EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Dimitris Christopoulos, Peter McAdam, Elias Tzavalis

Threshold endogeneity in vector autoregressions: reassessing monetary state dependence

No 3263

Abstract

We develop an endogenous threshold VAR that addresses contemporaneous dependence between the threshold variable and reduced-form innovations, a pervasive issue when regime indicators are jointly determined with system dynamics. A regime-specific copula-based control function removes this dependence without requiring external instruments or parametric assumptions on the threshold's marginal distribution, while preserving the linear regime-wise least-squares structure. We demonstrate that omitting the control generates "excess sensitivity" and "excess propagation" errors in impulse responses, clarify structural and proxy-SVAR identification under endogenous regimes, and establish conditions under which Chan-type threshold asymptotics remain valid with generated con-trols. A Hermite sieve extension accommodates tail-dependent and asymmet-ric dependence. Monte Carlo exercises document the size of the biases from ig-noring endogeneity across different designs. Applied to the analysis of monetary transmission, the framework avoids the price and persistence puzzles displayed by a linear VAR, delivers regime-dependent sacrifice ratios, and aligns estimated regimes with historical inflation episodes.

KEYWORDS. Endogenous Threshold, Monte Carlo, Monetary Policy, Impulse Response.

JEL CLASSIFICATION. C32, C34, E52.

NON-TECHNICAL SUMMARY

The paper studies what happens when the “state” that splits the economy into regimes is itself reacting to shocks, and proposes a way to fix this in threshold VARs. In standard threshold VARs, the sample is divided into regimes using a threshold variable (for example, an inflation measure). The literature almost always assumes this threshold variable is predetermined: it can depend on past data, but not on the current shocks hitting the system. The authors argue that this is unrealistic in many macro settings, where agents and policymakers react quickly and thresholds such as inflation, credit spreads, or risk indicators move contemporaneously with shocks. In that case, the threshold is endogenous to the VAR and traditional methods can misclassify regimes and distort the estimated dynamics.

The paper develops an “endogenous-threshold” VAR (TVARN) that explicitly corrects for this contemporaneous dependence between the threshold variable and the VAR innovations. The key device is a copula-based control function. The joint distribution of the threshold and the VAR shocks is modeled via a copula, which captures their dependence structure independently of their marginal distributions. The authors show that, by applying a simple Gaussian copula transformation to the threshold and adding the resulting term as an extra regressor in each regime, one can purge the correlation between the threshold and the residuals. This “copula control” orthogonalizes threshold and innovations without needing instruments or specifying how the threshold itself is distributed. It preserves the usual linear-within-regime VAR structure and scales well with sample size and dimensionality.

Neglecting threshold endogeneity leads to two systematic errors in impulse response analysis. First, excess sensitivity: when the threshold and shocks move together, omitting the copula control makes the reduced-form shocks mechanically too volatile, because they incorporate a missing factor linked to the threshold. As a result, the on-impact responses to a given “structural” shock look too large. Second, excess propagation: the same omission induces omitted-variable bias in the VAR coefficients, so the estimated dynamics exaggerate persistence or even generate spurious oscillations. These problems arise under common identification schemes (recursive, sign restrictions, long-run restrictions, external instruments) because they all build on the same biased reduced-form moments. The copula-based correction is designed to remove both channels.

The authors assess performance in an extensive Monte Carlo study. They generate data from a TVAR with an endogenous threshold and then estimate two models: the correctly specified endogenous-threshold TVARN and a conventional exogenous-threshold TVAR ($\mathrm{TVAR_X}$). They vary the strength and sign of dependence between the threshold and shocks, allow for Gaussian, Student-t and skewed-t innovations, and consider both linear and nonlinear (Archimedean) copulas such as Clayton, Joe, and Gumbel. Across designs and for sample sizes that mimic typical macro applications ($T = 200$ and 500 quarters), ignoring threshold endogeneity produces substantial bias and mean squared error in the estimated threshold and in regime-specific covariance matrices. In turn, the impulse responses from the exogenous model display the excess sensitivity and propagation predicted by the theory. By contrast, the endogenous-threshold model substantially reduces both bias and MSE and closely tracks the true impulse responses, even when the true dependence structure is non-Gaussian and innovations are heavy-tailed.

The empirical application revisits U.S. monetary policy transmission over 1973–2007. The VAR includes inflation, the output gap, the federal funds rate, and the excess bond premium. The threshold variable is a three-quarter moving average of inflation, interpreted as a proxy for inflation expectations. The endogenous-threshold VAR estimates a threshold around 3.34 percent, implying that the economy is in a "high inflation" regime roughly half of the time. This regime classification aligns well with standard narratives: a prolonged high-inflation period in the 1970s and early 1980s, a subsequent Volcker disinflation, a long low-inflation regime during the Great Moderation, and brief episodes around the global financial crisis.

When the authors compute impulse responses to a contractionary monetary policy shock, the differences across models are stark. In the linear VAR, and in the mis-specified exogenous-threshold VAR, responses tend to show price puzzles, overly persistent dynamics, and sluggish disinflation. Once threshold endogeneity is controlled for, the responses become more economically plausible: prices fall more promptly, output adjusts in a way consistent with standard New Keynesian reasoning, and the uncertainty bands are tighter. The high-inflation regime exhibits somewhat stronger and faster disinflation, but not the extreme inertia implied by mis-specified models. A complementary New Keynesian sketch shows how regime-dependent volatility, when combined with nonlinear solution methods, can rationalize larger real effects in more volatile high-inflation states, linking the empirical findings to theory.

Overall, the paper's message is that, in threshold VAR applications, it is not enough to let coefficients change across regimes; one must also take seriously the possibility that the regime-switching variable is itself endogenous. The proposed copula-based control offers a practical way to do so. It delivers more reliable regime classification and more credible impulse responses, and it can be implemented with standard tools. Given how widely VARs are used to study state dependence and nonlinear dynamics, the approach is likely to be useful beyond the specific monetary policy application in the paper.

1. INTRODUCTION

This paper studies threshold VAR (TVAR) models in which regime switching is governed by an observed threshold variable (e.g., Tong and Lim, 1980, Chan, 1993, Hansen, 1999, 2000, Caner and Hansen, 2004). The tractability of such models notably rests on the assumption that the threshold variable is exogenous (or at least predetermined) with respect to the reduced-form innovations of the system. While this assumption may be defensible in natural-science applications where the threshold reflects a physical or geological feature, it is far less compelling in economics, where the indicator classifying the regime – e.g., inflation, financial stress – is typically determined jointly with the variables whose dynamics the VAR is meant to describe.

In such settings, regime assignment and shock realizations are simultaneous rather than sequentially ordered, and treating the threshold as exogenous misspecifies the conditional distribution of the innovations within each regime. The problem is thus one of "endogenous sample splitting" with a contemporaneous state variable: the threshold rule simultaneously determines (i) the partition of the sample into regimes and (ii) the conditional distribution of the innovations within each regime.

We propose a regime-specific copula-based control function that removes the threshold-selection component from the conditional mean of the reduced-form innovations without imposing a parametric form on the marginal distribution of the threshold variable, and preserving the linear regime-wise least-squares structure of the TVAR. To our knowledge, this is the first paper to address threshold endogeneity in a multivariate VAR setting.

Contributions. Our paper makes four main contributions. First, we derive an exact Gaussian-copula benchmark for endogenous sample splitting in threshold VARs. The contemporaneous threshold–innovation dependence is captured by a regime-specific PIT score, yielding a linear conditional-mean correction that allows correlations, variances, and selection loadings to differ across regimes while preserving least-squares estimation.

Second, we characterize the misspecification induced by ignoring the endogenous split. Omitting the copula control produces nonzero conditional innovation means, covariance inflation, and omitted-variable bias in the autoregressive coefficients. These distortions translate into impulse-response errors through *excess sensitivity*, operating through impact matrices, and *excess propagation*, operating through biased dynamic multipliers.

Third, we clarify structural identification under endogenous regimes. For covariance-based schemes, the correction removes the covariance inflation that feeds into impact identification. For proxy-SVAR identification, corrected innovations identify the orthogonalized target shock, while uncorrected innovations may identify a composite direction mixing the target shock with threshold-selection variation.

Fourth, we establish threshold-estimation validity with generated control scores. Because the threshold objective is locally one-sided rather than smoothly quadratic, threshold bias is characterized as a change in the one-sided slopes driven by a residualized omission cost. We then show that the generated, winsorized PIT controls and their threshold-dependent oracle counterparts are asymptotically negligible at the global and local rates needed for consistency and for the Chan-type threshold limit.

The paper makes two ancillary methodological points. First, the Gaussian-copula benchmark can be embedded in a sieve control built from Hermite polynomials in the PIT score, thus accommodating non-elliptical (e.g., tail-dependent) threshold–innovation dependence under standard sieve rate conditions without disturbing the limits for slope and IRF estimators. Second, we derive divergence rates (for the threshold, for regime-specific covariances) for the relative mean-squared-error penalty from ignoring threshold endogeneity, providing a sharp quantitative benchmark against which to interpret simulation evidence.

Monte Carlo and Empirical Application. We validate the method through Monte Carlo experiments that vary the sign and strength of threshold–innovation dependence, the distribution of innovations (Gaussian, Student-t , skewed-t), and the copula family generating the dependence (Gaussian, Clayton, Joe, Gumbel). The uncorrected TVAR exhibits large and systematic distortions in the threshold estimate, regime-specific covariances, and impulse responses; the corrected TVAR eliminates or sharply reduces them, with relative-MSE gains that increase with sample size in line with the misspecification mechanisms derived in the theory.

As an illustration, we apply the framework to the transmission of monetary policy shocks. This is a context in which state dependence is theoretically well motivated and yet empirical VAR analyses have remained predominantly linear, producing fragile and sometimes puzzling findings across specifications. The endogenous-threshold VAR delivers a historically plausible regime classification, more theoretically-cogent impulse responses, and substantially different sacrifice ratios across regimes - differences obscured by both the linear VAR and the exogenous-threshold TVAR.

Existing Literature. Several literatures address problems related to threshold endogeneity and provide important foundations for our approach. However, none deliver the combination of features required in this environment.

In single-equation threshold regressions, endogenous regime assignment has been treated using control-function and selectivity-type corrections in the spirit of Maddala (1983), Lee (1982, 1983), and Vella (1998). More recent contributions adapt these ideas directly to threshold regression models, using inverse-Mills-ratio, semiparametric, or copula-based corrections to absorb the conditional mean of the disturbance given the threshold variable (Kourtellos et al., 2016, Christopoulos et al., 2021, Kourtellos et al., 2022, Chen et al., 2023, Zhang et al., 2025). These papers establish threshold endogeneity as a distinct econometric problem and show that correcting the selection channel can be essential for valid inference.[1]

Extending this logic to a threshold VAR, however, is neither trivial nor mechanical. In a multivariate dynamic setting, the threshold variable may be contemporaneously related to an entire vector of innovations; regime-specific covariance matrices enter structural identification directly; impulse responses propagate through regime-specific dynamic multipliers; and the threshold parameter must be estimated jointly with the correction under a discontinuous, *kinked* population objective. Consequently, omitted threshold endogeneity distorts not only slope and threshold estimates, as in single-equation models, but also covariance matrices, impact effects, and the dynamic propagation of structural shocks. Recently Duffy and Mavroeidis (2026) studied identification in endogenously nonlinear SVARs, where regimes may be determined jointly with current endogenous variables and contemporaneous structural shocks. This reinforces our use of a contemporaneous threshold, but our focus is different: we retain the TVAR sample-splitting estimator and correct the conditional-mean misspecification induced by threshold-innovation dependence using a regime-specific copula control before applying standard structural identification schemes.

Other approaches address adjacent issues but fall short on at least one of these dimensions. IV methods require valid and relevant instruments, seldom available for threshold variables in macroeconomic settings (Gabaix and Koijen, 2024), and, even when available, do not purge the covariance distortion that feeds into impact identification. Smooth-transition VARs replace the discrete regime indicator with a continuous

[1] Zhang et al. (2025) study endogenous kink designs in a single-equation framework, which are related to our one-sided threshold objective in emphasizing non-smooth threshold behavior. They do not address the multivariate VAR setting in which regime-specific covariance matrices, structural identification, and impulse responses are directly affected by the endogenous sample split.

transition function, mitigating abrupt classification but not orthogonalizing the transition variable with respect to contemporaneous innovations. Markov-switching VARs, including time-varying transition-probability extensions, relocate regime variation to a latent state and typically treat observable conditioning variables as exogenous. State-dependent local projections (Miranda-Agrippino and Ricco, 2021) allow heterogeneous responses by interacting shocks with observed states but inherit analogous impact and propagation distortions when the state and innovations are correlated. Full-information likelihood methods could in principle model the joint distribution of the threshold variable and innovations, but they require correct specification of the entire dependence structure and involve non-convex optimization that scales poorly in VAR systems.

Organization. Section 2 presents the standard threshold VAR, followed by our endogenous threshold VAR. Section 3 demonstrates how neglected threshold endogeneity imparts systematic static and dynamic errors. Section 4 describes estimation. Section 5 considers departures from elliptical dependence structures through approximations and sieve-based control corrections. Section 6 contains the Monte Carlo. Section 7 applies the framework to understanding the impact of unanticipated monetary shocks. Section 8 concludes.

2. An Endogenous Threshold VAR

Our point of departure is the following threshold Vector Autoregression (TVAR) model, with up to p lags and i regimes (hereafter, assumed to be two):[2]

$$y_t = \begin{cases} \mathrm{A}^{(1)}\mathrm{Y}_{t-1} + u_t^{(1)}, & \text{if } z_t \leq \delta \\ \\ \mathrm{A}^{(2)}\mathrm{Y}_{t-1} + u_t^{(2)}, & \text{if } z_t > \delta \end{cases} \tag{1}$$

where y_t is a $\mathrm{K} \times 1$ vector of variables; $\mathrm{Y}_{t-1} = (1, y'_{t-1}, \dots, y'_{t-p})'$, $\mathrm{A}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times(\mathrm{Kp}+1)}$ denote regime-specific coefficients; $t = \{p+1, \dots, \mathrm{T}\}$; and where $u_t^{(\mathrm{i})} = \big(u_{1t}^{(\mathrm{i})}, \cdots, u_{\mathrm{K}t}^{(\mathrm{i})}\big)'$ is a $\mathrm{K} \times 1$ vector of the reduced form residuals with regime specific unconditional variances $\Sigma_{\mathrm{uu}}^{(\mathrm{i})}$. Variable z_t is the threshold variable, whose support can be partitioned into $\mathrm{Z}^{(1)} = \{z_t : z_t \leq \delta\}$ and $\mathrm{Z}^{(2)} = \{z_t : z_t > \delta\}$, implying that the model is segmented over the values of z_t. Superscript $'$ indicates transpose.

Threshold endogeneity arises when the threshold variable is contemporaneously correlated with the innovation vector, that is $\mathbb{E}\big(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}\big) \neq 0$, where $z_t^{(\mathrm{i})}$ denotes the re-

[2] For convenience, Appendix D collects our main notation.

alization of the threshold variable z_t conditional on $z_t \in \mathrm{Z}^{(\mathrm{i})}$, so that both $u_t^{(\mathrm{i})}$ and $z_t^{(\mathrm{i})}$ are evaluated within regime i. In this case the regime indicator is correlated with the innovations driving the VAR dynamics. Such dependence may arise if the threshold variable reflects forward-looking or market-based indicators, such as expectations, financial or macroeconomic conditions that incorporate information about current shocks. Alternatively, sufficiently large structural shocks may move the threshold variable across the threshold, so that regime switching becomes endogenous and shocks determine which regime governs the system *at impact*. This mechanism may arise even when the standard VAR orthogonality condition holds, $\mathbb{E}(u_t^{(\mathrm{i})} \mid Y_{t-1}) = 0$, because the threshold variable is related to contemporaneous economic conditions affected by the innovations.

This idea relates to a broad literature on nonlinear macroeconomic dynamics in which regime transitions depend on underlying economic conditions. Early endogenous switching frameworks (Maddala, 1983, Lee, 1983) allow regime selection to depend on observables, while in VAR settings Sims and Zha (2006) document state dependence consistent with shocks interacting with the business cycle. In models with occasionally binding constraints (e.g., Guerrieri and Iacoviello, 2017), sufficiently large shocks can trigger regime changes when constraints bind. Similarly, state-dependent pricing models (Golosov and Lucas, 2007, Alvarez and Lippi, 2014) imply that adjustment decisions respond to contemporaneous shocks, even without an explicit threshold variable. The structural break literature (Engle and Smith, 1999, Kapetanios and Tzavalis, 2010) likewise associates large shocks with discrete regime changes.

These contributions suggest that shocks and regime selection need not be independent, motivating a specification in which the threshold variable is contemporaneously correlated with VAR innovations. By contrast, standard threshold VARs typically employ lagged transition variables to mitigate endogeneity; while convenient, this may attenuate or delay regime changes driven by contemporaneous shocks and rules out impact regime switching by construction.

2.1 *Correcting For Threshold Endogeneity Using Copulas*

To model the dependence between z_t and $u_t^{(\mathrm{i})}$, we use copulas to define the conditional mean correction $\mathbb{E}[u_t^{(\mathrm{i})} \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}]$. Copulas are designed to describe the joint dependence structure between random variables. Sklar (1959)'s theorem states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula.

Copulas have emerged as a pivotal framework for capturing complex dependency structures in fields as varied as quantitative finance, predictive meteorology, and system reliability engineering. In a univariate context, copulas have been suggested by Park and Gupta (2012) to tackle the endogeneity problem. Moreover, within the *single-equation* framework – see Kourtellos et al. (2016), Kourtellos et al. (2022), Christopoulos et al. (2021), Chen et al. (2023) – threshold endogeneity bias has been addressed using control variables to adjust the conditional mean for the bias. This approach is in the spirit of binary-choice selectivity models that allow for endogenous regime switching, e.g., Maddala (1983), Lee (1982, 1983), Vella (1998). It relies on a limited information estimation framework to adjust for the endogeneity of the threshold variable. However, the extension to threshold VARs has not yet been developed. Given the growing popularity and proven efficacy of threshold VARs, bridging this gap represents a critical next step for the field.

The next assumption formalizes the key building block.

Assumption 1 (Relation between threshold and innovations).

1. *Threshold variable z_t has a continuous, strictly increasing* CDF F_z.[3]
2. *Threshold endogeneity:* $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) \neq 0_{\mathrm{K}}$, *for at least one regime* i.
3. *Conditional mean independence from lagged information and corrected innovation:* $\mathbb{E}(u_t^{(\mathrm{i})} \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}) = \mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$. *Define* $\varepsilon_t^{(\mathrm{i})} = u_t^{(\mathrm{i})} - \mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$, *then* $\mathbb{E}(\varepsilon_t^{(\mathrm{i})} \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}) = 0_{\mathrm{K}}$.
4. *Serially uncorrelated corrected innovations:* $\mathbb{E}(\varepsilon_t^{(\mathrm{i})} \varepsilon_s^{(\mathrm{i})\prime}) = 0_{\mathrm{K}\times\mathrm{K}}$, *for* $s \neq t$.

The conditions in Assumption 1 define the source of threshold endogeneity and the innovation object used for estimation. Condition 1 requires the threshold variable z_t to have a continuous and strictly increasing distribution function. Condition 2 allows the contemporaneous threshold variable to contain information about the reduced-form innovation, so that $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) \neq 0_{\mathrm{K}}$; hence z_t is not assumed to be exogenous.

Condition 3 is the key conditional-mean sufficiency restriction. It requires that, once $z_t^{(\mathrm{i})}$ is observed, the lagged VAR information Y_{t-1} does not add further predictive content for the innovation mean. Equivalently, the contemporaneous threshold-related

[3]The continuity requirement of z_t precludes using discrete or categorical variables, such as a business-cycle chronology.

component of the innovation mean is summarized by $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$. Subtracting this component defines the corrected innovation $\varepsilon_t^{(\mathrm{i})}$, which satisfies $\mathbb{E}(\varepsilon_t^{(\mathrm{i})} \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}) = 0_{\mathrm{K}}$.

Condition 4 imposes serial uncorrelation on the corrected innovations, rather than on the uncentered $u_t^{(\mathrm{i})}$. This formulation is most relevant when z_t captures the contemporaneous economic state and may therefore be correlated with the same-period innovation. A lagged threshold $z_{t-\mathrm{L}}$, $\mathrm{L} \geq 1$, would typically make regime assignment predetermined and mechanically avoid this contemporaneous endogeneity channel, but it would also rule out impact regime switching and may delay or attenuate the state dependence that the model is designed to capture. Below, we impose a Gaussian-copula benchmark under which $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$ admits a linear control-function representation.

To implement copulas in the two-regime context, define the regime-specific truncations (around threshold δ):[4]

$$\mathrm{F}_{z_t}^{(1)} \equiv \frac{\mathrm{F}_{z_t}}{p_\delta} \quad \text{for } z_t \leq \delta \quad \text{and} \quad \mathrm{F}_{z_t}^{(2)} \equiv \frac{\mathrm{F}_{z_t} - p_\delta}{1 - p_\delta} \quad \text{for } z_t > \delta, \tag{2}$$

where, for brevity, we denote $\mathrm{F}_{z_t} = \mathrm{F}_z(z_t)$, where $\mathrm{F}_z(\bullet)$ denotes the marginal CDF of z_t. This convention extends to all (marginal, conditional, and joint) distribution functions used throughout the paper. These distributions are obtained by scaling the CDF by the associated probability, $p_\delta = \Pr(z_t \leq \delta) \in (0,1)$, so that they integrate to unity and constitute proper distribution functions. The definition below defines copula functions to represent the conditional distribution.

Definition 1 (Copula Representation of Conditional Joint Distribution). *Consider two $\mathrm{K}+1$-dimension copulas $\mathrm{C}^{(\mathrm{i})} : [0,1]^{\mathrm{K}+1} \to [0,1]$:*

$$\mathrm{F}_{u_t|z_t}^{(\mathrm{i})} \equiv \frac{\partial}{\partial \mathrm{F}_{z_t}^{(\mathrm{i})}} \mathrm{C}^{(\mathrm{i})}\left(\mathrm{F}_{u_t}^{(\mathrm{i})}, \mathrm{F}_{z_t}^{(\mathrm{i})}\right), \tag{3}$$

where $u_t^{(\mathrm{i})} = \left(u_{1t}^{(\mathrm{i})}, \ldots, u_{\mathrm{K}t}^{(\mathrm{i})}\right)'$, $\mathrm{F}_{u_t}^{(\mathrm{i})} = \left(\mathrm{F}_{u_{1t}}^{(\mathrm{i})}, \ldots, \mathrm{F}_{u_{\mathrm{K}t}}^{(\mathrm{i})}\right)'$ and $\mathrm{F}_{u_{kt}}$ denote marginal distributions of innovation terms $u_{kt}^{(\mathrm{i})}$.[5] See also Siburg and Stoimenov (2008).

[4] Under continuity of F_z, both $\mathrm{F}_{z_t}^{(1)}$ and $\mathrm{F}_{z_t}^{(2)}$ are non-decreasing and right-continuous, with

$$\lim_{z \to -\infty} \mathrm{F}_{z_t}^{(1)}(z) = 0, \quad \mathrm{F}_{z_t}^{(1)}(\delta) = 1, \quad \mathrm{F}_{z_t}^{(2)}(\delta^+) = 0, \quad \lim_{z \to \infty} \mathrm{F}_{z_t}^{(2)}(z) = 1,$$

and are therefore proper CDFs on the supports $(-\infty, \delta]$ and (δ, ∞), respectively.

[5] See, for example Erdely (2017). The copula functions C_i, defined by (3), can be thought of as scaled functions of gluing copulas along the support interval of threshold variable for $z_t \leq \delta$ and $z_t > \delta$.

Based on (3), we can write the conditional density of vector $u_t^{(\mathrm{i})}$ on $z_t^{(\mathrm{i})}$ as,

$$f_{u_t|z_t}^{(\mathrm{i})} = c^{(\mathrm{i})}\left(\mathrm{F}_{u_t}^{(\mathrm{i})}, \mathrm{F}_{z_t}^{(\mathrm{i})}\right) \times \prod_{k=1}^{\mathrm{K}} f_{u_{kt}}^{(\mathrm{i})}, \tag{4}$$

where

$$c^{(\mathrm{i})}\left(\mathrm{F}_{u_t}^{(\mathrm{i})}, \mathrm{F}_{z_t}^{(\mathrm{i})}\right) \equiv \frac{\partial^{\mathrm{K}+1}}{\partial \mathrm{F}_{u_{1t}}^{(\mathrm{i})} \dots \partial \mathrm{F}_{u_{\mathrm{K}t}}^{(\mathrm{i})} \, \partial \mathrm{F}_{z_t}^{(\mathrm{i})}} C^{(\mathrm{i})}\left(\mathrm{F}_{u_t}^{(\mathrm{i})}, \mathrm{F}_{z_t}^{(\mathrm{i})}\right) \tag{5}$$

is the copula density of $u_t^{(\mathrm{i})}$ and $z_t^{(\mathrm{i})}$, and $f_{u_{kt}}^{(\mathrm{i})}$ are the marginal densities of error terms $u_{kt}^{(\mathrm{i})}$. The conditional density $f_{u_t|z_t}^{(\mathrm{i})}$, defined in (4), enables us to derive an analytic expression, under a known copula distribution $\mathrm{C}^{(\mathrm{i})}$ (or density $c^{(\mathrm{i})}$) and the marginal distribution of error terms $u_{kt}^{(\mathrm{i})}$, for the conditional expectation $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$.

Assume $u_{kt}^{(\mathrm{i})} \sim \mathcal{N}(0, (\sigma_{uk}^{(\mathrm{i})})^2)$ and that $\mathrm{C}^{(\mathrm{i})}$ is the Gaussian copula. Then, define the copula control (Probability Integral Transform, PIT) conditional on each regime i by,

$$\begin{aligned} z_t^{(\mathrm{i})\star} &\equiv \Phi^{-1}(\mathrm{F}_{z_t}^{(\mathrm{i})}) \ \sim \mathcal{N}(0,1), \\ u_{kt}^{(\mathrm{i})\star} &\equiv \Phi^{-1}(\mathrm{F}_{u_{kt}}^{(\mathrm{i})}) \sim \mathcal{N}(0,1). \end{aligned} \tag{6}$$

Thus, both are transformed to be standard normal regardless of their original distribution.[6]

Under the regime-specific Gaussian-copula benchmark maintained in this paper, the PIT-transformed innovation vector $u_t^{(\mathrm{i})\star}$ and the threshold score $z_t^{(\mathrm{i})\star}$ are jointly normal within each regime. This joint-normal representation yields the following single-factor control representation:

$$u_t^{(\mathrm{i})} = \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star} + \varepsilon_t^{(\mathrm{i})}, \tag{7}$$

where $\varepsilon_t^{(\mathrm{i})}$ is the *copula-control adjusted* innovation, and $\Lambda^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times 1}$ is a regime-specific loading vector. Conditional on $z_t^{(\mathrm{i})}$, we then have the distributions,

$$\begin{aligned} u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})} &\sim \mathcal{N}\left(\Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}, \ \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right), \\ \varepsilon_t^{(\mathrm{i})} &\sim \mathcal{N}\left(0, \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right), \end{aligned} \tag{8}$$

given that

[6] Strictly speaking, given our assumption on $u_{kt}^{(\mathrm{i})}$, this would imply $u_{kt}^{(i)\star} \equiv \Phi^{-1}\left(\Phi\left(\frac{u_{kt}^{(\mathrm{i})}}{\sigma_{uk}^{(\mathrm{i})}}\right)\right) \equiv \frac{u_{kt}^{(\mathrm{i})}}{\sigma_{uk}^{(\mathrm{i})}}$.

$$\begin{aligned} \mathbb{E}\left(\varepsilon_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}\right) &= 0_{\mathrm{K}} \\ \mathbb{V}\mathrm{ar}\left(\varepsilon_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}\right) &= \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}}, \end{aligned} \tag{9}$$

with regime-specific covariance matrix $\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$ being constant within regime i. The following proposition formalizes this representation.

Proposition 1 (Single-Factor Control Under Gaussian Copula). *Define the regime-i PIT score of the threshold variable as* $z_t^{(\mathrm{i})\star} = \Phi^{-1}\left(\mathrm{F}_z^{(\mathrm{i})}\right)$, *and likewise define* $u_{kt}^{(\mathrm{i})\star} = \Phi^{-1}\left(\mathrm{F}_{u_{kt}}^{(\mathrm{i})}\right)$, $k = 1, \ldots, \mathrm{K}$. *Collect*

$$u_t^{(\mathrm{i})\star} \equiv \left(u_{1t}^{(\mathrm{i})\star}, \ldots, u_{\mathrm{K}t}^{(\mathrm{i})\star}\right)'. \tag{10}$$

Assume the marginals of $u_{kt}^{(\mathrm{i})}$ *are Gaussian* $\mathcal{N}(0, (\sigma_{u_k}^{(\mathrm{i})})^2)$, *and set* $\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} = diag\left(\sigma_{u_1}^{(\mathrm{i})}, \ldots, \sigma_{u_K}^{(\mathrm{i})}\right)$ *so that* $u_t^{(\mathrm{i})\star} = \left(\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}\right)^{-1} u_t^{(\mathrm{i})}$ *component-wise. Define* $\mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})} = \mathbb{C}\mathrm{orr}\left(u_t^{(\mathrm{i})\star}\right) \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}}$ *and* $r_{uz}^{(\mathrm{i})} = corr\left(u_t^{(\mathrm{i})\star}, z_t^{(\mathrm{i})\star}\right) \in \mathbb{R}^{\mathrm{K}\times 1}$. *Assume the copula between* $u_t^{(\mathrm{i})\star}$ *and* $z_t^{(\mathrm{i})\star}$ *is Gaussian, and assume only continuity for* $\mathrm{F}_z^{(\mathrm{i})}$. *Other than continuity, no other parametric restriction is required for* $\mathrm{F}_z^{(\mathrm{i})}$. *Then, the regime-$i$ innovations admit the single-factor representation,*

$$u_t^{(\mathrm{i})} = \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star} + \varepsilon_t^{(\mathrm{i})}, \tag{11}$$

with $\mathbb{E}\left(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}\right) = \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}$, *since* $\mathbb{E}\left(\varepsilon_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}\right) = 0_{\mathrm{K}}$, *where* $\Lambda^{(\mathrm{i})} = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} r_{uz}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times 1}$, *with*[7]

$$\mathbb{V}\mathrm{ar}(\varepsilon_t^{(\mathrm{i})}) = \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}\left(\mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})} - r_{uz}^{(\mathrm{i})} r_{uz}^{(\mathrm{i})\prime}\right)\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} \quad \textit{and} \quad \mathbb{V}\mathrm{ar}(u_t^{(\mathrm{i})}) = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}\mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})}\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}. \tag{12}$$

Invariance of the z_t distribution. Note that, whilst we exploit the Gaussian-copula dependence to obtain the relationship $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) = \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}$ in (7), we impose *no parametric assumption* on the marginal distribution of z_t. We estimate F_z nonparametrically; the regime-specific CDFs $\mathrm{F}_z^{(\mathrm{i})}$ and the PIT-Gaussian transform then imply $z_t^{(\mathrm{i})\star} \sim \mathcal{N}(0,1)$ by construction. The distribution $\mathrm{F}_z^{(\mathrm{i})}$ can be efficiently estimated from the data, in a first step, by a nonparametric method (e.g., Silverman, 1986), or by using the empirical cumulative distribution function (ECDF), implying that the sample estimates of $z_t^{(\mathrm{i})\star}$ are consistent (see Lemma A.2, Appendix A).

[7] Note $\mathrm{D} \equiv \mathrm{D}'$.

That there is no need to know the distribution of z_t constitutes an attractive feature of the copula method. Furthermore, in some applications, one might imagine z_t to follow different distributions across the regions $\mathrm{Z}^{(1)} = \{-\infty < z_t \leq \delta\}$ and $\mathrm{Z}^{(2)} = \{\delta < z_t < \infty\}$, which may be characterized by different degrees of asymmetry or tail behavior. This consideration arises because, whilst the transformed variable $z_t^{(\mathrm{i})\star}$ is normal, this *need not necessarily be true* for the original variable z_t. The assumption that $u_{kt}^{(\mathrm{i})}$, for all k, are normally distributed is commonly adopted in economics, both in single and multivariate regression models, primarily for analytical convenience; see, for example, Spanos (2018). This assumption is particularly useful in small samples and implies two appealing properties of the error distribution: symmetry and a reasonable amount of probability mass in the tails, both of which are typically regarded as plausible features of well-specified econometric models.

2.2 *Endogenous Threshold VAR form*

Given Proposition 1, and the single factor representation (7), we can augment the TVAR model (1) to control for threshold endogeneity as follows:

$$y_t = \begin{cases} \mathrm{A}^{(1)}\mathrm{Y}_{t-1} + \Lambda^{(1)} z_t^{(1)\star} + \varepsilon_t^{(1)}, & \text{if } z_t \leq \delta \\ \\ \mathrm{A}^{(2)}\mathrm{Y}_{t-1} + \Lambda^{(2)} z_t^{(2)\star} + \varepsilon_t^{(2)}, & \text{if } z_t > \delta \end{cases} \tag{13}$$

By Proposition 1, the copula control is orthogonal to the corrected innovation in each regime, $\mathbb{E}\big[z_t^{(\mathrm{i})\star}\, \varepsilon_t^{(\mathrm{i})'}\big] = 0$. The conditional mean of y_t given Y_{t-1} and $\mathrm{Z}^{(\mathrm{i})}$ is then,

$$\mathbb{E}\big(y_t \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}\big) \; = \; \mathrm{A}^{(\mathrm{i})}\mathrm{Y}_{t-1} + \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}, \tag{14}$$

where the copula transformation term $z_t^{(\mathrm{i})\star}$ serves as a control function to capture the correlation structure between the threshold variable and the TVAR vector of innovations $u_t^{(\mathrm{i})}$.

The vector of loading coefficients $\Lambda^{(\mathrm{i})}$, moreover, admits a clear interpretation, namely the strength of the selection channel across regimes, while retaining the linearity and parsimony of the VAR presentation. The corrected innovation term $\varepsilon_t^{(\mathrm{i})}$ is orthogonal to the regime-specific threshold score $\mathrm{Z}^{(\mathrm{i})}$. Two additional points are worth noting. First, the conditional mean representation in (14) depends only on the contemporaneous realization of z_t and therefore does not require specifying the dynamic process governing

z_t. Second,(14) nests the exogenous TVAR as the special case $\Lambda^{(\mathrm{i})} = 0\ \forall i$.[8] Hereafter, we label these endogenous and exogenous TVAR cases as $\mathrm{TVAR_N}$ and $\mathrm{TVAR_X}$, respectively.

Estimates of $\varepsilon_t^{(\mathrm{i})}$ from (13) can be used to obtain the structural errors of a structural TVAR model (say $e_t^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times 1}$) net of the influence of an endogenous threshold variable z_t. We can then use these structural errors to perform IRF analysis of y_t with respect to $\varepsilon^{(\mathrm{i})}$ based on a triangular, or a structural, factorization of its variance-covariance, $\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$. However, as we shall discuss, this could also be applied to essentially any other identification scheme, e.g., the external instruments approach proposed by the foundational papers of Stock and Watson (2018) and Mertens and Ravn (2013, 2014).

3. Misspecification Effects of Ignoring Threshold Endogeneity

Ignoring threshold endogeneity gives rise to several misspecification errors. As shown above, if the copula control is omitted, the $\mathrm{TVAR_X}$ residual contains the component $\Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}$. Hence, first, the conditional mean of the uncorrected innovation is not zero, $\mathbb{E}\left(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}\right) = \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}$. Second, omitting the control inflates the regime-specific covariance matrix of the reduced-form innovations. Third, if $z_t^{(\mathrm{i})\star}$ is correlated with the lagged VAR regressors, the omission induces bias in the regime-specific slope coefficients and distorts dynamic propagation. These channels affect structural identification, impulse responses, threshold estimation, and finite-sample stability. We examine them in turn.

We first formalize the covariance implication. The following lemma shows that omitting the copula control always makes the shocks excessively volatile within each regime i.

Lemma 1 (Variance Inflation under Misspecification). *Assume that the* $\mathrm{TVAR_N}$ *is the true model. Then, the covariance of the misspecified* $\mathrm{TVAR_X}$ *innovations,* $\Sigma_{\mathrm{uu}}^{(\mathrm{i})}$, *satisfies*

$$\Sigma_{\mathrm{uu}}^{(\mathrm{i})} = \Lambda^{(\mathrm{i})}\Lambda^{(\mathrm{i})\prime} + \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} \succeq \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}, \tag{15}$$

with equality if $\Lambda^{(\mathrm{i})} = 0$. *Hence, the corrected innovations* $\varepsilon_t^{(\mathrm{i})}$ *are (weakly) less variable. The Loewner ordering* $(\succeq)$ *follows because* $\Lambda^{(\mathrm{i})}\Lambda^{(i)\prime}$ *is positive semi-definite (psd).*

Proof. Appendix A.2 □

We next turn to the dynamic-coefficient implication. If $\Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star}$ is omitted from (13), the misspecified $\mathrm{TVAR_X}$ absorbs part of the threshold-selection component into the

[8] Standard threshold VARs often use a lagged threshold variable, making regime assignment predetermined. This avoids contemporaneous threshold endogeneity mechanically, but rules out impact regime switching.

autoregressive coefficients. The population least-squares limit in regime i therefore satisfies [9]

$$\Delta \mathrm{A}^{(\mathrm{i})} \equiv (\mathrm{A}_{\mathrm{X}}^{(\mathrm{i})} - \mathrm{A}_{\mathrm{N}}^{(\mathrm{i})}) = \Lambda^{(\mathrm{i})}\gamma^{(\mathrm{i})}, \qquad \gamma^{(\mathrm{i})} = \mathbb{E}\Big(z_t^{(\mathrm{i})\star}\mathrm{Y}_{t-1}' \mid \mathrm{Z}^{(\mathrm{i})}\Big)\left(Q_{\mathrm{YY}}^{(\mathrm{i})}\right)^{-1}, \tag{16}$$

where $Q_{\mathrm{YY}}^{(\mathrm{i})} = \mathbb{E}\Big(\mathrm{Y}_{t-1}\mathrm{Y}_{t-1}' \mid \mathrm{Z}^{(\mathrm{i})}\Big)$.[10] Here $\mathrm{A}_{\mathrm{N}}^{(\mathrm{i})}$ and $\mathrm{A}_{\mathrm{X}}^{(\mathrm{i})}$ denote, respectively, the population slope matrices under the correctly specified $\mathrm{TVAR}_{\mathrm{N}}$ and the misspecified $\mathrm{TVAR}_{\mathrm{X}}$.

Equation (16) follows from the regime-i population least-squares projection

$$\mathrm{A}_{\mathrm{X}}^{(\mathrm{i})} = \mathbb{E}\Big(y_t\mathrm{Y}_{t-1}' \mid \mathrm{Z}^{(\mathrm{i})}\Big)\left(Q_{\mathrm{YY}}^{(\mathrm{i})}\right)^{-1},$$

after substituting $y_t = \mathrm{A}_{\mathrm{N}}^{(\mathrm{i})}\mathrm{Y}_{t-1} + \Lambda^{(\mathrm{i})}z_t^{(\mathrm{i})\star} + \varepsilon_t^{(\mathrm{i})}$ and using $\mathbb{E}\Big(\varepsilon_t^{(\mathrm{i})}\mathrm{Y}_{t-1}' \mid \mathrm{Z}^{(\mathrm{i})}\Big) = 0$.

The covariance inflation in Lemma 1 and the dynamic coefficient bias in (16) have implications for system stability in finite samples. We formalize these implications using the companion matrix representation and the notion of a stability margin.[11]:

$$x_t = \mathcal{A}^{(\mathrm{i})}x_{t-1} + J'u_t, \qquad y_t = Jx_t, \tag{17}$$

where $x_t = \left(y_t', y_{t-1}', \ldots, y_{t-p+1}'\right)' \in \mathbb{R}^{\mathrm{K}p}$, $\mathcal{A}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}p\times\mathrm{K}p}$ is the regime-i companion matrix, and $J' \in \mathbb{R}^{\mathrm{K}p\times\mathrm{K}}$ maps the innovation vector into the companion state.

Definition 2 (Stability margin). *Let $\mathcal{A}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}p\times\mathrm{K}p}$ denote a regime-i companion matrix and let $\rho(\cdot)$ denote its spectral radius. For an arbitrary set, $\{\widetilde{A} : \rho(\widetilde{A}) \geq 1\}$, define*

$$\mathfrak{r}_{\mathrm{stab}}^{(\mathrm{i})} \equiv \inf_{\widetilde{\mathcal{A}}:\rho(\widetilde{\mathcal{A}})\geq 1} \left\|\widetilde{\mathcal{A}} - \mathcal{A}^{(\mathrm{i})}\right\|_F. \tag{18}$$

If $\rho(\mathcal{A}^{(\mathrm{i})}) < 1$, then $\mathfrak{r}_{\mathrm{stab}}^{(\mathrm{i})} > 0$, where $\|\cdot\|_F$ denotes the Frobenius norm.

Lemma 2 (Variance-based finite-sample stability bounds). *Let $\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}$ denote the regime-i companion matrix under the correctly specified $\mathrm{TVAR}_{\mathrm{N}}$ model, and let*

[9] Here $\mathrm{Z}^{(\mathrm{i})}$ denotes the regime event/support region, while $z_t^{(\mathrm{i})}$ denotes the threshold realization within regime i. We condition on $\mathrm{Z}^{(\mathrm{i})}$ for regime-level projection moments, and on $z_t^{(\mathrm{i})}$ for pointwise conditional-mean statements.

[10] We assume $Q_{\mathrm{YY}}^{(\mathrm{i})}$ is invertible.

[11] Since stability is governed by the autoregressive companion matrix, we omit the intercept from the companion-form notation for simplicity; the intercept is included in estimation but does not affect the spectral-radius stability condition

$\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}$ denote the pseudo-true companion matrix under the misspecified $\mathrm{TVAR}_{\mathrm{X}}$ *model. Let*

$$\mathfrak{r}_{\mathrm{N,stab}}^{(\mathrm{i})} = \inf_{\widetilde{\mathcal{A}}:\rho(\widetilde{\mathcal{A}})\geq 1} \left\| \widetilde{\mathcal{A}} - \mathcal{A}_{\mathrm{N}}^{(\mathrm{i})} \right\|_F \tag{19}$$

be the stability margin around the correctly specified companion matrix. For any radius $\mathfrak{r} \in (0, \mathfrak{r}_{\mathrm{N,stab}}^{(\mathrm{i})})$, *the corrected estimator satisfies*

$$\Pr\Big(\rho(\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})}) < 1\Big) \geq 1 - \frac{V_{\mathrm{N}}^{(\mathrm{i})}}{\mathfrak{r}^2}, \tag{20}$$

where $V_{\mathrm{N}}^{(\mathrm{i})} = \frac{1}{\mathrm{T}^{(\mathrm{i})}}\mathrm{tr}\left\{\mathcal{R}\left[\left(Q_{xx,\mathrm{N}}^{(\mathrm{i})}\right)^{-1} \otimes \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right]\mathcal{R}'\right\}$ *is the first-order LS approximation to* $\mathbb{E}\|\widehat{A}_N^{(i)} - A_N^{(i)}\|_F^2$, *after embedding the slope-coefficient error into companion form. The misspecified estimator satisfies*

$$\Pr\Big(\rho(\widehat{\mathcal{A}}_{\mathrm{X}}^{(\mathrm{i})}) < 1\Big) \geq 1 - \frac{V_{\mathrm{X}}^{(\mathrm{i})} + \left\|\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})} - \mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F^2}{\mathfrak{r}^2}, \tag{21}$$

where $V_{\mathrm{X}}^{(\mathrm{i})} = \frac{1}{\mathrm{T}^{(\mathrm{i})}}\mathrm{tr}\left\{\mathcal{R}\left[\left(Q_{xx,\mathrm{X}}^{(\mathrm{i})}\right)^{-1} \otimes \Sigma_{\mathrm{uu}}^{(\mathrm{i})}\right]\mathcal{R}'\right\}$ *is the approximation to* $\mathbb{E}\|\widehat{A}_X^{(i)} - A_X^{(i)}\|_F^2$. *Here,* $Q_{xx,\bullet}^{(\mathrm{i})} = \mathbb{E}\Big(x_{t-1}x_{t-1}' \mid \mathrm{Z}^{(\mathrm{i})}\Big)$ *with* $\bullet \in \{\mathrm{N},\mathrm{X}\}$, *denotes the population second-moment matrix of the companion-form regressors under the corrected and misspecified specifications. Moreover,* $\mathcal{R} \in \mathbb{R}^{(\mathrm{K}p)^2\times\mathrm{K}^2p}$ *is the fixed linear selection matrix satisfying* $\mathrm{vec}\left(\widehat{\mathcal{A}}^{(\mathrm{i})} - \mathcal{A}^{(\mathrm{i})}\right) = \mathcal{R}\,\mathrm{vec}\left(\widehat{B}^{(\mathrm{i})} - B^{(\mathrm{i})}\right)$, *where* $B^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}p}$ *denotes the regime-i VAR slope matrix entering the first block row of the companion matrix.*

PROOF. Appendix A.3. □

Lemma 2 shows that ignoring threshold endogeneity weakens finite-sample stability guarantees through two channels. First, the misspecified model has a larger innovation covariance matrix, since $\Sigma_{\mathrm{uu}}^{(\mathrm{i})} = \Lambda^{(\mathrm{i})}\Lambda^{(\mathrm{i})\prime} + \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} \succeq \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$. This increases the sampling-variation component of the stability bound through $V_{\mathrm{X}}^{(\mathrm{i})}$. Second, omitting the copula control shifts the pseudo-true companion matrix away from the correctly specified one, as captured by $\left\|\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})} - \mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F^2$. Thus, the misspecified $\mathrm{TVAR}_{\mathrm{X}}$ may display spurious finite-sample instability either because innovation variance is inflated or because the autoregressive dynamics are biased. By contrast, the corrected $\mathrm{TVAR}_{\mathrm{N}}$ removes the threshold-selection component from the reduced-form innovation. This reduces the covariance matrix from $\Sigma_{\mathrm{uu}}^{(\mathrm{i})}$ to $\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$ and removes the companion-form omitted-control bias relative to the corrected dynamics. Consequently, the stability bound is tightened when threshold endogeneity is controlled for.

3.1 *Impulse Responses under Misspecification*

Let $\mathbb{M}$ denote an identification procedure that maps the relevant reduced-form objects into an impact matrix $P_\bullet^{(\mathrm{i})}$. For covariance-based schemes, such as recursive or sign-restricted identification, the impact matrix satisfies $P_\bullet^{(\mathrm{i})} P_\bullet^{(\mathrm{i})\prime} = \Sigma_\bullet^{(\mathrm{i})}, \bullet \in \{\varepsilon\varepsilon, \mathrm{uu}\}$. For proxy-SVAR identification, the same notation refers to the impact column identified from the corrected or uncorrected reduced-form innovations and the external instrument. In what follows, when the subscript is applied to objects other than covariance matrices, such as slope coefficients, companion matrices, multipliers, or IRFs, we use $\bullet \in \{\mathrm{N}, \mathrm{X}\}$ to distinguish the copula-corrected $\mathrm{TVAR_N}$ from the uncorrected $\mathrm{TVAR_X}$.

Consider the companion form of the TVAR model given by (17), with reduced-form multiplier $\Psi_\bullet^{(\mathrm{i})}(h) = J(\mathcal{A}_\bullet^{(\mathrm{i})})^h J'$. Structural IRFs summarize the dynamic effects of a shock e_j at horizon h:

$$\mathrm{IRF}_\bullet^{(\mathrm{i})}(h, j) \equiv \Psi_\bullet^{(\mathrm{i})}(h) \mathrm{P}_\bullet^{(\mathrm{i})} e_j. \tag{22}$$

Two systematic distortions arise over horizon h if the maintained model erroneously omits the copula control, which we term "excess sensitivity" and "excess propagation" errors.

Excess Sensitivity refers to the impact distortion induced by covariance inflation. Because the reduced-form covariance matrix is inflated when the copula control is omitted, the associated impact matrix must reflect this excess variance. For regime i, let identification map $\mathbb{M}$ return $\mathrm{P}_\bullet \mathrm{P}_\bullet{}' = \Sigma_\bullet$. If the maintained model erroneously omits the copula control, then, following condition (15) and Lemma 1, we have:

$$\left\|\mathrm{P}_{\mathrm{uu}}^{(\mathrm{i})}\right\|_F^2 - \left\|\mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})}\right\|_F^2 = \mathrm{tr}\left(\Sigma_{\mathrm{uu}}^{(\mathrm{i})} - \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right) = \left\|\Lambda^{(\mathrm{i})}\right\|_F^2 \geq 0. \tag{23}$$

Consequently, the misspecified model assigns greater total impact variance to the identified shocks; in particular, at least one identified impact column must have larger norm at $h = 0$ under $\mathrm{TVAR_X}$, with equality only if $\Lambda^{(\mathrm{i})} = 0_{\mathrm{K}}$.

The distortion arises because omitting the copula control inflates the reduced-form variance: $\Sigma_{\mathrm{uu}}^{(\mathrm{i})} - \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} = \Lambda^{(\mathrm{i})}\Lambda^{(\mathrm{i})\prime}$, which is psd. Any identification map that rationalizes the inflated reduced-form covariance $\Sigma_{\mathrm{uu}}^{(\mathrm{i})}$ must reflect that inflation in the associated impact matrix, although the allocation across individual columns depends on the admissible normalization or rotation.[12]

Excess Propagation. The second distortion arises since, as established earlier, omitting the copula control biases slope coefficients (recalling equation (16)). This perturbs companion matrix $\mathcal{A}^{(\mathrm{i})}$ and moving-average coefficients $\Psi^{(\mathrm{i})}(h)$.

A Unified Decomposition. Given these two types of errors, we can now present a unified decomposition for dynamic analyses, nesting their separate channels. Let $\mathrm{IRF}_{\bullet}^{(\mathrm{i})}(h,j)$ be obtained from reduced-form inputs via the identification map $\mathbb{M}$ satisfying $\Sigma_{\bullet}$, with $\mathrm{P}_{\bullet}\mathrm{P}_{\bullet}' = \Sigma_{\bullet}$. Accordingly, we then consider the difference between the impulse response paths implied by the maintained and the true (copula corrected) models that allows us to separate out these two distortions:

$$\begin{aligned}\Delta\mathrm{IRF}^{(\mathrm{i})}(h,j) &\equiv \mathrm{IRF}_{\mathrm{X}}^{(\mathrm{i})}(h,j) - \mathrm{IRF}_{\mathrm{N}}^{(\mathrm{i})}(h,j) \\ &= \Bigg(\underbrace{\Big[\Psi_{\mathrm{X}}^{(\mathrm{i})}(h) - \Psi_{\mathrm{N}}^{(\mathrm{i})}(h)\Big]\mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})}}_{\text{Excess Propagation}} + \underbrace{\Psi_{\mathrm{X}}^{(\mathrm{i})}(h)\Big(\mathrm{P}_{\mathrm{uu}}^{(\mathrm{i})} - \mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})}\Big)}_{\text{Excess Sensitivity}}\Bigg) e_j\end{aligned} \tag{24}$$

where e_j denotes the canonical selector for the $\mathrm{j_{th}}$ identified column under the maintained identification scheme. The proof follows immediately by adding and subtracting $\Psi_{\mathrm{X}}^{(\mathrm{i})}(h)\,\mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})} e_j$.

The Excess Propagation error. Let $\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}$ and $\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}$ denote the regime-i companion matrices implied by the maintained and true models, and recall $\Psi_{\bullet}^{(\mathrm{i})}(h) = J\big(\mathcal{A}_{\bullet}^{(\mathrm{i})}\big)^h J'$. Define the companion-matrix perturbation $\Delta\mathcal{A}^{(\mathrm{i})} = \mathcal{A}_{\mathrm{X}}^{(\mathrm{i})} - \mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}$. Then the identity for matrix powers yields

$$\big(\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}\big)^h - \big(\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\big)^h = \sum_{s=0}^{h-1}\big(\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}\big)^s\,\Delta\mathcal{A}^{(\mathrm{i})}\,\big(\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\big)^{h-1-s}. \tag{25}$$

For fixed horizon h, or uniformly over bounded h, substituting $\big(\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}\big)^s = \big(\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})} + \Delta\mathcal{A}^{(\mathrm{i})}\big)^s = \big(\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\big)^s + O(\|\Delta\mathcal{A}^{(\mathrm{i})}\|)$ into the canceling sum and retaining only first-order

[12] For any orthogonal matrix $\mathcal{W}$, $\mathrm{P}_{\bullet}^{(\mathrm{i})}\mathcal{W}$ also identifies $\Sigma_{\bullet}^{(\mathrm{i})}$ and $\|\mathrm{P}_{\bullet}^{(\mathrm{i})}\mathcal{W}\|_F = \|\mathrm{P}_{\bullet}^{(\mathrm{i})}\|_F$. Hence the total impact-variance result is invariant to rotations within the class of decompositions satisfying $\mathrm{P}_{\bullet}^{(\mathrm{i})}\mathrm{P}_{\bullet}^{(\mathrm{i})\prime} = \Sigma_{\bullet}^{(\mathrm{i})}$.

terms yields

$$\Psi_{\mathrm{X}}^{(\mathrm{i})}(h) - \Psi_{\mathrm{N}}^{(\mathrm{i})}(h) = \sum_{s=0}^{h-1} J\big(\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\big)^{s}\, \Delta\mathcal{A}^{(\mathrm{i})}\, \big(\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\big)^{h-1-s} J' \;+\; O(\|\Delta\mathcal{A}^{(\mathrm{i})}\|^2), \tag{26}$$

where $\|\cdot\|$ is a matrix norm.

Expression (26) shows that the propagation component is first order in the companion-matrix perturbation $\Delta\mathcal{A}^{(\mathrm{i})}$. Since the omitted copula control induces the slope perturbation $\Delta\mathrm{A}^{(\mathrm{i})} = \Lambda^{(\mathrm{i})}\gamma^{(\mathrm{i})}$, the propagation channel is typically first order in the loading $\Lambda^{(\mathrm{i})}$, provided $\gamma^{(\mathrm{i})} \neq 0$. The sensitivity component has a different local order. The covariance distortion is exact:

$$\Delta\Sigma_{\mathrm{uu}}^{(\mathrm{i})} \;\equiv\; \Sigma_{\mathrm{uu}}^{(\mathrm{i})} - \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} \;=\; \Lambda^{(\mathrm{i})}\Lambda^{(\mathrm{i})\prime}. \tag{27}$$

If the impact matrix varies smoothly with the reduced-form covariance matrix, its local change can be written as

$$P_{\mathrm{uu}}^{(\mathrm{i})} - (P_{\varepsilon\varepsilon}^{(\mathrm{i})}) \;=\; \mathcal{L}_{P}^{(\mathrm{i})}\left[\Delta\Sigma_{\mathrm{uu}}^{(\mathrm{i})}\right] + o\Big(\|\Delta\Sigma_{\mathrm{uu}}^{(\mathrm{i})}\|\Big), \tag{28}$$

where $\mathcal{L}_{P}^{(\mathrm{i})}[\cdot]$ denotes the linear approximation to the impact matrix induced by the identification scheme. Therefore, Excess Sensitivity is first order in the covariance perturbation $\Delta\Sigma_{\mathrm{uu}}^{(\mathrm{i})}$, but locally second order in the primitive loading $\Lambda^{(\mathrm{i})}$. By contrast, excess propagation is first order in $\Delta\mathcal{A}^{(\mathrm{i})}$ and therefore typically first order in $\Lambda^{(\mathrm{i})}$. At impact, $h = 0$, the propagation term is absent because $\Psi_{\mathrm{X}}^{(\mathrm{i})}(0) = \Psi_{\mathrm{N}}^{(\mathrm{i})}(0)$. Hence impact differences are driven only by Excess Sensitivity. For horizons $h \geq 1$, both sensitivity and propagation channels may contribute.

3.2 *External Instruments*

We next examine identification with external instruments. In the proxy-SVAR framework, the instrument isolates a structural shock through its covariance with the reduced-form innovation. In the presence of threshold endogeneity, however, the relevant innovation depends on the estimand. If the target is the structural shock net of contemporaneous threshold selection, the proxy step should be conducted using the copula-corrected innovations of the $\mathrm{TVAR}_{\mathrm{N}}$. These innovations remove the threshold-selection component and are orthogonal to the threshold score.

This distinction is one of estimand. Proxy moments based on the corrected innovation identify the impact vector of the orthogonalized structural shock in $\varepsilon_t^{(\mathrm{i})}$. By contrast, proxy moments based on the uncorrected innovation $u_t^{(\mathrm{i})}$ identify a composite direction whenever the instrument is correlated with the threshold score. The following proposition formalizes this distinction.

Proposition 2 (Proxy–SVAR impact identification with and without orthogonality to the control). *Let the reduced-form innovation decompose as*

$$u_t^{(\mathrm{i})} \;=\; \Lambda^{(\mathrm{i})} z_t^{(i)\star} + \varepsilon_t^{(\mathrm{i})}, \qquad \mathbb{E}(\varepsilon_t^{(\mathrm{i})} \mid z_t^{(i)\star}) = 0_{\mathrm{K}}, \tag{29}$$

and let the structural mapping be $\varepsilon_t^{(\mathrm{i})} = \mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})} \cdot e_t^{(\mathrm{i})}$, *where* $\mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})} = (p_1^{(\mathrm{i})} \cdots p_{\mathrm{K}}^{(\mathrm{i})})$ *and* $\mathbb{E}(e_t^{(\mathrm{i})} e_t^{(\mathrm{i})\prime}) = \mathrm{I}_{\mathrm{K}}$. *Let* $v_t^{(\mathrm{i})}$ *be a regime-specific external instrument that is relevant and exogenous for the target shock:*

$$\lambda_v^{(\mathrm{i})} \;\equiv\; \mathbb{C}\mathrm{ov}(e_{1t}^{(\mathrm{i})}, v_t^{(\mathrm{i})}) \neq 0, \qquad \mathbb{C}\mathrm{ov}(e_{jt}^{(\mathrm{i})}, v_t^{(\mathrm{i})}) = 0 \;\; (j \neq 1). \tag{30}$$

Note that conditions (30) are silent on whether there is correlation between the threshold variable and the external instrument. To that end, define the threshold-instrument covariance as

$$\omega^{(\mathrm{i})} \;\equiv\; \mathbb{C}\mathrm{ov}(z_t^{(i)\star}, v_t^{(\mathrm{i})}).$$

In our context, accordingly, we have:

i) $\mathbf{TVAR_N}$. *The proxy moment based on the controlled innovations is*

$$\begin{aligned} \Sigma_{\varepsilon v}^{(\mathrm{i})} \;&\equiv\; \mathbb{C}\mathrm{ov}(\varepsilon_t^{(\mathrm{i})}, v_t^{(\mathrm{i})}) \;=\; \mathbb{C}\mathrm{ov}(\mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})} e_t^{(\mathrm{i})}, v_t^{(\mathrm{i})}) \\ &=\; \mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})}\, \mathbb{C}\mathrm{ov}(e_t^{(\mathrm{i})}, v_t^{(\mathrm{i})}) \;=\; \lambda_v^{(\mathrm{i})}\, p_1^{(\mathrm{i})} \end{aligned} \tag{31}$$

Thus the proxy moment based on the corrected innovations identifies the impact column $p_1^{(\mathrm{i})}$ *of the orthogonalized structural shock, up to normalization, regardless of* $\omega^{(\mathrm{i})}$.

ii) $\mathbf{TVAR_X}$. *The proxy moment based on* $u_t^{(\mathrm{i})}$ *is*

$$\begin{aligned} \Sigma_{uv}^{(\mathrm{i})} \;=\; \mathbb{C}\mathrm{ov}\big(u_t^{(\mathrm{i})},\, v_t^{(\mathrm{i})}\big) \;&=\; \mathbb{C}\mathrm{ov}\big(\Lambda^{(\mathrm{i})} z_t^{(i)\star} + \varepsilon_t^{(\mathrm{i})},\, v_t^{(\mathrm{i})}\big) \\ &=\; \lambda_v^{(\mathrm{i})}\, p_1^{(\mathrm{i})} + \Lambda^{(\mathrm{i})}\, \omega^{(\mathrm{i})} \end{aligned} \tag{32}$$

Equivalently, when $\lambda_v^{(\mathrm{i})} \neq 0$, *the uncorrected proxy moment is proportional to the (comp)osite direction*

$$p_{\mathrm{comp}}^{(\mathrm{i})} \;=\; p_1^{(\mathrm{i})} + \frac{\omega^{(\mathrm{i})}}{\lambda_v^{(\mathrm{i})}} \Lambda^{(\mathrm{i})}. \tag{33}$$

If the target is the orthogonalized shock $p_1^{(\mathrm{i})}$, the second term represents threshold-selection contamination. If the target is instead the full event that also moves the threshold variable contemporaneously, then $p_{\mathrm{comp}}^{(\mathrm{i})}$ is the relevant estimand.[13]

PROOF. Appendix A.4. □

The Proposition clarifies the estimand delivered by proxy-SVAR identification under threshold endogeneity. When the proxy step is applied to the corrected innovations $\varepsilon_t^{(\mathrm{i})}$, the moment $\mathbb{C}\mathrm{ov}(\varepsilon_t^{(\mathrm{i})}, v_t^{(\mathrm{i})})$ identifies the impact vector of the orthogonalized structural shock, net of contemporaneous threshold selection. This result holds even when the instrument is correlated with the threshold score, because that selection component has been removed from $\varepsilon_t^{(\mathrm{i})}$.

By contrast, applying the proxy step to the uncorrected innovations $u_t^{(\mathrm{i})}$ mixes structural-shock variation with threshold-selection variation whenever $\omega^{(\mathrm{i})} \neq 0$. If the target is the orthogonalized shock $p_1^{(\mathrm{i})}$, the additional component $(\omega^{(\mathrm{i})}/\lambda_v^{(\mathrm{i})})\Lambda^{(\mathrm{i})}$ represents threshold-selection contamination. If the target is instead the full threshold-moving event, the same moment identifies the composite direction $p_{\mathrm{comp}}^{(\mathrm{i})}$. Thus, the correction does not rule out studying full event responses. It makes explicit whether the researcher is identifying an orthogonalized shock or a composite threshold-moving event.

When $\omega^{(\mathrm{i})} = 0$, the threshold score is orthogonal to the instrument, and the corrected and uncorrected proxy moments identify the same impact direction. Any $\mathrm{TVAR_X} - \mathrm{TVAR_N}$ difference in the corresponding IRFs then reflects the dynamic propagation channel. When $\omega^{(\mathrm{i})} \neq 0$, the uncorrected proxy moment also contains $\omega^{(\mathrm{i})}\Lambda^{(\mathrm{i})}$, generating an additional impact discrepancy at $h = 0$.

3.3 *The Threshold Parameter*

Finally, we turn to the threshold parameter itself. Because the threshold determines regime allocation, misspecification in the regime-specific VARs feeds directly into its estimation by distorting the population least-squares objective that governs threshold selection. The identification of the threshold is central in threshold models because it

[13] A closed-form expression for the angle between the uncorrected proxy moment $\Sigma_{uv}^{(i)}$ and the target impact column $p_1^{(i)}$, together with a sharp first-order bound in the ratio $\omega^{(i)}/\lambda_v^{(i)}$ that quantifies the magnitude of the threshold-selection contamination, is provided in Corollary A.1, Appendix A.

materially shapes policy or behavioral conclusions.[14] Both our Monte Carlo experiments and empirical applications suggest that, when threshold endogeneity is ignored, the resulting threshold distortion can be systematic.

The population least-squares criterion in discontinuous threshold models is typically locally one-sided rather than smoothly quadratic. Local movements of the threshold reallocate observations across regimes, producing a switcher-driven objective; this feature underlies the nonstandard threshold asymptotics in Chan (1993), Tsay (1998) and Hansen (2000), and is closely related to change-point least-squares arguments in Bai (1997). Accordingly, we characterize the effect of omitting the copula control as a local tilt in the one-sided slopes of the population threshold objective, rather than as a smooth quadratic displacement of the minimizer.

The omission cost entering the concentrated least-squares criterion is the residualized cost of the omitted control. Once the included VAR regressors are re-estimated, only the component of the omitted copula control that remains after projection on the included regressors contributes to the concentrated residual sum of squares, as in Frisch–Waugh–Lovell partialling-out arguments. When the dependence on the candidate threshold matters, the objects $z_t^{(\mathrm{i})\star}$, $\gamma^{(\mathrm{i})}$, and $\mathrm{Z}^{(\mathrm{i})}$ should be read as functions of δ, although we suppress this dependence when no ambiguity arises.

Given the definitions of $\gamma^{(\mathrm{i})}$ and $Q_{\mathrm{YY}}^{(\mathrm{i})}$ in (16), define the residualized threshold score as

$$r_t^{(\mathrm{i})} = z_t^{(\mathrm{i})\star} - \gamma^{(\mathrm{i})}\mathrm{Y}_{t-1}. \tag{34}$$

The residualized "omission cost" is

$$D_R(\delta) = p_\delta\,\mathbb{E}\left[\left\|\Lambda^{(1)} r_t^{(1)}\right\|^2 \mid \mathrm{Z}^{(1)}\right] + (1-p_\delta)\,\mathbb{E}\left[\left\|\Lambda^{(2)} r_t^{(2)}\right\|^2 \mid \mathrm{Z}^{(2)}\right], \tag{35}$$

where $p_\delta = \Pr(z_t \leq \delta) = \mathrm{F}_z(\delta)$, as defined in (2). The Proposition below shows how this residualized omission cost changes the local one-sided slopes of the population threshold objective.

Proposition 3 (Local Threshold Bias under a Kinked Objective). *Consider the population concentrated objectives $Q_\bullet(\delta)$, $\bullet \in \{\mathrm{N}, \mathrm{X}\}$, where $Q_{\mathrm{N}}(\delta)$ corresponds to the correctly specified $\mathrm{TVAR}_{\mathrm{N}}$ and $Q_{\mathrm{X}}(\delta)$ to the misspecified $\mathrm{TVAR}_{\mathrm{X}}$. Let δ_0 denote the unique minimizer*

[14] As emphasized by Hansen (2000) and Tsay (1998), threshold estimates are subject to substantial finite-sample uncertainty, leading to regime misclassification. In applied macroeconomic settings, such misclassification has been shown to materially distort estimated policy responses, particularly near regime boundaries; see, for example, Lo and Piger (2005) and Auerbach and Gorodnichenko (2012).

of $Q_{\mathrm{N}}(\delta)$*. Assume that, in a neighborhood of* δ_0*, the correctly specified objective admits the one-sided expansion*

$$Q_{\mathrm{N}}(\delta) - Q_{\mathrm{N}}(\delta_0) \;=\; h^{+}(\delta-\delta_0)_{+} + h^{-}(\delta_0-\delta)_{+} + o(|\delta-\delta_0|), \tag{36}$$

where $h^{+} > 0$*,* $h^{-} > 0$*,*[15] *and* $x_{+} = \max\{x, 0\}$*. Suppose further that the misspecification component satisfies*

$$\big[Q_{\mathrm{X}}(\delta) - Q_{\mathrm{N}}(\delta)\big] - \big[Q_{\mathrm{X}}(\delta_0) - Q_{\mathrm{N}}(\delta_0)\big] \;=\; D_R(\delta) - D_R(\delta_0) + o(|\delta-\delta_0|) \tag{37}$$

locally around δ_0*, and that* $D_R(\delta)$ *is differentiable at* δ_0*, with derivative* $D'_R(\delta_0) = \frac{d}{d\delta}D_R(\delta)\Big|_{\delta=\delta_0}$*. Then, for* $\delta > \delta_0$*,*

$$Q_{\mathrm{X}}(\delta) - Q_{\mathrm{X}}(\delta_0) \;=\; \big[h^{+} + D'_R(\delta_0)\big](\delta-\delta_0) + o(|\delta-\delta_0|), \tag{38}$$

whereas for $\delta < \delta_0$*,*

$$Q_{\mathrm{X}}(\delta) - Q_{\mathrm{X}}(\delta_0) \;=\; \big[h^{-} - D'_R(\delta_0)\big](\delta_0-\delta) + o(|\delta-\delta_0|), \tag{39}$$

where $D_R(\delta) \equiv p_\delta C_1(\delta) + (1-p_\delta)C_2(\delta)$*. Thus, omitted threshold endogeneity changes the one-sided slopes of the population threshold objective. If*

$$h^{+} + D'_R(\delta_0) > 0 \quad \textit{and} \quad h^{-} - D'_R(\delta_0) > 0, \tag{40}$$

then δ_0 *remains a local minimizer of the misspecified objective. If*

$$h^{+} + D'_R(\delta_0) < 0, \tag{41}$$

the misspecified objective is locally tilted to the right; and to the left if

$$h^{-} - D'_R(\delta_0) < 0. \tag{42}$$

PROOF. See Appendix A.5. □

Proposition 3 should be interpreted as a local one-sided-slope result. In a discontinuous threshold model, the population criterion is generally kinked at the true threshold, because moving the threshold changes which observations are assigned to each regime. Omitting the copula control does not necessarily generate a smooth quadratic displacement of the minimizer.[16] Instead, omission of the copula control changes the left and right slopes of the population criterion through $D'_R(\delta_0)$. To interpret this deriva-

[15] Where h^{+} is the right-hand (one-sided) slope of the population concentrated CLS criterion Q_N at δ_0, i.e. the rate at which the objective increases as δ moves above δ_0: $h^{+} = \lim_{\delta\downarrow\delta_0}\frac{Q_N(\delta)-Q_N(\delta_0)}{\delta-\delta_0}$. Likewise for h^{-}. Both are strictly positive because δ_0 is the unique minimizer of Q_N.

[16] A smooth heuristic can be obtained by replacing the kinked objective with $Q_{\mathrm{N}}(\delta) - Q_{\mathrm{N}}(\delta_0) \simeq \frac{1}{2}H_0(\delta-\delta_0)^2$. Together with $D_R(\delta) - D_R(\delta_0) \simeq D'_R(\delta_0)(\delta-\delta_0)$, this gives $\delta_{\mathrm{X}} - \delta_0 \simeq -D'_R(\delta_0)/H_0$. We do not use this

tive, write

$$D_R(\delta) \equiv p_\delta C_1(\delta) + (1 - p_\delta) C_2(\delta), \tag{43}$$

where $C_i(\delta) \equiv \mathbb{E}\left[\left\|\Lambda^{(\mathrm{i})} r_t^{(\mathrm{i})}\right\|^2 \mid \mathrm{Z}^{(\mathrm{i})}\right]$. Since $r_t^{(\mathrm{i})}$ is scalar, we have $C_i(\delta) = \|\Lambda^{(\mathrm{i})}\|^2 \mathbb{E}\left[(r_t^{(\mathrm{i})})^2 \mid \mathrm{Z}^{(\mathrm{i})}\right]$. Assuming that F_z is differentiable at δ_0, with density $f_z(\delta_0) > 0$, we have

$$D_R'(\delta_0) = f_z(\delta_0) \left[C_1(\delta_0) - C_2(\delta_0)\right] + p_{\delta_0} C_1'(\delta_0) + (1 - p_{\delta_0}) C_2'(\delta_0). \tag{44}$$

The first term is a composition effect: moving the threshold reallocates marginal observations across regimes at rate $f_z(\delta_0)$, and the effect depends on the difference between the residualized omission costs in the two regimes. The second and third terms are within-regime omission-cost effects, weighted by the regime probabilities p_{δ_0} and $1 - p_{\delta_0}$.

This decomposition clarifies the direction of the local shift. If the composition effect dominates and $C_1(\delta_0) < C_2(\delta_0)$, then $D_R'(\delta_0) < 0$, so moving the threshold to the right lowers the residualized omission cost. In that case, the right-hand slope $h^+ + D_R'(\delta_0)$ is reduced and, if it becomes negative, the misspecified objective is locally shifted toward larger threshold values. Conversely, if the composition effect dominates and $C_1(\delta_0) > C_2(\delta_0)$, then $D_R'(\delta_0) > 0$, reducing the left-hand slope $h^- - D_R'(\delta_0)$ and potentially shifting the objective toward smaller threshold values.

Hence, the pseudo-true threshold under TVAR_X adjusts by reallocating marginal observations toward the side that reduces the residualized omission cost. This result is local. For fixed and large omitted-control effects, the pseudo-true threshold

$$\delta_\mathrm{X} = \arg\min_\delta Q_\mathrm{X}(\delta), \tag{45}$$

remains well defined, but its displacement from δ_0 depends on the global shape of the misspecified objective rather than only on the local one-sided slopes. The Monte Carlo experiments illustrate this nonlocal behavior under finite misspecification.

as the formal result because discontinuous threshold objectives are generally locally one-sided rather than quadratic.

4. Estimation and Implementation

Let $\mathcal{Z}_T = \{z_t : t = 1, \ldots, T\}$ denote the sample support of the threshold variable. Following the standard practice in threshold models, the threshold parameter is searched over a trimmed subset of this support,

$$l_\delta \equiv \{\delta \in \mathcal{Z}_T : \tau \leq \widehat{F}_z(\delta) \leq 1 - \tau\}, \tag{46}$$

where $\tau \in (0, 1/2)$ is a trimming parameter, say $\tau = 0.1$.

If the threshold value δ were known, estimation of the augmented TVAR model would proceed by least squares applied separately within each regime. When δ is unknown, it is estimated by concentrated least squares (CLS), following the standard threshold-estimation approach of Chan (1993). For each candidate threshold $\delta \in l_\delta$, the regime partition and the corresponding copula controls are recomputed, and the parameters are estimated by least squares within each regime.

For each $\delta \in l_\delta$, compute the residual sum of squares

$$RSS(\delta) = \sum_{i=1}^{2} \sum_{t \in \mathcal{T}_i(\delta)} \left\| \widehat{\varepsilon}_t^{(i)}(\delta) \right\|_F^2, \tag{47}$$

where $\mathcal{T}_1(\delta) = \{t : z_t \leq \delta\}$, $\mathcal{T}_2(\delta) = \{t : z_t > \delta\}$, and

$$\widehat{\varepsilon}_t^{(i)}(\delta) = y_t - \widehat{A}^{(i)} Y_{t-1} - \widehat{\Lambda}^{(i)} \widehat{z}_t^{(i)\star}. \tag{48}$$

The threshold estimator is then defined as

$$\widehat{\delta} = \arg\min_{\delta \in l_\delta} RSS(\delta). \tag{49}$$

Given $\widehat{\delta}$, least squares estimation of the remaining coefficients, $\Theta = \left(A^{(1)}, A^{(2)}, \Lambda^{(1)}, \Lambda^{(2)}\right)$, and of the regime-specific covariance matrices $\Sigma_{\varepsilon\varepsilon}^{(i)}$ follows in a second step. In the standard threshold VAR setting, the asymptotic results of Chan (1993), Tsay (1998), and Samia and Chan (2011) imply strong consistency of the threshold estimator and $\sqrt{T}$-consistency of the remaining slope and covariance estimators under stationarity and distinct threshold effects. In the present model, the copula correction introduces generated and threshold-dependent controls. The assumptions and lemmas below show that these additional components are asymptotically negligible at the global and local scales relevant for consistency and the standard threshold limit theory.

Inference for Θ can be based on its limiting distribution via the *delta method* or, more robustly, via a moving-block bootstrap that accounts for serial dependence and the generated-regressor structure. Our Monte Carlo exercises in Section 6 demonstrate that the estimator performs well, even in small samples. [17]

The following assumption collects regularity conditions ensuring that the estimated copula controls are asymptotically close to their infeasible population counterparts over the trimmed threshold grid. It also imposes a local counting regularity condition near the true threshold, which is needed to control local objective increments on the $1/\mathrm{T}$ neighborhoods relevant for the threshold asymptotics.

The regime-specific truncated CDFs $\mathrm{F}_{z_t}^{(\mathrm{i})}$ and their estimators $\widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})}$ are deterministic transforms of the marginal CDF indexed by the candidate threshold δ; we suppress this dependence when no ambiguity arises.

Assumption 2 (Plug-in Regularity for the Copula Control). *The following conditions hold.*

1. ***Trimmed grid.*** *Search over* $I_\delta = [\underline{\delta}, \overline{\delta}]$ *with* $\Pr(z_t \le \underline{\delta}) \ge \tau$ *and* $\Pr(z_t \le \overline{\delta}) \le 1-\tau$ *for some* $\tau \in (0, 1/2)$.

2. ***Smooth marginal.*** F_z *is continuous and strictly increasing with density* f_z *bounded away from zero on an open neighborhood of* I_δ.

3. ***Uniform CDF consistency under weak dependence.*** *Let* $\widehat{\mathrm{F}}_z$ *denote the empirical CDF, or a kernel CDF estimator, of* z_t. *Then*
$$\zeta_\mathrm{T} \equiv \sup_{z\in\mathbb{R}} \left| \widehat{\mathrm{F}}_z - \mathrm{F}_z \right| = o_p(1). \tag{50}$$
A sufficient primitive condition is that $\{z_t\}$ *is strictly stationary and weakly dependent, for example strongly mixing with summable mixing coefficients, so that a uniform Glivenko–Cantelli law holds for the ECDF. For kernel CDF estimators, the usual bandwidth conditions are assumed.*

4. ***Local counting regularity near the threshold.*** *Let* δ_0 *denote the true threshold. For every fixed* $C < \infty$,
$$\sup_{|x|\le C} \sum_{t=1}^{\mathrm{T}} \mathbf{1}\left\{ \min\left(\delta_0, \delta_0 + \frac{x}{\mathrm{T}}\right) < z_t \le \max\left(\delta_0, \delta_0 + \frac{x}{\mathrm{T}}\right) \right\} \;=\; O_p(1). \tag{51}$$

5. ***Winsorized PIT transform.*** *Let* $\epsilon_\mathrm{T} \downarrow 0$ *and define the winsorization map*
$$\kappa_\mathrm{T}(u) \;=\; \min\{1-\epsilon_\mathrm{T}, \max\{\epsilon_\mathrm{T}, u\}\}. \tag{52}$$

[17] Appendix F discusses computational costs and scaling aspects of the $\mathrm{TVAR_N}$ framework.

In the feasible CLS criterion, the regime-specific PIT controls are computed as

$$\widehat{z}_t^{(\mathrm{i})\star} \equiv \Phi^{-1}\left(\kappa_{\mathrm{T}}\left(\widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})}\right)\right), \tag{53}$$

with the corresponding infeasible controls defined analogously by replacing $\widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})}$ *with* $\mathrm{F}_{z_t}^{(\mathrm{i})}$. *Moreover, if*

$$L_{\mathrm{T}} \equiv \sup_{u \in [\epsilon_{\mathrm{T}}, 1-\epsilon_{\mathrm{T}}]} \left| \frac{d}{du}\Phi^{-1}(u) \right|, \tag{54}$$

then $L_{\mathrm{T}}\zeta_{\mathrm{T}} = o_p(1)$ *and* $\frac{L_{\mathrm{T}}^2}{\mathrm{T}} \to 0$.

This observation-level winsorization ensures that the arguments of $\Phi^{-1}(\bullet)$ *remain in* $[\epsilon_{\mathrm{T}}, 1-\epsilon_{\mathrm{T}}]$, *so that the estimated and infeasible PIT controls are uniformly close over the trimmed threshold grid. The additional rate condition* $L_{\mathrm{T}}^2/\mathrm{T} \to 0$ *ensures that the local change in the oracle controls induced by moving the threshold within* $1/\mathrm{T}$ *neighborhoods is negligible for the RSS-scale local objective increments.*

6. ***Stationarity and moments.*** *The stationarity and moment conditions in Proposition 4 Condition (i) hold.*

Assumption 2 ensures that the generated copula controls are asymptotically well behaved over the trimmed threshold grid. Condition 3 gives uniform consistency of the marginal CDF estimator, while Condition 5 applies an observation-level winsorization before the Gaussian quantile transform. This is important because trimming the threshold grid bounds the regime probabilities p_δ and $1-p_\delta$, but it does not by itself bound the individual regime-specific PIT values away from zero and one. The winsorized transform therefore keeps the arguments of $\Phi^{-1}(\bullet)$ in a region where the quantile transform is Lipschitz. The restriction $L_{\mathrm{T}}\zeta_{\mathrm{T}} = o_p(1)$ makes the feasible and infeasible PIT controls uniformly close, while $L_{\mathrm{T}}^2/\mathrm{T} \to 0$ ensures that threshold-induced local variation of the oracle controls is negligible at the RSS scale relevant for the $1/\mathrm{T}$ threshold limit.

Appendix A formalizes these approximation steps. In particular, subsection A.3 establishes feasible–oracle equivalence for the CLS criterion. Part (a) gives the global approximation over the trimmed grid l_δ, which is sufficient for consistency of the threshold estimator in Proposition 4. Part (b) gives the corresponding approximation for local objective increments in $1/\mathrm{T}$ neighborhoods of the true threshold δ_0.

For the local asymptotic analysis, one additional issue must be controlled: the regime-specific PIT controls are themselves indexed by the candidate threshold. Hence, changing δ changes not only regime membership but also the values of the oracle controls within each regime. Condition 4 controls the number of observations that switch regimes in $1/\mathrm{T}$ neighborhoods of δ_0, while subsection A.4 shows that the additional threshold-dependence of the oracle controls is asymptotically negligible for the local increments of the CLS objective. Combined, A.3 and A.4 imply that the feasible local objective has the same switcher-driven limit as the standard threshold objective. Consequently, the Chan-type nonstandard limit for $\mathrm{T}(\widehat{\delta}-\delta_0)$ continues to apply, as stated in **Remark 1**.

Proposition 4 (Consistency of the Threshold Estimator). *Let δ_0 denote the true threshold value and let*

$$\widehat{\delta} = \arg\min_{\delta \in l_\delta} \widehat{Q}_{\mathrm{T}}(\delta) \tag{55}$$

be the minimizer of the feasible winsorized CLS objective over the compact, trimmed grid l_δ. Suppose:

(i) *The piecewise VAR is regime-wise stable, so that $\{y_t\}$ is strictly stationary and ergodic with $\mathbb{E}\|y_t\|^{2+\varsigma} < \infty$ for some $\varsigma > 0$.*

(ii) *The threshold variable z_t has a continuous, strictly increasing distribution with positive density in a neighborhood of δ_0, and δ_0 lies in the interior of l_δ.*

(iii) *The threshold is identified: the population infeasible objective $Q_{\mathrm{N}}(\delta)$ has a unique minimum at δ_0. A sufficient condition is that the regime-specific conditional mean is discontinuous at δ_0, for example through distinct regime coefficients $\mathrm{A}^{(1)} \neq \mathrm{A}^{(2)}$ or distinct control loadings $\Lambda^{(1)} \neq \Lambda^{(2)}$.*

(iv) *Assumption 1 and Proposition 1 hold, so that the corrected innovations satisfy*

$$\mathbb{E}\left(\varepsilon_t^{(\mathrm{i})} \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}\right) = 0_{\mathrm{K}}. \tag{56}$$

(v) *Assumption 2 holds.*

Then

$$\widehat{\delta} \xrightarrow{p} \delta_0, \qquad \text{as } \mathrm{T} \to \infty. \tag{57}$$

PROOF. Appendix A.10. □

Proposition 4 establishes consistency of the threshold estimator in the presence of the copula control. The proof relies only on Appendix A.3 part (a), which shows that the feasible and infeasible CLS criteria are uniformly $o_p(1)$-close over the trimmed search grid l_δ. This global approximation is sufficient for consistency in probability. Under stronger almost-sure versions of the uniform convergence conditions in Assumption 2, the convergence can be strengthened to almost sure convergence.

The following remark concerns the local asymptotic distribution of the threshold estimator. This requires stronger local approximations than consistency. First, Appendix A.3 part (b) ensures that the feasible and infeasible objective increments are asymptotically equivalent on $1/\mathrm{T}$ neighborhoods of the true threshold δ_0. Second, Appendix A.4 shows that the additional dependence of the oracle PIT controls on the candidate threshold is locally negligible. Together, these two local results imply that the feasible local objective has the same switcher-driven limit as in the standard threshold theory of Chan (1993).

Remark 1 (Nonstandard Limit with Plug-in $\mathrm{F}_{z_t}^{(\mathrm{i})}$). *Under the conditions of Proposition 4, Appendix A.3 part (b), and Appendix A.4, the feasible local CLS objective has the same limit as the oracle local objective in the standard discontinuous-threshold problem. Hence the threshold estimator retains the Chan-type nonstandard limit,*

$$\mathrm{T}(\widehat{\delta}-\delta_0) \Rightarrow \arg\min_{x\in\mathbb{R}} \Upsilon(x), \tag{58}$$

where $\Upsilon(x)$ is the compound-Poisson limit process generated by the local switchers around δ_0. Accordingly, inference for δ should be based on a profile-objective likelihood-ratio set,

$$\{\delta : RSS(\delta) - RSS(\widehat{\delta}) \leq c_\alpha\}, \tag{59}$$

with c_α obtained from a moving-block bootstrap that recomputes $\widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})}$, the winsorized copula controls, and $\widehat{\delta}$ in each draw.

PROOF. Appendix A.11. □

Summing up, the copula correction preserves the large-sample behavior of the threshold estimator relative to the standard TVAR framework. Appendix A.3 and Appendix A.4 show that the generated, winsorized PIT controls and the threshold-dependent oracle controls are asymptotically negligible for consistency and for the Chan-type local threshold limit. Consequently, the standard threshold asymptotic theory can be applied after the copula-control step.

5. Extensions Beyond the Gaussian Benchmark

The formal results in the previous sections are developed for the regime-specific Gaussian-copula benchmark. Under this benchmark, the conditional mean $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$

is linear in the Gaussian PIT score $z_t^{(\mathrm{i})\star}$, and the resulting control can be included as an additional regressor in the TVAR. This yields the closed-form representation used for estimation, misspecification analysis, and threshold asymptotics.

The purpose of this section is more modest. We discuss how the same control-function logic can be extended or approximated when the dependence between the threshold variable and the reduced-form innovations departs from the Gaussian benchmark. These extensions should not be interpreted as part of the exact least-squares theory established above. Rather, they clarify how the baseline method relates to dependence structures in which the conditional mean correction may be nonlinear. These extensions are useful in applied work because they provide a parsimonious way to detect and absorb residual nonlinear threshold-selection effects that may remain after the baseline Gaussian PIT control.[18]

For non-elliptical copulas, the conditional mean $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$ is generally not linear in the Gaussian PIT score. In such cases, the single linear control $z_t^{(\mathrm{i})\star}$ may capture only part of the threshold-selection component. A flexible way to extend the correction is to approximate the unknown conditional mean using a sieve expansion in the Gaussian PIT score:

$$\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) \approx \sum_{j=1}^{d_{\mathrm{T}}} \beta_j^{(\mathrm{i})} s_j(z_t^{(\mathrm{i})\star}), \tag{60}$$

where $\{s_j(\cdot)\}_{j=1}^{d_{\mathrm{T}}}$ is a centered and scaled basis. Substituting this approximation into the regime-specific TVAR gives

$$y_t = A^{(\mathrm{i})} Y_{t-1} + \sum_{j=1}^{d_{\mathrm{T}}} \beta_j^{(\mathrm{i})} s_j(z_t^{(\mathrm{i})\star}) + \widetilde{\varepsilon}_t^{(\mathrm{i})}. \tag{61}$$

Here $\widetilde{\varepsilon}_t^{(\mathrm{i})}$ contains the mean-zero innovation $u_t^{(\mathrm{i})} - \mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$ and the approximation error from the finite sieve. See Appendix B for the derivation.

A convenient basis is given by Hermite polynomials evaluated at $z_t^{(\mathrm{i})\star}$. Under the Gaussian-copula benchmark, only the linear term is needed in population. Departures from the Gaussian benchmark may generate nonlinear components in the conditional

[18] Heavy-tailed elliptical specifications, such as a Student-t copula, provide another possible departure from the Gaussian benchmark. Under elliptical dependence, a Student-t PIT score can preserve a linear conditional-mean control, although the associated conditional dispersion may depend on the threshold score. Thus, Student-t departures can also be treated within the broader control-function logic considered here, either by using the corresponding PIT score or by approximating residual departures from the Gaussian PIT control with the sieve terms discussed below.

mean, which can be approximated by higher-order terms. Thus, the Hermite extension should be interpreted as a sieve-style robustness device: it asks whether enriching the PIT-score control improves finite-sample performance when the true threshold–innovation dependence is nonlinear. The Monte Carlo experiments in the next section use nonlinear copula designs to evaluate this robustness channel. They do not treat the non-elliptical extension as an exact analogue of Proposition 1; instead, they assess whether additional PIT-score terms help absorb nonlinear threshold-selection effects in finite samples.

6. Monte Carlo Analysis

We now study the finite-sample properties of the exogenous (misspecified) and endogenous (true) threshold VAR specifications using Monte Carlo (MC) simulations. We examine robustness to cases where the dependence between the TVAR innovations and the threshold variable is linear or nonlinear (using Clayton, Joe, and Gumbel copulas), with correlations of varying signs and strengths, and over different sample sizes.

6.1 *The Data Generating Process (DGP)*

Assume the true DGP is the following bivariate threshold VAR:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0.7 & 0.1 \\ 0.1 & 0.7 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t}^{(1)} \\ u_{2t}^{(1)} \end{pmatrix}, & \text{if } z_t \leq \delta \\ \begin{pmatrix} 0.1 & 0.7 \\ 0.7 & 0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t}^{(2)} \\ u_{2t}^{(2)} \end{pmatrix}, & \text{if } z_t > \delta \end{cases} \tag{62}$$

where the *true* value of the threshold parameter δ_0 is set to the third quantile of the assumed distribution of z_t. The initial values of y_{1t} and y_{2t} are set to zero. We consider two marginals for z_t, chosen to have the same variance for valid comparisons:[19]

$$z_t \sim \mathcal{U}(-4,3) \quad \Rightarrow \mathbb{V}\mathrm{ar}(z_t) = 4.08, \;\; \delta_0 = \mathrm{Q}_3 = 1.25,$$
$$z_t \sim \mathcal{N}(0,4.08) \Rightarrow \mathbb{V}\mathrm{ar}(z_t) = 4.08, \;\; \delta_0 = \mathrm{Q}_3 = 1.36.$$

[19] For $z \sim \mathcal{U}(a,b)$, $\mathbb{V}\mathrm{ar}(z) = (b-a)^2/12$.

The covariance structure among the reduced-form innovations is assumed common across regimes. For $u_t^{(i)} = (u_{1t}^{(i)}, u_{2t}^{(i)})'$, let

$$\Sigma_{uu}^{(i)} = \mathcal{B}^{(i)}\mathcal{B}^{(i)\prime}, \qquad \mathcal{B}^{(i)} = \begin{pmatrix} 1 & 0 \\ 0.8 & 1 \end{pmatrix}, \ \ i = 1, 2.$$

Thus, $\mathcal{B}^{(1)} = \mathcal{B}^{(2)}$ is the common lower-triangular Cholesky factor capturing the contemporaneous covariance between the elements of $u_t^{(i)}$. Equivalently, $u_t^{(i)} = \mathcal{B}^{(i)} e_t^{(i)}$, with $e_t^{(i)} \sim N(0, I_2)$.

Let $\operatorname{corr}(u_{jt}^{(i)}, z_t)$ denote the correlation coefficient between $u_{jt}^{(i)}$ and z_t, $j = 1, 2$. In the linear-Gaussian Scenarios 1 and 2, see below, the innovations and the threshold variable are jointly Gaussian,

$$\left(u_t^{(i)\prime}, z_t\right)' \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{uu}^{(i)} & \sigma_{uz}^{(i)} \\ \sigma_{uz}^{(i)\prime} & \sigma_{zz} \end{pmatrix}\right), \tag{63}$$

with the cross-block calibrated from the targets $\operatorname{corr}(u_{jt}^{(1)}, z_t)$ via

$$\left[\sigma_{uz}^{(i)}\right]_j = \operatorname{corr}(u_{jt}^{(i)}, z_t) \sqrt{\sigma_{zz} \left[\Sigma_{uu}^{(i)}\right]_{jj}}, \qquad j = 1, 2. \tag{64}$$

The resulting joint covariance matrix is chosen to be positive semidefinite in each regime. To keep δ_0 invariant across scenarios we draw z_t from its fixed marginal and then $u_t^{(i)} \mid z_t$ from the Gaussian conditional implied by (63)–(64); the regime i is selected ex post by $\mathbf{1}\{z_t > \delta_0\}$.

We consider four scenarios.

- **Scenario 1 (Common Linear Endogeneity)**. $\operatorname{corr}(u_{jt}^{(i)}, z_t) = 0.8$ for $j = 1, 2$.
- **Scenario 2 (Regime-dependent Linear Endogeneity)**. $\operatorname{corr}(u_{jt}^{(1)}, z_t) = 0.8$ and $\operatorname{corr}(u_{jt}^{(2)}, z_t) = -0.7$.
- **Scenario 3 (Heavy-Tailed / Skewed Innovations; Robustness to Marginals)**. The Gaussian-copula dependence structure of Scenario 1 is preserved (i.e. the copula implied by (63)–(64)), but the marginals of $u_{jt}^{(i)}$ are replaced by: (a) Student-t_5, generating excess kurtosis; and (b) skewed-t_5 of Fernandez and Steel (2012), adding skewness. Both features are common in financial data (see, e.g., Harvey and Siddique, 2000, Papantonis et al., 2023). Because rank dependence is copula-invariant, this scenario isolates the effect of non-Gaussian marginals while holding rank dependence fixed.

– **Scenario 4 (Nonlinear Dependence).** The Gaussian coupling between $u_{jt}^{(i)}$ and z_t is replaced, pair by pair, by one of three Archimedean copulas – *Clayton, Gumbel, Joe* – applied to the PIT pair $\big(\mathrm{F}_{u_j^{(i)}}, \mathrm{F}_z\big)$. The within-$u$ dependence is kept Gaussian via $\mathfrak{B}^{(i)}$. The three copulas capture, respectively, lower-tail dependence (Clayton) and upper-tail dependence (Gumbel and Joe), with Joe placing more mass deeper in the up per tail at matched overall dependence.[20] The dependence parameters are calibrated so that $\mathrm{corr}(u_{jt}^{(i)}, z_t) \approx 0.70$ after the marginal transforms, giving $\psi = 3.5$ (Clayton), $\psi = 4.0$ (Joe), $\psi = 5.0$ (Gumbel).

Moreover, to allow for flexible approximation of nonlinear dependence, we augment the $\mathrm{TVAR_N}$ specification with Hermite polynomial controls $\mathcal{H}_j(z_t^{(i)\star})$. The Gaussian copula transformation $z_t^{(i)\star}$ corresponds to a sieve approximation of the conditional mean, with $j = 1$ capturing linear dependence, while higher-order terms (e.g., $j = 2$) capture nonlinear features of the threshold–innovation relationship, according to equation (61). This exercise therefore shows whether higher-order sieve terms can improve meaningfully upon the baseline Gaussian-copula PIT control when the true threshold–innovation dependence is generated by nonlinear copulas.

Given the above, we perform $1{,}000$ MC replications at $\mathrm{T} \in \{200, 500\}$, estimating δ and the remaining TVAR parameters via the procedure outlined in Section 4.[21] The densities $f_{z_t}^{(i)}$ are estimated nonparametrically with a Gaussian kernel, and the corresponding CDFs $\mathrm{F}_{z_t}^{(i)}$ by integration. For the estimation of δ, we trim 10% of the observations of z_t from each end of its sample to increase the degrees of freedom in the estimation of the model for both regimes. The tables report the average bias, $\mathrm{BIAS} = \mathbb{E}(\widehat{\delta} - \delta_0)$, and mean-squared error (MSE) of the estimates of the threshold parameter δ and of the variance–covariance parameters of the TVAR innovations. The full set of results is relegated to Appendix C; the following two subsections compare dynamic responses and relative-efficiency metrics.

[20] For $(U, V) \in [0,1]^2$, see, e.g., Nelsen (2006):

$$C_{\text{Clayton}}(U,V;\psi) = \big(U^{-\psi} + V^{-\psi} - 1\big)^{-1/\psi}, \quad \psi > 0;$$

$$C_{\text{Gumbel}}(U,V;\psi) = \exp\big\{-\big[(-\ln U)^{\psi} + (-\ln V)^{\psi}\big]^{1/\psi}\big\}, \quad \psi \geq 1;$$

$$C_{\text{Joe}}(U,V;\psi) = 1 - \big[(1-U)^{\psi} + (1-V)^{\psi} - (1-U)^{\psi}(1-V)^{\psi}\big]^{1/\psi}, \quad \psi \geq 1.$$

[21] In the context of quarterly data, $\mathrm{T} = 200$ would represent 50 years; this is a typical aggregate macro sample size.

6.2 *Monte Carlo Results: Dynamic Responses*

Figure 1 presents the dynamic responses of y_{1t} and y_{2t} to the structural errors $e_{2t}^{(1)}$ and $e_{1t}^{(2)}$ across both regimes for the true DGP (in green solid), the $\mathrm{TVAR_X}$ (red dash) and the $\mathrm{TVAR_N}$ (blue dot-dash). Aside from the former, the IRFs reported are based on the average estimates of the model over all iterations for the simulation scenario where $\mathrm{T} = 500, \mathrm{Z} \sim \mathcal{U}(-4, 3)$ and $corr(u_{jt}^{(1)}, z_t) = 0.8$ and $corr(u_{jt}^{(2)}, z_t) = -0.7$ (corresponding to **Panel B**, Table C.2).

Whereas the $\mathrm{TVAR_N}$ tracks the true IRFs almost imperceptibly, the $\mathrm{TVAR_X}$ is characterized by the earlier-identified errors: excessive initial jumps in the endogenous variables, and a dynamic error is manifest for a horizon of roughly 10-15 periods before dissipating, consistent with the finite-order VAR dynamics. In addition, there are large biases in the threshold variable: $\left|\mathbb{E}[\widehat{\delta}_{\mathrm{X}}] - \delta_0\right| \approx 2.6$; $\left|\mathbb{E}[\widehat{\delta}_{\mathrm{N}}] - \delta_0\right| \approx 0$.

6.3 *Monte Carlo Results: Relative Efficiency*

To assess the relative efficiency gains from accounting for threshold endogeneity, we compare the MSE performance of the misspecified $\mathrm{TVAR_X}$ model with that of the correctly specified $\mathrm{TVAR_N}$ model using a relative efficiency (RE) metric. The following lemma characterizes the asymptotic behavior of this metric under threshold endogeneity.

Lemma 3 (Divergence of Relative MSE under Threshold Endogeneity). *Defining Relative Efficiency as the ratio of* MSE*s for both models, for* $\mathrm{T} \in \{200, 500\}$.[22] *Then, as* $\mathrm{T} \to \infty$*:*

1. ***Threshold parameter*** $\widehat{\delta}_N$ *is* T*-consistent and asymptotically unbiased, whereas* $\widehat{\delta}_X$ *converges to a pseudo-true value* $\delta_X^\star \neq \delta_0$*. Consequently,*

$$\mathrm{RE}_{\delta,\mathrm{T}} \equiv \frac{\mathrm{MSE}_\delta(\mathrm{TVAR_X})}{\mathrm{MSE}_\delta(\mathrm{TVAR_N})} = C_\delta \mathrm{T}^2 + O(1), \qquad C_\delta > 0, \tag{65}$$

so $\mathrm{RE}_{\delta,\mathrm{T}} \to \infty$ *at rate* T^2 *(where* C_δ *is the asymptotic* MSE *constant of the correctly specified threshold estimator).*

[22] Note that the slope and loading coefficients $(A^{(i)}, \Lambda^{(i)})$ are not included in the RE metric since their performance differences manifest through the implied IRFs rather than through a single–parameter MSE comparison, whereas the threshold and regime variances admit scalar bias terms that directly govern the relative MSE behavior.

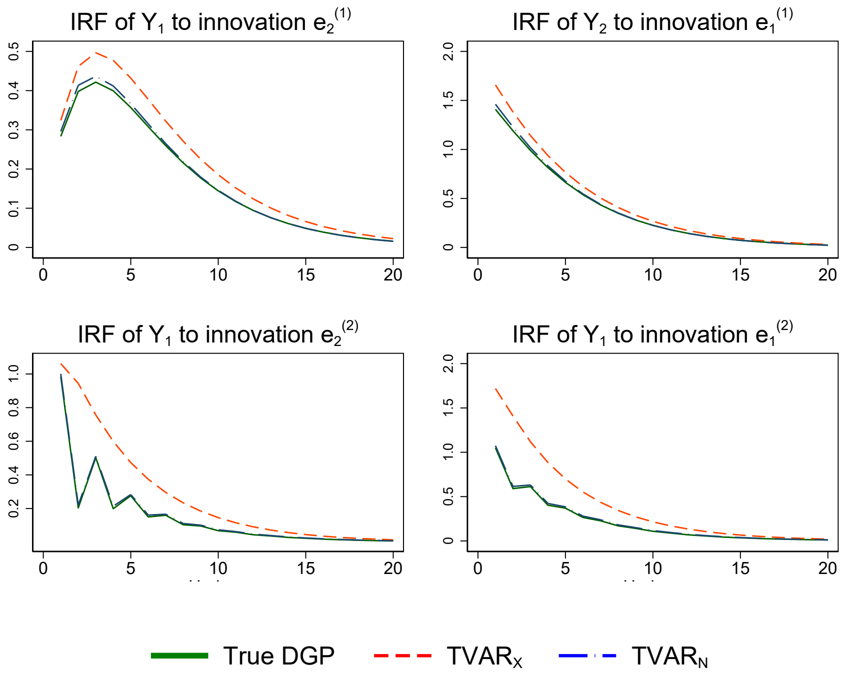

FIGURE 1. Monte Carlo: Dynamics Responses

This figure shows dynamic responses under an endogenous threshold and exogenous threshold VAR alongside those of the true data generating model. The panels plot replication-average IRFs of $y_{1,t+h}$ and $y_{2,t+h}$ to cross-equation shocks $e_t^{(1)}$ and $e_t^{(2)}$ in the two regimes. The DGP is the TVAR of the main text with $\mathrm{T} = 500$, $z_t \sim U(-4,3)$, with correlations 0.8 and -0.7. "True" denotes the theoretical IRFs from the known DGP; $\mathrm{TVAR_X}$ omits the copula control; $\mathrm{TVAR_N}$ includes it. $\mathrm{TVAR_X}$ exhibits impact overshoot and shape distortion, consistent with equations (16), (23), while $\mathrm{TVAR_N}$ closely matches the true DGP responses with tighter dispersion, consistent with equation (15).

2. ***Regime-specific variances*** *$\widehat{\Sigma}_{\varepsilon\varepsilon}^{(\mathrm{i})}$ are $\sqrt{\mathrm{T}}$-consistent and asymptotically unbiased, while $\widehat{\Sigma}_{\mathrm{uu}}^{(\mathrm{i})}$ converge to $\Sigma_{\mathrm{uu}}^{(\mathrm{i})} = \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} + \Lambda^{(\mathrm{i})}\Lambda^{(i)\prime} \neq \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$. Hence*

$$\mathrm{RE}_{\Sigma,\mathrm{T}} \equiv \frac{\mathrm{MSE_X}\left(\Sigma_{\mathrm{uu}}^{(\mathrm{i})}\right)}{\mathrm{MSE_N}\left(\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right)} = C_\Sigma \mathrm{T} + O(1), \qquad C_\Sigma > 0. \tag{66}$$

so $\mathrm{RE}_{\Sigma,\mathrm{T}} \to \infty$ at rate T.

In particular, under threshold endogeneity, the MSE of $\mathrm{TVAR_X}$ plateaus at a positive constant (the squared asymptotic bias), while the MSE of $\mathrm{TVAR_N}$ shrinks to zero. As T increases, the correctly specified $\mathrm{TVAR_N}$ therefore pulls away from $\mathrm{TVAR_X}$ in MSE terms.

PROOF. Appendix A.12. □

The divergence results in Lemma 3 follow directly from the variance–bias decomposition for δ:

$$\mathrm{MSE}(\widehat{\delta}) \equiv \mathbb{V}\mathrm{ar}(\widehat{\delta}) + \left(\mathbb{E}[\widehat{\delta}] - \delta_0\right)^2. \tag{67}$$

Under the correctly-specified $\mathrm{TVAR_N}$, both the threshold and the regime-specific variances are consistent, so their MSEs are driven by sampling variance and decay at rates T^{-2} for δ, and T^{-1} for the variances. In contrast, $\mathrm{TVAR_X}$ omits the copula control and its estimators converge to pseudo-true values shifted away from the truth by amounts proportional to $\Lambda^{(\mathrm{i})}$ (threshold) and $\Lambda^{(\mathrm{i})}\Lambda^{(i)\prime}$ (variances). These asymptotic biases, note, do not vanish with T, so the corresponding $\mathrm{TVAR_X}$ MSEs converge to positive constants. Taking the ratio therefore produces the T^2 and T divergence rates for the threshold and the variances, respectively.

Table 1 condenses our finite–sample evidence ($1,000$ replications across designs), comparing the relative efficiency based (RE) on MSE performance across the threshold parameter and reduced-form covariances. Values above one therefore quantify the MSE penalty from ignoring threshold endogeneity. Consistent with the theory, these penalties are modest in the Gaussian benchmark but become very large when the threshold–innovation dependence is heterogeneous across regimes, non-Gaussian, or non-elliptical, reflecting systematic regime misclassification and distorted innovation second moments under $\mathrm{TVAR_X}$.

In the table, $\mathcal{H}_2$ denotes the inclusion of second-order Hermite terms in the copula control, corresponding to a quadratic sieve expansion. The results indicate that even the baseline Gaussian copula control (first-order sieve) captures most of the relevant dependence, delivering substantial efficiency gains relative to $\mathrm{TVAR_X}$. The inclusion of higher-order terms provides additional (though generally moderate) refinements, particularly in settings with pronounced nonlinear or asymmetric dependence.

Overall, the Monte Carlo evidence indicates that relatively simple copula-based control functions provide a robust approximation to threshold endogeneity across a wide range of dependence structures. In particular, the Gaussian copula control, corresponding to a first-order sieve, captures most of the relevant dependence, delivering accurate impulse responses and substantial efficiency gains even when the true dependence is nonlinear or non-elliptical. Higher-order Hermite terms offer incremental improvements, especially in settings with strong asymmetries or tail dependence. These findings

TABLE 1. Relative Efficiency for Threshold Scalar and Variances

Scenario	$corr(u_{jt}^{(i)}, z_t)$	z_t distribution	RE				Comment[A]
			$\widehat{\delta}$		$\sum_{i=1,2} \mathrm{tr}(\widehat{\Sigma}_{\bullet}^{(i)})$		
			T = 200	T = 500	T = 200	T = 500	
1	0.8	$z_t \sim \mathcal{U}$	8	3	1	4	a
		$z_t \sim \mathcal{N}$	66	143	18	46	a
2	0.8; −0.7	$z_t \sim \mathcal{U}$	371	692	9	24	c
		$z_t \sim \mathcal{N}$	789	3490	5	9	d
3		$z_t \sim \mathbf{t}(5)$	9	3	2	4	e
		$z_t \sim$ skew-$\mathbf{t}(;\ 5)$	3	4	2	2	f
4	Clayton		144	760	4	9	g
	Clayton+$\mathcal{H}_2$		145	797	4	9	g
	Joe		50	463	5	5	h
	Joe+$\mathcal{H}_2$		60	493	5	5	h
	Gumbel		92	380	16	32	i
	Gumbel+$\mathcal{H}_2$		78	368	16	32	i

This table calculates the relative MSE *for the estimated threshold scalar and the sum of error variances (relative to the corresponding* $\mathrm{TVAR_X}$ *results). Full MC results are tabulated in Appendix C. Comments*[A]*: (a) Linear but non-Gaussian z: large* δ *and variance* MSE *gains from* $\mathrm{TVAR_N}$*; (b) Gaussian benchmark: modest* δ *gains; variance gains small but positive; (c) Regime-dependent correlation plus non-Gaussian* z_t*:* $\mathrm{TVAR_X}$ *essentially unusable for* δ*; (d) Regime-dependent correlation: very large threshold* MSE *cost if endogeneity ignored; (e) Heavy-tailed shocks: moderate* δ *gains, persistent variance gains; (f) Heavy-tailed and skewed shocks: similar moderate gains for* δ *and variances; (g) Lower-tail dependence: strong threshold efficiency gains particularly for large samples; (h) Upper-tail dependence: very large* δ *gains; variance gains smaller in small samples due to regime imbalance; (i) Upper-tail dependence (extreme-value structure): large and increasing gains for both* δ *and variances.*

are consistent across simulation scenarios; additional results reported in Appendix C confirm the accuracy and robustness of the approach.

7. EFFECTS OF MONETARY POLICY SHOCKS

As an application, we assess the effect of unanticipated monetary policy shocks. The choice is attractive for several reasons. First, VAR methods have dominated the empirical monetary literature, constituting a natural benchmark and entry point (e.g., Ramey, 2016, Miranda-Agrippino and Ricco, 2021). Second, there is a vibrant theoretical and empirical literature indicating that price-setting, expectation-formation, and policy responses may be characterized by nonlinearities and regime dependencies – e.g., in high versus low-inflation regimes (e.g., Golosov and Lucas, 2007, Davig and Leeper, 2008, Alvarez and Lippi, 2020). However, that literature notwithstanding, the *linear* VAR remains the core approach. This tension may partly explain the fragility of some empirical findings, such as price and output puzzles, excessive persistence, and sensitivity to different instruments, samples, and specifications (e.g., Coibion, 2012, Ramey, 2016, Bu et al., 2020, Herwartz et al., 2022).

7.1 *Inflation State Dependency in Policy Models*

The standard New Keynesian (NK) model relies on Calvo pricing, implying a constant degree of nominal rigidity and a linear Phillips curve. However, in models with menu costs or (S, s)-type pricing, the frequency of price adjustment is endogenous and governed by the cross-sectional distribution of price gaps (the deviations of current from desired prices, Khan et al., 1999, Alvarez and Lippi, 2014, Gagliardone et al., 2025). Because higher trend inflation erodes relative prices and widens these gaps more rapidly, aggregate price dynamics become highly sensitive to the inflation environment (Golosov and Lucas, 2007).

To formalize this, assume

$$\pi_t = \beta \, \mathbb{E}_t \pi_{t+1} + \mathcal{F}(z_t) \, \widetilde{y}_t, \tag{68}$$

where π_t is inflation, $\beta \in (0, 1)$ is the discount factor, $\widetilde{y}_t$ is the output gap, and function $\mathcal{F}(z_t)$ represents the slope of the Phillips curve which depends on the inflation regime (e.g., Gertler and Leahy, 2008):

$$\mathcal{F}(z_t) \equiv \begin{cases} \mathcal{F}^L & \text{if } z_t \leq \delta, \\ \mathcal{F}^H & \text{if } z_t > \delta. \end{cases} \tag{69}$$

Parameter δ denotes the threshold inflation level separating low- and high-inflation regimes. All else equal, $\mathcal{F}^H > \mathcal{F}^L$ implies a lower sacrifice ratio (where the sacrifice ratio is given by $1/\mathcal{F}$) in the high-inflation regime.

When many firms are clustered near adjustment boundaries, sufficiently large shocks or elevated trend inflation can trigger widespread simultaneous repricing (Blanco et al., 2024). This threshold specification parsimoniously approximates these extensive-margin nonlinearities. Because the true distribution of price gaps is unobservable, we proxy the inflation environment z_t using expected inflation. Following standard forecast-combination approaches (Stock and Watson, 2007, 2009, Faust and Wright, 2013), we construct the threshold (inflation state) variable as a fitted inflation-expectations measure that combines survey expectations with lagged realized inflation:

$$z_t = a_0 + a_S S_t + \sum_{j=1}^{4} a_{\pi,j}\, \pi_{t-j}, \tag{70}$$

where S_t is the expected inflation 2 quarters ahead made in period t (taken from the Livingston Survey), and π_{t-j} denotes lagged realized inflation.[23]

The fitted value, $\widehat{z}_t$, serves as our threshold variable. By combining forward-looking expectations with historical persistence, it filters out transitory noise, providing a stable, reduced-form proxy for the cumulative buildup of price misalignments that govern firms' adjustment incentives (Coibion and Gorodnichenko, 2015). Crucially, because z_t incorporates anticipated and recent inflation outcomes, it responds contemporaneously to the structural shocks driving the economy. Large monetary or cost-push shocks can directly shift the economy across the threshold δ. This inherent endogeneity of the regime state provides the foundational econometric motivation for employing an *endogenous* threshold VAR framework.

7.2 *Empirical Exercise*

Our empirical analogue of the NK model is a VAR in the *Output Gap* ($\widetilde{y}$: the percentage gap between real and potential output as calculated by the Congressional Budget Office); the rate of *Inflation* (π: the percentage change in the chain-weighted PCE price deflator at an annual rate); the *Federal Funds Rate* (i_f); and the *Excess Bond Premium* (ebp). The latter indicator is the average corporate bond spread purged of default compensation to capture financial channels, following Gilchrist and Zakrajšek (2012) (see also Caldara and Herbst (2019)). The monetary shock is the canonical Romer and Romer (2004) narrative shock using the Wieland and Yang (2020) update, which serves as a proxy for the change in the federal funds rate.[24] The sample is from 1973q1 to 2007q4, reflect-

[23] This is the longest backdated US survey of inflation expectations, dating back to 1946. The Michigan Survey and the Survey of Professional Forecasters (SPF) begin in 1978 and 1981, respectively.

[24] The monetary policy shock measure constructed by Romer and Romer (2004) is widely used as an external instrument in structural VARS (SVAR-IV). These shocks are derived in two steps: first, narrative records are used to construct a series of intended changes in the Federal Funds rate; second, these intended changes are regressed on the Fed's internal GreenBook forecasts of inflation, real output, and unemployment. The residuals from this regression isolate exogenous variations in monetary policy, capturing unexpected deviations from the Fed's systematic reaction function. In the Proxy SVAR framework (e.g., Stock and Watson, 2012, Mertens and Ravn, 2013), a reduced-form VAR is initially estimated, and the Romer and Romer series is utilized in a second stage as an external instrument. This provides the covariance restrictions necessary to identify the structural column vector associated with the monetary policy shock, allowing for the estimation of causal impulse response functions.

ing the joint time span of the monetary shock series and the premium (see Data Appendix G).

7.3 *Results*

Table 2 shows that the endogenous threshold VAR locates the threshold at $\widehat{\delta} = 3.5\%$.[25] Loosely interpreted, this suggests that the transmission of unexpected monetary policy changes is constant if z_t is at or below that value. This threshold is 150bp above the sample median and significantly different from the conventionally-understood 2% target.[26] The economy spent 51.45% of the sample in the low inflation regime, and 48.55% in the high-inflation regime. The middle block shows the regime-specific covariance matrices, showing that output and inflation volatility is around 3- and 2-times higher, respectively, and funds rate volatility around 10 times higher in the upper regime. The corresponding exogenous-threshold specification gives a substantially higher cutoff, about $\widehat{\delta}_{\mathrm{X}} = 4.8\%$.[27]

To test for the presence of significant threshold effects, we report a likelihood ratio LR statistic testing the null of a linear VAR (i.e., without threshold effects) against the alternative of the $\mathrm{TVAR}_{\mathrm{N}}$ specification. Since the two threshold regimes cannot be identified under the null, bootstrapped probability values of the statistic are reported, calculated based on a non-overlapping block-bootstrap of fixed-length using $5,000$ iterations. The linear VAR is decisively rejected against the $\mathrm{TVAR}_{\mathrm{N}}$.

7.4 *Inflation Regimes.*

Figure 2 presents the inflation regimes inferred from the $\mathrm{TVAR}_{\mathrm{N}}$ model. The solid black line denotes the observed inflation rate, while the dashed line plots the threshold inflation expectations variable and the horizontal red line marks the estimated threshold scalar. Shaded green bars indicate "high"-inflation episodes: $z_t(\pi) > \widehat{\delta}$.

The classifications closely align with established monetary and economic narratives (inter alia, King and Lu, 2021). The 1970s and early 1980s exhibit a prolonged high-inflation regime for most of the period 1974–1981, coinciding with the twin oil shocks

[25] Appendix H tabulates the full bootstrapped $\mathrm{TVAR}_{\mathrm{N}}$ and VAR parameter estimates.

[26] The 2% inflation target was explicitly adopted by the FOMC in January 2012. Bullard (2018) argued that policy makers used an implicit 2% inflation target since 1995.

[27] In terms of Proposition 3, this upward displacement is consistent with a negative contribution from $D_R'(\delta_0)$ to the right-hand slope of the misspecified objective. Economically, the exogenous-threshold TVAR shifts the regime boundary upward, reallocating marginal observations toward the lower-inflation regime to absorb part of the omitted contemporaneous threshold–innovation dependence.

TABLE 2. Endogenous Threshold VAR Estimation Results

Coefficient/Test			Value		
$\widehat{\delta}$			3.533		
			{2.403 : 4.660}		
$\widehat{\delta} = 2.00$			[0.000]		
$\widehat{\delta} = 4.80$			[0.000]		
$\mathbf{1}\{z_t > \widehat{\delta}\}/T$			0.486		
		$\widetilde{y}$	π	i_f	ebp
$\widehat{\Sigma}$(L)	$\widetilde{y}$	0.195			
	π	-0.023	0.896		
	i_f	0.066	0.062	0.152	
	ebp	-0.002	0.014	0.004	0.055
$\widehat{\Sigma}$(H)	$\widetilde{y}$	0.640			
	π	0.077	1.923		
	i_f	0.381	0.094	1.613	
	ebp	-0.010	-0.014	0.059	0.028
AIC			950.666		
LR (vs. linear VAR)			115.120		
			[0.000]		

This table shows the estimated threshold value, its duration characteristics, the covariance matrix in each regime (“(H)”igh and “(L)”ow), and some diagnostics. Numbers in braces and squared brackets indicate, respectively, 90% confidence intervals and probability-values. AIC denotes the $\mathrm{TVAR_N}$ *Information Criterion (we report the regime summed AIC). Lag orders are chosen by AIC. Full estimation results are reported in Appendix H.*

and the accommodative Burns-Miller era, which allowed inflation expectations to become entrenched. The subsequent Volcker disinflation represents a sharp transition, as the high inflation and high inflation expectation fall sharply in the early 1980s following a series of aggressive policy tightening episodes.

From the mid-1980s through the early 2000s, we identify a persistent “low” inflation regime, consistent with the Great Moderation, and with the anchoring of inflation expectations such that once the low-inflation regime was established, it became highly persistent, thus explaining why our threshold indicator remains in the low state for the rest of the sample, tying in with the notion of policy credibility and stable inflation targets.

7.5 *Dynamic Analysis*

Having estimated the models, we now examine their dynamic implications for monetary transmission. Figure 3 reports responses to the monetary policy shock over a 20-quarter horizon under the low-inflation $\mathrm{TVAR_N}$ regime (blue, solid), the linear VAR (black,

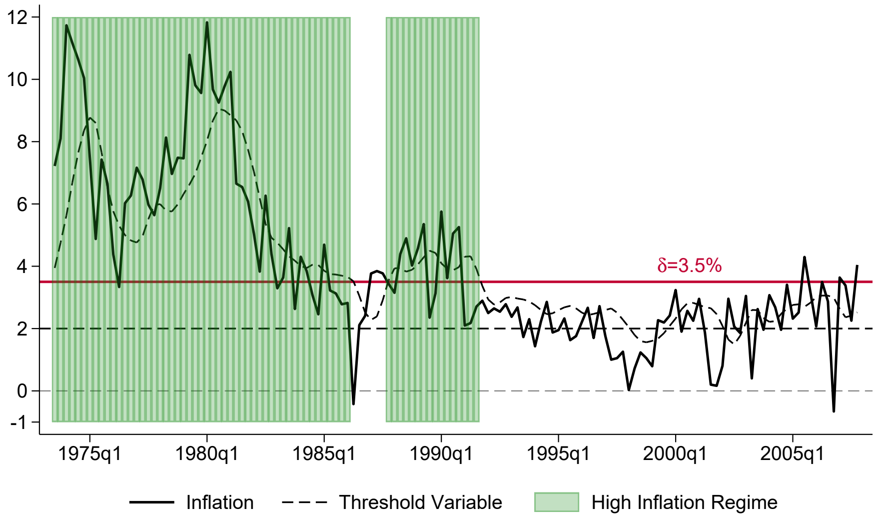

FIGURE 2. Inflation Regimes

This figure shows the inflation rate (solid line), the threshold variable (dashed), and the threshold scalar at $\widehat{\delta} = 3.53\%$ (horizontal red solid line). The bars marked in green shading demarcate high-inflation regimes given by $\mathbf{1}\{z_t(\pi) > \widehat{\delta}\}$. Two additional horizontal lines of 0 and 2 percent are drawn for visual convenience.

dashed), and the high-inflation regime (red, solid), with confidence bands. The estimated one-standard-deviation structural shocks are regime- and specification-specific. Furthermore, the reduced-form volatility of the policy equation differs across regime: larger in the high-inflation regime, smaller in the low-inflation regime. Because these differing shock sizes obscure direct comparisons, Figure 3 focuses strictly on the qualitative shape, sign, persistence, and uncertainty of the responses. A strictly normalized comparison is deferred to Figure 4.

The qualitative picture aligns tightly with regime-dependent transmission. For the output gap, both regimes display a hump-shaped contraction, but the high-regime response is sharper and troughs earlier (5/6 quarters) than the low-regime one (9/10 quarters), suggesting real activity reacts faster to tightening when inflation is elevated; the linear VAR's intermediate trough masks this timing asymmetry. For inflation, the high regime delivers a clean disinflation with tight bands excluding zero, whereas the low regime is essentially flat or mildly positive with bands straddling zero — monetary policy in low-inflation states has no statistically discernible price effect.

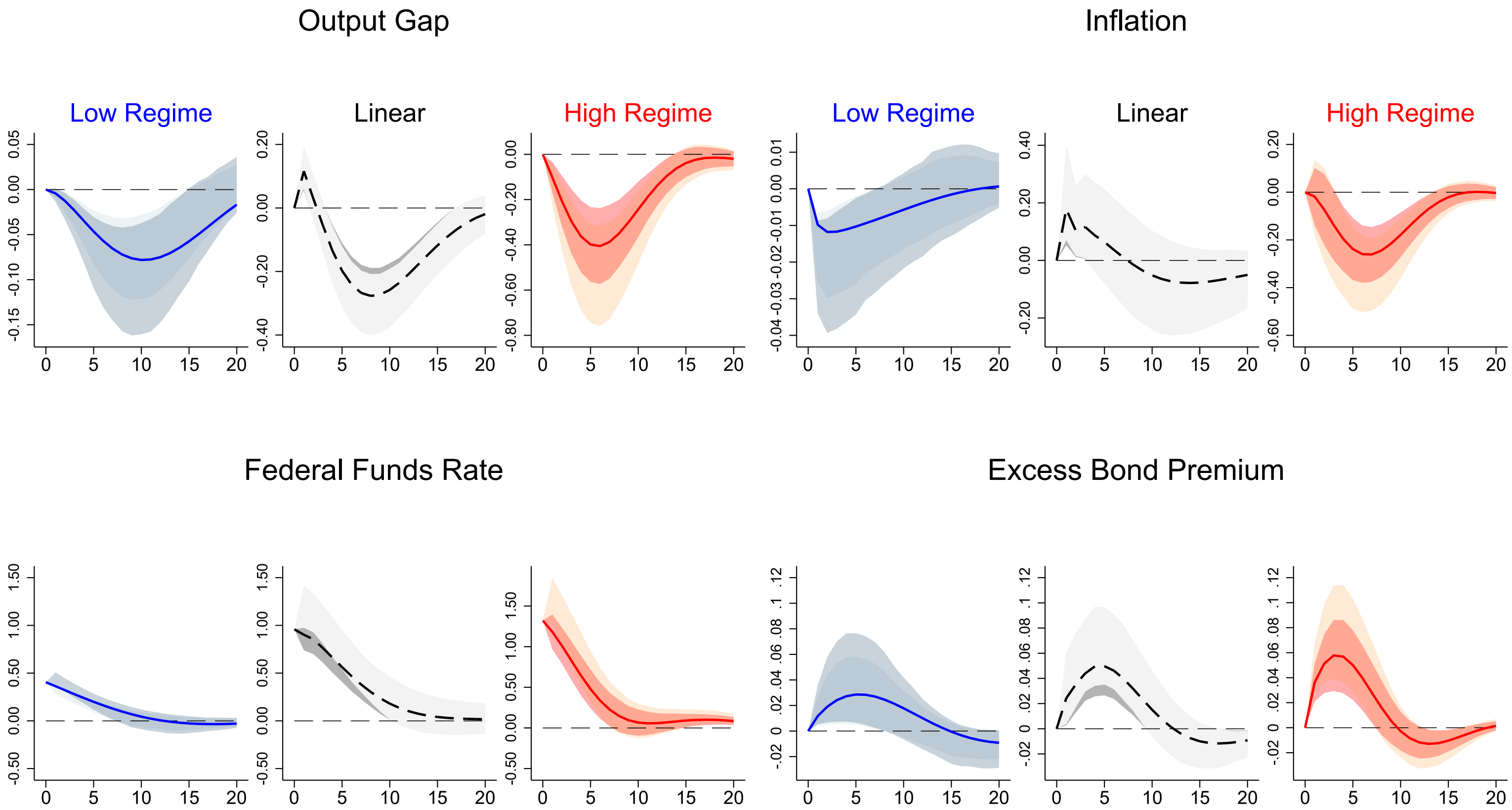

FIGURE 3. TVAR$_N$ and linear VAR Impulse Responses to a Monetary Shock

This figure reports impulse responses to an identified one-standard-deviation structural monetary policy shock over a 20-quarter horizon (where the standard deviation is estimated separately within each specification). Responses are shown for the output gap, inflation, the federal funds rate, and the excess bond premium. Blue and red lines denote responses in the low and high inflation regime, respectively for the endogenous TVAR model. Black dashed lines denote the corresponding linear VAR responses. The shaded areas indicate 68% and 90% (dark and lighter, respectively) confidence bands. For visual convenience, we overlay a horizontal dashed line marking the zero response.

The linear VAR produces the familiar price puzzle. The federal funds rate decays in all three cases but with regime-specific persistence: the high regime has the largest impact and fastest mean reversion, the low regime is smaller and more drawn out. The excess bond premium rises in a hump shape in both regimes, confirming that contractionary shocks tighten financial conditions regardless of state, but the high-regime response is sharper and more precisely estimated. Accordingly, the linear VAR is not merely a less precise version of the regime-specific responses but a qualitatively misleading one. See Figure H.1 for the $\mathrm{TVAR_X}$ IRFs.

Normalized IRFs Figure 4 isolates the regime-dependent transmission mechanism by rescaling responses to a 25-basis-point tightening. In the high-inflation regime, the shock transmits efficiently to nominal variables: inflation falls markedly (troughing near -0.05 percentage points) alongside a peak output gap contraction of -0.08, implying a relatively low dynamic sacrifice ratio[28] of ≈ 1.6. Conversely, nominal rigidities bind more strongly in the low-inflation regime; the real economy absorbs the shock via a protracted output contraction (-0.05) while the inflation response is muted (dropping by around 0.01), yielding a steep sacrifice ratio of 5.0.

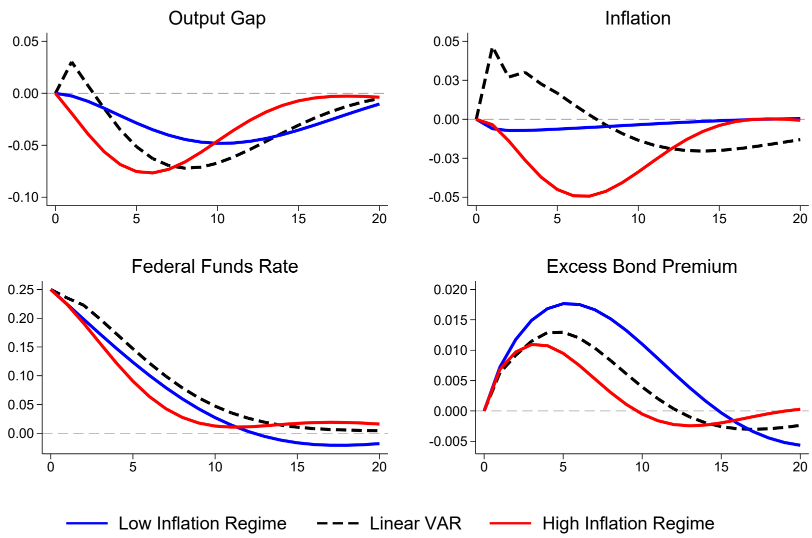

FIGURE 4. Normalized IRF Comparisons

This figure reports impulse responses to an identified monetary policy shock over a 20-quarter horizon. The responses are rescaled to show a within-regime initial increase in the funds rate of 25bp.

[28] Calculated here as the ratio of the peak output contraction to the peak inflation decline.

The funds rate mean-reverts fastest in the high regime, more slowly in the linear VAR, and slowest in the low regime, consistent with a more aggressive systematic policy response in inflationary periods. The comovement between the funds rate and ebp is likewise state dependent: the high-regime response is larger in levels, but the low-regime response is larger per basis point. A natural interpretation is that higher inflation erodes real liabilities and moves firms away from the default boundary, dampening credit-spread sensitivity to tightening; in low-inflation regimes, higher real debt burdens amplify bond premia responses through the convexity of the Merton (1974) default schedule. The risk-taking channel (Borio and Zhu, 2012, Drechsler et al., 2018) reinforces this asymmetry.

The linear VAR's magnitudes are revealing in their own right: its responses tend to lie closer to the low-regime lines, reflecting the greater sample share of low-regime observations in the sample, and its inflation response in particular fails to recover either regime's amplitude: overstating the low-regime effect and grossly understating the high-regime one. The results establish that regime asymmetries are not artifacts of differing shock sizes but reflect quantitatively large differences in transmission, and that the linear specification understates the dispersion in policy potency across regimes.

8. Conclusions

This paper develops a new method for controlling for threshold endogeneity in threshold VARs. The method uses a regime-specific copula control to remove the threshold-selection component from the conditional mean of the reduced-form innovations. It is instrument-free, does not impose a parametric form on the marginal distribution of the threshold variable, and preserves the regime-wise least-squares structure of the TVAR.

The analysis shows that threshold endogeneity is not only a threshold-estimation problem. When the contemporaneous threshold variable is correlated with the VAR innovations, the sample split changes the conditional mean and covariance structure of the reduced-form shocks, contaminates autoregressive dynamics, and distorts structural impulse responses. The copula correction addresses this conditional-mean misspecification while preserving the least-squares threshold-search structure and the Chan-type large-sample behavior of the estimator.

The Monte Carlo evidence confirms the quantitative importance of the above results. Ignoring threshold endogeneity generates sizable distortions in the estimated threshold, regime-specific covariance matrices, and impulse responses, whereas the corrected TVAR substantially reduces them across Gaussian, non-Gaussian, and nonlinear-copula designs. Thus, the simulations show that the proposed correction is not only theoretically valid under the Gaussian benchmark but also practically robust in settings that depart from it.

As an illustration, we examine the threshold characteristics of unanticipated monetary shocks. In contrast to linear VARs, the endogenous threshold VAR reveals pronounced regime dependence and variable sacrifice ratios, delivering more coherent impulse responses and historically regime characteristics. Given the widespread popularity of VARs, and the growing interest in nonlinearity and regime dependence, our method should prove valuable across a variety of applications commonly addressed in the threshold literature.

References

Alvarez, Fernando and Francesco Lippi (2014), "Price setting with menu cost for multi-product firms." *Econometrica*, 82 (1), 89–135. [7, 37]

Alvarez, Fernando and Francesco Lippi (2020), "Temporary price changes, inflation regimes, and the propagation of monetary shocks." *American Economic Journal: Macroeconomics*, 12 (1), 104–52. [36]

Auerbach, Alan J. and Yuriy Gorodnichenko (2012), "Measuring the output responses to fiscal policy." *American Economic Journal: Economic Policy*, 4 (2), 1–27. [21]

Bai, Jushan (1997), "Estimation of a change point in multiple regression models." *Review of Economics and Statistics*, 79 (4), 551–563. [21]

Blanco, Andrés, Corina Boar, Callum Jones, and Virgiliu Midrigan (2024), "Nonlinear inflation dynamics in menu cost economies." Finance and Economics Discussion Series 2024-005, Board of Governors of the Federal Reserve System. [37]

Borio, Claudio and Haibin Zhu (2012), "Capital regulation, risk-taking and monetary policy: A missing link in the transmission mechanism?" *Journal of Financial Stability*, 8 (4), 236–251. [44]

Bu, Chunya, John H. Rogers, and Wenbin Wu (2020), "Forward-looking monetary policy and the transmission of conventional monetary policy shocks." Discussion Paper 2020-014, Federal Reserve Board, Washington, D.C. [37]

Bullard, James (2018), "What Is the Best Strategy for Extending the U.S. Economy's Expansion?" Speech at CFA Society Chicago, Distinguished Speaker Series Breakfast, Sept. 12, Chicago, Ill. [39]

Caldara, Dario and Edward Herbst (2019), "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs." *American Economic Journal: Macroeconomics*, 11 (1), 157–192. [38]

Caner, Mehmet and Bruce E. Hansen (2004), "Instrumental variable estimation of a threshold model." *Econometric Theory*, 20 (5), 813–843. [3]

Chan, Kung-Sik (1993), "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model." *The Annals of Statistics*, 21 (1), 520–533. [3, 21, 24, 28, –A14– , SupApp-4]

Chen, Chaoyi, Thanasis Stengos, and Yiguo Sun (2023), "Endogeneity in semiparametric threshold regression models with two threshold variables." *Econometric Reviews*, 42 (9–10), 758–779. [5, 8]

Christopoulos, Dimitris K., Peter McAdam, and Elias Tzavalis (2021), "Dealing with endogeneity in threshold models using copulas." *Journal of Business & Economic Statistics*, 39, 166–178. [5, 8]

Coibion, Olivier (2012), "Are the effects of monetary policy shocks big or small?" *American Economic Journal: Macroeconomics*, 4 (2), 1–32. [37]

Coibion, Olivier and Yuriy Gorodnichenko (2015), "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 105 (8), 2644–2678. [38]

Davig, Troy and Eric M. Leeper (2008), "Endogenous monetary policy regime change." In *NBER International Seminar on Macroeconomics 2006*, 345–391, National Bureau of Economic Research, Cambridge, MA. [36]

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl (2018), "A model of monetary policy and risk premia." *The Journal of Finance*, 73 (1), 317–373. [44]

Duffy, James A. and Sophocles Mavroeidis (2026), "Identification in (Endogenously) Nonlinear SVARs Is Easier Than You Think." URL https://arxiv.org/abs/2604.07718. [5]

Engle, Robert F. and Aaron D. Smith (1999), "Stochastic permanent breaks." *Review of Economics and Statistics*, 81 (4), 553–574. [7]

Erdely, Arturo (2017), "Copula-based piecewise regression." In *Copulas and Dependence Models with Applications* (Manuel Úbeda Flores, Enrique de Amo Artero, Fabrizio Durante, and Juan Fernández-Sánchez, eds.), 69–81, Springer, New York. [9]

Faust, Jon and Jonathan H. Wright (2013), "Forecasting inflation." In *Handbook of Economic Forecasting* (Graham Elliott and Allan Timmermann, eds.), volume 2A, 2–56, Elsevier, Amsterdam. [37]

Favara, Giovanni, Simon Gilchrist, Kurt F. Lewis, and Egon Zakrajšek (2016), "Updating the recession risk and the excess bond premium." Feds notes, Board of Governors of the Federal Reserve System, Washington. [SupApp-12]

Fernandez, Carmen and Mark F. J. Steel (2012), "On bayesian modeling of fat tails and skewness." *Journal of the American Statistical Association*, 93, 359–371. [31]

Gabaix, Xavier and Ralph S. J. Koijen (2024), "Granular instrumental variables." *Journal of Political Economy*, 132 (7), 2274–2303. [5]

Gagliardone, Luca, Mark Gertler, Simone Lenzu, and Joris Tielens (2025), "Micro and macro cost-price dynamics in normal times and during inflation surges." Working Paper 33478, National Bureau of Economic Research. [37]

Gertler, Mark and John Leahy (2008), "A phillips curve with an ss foundation." *Journal of Political Economy*, 116 (3), 533–572. [37]

Gilchrist, Simon and Egon Zakrajšek (2012), "Credit spreads and business cycle fluctuations." *American Economic Review*, 102 (4), 1692–1720. [38, SupApp-5, SupApp-12]

Golosov, Mikhail and Robert E. Lucas (2007), "Menu Costs and Phillips Curves." *Journal of Political Economy*, 115 (2), 171–199. [7, 36, 37]

Guerrieri, Luca and Matteo Iacoviello (2017), "Collateral constraints and macroeconomic asymmetries." *Journal of Monetary Economics*, 90 (C), 28–49. [7]

Hansen, Bruce E. (1999), "Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference." *Journal of Econometrics*, 93 (2), 345–368. [3]

Hansen, Bruce E. (2000), "Sample splitting and threshold estimation." *Econometrica*, 68 (3), 575–603. [3, 21]

Harvey, Campbell R. and Akhtar Siddique (2000), "Conditional skewness in asset pricing tests." *Journal of Finance*, 55, 1263–1295. [31]

Herwartz, Helmut, Hannes Rohloff, and Shu Wang (2022), "Proxy SVAR identification of monetary policy shocks - Monte Carlo evidence and insights for the US." *Journal of Economic Dynamics and Control*, 139, 104457. [37]

Kapetanios, George and Elias Tzavalis (2010), "Modeling structural breaks in economic relationships using large shocks." *Journal of Economic Dynamics and Control*, 34 (3), 417–436. [7]

Khan, Aubhik, Robert G. King, and Alexander L. Wolman (1999), "Optimal monetary policy." *Review of Economic Studies*, 66 (4), 825–860. [37]

King, Robert G. and Yang K. Lu (2021), "The Rise, Fall and Stabilization of U.S. Inflation: Shifting Regimes and Evolving Reputation." NBER Working Papers 29585, National Bureau of Economic Research, Inc. [39]

Kourtellos, Andros, Thanasis Stengos, and Yiguo Sun (2022), "Endogeneity in semiparametric threshold regression." *Econometric Theory*, 38 (3), 562–595. [5, 8]

Kourtellos, Andros, Thanasis Stengos, and Chih Ming Tan (2016), "Structural threshold regression." *Econometric Theory*, 32, 827–860. [5, 8]

Lee, Lung-Fei (1982), "Some approaches to the correction of selectivity bias." *Review of Economic Studies*, 49, 355–372. [5, 8]

Lee, Lung-Fei (1983), "Generalized econometric models with selectivity." *Econometrica*, 51, 507–512. [5, 7, 8]

Lo, Ming Chien and Jeremy Piger (2005), "Is the response of output to monetary policy asymmetric? evidence from a regime-switching var." *Journal of Money, Credit and Banking*, 37 (5), 865–886. [21]

Maddala, Gangadharrao Soundalyarao (1983), *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge. [5, 7, 8]

Mertens, Karel and Morten O. Ravn (2013), "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review*, 103 (4), 1212–1247. [13, 38]

Mertens, Karel and Morten O. Ravn (2014), "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics*, 68 (S), 1–19. [13]

Merton, Robert C. (1974), "On the pricing of corporate debt: The risk structure of interest rates." *The Journal of Finance*, 29 (2), 449–470. [44]

Miranda-Agrippino, Silvia and Giovanni Ricco (2021), "The transmission of monetary policy shocks." *American Economic Journal: Macroeconomics*, 13 (3), 74–107. [6, 36]

Nelsen, Roger B. (2006), *An Introduction to Copulas*, 2nd edition. Springer, New York. [32]

Papantonis, Ioannis, Leonidas Rompolis, and Elias Tzavalis (2023), "Improving variance forecasts: The role of realized variance features." *International Journal of Forecasting*, 39 (3), 1221–1237. [31]

Park, Sungho and Sachin Gupta (2012), "Handling endogenous regressors by joint estimations using copulas." *Marketing Science*, 31, 567–586. [8]

Ramey, Valerie A. (2016), "Macroeconomic shocks and their propagation." In *Handbook of Macroeconomics*, volume 2, 71–162, Elsevier. [36, 37]

Romer, Christina D. and David H. Romer (2004), "A new measure of monetary shocks: Derivation and implications." *American Economic Review*, 94 (4), 1055–1084. [38, SupApp-12]

Samia, Noelle I. and Kung-Sik Chan (2011), "Maximum likelihood estimation of a generalized threshold stochastic regression model." *Biometrika*, 98 (2), 433–448. [24]

Siburg, Karl Friedrich and Pavel A. Stoimenov (2008), "Gluing copulas." *Communications in Statistics, Theory and Methods*, 37, 3124–3134. [9]

Silverman, Bernard W. (1986), *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London. [11]

Sims, Christopher A. and Tao Zha (2006), "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96 (1), 54–81. [7]

Sklar, Abe (1959), "Fonctions de répartition à n dimensions et leurs marges." *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231. [7, –A1–]

Spanos, Aris (2018), "Mis-specification testing in retrospect." *Journal of Economic Surveys*, 32, 541–577. [12]

Stock, James H. and Mark W. Watson (2007), "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 39 (1), 3–33. [37]

Stock, James H. and Mark W. Watson (2009), "Phillips curve inflation forecasts." In *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective* (Jeffrey Fuhrer, Yolanda Kodrzycki, Jane Little, and Giovanni Olivei, eds.), 99–202, MIT Press, Cambridge, MA. [37]

Stock, James H. and Mark W. Watson (2012), "Disentangling the channels of the 2007-09 recession." *Brookings Papers on Economic Activity*, 42 (1), 81–135. [38]

Stock, James H and Mark W Watson (2018), “Identification and estimation of dynamic causal effects in macroeconomics using external instruments.” *The Economic Journal*, 128 (610), 917–948. [13]

Tong, Howell and K. S. Lim (1980), “Threshold autoregressions, limit cycles and cyclical data.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 42 (3), 245–292. [3]

Tsay, Ruey S. (1998), “Testing and modeling multivariate threshold models.” *Journal of the American Statistical Association*, 93, 1188–1202. [21, 24]

Vella, Francis (1998), “Estimating Models With Sample Selection Bias: A Survey.” *Journal of Human Resources*, 33, 127–169. [5, 8]

Wieland, Johannes F. and Mu-Jeung Yang (2020), “Financial dampening.” *Journal of Money, Credit and Banking*, 52 (1), 79–113. [38, SupApp-12]

Zhang, Jianhan, Chaoyi Chen, Yiguo Sun, and Thanasis Stengos (2025), “Endogenous kink threshold regression.” *Journal of Business & Economic Statistics*, 43 (3), 556–567. [5]

APPENDIX A: PROOFS

This appendix presents detailed proofs of the main results and collects auxiliary lemmas used in these proofs. Throughout, all stochastic orders are understood uniformly over the parameter space.

A.1 *Proof of Proposition 1*

PROOF. By the PIT, for regime i,

$$z_t^{(\mathrm{i})\star} = \Phi^{-1}\left(\mathrm{F}_{z_t}^{(\mathrm{i})}\right) \sim \mathcal{N}(0,1), \qquad u_{kt}^{(i)\star} = \Phi^{-1}\left(\mathrm{F}_{u_{kt}}^{(\mathrm{i})})\right) \sim \mathcal{N}(0,1), \quad k=1,\dots,\mathrm{K}, \tag{A.1}$$

where $\mathrm{F}_{z_t}^{(\mathrm{i})}$ is the regime-i truncated CDF of z_t and $\mathrm{F}_{u_{kt}}^{(\mathrm{i})}$ are the marginals of the $u_{kt}^{(\mathrm{i})}, k = 1,2,...,K$. By Sklar (1959)'s theorem and the Gaussian-copula assumption, the K-vector $u_t^{(i)\star} = (u_{1t}^{(i)\star},\dots,u_{\mathrm{K}t}^{(i)\star})'$ and $z_t^{(\mathrm{i})\star}$ are jointly Gaussian with mean zero and correlation matrix,

$$\begin{pmatrix} u_t^{(i)\star} \\ z_t^{(i)\star} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})} & r_{uz}^{(\mathrm{i})} \\ r_{uz}^{(i)\prime} & 1 \end{pmatrix}\right) \tag{A.2}$$

where, as before, we have the correlation objects, $\mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}}$, $r_{uz}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times 1}$. The last result implies that

$$\mathbb{E}\big(u_t^{(i)\star} \mid z_t^{(\mathrm{i})\star}\big) = r_{uz}^{(\mathrm{i})} z_t^{(\mathrm{i})\star}, \qquad \mathbb{V}\mathrm{ar}\left(u_t^{(i)\star} \mid z_t^{(\mathrm{i})\star}\right) = \mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})} - r_{uz}^{(\mathrm{i})} r_{uz}^{(i)\prime}. \tag{A.3}$$

Equivalently, by joint normality, there exists a K-vector $\eta_t^{(\mathrm{i})}$, independent of $z_t^{(\mathrm{i})\star}$, such that

$$u_t^{(i)\star} = r_{uz}^{(\mathrm{i})} z_t^{(i)\star} + \eta_t^{(\mathrm{i})}, \qquad \eta_t^{(\mathrm{i})} \perp z_t^{(i)\star}, \qquad \eta_t^{(\mathrm{i})} \sim \mathcal{N}\big(0,\ R_{\mathrm{uu}}^{(\mathrm{i})} - r_{uz}^{(\mathrm{i})} r_{uz}^{(i)\prime}\big). \tag{A.4}$$

Since the marginals of $u_{kt}^{(\mathrm{i})}$ are $\mathcal{N}\big(0,(\sigma_{uk}^{(\mathrm{i})})^2\big)$, write $\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} = \mathrm{diag}(\sigma_{u1}^{(\mathrm{i})},\dots,\sigma_{uK}^{(\mathrm{i})})$ so that $u_t^{(\mathrm{i})} = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} u_t^{(i)\star}$ component-wise. Multiplying (A.4) by $\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}$ and defining

$$\Lambda^{(\mathrm{i})} = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} r_{uz}^{(\mathrm{i})}, \qquad \varepsilon_t^{(\mathrm{i})} = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} \eta_t^{(\mathrm{i})}, \tag{A.5}$$

we obtain the single-factor representation

$$u_t^{(\mathrm{i})} = \Lambda^{(\mathrm{i})} z_t^{(\mathrm{i})\star} + \varepsilon_t^{(\mathrm{i})}, \qquad \varepsilon_t^{(\mathrm{i})} \perp z_t^{(\mathrm{i})\star}. \tag{A.6}$$

Its covariance follows immediately:

$$\mathbb{V}\mathrm{ar}(\varepsilon_t^{(\mathrm{i})}) = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})} \left(\mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})} - r_{uz}^{(\mathrm{i})} r_{uz}^{(i)\prime}\right) \mathrm{D}_{\mathrm{uu}}^{(i)\prime} = \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}. \tag{A.7}$$

The conditional density $f_{u_t|z_t}^{(\mathrm{i})}$ in (4) factorizes via Sklar's theorem into a Gaussian copula and Gaussian marginal densities for $u_t^{(\mathrm{i})}$. In PIT–Gaussian coordinates, this implies joint normality, so the conditional mean is linear and the conditional variance is constant, as in (A.3).

Because $z_t^{(i)\star} = \Phi^{-1}\left(F_{z_t}^{(i)}\right)$ is a measurable function of z_t within regime i, we have

$$\mathbb{E}\big(u_t^{(i)} \mid z_t^{(i)}\big) \;=\; \mathbb{E}\big(\Lambda^{(i)} z_t^{(i)\star} + \varepsilon_t^{(i)} \,\big|\, z_t^{(i)}\big) \;=\; \Lambda^{(i)} z_t^{(i)\star} + \mathbb{E}\big(\varepsilon_t^{(i)} \mid z_t^{(i)}\big) = \Lambda^{(i)} z_t^{(i)\star}, \tag{A.8}$$

where the last equality follows from the independence of $\varepsilon_t^{(i)}$ and $z_t^{(i)\star}$. This proves both the orthogonal decomposition and the linear selection term. The exogenous-threshold case corresponds to $r_{uz}^{(i)} = 0$ hence $\Lambda^{(i)} = 0$ and $\Sigma_{\varepsilon\varepsilon}^{(i)} = D_{uu}^{(i)} R_{uu}^{(i)} D_{uu}^{(i)\prime}$. □

A.2 *Proof of Lemma 1*

PROOF. From $u_t^{(i)} = \Lambda^{(i)} z_t^{(i)\star} + \varepsilon_t^{(i)}$, with $\mathbb{E}\left(z_t^{(i)\star} \varepsilon_t^{(i)\prime}\right) = 0$ and $\mathbb{V}\mathrm{ar}(z_t^{(i)\star}) = 1$, it follows that

$$\Sigma_{uu}^{(i)} \;=\; \mathbb{V}\mathrm{ar}(u_t^{(i)}) = \Lambda^{(i)}\Lambda^{(i)\prime} + \Sigma_{\varepsilon\varepsilon}^{(i)}. \tag{A.9}$$

Hence,

$$\Sigma_{uu}^{(i)} - \Sigma_{\varepsilon\varepsilon}^{(i)} \;=\; \Lambda^{(i)}\Lambda^{(i)\prime}. \tag{A.10}$$

For any $a \in \mathbb{R}^{K}$,

$$a'\left(\Sigma_{uu}^{(i)} - \Sigma_{\varepsilon\varepsilon}^{(i)}\right) a \;=\; a'\Lambda^{(i)}\Lambda^{(i)\prime} a \;=\; \left(\Lambda^{(i)\prime} a\right)^2 \geq 0. \tag{A.11}$$

Thus, $\Sigma_{uu}^{(i)} - \Sigma_{\varepsilon\varepsilon}^{(i)}$ is positive semi-definite and therefore $\Sigma_{uu}^{(i)} \succeq \Sigma_{\varepsilon\varepsilon}^{(i)}$ in the Loewner order. As a corollary,

$$\mathrm{tr}(\Sigma_{uu}^{(i)}) - \mathrm{tr}(\Sigma_{\varepsilon\varepsilon}^{(i)}) \;=\; \mathrm{tr}(\Lambda^{(i)}\Lambda^{(i)\prime}) = \|\Lambda^{(i)}\|_F^2 \geq 0, \tag{A.12}$$

which shows that the total innovation variance increases when the copula control is omitted. □

A.3 *Proof of Lemma 2*

PROOF. We prove the two bounds separately. Let $B^{(i)}$ denote the regime-i VAR slope matrix entering the first block row of the companion matrix. Since the lower block of the companion matrix is fixed, the companion-matrix error is a linear function of the slope-coefficient error. Hence there exists a fixed embedding matrix $\mathcal{R} \in \mathbb{R}^{(Kp)^2 \times K^2 p}$ such that

$$\mathrm{vec}\left(\widehat{\mathcal{A}}^{(i)} - \mathcal{A}^{(i)}\right) \;=\; \mathcal{R}\,\mathrm{vec}\left(\widehat{B}^{(i)} - B^{(i)}\right). \tag{A.13}$$

First consider the correctly specified $\mathrm{TVAR_N}$ estimator. Standard least-squares theory gives

$$\mathbb{V}\mathrm{ar}\left[\mathrm{vec}\left(\widehat{B}_{\mathrm{N}}^{(\mathrm{i})}-B_{\mathrm{N}}^{(\mathrm{i})}\right)\right] \;=\; \frac{1}{\mathrm{T}^{(\mathrm{i})}}\left[\left(Q_{xx,\mathrm{N}}^{(\mathrm{i})}\right)^{-1}\otimes\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right]. \tag{A.14}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left\|\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F^2 &= \mathrm{tr}\left\{\mathbb{V}\mathrm{ar}\left[\mathrm{vec}\left(\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right)\right]\right\}\\ &= \frac{1}{\mathrm{T}^{(\mathrm{i})}}\mathrm{tr}\left\{\mathcal{R}\left[\left(Q_{xx,\mathrm{N}}^{(\mathrm{i})}\right)^{-1}\otimes\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}\right]\mathcal{R}'\right\}\equiv V_{\mathrm{N}}^{(\mathrm{i})}.\end{aligned} \tag{A.15}$$

By Markov's inequality,

$$\Pr\left(\left\|\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F\geq\mathfrak{r}\right)\leq\frac{V_{\mathrm{N}}^{(\mathrm{i})}}{\mathfrak{r}^2}. \tag{A.16}$$

Since $\mathfrak{r}<\mathfrak{r}_{\mathrm{N,stab}}^{(\mathrm{i})}$, the event $\left\|\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F<\mathfrak{r}$ implies $\rho(\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})})<1$. Hence

$$\Pr\left(\rho(\widehat{\mathcal{A}}_{\mathrm{N}}^{(\mathrm{i})})<1\right)\geq 1-\frac{V_{\mathrm{N}}^{(\mathrm{i})}}{\mathfrak{r}^2}. \tag{A.17}$$

Now consider the misspecified $\mathrm{TVAR_X}$ estimator. Its least-squares fluctuations are centered around the pseudo-true misspecified companion matrix $\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}$. Thus,

$$\mathbb{V}\mathrm{ar}\left[\mathrm{vec}\left(\widehat{B}_{\mathrm{X}}^{(\mathrm{i})}-B_{\mathrm{X}}^{(\mathrm{i})}\right)\right] \;=\; \frac{1}{\mathrm{T}^{(\mathrm{i})}}\left[\left(Q_{xx,\mathrm{X}}^{(\mathrm{i})}\right)^{-1}\otimes\Sigma_{\mathrm{uu}}^{(\mathrm{i})}\right], \tag{A.18}$$

and therefore

$$\mathbb{E}\left\|\widehat{\mathcal{A}}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}\right\|_F^2 = \frac{1}{\mathrm{T}^{(\mathrm{i})}}\mathrm{tr}\left\{\mathcal{R}\left[\left(Q_{xx,\mathrm{X}}^{(\mathrm{i})}\right)^{-1}\otimes\Sigma_{\mathrm{uu}}^{(\mathrm{i})}\right]\mathcal{R}'\right\}\equiv V_{\mathrm{X}}^{(\mathrm{i})}. \tag{A.19}$$

Relative to the correctly specified companion matrix, $\widehat{\mathcal{A}}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}=\left(\widehat{\mathcal{A}}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}\right)+\left(\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right)$, the first term is centered around zero. Hence,

$$\mathbb{E}\left\|\widehat{\mathcal{A}}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F^2 \;=\; V_{\mathrm{X}}^{(\mathrm{i})}+\left\|\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F^2. \tag{A.20}$$

Applying Markov's inequality gives

$$\Pr\left(\left\|\widehat{\mathcal{A}}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F\geq\mathfrak{r}\right)\leq\frac{V_{\mathrm{X}}^{(\mathrm{i})}+\left\|\mathcal{A}_{\mathrm{X}}^{(\mathrm{i})}-\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}\right\|_F^2}{\mathfrak{r}^2}. \tag{A.21}$$

Again, since $\mathfrak{r} < \mathfrak{r}^{(\mathrm{i})}_{\mathrm{N,stab}}$, the event $\left\| \widehat{\mathcal{A}}^{(\mathrm{i})}_{\mathrm{X}} - \mathcal{A}^{(\mathrm{i})}_{\mathrm{N}} \right\|_F < \mathfrak{r}$ implies $\rho(\widehat{\mathcal{A}}^{(\mathrm{i})}_{\mathrm{X}}) < 1$. Therefore,

$$\Pr\left(\rho(\widehat{\mathcal{A}}^{(\mathrm{i})}_{\mathrm{X}}) < 1\right) \geq 1 - \frac{V^{(\mathrm{i})}_{\mathrm{X}} + \left\| \mathcal{A}^{(\mathrm{i})}_{\mathrm{X}} - \mathcal{A}^{(\mathrm{i})}_{\mathrm{N}} \right\|_F^2}{\mathfrak{r}^2}. \tag{A.22}$$

This proves the result. □

A.4 *Proof of Proposition 2*

PROOF. The proof follows from the decomposition of the reduced-form innovation,

$$u^{(\mathrm{i})}_t \; = \; \Lambda^{(\mathrm{i})} z^{(\mathrm{i})\star}_t + \varepsilon^{(\mathrm{i})}_t. \tag{A.23}$$

Taking covariances with the external instrument $v^{(\mathrm{i})}_t$ gives

$$\mathbb{C}\mathrm{ov}\left(u^{(\mathrm{i})}_t, v^{(\mathrm{i})}_t\right) = \Lambda^{(\mathrm{i})} \mathbb{C}\mathrm{ov}\left(z^{(\mathrm{i})\star}_t, v^{(\mathrm{i})}_t\right) + \mathbb{C}\mathrm{ov}\left(\varepsilon^{(\mathrm{i})}_t, v^{(\mathrm{i})}_t\right). \tag{A.24}$$

By definition, we have $\omega^{(\mathrm{i})} = \mathbb{C}\mathrm{ov}\left(z^{(\mathrm{i})\star}_t, v^{(\mathrm{i})}_t\right)$. Using the structural mapping

$$\varepsilon^{(\mathrm{i})}_t \; = \; \mathrm{P}^{(\mathrm{i})}_{\varepsilon\varepsilon} e^{(\mathrm{i})}_t, \qquad \mathrm{P}^{(\mathrm{i})}_{\varepsilon\varepsilon} \; = \; \left(p^{(\mathrm{i})}_1, \ldots, p^{(\mathrm{i})}_{\mathrm{K}}\right), \tag{A.25}$$

and the proxy relevance and exogeneity conditions,

$$\mathbb{C}\mathrm{ov}(e^{(\mathrm{i})}_{1t}, v^{(\mathrm{i})}_t) = \lambda^{(\mathrm{i})}_v \neq 0, \qquad \mathbb{C}\mathrm{ov}(e^{(\mathrm{i})}_{jt}, v^{(\mathrm{i})}_t) = 0, \quad j \neq 1, \tag{A.26}$$

we have

$$\mathbb{C}\mathrm{ov}\left(\varepsilon^{(\mathrm{i})}_t, v^{(\mathrm{i})}_t\right) \; = \; \mathrm{P}^{(\mathrm{i})}_{\varepsilon\varepsilon} \mathbb{C}\mathrm{ov}(e^{(\mathrm{i})}_t, v^{(\mathrm{i})}_t) \; = \; \lambda^{(\mathrm{i})}_v p^{(\mathrm{i})}_1. \tag{A.27}$$

Case i) $\mathbf{TVAR_N}$. When the proxy step is applied to the corrected innovations $\varepsilon^{(\mathrm{i})}_t$, the proxy moment is

$$\Sigma^{(\mathrm{i})}_{\varepsilon v} \; = \; \mathbb{C}\mathrm{ov}(\varepsilon^{(\mathrm{i})}_t, v^{(\mathrm{i})}_t) \; = \; \lambda^{(\mathrm{i})}_v p^{(\mathrm{i})}_1. \tag{A.28}$$

Thus, the corrected proxy moment identifies (up to scale/sign) the impact column $p^{(\mathrm{i})}_1$ of the orthogonalized structural shock in $\varepsilon^{(\mathrm{i})}_t$, up to normalization. This result is unaffected by $\omega^{(\mathrm{i})}$, because the threshold-selection component has been removed from $\varepsilon^{(\mathrm{i})}_t$.

Case ii) $\mathbf{TVAR_X}$. When the proxy step is instead applied to the uncorrected innovation $u^{(\mathrm{i})}_t$, substituting (A.27) into (A.24) gives

$$\Sigma^{(\mathrm{i})}_{uv} \; = \; \mathbb{C}\mathrm{ov}(u^{(\mathrm{i})}_t, v^{(\mathrm{i})}_t) \; = \; \lambda^{(\mathrm{i})}_v p^{(\mathrm{i})}_1 + \Lambda^{(\mathrm{i})} \omega^{(\mathrm{i})}. \tag{A.29}$$

If $\lambda_v^{(\mathrm{i})} \neq 0$, this moment is proportional to the composite direction

$$p_{\mathrm{comp}}^{(\mathrm{i})} \equiv p_1^{(\mathrm{i})} + \frac{1}{\lambda_v^{(\mathrm{i})}} \Lambda^{(\mathrm{i})} \omega^{(\mathrm{i})}. \tag{A.30}$$

Hence, if the target is the orthogonalized shock $p_1^{(\mathrm{i})}$, the term $(\omega^{(\mathrm{i})}/\lambda_v^{(\mathrm{i})})\Lambda^{(\mathrm{i})}$ is threshold-selection contamination. If the target is instead the full event that also moves the threshold variable contemporaneously, then $p_{\mathrm{comp}}^{(\mathrm{i})}$ is the relevant estimand. When $\omega^{(\mathrm{i})} = 0$, the threshold score is orthogonal to the instrument, and the corrected and uncorrected proxy moments identify the same impact direction. □

Corollary A.1 (Magnitude of threshold-selection contamination). *Under the assumptions of Proposition 2, with $\lambda_v^{(i)} \neq 0$, decompose the threshold loading as $\Lambda^{(i)} = \Lambda_{\parallel}^{(i)} + \Lambda_{\perp}^{(i)}$, where*

$$\Lambda_{\parallel}^{(i)} = \frac{p_1^{(i)\prime} \Lambda^{(i)}}{\|p_1^{(i)}\|_2^2} p_1^{(i)}, \qquad \Lambda_{\perp}^{(i)} = \Lambda^{(i)} - \Lambda_{\parallel}^{(i)}, \tag{A.31}$$

are, respectively, the components of $\Lambda^{(i)}$ parallel and orthogonal to the target impact column $p_1^{(i)}$. Then the angle $\theta^{(i)} \in [0, \pi/2]$ between the uncorrected proxy moment $\Sigma_{uv}^{(i)} \neq 0$ and $p_1^{(i)} \neq 0$ satisfies the exact identities

$$\tan\theta^{(i)} = \frac{|\omega^{(i)}| \, \|\Lambda_{\perp}^{(i)}\|_2}{\left\|\lambda_v^{(i)} p_1^{(i)} + \omega^{(i)} \Lambda_{\parallel}^{(i)}\right\|_2}, \tag{A.32}$$

$$\sin\theta^{(i)} = \frac{|\omega^{(i)}| \, \|\Lambda_{\perp}^{(i)}\|_2}{\sqrt{\left\|\lambda_v^{(i)} p_1^{(i)} + \omega^{(i)} \Lambda_{\parallel}^{(i)}\right\|_2^2 + (\omega^{(i)})^2 \, \|\Lambda_{\perp}^{(i)}\|_2^2}}. \tag{A.33}$$

In particular, in the small-contamination regime $|\omega^{(i)}| \, \|\Lambda^{(i)}\|_2 \ll |\lambda_v^{(i)}| \, \|p_1^{(i)}\|_2$,

$$\theta^{(i)} = \frac{|\omega^{(i)}| \, \|\Lambda_{\perp}^{(i)}\|_2}{|\lambda_v^{(i)}| \, \|p_1^{(i)}\|_2} + O\left(\left[\frac{|\omega^{(\mathrm{i})}| \|\Lambda^{(\mathrm{i})}\|_2}{|\lambda_v^{(\mathrm{i})}| \|p_1^{(\mathrm{i})}\|_2}\right]^2\right) \tag{A.34}$$

Two limiting cases deserve emphasis. (a) If $\omega^{(i)} = 0$, then $\theta^{(i)} = 0$ and the uncorrected proxy moment identifies $p_1^{(i)}$ exactly, recovering Case (i) of Proposition 2. (b) If $\Lambda^{(i)} \propto p_1^{(i)}$, then $\Lambda_{\perp}^{(i)} = 0$ and $\theta^{(i)} = 0$: threshold selection rescales the impact moment but does not rotate it, so directional identification is preserved even when $\omega^{(i)} \neq 0$.

PROOF. From (32), $\Sigma_{uv}^{(i)} = \lambda_v^{(i)} p_1^{(i)} + \omega^{(i)} \Lambda^{(i)}$. Substituting $\Lambda^{(i)} = \Lambda_{\parallel}^{(i)} + \Lambda_{\perp}^{(i)}$ and noting $\Lambda_{\parallel}^{(i)} \parallel p_1^{(i)}$ while $\Lambda_{\perp}^{(i)} \perp p_1^{(i)}$, the orthogonal decomposition of $\Sigma_{uv}^{(i)}$ relative to $p_1^{(i)}$ is

$$\underbrace{\lambda_v^{(i)} p_1^{(i)} + \omega^{(i)} \Lambda_{\parallel}^{(i)}}_{\text{parallel to } p_1^{(i)}} + \underbrace{\omega^{(i)} \Lambda_{\perp}^{(i)}}_{\text{orthogonal to } p_1^{(i)}} . \tag{A.35}$$

Identity (A.32) follows from $\tan\theta^{(i)} = \|\text{orthogonal component}\|_2 / \|\text{parallel component}\|_2$, and (A.33) from $\sin\theta^{(i)} = \|\text{orthogonal component}\|_2 / \|\Sigma_{uv}^{(i)}\|_2$ with $\|\Sigma_{uv}^{(i)}\|_2^2 = \|\lambda_v^{(i)} p_1^{(i)} + \omega^{(i)} \Lambda_{\parallel}^{(i)}\|_2^2 + (\omega^{(i)})^2 \|\Lambda_{\perp}^{(i)}\|_2^2$ by Pythagoras. For (A.34), write $\delta^{(i)} = \omega^{(i)} / \lambda_v^{(i)}$. The denominator in (A.33) equals $|\lambda_v^{(i)}| \, \|p_1^{(i)}\|_2 \cdot \big(1 + O(\delta^{(i)})\big)$, and a first-order Taylor expansion combined with $\theta^{(i)} = \sin\theta^{(i)} + O\big((\sin\theta^{(i)})^3\big)$ yields the stated expansion. The two limiting cases are immediate from (A.32). □

A.5 *Proof of Proposition 3*

PROOF. Let

$$G(\delta) \;=\; Q_{\mathrm{X}}(\delta) - Q_{\mathrm{N}}(\delta) \tag{A.36}$$

denote the misspecification gap. By assumption,

$$G(\delta) - G(\delta_0) \;=\; D_R(\delta) - D_R(\delta_0) + o(|\delta - \delta_0|). \tag{A.37}$$

Since D_R is differentiable at δ_0,

$$G(\delta) - G(\delta_0) \;=\; D_R'(\delta_0)(\delta - \delta_0) + o(|\delta - \delta_0|). \tag{A.38}$$

For $\delta > \delta_0$, the one-sided expansion of Q_{N} gives

$$Q_{\mathrm{N}}(\delta) - Q_{\mathrm{N}}(\delta_0) \;=\; h^{+}(\delta - \delta_0) + o(|\delta - \delta_0|). \tag{A.39}$$

Therefore,

$$Q_{\mathrm{X}}(\delta) - Q_{\mathrm{X}}(\delta_0) \;=\; \{Q_{\mathrm{N}}(\delta) - Q_{\mathrm{N}}(\delta_0)\} + \{G(\delta) - G(\delta_0)\} \tag{A.40}$$

$$\;=\; \big[h^{+} + D_R'(\delta_0)\big]\,(\delta - \delta_0) + o(|\delta - \delta_0|). \tag{A.41}$$

For $\delta < \delta_0$, the one-sided expansion of Q_{N} gives

$$Q_{\mathrm{N}}(\delta) - Q_{\mathrm{N}}(\delta_0) \;=\; h^{-}(\delta_0 - \delta) + o(|\delta - \delta_0|). \tag{A.42}$$

Using

$$D_R'(\delta_0)(\delta - \delta_0) \;=\; -D_R'(\delta_0)(\delta_0 - \delta), \tag{A.43}$$

we obtain

$$Q_{\mathrm{X}}(\delta) - Q_{\mathrm{X}}(\delta_0) = \{Q_{\mathrm{N}}(\delta) - Q_{\mathrm{N}}(\delta_0)\} + \{G(\delta) - G(\delta_0)\} \tag{A.44}$$

$$= \left[h^{-} - D'_R(\delta_0)\right](\delta_0 - \delta) + o(|\delta - \delta_0|). \tag{A.45}$$

The local-minimizer and local-tilt statements follow directly from the signs of the two one-sided coefficients. □

A.6 *Lemma A.1*

Lemma A.1 (Uniform Consistency of Truncated CDFs). *Under Assumption 2,*

$$\sup_{\delta \in l_\delta} \sup_{z \le \delta} \left|\widehat{\mathrm{F}}_z^{(1)} - \mathrm{F}_z^{(1)}\right| = o_p(1), \qquad \sup_{\delta \in l_\delta} \sup_{z > \delta} \left|\widehat{\mathrm{F}}_z^{(2)} - \mathrm{F}_z^{(2)}\right| = o_p(1). \tag{A.46}$$

Proof. Let $p_\delta = \mathrm{F}_z(\delta) = \Pr(z_t \le \delta), \widehat{p}_\delta = \widehat{\mathrm{F}}_z(\delta)$, and define

$$\Delta \mathrm{F}_z = \widehat{\mathrm{F}}_z - \mathrm{F}_z, \qquad \Delta p_\delta = \widehat{p}_\delta - p_\delta. \tag{A.47}$$

By Assumption 2, Conditions 1 and 3,

$$\inf_{\delta \in l_\delta} p_\delta \ge \tau, \qquad \sup_{\delta \in l_\delta} |\Delta p_\delta| \le \sup_{z \in \mathbb{R}} \left|\widehat{\mathrm{F}}_z - \mathrm{F}_z\right| = \zeta_{\mathrm{T}} = o_p(1). \tag{A.48}$$

Hence, with probability approaching one,

$$\inf_{\delta \in l_\delta} \widehat{p}_\delta \ge \tau/2, \qquad \inf_{\delta \in l_\delta} (1 - \widehat{p}_\delta) \ge \tau/2. \tag{A.49}$$

For the lower regime,

$$\mathrm{F}_z^{(1)} = \frac{\mathrm{F}_z}{p_\delta}, \qquad \widehat{\mathrm{F}}_z^{(1)} = \frac{\widehat{\mathrm{F}}_z}{\widehat{p}_\delta}. \tag{A.50}$$

Therefore, uniformly over $\delta \in l_\delta$ and $z \le \delta$,

$$\begin{aligned} \left|\widehat{\mathrm{F}}_z^{(1)} - \mathrm{F}_z^{(1)}\right| &= \left|\frac{\widehat{\mathrm{F}}_z}{\widehat{p}_\delta} - \frac{\mathrm{F}_z}{p_\delta}\right| \\ &= \frac{|(\Delta \mathrm{F}_z) p_\delta - \mathrm{F}_z(\Delta p_\delta)|}{\widehat{p}_\delta p_\delta}. \end{aligned} \tag{A.51}$$

On the event above, the denominator is bounded away from zero, and since $0 \le \mathrm{F}_z \le 1$ and $0 \le p_\delta \le 1$,

$$\sup_{\delta \in l_\delta} \sup_{z \le \delta} \left|\widehat{\mathrm{F}}_z^{(1)} - \mathrm{F}_z^{(1)}\right| \le C_\tau \left(\sup_{z \in \mathbb{R}} |\Delta \mathrm{F}_z| + \sup_{\delta \in l_\delta} |\Delta p_\delta| \right) = o_p(1), \tag{A.52}$$

for a finite constant C_τ.

For the upper regime,

$$\mathrm{F}_z^{(2)} = \frac{\mathrm{F}_z - p_\delta}{1 - p_\delta}, \qquad \widehat{\mathrm{F}}_z^{(2)} = \frac{\widehat{\mathrm{F}}_z - \widehat{p}_\delta}{1 - \widehat{p}_\delta}. \tag{A.53}$$

Thus,

$$\widehat{\mathrm{F}}_z^{(2)} - \mathrm{F}_z^{(2)} = \frac{(\Delta \mathrm{F}_z)(1 - p_\delta) - (1 - \mathrm{F}_z)(\Delta p_\delta)}{(1 - \widehat{p}_\delta)(1 - p_\delta)}. \tag{A.54}$$

Again, the denominator is bounded away from zero with probability approaching one, while the numerator is bounded by

$$\sup_{z \in \mathbb{R}} |\Delta \mathrm{F}_z| + \sup_{\delta \in l_\delta} |\Delta p_\delta|. \tag{A.55}$$

Therefore,

$$\sup_{\delta \in l_\delta} \sup_{z > \delta} \left| \widehat{\mathrm{F}}_z^{(2)} - \mathrm{F}_z^{(2)} \right| = o_p(1). \tag{A.56}$$

□

A.7 *Lemma A.2*

Lemma A.2 (Uniform Accuracy of the Copula Control). *Let* $z_t^{(i)\star} = \Phi^{-1}\left(\kappa_\mathrm{T}\left(\mathrm{F}_{z_t}^{(\mathrm{i})}\right)\right)$ *and* $\widehat{z}_t^{(i)\star} = \Phi^{-1}\left(\kappa_\mathrm{T}\left(\widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})}\right)\right)$. *Under Assumption 2,*

$$\sup_{\delta \in l_\delta} \max_{t \in \mathcal{T}_i(\delta)} \left| \widehat{z}_t^{(i)\star} - z_t^{(i)\star} \right| = o_p(1), \qquad i = 1, 2. \tag{A.57}$$

PROOF. By Assumption 2, Condition 5, the winsorized PIT arguments lie in $[\epsilon_\mathrm{T}, 1 - \epsilon_\mathrm{T}]$. Hence, on this interval, $\Phi^{-1}(\bullet)$ is Lipschitz with constant

$$L_\mathrm{T} = \sup_{u \in [\epsilon_\mathrm{T}, 1 - \epsilon_\mathrm{T}]} \left| \frac{d}{du} \Phi^{-1}(u) \right|. \tag{A.58}$$

Moreover, the winsorization map $\kappa_\mathrm{T}(\bullet)$ is non-expansive, so that

$$|\kappa_\mathrm{T}(a) - \kappa_\mathrm{T}(b)| \leq |a - b| \tag{A.59}$$

for all $a, b \in [0, 1]$. Therefore, for each regime i,

$$\begin{aligned} \left| \widehat{z}_t^{(i)\star} - z_t^{(i)\star} \right| &\leq L_\mathrm{T} \left| \kappa_\mathrm{T}\left(\widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})}\right) - \kappa_\mathrm{T}\left(\mathrm{F}_{z_t}^{(\mathrm{i})}\right) \right| \\ &\leq L_\mathrm{T} \left| \widehat{\mathrm{F}}_{z_t}^{(\mathrm{i})} - \mathrm{F}_{z_t}^{(\mathrm{i})} \right|. \end{aligned} \tag{A.60}$$

Taking the supremum over $\delta \in l_\delta$ and the maximum over $t \in \mathcal{T}_i(\delta)$ gives

$$\sup_{\delta\in l_\delta} \max_{t\in\mathcal{T}_i(\delta)} \left|\widehat{z}_t^{(i)\star} - z_t^{(i)\star}\right| \leq L_{\mathrm{T}} \sup_{\delta\in l_\delta} \sup_{z\in \mathrm{Z}^{(\mathrm{i})}} \left|\widehat{\mathrm{F}}_z^{(\mathrm{i})} - \mathrm{F}_z^{(\mathrm{i})}\right|. \tag{A.61}$$

By Lemma A.1 and the proof of that lemma, the last supremum is $O_p(\zeta_{\mathrm{T}})$. Hence

$$\sup_{\delta\in l_\delta} \max_{t\in\mathcal{T}_i(\delta)} \left|\widehat{z}_t^{(i)\star} - z_t^{(i)\star}\right| = O_p(L_{\mathrm{T}}\zeta_{\mathrm{T}}) \ = \ o_p(1), \tag{A.62}$$

by Assumption 2, Condition 5. □

Lemma A.1 and Lemma A.2 establish the consistency properties of the estimated, winsorized copula controls. Appendix A.3 uses these results to prove feasible–oracle equivalence of the CLS criterion, both globally over the trimmed grid and locally on $1/\mathrm{T}$ neighborhoods of δ_0. The next lemma then handles a separate issue: even with the population CDF, the regime-specific PIT controls depend on the candidate threshold. It shows that this additional threshold dependence is negligible for the local objective increments underlying the Chan-type limit.

A.8 *Appendix A.3*

Lemma A.3 (Generated-regressor effect on the CLS objective).*Let*

$$Q_{\mathrm{T}}(\delta) = \mathrm{T}^{-1} RSS_{\mathrm{T}}(\delta), \qquad \widehat{Q}_{\mathrm{T}}(\delta) = \mathrm{T}^{-1}\widehat{RSS}_{\mathrm{T}}(\delta), \tag{A.63}$$

where $Q_{\mathrm{T}}(\delta)$ denotes the infeasible concentrated CLS criterion computed with the infeasible winsorized copula controls $z_t^{(\mathrm{i})\star}$, and $\widehat{Q}_{\mathrm{T}}(\delta)$ denotes the feasible criterion obtained by replacing $z_t^{(\mathrm{i})\star}$ with $\widehat{z}_t^{(\mathrm{i})\star}$. Under Assumption 2, the following hold.

(a) ***Global criterion closeness.***

$$\sup_{\delta\in l_\delta} \left|\widehat{Q}_{\mathrm{T}}(\delta) - Q_{\mathrm{T}}(\delta)\right| = o_p(1). \tag{A.64}$$

(b) ***Local increment closeness.*** *For every fixed $C < \infty$,*

$$\sup_{|x|\leq C} \left|\mathrm{T}\left[\left(\widehat{Q}_{\mathrm{T}}(\delta_0 + x/\mathrm{T}) - \widehat{Q}_{\mathrm{T}}(\delta_0)\right) - \left(Q_{\mathrm{T}}(\delta_0 + x/\mathrm{T}) - Q_{\mathrm{T}}(\delta_0)\right)\right]\right| = o_p(1). \tag{A.65}$$

PROOF. For a candidate threshold δ, let $\mathcal{T}_i(\delta)$ denote the regime-i index set. Let $X^{(\mathrm{i})}(\delta)$ be the infeasible regime-i design matrix constructed using the infeasible winsorized control $z_t^{(\mathrm{i})\star}$, and let $\widehat{X}^{(\mathrm{i})}(\delta)$ be the corresponding feasible design matrix obtained by replacing $z_t^{(\mathrm{i})\star}$ with $\widehat{z}_t^{(\mathrm{i})\star}$. The two design matrices differ only in the copula-control column. Let

$$P^{(\mathrm{i})}(\delta) \;=\; X^{(\mathrm{i})}(\delta)\big(X^{(\mathrm{i})}(\delta)'X^{(\mathrm{i})}(\delta)\big)^{-1}X^{(\mathrm{i})}(\delta)', \tag{A.66}$$

and define $\widehat{P}^{(\mathrm{i})}(\delta)$ analogously using $\widehat{X}^{(\mathrm{i})}(\delta)$. For $y^{(\mathrm{i})}(\delta) = (y_t)_{t\in\mathcal{T}_i(\delta)}$, the regime-specific concentrated residual sums of squares are

$$RSS_{\mathrm{T}}^{(\mathrm{i})}(\delta) \;=\; \|(I - P^{(\mathrm{i})}(\delta))y^{(\mathrm{i})}(\delta)\|^2, \qquad \widehat{RSS}_{\mathrm{T}}^{(\mathrm{i})}(\delta) = \|(I - \widehat{P}^{(\mathrm{i})}(\delta))y^{(\mathrm{i})}(\delta)\|^2. \tag{A.67}$$

Hence

$$\widehat{RSS}_{\mathrm{T}}(\delta) - RSS_{\mathrm{T}}(\delta) \;=\; \sum_{i=1}^{2} y^{(\mathrm{i})}(\delta)'\big(P^{(\mathrm{i})}(\delta) - \widehat{P}^{(\mathrm{i})}(\delta)\big)y^{(\mathrm{i})}(\delta). \tag{A.68}$$

By Lemma A.2,

$$d_{\mathrm{T}} \equiv \sup_{\delta\in l_\delta}\max_{i=1,2}\max_{t\in\mathcal{T}_i(\delta)}\left|\widehat{z}_t^{(\mathrm{i})\star} - z_t^{(\mathrm{i})\star}\right| \;=\; o_p(1). \tag{A.69}$$

The uniform nonsingularity of the regime-specific Gram matrices and the moment conditions in Assumption 2 imply that the projection operator is locally Lipschitz in the design matrix, uniformly over $\delta \in l_\delta$. Therefore,

$$\sup_{\delta\in l_\delta}\max_{i=1,2}\left\|\widehat{P}^{(\mathrm{i})}(\delta) - P^{(\mathrm{i})}(\delta)\right\|_{op} \;=\; O_p(d_{\mathrm{T}}) \;=\; o_p(1). \tag{A.70}$$

Also, by stationarity and the moment condition,

$$\sup_{\delta\in l_\delta}\max_{i=1,2}\frac{\|y^{(\mathrm{i})}(\delta)\|^2}{\mathrm{T}} \;=\; O_p(1). \tag{A.71}$$

It follows that

$$\sup_{\delta\in l_\delta}\left|\widehat{Q}_{\mathrm{T}}(\delta) - Q_{\mathrm{T}}(\delta)\right| \le \sup_{\delta\in l_\delta}\sum_{i=1}^{2}\frac{\|y^{(\mathrm{i})}(\delta)\|^2}{\mathrm{T}}\left\|\widehat{P}^{(\mathrm{i})}(\delta) - P^{(\mathrm{i})}(\delta)\right\|_{op} = o_p(1), \tag{A.72}$$

which proves part (a).

For part (b), define the RSS-scale generated-control error

$$H_{\mathrm{T}}(\delta) \;=\; \widehat{RSS}_{\mathrm{T}}(\delta) - RSS_{\mathrm{T}}(\delta). \tag{A.73}$$

We need to show that, uniformly over $|x| \le C$,

$$H_{\mathrm{T}}(\delta_0 + x/\mathrm{T}) - H_{\mathrm{T}}(\delta_0) = o_p(1). \tag{A.74}$$

The preceding projection-matrix expansion gives a first-order representation of $H_{\mathrm{T}}(\delta)$ in terms of the control-column perturbation $\widehat{z}_t^{(\mathrm{i})\star} - z_t^{(\mathrm{i})\star}$, with a remainder of smaller order uniformly over the local neighborhood. When δ moves from δ_0 to $\delta_0 + x/\mathrm{T}$, two changes occur. First, only observations satisfying

$$\min(\delta_0, \delta_0 + x/\mathrm{T}) < z_t \leq \max(\delta_0, \delta_0 + x/\mathrm{T}) \tag{A.75}$$

can switch regimes. By Assumption 2, Condition 4, the number of such observations is $O_p(1)$ uniformly over $|x| \leq C$. Their contribution to the difference $H_\mathrm{T}(\delta_0+x/\mathrm{T}) - H_\mathrm{T}(\delta_0)$ is therefore $O_p(1)d_\mathrm{T} = o_p(1)$.

Second, for observations that do not switch regimes, the local change in the regime-specific PIT transformation is induced only by the change in the candidate threshold from δ_0 to $\delta_0 + x/\mathrm{T}$. The truncated CDFs are smooth functions of p_δ, and

$$p_{\delta_0+x/\mathrm{T}} - p_{\delta_0} = O(1/\mathrm{T}) \tag{A.76}$$

uniformly over $|x| \leq C$ by the smoothness of F_z near δ_0. After winsorization, the Gaussian quantile transform is Lipschitz with constant L_T, and Assumption 2, Condition 5 gives the required uniform control of the feasible–oracle PIT error. Hence the non-switcher contribution to $H_\mathrm{T}(\delta_0 + x/\mathrm{T}) - H_\mathrm{T}(\delta_0)$ is also $o_p(1)$.

Combining the switcher and non-switcher contributions yields

$$\sup_{|x|\leq C} |H_\mathrm{T}(\delta_0 + x/\mathrm{T}) - H_\mathrm{T}(\delta_0)| = o_p(1), \tag{A.77}$$

which is equivalent to part (b), since $H_\mathrm{T}(\delta) = \mathrm{T}\{\widehat{Q}_\mathrm{T}(\delta) - Q_\mathrm{T}(\delta)\}$. □

A.9 *Appendix A.4*

Lemma A.4 (Local negligibility of threshold-dependent oracle controls). *Let $Q_\mathrm{T}(\delta)$ denote the infeasible CLS criterion computed with the oracle winsorized controls $z_t^{(\mathrm{i})\star}$, where the regime-specific PIT controls are recomputed at each candidate threshold δ. For $\delta_x = \delta_0 + x/\mathrm{T}$, define $\widetilde{Q}_\mathrm{T}(\delta_x)$ as the same infeasible CLS criterion, except that for observations whose regime assignment does not change between δ_0 and δ_x, the oracle controls are held fixed at their δ_0 values. For observations that switch regimes, the candidate-threshold controls are retained.*

Under Assumption 2, for every fixed $C < \infty$,

$$\sup_{|x|\leq C} \left| \mathrm{T}\left[\left(Q_\mathrm{T}(\delta_0 + x/\mathrm{T}) - Q_\mathrm{T}(\delta_0)\right) - \left(\widetilde{Q}_\mathrm{T}(\delta_0 + x/\mathrm{T}) - \widetilde{Q}_\mathrm{T}(\delta_0)\right)\right]\right| = o_p(1) \tag{A.78}$$

Proof. Let $C < \infty$ and write $\delta_x = \delta_0 + x/\mathrm{T}$, with $|x| \leq C$. The only difference between $Q_\mathrm{T}(\delta_x)$ and $\widetilde{Q}_\mathrm{T}(\delta_x)$ comes from observations whose regime assignment is unchanged between δ_0 and δ_x but whose oracle regime-specific PIT controls are recomputed at δ_x rather than held fixed at δ_0.

Consider first a non-switching observation in regime i. By Assumption 2, Conditions 1 and 2, the regime probabilities are bounded away from zero and one, and F_z is differentiable with bounded density on a neighborhood of l_δ. Hence the truncated

CDF transforms satisfy, uniformly over $|x| \le C$ and over non-switching observations,

$$\left| \mathrm{F}_{z_t}^{(\mathrm{i})}(\delta_x) - \mathrm{F}_{z_t}^{(\mathrm{i})}(\delta_0) \right| = O\left(\frac{1}{\mathrm{T}} \right). \tag{A.79}$$

The winsorization map κ_{T} is non-expansive, and $\Phi^{-1}(\bullet)$ is Lipschitz on $[\epsilon_{\mathrm{T}}, 1 - \epsilon_{\mathrm{T}}]$ with constant L_{T}. Therefore,

$$\sup_{|x| \le C} \max_{\text{non-switchers}} \left| z_t^{(\mathrm{i})\star}(\delta_x) - z_t^{(\mathrm{i})\star}(\delta_0) \right| = O\left(\frac{L_{\mathrm{T}}}{\mathrm{T}} \right). \tag{A.80}$$

Let $\Delta Z_i(x)$ denote the vector collecting the differences in (A.80) for the non-switching observations in regime i. Then

$$\sup_{|x| \le C} \|\Delta Z_i(x)\|^2 = O_p\left(\mathrm{T} \frac{L_{\mathrm{T}}^2}{\mathrm{T}^2} \right) = O_p\left(\frac{L_{\mathrm{T}}^2}{\mathrm{T}} \right) = o_p(1), \tag{A.81}$$

where the last equality follows from the winsorization-rate condition in Assumption 2.

Because the CLS criterion is concentrated over the regression coefficients, the first-order effect of a small perturbation in an included regressor column is absorbed by least-squares re-estimation. Equivalently, by standard projection-matrix or Frisch–Waugh–Lovell algebra, the resulting change in the concentrated residual sum of squares is of quadratic order in the perturbation of the included control column, uniformly over the local neighborhood, provided the regime-specific Gram matrices are uniformly nonsingular. Hence,

$$\sup_{|x| \le C} \left| RSS_{\mathrm{T}}(\delta_x) - \widetilde{RSS}_{\mathrm{T}}(\delta_x) \right| = o_p(1). \tag{A.82}$$

At δ_0, the two criteria coincide by construction,

$$Q_{\mathrm{T}}(\delta_0) = \widetilde{Q}_{\mathrm{T}}(\delta_0). \tag{A.83}$$

Moreover, observations that switch regimes between δ_0 and δ_x are treated identically in $Q_{\mathrm{T}}(\delta_x)$ and $\widetilde{Q}_{\mathrm{T}}(\delta_x)$. Condition 4 of Assumption 2 ensures that the number of such switchers is $O_p(1)$ uniformly over $|x| \le C$, so they are part of the usual local threshold objective and do not contribute to the difference between the two criteria.

Since $Q_{\mathrm{T}}(\delta) = \mathrm{T}^{-1} RSS_{\mathrm{T}}(\delta)$, the preceding RSS-scale bound gives

$$\sup_{|x| \le C} \left| \mathrm{T} \left[\left(Q_{\mathrm{T}}(\delta_x) - Q_{\mathrm{T}}(\delta_0) \right) - \left(\widetilde{Q}_{\mathrm{T}}(\delta_x) - \widetilde{Q}_{\mathrm{T}}(\delta_0) \right) \right] \right| = o_p(1). \tag{A.84}$$

This proves the lemma. □

A.10 *Proof of Proposition 4*

PROOF. Consider the infeasible concentrated CLS criterion $Q_{\mathrm{T}}(\delta)$ and its feasible counterpart $\widehat{Q}_{\mathrm{T}}(\delta)$, as defined in Appendix A.3. Let $Q_{\mathrm{N}}(\delta)$ denote the population counterpart of the infeasible criterion. Under Condition (i) of Proposition 4, the process is stationary and ergodic with sufficient moments. Together with the compact trimmed grid in Assumption 2, this gives the standard uniform law of large numbers for threshold regressions:

$$\sup_{\delta \in l_\delta} |Q_{\mathrm{T}}(\delta) - Q_{\mathrm{N}}(\delta)| \;=\; o_p(1). \tag{A.85}$$

Condition (iii), together with the orthogonality implied by Assumption 1 and Proposition 1,

$$\mathbb{E}\Big(\varepsilon_t^{(\mathrm{i})} \mid \mathrm{Y}_{t-1}, z_t^{(\mathrm{i})}\Big) \;=\; 0_{\mathrm{K}}, \tag{A.86}$$

and the positive density of z_t around δ_0, implies that assigning observations to the wrong regime raises the population CLS objective. Hence $Q_{\mathrm{N}}(\delta)$ is uniquely minimized at δ_0.

It follows from the argmin consistency theorem that the infeasible threshold estimator

$$\delta_{\mathrm{T}}^{\star} \;=\; \arg\min_{\delta \in l_\delta} Q_{\mathrm{T}}(\delta) \tag{A.87}$$

satisfies

$$\delta_{\mathrm{T}}^{\star} \xrightarrow{p} \delta_0. \tag{A.88}$$

It remains to transfer this result to the feasible estimator. By Appendix A.3 part (a),

$$\sup_{\delta \in l_\delta} \left|\widehat{Q}_{\mathrm{T}}(\delta) - Q_{\mathrm{T}}(\delta)\right| \;=\; o_p(1). \tag{A.89}$$

Therefore,

$$\sup_{\delta \in l_\delta} \left|\widehat{Q}_{\mathrm{T}}(\delta) - Q_{\mathrm{N}}(\delta)\right| \;=\; o_p(1). \tag{A.90}$$

Since $Q_{\mathrm{N}}(\delta)$ is uniquely minimized at δ_0 on the compact trimmed grid l_δ, another application of the argmin consistency theorem yields

$$\widehat{\delta} \;=\; \arg\min_{\delta \in l_\delta} \widehat{Q}_{\mathrm{T}}(\delta) \xrightarrow{p} \delta_0. \tag{A.91}$$

This proves consistency of the feasible threshold estimator. □

A.11 *Proof of **Remark 1***

PROOF. Define the feasible local objective process

$$\widehat{\mathcal{M}}_{\mathrm{T}}(x) \;=\; \mathrm{T}\left[\widehat{Q}_{\mathrm{T}}\Big(\delta_0 + \frac{x}{\mathrm{T}}\Big) - \widehat{Q}_{\mathrm{T}}(\delta_0)\right], \qquad x \in \mathbb{R}, \tag{A.92}$$

and the corresponding infeasible local process

$$\mathcal{M}_{\mathrm{T}}(x) \;=\; \mathrm{T}\left[Q_{\mathrm{T}}\Big(\delta_0 + \frac{x}{\mathrm{T}}\Big) - Q_{\mathrm{T}}(\delta_0)\right]. \tag{A.93}$$

By Appendix A.3 part (b), for every fixed $C < \infty$,

$$\sup_{|x|\le C} \left| \widehat{\mathcal{M}}_{\mathrm{T}}(x) - \mathcal{M}_{\mathrm{T}}(x) \right| \;=\; o_p(1). \tag{A.94}$$

Next, let $\widetilde{\mathcal{M}}_{\mathrm{T}}(x)$ denote the oracle local objective process defined from $\widetilde{Q}_{\mathrm{T}}$ in Appendix A.4, namely

$$\widetilde{\mathcal{M}}_{\mathrm{T}}(x) \;=\; \mathrm{T}\left[\widetilde{Q}_{\mathrm{T}}\left(\delta_0 + \frac{x}{\mathrm{T}}\right) - \widetilde{Q}_{\mathrm{T}}(\delta_0)\right]. \tag{A.95}$$

By Appendix A.4, again for every fixed $C < \infty$,

$$\sup_{|x|\le C} \left| \mathcal{M}_{\mathrm{T}}(x) - \widetilde{\mathcal{M}}_{\mathrm{T}}(x) \right| = o_p(1). \tag{A.96}$$

Therefore,

$$\sup_{|x|\le C} \left| \widehat{\mathcal{M}}_{\mathrm{T}}(x) - \widetilde{\mathcal{M}}_{\mathrm{T}}(x) \right| \;=\; o_p(1). \tag{A.97}$$

The process $\widetilde{\mathcal{M}}_{\mathrm{T}}(x)$ is the standard local threshold objective: over $1/\mathrm{T}$ neighborhoods of δ_0, only the observations that switch regimes contribute to the first-order local criterion, while the remaining regressors and oracle controls are held fixed to first order. Under the stationarity, moment, positive-density, and distinct-regime conditions in Proposition 4, the standard threshold limit theory of Chan (1993) yields

$$\widetilde{\mathcal{M}}_{\mathrm{T}}(\bullet) \Rightarrow \Upsilon(\bullet) \tag{A.98}$$

on compact subsets of $\mathbb{R}$, where Υ is the compound-Poisson limit process.

Because the feasible and oracle local processes are uniformly $o_p(1)$-close on compact sets, the argmin continuous mapping theorem gives

$$\mathrm{T}(\widehat{\delta} - \delta_0) \;=\; \arg\min_{x\in\mathbb{R}} \widehat{\mathcal{M}}_{\mathrm{T}}(x) \Rightarrow \arg\min_{x\in\mathbb{R}} \Upsilon(x). \tag{A.99}$$

This proves the stated limit. □

A.12 *Proof of Lemma 3*

PROOF. We proceed in two steps. First, we establish the result for the threshold parameter δ, and then for the regime-specific variances. The two cases must be treated separately because the correctly specified threshold estimator is T-consistent, whereas the correctly specified variance estimators are only $\sqrt{\mathrm{T}}$-consistent.

Part 1: Threshold parameter. Let

$$\mathrm{RE}_{\delta,\mathrm{T}} \;=\; \frac{\mathrm{MSE}_{\delta}(\mathrm{TVAR}_{\mathrm{X}})}{\mathrm{MSE}_{\delta}(\mathrm{TVAR}_{\mathrm{N}})}. \tag{A.100}$$

By assumption, under correct specification the $\mathrm{TVAR_N}$ threshold estimator is T-consistent and asymptotically unbiased. Hence there exists a finite constant $V_\delta > 0$ such that

$$\mathrm{T}(\widehat{\delta}_N - \delta_0) \xrightarrow{d} Z_\delta, \qquad \mathbb{E}(Z_\delta) = 0, \qquad \mathbb{E}(Z_\delta^2) = V_\delta, \tag{A.101}$$

which implies

$$\mathbb{E}\left[(\widehat{\delta}_N - \delta_0)^2\right] = \frac{V_\delta}{\mathrm{T}^2} + o\left(\frac{1}{\mathrm{T}^2}\right). \tag{A.102}$$

Therefore,

$$\mathrm{MSE}_\delta(\mathrm{TVAR_N}) = \frac{V_\delta}{\mathrm{T}^2} + o\left(\frac{1}{\mathrm{T}^2}\right). \tag{A.103}$$

Under misspecification, the $\mathrm{TVAR_X}$ threshold estimator converges to a pseudo-true value $\delta_X^\star$ where $\delta_X^\star \neq \delta_0$. Define the asymptotic bias

$$b_\delta = \delta_X^\star - \delta_0 \neq 0. \tag{A.104}$$

Then

$$\widehat{\delta}_X - \delta_0 = b_\delta + o_p(1), \tag{A.105}$$

so that

$$\mathbb{E}\left[(\widehat{\delta}_X - \delta_0)^2\right] = b_\delta^2 + o(1). \tag{A.106}$$

Hence

$$\mathrm{MSE}_\delta(\mathrm{TVAR_X}) = b_\delta^2 + o(1). \tag{A.107}$$

Taking the ratio gives

$$\mathrm{RE}_{\delta,\mathrm{T}} = \left(\frac{b_\delta^2 + o(1)}{V_\delta/\mathrm{T}^2 + o(\mathrm{T}^{-2})}\right) = \frac{b_\delta^2}{V_\delta}\mathrm{T}^2 + O(1). \tag{A.108}$$

Thus

$$\mathrm{RE}_{\delta,\mathrm{T}} = C_\delta \mathrm{T}^2 + O(1), \qquad C_\delta = \frac{b_\delta^2}{V_\delta} > 0, \tag{A.109}$$

and therefore

$$\mathrm{RE}_{\delta,\mathrm{T}} \to \infty \qquad \text{at rate } \mathrm{T}^2. \tag{A.110}$$

Part 2: Regime-specific variances. Now define

$$\mathrm{RE}_{\Sigma,\mathrm{T}} = \frac{\mathrm{MSE}_{\Sigma^{(\mathrm{i})}}(\mathrm{TVAR_X})}{\mathrm{MSE}_{\Sigma^{(\mathrm{i})}}(\mathrm{TVAR_N})}. \tag{A.111}$$

Under correct specification, the regime-specific variance estimators are $\sqrt{\mathrm{T}}$-consistent and asymptotically unbiased. Hence, for each regime i, there exists a finite constant

$V_\Sigma^{(\mathrm{i})} > 0$ such that

$$\sqrt{\mathrm{T}}\left(\widehat{\Sigma}^{(\mathrm{i})}_{\varepsilon\varepsilon,N} - \Sigma^{(\mathrm{i})}_{\varepsilon\varepsilon}\right) \xrightarrow{d} Z_\Sigma^{(\mathrm{i})}, \qquad \mathbb{E}\left[Z_\Sigma^{(\mathrm{i})}\right] = 0, \qquad \mathbb{E}\left[(Z_\Sigma^{(\mathrm{i})})^2\right] = V_\Sigma^{(\mathrm{i})}, \tag{A.112}$$

where, for notational simplicity, the MSE is understood componentwise or for a scalar summary of the variance object. It follows that

$$\mathrm{MSE}_{\Sigma^{(\mathrm{i})}}(\mathrm{TVAR_N}) = \frac{V_\Sigma^{(\mathrm{i})}}{\mathrm{T}} + o\left(\frac{1}{\mathrm{T}}\right). \tag{A.113}$$

Under misspecification, $\mathrm{TVAR_X}$ ignores the copula control and the corresponding variance estimator converges to

$$\Sigma^{(\mathrm{i})}_{\mathrm{uu}} = \Sigma^{(\mathrm{i})}_{\varepsilon\varepsilon} + \Lambda^{(\mathrm{i})}\Lambda^{(i)\prime} \neq \Sigma^{(\mathrm{i})}_{\varepsilon\varepsilon}, \tag{A.114}$$

as stated in the main text. Therefore the estimator is asymptotically biased relative to the target $\Sigma^{(\mathrm{i})}_{\varepsilon\varepsilon}$. Let

$$b_\Sigma^{(\mathrm{i})} \equiv \Sigma^{(\mathrm{i})}_{\mathrm{uu}} - \Sigma^{(\mathrm{i})}_{\varepsilon\varepsilon} = \Lambda^{(\mathrm{i})}\Lambda^{(i)\prime} \neq 0. \tag{A.115}$$

Then

$$\widehat{\Sigma}^{(\mathrm{i})}_{uu,X} - \Sigma^{(\mathrm{i})}_{\varepsilon\varepsilon} = b_\Sigma^{(\mathrm{i})} + o_p(1), \tag{A.116}$$

and hence

$$\mathrm{MSE}_{\Sigma^{(\mathrm{i})}}(\mathrm{TVAR_X}) = \left(b_\Sigma^{(\mathrm{i})}\right)^2 + o(1), \tag{A.117}$$

again interpreted componentwise or through the scalar variance summary used in the lemma.

Therefore,

$$\mathrm{RE}_{\Sigma,\mathrm{T}} = \left(\frac{\left(b_\Sigma^{(\mathrm{i})}\right)^2 + o(1)}{V_\Sigma^{(\mathrm{i})}/\mathrm{T} + o(\mathrm{T}^{-1})}\right) = \frac{\left(b_\Sigma^{(\mathrm{i})}\right)^2}{V_\Sigma^{(\mathrm{i})}}\mathrm{T} + O(1). \tag{A.118}$$

Thus

$$\mathrm{RE}_{\Sigma,\mathrm{T}} = C_\Sigma \mathrm{T} + O(1), \qquad C_\Sigma = \frac{\left(b_\Sigma^{(\mathrm{i})}\right)^2}{V_\Sigma^{(\mathrm{i})}} > 0, \tag{A.119}$$

and therefore $\mathrm{RE}_{\Sigma,\mathrm{T}} \to \infty$ at rate T.

Since the misspecified $\mathrm{TVAR_X}$ estimator has a nonvanishing asymptotic bias in both cases, its MSE converges to a strictly positive constant, whereas the correctly specified $\mathrm{TVAR_N}$ MSE shrinks to zero at rate T^{-2} for the threshold and at rate T^{-1} for the regime-specific variances. This proves the lemma. □

Appendix B: Beyond Elliptical Dependence

We now provide details on how the Gaussian-copula control can be extended to non-elliptical dependence structures. Under the Gaussian benchmark, the conditional mean of the reduced-form innovations is linear in the pseudo-normal score. This linearity need not hold under more general copulas. We therefore treat the non-elliptical case as an approximation exercise: a first-order expansion of the copula density around independence motivates a nonlinear control function, which can be approximated using sieve terms in the pseudo-normal score. The appendix is organized as follows. We first present the local copula-density expansion, then derive the implied conditional-mean representation, and finally describe its implementation using Hermite polynomial controls.

B.1 *Local Copula-Density Expansion*

Let $U_t^{(i)} = F_{u_t}^{(i)} \text{ß} \mathcal{N}(0,1)^{\mathrm{K}}$ and $V_t^{(i)} = F_{z_t}^{(i)} \text{ß} \mathcal{N}(0,1)$ denote the regime-specific probability integral transforms of the reduced-form innovation vector and the threshold variable. Consider a copula density $c_\theta^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})$, where $\theta^{(i)} \in \mathbb{R}^m$ indexes the dependence between $u_t^{(i)}$ and z_t. We normalize independence to $\theta^{(i)} = 0$, so that $c_0^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}) = 1$. Assume that $c_\theta^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})$ is continuously differentiable in $\theta^{(i)}$ in a neighborhood of zero. A first-order Taylor expansion around independence gives

$$c_\theta^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}) = 1 + \xi^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})'\theta^{(i)} + o(\|\theta^{(i)}\|). \tag{B.1}$$

where

$$\xi^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}) = \left(\xi_1^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}), \ldots, \xi_m^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})\right)', \tag{B.2}$$

with

$$\xi_j^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}) = \frac{\partial}{\partial \theta_j^{(i)}} c_\theta^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}; \theta^{(i)}) \Bigg|_{\theta^{(i)}=0}, \qquad j = 1, \ldots, m. \tag{B.3}$$

The vector $\xi^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})$ is the copula-density score at independence. Because $c_\theta^{(i)}$ is a copula density for each $\theta^{(i)}$, the score satisfies the marginal normalization restrictions

$$\int \xi^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})\, dF_{u_t}^{(i)} = 0 \quad \text{for all } F_{z_t}^{(i)}, \qquad \int \xi^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)})\, dF_{z_t}^{(i)} = 0 \quad \text{for all } F_{u_t}^{(i)}.$$

These restrictions ensure that the expansion changes the dependence structure without changing the marginal distributions.

B.2 *Conditional-Mean Representation*

From the conditional-density representation in (4), we have

$$f_{u_t|z_t}^{(i)} = c_\theta^{(i)}(F_{u_t}^{(i)}, F_{z_t}^{(i)}) f_{u_t}^{(i)}.$$

Substituting (B.1) gives

$$f_{u_t|z_t}^{(\mathrm{i})} = \left[1 + \xi^{(\mathrm{i})}(F_{u_t}^{(\mathrm{i})}, F_{z_t}^{(\mathrm{i})})'\theta^{(\mathrm{i})} + o(\|\theta^{(\mathrm{i})}\|)\right] f_{u_t}^{(\mathrm{i})}.$$

Taking expectations and using $\mathbb{E}(u_t^{(\mathrm{i})}) = 0_{\mathrm{K}}$ yields

$$\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) = B^{(\mathrm{i})}(z_t^{(\mathrm{i})})\theta^{(\mathrm{i})} + o(\|\theta^{(\mathrm{i})}\|), \tag{B.4}$$

where

$$B^{(\mathrm{i})}(z_t^{(\mathrm{i})}) = \int u\,\xi^{(\mathrm{i})}(F_{u_t}^{(\mathrm{i})}, F_{z_t}^{(\mathrm{i})})' f_{u_t}^{(\mathrm{i})}\,du. \tag{B.5}$$

In (B.5), u denotes the integration argument, while $F_{u_t}^{(\mathrm{i})}$ and $f_{u_t}^{(\mathrm{i})}$ follow the notation of the main text and denote the regime-specific marginal distribution and density evaluated at that argument. Thus, away from the Gaussian benchmark, the conditional mean correction is generally a nonlinear function of the threshold variable. The Gaussian-copula case corresponds to the special case in which this function is linear in the pseudo-normal threshold score.

B.3 *Sieve Approximation in the Gaussian PIT Score*

Let $z_t^{(\mathrm{i})\star} = \Phi^{-1}(F_{z_t}^{(\mathrm{i})})$ denote the regime-specific Gaussian PIT score used in the baseline model. Since $B^{(\mathrm{i})}(z_t^{(\mathrm{i})})$ is an unknown function of the threshold variable, we approximate the conditional mean in (B.4) using a sieve basis in $z_t^{(\mathrm{i})\star}$:

$$\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) \approx \sum_{j=1}^{d_{\mathrm{T}}} \beta_j^{(\mathrm{i})} s_j(z_t^{(\mathrm{i})\star}), \tag{B.6}$$

where $\{s_j(\cdot)\}_{j=1}^{d_{\mathrm{T}}}$ is a centered and scaled basis. Substituting (B.6) into the regime-specific TVAR gives

$$y_t = A^{(\mathrm{i})}Y_{t-1} + \sum_{j=1}^{d_{\mathrm{T}}} \beta_j^{(\mathrm{i})} s_j(z_t^{(\mathrm{i})\star}) + \widetilde{\varepsilon}_t^{(\mathrm{i})}. \tag{B.7}$$

The error term $\widetilde{\varepsilon}_t^{(\mathrm{i})}$ contains the mean-zero innovation $u_t^{(\mathrm{i})} - \mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})})$, the finite-sieve approximation error, and the higher-order remainder from the local copula expansion.

Equation (B.7) is therefore not an exact analogue of the Gaussian-copula representation in Proposition 1. Rather, it is a flexible control-function approximation designed to absorb nonlinear threshold-selection effects when the true copula is non-elliptical.

B.4 *Hermite Polynomial Implementation*

A convenient choice of basis is given by orthonormal Hermite polynomials evaluated at the Gaussian PIT score. The first few basis functions are

$$\mathcal{H}_1(x) = x, \qquad \mathcal{H}_2(x) = \frac{x^2 - 1}{\sqrt{2}}, \qquad \mathcal{H}_3(x) = \frac{x^3 - 3x}{\sqrt{6}},$$

and

$$\mathcal{H}_4(x) = \frac{x^4 - 6x^2 + 3}{\sqrt{24}}.$$

With this basis, the sieve-augmented TVAR becomes

$$y_t = A^{(\mathrm{i})} Y_{t-1} + \sum_{j=1}^{d_{\mathrm{T}}} \beta_j^{(\mathrm{i})} \mathcal{H}_j(z_t^{(\mathrm{i})\star}) + \tilde{\varepsilon}_t^{(\mathrm{i})}. \tag{B.8}$$

The baseline Gaussian-copula control corresponds to the first-order term $\mathcal{H}_1(z_t^{(\mathrm{i})\star}) = z_t^{(\mathrm{i})\star}$. Under the Gaussian benchmark, no higher-order Hermite terms are needed in population. Under non-elliptical dependence, however, the conditional mean may contain nonlinear components, and the higher-order Hermite terms provide a parsimonious way to approximate them.

APPENDIX C: **MONTE CARLO EXERCISES IN FULL**

We now present additional details on the results of our Monte Carlo experiments. We consider the following simulation scenarios: (i) common correlations across regimes (Scenario 1) versus regime-specific correlations (Scenario 2), for both Gaussian and uniform z_t; (ii) a non-normal threshold variable (specifically, $z_t \sim U(-4,3)$); (iii) non-Gaussian VAR innovations (Scenario 3); and (iv) non-elliptical dependence between z_t and the innovations (Scenario 4).

The following tables report Monte Carlo results, that is to say, the average *BIAS* and *MSE* over 1,000 replications (with trimming at 10% when estimating δ) for the threshold estimate $\widehat{\delta}$ and regime-specific innovation variance-covariance parameters $\{(\sigma_{u_1}^{(\mathrm{i})})^2, (\sigma_{u_2}^{(\mathrm{i})})^2, \sigma_{u_1 u_2}^{(\mathrm{i})}\}$, $i \in \{1,2\}$. Results are shown for the following estimators: (i) the misspecified TVAR that ignores threshold endogeneity ($\mathrm{TVAR_X}$), (ii) the copula control specification ($\mathrm{TVAR_N}$), and (iii) for the final scenario $\mathrm{TVAR_N}$ with Section 5 Hermite term approximation.

- Table C.1: Scenarios 1. and 2. for $\mathrm{T} = 200$.
- Table C.2: Scenarios 1. and 2. for $\mathrm{T} = 500$.
- Table C.3: Scenarios 3. for both $\mathrm{T} = 200$ and $\mathrm{T} = 500$.
- Table C.4: Scenarios 4. for $\mathrm{T} = 200$.
- Table C.5: Scenarios 4. for $\mathrm{T} = 500$.

Overall Summary

Across all scenarios, neglecting threshold endogeneity ($\mathrm{TVAR_X}$) produces biased estimates and elevated MSE for the threshold δ and the regime-wise innovation second moments. These distortions are most pronounced when (i) the correlation structure between z_t and the shocks differs across regimes (Scenario 2), (ii) the threshold variable is non-normal (here $z_t \sim U(-4,3)$), (iii) the VAR innovations depart from normality (Scenario 3), or (iv) the dependence between z_t and innovations is non-elliptical (Scenario 4). In contrast, the copula control specification ($\mathrm{TVAR_N}$) substantially reduces both bias and MSE for δ and typically also improves estimation of the innovation variance-covariance parameters, even in relatively small samples ($\mathrm{T} = 200$).

Scenarios 1-2 (linear dependence)

Table C.1 ($\mathrm{T} = 200$) and Table C.2 ($\mathrm{T} = 500$) consider cases in which dependence between z_t and structural TVAR shocks is linear, either with common correlations across regimes (Scenario 1) or with correlations that differ across regimes (Scenario 2), and for both Gaussian and uniform z_t. Two regularities stand out. First, the uniform threshold designs are substantially more challenging for $\mathrm{TVAR_X}$: bias and MSE in $\widehat{\delta}$ and in several regime-specific second moments remain large, particularly when correlations differ across regimes. Second, increasing the sample size from $\mathrm{T} = 200$ to $\mathrm{T} = 500$ reduces sampling variability but does not remove misspecification bias in $\mathrm{TVAR_X}$ in the more demanding designs; by contrast, $\mathrm{TVAR_N}$ delivers tight performance already at $\mathrm{T} = 200$ and remains accurate at $\mathrm{T} = 500$.

A concrete illustration is Scenario 2 with $z_t \sim U(-4, 3)$: at $\mathrm{T} = 500$ (Table C.2, Panel B), $\mathrm{TVAR_X}$ exhibits a large threshold bias (about $|\widehat{\delta}_X - \delta_0| \approx 2.6$) and correspondingly large MSE, while $\mathrm{TVAR_N}$ drives the threshold bias close to zero and sharply reduces MSE. This ranking carries over to several innovation second moments, where $\mathrm{TVAR_N}$ typically dominates $\mathrm{TVAR_X}$ in both bias and MSE.

Scenario 3 (non-normal innovations)

Table C.3 reports results when the VAR innovations are Student-t or skewed-t (with $z_t \sim U(-4, 3)$) for both $\mathrm{T} = 200$ and $\mathrm{T} = 500$. Under heavy tails, $\mathrm{TVAR_X}$ continues to display non-trivial bias in $\widehat{\delta}$ and in innovation second moments, while $\mathrm{TVAR_N}$ reduces threshold bias toward zero and typically lowers MSE. These results indicate that the Gaussian copula control provides a satisfactory approximation even when the underlying innovations are non-Gaussian. The gains are generally larger under symmetric heavy tails (Student-t) than under skewness (skewed-t), consistent with skewness exacerbating small-sample regime-wise moment estimation.

Scenario 4 (Nonlinear Dependence)

Table C.4 and Table C.5 examine cases in which the dependence between the threshold variable and the innovations is non-elliptical and generated by Archimedean copulas (Clayton, Joe, and Gumbel). In both sample sizes, the exogenous-threshold specification $\mathrm{TVAR_X}$ exhibits sizable bias and MSE for the threshold parameter and for several regime-specific second moments, reflecting the failure to account for nonlinear threshold–innovation dependence. Introducing a copula-based control function substantially improves performance.

Consistent with earlier findings, the Gaussian copula control provides a satisfactory approximation even under non-elliptical dependence. In particular, even in the smaller sample ($\mathrm{T} = 200$), the linear control $\mathcal{H}_1\left(z_t^{(i)\star}\right) = z_t^{(i)\star}$ removes most of the threshold bias across copulas, while the augmented specification including both linear and quadratic terms, $\mathcal{H}_2\left(z_t^{(i)\star}\right)$, delivers additional—though generally modest—efficiency gains. These gains become more pronounced at $\mathrm{T} = 500$, where the augmented $\mathrm{TVAR_N}$ specification typically attains the lowest MSE for δ, indicating that the quadratic term helps capture residual nonlinear dependence.

The improvements are particularly pronounced for the Clayton and Gumbel copulas. In these cases, both the baseline $\mathrm{TVAR_N}$ model and its extension with $\mathcal{H}_2\left(z_t^{(i)\star}\right)$

reduce threshold bias essentially to zero and markedly lower MSE at both sample sizes, with the extended specification providing a systematic refinement in larger samples. The gains are comparatively more modest for the Joe copula. Although the inclusion of $\mathcal{H}_1$ and $\mathcal{H}_2$ controls effectively for threshold endogeneity, reductions in the bias and MSE of some elements of $\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$—particularly in regime 2—are more limited in smaller samples.[1]

Overall, the Monte Carlo results indicate that relatively simple semiparametric copula-based control functions are effective in controlling for threshold endogeneity across a broad range of nonlinear dependence structures. Their gains are evident in small samples and become more pronounced as the sample size increases, particularly in the presence of strong or asymmetric nonlinear dependence.

Mapping Results and Theory

The Monte Carlo metrics in the tables directly rationalize the dynamic distortions documented in Section 6.2. In the design used for Figure 1 (the $\mathrm{T} = 500$, $z_t \sim U(-4, 3)$, regime-varying correlation case corresponding to Table C.2, Panel B), $\mathrm{TVAR_X}$ exhibits a large threshold error (about $|\widehat{\delta}_X - \delta_0| \approx 2.6$), whereas $\mathrm{TVAR_N}$ yields $|\widehat{\delta}_N - \delta_0| \approx 0$. This threshold misclassification is consequential for impulse responses because it changes (i) which regime-specific dynamics are applied along a simulated path and (ii) the regime-wise reduced-form innovation covariance that feeds into structural identification. Empirically in that figure, these channels manifest as (a) excessive impact responses at $h = 0$ under $\mathrm{TVAR_X}$ ("impact overshoot") and (b) a pronounced shape distortion that persists for roughly 10-15 horizons before dissipating, whereas $\mathrm{TVAR_N}$ tracks the true DGP IRFs closely.

The link between these dynamic errors and the table evidence is mechanical: the large δ bias in $\mathrm{TVAR_X}$ implies systematic regime mis-assignment, while the inflated/biased regime-wise second moments reported across the tables imply distorted impact matrices and (through biased reduced-form coefficients) distorted propagation. By removing the endogenous component of $\mathbb{E}(u_t^{(\mathrm{i})} \mid z_t)$ via the copula control, $\mathrm{TVAR_N}$ restores near-orthogonality between the threshold and innovations, which in turn yields accurate regime classification, tighter innovation covariance estimates, and IRFs that align with the true DGP in both impact and medium-horizon shape.

[1] This behavior likely reflects the strong upper-tail dependence of the Joe copula, which interacts with a relatively small number of observations in regime 2. In such settings, finite-sample noise limits the ability of low-order polynomial controls to fully capture the nonlinear dependence structure, although precision improves with sample size. Higher-order terms $\mathcal{H}_j(\bullet)$ or larger samples may further enhance performance.

TABLE C.1. Monte Carlo Results, $\mathrm{T}=200$, Scenarios 1. and 2.

	$z_t \sim \mathcal{U}(-4,3)$							$z_t \sim \mathcal{N}(0,4.08)$						
	δ	$(\sigma^{(1)}_{u_1})^2$	$(\sigma^{(1)}_{u_2})^2$	$\sigma^{(1)}_{u_1u_2}$	$(\sigma^{(2)}_{u_1})^2$	$(\sigma^{(2)}_{u_2})^2$	$\sigma^{(2)}_{u_1u_2}$	δ	$(\sigma^{(1)}_{u_1})^2$	$(\sigma^{(1)}_{u_2})^2$	$\sigma^{(1)}_{u_1u_2}$	$(\sigma^{(2)}_{u_1})^2$	$(\sigma^{(2)}_{u_2})^2$	$\sigma^{(2)}_{u_1u_2}$
Panel A: $corr(e^{(\mathrm{i})}_j z)=0.8$ for $j=1,2$ equations and $i=1,2$ regimes														
$\mathrm{TVAR_X}$														
BIAS	-0.243	0.316	0.188	0.000	0.141	0.061	-0.060	-1.064	1.184	1.159	0.107	0.591	0.575	-0.056
MSE	0.398	0.101	0.065	0.010	0.033	0.063	0.053	1.644	1.619	1.633	0.011	0.465	0.515	0.003
$\mathrm{TVAR_N}$														
BIAS	0.024	0.008	-0.013	0.000	0.060	-0.018	-0.009	-0.015	0.067	-0.054	0.013	0.034	-0.000	0.006
MSE	0.049	0.002	0.021	0.010	0.012	0.058	0.054	0.025	0.019	0.042	0.000	0.050	0.119	0.000
Panel B: $corr(e^{(1)}_j z)=0.8$ and $corr(e^{(2)}_j z)=-0.7$ for $j=1,2$ equations														
$\mathrm{TVAR_X}$														
BIAS	-1.524	0.821	0.825	0.097	0.193	0.163	-0.023	2.616	0.394	0.388	0.058	0.523	0.506	-0.144
MSE	8.900	0.838	0.885	0.016	0.107	0.157	0.022	7.105	0.226	0.271	0.015	0.323	0.343	0.031
$\mathrm{TVAR_N}$														
BIAS	0.000	0.028	0.029	0.015	0.008	-0.039	0.005	0.000	0.076	0.064	0.016	-0.021	-0.056	-0.001
MSE	0.024	0.016	0.037	0.006	0.048	0.107	0.022	0.009	0.021	0.043	0.006	0.048	0.121	0.023

The table presents the average values of the bias (denoted as $BIAS=\mathbb{E}[\widehat{\delta}_\bullet-\delta_0]$)) and mean square error (MSE) of the estimates of threshold parameter δ and the variance-covariance parameters of the TVAR innovations $u^{(\mathrm{i})}_{1t}$ and $u^{(\mathrm{i})}_{2t}$, denoted $\sigma^{(i)2}_{u_1}$, $\sigma^{(i)2}_{u_2}$ and $\sigma^{(\mathrm{i})}_{u_1u_2}$, across the two regimes $i=1,2$. This is done for the following cases: (i) ignoring the threshold endogeneity, (ii) controlling for it based on the copula approach. For all the above cases, the sample size is set to $\mathrm{T}=200$. *We present results for the case that the threshold variables is distributed as $z_t\sim\mathcal{U}(-4,3)$ and the case that $z_t\sim\mathcal{N}(0,4.08)$. **Panel A** presents results for the case that the correlation between z_t and $u^{(\mathrm{i})}_j$ is linear, given as $corr\big(u^{(\mathrm{i})}_j,z\big)=0.8$, for $j=1,2$ equations and $i=1,2$ regimes, while **Panel B** is for the case that the correlation structure changes across the two regimes, i.e., $corr\big(u^{(1)}_j,z\big)=0.8$ and $corr\big(u^{(2)}_j,z\big)=-0.7$ for $j=1,2$ equations. Finally, note all metrics represent averages over $1,000$ replications; trimming at 10% was used when estimating δ.*

TABLE C.2. Monte Carlo Results, $\mathrm{T} = 500$, Scenarios 1. and 2.

	$z_t \sim \mathcal{U}(-4,3)$							$z_t \sim \mathcal{N}(0,4.08)$						
	δ	$(\sigma^{(1)}_{u_1})^2$	$(\sigma^{(1)}_{u_2})^2$	$\sigma^{(1)}_{u_1u_2}$	$(\sigma^{(2)}_{u_1})^2$	$(\sigma^{(2)}_{u_2})^2$	$\sigma^{(2)}_{u_1u_2}$	δ	$(\sigma^{(1)}_{u_1})^2$	$(\sigma^{(1)}_{u_2})^2$	$\sigma^{(1)}_{u_1u_2}$	$(\sigma^{(2)}_{u_1})^2$	$(\sigma^{(2)}_{u_2})^2$	$\sigma^{(2)}_{u_1u_2}$
Panel A: $corr(e^{(i)}_j z) = 0.8$ for $j = 1,2$ equations and $i = 1,2$ regimes														
$\mathrm{TVAR_X}$														
BIAS	−0.050	0.333	0.208	−0.002	0.148	0.083	−0.016	−1.017	1.266	1.263	0.111	0.647	0.629	−0.070
MSE	0.026	0.114	0.055	0.004	0.027	0.033	0.019	1.434	1.763	1.733	0.013	0.501	0.507	0.008
$\mathrm{TVAR_N}$														
BIAS	0.010	0.013	0.001	−0.006	0.037	0.001	−0.006	−0.001	0.080	−0.054	0.015	0.035	0.014	0.012
MSE	0.009	0.001	0.008	0.007	0.004	0.022	0.021	0.010	0.013	0.021	0.000	0.020	0.045	0.000
Panel B: $corr(e^{(1)}_j z) = 0.8$ and $corr(e^{(2)}_j z) = -0.7$ for $i = 1,2$														
$\mathrm{TVAR_X}$														
BIAS	−0.603	0.981	0.977	0.099	0.129	0.127	0.003	−2.616	0.443	0.435	0.060	0.535	0.537	−0.145
MSE	3.462	1.028	1.047	0.011	0.048	0.075	0.127	6.980	0.227	0.244	0.007	0.302	0.321	0.025
$\mathrm{TVAR_N}$														
BIAS	−0.003	0.038	0.004	0.007	0.001	−0.001	0.006	−0.004	0.088	0.086	0.010	0.006	0.013	−0.005
MSE	0.005	0.008	0.017	0.003	0.017	0.046	0.008	0.002	0.014	0.023	0.003	0.023	0.048	0.008

*The table presents the average values of the bias (denoted as $BIAS = \mathbb{E}[\widehat{\delta}_{\bullet} - \delta_0]$) and mean square error (MSE) of the estimates of threshold parameter δ and the variance-covariance parameters of the TVAR innovations $u^{(i)}_{1t}$ and $u^{(i)}_{2t}$, denoted $\sigma^{(i)2}_{u_1}$, $\sigma^{(i)2}_{u_2}$ and $\sigma^{(i)}_{u_1u_2}$, across the two regimes $i = 1,2$. This is done for the following cases: (i) ignoring the threshold endogeneity, (ii) controlling for it based on our suggested copula approach. For all the above cases, the sample size is set to $\mathrm{T} = 500$. We present results for the case that the threshold variables is distributed as $z_t \sim \mathcal{U}(-4,3)$ and the case that $z_t \sim \mathcal{N}(0,4.08)$. **Panel A** presents results for the case that the correlation between z_t and $u^{(i)}_j$ is linear, given as $corr(u^{(i)}_j, z) = 0.8$, for $j = 1,2$ equations and $i = 1,2$ regimes, while **Panel B** is for the case that the correlation structure changes across the two regimes, i.e., $corr(u^{(1)}_j, z) = 0.8$ and $corr(u^{(2)}_j, z) = -0.7$ for $j = 1,2$ equations. Finally, note all metrics represent averages over $1,000$ replications; trimming at 10% was used when estimating δ.*

TABLE C.3. Monte Carlo Results – Non-Normal VAR Threshold, Scenario 3.

	δ	$\sigma_{u_1}^{(1)2}$	$\sigma_{u_2}^{(1)2}$	$\sigma_{u_1u_2}^{(1)}$	$\sigma_{u_1}^{(2)2}$	$\sigma_{u_2}^{(2)2}$	$\sigma_{u_1u_2}^{(2)}$	δ	$\sigma_{u_1}^{(1)2}$	$\sigma_{u_2}^{(1)2}$	$\sigma_{u_1u_2}^{(1)}$	$\sigma_{u_1}^{(2)2}$	$\sigma_{u_2}^{(2)2}$	$\sigma_{u_1u_2}^{(2)}$
	TVAR_X							TVAR_N						
Panel A: $\text{T} = 200$ Student-t Threshold														
BIAS	−0.075	0.378	0.322	0.352	0.188	0.015	0.015	0.010	0.002	−0.021	−0.009	0.080	−0.047	0.010
MSE	0.033	0.151	0.141	0.136	0.045	0.050	0.010	0.004	0.001	0.019	0.002	0.012	0.048	0.007
Panel B: $\text{T} = 500$ Student-t Threshold														
BIAS	−0.026	0.398	0.355	0.374	0.191	0.025	0.021	0.002	−0.001	−0.003	−0.008	0.040	−0.005	−0.002
MSE	0.005	0.162	0.140	0.144	0.040	0.018	0.005	0.001	0.001	0.008	0.001	0.003	0.017	0.002
Panel C: $\text{T} = 200$ Skewed Student-t Threshold														
BIAS	−0.089	0.313	0.265	0.291	0.118	0.001	0.003	0.017	−0.066	−0.081	−0.072	0.012	−0.057	−0.003
MSE	0.047	0.106	0.107	0.098	0.023	0.048	0.009	0.006	0.006	0.026	0.008	0.005	0.047	0.006
Panel D: $\text{T} = 500$ Skewed Student-t Threshold														
BIAS	−0.030	0.327	0.292	0.308	0.116	0.015	0.007	0.000	−0.073	−0.068	−0.074	−0.032	−0.013	−0.018
MSE	0.008	0.111	0.099	0.100	0.017	0.018	0.004	0.002	0.006	0.013	0.007	0.003	0.017	0.003

The table presents the average values of the bias (denoted $BIAS = \mathbb{E}[\widehat{\delta}_\bullet - \delta_0]$) and mean square error (MSE) of the estimates of threshold parameter δ and the variance-covariance parameters of the TVAR innovations $u_{1t}^{(\text{i})}$ and $u_{2t}^{(\text{i})}$, denoted $\sigma_{u_1}^{(i)2}$, $\sigma_{u_2}^{(i)2}$ and $\sigma_{u_1u_2}^{(\text{i})}$, across the two regimes $i = 1, 2$, for cases that the distribution of the error terms $u_{1t}^{(\text{i})}$ and $u_{2t}^{(\text{i})}$ is Student-t with five degrees of freedom and skewed-t, with the same degrees of freedom. The threshold variable is $z_t \sim \mathcal{U}(-4, 3)$. The table presents results for the following two cases: (i) ignoring the threshold endogeneity, (ii) controlling for it based on our suggested method relying on the Gaussian copula approach. The table considers the cases of $\text{T} = 200$ and $\text{T} = 500$. Finally, note all metrics represent averages over $1,000$ replications; trimming at 10% was used when estimating δ.

TABLE C.4. MONTE CARLO RESULTS, $\mathrm{T} = 200$, SCENARIO 4.

	δ	$(\sigma^{(1)}_{u_1})^2$	$(\sigma^{(1)}_{u_2})^2$	$\sigma^{(1)}_{u_1u_2}$	$(\sigma^{(2)}_{u_1})^2$	$(\sigma^{(2)}_{u_2})^2$	$\sigma^{(2)}_{u_1u_2}$
				Clayton : $\psi = 3.5$			
				$\mathbf{TVAR_X}$			
BIAS	-1.1133	0.1374	0.0908	0.1387	0.3069	0.3253	0.1375
MSE	1.5577	0.0672	0.0996	0.0714	0.1098	0.1639	0.0301
				$\mathbf{TVAR_N}$			
BIAS	-0.0229	0.0184	-0.0078	0.0271	0.1231	-0.0302	0.0956
MSE	0.0108	0.0031	0.0257	0.0061	0.0256	0.0447	0.0201
				$\mathbf{TVAR_N} + \mathcal{H}_2(z_t^{(\mathrm{i})\star})$			
BIAS	-0.0159	0.0117	-0.0198	0.0208	0.1082	-0.0562	0.0935
MSE	0.0107	0.0028	0.0252	0.0056	0.0216	0.0445	0.0192
				Joe : $\psi = 4.0$			
				$\mathbf{TVAR_X}$			
BIAS	-1.0464	0.8139	0.6740	0.7609	0.7531	0.8297	0.4404
MSE	1.5964	0.8123	0.6200	0.7180	0.5921	1.0099	0.2402
				$\mathbf{TVAR_N}$			
BIAS	-0.0304	0.2144	0.1743	0.2080	0.6771	0.0473	0.2248
MSE	0.0316	0.0534	0.0634	0.0538	0.4935	0.0610	0.0744
				$\mathbf{TVAR_N} + \mathcal{H}_2(z_t^{(\mathrm{i})\star})$			
BIAS	-0.0314	0.2052	0.1586	0.2001	0.6525	0.0117	0.2172
MSE	0.0264	0.0490	0.0563	0.0497	0.4601	0.0524	0.0690
				Gumbel : $\psi = 5.0$			
				$\mathbf{TVAR_X}$			
BIAS	-1.0212	0.5722	0.4689	0.5272	0.5062	0.6500	0.2532
MSE	1.1693	0.3484	0.2734	0.2998	0.2699	0.5137	0.0789
				$\mathbf{TVAR_N}$			
BIAS	-0.0018	0.0635	0.0591	0.0606	0.1363	-0.0286	0.0325
MSE	0.0127	0.0045	0.0213	0.0048	0.0222	0.0418	0.0051
				$\mathbf{TVAR_N} + \mathcal{H}_2(z_t^{(\mathrm{i})\star})$			
BIAS	-0.0074	0.0612	0.0471	0.0581	0.1323	-0.0496	0.0290
MSE	0.0149	0.0042	0.0200	0.0044	0.0209	0.0430	0.0047

The table presents the average values of the bias ($BIAS = \mathbb{E}[\widehat{\delta}_\bullet - \delta_0]$) and mean square error (MSE) of the estimates of the threshold parameter δ and the variance–covariance parameters of the TVAR innovations $u^{(\mathrm{i})}_{1t}$ and $u^{(\mathrm{i})}_{2t}$, denoted $(\sigma^{(\mathrm{i})}_{u_1})^2$, $(\sigma^{(\mathrm{i})}_{u_2})^2$, and $\sigma^{(\mathrm{i})}_{u_1u_2}$, across the two regimes $i = 1, 2$, when the dependence structure between z_t and $u^{(\mathrm{i})}_j$ is nonlinear and governed by the Clayton, Joe, and Gumbel copulas. The copula parameters ψ are reported in parentheses and imply a Pearson correlation coefficient of about 0.7 between z_t and $u^{(\mathrm{i})}_{jt}$ for all j, i. Results are shown for: (i) ignoring threshold endogeneity, (ii) controlling for it via the copula approach, and (iii) adding a Hermite approximation. The sample size is $\mathrm{T} = 200$. Metrics are averages over 1,000 replications; trimming at 10% was used when estimating δ.

TABLE C.5. Monte Carlo Results, $\mathrm{T} = 500$, Scenario 4.

	δ	$(\sigma^{(1)}_{u_1})^2$	$(\sigma^{(1)}_{u_2})^2$	$\sigma^{(1)}_{u_1u_2}$	$(\sigma^{(2)}_{u_1})^2$	$(\sigma^{(2)}_{u_2})^2$	$\sigma^{(2)}_{u_1u_2}$
				Clayton : $\psi = 3.5$			
				$\mathbf{TVAR_X}$			
BIAS	−1.2027	0.1345	0.0892	0.1422	0.3582	0.3865	0.1356
MSE	1.6730	0.0519	0.0536	0.0546	0.1383	0.1849	0.0231
				$\mathbf{TVAR_N}$			
BIAS	−0.0111	0.0202	0.001	0.0327	0.14722e − 04		0.0984
MSE	0.0022	0.0014	0.0092	0.003	0.0258	0.0169	0.0144
				$\mathbf{TVAR_N} + \mathcal{H}_2(z^{(i)\star}_t)$			
BIAS	−0.0077	0.0143	−0.0088	0.0264	0.1409	−0.0107	0.0978
MSE	0.0021	0.0012	0.0090	0.0026	0.0238	0.0165	0.0141
				$Joe : \psi = 4$			
				$\mathbf{TVAR_X}$			
BIAS	−0.9993	0.8704	0.7417	0.8285	0.8060	0.8987	0.4648
MSE	1.4797	0.9025	0.6806	0.8184	0.6635	1.1474	0.2544
				$\mathbf{TVAR_N}$			
BIAS	−0.0019	0.2156	0.1726	0.2089	0.7198	0.061	0.2335
MSE	0.0032	0.0494	0.0403	0.0474	0.5341	0.0284	0.0649
				$\mathbf{TVAR_N} + \mathcal{H}_2(z^{(i)\star}_t)$			
BIAS	−0.0065	0.2038	0.1595	0.1995	0.7097	0.0487	0.2294
MSE	0.0030	0.0443	0.0356	0.0433	0.5196	0.0250	0.0627
				$Gumbell : \psi = 5.0$			
				$\mathbf{TVAR_X}$			
BIAS	−1.0739	0.5639	0.4752	0.5308	0.5519	0.7411	0.2676
MSE	1.2173	0.3289	0.2489	0.2936	0.3118	0.5961	0.0794
				$\mathbf{TVAR_N}$			
BIAS	−0.0047	0.0458	0.0285	0.043	0.1389	0.0015	0.0298
MSE	0.0032	0.0023	0.0065	0.0022	0.0207	0.0172	0.0026
				$\mathbf{TVAR_N} + \mathcal{H}_2(z^{(i)\star}_t)$			
BIAS	−0.0073	0.0448	0.0244	0.042	0.1369	−0.0073	0.0286
MSE	0.0033	0.0022	0.0063	0.0021	0.0201	0.0171	0.0025

The table presents the average values of the bias (denoted as $BIAS = \mathbb{E}[\widehat{\delta}_{\bullet} - \delta_0]$) and mean square error (MSE) of the estimates of threshold parameter δ and the variance-covariance parameters of the TVAR innovations $u^{(i)}_{1t}$ and $u^{(i)}_{2t}$, denoted $\sigma^{(i)2}_{u_1}$, $\sigma^{(i)2}_{u_2}$ and $\sigma^{(i)}_{u_1u_2}$, across the two regimes $i = 1, 2$, when the dependence structure between z_t and $u^{(i)}_j$ is nonlinear governed by the Clayton, Joe, and Gumbel copulas. The values of the copula parameters ψ considered are reported in parentheses. These imply a Pearson correlation coefficient of about 0.7 between z_t and $u^{(i)}_{jt} \forall j, i$. This is done for the following cases: (i) ignoring the threshold endogeneity, (ii) controlling for it based on our suggested copula approach, and (iii) the controlling case plus a Hermite approximation. For all the above cases, the size of the sample is set to $\mathrm{T} = 500$. Finally, note all metrics represent averages over $1{,}000$ replications; trimming at 10% was used when estimating δ.

Supplemental Internet Appendix for

Threshold Endogeneity in Vector Autoregressions

by Dimitris Christopoulos, Peter McAdam, and Elias Tzavalis

Abstract

This Internet Appendix provides additional material such as proofs, graphical presentations, robustness exercises, and elaborations on the theory in the main text.

Appendix D: Summary of Notation

Table D.1. Summary of notation (Part 1 of 4)

Symbol	Description
INDICES, SETS, AND BASIC OPERATORS	
K	Number of endogenous variables in the VAR (dimension of y_t).
p	VAR lag order (stacked in Y_{t-1}).
T	Sample size; $\mathrm{T}^{(\mathrm{i})}$ = regime-i sample size.
i	Regime index ($i=1$ low, $i=2$ high).
t	Time index ($t=1,\ldots,\mathrm{T}$).
h	IRF horizon, $h\in\mathbb{N}_0$. Maximum horizon: $\bar{H}$.
j	Shock index for IRFs, $j\in\{1,\ldots,\mathrm{K}\}$.
k	Response-variable index for IRFs, $k\in\{1,\ldots,\mathrm{K}\}$.
$\mathbf{1}\{\bullet\}$	Indicator function in $\{0,1\}$.
e_j	jth canonical basis vector in $\mathbb{R}^{\mathrm{K}}$.
I_n	$n\times n$ identity matrix.
$\mathrm{tr}(\bullet)$	Trace of a square matrix.
$\mathrm{vec}(\bullet)$	Vectorization (column–stacking) operator.
$\mathrm{diag}(\bullet)$	Diagonal matrix with given entries.
$\otimes$	Kronecker product.
$\perp$	Statistical independence.
$\succeq,\preceq$	Löwner (semidefinite) order.
$\rho(\bullet)$	Spectral radius.
$\\|\cdot\\|$, $\\|\cdot\\|_2$, $\\|\cdot\\|_{\mathrm{op}}$, $\\|\cdot\\|_F$	Euclidean (also ℓ_2), operator, and Frobenius norms.
$O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$, $o_p(\cdot)$	Deterministic and stochastic order-of-magnitude symbols.
$\xrightarrow{p}$, $\Rightarrow$	Convergence in probability; weak convergence.
PROBABILITY, DISTRIBUTIONS, AND OPERATORS	
$\mathbb{E}$, $\mathbb{V}\mathrm{ar}$, $\mathbb{C}\mathrm{ov}$, $\mathbb{C}\mathrm{orr}$	Expectation, variance, covariance, correlation operators.
$\mathcal{N}(\mu,\Sigma)$	(Multivariate) normal distribution.
$\mathcal{U}(a,b)$	Uniform distribution on $[a,b]$.
$\Phi(\cdot)$, $\phi(\cdot)$	Standard normal CDF and density.
$\Phi^{-1}(\cdot)$	Standard normal quantile function (PIT–Gaussian map).
$T_\nu(\cdot)$	Univariate Student-t CDF with ν degrees of freedom.
ς	Moment-existence exponent: $\mathbb{E}\\|y_t\\|^{2+\varsigma}<\infty$.
THRESHOLD VARIABLE, REGIMES, AND CDFs	
δ	Threshold level (parameter).
δ_0	True threshold level.
$\widehat{\delta}$	Estimated threshold.
$\delta^\star_{\mathrm{X}}$	Pseudo-true threshold under the misspecified $\mathrm{TVAR}_{\mathrm{X}}$.
$\mathrm{Z}^{(1)}$	Lower–regime event $\{z_t\le\delta\}$.
$\mathrm{Z}^{(2)}$	Upper–regime event $\{z_t>\delta\}$.
p_δ	Regime mass $p_\delta=\Pr(z_t\le\delta)=\mathrm{F}_z(\delta)\in(0,1)$.
z_t	Threshold variable.
$z_t^{(i)}$	Realization of z_t conditional on $z_t\in\mathrm{Z}^{(\mathrm{i})}$.
$\mathrm{F}_z(\cdot)$, F_{z_t}	Marginal CDF of z_t; shorthand $\mathrm{F}_{z_t}\equiv\mathrm{F}_z(z_t)$ used throughout (also applies to all marginal/conditional/joint CDFs).
$\mathrm{F}_z^{(i)}$, $\mathrm{F}_{z_t}^{(i)}$	Regime-i truncated CDF, eq. (2).
$f_z(\cdot)$	Density of z_t (assumed positive on a neighborhood of δ_0).
$\widehat{\mathrm{F}}_z$, $\widehat{\mathrm{F}}_{z_t}^{(i)}$	Empirical (or kernel) CDF and its regime-i truncated counterpart.
$z_t^{(i)\star}$	Regime-i Gaussian PIT score: $z_t^{(i)\star}=\Phi^{-1}(\mathrm{F}_{z_t}^{(i)})\sim\mathcal{N}(0,1)$.
$\widehat{z}_t^{(i)\star}$	Feasible (winsorized plug-in) PIT score: $\Phi^{-1}\big(\kappa_{\mathrm{T}}(\widehat{\mathrm{F}}_{z_t}^{(i)})\big)$.
$\kappa_{\mathrm{T}}(\cdot)$, ϵ_{T}	Winsorization map and threshold; $\kappa_{\mathrm{T}}(u)=\min\{1-\epsilon_{\mathrm{T}},\max\{\epsilon_{\mathrm{T}},u\}\}$.
L_{T}, ζ_{T}	Lipschitz bound for Φ^{-1} on $[\epsilon_{\mathrm{T}},1-\epsilon_{\mathrm{T}}]$; uniform CDF approximation error.

TABLE D.2. Summary of notation (Part 2 of 4)

Symbol	Description
VAR REPRESENTATION AND REDUCED-FORM INNOVATIONS	
Y_{t-1}	Regressor stack $(1, y'_{t-1}, \ldots, y'_{t-p})' \in \mathbb{R}^{\mathrm{K}p+1}$.
$\mathrm{A}^{(\mathrm{i})}$	Regime-i slope matrix $\mathrm{A}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times(\mathrm{K}p+1)}$.
$\mathrm{A}_{\mathrm{N}}^{(\mathrm{i})}$, $\mathrm{A}_{\mathrm{X}}^{(\mathrm{i})}$	Population slope matrices under $\mathrm{TVAR}_{\mathrm{N}}$ and $\mathrm{TVAR}_{\mathrm{X}}$.
$\Delta\mathrm{A}^{(\mathrm{i})}$	Omitted-control slope bias: $\Delta\mathrm{A}^{(\mathrm{i})} = \mathrm{A}_{\mathrm{X}}^{(\mathrm{i})} - \mathrm{A}_{\mathrm{N}}^{(\mathrm{i})} = \Lambda^{(\mathrm{i})}\gamma^{(\mathrm{i})}$.
$u_t^{(\mathrm{i})}$	Reduced-form innovation in regime i: $u_t^{(\mathrm{i})} = y_t - \mathrm{A}^{(\mathrm{i})}\mathrm{Y}_{t-1}$.
$\Sigma_{\mathrm{uu}}^{(\mathrm{i})}$	Uncorrected innovation covariance: $\Sigma_{\mathrm{uu}}^{(\mathrm{i})} = \mathbb{V}\mathrm{ar}(u_t^{(\mathrm{i})}) = \Lambda^{(\mathrm{i})}\Lambda^{(\mathrm{i})\prime} + \Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$ (also denotes the $\mathrm{TVAR}_{\mathrm{X}}$ population covariance).
$\varepsilon_t^{(\mathrm{i})}$	Endogeneity-corrected innovation: $\varepsilon_t^{(\mathrm{i})} = u_t^{(\mathrm{i})} - \mathbb{E}(u_t^{(\mathrm{i})} \mid z_t^{(\mathrm{i})}) = u_t^{(\mathrm{i})} - \Lambda^{(\mathrm{i})} z_t^{(i)\star}$.
$\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})}$	Corrected innovation covariance: $\Sigma_{\varepsilon\varepsilon}^{(\mathrm{i})} = \mathbb{V}\mathrm{ar}(\varepsilon_t^{(\mathrm{i})})$.
$\Lambda^{(\mathrm{i})}$	Regime-i copula-control loading vector ($\in \mathbb{R}^{\mathrm{K}\times 1}$).
$e_t^{(\mathrm{i})}$	Orthogonalized structural shock vector, $\mathbb{E}(e_t^{(\mathrm{i})} e_t^{(\mathrm{i})\prime}) = \mathrm{I}_{\mathrm{K}}$.
COPULA CONSTRUCTION AND PIT MACHINERY	
$\mathrm{C}^{(\mathrm{i})}(\cdot)$	Regime-i (K+1)-dimensional copula, eq. (3).
$c^{(\mathrm{i})}(\cdot)$	Corresponding copula density, eq. (5).
$\mathrm{F}_{u_t}^{(\mathrm{i})}$, $\mathrm{F}_{u_{kt}}^{(\mathrm{i})}$	Regime-i marginal CDFs of the innovation vector and its kth component.
$f_{u_{kt}}^{(\mathrm{i})}$	Regime-i marginal density of $u_{kt}^{(\mathrm{i})}$.
$\sigma_{uk}^{(\mathrm{i})}$	Regime-i standard deviation of $u_{kt}^{(\mathrm{i})}$.
$\mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}$	Diagonal matrix $\mathrm{diag}(\sigma_{u1}^{(\mathrm{i})}, \ldots, \sigma_{u\mathrm{K}}^{(\mathrm{i})})$.
$u_t^{(i)\star}$, $u_{kt}^{(i)\star}$	PIT-Gaussian innovations: $u_{kt}^{(i)\star} = \Phi^{-1}(\mathrm{F}_{u_{kt}}^{(\mathrm{i})})$.
$\mathrm{R}_{\mathrm{uu}}^{(\mathrm{i})}$	Latent correlation matrix of $u_t^{(i)\star}$.
$r_{uz}^{(\mathrm{i})}$	Latent correlation vector: $r_{uz}^{(\mathrm{i})} = \mathbb{C}\mathrm{orr}(u_t^{(i)\star}, z_t^{(i)\star}) \in \mathbb{R}^{\mathrm{K}\times 1}$.
$\eta_t^{(\mathrm{i})}$	Latent residual orthogonal to $z_t^{(i)\star}$, with $\varepsilon_t^{(\mathrm{i})} = \mathrm{D}_{\mathrm{uu}}^{(\mathrm{i})}\eta_t^{(\mathrm{i})}$.
COMPANION FORM, IRFS, AND STABILITY	
$\bullet$	Subscript tag. For covariance-class objects, $\bullet \in \{\mathrm{uu}, \varepsilon\varepsilon\}$. For dynamic objects (slopes, companion matrices, multipliers, IRFs), $\bullet \in \{\mathrm{N}, \mathrm{X}\}$ distinguishing $\mathrm{TVAR}_{\mathrm{N}}$ from $\mathrm{TVAR}_{\mathrm{X}}$.
x_t	Companion state $x_t = (y'_t, \ldots, y'_{t-p+1})' \in \mathbb{R}^{\mathrm{K}p}$.
$\mathcal{A}_\bullet^{(\mathrm{i})}$	Regime-i companion matrix under specification $\bullet$.
J	Selection matrix $J = [\mathrm{I}_{\mathrm{K}}\ 0 \cdots 0] \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}p}$.
$\Psi_\bullet^{(\mathrm{i})}(h)$	Reduced-form multiplier: $\Psi_\bullet^{(\mathrm{i})}(h) = J(\mathcal{A}_\bullet^{(\mathrm{i})})^h J'$.
$\mathrm{P}_\bullet^{(\mathrm{i})}$	Impact / identification matrix: $\mathrm{P}_\bullet^{(\mathrm{i})}(\mathrm{P}_\bullet^{(\mathrm{i})})' = \Sigma_\bullet^{(\mathrm{i})}$.
$p_1^{(\mathrm{i})}$	First column of $\mathrm{P}_{\varepsilon\varepsilon}^{(\mathrm{i})}$ (impact vector of the target shock).
$\mathrm{IRF}_\bullet^{(\mathrm{i})}(h, j)$	IRF response at horizon h to shock j: $\Psi_\bullet^{(\mathrm{i})}(h)\mathrm{P}_\bullet^{(\mathrm{i})} e_j$.
$B^{(\mathrm{i})}$ (slope)	Regime-i first-block-row VAR slope matrix in companion form; $B^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}p}$.
$\mathcal{R}$	Fixed selection matrix linking $\mathrm{vec}(\widehat{\mathcal{A}} - \mathcal{A}) = \mathcal{R}\,\mathrm{vec}(\widehat{B} - B)$.
$Q_{xx,\bullet}^{(\mathrm{i})}$	Companion-form second-moment matrix: $\mathbb{E}(x_{t-1}x'_{t-1} \mid \mathrm{Z}^{(\mathrm{i})})$ under specification $\bullet$.
$V_{\mathrm{N}}^{(\mathrm{i})}$, $V_{\mathrm{X}}^{(\mathrm{i})}$	Variance terms entering the finite-sample stability bound (Lemma 2).
$\mathfrak{r}_{\mathrm{stab}}^{(\mathrm{i})}$, $\mathfrak{r}_{\mathrm{N,stab}}^{(\mathrm{i})}$	Stability margin around $\mathcal{A}^{(\mathrm{i})}$; subscript N indicates margin around the correctly specified $\mathcal{A}_{\mathrm{N}}^{(\mathrm{i})}$.

TABLE D.3. Summary of notation (Part 3 of 4)

Symbol	Description
PROXY-SVAR IDENTIFICATION AND CONTAMINATION	
$v_t^{(\mathrm{i})}$	Regime-i external instrument / proxy variable.
$\lambda_v^{(\mathrm{i})}$	Proxy relevance: $\lambda_v^{(\mathrm{i})} = \mathbb{C}\mathrm{ov}(e_{1t}^{(\mathrm{i})}, v_t^{(\mathrm{i})}) \neq 0$, with $\mathbb{C}\mathrm{ov}(e_{jt}^{(\mathrm{i})}, v_t^{(\mathrm{i})}) = 0$ for $j \neq 1$.
$\omega^{(\mathrm{i})}$	Instrument–threshold covariance: $\omega^{(\mathrm{i})} = \mathbb{C}\mathrm{ov}(z_t^{(i)\star}, v_t^{(\mathrm{i})})$.
$\Sigma_{\varepsilon v}^{(\mathrm{i})}$	Proxy moment using corrected innovations: $\Sigma_{\varepsilon v}^{(\mathrm{i})} = \mathbb{C}\mathrm{ov}(\varepsilon_t^{(\mathrm{i})}, v_t^{(\mathrm{i})}) = \lambda_v^{(\mathrm{i})} p_1^{(\mathrm{i})}$.
$\Sigma_{uv}^{(\mathrm{i})}$	Proxy moment using uncorrected innovations: $\Sigma_{uv}^{(\mathrm{i})} = \lambda_v^{(\mathrm{i})} p_1^{(\mathrm{i})} + \omega^{(\mathrm{i})} \Lambda^{(\mathrm{i})}$.
$p_{\mathrm{comp}}^{(\mathrm{i})}$	Composite impact direction: $p_{\mathrm{comp}}^{(\mathrm{i})} = p_1^{(\mathrm{i})} + (\omega^{(\mathrm{i})}/\lambda_v^{(\mathrm{i})})\Lambda^{(\mathrm{i})}$.
$\Lambda_{\parallel}^{(\mathrm{i})}$, $\Lambda_{\perp}^{(\mathrm{i})}$	Components of $\Lambda^{(\mathrm{i})}$ parallel / orthogonal to $p_1^{(\mathrm{i})}$ (Corollary A.1).
$\theta^{(\mathrm{i})}$	Angle between $\Sigma_{uv}^{(\mathrm{i})}$ and $p_1^{(\mathrm{i})}$ (contamination angle).
$\Pi_{\perp v}$	Orthogonal projector onto the complement of $\mathrm{span}(v)$: $\Pi_{\perp v} = \mathrm{I} - vv'/\lVert v\rVert^2$.
ESTIMATION, CLS OBJECTIVES, AND GENERATED REGRESSORS	
$\mathcal{Z}_{\mathrm{T}}$	Sample support of the threshold variable.
l_δ	Trimmed search grid for δ: $\{\delta : \tau \leq \widehat{\mathrm{F}}_z(\delta) \leq 1-\tau\}$.
τ	Trimming fraction (e.g., $\tau = 0.10$).
$\mathcal{T}_i(\delta)$	Regime-i index set: $\mathcal{T}_1(\delta) = \{t : z_t \leq \delta\}$, $\mathcal{T}_2(\delta) = \{t : z_t > \delta\}$.
$\mathrm{RSS}(\delta)$	Residual sum of squares at candidate δ.
$Q_{\mathrm{N}}(\delta)$, $Q_{\mathrm{X}}(\delta)$	Population concentrated CLS objectives under $\mathrm{TVAR_N}$ and $\mathrm{TVAR_X}$.
$Q_{\mathrm{T}}(\delta)$, $\widehat{Q}_{\mathrm{T}}(\delta)$, $\widetilde{Q}_{\mathrm{T}}(\delta)$	Infeasible (oracle), feasible (winsorized plug-in), and oracle with frozen non-switcher controls CLS criteria.
$\mathcal{M}_{\mathrm{T}}(x)$, $\widehat{\mathcal{M}}_{\mathrm{T}}(x)$, $\widetilde{\mathcal{M}}_{\mathrm{T}}(x)$	Local objective processes evaluated at $\delta_0 + x/\mathrm{T}$.
$\Upsilon(\cdot)$	Compound-Poisson limit process of Chan (1993).
$T(\hat\delta - \delta_0) \Rightarrow \arg\min_x \Upsilon(x)$	Limit process governing the local threshold objective in Remark 1
$\gamma^{(\mathrm{i})}$	Projection of $z_t^{(i)\star}$ on Y_{t-1} in regime i.
$Q_{\mathrm{YY}}^{(\mathrm{i})}$	Conditional second-moment matrix: $\mathbb{E}(\mathrm{Y}_{t-1}\mathrm{Y}_{t-1}' \mid \mathrm{Z}^{(\mathrm{i})})$.
$r_t^{(\mathrm{i})}$	Residualized threshold score: $r_t^{(\mathrm{i})} = z_t^{(i)\star} - \gamma^{(\mathrm{i})}\mathrm{Y}_{t-1}$.
$D_R(\delta)$, $C_i(\delta)$	Residualized omission cost and its within-regime components.
h^+, h^-	Right/left one-sided slopes of the population threshold objective at δ_0.
H_0	Local quadratic curvature (smooth heuristic, Proposition 3 footnote).
c_α	LR critical value at significance level α for the threshold profile set, Remark 1.
MSE / RELATIVE-EFFICIENCY METRICS	
BIAS, MSE	Bias ($\mathbb{E}[\widehat{\theta}] - \theta_0$) and mean squared error of an estimator.
$\mathrm{RE}_{\delta,\mathrm{T}}$, $\mathrm{RE}_{\Sigma,\mathrm{T}}$	Relative MSE: ratio of $\mathrm{TVAR_X}$ to $\mathrm{TVAR_N}$ for threshold and variance objects.
b_δ, $b_\Sigma^{(\mathrm{i})}$	Asymptotic biases of $\mathrm{TVAR_X}$ estimators: $b_\delta = \delta_{\mathrm{X}}^\star - \delta_0$, $b_\Sigma^{(\mathrm{i})} = \Lambda^{(\mathrm{i})}\Lambda^{(\mathrm{i})\prime}$.
V_δ, $V_\Sigma^{(\mathrm{i})}$	Asymptotic variances of the correctly specified $\mathrm{TVAR_N}$ estimators.
C_δ, C_Σ	RE divergence constants: $\mathrm{RE}_{\delta,\mathrm{T}} \sim C_\delta \mathrm{T}^2$, $\mathrm{RE}_{\Sigma,\mathrm{T}} \sim C_\Sigma \mathrm{T}$.

TABLE D.4. Summary of notation (Part 4 of 4)

Symbol	Description
SIEVE / NON-ELLIPTICAL EXTENSION	
$\vartheta^{(\mathrm{i})}$	Copula dependence parameter indexing departures from independence.
$\xi^{(\mathrm{i})}(\cdot)$, $\xi_j^{(\mathrm{i})}(\cdot)$	Copula-density score at independence and its jth component.
$B^{(\mathrm{i})}(z_t^{(\mathrm{i})})$ (kernel)	Nonlinear conditional-mean kernel arising in the non-elliptical case, eq. (B.5).
$s_j(\cdot)$, $\beta_j^{(\mathrm{i})}$	Sieve basis functions and regime-i coefficients.
d_{T}	Sieve dimension (number of basis terms; $d_{\mathrm{T}} = 1$ recovers the Gaussian one-factor control).
$\mathcal{H}_j(\cdot)$	Orthonormal Hermite polynomial of order j evaluated at $z_t^{(i)\star}$.
$\widetilde{\varepsilon}_t^{(\mathrm{i})}$	Sieve-augmented residual innovation.
COMPUTATIONAL PARAMETERS	
G	Number of threshold grid points searched.
$\mathcal{B}$	Number of bootstrap replications.
$\bar{H}$	Maximum horizon for IRFs.
q	Regressors per equation including controls: $q = \mathrm{K}p + 1 + d$.
d	Copula-control dimension (sieve order).
MONTE CARLO DGP AND COPULAS	
$\mathfrak{B}$	Cholesky factor of the reduced-form covariance in the MC DGP: $\Sigma_{\mathrm{uu}}^{(\mathrm{i})} = \mathfrak{B}\mathfrak{B}'$.
L	Contemporaneous structural loading matrix mapping e_t into u_t.
ψ	Archimedean copula dependence parameter (Clayton/Joe/Gumbel).
U, V	Uniform $[0,1]$ arguments of copula functions.
$\mathrm{C}_{\mathrm{Clayton}}, \mathrm{C}_{\mathrm{Joe}}, \mathrm{C}_{\mathrm{Gumbel}}$	Archimedean copulas used in Scenario 4.
EMPIRICAL APPLICATION (NK MODEL)	
$\widetilde{y}_t$	Output gap.
π_t	Inflation rate (annualized PCE deflator change).
ebp_t	Excess Bond Premium of Gilchrist and Zakrajšek (2012).
$\mathcal{F}(z_t)$, $\mathcal{F}^L$, $\mathcal{F}^H$	State-dependent Phillips-curve slope and its low/high-regime values.
ε_t^m	Monetary policy shock.
A_0	VAR intercept

APPENDIX E: ADDITIONAL RESULTS AND TECHNICAL NOTES

This appendix presents Supplemental results and technical details related to the main sections of the paper. The material is not included in the main text but may be of independent interest.

E.1 *Impact contamination and weak identification instruments*

Lemma E.1 (Size of the impact contamination). *If $\omega^{(\mathrm{i})} \neq 0$ and $p_1^{(\mathrm{i})}$ is not collinear with $\Lambda^{(\mathrm{i})}$, the (small-angle, denoted by "$\angle$") deviation between the $\mathrm{TVAR_X}$ impact direction and $p_1^{(\mathrm{i})}$ obeys*

$$\angle\big(\, \lambda_v^{(\mathrm{i})} p_1^{(\mathrm{i})} + \omega^{(\mathrm{i})}\Lambda^{(\mathrm{i})},\, p_1^{(\mathrm{i})}\,\big) \;\approx\; \frac{|\omega^{(\mathrm{i})}|}{|\lambda_v^{(\mathrm{i})}|}\,\frac{\|\Pi_{p_1}^{\perp(\mathrm{i})}\Lambda^{(\mathrm{i})}\|}{\|p_1^{(\mathrm{i})}\|}, \tag{E.1}$$

where $\Pi_{p_1}^{\perp(\mathrm{i})}$ projects onto the orthogonal complement of $p_1^{(\mathrm{i})}$. Note that if $\Lambda^{(\mathrm{i})}$ is collinear with p_1 then $\Pi_{p_1}^{\perp(\mathrm{i})}\Lambda^{(\mathrm{i})} = 0$ and the angle is correctly zero.

PROOF. Let $v \neq 0$ and w be vectors in $\mathbb{R}^n$, and let $\varepsilon \in \mathbb{R}$. Denote by $\angle(v+\varepsilon w,\, v)$ the angle between $v+\varepsilon w$ and v. Let

$$\Pi_v^{\perp} \;\equiv\; I - \frac{vv'}{\|v\|^2} \tag{E.2}$$

be the orthogonal projector onto the complement of $span(v)$. We begin by decomposing $v+\varepsilon w$ into components parallel and orthogonal to v:

$$v+\varepsilon w = \underbrace{\left(v + \frac{vv'}{\|v\|^2}\varepsilon w\right)}_{\text{component along } v} + \underbrace{\Pi_v^{\perp}(\varepsilon w)}_{\text{component orthogonal to } v} \tag{E.3}$$

This decomposition is exact and separates the perturbation εw into its projection along v and its projection orthogonal to v. The two components form a right triangle whose hypotenuse is $v+\varepsilon w$ and whose side opposite the angle $\angle(v+\varepsilon w,\, v)$ has length $\|\Pi_v^{\perp}(\varepsilon w)\|$. Hence,

$$\sin\angle(v+\varepsilon w,\, v) = \frac{\|\Pi_v^{\perp}(\varepsilon w)\|}{\|v+\varepsilon w\|}. \tag{E.4}$$

As $\varepsilon \to 0$, we have

$$\|v+\varepsilon w\| = \|v\| + O(|\varepsilon|), \qquad \|\Pi_v^{\perp}(\varepsilon w)\| = |\varepsilon|\,\|\Pi_v^{\perp} w\|. \tag{E.5}$$

Moreover, for small angles ϕ, $\sin\phi \sim \phi$. Combining these relations yields the first-order expansion

$$\angle(v+\varepsilon w,\, v) = \frac{\|\Pi_v^\perp w\|}{\|v\|}\,|\varepsilon| + O(\varepsilon^2), \qquad \varepsilon \to 0. \tag{E.6}$$

Applying this result with $v^{(\mathrm{i})} = \lambda_v^{(\mathrm{i})} p_1^{(\mathrm{i})}$ and $w^{(\mathrm{i})} = \omega^{(\mathrm{i})}\Lambda^{(\mathrm{i})}$ delivers the statement of the lemma. □

The quantity

$$\angle\Big(\lambda^{(\mathrm{i})} p_1^{(\mathrm{i})} + \omega^{(\mathrm{i})}\Lambda^{(\mathrm{i})},\, p_1^{(\mathrm{i})}\Big) \tag{E.7}$$

provides a first-order approximation to the distortion in the identified impact direction. The lemma shows that, when the instrument loads on the threshold variable, proxy-VAR identification is rotated toward the threshold-selection direction. The distortion increases in the instrument–threshold correlation $\omega^{(\mathrm{i})}$ and decreases in the instrument strength $\lambda^{(\mathrm{i})}$, reflecting the fact that weaker instruments are more susceptible to contamination by omitted selection effects.

E.2 *Convergence rates of the estimators: A summary*

Table E.1 summarizes the convergence rates of the estimators considered in the paper.

TABLE E.1. Large-sample rates and limits for key $\mathrm{TVAR_N}$ objects

Object	Estimator	Rate	Limit law
Regime slopes $A^{(\mathrm{i})}$, loadings $\Lambda^{(\mathrm{i})}$	OLS within regime	$\sqrt{\mathrm{T}}$	Normal
IRFs $\Psi^{(\mathrm{i})}(h)$ (fixed h)	Plug-in and "delta" method	$\sqrt{\mathrm{T}}$	Normal
Threshold δ	CLS over trimmed grid	T	$\arg\min$ of compound-Poisson process
Sieve-$\mathrm{TVAR_N}$ slopes/IRFs	OLS with d_{T} control	$\sqrt{\mathrm{T}}$ if $d_{\mathrm{T}}^2/\mathrm{T}\to 0$	Normal
Proxy–SVAR (strong IV)	Moment on controlled residual	$\sqrt{\mathrm{T}}$	Normal
Proxy–SVAR (weak IV)	As above, $\lambda/\sqrt{\mathrm{T}}$	—	Non-Gaussian (Cauchy-type)

This table collects the large-sample behavior of the main $\mathrm{TVAR_N}$ estimators. Regime-specific slope and loading estimates are standard $\sqrt{\mathrm{T}}$ objects (as are fixed-horizon IRFs via plug-in/delta-method arguments), whereas the threshold estimator is nonstandard, converging at rate T to an argmin of a compound-Poisson process; inference for the threshold should therefore be based on a profile-LR set or a block bootstrap that re-estimates the control and the threshold. When the copula control is implemented via a sieve with dimension d_{T}, the usual $\sqrt{\mathrm{T}}$ limits for slopes and IRFs continue to apply provided d_{T} grows slowly enough (e.g. $d_{\mathrm{T}}^2/\mathrm{T} \to 0$). For proxy-SVAR identification, the impact column is invariant to including the control under strong instruments, but under weak (local-to-zero) relevance the limit becomes non-Gaussian and weak-IV-robust inference is required.

APPENDIX F: COMPUTATIONAL COST AND SCALING OF $\mathrm{TVAR_N}$

As before, T is the sample length; K is the number of endogenous variables; p is the lag order; G is the number of threshold grid points (post trimming); $\mathcal{B}$ is the number of bootstrap replications; $\bar{H}$ is the maximum horizon for IRFs; d is the control dimension ($d = 1$ for the Gaussian one-factor; sieve uses $d = d_{\mathrm{T}}$); q is the number of regressors per equation: $q = (\mathrm{K}p + 1 + d)$ (the $+1$ is for the intercept and the d is for the controls), and here O denotes complexity, order-of-growth notation.

Per-step complexity (feasible CLS implementation)

We decompose the runtime and memory in the canonical implementation (grid search over δ, regime-wise OLS, optional bootstrap).[1]

Stage (1) Preprocessing. Sort $\{z_t\}_{t=1}^{\mathrm{T}}$ and build the empirical CDF (for fast truncated CDF evaluations):

$$\text{time } = O\left(\mathrm{T}\log\mathrm{T}\right), \qquad \text{memory } = O\left(\mathrm{T}\right). \tag{F.1}$$

Stage (2) Per grid value δ.

(a) Build controls. For each candidate δ, form regime-specific PIT-Gaussian controls $z_t^{(i)\star} = \Phi^{-1}\big(\mathrm{F}_{\mathrm{Z}}^{(\mathrm{i})}(z_t;\delta)\big)$ in a single pass:

$$\text{time } = O\left(\mathrm{T}d\right) \quad \text{with } O\left(\mathrm{T}\right) \text{ if } d = 1. \tag{F.2}$$

(b) Regime-wise OLS. With $X \in \mathbb{R}^{\mathrm{T}_i \times q}$ and $Y \in \mathbb{R}^{\mathrm{T}_i \times \mathrm{K}}$ in regime i, a QR/normal equations solve with K right-hand sides costs:

$$O\left(\mathrm{T}_i q^2 + \mathrm{T}_i q\mathrm{K} + q^3 + q^2\mathrm{K}\right) \quad \text{per regime,} \quad O\left(\mathrm{T}q^2 + \mathrm{T}q\mathrm{K} + q^3 + q^2\mathrm{K}\right) \text{ across both.} \tag{F.3}$$

When $\mathrm{K} \le q$ (which is typical VAR case with $q \approx \mathrm{K}p$ and $p \ge 1$) the $q^2\mathrm{K}$ term is dominated by q^3.

Summary (per δ):

$$\text{time per } \delta \;=\; O\left(\mathrm{T}\,q(q+\mathrm{K}) \;+\; q^3 \;+\; q^2\mathrm{K} \;+\; \mathrm{T}d\right). \tag{F.4}$$

Stage (3) Selection and refit at $\widehat{\delta}$. One additional two-regime OLS:

$$O\left(\mathrm{T}q^2 + \mathrm{T}q\mathrm{K} + q^3 + q^2\mathrm{K}\right). \tag{F.5}$$

Stage (4) IRFs at $\bar{H}$ horizons. Using the $\mathrm{K}p \times \mathrm{K}p$ companion:[2]

$$\text{time } = O\left(\bar{H}\,(\mathrm{K}p)^3\right) \quad \text{(reduced in practice via recursions).} \tag{F.6}$$

[1] A Caveat: Storing G grid points of precomputed statistics (e.g., $X'X$) reduces runtime but increases memory to $O(Gq^2)$. Thus, memory may increase to $O(Gq^2)$ if sufficient statistics are cached for grid reuse.

[2] In practice, Krylov/recurrence methods or eigen-decomposition can reduce the per-horizon cost to $O((\mathrm{K}p)^2)$ after an $O((\mathrm{K}p)^3)$ initialization; so the bound reported is conservative.

Stage (5) Bootstrap. A moving-block bootstrap that *recomputes* the controls and $\widehat{\delta}$ recomputes steps (2)–(4) for each draw, i.e. adds a factor $\mathcal{B}$ to those costs. However, IRFs tend to be only computed once post-bootstrap (rather than per draw). Accordingly, we can specify two relevant cases: (I) Threshold-only bootstrap: $O(BG\mathrm{T}(q^2+q^3))$; (II) Full IRF bootstrap: add $O(B\bar{H}(Kp)^2)$.

Total runtime and memory

Collecting terms in Table F.1:

$$\text{Time} = O\Big(\mathrm{T}\log\mathrm{T} + G\left[\mathrm{T}\,q(q+\mathrm{K})+q^3+q^2\mathrm{K}+\mathrm{T}d\right] \tag{F.7}$$

$$+\ \bar{H}\,(\mathrm{K}p)^3 + \mathcal{B}\Big(G\left[\mathrm{T}\,q(q+\mathrm{K})+q^3+q^2\mathrm{K}+\mathrm{T}d\right]+\bar{H}\,(\mathrm{K}p)^3\Big)\Big). \tag{F.8}$$

In the common regime $\mathrm{T}\gg q^2$ and $d\le q$, we have

$$\text{leading term} \approx O\left((\mathcal{B}{+}1)\,G\,\mathrm{T}\,q(q+\mathrm{K})\right), \tag{F.9}$$

and if additionally $\mathrm{K}\le q$, the $q^2\mathrm{K}$ term is absorbed by q^3.

Special case for the Gaussian one-factor control.

For the Gaussian one-factor control $(d=1)$, $q=\mathrm{K}p+1$ and

$$\text{Time} \approx O\big((\mathcal{B}{+}1)\,G\,\mathrm{T}\,(\mathrm{K}p+1)\,(\mathrm{K}p+\mathrm{K}+1)\big) = O\big((\mathcal{B}{+}1)\,G\,\mathrm{T}\,\mathrm{K}^2\,p(p+1)\big). \tag{F.10}$$

IRFs add $O\big((\mathcal{B}{+}1)\,\bar{H}\,(\mathrm{K}p)^3\big)$ if IRFs are computed for each bootstrap draw (e.g. for bands), and $O\big(\bar{H}\,(\mathrm{K}p)^3\big)$ as a one-off otherwise.

Memory. Storing Y, X and the control columns requires

$$\text{Memory} = O\left(\mathrm{T}\,(\mathrm{K}+q)\right). \tag{F.11}$$

Remarks

- *Sieve control.* If a d_T-term sieve is used, replace d by d_T in q; the control-build pass adds $O\left(\mathrm{T}d_\mathrm{T}\right)$ and OLS terms scale with q.

- *When* $\mathrm{T}\gg q^2$, the q^3 term is dominated and total cost is linear in T, linear in G, and linear in $\mathcal{B}$.

- *Practical speedups (same asymptotics).* One QR per regime with K right-hand side, accumulate $X'X$ and $X'Y$ via BLAS-3 (Level-3 routines in the Basic Linear Algebra Subprograms); reuse sufficient statistics along a sorted grid; coarse-to-fine δ grids.

TABLE F.1. Per-operation summary (asymptotic order)

Operation	**Time**	**Memory**	**Notes**
Sort & empirical CDF of z_t	$O\,(\mathrm{T}\log\mathrm{T})$	$O\,(\mathrm{T})$	one-off
Controls per δ	$O\,(\mathrm{T}d)$	$O\,(\mathrm{T})$	$d{=}1$ for Gaussian one-factor
Two-regime OLS per δ	$O\left(\mathrm{T}q^2+\mathrm{T}q\mathrm{K}+q^3+q^2\mathrm{K}\right)$	$O\,(\mathrm{T}(\mathrm{K}{+}q))$	$q=\mathrm{K}p{+}1{+}d$
Select $\widehat{\delta}$ and refit	$O\left(\mathrm{T}q^2+\mathrm{T}q\mathrm{K}+q^3+q^2\mathrm{K}\right)$	$O\,(\mathrm{T}(\mathrm{K}{+}q))$	one-off
IRFs (all horizons)	$O\left(\bar{H}\,(\mathrm{K}p)^3\right)$	$O\left((\mathrm{K}p)^2\right)$	companion powers
Bootstrap ($\mathcal{B}$ draws)	add $\mathcal{B}\times$ (steps 2–4)	$\times\,1$	recompute controls and $\widehat{\delta}$; IRFs if bands

Accordingly, for fixed (K,p,d) and without the bootstrap, $\mathrm{TVAR_N}$'s feasible CLS scales essentially as

$$O\big(G\,\mathrm{T}\,q(q{+}\mathrm{K})\big) \quad\Rightarrow\quad \text{linear in } \mathrm{T}, \tag{F.12}$$

with IRFs adding one $O\big(\bar{H}\,(\mathrm{K}p)^3\big)$ pass – or $(\mathcal{B}{+}1)$ such passes if IRFs are bootstrapped.

Appendix G: Data, Data Sources, and Construction

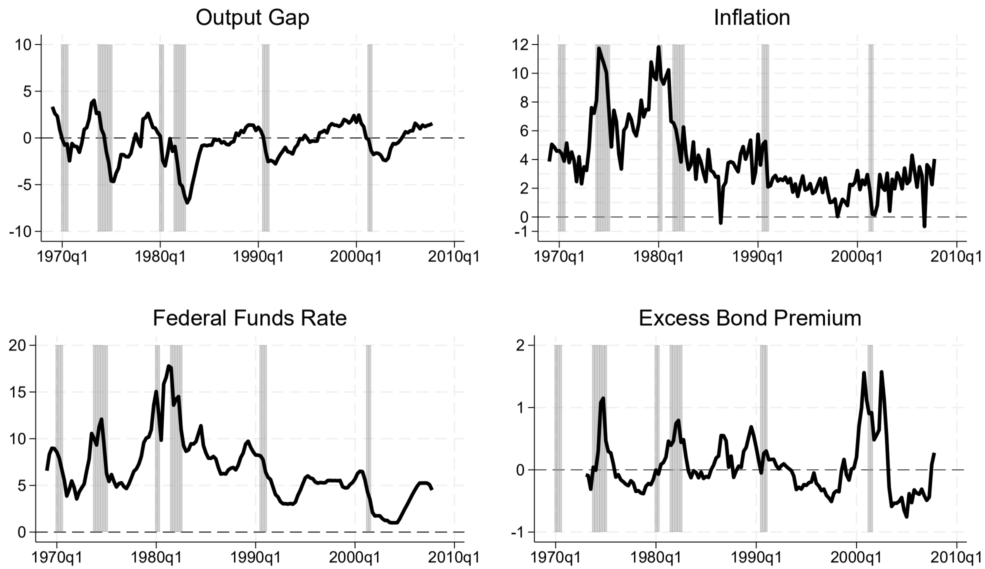

FIGURE G.1. Estimation Data

This figure shows the estimation variables: the Output Gap (the percentage gap between real and potential output as calculated by the Congressional Budget Office); the rate of Inflation (the percentage change in the chain-weighted PCE price deflator at an annual rate); and the Federal Funds rate; and the Excess Bond Premia. The sample shown is over 1969q1-2007q4. For visual convenience, a dashed reference line of zero is drawn, as are the NBER recession dates (in gray vertical bars).

TABLE G.1. Data Sources and Transformations

Series	**FRED Mnemonic**	**Description**	**Transformation**	**Sample Restrictions**
Potential GDP	`GDPPOT`	Potential Output	Billions of Chained 2017 Dollars SAAR	–
Real GDP	`GDPC1`	Realized Real GDP	Billions of Chained 2017 Dollars, SAAR	–
Price	`PCEPI`	Personal Consumption Expenditures	Chain-type Price Index	–
Recession	`USRECD`	NBER Recession Indicator $\in \{0,1\}$	End of Period	
Federal Funds Rate	`FEDFUNDS`	Effective Federal Funds Rate	Quarterly Average	–
Inflation		Inflation Rate	$\pi = 400 \times (\ln \mathtt{PCEPI}_t - \ln \mathtt{PCEPI}_{t-1})$	–
z_t		Livingston Inflation Expectations Survey	MA(2)[a]	
Output Gap		Output Gap	$100 \times \frac{GDPC1-GDPPOT}{GDPPOT}$	–
Excess Bond Premium		Gilchrist and Zakrajšek (2012), Favara et al. (2016)		From 1973q1
Monetary Shocks		Romer and Romer (2004) Wieland and Yang (2020)		1969-1996 1969-2007

This table shows the sources, definitions and transformations for our empirical exercises. Where the data is provided by the Federal Reserve Bank of St. Louis, we indicate the FRED mnemonic in the second column. [a] *To reduce short-term volatility and isolate the underlying expectations trend, a two-quarter moving average is constructed.*

APPENDIX H: ADDITIONAL MATERIALS

TABLE H.1. TVAR_N Estimation and Bootstrapped Probability Values

Regressor	Low-Inflation Regime Estimate	Low-Inflation Regime p-value	High-Inflation Regime Estimate	High-Inflation Regime p-value
	Equation: $\widetilde{y}$			
$\widetilde{y}_{t-1}$	0.871	0.001	0.857	0.001
π_{t-1}	−0.016	0.774	−0.121	0.063
$i_{f,t-1}$	−0.008	0.874	−0.065	0.354
ebp_{t-1}	−0.521	0.222	−1.221	0.001
A_0	−0.001	0.998	0.996	0.001
λ	−0.137	0.512	0.273	0.057
	Equation: π			
$\widetilde{y}_{t-1}$	0.021	0.862	0.197	0.274
π_{t-1}	0.257	0.028	0.527	0.001
$i_{f,t-1}$	−0.024	0.820	−0.004	0.978
ebp_{t-1}	−0.035	0.966	−1.072	0.005
A_0	1.917	0.004	1.818	0.006
λ	0.202	0.525	1.518	0.001
	Equation: i_f			
$\widetilde{y}_{t-1}$	0.129	0.253	0.064	0.638
π_{t-1}	0.079	0.428	−0.040	0.683
$i_{f,t-1}$	0.899	0.001	0.911	0.001
ebp_{t-1}	−0.345	0.690	−1.644	0.001
A_0	0.219	0.744	0.646	0.202
λ	−0.015	0.964	0.696	0.001
	Equation: ebp			
$\widetilde{y}_{t-1}$	0.036	0.013	0.011	0.804
π_{t-1}	0.036	0.008	0.027	0.325
$i_{f,t-1}$	0.025	0.019	0.019	0.526
ebp_{t-1}	0.889	0.001	0.827	0.001
A_0	−0.246	0.001	−0.215	0.137
λ	−0.081	0.098	−0.107	0.134

This table shows the estimation results of the bootstrapped estimation for the TVAR_N *model. The first block refers to results in the low regime,* $z_t \leq \delta$, and the second block corresponds to the high inflation regime.

TABLE H.2. Linear VAR Estimation and Bootstrapped Probability Values

Regressor	Estimate	p-value
	Equation: $\widetilde{y}$	
$\widetilde{y}_{t-1}$	0.923	0.000
π_{t-1}	−0.057	0.170
$i_{f,t-1}$	0.127	0.049
ebp_{t-1}	−0.669	0.013
$\widetilde{y}_{t-2}$	−0.093	0.271
π_{t-2}	0.020	0.641
$i_{f,t-2}$	−0.175	0.006
ebp_{t-2}	0.159	0.564
A_0	0.403	0.002
	Equation: π	
$\widetilde{y}_{t-1}$	−0.033	0.866
π_{t-1}	0.599	0.000
$i_{f,t-1}$	0.188	0.165
ebp_{t-1}	−0.160	0.774
$\widetilde{y}_{t-2}$	0.074	0.676
π_{t-2}	0.272	0.003
$i_{f,t-2}$	−0.170	0.200
ebp_{t-2}	−0.138	0.810
A_0	0.402	0.125
	Equation: i_f	
$\widetilde{y}_{t-1}$	0.266	0.061
π_{t-1}	0.068	0.281
$i_{f,t-1}$	0.944	0.000
ebp_{t-1}	−0.388	0.337
$\widetilde{y}_{t-2}$	−0.208	0.107
π_{t-2}	0.058	0.370
$i_{f,t-2}$	−0.031	0.749
ebp_{t-2}	−0.183	0.661
A_0	0.100	0.596
	Equation: ebp	
$\widetilde{y}_{t-1}$	−0.014	0.644
π_{t-1}	0.014	0.298
$i_{f,t-1}$	0.016	0.454
ebp_{t-1}	1.004	0.000
$\widetilde{y}_{t-2}$	0.034	0.225
π_{t-2}	−0.018	0.205
$i_{f,t-2}$	−0.003	0.885
ebp_{t-2}	−0.169	0.066
A_0	−0.049	0.238

This table shows the estimation results of the bootstrapped estimation for the linear VAR model.

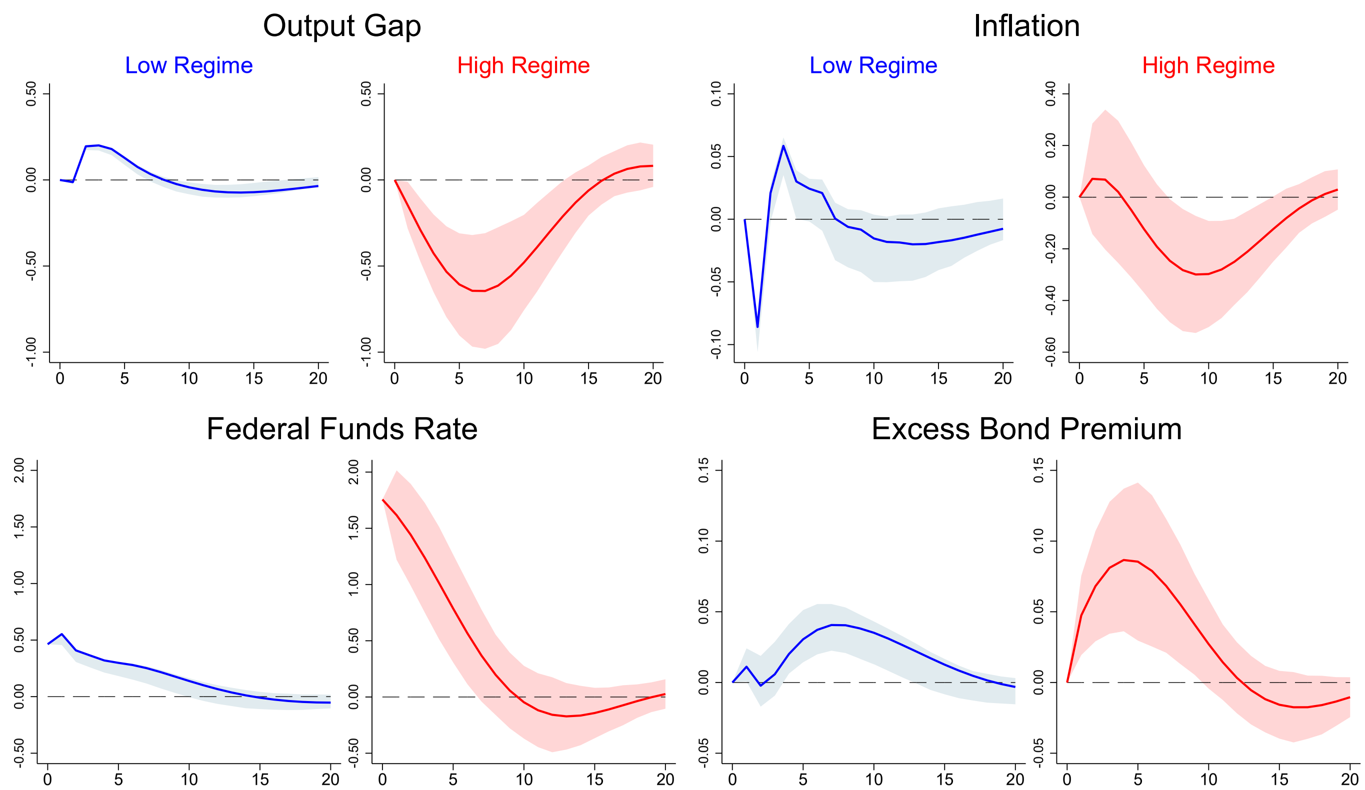

FIGURE H.1. TVAR_X Impulse Response to a Monetary Shock

This figure reports impulse responses to an identified monetary policy shock over a 20 quarter horizon. Responses are shown for the output gap, inflation, the federal funds rate, and the excess bond premium. Blue and red lines denote responses in the low and high inflation regime, respectively, for the exogenous TVAR model. The high-inflation regime occurs for the latter model when $z_t > 4.8$. A reference (dashed) line of zero is drawn for visual convenience. The shaded areas indicate bootstrapped 68% confidence bands. For visual convenience, we overlay a horizontal dashed line marking the zero response.

Acknowledgements

The paper is dedicated to the memory of Ivan Jaccard.

Dimitris Christopoulos
European Central Bank, Frankfurt am Main, Germany; email: dk.christopoulos@gmail.com

Peter McAdam
European Central Bank, Frankfurt am Main, Germany; email: peter.mcadam@ecb.europa.eu

Elias Tzavalis
Athens University of Economics & Business, Athens, Greece; email: etzavalis@aueb.gr

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

This paper can be downloaded without charge from www.ecb.europa.eu, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-7997-9 ISSN 1725-2806 doi:10.2866/5071779 QB-01-26-191-EN-N