint q^σ 6.30, decoupled μ 5.29, decoupled σ 5.67.						
DGP6	200	joint q^μ	0.033	0.021	$-0.009 / 0.032$	
		joint q^σ	0.094	0.050	$+0.007 / 0.094$	
		decoupled $\hat{\gamma}^\mu$	0.027	0.018	$-0.009 / 0.025$	
		SSR q^μ	0.032	0.021	$-0.009 / 0.030$	
	1000	joint q^μ	0.006	0.004	$-0.001 / 0.006$	
		joint q^σ	0.013	0.008	$+0.002 / 0.013$	
		decoupled $\hat{\gamma}^\mu$	0.005	0.003	$-0.001 / 0.005$	
		SSR q^μ	0.006	0.004	$-0.001 / 0.005$	
DGP6 decoupled $\hat{\gamma}^\sigma$: $T = 200$ RMSE = 0.052, $T = 1000$ RMSE = 0.009.						
DGP6 T-rate ratios: joint q^μ 5.92, joint q^σ 7.18, decoupled μ 5.63, decoupled σ 5.97.						

Notes. “joint q^μ ” is the joint estimator with state variable q^μ , comparable to the decoupled estimator on the mean dimension; analogously for “joint q^σ ”. The full decoupled estimator returns both $\hat{\gamma}^\mu$ and $\hat{\gamma}^\sigma$; their statistics are reported separately. Statistics for $\hat{\gamma}^\sigma$ are summarized in the comment rows for readability.

C The Interest-Rate (ZLB) Exercise

This appendix reports the interest-rate counterpart of the slack exercise of Section 4, with the lagged three-month T-bill rate as the mean state and RZ2018’s imposed threshold of 0.25%. The post-WWII sample is infeasible (too few low-rate observations before 2009), so two samples are reported.

Thresholds. The joint estimator places the threshold at 3.48% on both samples (Table 4): a broad low-rate environment, not a zero lower bound, and again a variance-driven split (the SSR criterion gives 0.21% and 1.22%). Once the fiscal-event regime is isolated, the decoupled

Table A-5: Decoupled MC – classification accuracy

DGP	T	Estimator	Classification accuracy	
			mean regime	variance regime
DGP5	200	joint q^μ	0.9946	—
		joint q^σ	—	0.9861
		decoupled	0.9970	0.9906
		SSR q^μ	0.9956	—
	1000	joint q^μ	0.9992	—
		joint q^σ	—	0.9979
		decoupled	0.9991	0.9983
		SSR q^μ	0.9993	—
	DGP6	joint q^μ	0.9955	—
		joint q^σ	—	0.9829
		decoupled	0.9969	0.9907
		SSR q^μ	0.9957	—
		joint q^μ	0.9992	—
		joint q^σ	—	0.9974
		decoupled	0.9992	0.9983
		SSR q^μ	0.9992	—

Notes: $S = 5,000$. Fraction of observations assigned to their true mean regime (under $\hat{\gamma}^\mu$) and true variance regime (under $\hat{\gamma}^\sigma$); “—” marks the dimension an estimator does not model. For the decoupled estimator the joint ($\mu \times \sigma$) cell accuracy is 0.9946 at $T = 200$ and 0.9988 at $T = 1000$ on DGP5, and 0.9944 and 0.9989 on DGP6. Parameter RMSEs are in Table 3.

Table A-6: Threshold estimators: finite-sample vs. limiting distributions

		Quantiles of $T(\hat{\gamma} - \gamma^0)$					KS
		q_{10}	q_{25}	q_{50}	q_{75}	q_{90}	
DGP7, $\hat{\gamma}$ (joint)	limit	−6.06	−3.32	−1.29	−0.12	2.93	
	$T = 200$	−6.05	−3.38	−1.29	−0.10	3.04	0.010
	$T = 1000$	−5.62	−3.12	−1.22	−0.06	3.44	0.024
DGP8, $\hat{\gamma}^\mu$	limit	−6.73	−3.80	−1.64	−0.40	1.34	
	$T = 200$	−6.73	−3.83	−1.61	−0.38	1.47	0.009
	$T = 1000$	−6.78	−3.79	−1.60	−0.34	1.23	0.025
DGP8, $\hat{\gamma}^\sigma$	limit	−9.95	−3.58	−0.02	6.14	15.64	
	$T = 200$	−12.42	−4.31	−0.13	6.49	17.89	0.031
	$T = 1000$	−10.02	−3.55	0.03	6.15	16.57	0.012

Notes: limits simulated with $R = 200,000$ draws of the smallest maximizer M_- of the corresponding two-sided compound Poisson process; finite-sample distributions from $S = 5,000$ ($T = 200$) and $S = 2,000$ ($T = 1000$) replications. KS is the two-sample Kolmogorov–Smirnov distance to the simulated limit; because S differs across columns, the entries have different sampling floors (5% critical values 0.019 and 0.030 respectively) and are not comparable across T .

Table A-7: Regime parameters, independence, and plug-in coverage

		$T = 200$	$T = 1000$
<i>Asymptotic standard deviations: empirical/theoretical ratio, $\sqrt{T} \text{vec}(\hat{B}_1 - B_1^0)$</i>			
DGP7	mean [range]	1.04 [1.02, 1.06]	1.00 [0.98, 1.01]
DGP8	mean [range]	1.05 [1.03, 1.06]	1.00 [0.99, 1.03]
<i>Independence checks (correlations)</i>			
DGP7	$\max_j \text{corr}(T\Delta\hat{\gamma}, \sqrt{T} \text{vec}(\hat{B}_1)_j) $	0.080	0.026
DGP8	$\max_j \text{corr}(T\Delta\hat{\gamma}^\mu, \sqrt{T} \text{vec}(\hat{B}_1)_j) $	0.072	0.053
DGP8	$\text{corr}(T\Delta\hat{\gamma}^\mu, T\Delta\hat{\gamma}^\sigma)$	-0.007	+0.013
<i>Remark 5 plug-in confidence intervals for $\gamma^{\sigma,0}$: empirical coverage</i>			
nominal 90%	oracle / feasible	0.859 / 0.860	0.894 / 0.899
nominal 95%	oracle / feasible	0.916 / 0.915	0.942 / 0.940

Notes: $\Delta\hat{\gamma} := \hat{\gamma} - \gamma^0$. Theoretical standard deviations from $S_1^{-1} \otimes \Sigma_1^0$ (DGP7) and the GLS information inverse (DGP8), with S_1 computed from a long simulated path. Feasible coverage: each replication re-simulates the variance-side limit law from its own $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ (1,500 draws), with the exogenous-state density known; evaluated on 1,000 replications (binomial standard error ≈ 0.010).

Table A-8: Nesting experiment: selection frequencies under alternative penalties and quantiles of the likelihood gap

DGP	T	Restriction selected, $c_T =$			Quantiles of $T(\hat{L}_T^F - \hat{L}_T^J)$					
		$\ln T$	$2 \ln T$	$3 \ln T$	min	50%	90%	95%	99%	max
DGP1	200	0.610	0.872	0.964	0.00	4.17	11.77	14.62	20.00	32.08
	1000	0.761	0.948	0.992	0.00	3.85	10.98	14.01	19.98	28.79
DGP2	200	0.813	0.990	0.999	0.00	3.35	6.40	7.83	10.55	18.05
	1000	0.946	0.998	1.000	0.00	2.73	5.93	7.06	9.99	19.01
DGP3	200	0.800	0.997	1.000	0.00	3.54	6.36	7.35	9.20	15.29
	1000	0.920	1.000	1.000	0.00	3.74	6.61	7.55	9.89	14.92
DGP4	200	0.992	1.000	1.000	0.00	0.54	2.38	3.24	5.03	11.46
	1000	0.999	1.000	1.000	0.00	0.00	1.61	2.40	4.19	12.89
DGP5	200	0.000	0.000	0.001	13.77	37.39	46.66	49.45	55.25	64.95
	1000	0.000	0.000	0.000	139.2	197.1	217.8	223.4	235.6	257.8
DGP6	200	0.000	0.000	0.000	18.82	43.36	55.76	59.47	65.92	75.49
	1000	0.000	0.000	0.000	157.6	221.3	245.9	253.1	268.7	301.7

Notes: $S = 5,000$ per cell. Frequencies are the share of replications with $T(\hat{L}_T^F - \hat{L}_T^J) < c_T$; $\ln T = 5.30$, 6.91 and $2 \ln T = 10.60$, 13.82 and $3 \ln T = 15.89$, 20.72 at $T = 200$, 1000. Monte Carlo standard errors are at most 0.007. The statistic is constructed as in Table 2; on DGP1–DGP4 it is non-negative by construction, and the frequencies are identical whether or not the unrestricted criterion is augmented with the restricted estimate. With $c_T = 3 \ln T$ the restriction is wrongly selected in 3 of 5,000 DGP5 samples at $T = 200$ (minimum gap 13.8 against a cut-off of 15.9), the only loss of power in the table.

mean threshold on the full sample is 0.24% – RZ2018’s 0.25% recovered from the data – while on the sample excluding WWII it is 2.73%; we therefore label the regimes low-rate and high-rate rather than ZLB and normal.

Table A-9: Decoupled TVAR, interest-rate exercise: thresholds and four-cell observation counts

Sample	Thresholds		Low rate		High rate		z share (%)
	$\hat{\gamma}^\mu$ (%)	$\hat{\gamma}^\sigma$	normal	fiscal-event [†]	normal	fiscal-event [†]	
Full sample	0.24*	0.0243	50 (12.5)	7 (1.8)	317 (79.3)	26 (6.5)	37.3
Excl. WWII	2.73	0.0241	140 (37.0)	14 (3.7)	213 (56.3)	11 (2.9)	45.7
Imposed (RZ2018)	0.25	—					—

Notes: As Table 5, with the lagged T-bill rate as mean state. *Essentially RZ2018’s imposed value.

Multipliers. Tables A-10 and A-11 and Figures A-1–A-2 report the normal-shock multipliers at the estimated and at the imposed thresholds. At the estimated thresholds, the full-sample multipliers are close to one in both regimes and the difference is never significant; on the sample excluding WWII the low-rate regime has the larger multiplier (0.98–1.37 against 0.15–0.23), significantly so only at $h = 3$. At the imposed 0.25% threshold the full-sample estimates essentially coincide with the free ones – the two thresholds differ by one basis point, an internal consistency check – while on the sample excluding WWII the low-rate regime’s multiplier is 2.5–2.8 against 0.6 in the high-rate regime, with the difference significant at every horizon ($\Delta = -2.0$ to -2.2). This mirrors RZ2018’s own finding of larger multipliers at the zero lower bound once WWII is excluded. Two caveats: the low-rate normal-shock cell contains 50 quarters (the 1930s and the post-2008 period), so the result is complementary to, not independent of, the slack results; and the difference between the two historical samples traces to the leverage points of Appendix D.

D Leverage Points and the Full Sample

On the full sample the imposed-threshold estimates are less sharp than on the sample excluding WWII (Tables 7 and A-11), and in the interest-rate exercise the free estimates differ markedly between the two samples although the fiscal-event regime is isolated in both. A decomposition that swaps, one at a time, the regime coefficients and the regime covariances estimated on the two samples attributes the entire difference to the mean-regime coefficients

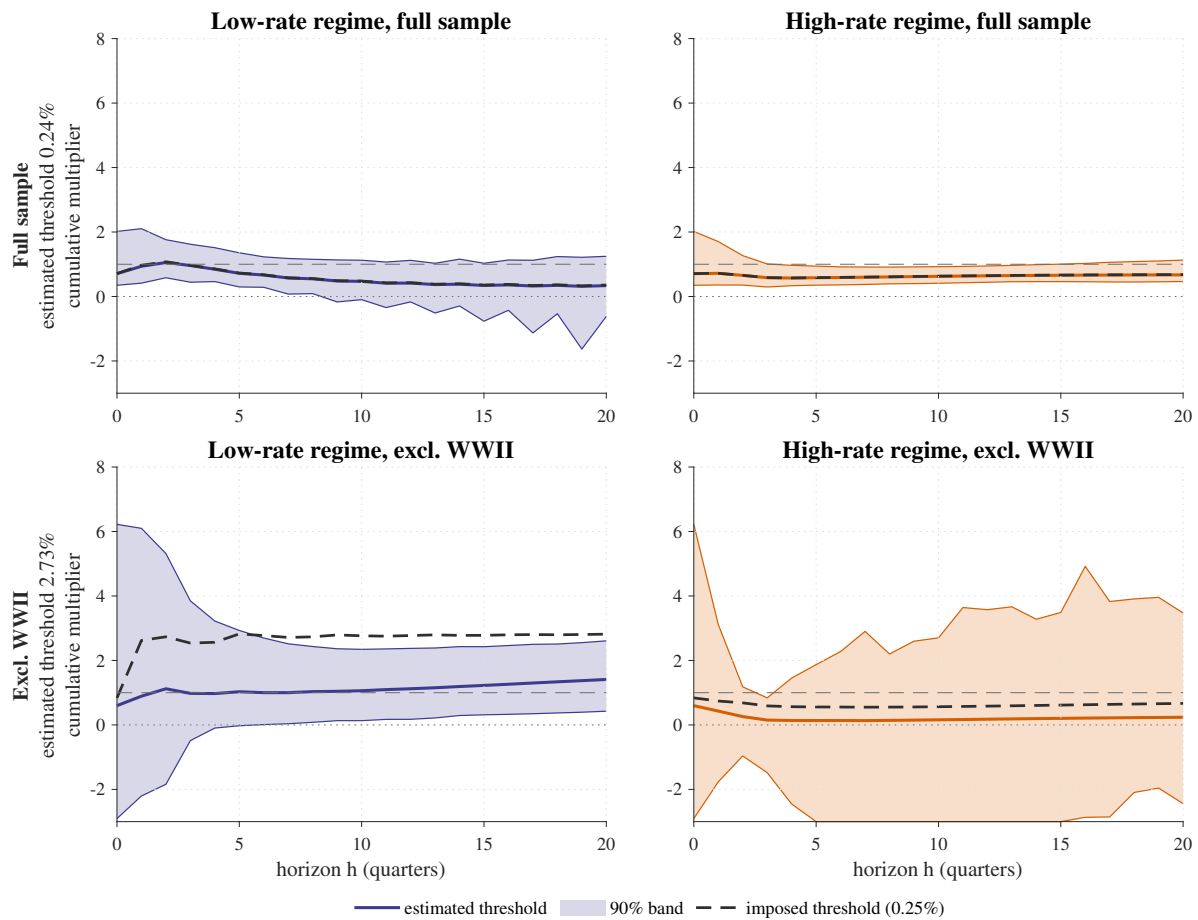


Figure A-1: Interest-rate exercise: cumulative multipliers in the normal-shock regime at the estimated thresholds (rows: full sample, excl. WWII; columns: low-rate, high-rate), 90% bands; dashed line: point estimate at the imposed 0.25%.

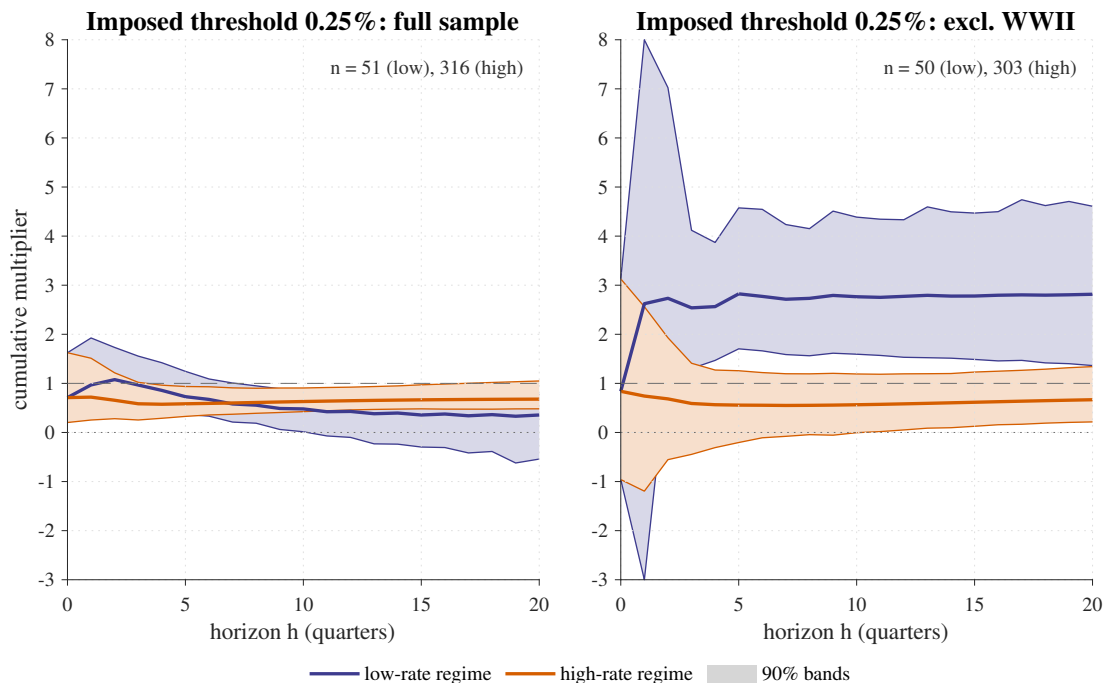


Figure A-2: Interest-rate exercise at the imposed 0.25% threshold, by sample: low-rate (violet) and high-rate (orange) regimes with 90% bands conditional on the fixed threshold pair.

Table A-10: Interest-rate exercise: normal-shock multipliers at the estimated thresholds

Sample	Regime	Horizon h (quarters)			
		$h = 3$	$h = 7$	$h = 11$	$h = 19$
Full sample	low rate	0.96 [0.44, 1.62]	0.57 [0.08, 1.18]	0.41 [-0.34, 1.07]	0.31 [-1.63, 1.22]
	high rate	0.58 [0.30, 1.01]	0.60 [0.37, 0.92]	0.63 [0.43, 0.93]	0.67 [0.46, 1.10]
	Δ (high – low)	-0.38 [-0.88, 0.22]	0.02 [-0.55, 0.59]	0.22 [-0.43, 1.10]	0.36 [-0.50, 2.72]
Excl. WWII	low rate	0.98 [-0.49, 3.85]	1.00 [0.04, 2.52]	1.10 [0.17, 2.36]	1.37 [0.39, 2.55]
	high rate	0.15 [-1.48, 0.84]	0.13 [-3.79, 2.90]	0.17 [-3.83, 3.64]	0.23 [-1.96, 3.96]
	Δ (high – low)	-0.83 [-3.95, -0.29]	-0.87 [-5.47, 2.06]	-0.93 [-4.88, 2.66]	-1.14 [-3.92, 2.55]

Notes: As Table 6; mean regimes at $\hat{\gamma}^\mu$ of Table A-9. $\Delta < 0$ indicates larger multipliers in the low-rate regime.

Table A-11: Interest-rate exercise: normal-shock multipliers at the imposed 0.25% threshold

Sample	Regime	Horizon h (quarters)			
		$h = 3$	$h = 7$	$h = 11$	$h = 19$
Full sample	low rate	0.97	0.58	0.42	0.33
		$[0.42, 1.56]$	$[0.21, 1.01]$	$[-0.07, 0.84]$	$[-0.62, 0.76]$
	high rate	0.58	0.60	0.64	0.68
		$[0.25, 1.02]$	$[0.37, 0.91]$	$[0.44, 0.92]$	$[0.48, 1.03]$
	Δ (high – low)	–0.38	0.02	0.22	0.35
		$[-0.81, 0.11]$	$[-0.36, 0.40]$	$[-0.19, 0.78]$	$[-0.10, 1.39]$
Excl. WWII	low rate	2.54	2.71	2.75	2.80
		$[1.30, 4.12]$	$[1.59, 4.24]$	$[1.57, 4.34]$	$[1.40, 4.71]$
	high rate	0.59	0.55	0.57	0.66
		$[-0.45, 1.41]$	$[-0.08, 1.20]$	$[0.02, 1.19]$	$[0.21, 1.32]$
	Δ (high – low)	–1.95	–2.16	–2.18	–2.15
		$[-3.40, -0.93]$	$[-3.69, -1.10]$	$[-3.82, -0.89]$	$[-4.04, -0.51]$

Notes: As Table 7, mean threshold imposed at 0.25%. Normal-shock cell counts (low / high rate): 51 / 316 (full), 50 / 303 (excl. WWII).

of the low-rate (and, in the slack exercise, high-unemployment) regime, not to the covariances; and within those coefficients to exactly three quarters, 1941Q3–1942Q1, the U.S. entry into WWII. These quarters carry the largest news realizations of the century while unemployment is still high and rates are at their floor, so they are extreme leverage points inside the regime. Two of the three fall in the fiscal-event regime, but the fiscal-event regime does not exclude them from the estimation of the mean-regime coefficients: under the decoupled model’s GLS structure, observations in the fiscal-event regime enter $\hat{B}_i$ with a smaller precision weight, not a zero weight. Excluding the sample removes them; isolating the shock regime downweights them. This is why the sample excluding WWII – the sample RZ2018 themselves emphasize for the interest-rate exercise, given the rate peg and rationing of the war years – delivers the cleaner comparison on both state variables.

E Instrument Strength by Cell

In the internal-proxy SVAR the news series is ordered first and its innovation is the identified shock, so the strength of the identifying variation in each cell can be measured as in an instrumental-variables first stage: a regression of g_t on z_t and the VAR controls x_t within the cell, with z_t the news series. Table A-12 reports the number of observations, the F -statistic on z_t with Newey–West (four lags) and with homoskedastic standard errors, the

effective F of [Montiel Olea and Pflueger \(2013\)](#), which for a single instrument coincides with the HAC F , the HAC t -statistic, and the share of the cell's $\sum_t z_t^2$ accounted for by its three and five largest realizations. Four layers are compared on each sample: the whole sample without regimes; RZ2018's own partition, lagged unemployment at 6.5% with no variance regimes; the imposed 6.5% threshold with the fiscal-event regime isolated at the estimated variance threshold, which is the exercise of Table 7; and the estimated mean and variance thresholds of Table 6. Cells with fewer than 16 observations are not estimated.

Three facts emerge. First, isolating the fiscal-event regime strengthens the first stage in the low-unemployment regime of every sample: relative to RZ2018's partition the HAC F rises by factors of 5.3, 17.6 and 2.1 on the full, excluding-WWII and post-WWII samples (from 2.2, 0.8 and 1.5 to 11.6, 13.7 and 3.1; homoskedastic F from 16 to 84 on the full sample). The fiscal-event cells carry a contemporaneous news-spending coefficient of the opposite sign ($t = -5.3$ on the full sample), so the pooled regime's first stage is a cancellation. Isolating the fiscal-event regime also strengthens both post-WWII cells and the full-sample high-unemployment cell at 6.5%; it weakens the high-unemployment cell on the sample excluding WWII ($F < 0.1$ with 170 observations), where the news variation outside the fiscal-event regime is negligible, as RZ2018 note for their high-unemployment state on this sample. Second, no cell in any layer, RZ2018's own partition included, reaches the [Montiel Olea and Pflueger \(2013\)](#) critical value for a 5% worst-case bias (37.4; 23.1 for 10%): the largest effective F is 13.7, and the top-three shares, between 0.56 and 1.00, show that in every cell the identifying variation is a handful of quarters. This is the property of the military news series that RZ2018 document, and it is shared by every specification in this literature; a VAR first stage is also contemporaneous, whereas the news series is anticipatory and RZ2018's first stages cumulate spending over the horizon, so the statistics here are conservative. The identifying moment of the decoupled model is the normal-shock covariance Σ_1 , which pools the two mean regimes: on it the homoskedastic F is 39, 50 and 3.6 on the three samples (HAC: 3.0, 6.0 and 4.1). Third, the two cells whose bootstrap draws produce the denominator problem noted under Tables 6 and 7 – the post-WWII low-unemployment cell at the estimated threshold (47 quarters) and the post-WWII high-unemployment cell at 6.5% (79 quarters) – are both weak by every measure, but so are seven other normal-shock cells in which the denominator never crosses zero. Weak identification is pervasive in these data; the denominator problem arises where it meets a small cell.

Table A-12: First-stage strength of the news instrument by cell (slack exercise)

Sample	Cell	n	F_{HAC}	F_{homosk}	eff. F	t_{HAC}	top-3	top-5
<i>L0: whole sample, no regimes</i>								
Full sample	all quarters	499	3.75	19.73	3.75	1.94	0.56	0.73
Excl. WWII	all quarters	477	0.90	8.76	0.90	0.95	0.66	0.85
Post-WWII	all quarters	272	1.25	1.26	1.25	-1.12	0.95	0.97
<i>L1: RZ2018 partition, lagged unemployment at 6.5%, no variance regimes</i>								
Full sample	low unemployment	318	2.21	16.45	2.21	1.49	0.64	0.83
	high unemployment	181	2.87	0.83	2.87	1.70	0.93	0.99
Excl. WWII	low unemployment	300	0.78	7.11	0.78	0.88	0.82	0.95
	high unemployment	177	0.42	0.34	0.42	0.65	0.96	0.98
Post-WWII	low unemployment	189	1.51	1.64	1.51	-1.23	0.96	0.98
	high unemployment	83	0.64	0.67	0.64	0.80	0.75	0.92
<i>L2: imposed 6.5% mean threshold, estimated variance threshold</i>								
Full sample	low-u $\times$ normal-shock	294	11.62	84.18	11.62	3.41	0.72	0.88
	low-u $\times$ fiscal-event	24	27.59	9.75	27.59	-5.25	0.95	1.00
	high-u $\times$ normal-shock ^a	172	6.10	2.67	6.10	2.47	0.99	0.99
	high-u $\times$ fiscal-event	9	—	—	—	—	1.00	1.00
	normal-shock, pooled	466	2.96	39.29	2.96	1.72	0.72	0.84
	fiscal-event, pooled	33	22.98	5.47	22.98	-4.79	0.82	0.99
Excl. WWII	low-u $\times$ normal-shock	282	13.66	68.93	13.66	3.70	0.90	0.94
	low-u $\times$ fiscal-event	18	0.04	0.01	0.04	0.20	0.98	1.00
	high-u $\times$ normal-shock ^a	170	0.00	0.00	0.00	-0.05	0.95	0.97
	high-u $\times$ fiscal-event	7	—	—	—	—	1.00	1.00
	normal-shock, pooled	452	5.97	50.44	5.97	2.44	0.88	0.91
	fiscal-event, pooled	25	3.03	1.02	3.03	-1.74	0.94	0.99
Post-WWII	low-u $\times$ normal-shock ^a	176	3.11	2.48	3.11	1.76	0.58	0.71
	low-u $\times$ fiscal-event	13	—	—	—	—	1.00	1.00
	high-u $\times$ normal-shock ^a	79	1.32	0.91	1.32	1.15	0.76	0.93
	high-u $\times$ fiscal-event	4	—	—	—	—	1.00	1.00
	normal-shock, pooled	255	4.05	3.62	4.05	2.01	0.47	0.64
	fiscal-event, pooled	17	1.15	0.32	1.15	1.07	1.00	1.00
<i>L3: estimated mean and variance thresholds</i>								
Full sample	low-u $\times$ normal-shock	382	11.46	101.4	11.46	3.39	0.64	0.83
	low-u $\times$ fiscal-event	28	16.89	5.17	16.89	-4.11	0.95	1.00
	high-u $\times$ normal-shock ^a	84	1.03	0.78	1.03	1.02	1.00	1.00
	high-u $\times$ fiscal-event	5	—	—	—	—	1.00	1.00
	normal-shock, pooled	466	2.96	39.29	2.96	1.72	0.72	0.84
	fiscal-event, pooled	33	22.98	5.47	22.98	-4.79	0.82	0.99
Excl. WWII	low-u $\times$ normal-shock ^a	37	0.05	0.01	0.05	-0.23	1.00	1.00
	low-u $\times$ fiscal-event	4	—	—	—	—	1.00	1.00
	high-u $\times$ normal-shock ^a	415	4.11	66.30	4.11	2.03	0.88	0.91
	high-u $\times$ fiscal-event	21	0.09	0.02	0.09	0.30	0.99	1.00
	normal-shock, pooled	452	5.97	50.44	5.97	2.44	0.88	0.91
	fiscal-event, pooled	25	3.03	1.02	3.03	-1.74	0.94	0.99
Post-WWII	low-u $\times$ normal-shock ^a	47	4.98	2.27	4.98	2.23	0.56	0.76
	low-u $\times$ fiscal-event	5	—	—	—	—	1.00	1.00
	high-u $\times$ normal-shock ^a	208	4.59	3.74	4.59	2.14	0.55	0.74
	high-u $\times$ fiscal-event	12	—	—	—	—	1.00	1.00
	normal-shock, pooled	255	4.05	3.62	4.05	2.01	0.47	0.64
	fiscal-event, pooled	17	1.15	0.32	1.15	1.07	1.00	1.00

Notes: First-stage regression of g_t on z_t and the VAR regressors x_t (constant and four lags of z , y , g) within each cell; slack exercise, mean state lagged unemployment. L2 and L3 cells are (mean regime) $\times$ (variance regime); “pooled” rows pool the two mean regimes within a variance regime. $F_{\text{HAC}} = t_{\text{HAC}}^2$ with a Newey–West (Bartlett, 4 lags) variance, lags in calendar quarters; F_{homosk} : classical first-stage F . Eff. F : Montiel Olea and Pflueger (2013) effective F with the same HAC variance; with one instrument it equals F_{HAC} . Critical value (simplified procedure, 5% test, worst-case bias 5% of the benchmark): 37.42 (10%: 23.11). Top-3 / top-5: share of the cell’s $\sum z_t^2$ due to its three / five largest $|z_t|$. — : cell too small for the regression ($n \leq 46$).

^a Normal-shock cell with $F_{\text{HAC}} < 10$.

F Posterior Bands as a Check on the Bootstrap

The wild-bootstrap intervals of Section 4 are wide in some cells, and a reader may ask whether the width is a property of the bootstrap or of the object. This appendix answers by computing the bands a second way. Conditional on the estimated threshold pair, the decoupled model is a set of linear regressions with cell-specific covariances, and a two-block Gibbs sampler – B_i given Σ from its Normal full conditional, Σ_ℓ given B from an inverse-Wishart – delivers a posterior for the regime parameters and, through the same impulse-response code, for the multipliers. We use two priors. The flat prior has no shrinkage and serves as the benchmark that should agree with the bootstrap if the bootstrap is doing its job. The second is a Minnesota prior with the usual tightness ($\lambda_1 = 0.2$, lag decay 1, cross-equation factor 0.5) on the macroeconomic equations, but a loose prior on the coefficients of the news series: the news column is the instrument, and shrinking it towards zero removes the identifying variation from the system. A standard Minnesota prior applied to all columns shrinks the summed coefficients of output on lagged news from 0.10 to 0.005 in the post-WWII low-unemployment cell at the estimated threshold, and the changes it then produces in the multipliers reflect that choice rather than the data. Each chain has 2,000 retained draws after 500 of burn-in, started at the FGLS estimates.

Three results. First, the the posterior under the flat prior reproduces the bootstrap. Posterior medians are within a few hundredths of the point estimates in the well-populated cells, and the 90% posterior intervals are 0.81–0.98 times the bootstrap width on the two historical samples and 0.65–0.72 times post-WWII, where the wild bootstrap’s resampling of heteroskedastic residuals produces heavier tails than a Gaussian likelihood. The two methods agree on which cells are not identified: in the post-WWII high-unemployment cell at the imposed threshold the flat 90% interval at $h = 3$ is $[-14.9, 13.6]$ against the bootstrap’s $[-23.5, 21.6]$, and the cumulative spending response changes sign in 17–20% of posterior draws against 23% of bootstrap draws. The width is the object. Second, under the Minnesota prior the results of Section 4 survive where they are claimed (Table A-13). At the imposed threshold on the sample excluding WWII the difference between regimes is 1.0–2.1 with the 90% interval excluding zero at $h = 3, 7$ and 11; at the estimated thresholds on the same sample the ranking is unchanged and the 68% interval excludes zero at every horizon; on the post-WWII sample at the estimated thresholds the low-unemployment multiplier is 2.9 at $h = 3$ against 1.5 in the high-unemployment regime, the 68% interval for the difference excluding zero at $h = 3, 11$ and 19 and the 90% interval never, as under the bootstrap. The imposed post-WWII high-unemployment cell remains unidentified under every prior.

Third, the shrinkage exposes one fragility the bootstrap does not: in the full-sample high-unemployment regime at the estimated threshold (84 quarters) the estimated within-regime dynamics are explosive, with a largest root of 1.58, and under a prior that shrinks the own lags while leaving the news column free the long-horizon denominator changes sign in 8–28% of draws for $h \geq 7$. The short-horizon multipliers in that cell are unaffected; the $h = 19$ values should be read with that caveat, which we note under Table 6. Where the posterior and the bootstrap differ, the difference traces to the prior; where they agree, the bands are what the data support.

G Fiscal-Event Cells

The fiscal-event cells are too small for inference – 4 to 28 quarters in the slack exercise and 7 to 26 in the interest-rate exercise (Tables 5 and A-9) – and their multipliers are dominated by a handful of historical episodes; the main text uses them only to isolate the normal-shock regime. For completeness, Table A-14 reports the fiscal-event cells of the AG2012 exercise, where the cell sizes are documented in Table A-16: every 90% interval includes zero, the fiscal-event environment carries larger point multipliers, the ranking across mean regimes matches the normal-shock cells, and the $h = 0$ multiplier is identical across mean regimes within the variance regime, as implied by the common impact vector.

H Spending Growth as the Variance State

This appendix re-estimates the decoupled model of Section 4.3 with the absolute lagged change in government spending, $|\Delta g_{t-1}|$, in place of the absolute lagged news series as the variance state, on the same samples, grid rule and bootstrap. Table A-15 compares the two states. The spending-growth state does not recover the fiscal-event regime: it isolates the build-ups *and* demobilizations of WWI, WWII and Korea together with pre-1940 peacetime spending volatility, misses the announcement quarters that precede spending and every fiscal event after 1968, and agrees with the news-based regime on 15 of its 92 quarters ($\kappa = 0.16$; 0.12 excluding WWII; 0.00 post-WWII, where peacetime spending changes are too small for the state to separate a shock environment and the above-threshold regime contains 91% of the sample). This is the pattern the timing argument of Section 4.3 predicts for a state defined on the response rather than on the identifying shock. Multipliers are interpretable only on the full sample, where the mean threshold moves to 3.56% – the joint estimate of

Table A-13: Posterior bands for the normal-shock multipliers under a Minnesota prior with a loose news column

Sample	Regime	Horizon h (quarters)			
		$h = 3$	$h = 7$	$h = 11$	$h = 19$
<i>Panel A: estimated thresholds (Table 6)</i>					
Full sample	low unemployment	0.41 [0.15, 0.69]	0.44 [0.26, 0.66]	0.46 [0.30, 0.66]	0.45 [0.30, 0.64]
	high unemployment	0.76 [0.21, 1.48]	1.08 [-1.31, 3.90]	1.19 [-5.09, 5.99]	1.14 [-6.71, 7.87]
	Δ (high – low)	0.34 [-0.18, 0.98]	0.65 [-1.77, 3.48]	0.73 [-5.49, 5.55]	0.70 [-7.10, 7.44]
Excl. WWII	low unemployment	0.28 [0.08, 0.52]	0.15 [-0.06, 0.39]	0.07 [-0.18, 0.43]	-0.07 [-0.46, 0.76]
	high unemployment	0.86 [0.06, 1.85]	1.05 [0.19, 2.12]	1.07 [0.21, 2.28]	0.97 [0.21, 2.08]
	Δ (high – low)	0.57 [-0.16, 1.47]	0.89 [0.01, 1.99]	0.99 [0.02, 2.21]	1.01 [-0.17, 2.27]
Post-WWII	low unemployment	2.85 [0.52, 7.20]	2.38 [0.54, 5.67]	2.50 [0.50, 7.71]	2.98 [-7.98, 21.32]
	high unemployment	1.53 [-0.38, 3.88]	1.26 [-0.42, 3.29]	1.05 [-0.55, 2.88]	0.81 [-0.49, 2.30]
	Δ (high – low)	-1.34 [-5.13, 0.79]	-1.14 [-4.21, 1.14]	-1.53 [-6.84, 1.04]	-2.27 [-20.91, 9.33]
<i>Panel B: imposed 6.5% mean threshold (Table 7)</i>					
Full sample	low unemployment	0.47 [0.18, 0.80]	0.48 [0.23, 0.78]	0.49 [0.27, 0.78]	0.47 [0.29, 0.75]
	high unemployment	0.45 [-0.05, 0.92]	0.67 [0.31, 1.03]	0.78 [0.37, 1.20]	0.82 [0.31, 1.37]
	Δ (high – low)	-0.01 [-0.46, 0.39]	0.19 [-0.20, 0.57]	0.28 [-0.20, 0.74]	0.33 [-0.25, 0.92]
Excl. WWII	low unemployment	0.40 [-0.05, 1.01]	0.27 [-0.04, 0.75]	0.22 [-0.04, 0.62]	0.17 [-0.04, 0.48]
	high unemployment	1.43 [0.38, 4.17]	2.05 [0.56, 6.67]	2.27 [0.45, 8.44]	2.31 [0.20, 10.01]
	Δ (high – low)	1.01 [0.05, 3.61]	1.76 [0.24, 6.44]	2.03 [0.22, 8.07]	2.14 [0.00, 9.86]
Post-WWII	low unemployment	1.17 [-0.68, 3.29]	0.73 [-0.46, 1.93]	0.59 [-0.43, 1.57]	0.49 [-0.42, 1.38]
	high unemployment	0.57 [-10.36, 11.07]	0.77 [-6.53, 8.49]	0.71 [-5.00, 6.24]	0.59 [-2.84, 3.51]
	Δ (high – low)	-0.66 [-11.90, 9.31]	0.06 [-7.23, 7.54]	0.16 [-5.71, 5.73]	0.14 [-3.60, 2.91]

Notes: Posterior medians with 90% posterior intervals below, from a two-block Gibbs sampler conditional on the threshold pair (2,000 draws after 500 of burn-in, started at the FGLS estimates), Minnesota prior with $\lambda = (0.2, 0.5, 1)$ on the macroeconomic coefficients and prior standard deviation $2/j$ on the lag- j coefficients of the news series in the output and spending equations. Multipliers as in Table 6; Δ rows are posterior differences across mean regimes, draw by draw. Point estimates and 90% wild-bootstrap intervals for the same cells are in Tables 6 (Panel A) and 7 (Panel B). Normal-shock cell counts (low / high unemployment): Panel A, 382 / 84 (full), 37 / 415 (excl. WWII), 47 / 208 (post-WWII); Panel B, 294 / 172, 282 / 170, 176 / 79.

Table A-14: Decoupled TVAR on AG2012 data: fiscal-event (VR2) cell multipliers

Cell	Horizon h (quarters)					
	$h = 0$	$h = 3$	$h = 7$	$h = 11$	$h = 15$	$h = 19$
MR1×VR2 (low growth, $n = 11$)	0.84 [-1.77, 10.25]	0.20 [-3.50, 10.30]	0.12 [-3.66, 8.17]	0.40 [-3.94, 6.84]	0.80 [-4.45, 6.28]	1.20 [-5.29, 5.63]
MR2×VR2 (high growth, $n = 4$)	0.84 [-1.77, 10.25]	1.79 [-0.93, 9.25]	2.50 [-1.53, 8.65]	3.05 [-2.21, 8.26]	3.58 [-2.79, 7.83]	4.02 [-3.12, 7.68]

Notes: Cumulative dollar multipliers in the fiscal-event variance regime (VR2) of the decoupled specification with $q_t^\sigma = |\text{news}_{t-1}|$; 90% wild-bootstrap intervals in brackets ($B = 499$, all draws valid). Estimates are descriptive only, given the cell sizes documented in Table A-16. The identical $h = 0$ entries across the two cells reflect the common impact vector within a variance regime.

Table 4, since the state leaves the wartime variance where the joint estimator found it – and the difference between the normal-shock cells vanishes ($\Delta = 0.06$ at $h = 19$); the other two samples yield boundary partitions with cells too thin for inference.²

Table A-15: Fiscal-event regime from news vs spending growth: thresholds, overlap and normal-shock multipliers ($h = 19$)

Sample	$(\hat{\gamma}^\mu, \hat{\gamma}^\sigma)$		VR2 quarters			News state, $h = 19$			Spending state, $h = 19$		
	News	Spending	News	Spending	κ	Low-u	High-u	Δ	Low-u	High-u	Δ
Full sample	(8.42, 2.434)	(3.56, 0.651)	33	92	0.16	0.45	0.98	0.53 [-0.69, 1.62]	0.49	0.55	0.06 [-0.88, 0.72]
Excl. WWII [†]	(3.44, 2.412)	(12.06, 0.674)	25	71	0.12	0.24	0.55	0.31 [-0.90, 1.61]	0.16	-36.75	-36.91 [-16.04, 15.34]
Post-WWII [†]	(4.23, 2.111)	(4.72, 0.023)	17	247	0.00	4.47	0.72	-3.75 [-25.59, 13.77]	1.37	-0.13	-1.50 [-11.66, 4.53]

Notes: Decoupled TVAR, mean state lagged unemployment (%). News state: $|z_{t-1}|$ (% of GDP); spending state: $|\Delta g_{t-1}|$, first difference of the VAR's spending series g (% of potential GDP). $\hat{\gamma}^\sigma$ in %. VR2 = fiscal-event regime (above the variance threshold). κ : Cohen's kappa between the two VR2 indicators, quarter by quarter on the common sample. Normal-shock (VR1) cumulative multipliers at $h = 19$; Δ = high-u – low-u with its 90% wild-bootstrap interval in brackets ($B = 499$).

[†] Spending state, Excl. WWII: point estimates lie outside their own 90% band at $h = 11, 19$ (cumulative-spending denominator near zero in a 17-quarter cell); these multipliers are not interpretable.

[†] Spending state, Post-WWII: the smallest normal-shock cell has only 6 quarters, so the multipliers are imprecise.

I The Auerbach–Gorodnichenko Exercise

This appendix documents the robustness exercise summarized in Section 4.7.

²The grid is built by the rule used for $|z_{t-1}|$, which has an atom at zero; $|\Delta g_{t-1}|$ is continuous, and a stricter minimum regime size would rule out the post-WWII corner solution. We keep one rule so that the comparison uses one code path.

Data, identification and states. We use the AG2012 replication data: real government purchases G_t , net real taxes T_t and real output Y_t , in log levels, 1947Q1–2008Q4 ($T = 248$), with $p = 4$ lags. Identification follows AG2012 and Blanchard and Perotti (2002): a Cholesky decomposition of the regime-specific covariance matrix with G ordered first identifies the government-spending shock by the restriction that G does not respond to within-quarter innovations in T or Y . We report cumulative integral dollar multipliers, $\sum_{s \leq h} \hat{\phi}_{Y,s} / \sum_{s \leq h} \hat{\phi}_{G,s}$ converted with the sample ratio $\bar{Y}/\bar{G} = 5.71$. AG2012’s state variable is a seven-quarter *centered* moving average of output growth dated at $t - 1$, which embeds $\Delta \ln Y_{t+1}$ and $\Delta \ln Y_{t+2}$ and is therefore not known at $t - 1$; since our consistency results require a predetermined state (Assumption B2), we use the backward seven-quarter moving average, $q_t^\mu = \frac{1}{7} \sum_{j=0}^6 \Delta \ln Y_{t-j}$. The variance state is the same Ramey (2011) military-news series as in Section 4, here expressed as a percent of GDP (the two normalizations differ by a factor of 100, verified on the common quarters). Inference uses the same wild bootstrap as in Section 4 ($B = 499$).

Table A-16: Decoupled TVAR on AG2012 data (1947Q1–2008Q4): thresholds and four-cell observation counts

Sample	Thresholds		Low growth		High growth		q^σ share (%)
	$\hat{\gamma}^\mu$ (%)	$\hat{\gamma}^\sigma$	normal	fiscal-event [†]	normal	fiscal-event [†]	
<i>Panel A: Ramey news, $q_t^\sigma = \text{news}_{t-1}$ (% of GDP)</i>							
$ \text{news}_{t-1} $	1.245	2.036	184 (76.7)	11 (4.6)	41 (17.1)	4 (1.7)	79.8
<i>Panel B: G-equation residual, $q_t^\sigma = \hat{u}_{t-1}^G$ (log-difference)</i>							
$ \hat{u}_{t-1}^G $	1.245	0.00326	46 (19.2)	149 (62.1)	5 (2.1) [‡]	40 (16.7)	97.3

Notes: $\hat{\gamma}^\mu$ is the threshold on the backward seven-quarter moving average of $\Delta \ln Y$ (quarterly percent; $1.245 \approx 5\%$ annualized); mean regimes: low-growth ($q_{t-1}^\mu \leq \hat{\gamma}^\mu$) and high-growth. $\hat{\gamma}^\sigma$ is the threshold on the variance state; variance regimes: normal-shock (VR1) and fiscal-event (VR2) in Panel A, quiet and dispersed in Panel B (Panel B’s split has no event structure and serves as a specificity check). The last column is the share of $\sum_t |q_t^\sigma|$ in the above-threshold variance regime. [†]Cells too small for inference; see Appendix G. [‡]Five observations; descriptive only.

Thresholds and cells. The news candidate selects $\hat{\gamma}^\sigma = 2.04\%$ of GDP, isolating 15 quarters that coincide, quarter for quarter, with the fiscal-event regime of Section 4 on the common 1949Q1–2008Q4 window (Cohen’s $\kappa = 1$; the two RZ2018 quarters outside this window, 2011Q4 and 2013Q2, fall after AG2012’s sample).³ The mean threshold is 1.25% per quarter ($\approx 5\%$ annualized), isolating a high-growth regime of 45 quarters.

³The agreement is not near-automatic: the news candidate induces 59 distinct nontrivial partitions on this sample, so agreement by chance is below two percent; the fifteen matched quarters span six episodes across six decades; and Panel B shows that a dense volatility candidate – the G-equation’s own absolute

Table A-17: Decoupled TVAR multipliers on AG2012 data, normal-shock regime

Cell	Horizon h (quarters)			
	$h = 3$	$h = 7$	$h = 11$	$h = 19$
<i>Panel A: Ramey news variance state ($\hat{\gamma}^\sigma = 2.04\%$ of GDP)</i>				
low-growth $\times$ normal-shock	0.18 [−0.90, 1.17]	−0.15 [−1.61, 0.85]	−0.06 [−2.06, 0.96]	0.31 [−2.98, 1.71]
high-growth $\times$ normal-shock	1.11 [0.32, 1.82]	1.54 [0.08, 3.18]	2.08 [−0.22, 4.17]	3.12 [−0.59, 4.78]
Δ (high − low)	0.93 [0.05, 1.79]	1.69 [0.15, 3.76]	2.14 [0.06, 5.04]	2.80 [−0.56, 5.89]
<i>Panel B: G-residual variance state ($\hat{\gamma}^\sigma = 0.00326$)</i>				
low-growth $\times$ normal-shock	−0.11 [−1.56, 1.12]	−0.57 [−2.21, 0.81]	−0.66 [−2.61, 0.83]	−0.67 [−3.57, 1.40]
high-growth $\times$ normal-shock [‡]	0.84 [−0.19, 1.69]	1.11 [−0.52, 2.92]	1.55 [−0.87, 3.90]	2.47 [−1.67, 4.63]

Notes: Cumulative integral dollar multipliers in the below-threshold variance regime only. 90% wild-bootstrap intervals in brackets ($B = 499$). Δ : difference across mean regimes, interval from the same bootstrap draws as the cell bands (the two cells share the normal-shock covariance, so the interval is tighter than a comparison of bands suggests); it excludes zero through $h = 11$. [‡]Cell with $n = 5$ observations; descriptive only. Fiscal-event cells in Appendix G.

Multipliers. Table A-17 and Figure A-3 report the normal-shock cells. The high-growth multiplier exceeds the low-growth multiplier at every horizon – 1.11, 1.54, 2.08, 3.12 against 0.18, −0.15, −0.06, 0.31 at $h = 3, 7, 11, 19$ – and the difference is significant through $h = 11$ ($\Delta = 0.93, 1.69, 2.14$, with 90% intervals [0.05, 1.79], [0.15, 3.76], [0.06, 5.04]), even where the individual bands overlap, since the two cells are positively correlated through the shared normal-shock covariance.

Mean regimes across the two exercises. The two mean states measure different things – a labor-market stock in RZ2018, a growth flow in AG2012 – and their partitions overlap only partially on the common window: 45 and 48 quarters respectively, 16 shared ($\kappa = 0.19$), with the growth state leading the labor-market state by about six quarters through 1970 and the two selecting different episodes afterwards. The ranking of multipliers is nonetheless the same under both (Section 4.7). Figure A-4 shows the three regime indicators on the common window.

residual, free to isolate the largest spending surprises – is not steered toward an event partition but selects a low-versus-high dispersion split. Recovering the fiscal-event regime as the variance break therefore requires a candidate that encodes the timing of fiscal events, and it is that specific partition that the macroeconomic covariance confirms.

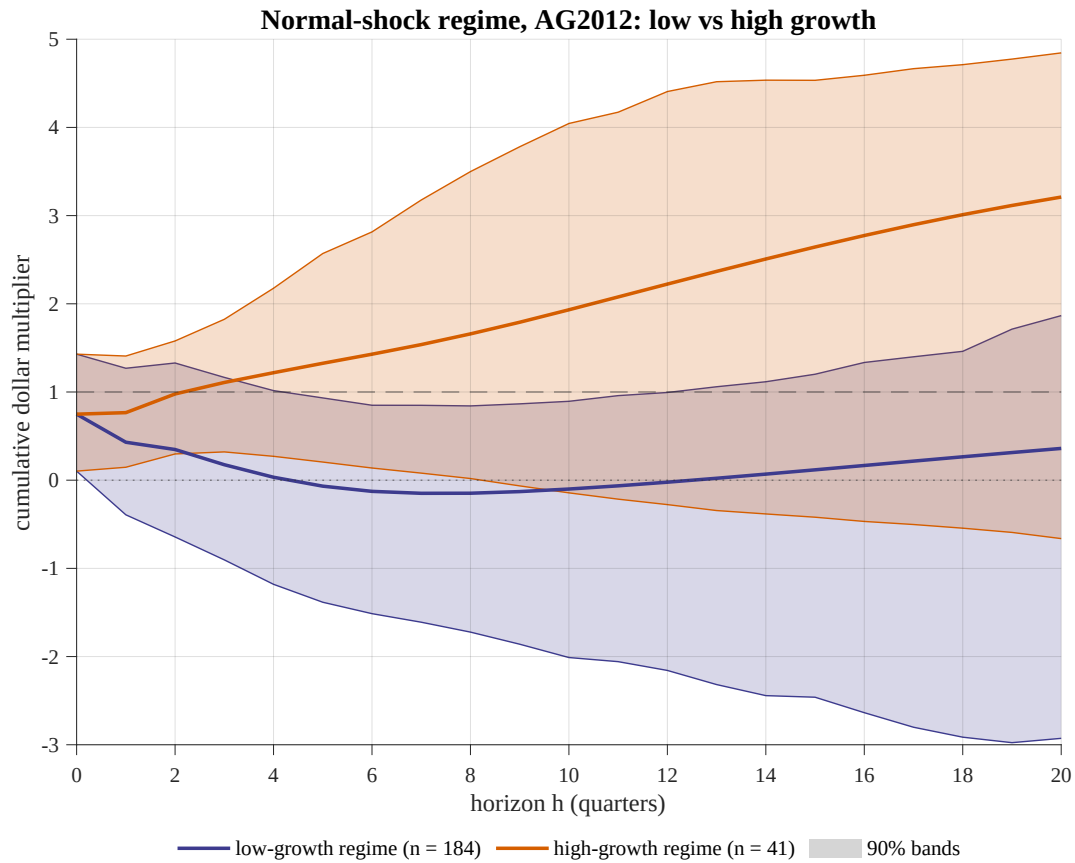


Figure A-3: AG2012 exercise: cumulative dollar multipliers in the normal-shock cells (high-growth and low-growth regimes) with 90% wild-bootstrap bands.

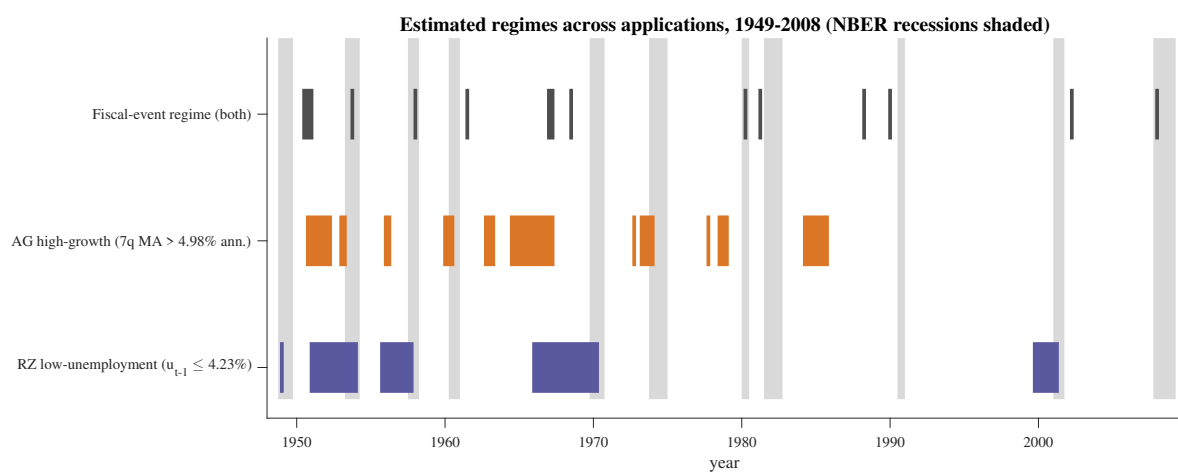


Figure A-4: Estimated regime assignments across the two exercises on the common window (1949Q1–2008Q4), NBER recessions shaded. *Bottom*: RZ2018 low-unemployment regime ($u_{t-1} \leq 4.23\%$). *Middle*: AG2012 high-growth regime. *Top*: the fiscal-event quarters, identical across the two exercises.