TI 2024-072/III
Tinbergen Institute Discussion Paper

Modeling Common Bubbles: A Mixed Causal Non-Causal Dynamic Factor Model

Revision: January 2025

Gabriele Mingoli[1]

1 Vrije Universiteit Amsterdam, Tinbergen Institute

Modeling Common Bubbles: A Mixed Causal Non-Causal Dynamic Factor Model

Gabriele Mingoli[(a,b)*]

(a) Vrije Universiteit Amsterdam, (b) Tinbergen Institute.

January 29, 2025

Abstract

This paper introduces a novel dynamic factor model designed to capture common locally explosive episodes, also known as common bubbles, within large-dimensional, potentially non-stationary time series. The model leverages a lower-dimensional set of common unobserved factors exhibiting locally explosive behavior to identify common extreme events. Modeling these explosive behaviors allows to predict systemic risk and test for the emergence of common bubbles. The dynamics of the explosive factors are modeled using mixed causal non-causal models, a class of heavy-tailed autoregressive models that allow processes to depend on their future values through a lead polynomial. The paper establishes the asymptotic properties of the model and provides sufficient conditions for consistency of the estimated factors and parameters. A Monte Carlo simulation confirms the good finite sample properties of the estimator, while an empirical analysis highlights its practical effectiveness. Specifically, the model accurately identifies the common explosive component in monthly stock prices of NASDAQ-listed energy companies during the financial crisis in

*Email : g.mingoli@vu.nl. An earlier version of this paper is presented at various seminars, including "12th SIdE Workshop for PhD students in Econometrics and Empirical Economics" in Bertinoro, "Financial Econometrics Seminar" at CREST, "2024 Job Market Jamboree" at Tinbergen Institute and Econometrics Seminar at Paris Dauphine University. The author is grateful for comments from Christian Francq, Christian Gouriéroux, Jean-Michel Zakoïan, Peter Boswijk, Arthur Thomas, Gilles de Truchis, Elena Dumitrescu and other colleagues.

2008 and predicts its evolution significantly outperforming alternative forecasting methods.

Key words: observation-driven filter, non-stationary time-series, mixed causal non-causal models, dynamic factor models.

1 Introduction

Speculative bubbles — defined as rapid, unsustainable price surges followed by sharp declines — pose a significant threat to economic stability. These locally explosive episodes occur in various domains, as financial markets, real estate and commodity prices. When such bubbles propagate simultaneously across multiple time series, they can become a critical source of instability for investors, governments, and consumers, as many elements of the same economy are exposed to the same extreme event. Indeed, many recent financial crises can be traced back to such common bubbles. Therefore, accurately predicting and modeling the development of these shared explosive episodes is crucial to protect economic and financial stability. This paper introduces a novel dynamic factor model designed to capture common bubbles in high-dimensional time series vectors. By allowing high-dimensional time series to be driven by a smaller set of unobserved common factors, some of which experience explosive episodes, this model captures the onset and propagation of common extreme events. Modeling the explosive dynamics of these factors the model allows to assess systemic risk, offering early detection of common bubbles and measures to evaluate shared vulnerabilities. The proposed dynamics for the explosive factors allow to disentangle the common unobserved components in their fundamental value and bubble elements. This novel approach is particularly suited for non-stationary settings like stock and commodity prices and large macroeconomic datasets, where traditional models for common bubbles fall short.

This paper extends the two-stage estimation procedure from Barigozzi et al. (2021) to account for explosive factor dynamics, providing theoretical results for the asymptotic behavior of the estimated factors and parameters in the context of common bubbles. In particular, the paper explores the performance of Principal Component Analysis (PCA) for processes with only two finite moments and in the presence of explosive dynamics, establishing convergence rates for the estimates using the framework of Davis and Resnick (1986) and Davis et al. (2014). Monte Carlo simulations confirm the reliability of this estimation procedure, showing good finite-sample properties for both the PCA estimates and maximum likelihood parameter estimates. An empirical application to NASDAQ-listed energy stocks demonstrates the model's ability to model explosive dynamics in high-dimensional data by successfully identifying a significant common bubble during the 2008 financial crisis. Furthermore, the application showcases the model's forecasting capability during explosive episodes and the possibility to test for the emergence of a common bubble during the relevant common explosive episodes of the sample, relying on the testing framework of Blasques et al. (2024). In this sense, the model is able to identify the most significant common bubble of the sample up to several months in advance with respect to the date of the crash of the bubble.

The existing literature provides two main approaches for detecting and modeling bubbles in univariate settings. The first focuses on testing for the presence of a bubble analyzing global, and more recently, local non-stationarity within the process under study. For instance, Diba and Grossman (1988) proposes the use of unit root and cointegration tests to detect such explosive non-stationarity in the data. However, bubbles appear mostly as short-lived episodes and their repetitive expansion and collapse may render the time series globally stationary (Evans, 1991). The mismatch between global and local non-stationarity is addressed by Phillips et al. (2011), Phillips et al. (2015) and Phillips and Shi (2018) who develop supremum tests on recursive right-side unit root test statistics. These tests focus on local non-stationarities within a time series, allowing to identify the start and end date of explosive bubble events. Empirical studies following this approach find evidence of bubbles in the NASDAQ index, the U.S. housing price index, the price of crude oil (Phillips et al., 2011; Phillips and Yu, 2011), commodity (Etienne et al., 2014; Gutierrez, 2013) and real estate prices (Chen and Funke, 2013; Yiu and Jin, 2012). The testing procedures just described assess the presence of a bubble,they do not model the process itself, making it impossible to determine the bubble dynamics and perform any forecast. The second approach models bubbles using mixed causal non-causal autoregressive (MAR) models. This class of models characterizes a process through a forward-looking autoregressive specification with heavy-tailed innovations (Lanne and Saikkonen, 2011; Gouriéroux and Zakoïan, 2013). The future-driven dynamics provided by these non-causal models enable a buildup towards a future extreme shock, generating bubble episodes. The MAR framework has been applied to model and forecast financial bubbles across a wide range of processes displaying explosive behaviors, such as Nickel (Hecq and Voisin, 2021), NASDAQ (Gouriéroux and Zakoïan, 2017), Bitcoin (Hencic and Gouriéroux, 2015), and various commodity prices (Fries and Zakoian, 2019).

Recently, the issue of modelling common bubbles has caught more attention in the literature. The MAR framework extends naturally to multivariate settings through non-causal vector autoregressive models, as shown by Lanne and Saikkonen (2013). Common bubbles can then be modeled by allowing the non-causal matrix polynomial to be of reduced-rank as in Cubadda et al. (2023). A similar approach has been explored by Gourieroux and Jasiak (2017) and Davis and Song (2020), who relies on a backward looking representation with roots that are allowed to be explosive. These explosive roots identify common unobserved non-causal components, that can generate common bubbles. The limitation of multivariate MAR models is that this a class of stationary models, conflicting with the inherent non-stationarity of processes that generally experience bubbles, forcing researchers to rely on detrending techniques (Hecq and Voisin, 2023). Moreover, multivariate MAR processes are not scalable to high-dimensional contexts. Estimating a multivariate MAR model and using it to forecast is computationally problematic even for relatively small sample sizes, greatly limiting the applicability of this framework. A different perspective from outside the MAR literature in the context of common bubbles comes from Chen et al. (2022), that allow a high-dimensional time series to be driven by several factors, some of which are assumed to be driven by an autoregressive process with explosive roots. Their method focuses on testing for the presence of an explosive root and provides no insight in the dynamics of the factors. Although conceptually similar to the model presented in this paper, their approach does not allow to model or forecast the explosive dynamics and the evolution of the overall system during an explosive episode.

This paper introduces a novel dynamic factor model capturing how common extreme events propagate through a system, by modeling common factors as mixed causal non-causal processes. This framework is the first creating a bridge between the literature on dynamic factor models and the MAR literature. The proposed model extends the factor literature through the introduction of bubble episodes in the unobserved components and by establishing the properties of the estimates of such unobserved components in the context of heavy tails and explosive dynamics. This paper presents also two main improvements with respect to multivariate MAR models. First, the stationarity requirement in MAR models — often violated by the nature of the data — is relaxed, thanks to the flexibility of the proposed approach. Second, the feasibility issues for estimation and forecasting in high dimensions faced by the multivariate MAR are solved through the dimensionality reduction of the non-causal dynamics offered by the factor structure. Finally, while being designed to detect and model common bubbles, the model also accommodates idiosyncratic components that may experience their own locally explosive episodes, providing a comprehensive view of both shared and individual bubbles within the data.

The paper is organized as follows: Section 2 presents the Dynamic Factor model with common explosive dynamics, establishing the model's properties and discussing the relevant assumptions. Section 3 outlines the estimation strategy and the asymptotic properties of the estimators. Section 4 examines the finite sample properties of our estimator through a Monte Carlo simulation, and Section 5 presents an application to monthly stock prices in the energy sector.

2 The Model

This section discusses the proposed non-stationary mixed causal non-causal (MAR) factor model, first introducing mixed causal non-causal models in general and then then describing the novel factor model specification and the relevant assumptions.

Mixed causal non-causal models are a class of autoregressive models that allows the process to depend on its future, rather than only on its past. This feature allows these processes to exhibit locally explosive patterns. For a univariate time series $\{z_t\}_{t\in\mathbb{Z}}$ a MAR(l, s) process depends on l lags, and s leads, and, following Lanne and Saikkonen (2011) specification, has the form:

$$\psi(L^{-1})\phi(L)z_t = \varepsilon_t,$$

with $L^{-1}z_t = z_{t+1}$, $\psi(z) = 1-\psi_1 z-...-\psi_s z^s$, $\phi(z) = 1-\phi_1 z-...-\phi_l z^l$, respectively the non-causal (forward looking) and causal (backward looking) polynomials, and ε_t are iid innovations from a non-gaussian distribution. Like AR models, MAR have an infinite MA representation,

$$z_t = \sum_{h=-\infty}^{\infty} \varphi_h \varepsilon_{t+h}.$$

The dependence on the future, through the forward looking polynomial, and the non-gaussianity allow the process to generate explosive episodes. Under the assumption,

$$\phi(z) = 0 \quad \text{for} \quad |z| > 1 \quad \text{and} \quad \psi(z) = 0 \quad \text{for} \quad |z| > 1,$$

$\{z_t\}_{t\in\mathbb{Z}}$ is a stationary and ergodic process by Lanne and Saikkonen (2011). By having the unobserved common components experience this dynamic we can observe these explosive episodes propagate throught the system.

Let us now introduce the proposed non-stationary MAR factor model. Consider $\boldsymbol{y}_t = (y_{1,t}, ..., y_{n,t})'$ a n-dimensional vector of time series. Assume that the series in this vector share r common factors:

$$\boldsymbol{y}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{v}_t, \tag{1}$$

where $\boldsymbol{f}_t = (f_{1,t}, ..., f_{r,t})'$ is the $r \times 1$ vector of common factors and Λ is the $n \times r$ matrix of loadings. Moreover, $\boldsymbol{v}_t = (v_{1,t}, ..., v_{n,t})$ is the vector of idiosyncratic components. The novelty of the proposed approach lies in allowing any of the $\boldsymbol{f}_t$, and $\boldsymbol{v}_t$, to experience explosive dynamics. The next section describes the modeling of these dynamics in details.

Common Factors

Let the common factors $\boldsymbol{f}_t$ in (1) be represented as,

$$\Delta \boldsymbol{f}_t = \boldsymbol{\Gamma}(L)\boldsymbol{\varepsilon}_t, \tag{2}$$

with $\Gamma(L) = \sum_{h=-\infty}^{\infty} \boldsymbol{\Gamma}_h L^{-h}$ being a two-sided sum of $r \times q$ matrices, with $q \leq r$ and $\boldsymbol{\varepsilon}_t$ a $q-$dimensional vector of iid noise.

Assumption 1

1. *$\boldsymbol{\varepsilon_t} = (\varepsilon_{1,t}, ..., \varepsilon_{q,t})$ is a vector of strong noise, $\boldsymbol{\varepsilon}_t$ is independent of $\boldsymbol{\varepsilon}_{t+k}$ for all $k \neq 0$ and $\mathbb{E}\boldsymbol{\varepsilon}_t = 0$, $\mathbb{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \mathbb{I}_r$.*

2. *The matrix polynomial $\boldsymbol{\Gamma}(L)$ is such that $\sum_{h=-\infty}^{\infty} h|\Gamma_{i,h}| \leq M_2 < \infty$ for all i and for $\Gamma_{i,h}$ the $1 \times q$ row of $\boldsymbol{\Gamma}_h$. Finally, $\boldsymbol{\Gamma}_h$ is a diagonal matrix for all $h < 0$.*

Assumption 1 extends the non-stationary common factors assumption by Barigozzi et al. (2021) to factors with mixed causal non-causal dynamics. The common factors

$\boldsymbol{f}_t$ are allowed to be I(1), coherently with what observed in the factor literature (Barigozzi et al., 2021) and in the mixed causal non-causal literature (Hecq and Voisin, 2023; Blasques et al., 2023). Assumption 1, however, allows also factors to be stationary, not restricting the model only to the non-stationary setting.

The common factors described in (2) can be divided in two categories, the the purely causal factors and the non-causal factors. The second category includes the factors that do not show any type of anticipative behavior and exhibit a more standard backward-looking dynamic. The second ones are the factors that exhibit explosive behaviors and are driven by a forward-looking dynamics. Define,

$$\Delta f_{i,t} = \sum_{h=0}^{\infty} \Gamma_{i,h} \boldsymbol{\varepsilon}_{t-h}, \tag{3}$$

a purely causal (backward-looking) factor, and,

$$\Delta f_{i,t} = \sum_{h=-\infty}^{\infty} \Gamma_{i,h} \boldsymbol{\varepsilon}_{t-h}, \tag{4}$$

a factor exhibiting non-causal (forward-looking) dynamics, or equivalently an explosive factor. A factor with purely backward looking dynamics is defined as a factor for which $\Gamma_{i,h} = \mathbf{0}$ for all $h < 0$ in model (2). At this stage, the number of both the causal and non-causal factors is assumed to be known. The identification strategy for the different types of factor will be presented in a later section.

Assumption 1.2 restricts the non-causal structure allowed for our model. In most scenarios the common factors describe few sizeable locally-explosive episodes implying a simple non-causal structure. In general one or at most two factors with non-causal dynamics are expected.

Remark 1 *While the common factors are assumed to follow (2), this is not the proposed estimating equation. As discussed in Hecq and Voisin (2023) the first differences of a process with non-causal dynamics lose their anticipative behavior and result in a series of spikes. This makes important to model the non-stationary part of the process if we want to estimate a MAR process in a I(1) setting. For this reason (2) is a representation assumption.*

Let us discuss the estimating equations for the factors. If $\{f_{i,t}\}_{t\in\mathbb{Z}}$ is a common explosive factor as in (4), let us assume it follows a MAR process with a stochastic trend with exogenous variables. This is an extension of the model proposed in Blasques et al. (2023) designed to allow for interaction between the factors.

Assumption 2 *Let $\{f_{i,t}\}_{t\in\mathbb{Z}}$ be a common explosive factor in (4), then,*

$$\begin{aligned} f_{i,t} &= \mu_{i,t} + z_{i,t} + \sum_{j\neq i} \phi_{ij}(L) f_{j,t}, \\ \mu_{i,t+1} &= \mu_{i,t} + \beta_i \varepsilon_{i,t-s}, \quad \psi_i(L^{-1})\phi_i(L) z_{i,t} = \varepsilon_{i,t}, \end{aligned} \tag{5}$$

where $\{\mu_{i,t}\}_{t\in\mathbb{Z}}$ is the stochastic trend component, $\{z_{i,t}\}_{t\in\mathbb{Z}}$ is the bubble component and the term $\sum_{j\neq i} \phi_{ij}(L) f_{j,t}$ drives the dependence between factors. The stochastic trend component $\{\mu_{i,t}\}_{t\in\mathbb{Z}}$ has a random walk structure where β_i drives the strength of the update. The lag $t-s$, with s the order of the lead polynomial $\psi_i(\cdot)$, ensures that the right hand side of the update equation for the stochastic trend includes only terms up to time t. The bubble component $\{z_{i,t}\}_{t\in\mathbb{Z}}$ follows a univariate MAR process, with $\psi_i(L^{-1})$ and $\phi_i(L)$ the non-causal and causal polynomials respectively.

For what concerns the causal factors, the model described in (3) allows for any ARIMA and, under the assumptions in Barigozzi et al. (2020), VECM specification. This is not surprising as the specification in (3) falls in the setting of Barigozzi et al. (2021) under Assumption 1. Since the focus of this paper is on modelling common bubbles and the risk connected with this type of extreme events, the dynamic of the causal factors will be left unspecified for the rest of the paper and only the relevant assumptions for the correct estimation of the factors and their parameters will be discussed.

The remainder of the section describes features and assumptions relevant for the proposed factor specification and its estimation. The potential lower rank $q \leq r$ of $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ in (2) is important because it allows for flexible specifications of the factor dynamics. A lower dimension of the error vector is coherent with the reduced rank static representation from a dynamic factor model, as in Bai and Ng (2007), allowing for dynamic propagation of the explosive episodes.

Remark 2 *The dimension $q \leq r$ of $\boldsymbol{\varepsilon}_t$, potentially lower than the factors, is crucial to allow for a dynamic factor structure. In a stationary setting, for simplicity of the representation, we can have:*

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\lambda}_1 f_t + \ldots + \boldsymbol{\lambda}_p f_{t-p} + \boldsymbol{v}_t \\ &= \begin{bmatrix} \boldsymbol{\lambda}_1 & \ldots & \boldsymbol{\lambda}_p \end{bmatrix} \boldsymbol{f}_t + \boldsymbol{v}_t, \end{aligned}$$

with $\boldsymbol{f}_t = (f_t, ..., f_{t-p})'$ *a stationary MAR*$(1,1)$ *factor:*

$$\left(\mathbb{I}_p - \begin{bmatrix} \psi & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 0 \end{bmatrix} L^{-1}\right)\left(\mathbb{I}_p - \begin{bmatrix} \phi & \dots & 0 \\ 1 & \ddots & 0 \\ 0 & \dots & 0 \end{bmatrix} L\right)\begin{bmatrix} f_t \\ \vdots \\ f_{t-p} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \varepsilon_t.$$

Note that this factor specification allows the model to capture explosive events potentially distributed over time. By doing so it is possible to identify leading processes that experience the bubble first and understand the temporal dynamics of the bubble diffusion.

The next assumptions discusses the necessary elements for the estimation of the presented common factors structure.

Assumption 3

1. $\boldsymbol{\varepsilon_t} = (\varepsilon_{1,t}, ..., \varepsilon_{q,t})$ *belongs to a distribution with regularly varying tail probability with index* $\alpha > 2$*.*

2. *The covariance matrix* $\Sigma_{\Delta f} = \mathbb{E}\big[\Delta \boldsymbol{f}_t \Delta \boldsymbol{f}_t'\big]$ *is a diagonal matrix of rank* r *with distinct elements on the diagonal.*

Assumption 3.1 requires the errors to follow an heavy-tailed distribution, differently from similar models in the literature, see Barigozzi et al. (2021). The assumption implies that, if ε_{it} has regularly varying tail probability with index α,

$$P(|\varepsilon_{it}| > x) = x^{-\alpha} L(x),$$

with $L(x)$ a slowly varying function. Moreover note that $\mathbb{E}|\varepsilon_{it}|^n = \infty$ for $n > \alpha$. This implies that Assumption 3.1 requires only the variance to be finite. The heavy tail assumptions is crucial for non-causal models. Any mixed causal non-causal process, or purely non-causal, has a purely causal representation, for example a MAR(1,1) process can be represented as an AR(2) where the errors are an all-pass process, see Fries and Zakoian (2019). Being able to disentangle the two is a challenge that the researcher always faces in the MAR literature as these are equivalent up to second order moments. This can be achieved only under non-gaussianity assumption (Lanne and Saikkonen, 2011). Moreover the extreme values in the tail of the distribution are what drives the bubble dynamics, as these events generate from a build up towards a sizeable shock. Generally assumptions on the errors of the common factors require at least four moments, see Barigozzi et al. (2021) and Bai (2004). An exception is Barigozzi et al. (2024) where for a similar non-stationary factor structure driven by a distribution in the domain of attraction of a stable law no moment is required for the convergence of PCA estimates. While a similar approach could be used in the setting presented by this paper, the second stage estimation, following Blasques et al. (2023), requires at least two moments so the setting is restricted to $\alpha > 2$.

For what concerns the loadings the following structure is assumed:

Assumption 4 *The loadings matrix* Λ *is such that* $\Lambda'\Lambda \to I_r$ *as* $n \to \infty$ *and* $||\boldsymbol{\lambda}_i|| \leq C < \infty$ *for all* i*.*

The estimation strategy for the factor dynamics in a non-stationary setting follows a block strategy. The models for the non-stationary factors that show non-causal dynamics are estimated univariately using an extension of the MAR model with a stochastic trend (MARST) from Blasques et al. (2023). The advantage of the MARST model is the possibility to jointly estimate the non-stationary and the non-causal components of the explosive factors. Such a model falls in the setting considered by Assumption 1.

Proposition 1 *Let* $f_{i,t}$ *follow a MAR process with a stochastic trend with exogenous variables as defined in (5). Then* Δf_i *has a double sided infinite MA representation compatible with Assumption 1.1 and 1.2.*

Note that the approach to model common non-causal dynamics is coherent with the models considered in Cubadda et al. (2023) and Gourieroux and Jasiak (2017). The mixed causal non-causal VAR specifications considered by these authors allow to represent the non-causality through dependence on a lower dimensional process, respectively through a reduced rank of the non-causal matrix coefficient in Cubadda et al. (2023) and through a decomposition around the explosive roots of the VAR coefficient in Gourieroux and Jasiak (2017). The proposed approach has the advantage of extending to non-stationary high dimensional settings and allowing also for idiosyncratic non-causal dynamics.

Idiosyncratic Components

Similarly, the idiosyncratic components v_{it}, for $i = 1, ..., n$, are modeled accounting for potential non-stationarity and idiosyncratic bubbles.

$$\Delta v_{it} = \sum_{h=-\infty}^{\infty} \psi_{ih}\eta_{it}, \quad \text{for all } i = 1, ..., n. \tag{6}$$

Then we assume:

Assumption 5

1. *$\boldsymbol{\eta}_t = (\eta_{1t}, ..., \eta_{nt})$ is a vector of strong noise, in other words $\boldsymbol{\eta}_t$ is independent of $\boldsymbol{\eta}_{t+k}$ for all $k \neq 0$ and $\mathbb{E}\boldsymbol{\eta}_t = 0$, $\mathbb{E}\boldsymbol{\eta}_t\boldsymbol{\eta}_t' = \Sigma_\eta$. Moreover $\mathbb{E}\big[|\eta_{it}|^a|\eta_{it}|^b\big] < \infty$ with $a + b = 4$.*

2. *Σ_η is positive definite and $\max_j \sum_{i=1}^n \big|\mathbb{E}\big[\eta_{it}\eta_{jt}\big]\big| < M$.*

3. *$\sum_{h=-\infty}^{\infty} h|\psi_{ih}| < \infty$ for all i.*

4. *ε_{it} and η_{js} are independent for all $i = 1, ..., r$, $j \in \mathbb{N}$ and $t, s = 1, ..., T$.*

The points in Asssumption 5 are in line with what assumed by Barigozzi et al. (2021). The model in (6) is coherent with some of the idiosyncratic component being I(1) and presenting locally explosive behaviors. The first assumption tightens the moment condition we imposed on the common factors requiring at least four moments. This is a limit on the idiosyncratic bubbles that such a model can generate. However this is coherent with what observed in practice. In multivariate settings the most relevant locally explosive episodes usually spread across series and become common episodes. The second assumption limits the amount of cross-sectional correlation that the model can accommodate.

3 Parameter Estimation

The model is estimated using a two-stage procedure. In the first stage, the common factors and the loadings are estimated through PCA on the first differences of the data. In the second stage the non-causal and the purely causal part of the factors are estimated by Maximum Likelihood. The non-causal and the backward-looking components of the factors are estimated separately, this allows the estimation to take into account non-stationarity and non-causality jointly. The next section describes the estimation procedure assuming that the number of common factors, r, and of the non-causal factors, are known, and then discusses how to estimate these quantities.

First stage estimation

Consider the first differences of the data,

$$\Delta \boldsymbol{y}_t = \Lambda \Delta \boldsymbol{f}_t + \Delta \boldsymbol{v}_t.$$

Since the loadings Λ are the same as in the non-stationary representation of the model in (1), these can be estimated through PCA on the differenced model following Barigozzi et al. (2021). The loadings estimator $\hat{\Lambda}$ is defined by,

$$\hat{\Lambda} = \sqrt{n}\hat{Q}^{\Delta y},$$

with $\hat{Q}^{\Delta y}$ the $r \times n$ matrix of standardized eigenvectors corresponding to the the first r eigenvalues of the sample covariance matrix of the first differenced data $\hat{\Sigma}^{\Delta y}$.

The factors estimate is then defined as,

$$\hat{\boldsymbol{f}}_t = \frac{1}{n}\hat{\Lambda}\boldsymbol{y}_t.$$

Proposition 2 *Define* $\vartheta = \max\left(T^{-\delta}, n^{-1}\right)$ *with* $\delta \leq (\alpha-2)/\alpha$ *such that* $\mathbb{E}|\varepsilon_t|^{\alpha} < \infty$. *Under assumptions 1-5 there is a* $r \times r$ *diagonal matrix* S *with entries of* 1 *and* -1 *such that the PCA estimate* $\hat{\Lambda}$ *of the loadings are,*

$$\left|\left|\hat{\Lambda} - S\Lambda\right|\right| = O_p(\vartheta).$$

Moreover define the estimated factors $\hat{\boldsymbol{f}}_t = \frac{1}{n}\hat{\Lambda}\boldsymbol{y}_t$ *then,*

$$\frac{1}{\sqrt{T}}\left|\left|\hat{\boldsymbol{f}}_t - S\boldsymbol{f}_t\right|\right| = O_p(\vartheta).$$

The proof for Proposition 2 follows Barigozzi et al. (2021). The main difference comes from the different convergence rate obtained by allowing for heavy-tailed distributions. The proof relies on the result of convergence for the covariance of infinite

MA processes with errors with regularly varying tail distributions from Davis and Resnick (1986). The absence of moments up to the fourth order influences how fast the sample covariance matrix $\hat{\Sigma}_{\Delta y}$ converges[1]. Note that if the innovations $\boldsymbol{\varepsilon}_t$ have regularly varying tail probability with index $\alpha \geq 4$, then $\delta = 1/2$, redirecting to the same setting as Barigozzi et al. (2021). While there exists results in the literature for heavy tailed non-stationary factors, see Barigozzi et al. (2024), this paper is the first, up to the knowledge of the author, in following this approach to establish consistence of the factors and loadings estimates.

Identifying the number of explosive factors

So far, the number of common factors r and of the non-causal factors has assumed to be known. This section presents results on the estimation of these two quantities, the estimation of the number of non-stationary factors and the identification of explosive factors.

The first step is to identify the number of common factors. For this purpose it is possible to rely on the standard methodology in the factor literature, see Bai and Ng (2002) and Onatski (2009).

The second step is to estimate the number of non-stationary factors. Given a consistent estimate of the spectral density of $\Delta \boldsymbol{y}_t$ and the corresponding eigenvalues $\hat{\nu}_j^{\Delta y}(\chi)$, for $j = 1, ..., r$, the number of non-stationary factors $\hat{m}$ can be estimated as,

$$\hat{m} = \arg \min_{h=1,\dots,m_{max}} \left[\log \left(\sum_{j=h+1}^{n} \hat{\nu}_j^{\Delta y}(0) \right) + hp(n,T) \right],$$

for $p(n, T)$ some penalty term, as proposed by Barigozzi et al. (2021).

Finally the last step is to estimate the number of the explosive factors. Davis and Song (2020) suggest to use BIC to identify the number of lags p and then analyze the roots of the estimated polynomial to see how many roots are outside the unit circle corresponding to a number of non-causal components. This approach, however, is available only for the stationary setting. If the common factors are non-stationary we need a different approach. The methodology proposed by this paper is to use factor by factor BIC to identify the factors presenting non-causal dynamics. Note that while it may be intuitive to fit the model on the first differences, in the case of non-causal

[1] As shown in Lemma 2.

dynamics the first differences lead to sever underestimation of the total order p. This is due to the fact that while it is possible to express the model as a mixed causal non-causal ARIMA, the most appropriate representation of these stationary processes is that of a fundamental value stochastic trend plus a bubble component. To estimate the correct total number of lags k this paper proposes using a BIC approach on a local level model plus AR(k) model. The procedure showed to perform really well in estimating the true k in a simulation study. Once p is established it is possible to estimate all combination of MARST(r, s) models such that $r + s = p$ and select the best combination, similarly to what is done in the stationary setting.

Another approach to detect non-causality in the factors is to use the test from Phillips et al. (2011) and then in case of evidence of explosive roots fit fit the MARST model. This second procedure is more robust in small samples.

Another possible evaluation for the model specification is the extreme clustering approach proposed by Fries and Zakoian (2019). Any misspecification of the (r, s) order, for example fitting a MAR(1,1) on a process that is purely non-causal MAR(0,2), leads to the errors being not iid anymore. Fries and Zakoian (2019) suggest testing for extremes clustering in the residuals, as only the correct specification has a representation with iid noise. Finally the last possible approach, to be developed in a further stage, is to rely on the higher-order cumulants. Spectrum and bispectrum of the factors carry information about non-causality as shown in Hecq and Velasquez Gaviria (2024). The problem of this approach is that it requires the innovation that drives the process to have at least three moments.

Second stage estimation

The parameters of the dynamic specification of the factors and the idiosyncratic components are estimated by Maximum Likelihood in the second stage. The parameters of the models for the factors that present non-causal dynamics are estimated separately from the ones of the purely causal factors. Estimating non-causal factors separately allows the model to address non-causality and non-stationarity jointly. Consistency of the MLE estimator is shown under the previous result of consistency of the factors estimates. Defining $\boldsymbol{\theta}_{i,0}, \gamma_{j,0}$ the true parameter vectors for respectively the $i-th$ factor and the $j-th$ indiosyncratic component,

Theorem 1 *Under assumptions 1-5 the MLE $\hat{\boldsymbol{\theta}}_i$ of the parameters for the i-th common factor model, for $i = 1, ..., r$, are consistent estimators of the true parameter* *vectors* $\boldsymbol{\theta}_{i,0}$*:*

$$||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i,0}|| \xrightarrow{p} 0, \quad \text{as } n, T \to \infty.$$

Moreover the MLE $\hat{\boldsymbol{\gamma}}_j$ *for the* $j-th$ *idiosyncratic component model is consistent for the true parameter vector* $\boldsymbol{\gamma}_{j,0}$

$$||\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0}|| \xrightarrow{p} 0, \quad \text{as } n, T \to \infty.$$

Theorem 1 ensures that the parameters driving the factor dynamics can be estimated consistently. This result guarantees the possibility to model, forecast and test the non-causal dynamics of the explosive factors, that is the ultimate goal of this paper. The proof for Theorem 1 follows the same steps of Blasques et al. (2023) while ensuring that the estimation of factors does not influence the consistency result.

Remark 3 *Theorem 1 does not require the model to be correctly specified. The only requirements are the ones from assumptions 1-4, that is being able to estimate consistently the factors. This means that the explosive factor could have a different dynamic for the stochastic trend while still maintaining its ability to identify the non-causal dynamics.*

Such a scenario will be part of the simulation study in the next section.

Remark 4 *The proposed approach focuses on the setting with non-stationary explosive factor. It is important to note that the approach remains valid if the common factors are stationary or if the explosive factors are stationary. The challenge rises by trying to combine non-stationarity and non-causal dynamics. In a stationary setting one can rely on the estimation procedures available in the MAR literature, see the AML of Lanne and Saikkonen (2013) and Davis and Song (2020) or the GCov of Gourieroux and Jasiak (2017).*

4 Simulation Study

This section evaluates the performance of the proposed model for factors and parameters estimation, with specific focus on common MAR process parameters. The simulation study presents n-dimensional vector of time series $\boldsymbol{y}_t^s = (y_{1,t}, ..., y_{n,t})$ of length T for a number of simulations $s = 1, ..., 200$, generated by the model,

$$\boldsymbol{y}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{v}_t,$$

$$\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ \Delta f_{3,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ 0 \\ 0 \end{pmatrix} + \begin{bmatrix} 0 & 0.4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.8 \end{bmatrix} \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ \Delta f_{3,t-1} \end{pmatrix} + \begin{bmatrix} z_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t}, \end{bmatrix}$$

and,

$$\begin{aligned} &\mu_{1,t+1} = \mu_{1,t} + \beta \kappa_t, \quad \kappa_t \sim N(0, \sigma) \\ &\text{and} \quad (1 - \phi L)(1 - \psi L^{-1}) z_{1,t} = \varepsilon_{1,t}, \quad \varepsilon_t \sim t_\nu \end{aligned}$$

for different values of n, T.[2] The idiosyncratic components $\boldsymbol{v}_t$ are divided uniformly into three data generating processes, respectively Gaussian noise,

$$v_{j,t}^{(1)} = \eta_{j,t} \sim N(0, \sigma_j),$$

a Gaussian random walk process,

$$v_{j,t}^{(2)} = v_{j,t-1}^{(2)} + \eta_{j,t}, \quad \eta_{j,t} \sim N(0, \sigma_j),$$

and an mixed causal non-causal process with a non-stationary stochastic trend process,

$$\begin{aligned} v_{j,t}^{(3)} = \mu_{j,t} + z_{j,t}, \quad \mu_{j,t} = \mu_{j,t-1} + \xi_{j,t} \sim N(0, \sigma_j) \\ \psi_j(L^{-1}) \phi_j(L) z_{j,t} = \eta_{j,t}, \quad \eta_{j,t} \sim t(\nu_j). \end{aligned}$$

Note that the non-stationary explosive dynamics is simulated as a MAR process with Gaussian random walk stochastic trend. The choice of a misspecified data generating process for the non-stationary explosive factor is motivated by the will to create a more realistic setting and to showcase the performance of the model in case of incorrect specification.

[2] Appendix D presents examples of the generated samples for the different data generating processes.

The considered data generating process for $\boldsymbol{y}_t$ creates an environment of non-stationary common factors, with one of the factors presenting explosive dynamics and heavy tailed errors. The idiosyncratic components are a mix of stationary and non-stationary processes with some of these presenting explosive behavior. The first goal of this simulation is to show how the first stage estimation behaves in small samples in the presence of heavy tails and explosive dynamics and how this performance changes as n and T grow. The second goal is to show that both the first stage estimation and the misspecification of the stochastic trend component have a limited influence on the second stage estimation. An important thing to highlight is that estimating common (or idiosyncratic) non-causal dynamics for the cross-sectional dimensions considered in this simulation and later on in the application is not computationally feasible with current approaches in the mixed causal non-causal literature. An element left for future study is to design more complex interaction between the factors as for example done in Barigozzi et al. (2021).

The RMSE for the estimated factors and loadings are presented, respectively

$$RMSE_{f_j} = \sqrt{\sum_{i=1}^{T} \left(\hat{f}_{i,j} - f_{i,j}\right)^2}, \quad \text{and} \quad RMSE_{\Lambda_j} = \sqrt{\sum_{i=1}^{N} \left(\hat{\lambda}_{i,j} - \lambda_{i,j}\right)^2},$$

where $\boldsymbol{f}_j$ and Λ_j are respectively the true j-th factor and the j-th column of the true loading matrix Λ. Table 1 reports the results in terms of RMSE for the presented data generating process.

RMSE for the estimated factors and loadings

	$T = 250$			$T = 500$		
RMSE	$n = 25$	$n = 50$	$n = 100$	$n = 25$	$n = 50$	$n = 100$
Loadings_1	0.09	0.08	0.08	0.07	0.06	0.05
Factor_1	1.19	0.84	0.60	1.17	0.82	0.58
Loadings_2	0.31	0.25	0.22	0.24	0.18	0.17
Factor_2	1.52	1.10	0.82	1.38	0.97	0.74
Loadings_3	0.41	0.28	0.22	0.33	0.20	0.17
Factor_3	1.66	1.16	0.86	1.50	1.03	0.76

Table 1: RMSE for the Loadings and Factor estimates for the three common factors considered in the data generating process.

The RMSE presented in Table 1 shows how the estimation improves, as expected, as T and n grow. In the second stage the parameters of the non-stationary explosive factor are estimated, with emphasis in Table 2 on the respectively causal and non-causal parameters.

Estimated Parameters for the Explosive Factor model

	$T = 250$	$T = 500$	$T = 1000$
$\hat{\boldsymbol{\phi}}$	0.61	0.62	0.64
$(\phi_0 = 0.7)$	(0.09)	(0.07)	(0.04)
$\hat{\boldsymbol{\psi}}$	0.82	0.83	0.82
$(\psi_0 = 0.8)$	(0.08)	(0.06)	(0.04)
$\hat{\boldsymbol{\phi}}_x$	0.35	0.36	0.37
$(\phi_{x,0} = 0.4)$	(0.11)	(0.07)	(0.06)

Table 2: Mean and standard deviation of second stage estimates for the causal (ϕ) and non-causal (ψ) parameters for a MAR/MARST model with $\phi_0 = 0.7, \psi_0 = 0.8$.

The estimates improve as T grows both in terms of their closeness to the true value and in terms of variance. From the table it is possible to observe that the misspecification creates an attenuation bias, especially in the backward looking part of the model. This is coherent with what happens in the detrending literature for MAR models, see Hecq and Voisin (2021) and Blasques et al. (2023). However the non-stationary MAR model presented from Blasques et al. (2023) and its multivariate extension here presented solve the issue by disentangling the process in its total backward-looking component and the purely non-causal component. In case of misspecification the total backward-looking component, being the sum of all causal parts of the process and the stochastic trend offsets the estimation bias in the individual parts correctly identifying the purely non-causal component, that is the relevant component in terms of forecasting and testing for explosive dynamics. This robustness is due to the desirable properties of the observation-driven filters framework. This decomposition into total backward-looking component and purely non-causal component is discussed more in detail in the application section.

5 Application

In this section the proposed model is applied to monthly price of 61 stocks of companies listed in the energy sector from the NASDAQ. [3] The dataset consists of 353 observations spanning the period from October 1994 to February 2024. Figure 1 gives an overview of the data.

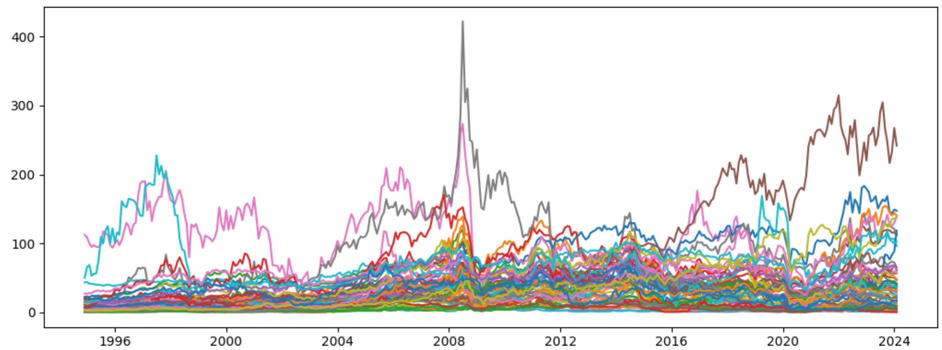


Figure 1: Monthly price of stocks from the energy sector.

Figure 2 shows the data scaled dividing each series by its standard deviation to avoid any scale effect, highlighting the comovement experienced between 2008 and 2009 and its pervasiveness through all the series of the sample. The proposed application has been considered also for the log of the data and the results are qualitatively the same in terms of factor estimates and explosiveness of the underlying components. In the application, the factors, loadings and second-stage estimates are re-estimated on an expanding window. These estimates are then used to test for forecasting performance using one-step-ahead forecasts on pseudo-out-of-sample test sample and to assess the risk of the emergence of a bubble in the last observation of the sample, the one observed at the moment of estimation.

[3]The data are obtained through yfinance API.

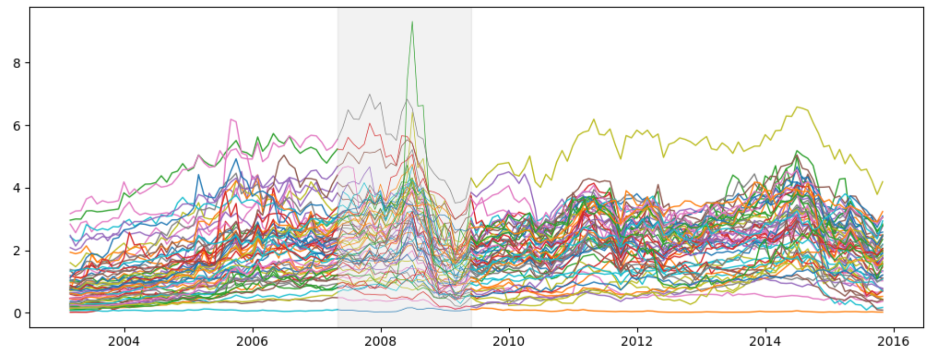


Figure 2: Detail of the period between 2004 and 2016 for the scaled series.

The first step in the estimation is to estimate the common factors and determine $\hat{r}$ the number of common factors. The number of factors is estimated to be $\hat{r} = 5$, using the methodology from Alessi et al. (2010). The standard methodology from Bai and Ng (2002) seemed to overestimate the number of common factors selecting always $\hat{r} = r_{max}$, probably due to the relatively small sample size. Moreover the factors are all identified as non-stationary common trends according to the procedure of Barigozzi et al. (2021).

The second step is then to identify which factors exhibit explosive dynamics, identifying the causal and non-causal orders of the factors according to procedure described in Section 3. The model selection steps identify only two factors as having non-causal dynamics. The remainder of the section focuses on the first explosive factor, that is responsible for the sizeable explosive episode in 2008.

The explosive common factor

This section concentrates on the first explosive common factor, showcasing a forecasting exercise of the common explosive dynamics and testing for the emergence of a common bubble relying on the methodology from Blasques et al. (2024). Figure 3 shows the explosive factor fitted by the first stage estimation.

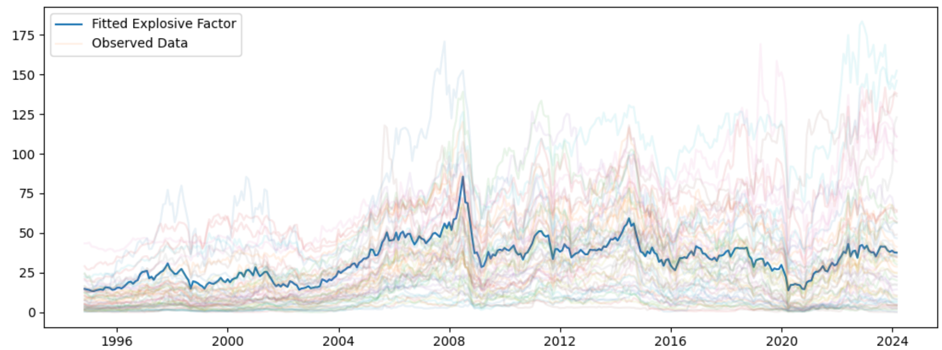


Figure 3: The explosive principal component obtained from the first stage of the estimation procedure.

For illustrative purposes the factor estimated in the first stage showcased by Figure 3 is obtained by applying PCA on the whole sample. Interestingly the re-estimation of the factors, especially before and after the sizeable common episode in 2008-2009, changes the results of the first stage estimation for some of the factors. However, the explosive factor in Figure Figure 3, that is central for this application, remains consistent over time, before, during and after the common explosive episode.

Over the expanding window re-estimation also the second stage estimates are influenced by the sizeable explosive episode of 2008-2009. This is a phenomenon often observed in MAR models, especially in small samples. The first explosive episode, or in this case the first explosive episode of a certain magnitude, influences the parameter estimates. In this case at the beginning of the test sample the model selection step selects only one non-causal root. While the process is experiencing the central explosive episode the number of selected non-causal roots switches to a higher order and then stays consistent for the rest of the sample. The final total order of the causal and non-causal polynomials selected by model selection for this estimated factor is $\hat{p} = 5$. Different orders of the MARST model are fitted to determine the causal and potential non-causal order of the MAR part of the process. In the final model selection step, the MAR(2,3) is preferred over the others.

	Likelihood	AIC	BIC
MAR(1,4)	- 332.74	683.91	718.70
MAR(2,3)	**- 326.36**	**670.73**	**705.53**
MAR(3,2)	- 331.29	680.58	715.38
MAR(4,1)	- 327.03	672.06	706.68

Table 3: Likelihood and Information Criteria for Model Selection. The results highlight the MAR$(2, 3)$ as the preferred specification.

When fitting the model it is possible to obtain two components, the purely non-causal component and the total backward-looking component, which this paper refers also to as the pseudo-fundamental value. The pseudo-fundamental value fv_t, defined as,

$$fv_t = \hat{\mu}_t + \hat{\phi}^*(L)\big(\hat{f}_t - \hat{\mu}_t\big), \quad \text{with} \quad \phi^*(L) = 1 - \phi(L) \tag{7}$$

represents the value at which the bubble collapses after the crash. The purely non-causal process reverts sharply to the unconditional mean after an explosive episode, letting the total process collapse to its pseudo-fundamental value. Importantly, by construction, this total backward-looking component is the conditional mean of the process, conditional on $t-1$, around which the purely non-causal component realizes. What distinguishes the fitted pseudo-fundamental value from its theoretical counterpart (the true fundamental value) is that the proposed quantity includes part of the bubble through the causal part of the process. In other words, after the crash of a bubble, also the pseudo-fundamental value goes through a deflation process, but this is predictable as it happens at deterministic rate, more specifically the rate of the causal polynomial $\phi(L)$.

The purely non-causal component is instead the forward-looking part of the process. It is the filtered version of the process defined as,

$$u_t = \sum_{h=0}^{\infty} \gamma_h \varepsilon_{t+h}.$$

This is the component that is responsible for the explosive dynamics around the non-stationary conditional mean denoted by the pseudo-fundamental value. Importantly, this component does not necessarily represent a bubble. As shown in Hecq and Voisin (2021), the predictive density of such a process is actually the same of a symmetric heavy-tailed random variable as long as the process does not deviate from its unconditional mean. This means that outside of explosive episodes the process as a whole behaves as a heavy tailed ARIMA, an heavy tailed random variable around the non-stationary stochastic trend. As soon as the process starts to substantially deviate from the stochastic trend it will exhibit the peculiar features of MAR dynamics.

Figure 4 shows the estimated factor and the fitted pseudo-fundamental value, Figure 5 shows the filtered purely non-causal process for the period that goes from the start of the sample until 2016.

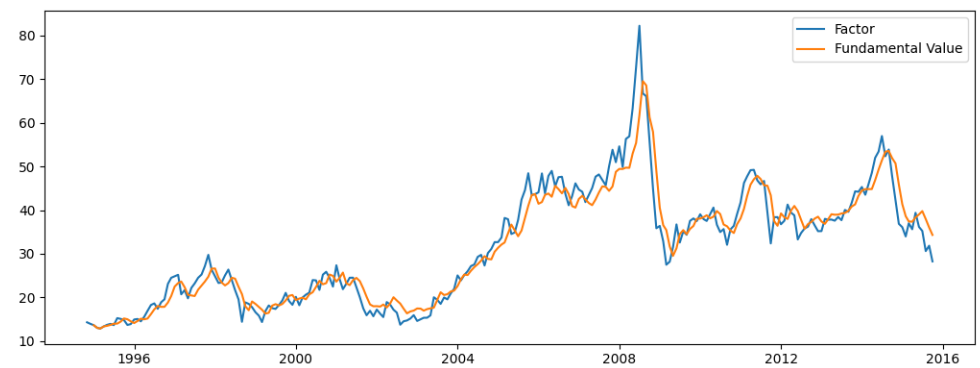


Figure 4: The estimated factor with non-causal dynamics and the pseudo-fundamental value for the estimated model.

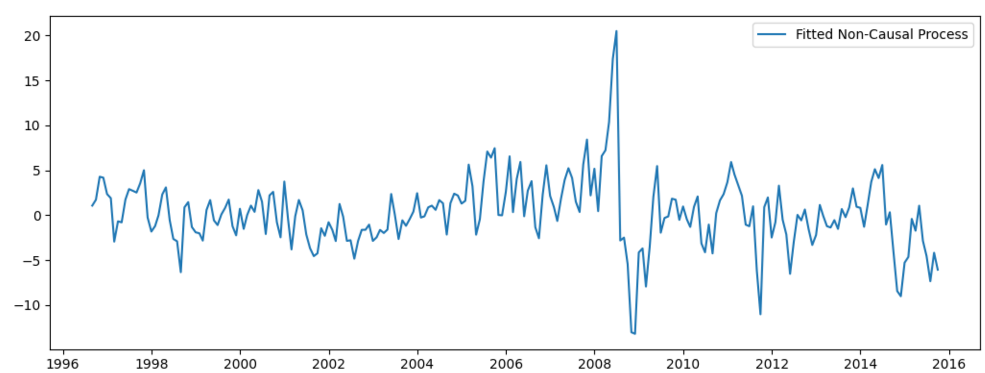


Figure 5: The filtered purely non-causal component of the process.

To test formally the predictive ability of such a model in the context of an explosive episodes a Diebold-Mariano test (Diebold and Mariano (1995)) is performed on the one step ahead out of sample predictive density[4]. The pseudo-out-of-sample test sample consists of the bubble episodes happening between September 2007 to October 2009 and from February 2014 to March 2015. To test predictive performance interval forecasts are used, using the Brier Score as test score.[5]

The model is compared against an heavy-tailed random walk, that is the outcome of the model selection steps for an heavy-tailed ARIMA$(p, 1, q)$, during the relevant bubble episodes of the sample. Figure 6 highlights the part of the sample used as test samples.

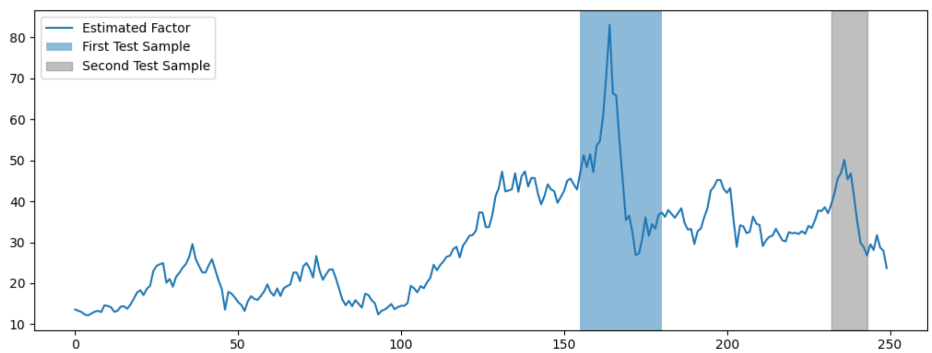


Figure 6: The fitted explosive factor and the test samples.

The forecasting performance is not evaluated in all the sample because the MAR with stochastic trend model, as mentioned before, is actually designed to fall back to a heavy-tailed ARIMA process in the absence of an explosive episode. The non-causal component acts as an heavy-tailed symmetric random variable when the process is not significantly deviating from the time varying conditional mean represented by the pseudo-fundamental value. Not surprisingly, the two methods not being significantly different in predictive power over the whole sample. On the other hand, the core of this empirical illustration is the correct prediction of the evolution of the time series during an explosive episodes. In the two considered test samples, highlighted

[4]The predictive density for the MAR model was produced using the simulation based approach of Lanne and Saikkonen (2013).

[5]More details on the testing procedure in the Appendix C.

in Figure 6 the MARST model significantly outperforms the alternative as shown in Table 4.

Models	First Test Sample		Second Test Sample	
	Model Score	Test Statistic	Model Score	Test Statistic
MARST	**0.85**	·	**0.65**	·
ARIMA	0.99	**-1.96**	0.88	**-1.76**

Table 4: Event prediction scores and test statistics against heavy-tailed ARIMA model.

The second part of the section revolves around testing for the emergence of a common bubble. The framework from Blasques et al. (2024) can be used on the fitted purely non-causal component of the process to test if the observed data at time t is compatible with a bubble of given size at an horizon $t + h$. By doing so the model is found to be able to identify the common episode as a bubble quite early in its development, as shown in Figure 7.

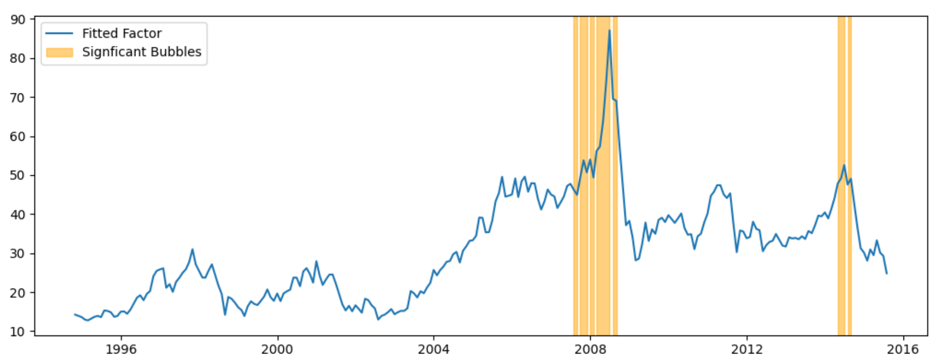
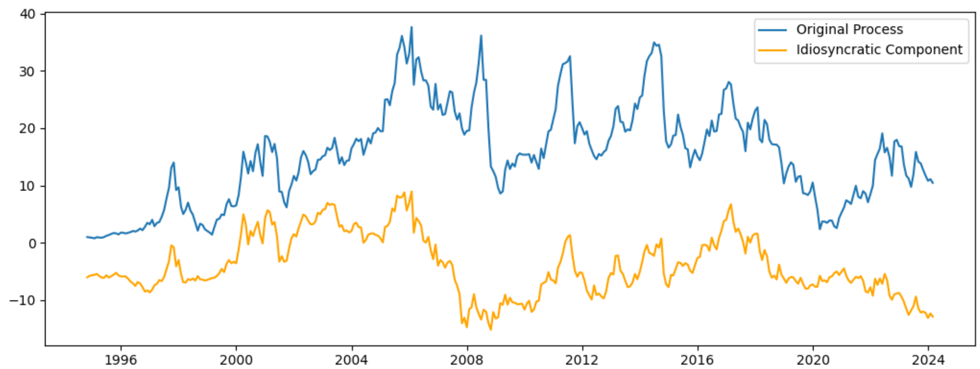


Figure 7: The highlighted observations are the ones that are compatible with a sizeable explosive episode at a short horizon.

Figure 7 highlights the observations that are tested to be significantly compatible with a sizeable bubble happening in a close time range. The most relevant result in this sense is that the start of the bubble happening in 2009 is being identified as compatible with an explosive episodes happening soon around 8 months before the crash of the bubble. This shows the empirical relevance of the proposed approach, that can provide valuable insights on the systemic risk in the considered dataset.

The Idiosyncratic Components

Finally, concerning the idiosyncratic components, it is relevant to understand how much of the explosive episode is explained by common factors and how much is left to idiosyncratic bubbles. In Figure 8 two out of the individual series y_{it} and their corresponding estimated $\hat{v}_{it}$ are presented.



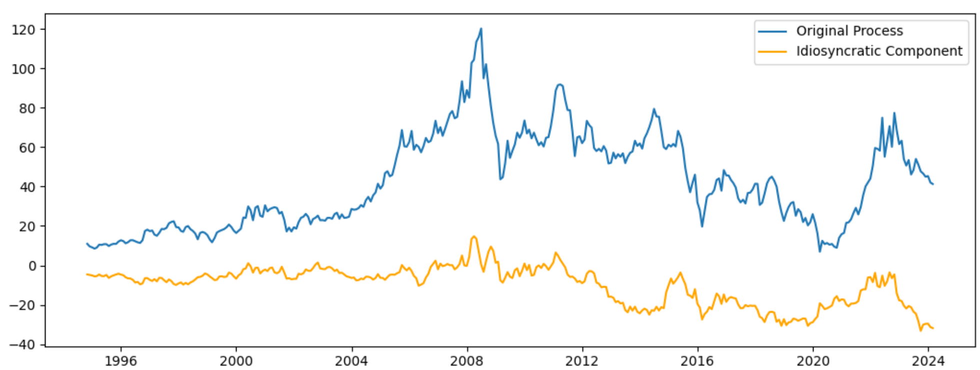


Figure 8: The figure shows the comparison between the original processes y_{it} and the estimated idiosyncratic processes $\hat{v}_{it}$ for PTEN (top) and DVN (bottom) prices.

Figure 8 shows two different cases, in the left figure the explosive episodes are mainly idiosyncratic and the filtered component still shows a non-causal dynamic. In the right figure most of the explosive episodes are the common bubbles captured by the common factors and the filtered component shows no relevant sign of explosive episodes.

6 Conclusions

This paper introduces a novel dynamic factor model that permits factors to exhibit locally explosive behavior, effectively capturing both common and idiosyncratic explosive episodes in a high-dimensional, non-stationary environment. This model is designed to address non-causal dynamics within non-stationary factors, but it is not restricted to the non-stationary setting. The paper explores the theoretical properties of the model and develops a method to estimate the unknown parameters, with a focus on the parameter of the non-causal model driving the potential common extreme events. Unlike existing models for common bubbles, the proposed approach is specifically designed for non-stationary processes and high-dimensional settings, enabling the detection and modeling of both common and idiosyncratic bubbles. The empirical application demonstrates that this method can reliably forecast during periods of high uncertainty, such as common explosive episodes, and serves as an effective tool in testing for the emergence of common bubbles.

References

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics Probability Letters*, 80(23):1806–1813.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):137–183.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Barigozzi, M., Cavaliere, G., and Trapani, L. (2024). Inference in heavy-tailed nonstationary multivariate time series. *Journal of the American Statistical Association*, 119(545):565–581.

Barigozzi, M., Lippi, M., and Luciani, M. (2020). Cointegration and error correction mechanisms for singular stochastic vectors. *Econometrics*, 8(1).

Barigozzi, M., Lippi, M., and Luciani, M. (2021). Large-dimensional dynamic factor models: Estimation of impulse–response functions with i(1) cointegrated factors. *Journal of Econometrics*, 221(2):455–482.

Blasques, F., Koopman, S., and Mingoli, G. (2023). Observation-driven filters for time-series with stochastic trends and mixed causal non-causal dynamics. Working Paper 23-065/III, Tinbergen Institute.

Blasques, F., Koopman, S., Mingoli, G., and Telg, S. (2024). A novel test for the presence of local explosive dynamics. Working Paper 24-036/III, Tinbergen Institute.

Blasques, F., van Brummelen, J., Koopman, S. J., and Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, 227(2):325–346.

Breid, F., Davis, R. A., Lh, K.-S., and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36(2):175–198.

Chen, X. and Funke, M. (2013). Real-time warning signs of emerging and collapsing chinese house price bubbles. *National Institute Economic Review*, 223(1):R39–R48.

Chen, Y., Phillips, P. C. B., and Shi, S. (2022). Common Bubble Detection in Large Dimensional Financial Systems*. *Journal of Financial Econometrics*, 21(4):989–1063.

Cubadda, G., Hecq, A., and Voisin, E. (2023). Detecting common bubbles in multivariate mixed causal–noncausal models. *Econometrics*, 11(1).

Davis, R. and Resnick, S. (1986). Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages. *The Annals of Statistics*, 14(2):533 – 558.

Davis, R. A., Pfaffel, O., and Stelzer, R. (2014). Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. *Stochastic Processes and their Applications*, 124(1):18–50.

Davis, R. A. and Song, L. (2020). Noncausal vector ar processes with application to economic time series. *Journal of Econometrics*, 216(1):246–267. Annals Issue in honor of George Tiao: Statistical Learning for Dependent Data.

Diba, B. T. and Grossman, H. I. (1988). Explosive rational bubbles in stock prices? *The American Economic Review*, 78(3):520–530.

Diebold, F. X. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–265.

Etienne, X. L., Irwin, S. H., and Garcia, P. (2014). Bubbles in food commodity markets: Four decades of evidence. *Journal of International Money and Finance*, 42:129–155.

Evans, G. W. (1991). Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, 81(4):922–930.

Fries, S. and Zakoian, J.-M. (2019). Mixed causal-noncausal ar processes and the modelling of explosive bubbles. *Econometric Theory*, 35(6):1234 – 1270. Cited by: 13; All Open Access, Green Open Access.

Gourieroux, C. and Jasiak, J. (2017). Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation. *Journal of Econometrics*, 200(1):118–134.

Gouriéroux, C. and Zakoïan, J.-M. (2013). Explosive bubble modelling by noncausal process. *CREST working paper*.

Gouriéroux, C. and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):737–756.

Gutierrez, L. (2013). Speculative bubbles in agricultural commodity markets. *European Review of Agricultural Economics*, 40(2):217–238.

Hecq, A. and Velasquez Gaviria, D. (2024). Non-causal and non-invertible arma models: Identification, estimation and application in equity portfolios. *Journal of Time Series Analysis*.

Hecq, A. and Voisin, E. (2021). Forecasting bubbles with mixed causal-noncausal autoregressive models. *Econometrics and Statistics*, 20:29–45.

Hecq, A. and Voisin, E. (2023). *Predicting Crashes in Oil Prices During The Covid-19 Pandemic with Mixed Causal-Noncausal Models*, volume 45B of *Advances in Econometrics*, pages 209–233. Emerald Publishing Limited.

Hencic, A. and Gouriéroux, C. (2015). Noncausal autoregressive model in application to Bitcoin/USD exchange rates. In *Econometrics of Risk*, pages 17–40. Springer, New York.

Krengel, U. (1985). *Ergodic theorems*. Walter de Gruyter, Berlin.

Lanne, M. and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3):1–32.

Lanne, M. and Saikkonen, P. (2013). Noncausal vector autoregression. *Econometric Theory*, 29(3):447–481.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447 – 1479. Cited by: 215.

Phillips, P. C., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International economic review*, 56(4):1043–1078.

Phillips, P. C. and Shi, S. P. (2018). Financial bubble implosion and reverse regression. *Econometric Theory*, 34(4):705–753.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52(1):201–226.

Phillips, P. C. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, 34(5):2449–2495.

Yiu, M. S. and Jin, L. (2012). Detecting bubbles in the hong kong residential property market: An explosive-pattern approach. *Hong Kong Institute for Monetary Research Working Paper*, (1).

A Proofs

Proof of Proposition 1

First of all we check what happens if we take the first differences of our stochastic trend. We have:

$$\Delta\mu_{i,t} = \beta_i \varepsilon_{i,t-s}.$$

Then if we consider the bubble component:

$$\phi_i(L)\psi_i(L^{-1})\Delta z_{i,t} = \Delta\varepsilon_{i,t}.$$

The first difference of the considered explosive common factor will be,

$$\Phi_i(L)\Delta \boldsymbol{f}_t = \phi_i(L)\beta_i\varepsilon_{i,t-s} + \psi_i(L^{-1})^{-1}\Delta\varepsilon_{i,t},$$

with $\Phi_i(L)$ a $1\times r$ causal polynomial vector that drives dependence across factors. Then, stacking these processes in a matrix we get,

$$\Phi(L)\Delta \boldsymbol{f}_t = \boldsymbol{C}(L)\boldsymbol{\varepsilon}_t,$$

with $\boldsymbol{C}_i(L)$ the i-th row of $\boldsymbol{C}(L)$ is such that $\boldsymbol{C}_i(L) = \phi_i(L)\beta_i L^s + \psi_i(L^{-1})^{-1}(1-L)$ if the i-th factor is an explosive factor. Then,

$$\begin{aligned}\Delta \boldsymbol{f}_t &= \Phi(L)^{-1}\boldsymbol{C}(L)\boldsymbol{\varepsilon}_t,\\ &= \boldsymbol{\Gamma}(L)\boldsymbol{\varepsilon}_t.\end{aligned}$$

The two-sided nature of the infinite sum in $\boldsymbol{\Gamma}(L)$ is ensured by $\psi_i(L^{-1})^{-1}$ inside the polynomial $\boldsymbol{C}_i(L)$.

Proof of Proposition 2

This proof follows the proof of Lemma 1 in Barigozzi et al. (2021), taking into account the different rate of convergence introduced by the heavy tails of the process obtained in Lemma 2. Define,

$$\hat{\Sigma}^{\Delta y} = T^{-1} \sum_{t=1}^{T} \Delta \boldsymbol{y}_t \Delta \boldsymbol{y}_t', \quad \Sigma^{\Delta y} = \mathbb{E}\big[\Delta \boldsymbol{y}_t \Delta \boldsymbol{y}_t'\big],$$

respectively the sample covariance matrix of $\Delta \boldsymbol{y}_t$ and the true covariance matrix. Moreover define and $\sigma^{\Delta}_{i,j}$ the generic element on the i-th row j-th column of $\Sigma^{\Delta y}$. Moreover define also,

$$\boldsymbol{\zeta}_t = \Lambda \boldsymbol{f}_t, \quad \text{and} \quad \Delta \boldsymbol{\zeta}_t = \Lambda \Delta \boldsymbol{f}_t.$$

Notice that we can express,

$$\begin{aligned} \Sigma^{\Delta y}_{ij} &= \Sigma^{\Delta \zeta} + \Sigma^{\Delta v}_{ij} \\ &= \boldsymbol{\lambda}_i \Sigma^{\Delta f} \boldsymbol{\lambda}_j + \Sigma^{\Delta v}_{ij}. \end{aligned}$$

with $\boldsymbol{\lambda}_i$ the i-th row of the loadings matrix, $\Sigma^{\Delta f} = \mathbb{E}\big[\Delta \boldsymbol{f}_t \Delta \boldsymbol{f}_t'\big]$ and $\Sigma^{\Delta v}_{ij} = \mathbb{E}\big[\Delta v_{i,t} \Delta v_{j,t}\big]$.

According to Lemma 2 we have that:

$$\begin{aligned} \left\| \frac{\hat{\Sigma}^{\Delta y}}{n} - \frac{\Sigma^{\Delta \zeta}}{n} \right\| &\leq \left\| \frac{\hat{\Sigma}^{\Delta y}}{n} - \frac{\Sigma^{\Delta y}}{n} \right\| + \left\| \frac{\hat{\Sigma}^{\Delta v}}{n} \right\| \\ &\leq \frac{1}{n} \max_j \sum_{i=1}^{n} \big| \hat{\Sigma}^{\Delta y}_{ij} - \Sigma^{\Delta y}_{ij} \big| + \frac{\nu_1^{\Delta v}}{n} \\ &= O_p\big(T^{-\delta}\big) + \frac{\overline{c}_2}{n} \\ &= O_p\big(T^{-\delta}\big) + O_p\big(n^{-1}\big) = O_p(\max\big(T^{-\delta}, n^{-1}\big)\big). \end{aligned}$$

Following Barigozzi et al. (2021) define $\boldsymbol{h}_i$ a n-dimensional vector with i-th entry 1 and all the other entries equal to 0. Then we have:

$$\begin{aligned} \left\| \frac{\boldsymbol{h}_i'}{\sqrt{n}} \big(\hat{\Sigma}^{\Delta y} - \Sigma^{\Delta \zeta}\big) \right\| &\leq \left\| \frac{\boldsymbol{h}_i'}{\sqrt{n}} \big(\hat{\Sigma}^{\Delta y} - \Sigma^{\Delta y}\big) \right\| + \left\| \frac{\boldsymbol{h}_i'}{\sqrt{n}} \Sigma^{\Delta v} \right\| \\ &\leq \frac{1}{\sqrt{n}} \max_j \big| \hat{\Sigma}^{\Delta y}_{ij} - \Sigma^{\Delta y}_{ij} \big| + \left\| \frac{\mu_1^{\Delta v}}{\sqrt{n}} \right\| \qquad (8) \\ &= O_p\big(\max\big(T^{-\delta}, n^{-1/2}\big)\big), \end{aligned}$$

where we used Lemma 1 for the boundedness of the eigenvalues of $\Sigma^{\Delta v}$. Note also that:

$$\left\|\frac{\boldsymbol{h}_i'}{\sqrt{n}}\Sigma^{\Delta v}\right\| = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\Sigma_{ij}^{\Delta v}\right)^2} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\Lambda_i\Sigma^{\Delta f}\Lambda_j'\right)^2} \leq C < \infty,$$

using the boundedness of the loadings and of the eigenvalues of the factors.

For what concerns the convergence of the eigenvalues, defining $\nu_i^{\Delta f}$ the $i-th$ eignenvalue of the matrix $\Sigma_{\Delta f}$ using Weyl's inequality we have for $i = 1, ..., r$:

$$\left\|\frac{\hat{\nu}_i^{\Delta y}}{n} - \frac{\nu_i^{\Delta\zeta}}{n}\right\| \leq \left\|\frac{\Sigma^{\Delta y}}{n} - \frac{\Sigma^{\Delta\zeta}}{n}\right\| = O_p(\max\left(T^{-\delta}, n^{-1}\right)). \tag{9}$$

Using the lower bound established for the eigenvalues of $\frac{\Sigma^{\Delta\zeta}}{n}$ by Lemma 1 and (9) we can establish a lower bound also for the r-th eigenvalue of the sample covariance matrix, that is:

$$\frac{\hat{\nu}_r^{\Delta y}}{n} \geq \underline{c}_1 + O_p(\max\left(T^{-\delta}, n^{-1}\right)). \tag{10}$$

If we define $N^{\Delta\zeta}$ and $\hat{N}^{\Delta y}$ the diagonal matrices of the first r eigenvalues of respectively $\Sigma_{\Delta\zeta}$ and $\Sigma_{\Delta y}$. By the previously established bounds we know that $N^{\Delta\zeta}$ is invertible and that $\hat{N}^{\Delta y}$ will be invertible with probability 1 as n and T grow to infinity. Moreover we have that:

$$\left\|\left(\frac{N^{\Delta\zeta}}{n}\right)^{-1}\right\| = \frac{n}{\lambda_r^{\Delta\zeta}} \leq C,$$

and,

$$\begin{aligned}
\left\|\left(\frac{N^{\Delta\zeta}}{n}\right)^{-1} - \left(\frac{\hat{N}^{\Delta y}}{n}\right)^{-1}\right\| &\leq \sqrt{\sum_{i=1}^{r}\left(\frac{n}{\nu_i^{\Delta\zeta}} - \frac{n}{\hat{\nu}_i^{\Delta y}}\right)^2} \\
&\leq \sum_{i=1}^{r} n\left|\frac{\hat{\nu}_i^{\Delta y} - \nu_i^{\Delta\zeta}}{\hat{\nu}_i^{\Delta y}\nu_i^{\Delta\zeta}}\right| \\
&\leq \frac{r\max_i\left|\hat{\nu}_i^{\Delta y} - \nu_i^{\Delta\zeta}\right|}{n\big(\underline{c}_1 + O_p(\max\left(T^{-\delta}, n^{-1}\right))\big)},
\end{aligned}$$

where in the last line we used the lower bounds from Lemma 2 and (10) for the eigenvalues of the two matrices. Then:

$$\left\|\left(\frac{N^{\Delta\zeta}}{n}\right)^{-1} - \left(\frac{\hat{N}^{\Delta y}}{n}\right)^{-1}\right\| \leq O_p(\max\left(T^{-\delta}, n^{-1}\right)).$$

Moreover from the identification constraint in Assumption 1.2 we have that $\Sigma^{\Delta f}$ is diagonal with values on the diagonal equal to $\mathbb{E}\big[\Delta f_{it}^2\big] = \nu_i^{\Delta\zeta}/n$, finite and bounded away from zero for $i = 1, ..., r$. from Lemma 1. Recall also that by Assumption 3 we have that $\Lambda'\Lambda/n = I_r$ for any $n \in \mathbb{N}$.

This allows us, using Lemma 1 of an upper and lower bound and Assumption 1.4 of distinct values on the diagonal of $\Sigma^{\Delta f}$, to conclude that also the first r eigenvalues of the covariance matrix of $\Delta\boldsymbol{\zeta}_t$ are distinct such that there are constants $\underline{k}_j, \overline{k}_j$ for $j = 1, ..., r-1$ $\underline{k}_j > \overline{k}_{j+1}$ and:

$$\underline{k}_j \leq \frac{\nu_i^{\Delta\zeta}}{n} \leq \overline{k}_j, \quad \text{for } j = 1, ..., r..$$

Moving now to the eigenvectors we have that defining $\boldsymbol{w}_j^{\Delta\zeta}$ and $\hat{\boldsymbol{w}}_j^{\Delta y}$ the normalized eigenvectors corresponding to the j-th largest eigenvalue of respectively $\Sigma_{\Delta\zeta}$ and $\hat{\Sigma}_{\Delta y}$ we can again use the same approach in Barigozzi et al. (2021), then defining $s = \text{sign}(\hat{\boldsymbol{w}}_j^{\Delta y}\boldsymbol{w}_j^{\Delta f})$ and $\nu_0^{\Delta\zeta} = \infty$:

$$||\hat{\boldsymbol{w}}_j^{\Delta y} - s\boldsymbol{w}_j^{\Delta\zeta}|| \leq 2^{3/2}\frac{\big|\big|\hat{\Sigma}_{\Delta y} - \Sigma_{\Delta\zeta}\big|\big|}{\min\big(|\nu_{j-1}^{\Delta\zeta} - \nu_j^{\Delta\zeta}|, |\nu_{j+1}^{\Delta\zeta} - \nu_j^{\Delta\zeta}|\big)}.$$

Given the previously established difference between the eigenvalues of $\Sigma^{\Delta\zeta}$ it is possible to establish a lower bound for the quantity at the denominator such that:

$$|\nu_{j-1}^{\Delta\zeta} - \nu_j^{\Delta\zeta}| \geq n\big(\underline{k}_{j-1} - \underline{k}_j\big) > 0$$
$$|\nu_{j+1}^{\Delta\zeta} - \nu_j^{\Delta\zeta}| \geq n\big(\underline{k}_{j+1} - \underline{k}_j\big) > 0$$

Then for the $n \times r$ matrices of normalized eigenvectors, respectively of the covariance matrix of $\Delta\boldsymbol{y}$, $\hat{\boldsymbol{W}}^{\Delta y} = \big(\hat{\boldsymbol{w}}_1^y, ..., \hat{\boldsymbol{w}}_r^y\big)$ and of $\Delta\boldsymbol{\zeta}_t$, $\boldsymbol{W}^{\Delta\zeta} = \big(\hat{\boldsymbol{w}}_1^f, ..., \hat{\boldsymbol{w}}_r^f\big)$. Then by the previous results defining S a diagonal matrix with s_j as j-th element on the diagonal, such that:

$$||\hat{\boldsymbol{W}}^{\Delta y} - \boldsymbol{W}^{\Delta f}S|| \leq \sqrt{\sum_{j=1}^{r}||\hat{\boldsymbol{w}}_j^{\Delta y} - s\boldsymbol{w}_j^{\Delta\zeta}||} = O_p(\max\big(T^{-\delta}, n^{-1}\big)\big).$$

The estimator for the loadings Λ is defined as,

$$\hat{\Lambda} = \sqrt{n}\hat{\boldsymbol{W}}^{\Delta y}.$$

Note that by this definition we have,

$$||\frac{\hat{\Lambda} - \Lambda}{\sqrt{n}}|| = O_p(\max\big(T^{-\delta}, n^{-1}\big)).$$

We can define $\hat{\boldsymbol{\lambda}}_i = \sqrt{n}\boldsymbol{h}_i'\hat{\boldsymbol{W}}^{\Delta y}$ and $\boldsymbol{\lambda}_i = \sqrt{n}\boldsymbol{h}_i'\boldsymbol{W}^{\Delta\zeta}$. Note that $\boldsymbol{W}^{\Delta\zeta}$ is the matrix of normalized eigenvalues of $\Delta\zeta$, then $\Sigma_{\Delta\zeta}\boldsymbol{W}^{\Delta\zeta}S = \boldsymbol{W}^{\Delta\zeta}SM_{\Delta\zeta}$. Then we have,

$$\begin{aligned}
\big|\big|\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}S\big|\big| &= \big|\big|\sqrt{n}\boldsymbol{h}_i'\hat{\boldsymbol{W}}^{\Delta y} - \sqrt{n}\boldsymbol{h}_i'\boldsymbol{W}^{\Delta\zeta}S\big|\big| \\
&= \left|\left|\frac{\boldsymbol{h}_i'}{\sqrt{n}}\left(\hat{\Sigma}_{\Delta y}\hat{\boldsymbol{W}}^{\Delta y}\big(\hat{N}_{\Delta y}/n\big)^{-1} - \Sigma_{\Delta\zeta}\boldsymbol{W}^{\Delta\zeta}S\big(N_{\Delta\zeta}/n\big)^{-1}\right)\right|\right| \\
&\leq \left|\left|\frac{\boldsymbol{h}_i'}{\sqrt{n}}\left(\hat{\Sigma}_{\Delta y} - \Sigma_{\Delta\zeta}\right)\right|\right| \left|\left|\big(N^{\Delta\zeta}/n\big)^{-1}\right|\right| \\
&+ \left|\left|\frac{\boldsymbol{h}_i'}{\sqrt{n}}\Sigma_{\Delta\zeta}\right|\right| \left|\left|\big(\hat{N}_{\Delta y}/n\big)^{-1} - \big(N_{\Delta\zeta}/n\big)^{-1}\right|\right| \\
&+ ||\hat{\boldsymbol{W}}^{\Delta y} - \boldsymbol{W}^{\Delta\zeta}S|| \left|\left|\frac{\boldsymbol{h}_i'}{\sqrt{n}}\Sigma_{\Delta\zeta}\right|\right| \left|\left|\big(N^{\Delta\zeta}/n\big)^{-1}\right|\right| + o_p(\max(1/\sqrt{n}, 1/\sqrt{T}) \\
&= O_p(\max\big(T^{-\delta}, n^{-1}\big)).
\end{aligned}$$

Now moving to the factors, we define:

$$\hat{\boldsymbol{f}}_t = \big(\hat{\Lambda}'\hat{\Lambda}\big)^{-1}\hat{\Lambda}'\boldsymbol{f}_t = \frac{\hat{\Lambda}'\boldsymbol{f}_t}{n}.$$

This implies the final point of the proposition:

$$\begin{aligned}
\frac{1}{\sqrt{T}}||\hat{\boldsymbol{f}}_t - S\boldsymbol{f}_t|| &= ||\frac{\hat{\Lambda}\Lambda}{n}\boldsymbol{f}_t - S\boldsymbol{f}_t + \frac{\hat{\Lambda}\boldsymbol{v}_t}{n}|| \\
&\leq ||\frac{\hat{\Lambda}\Lambda}{n} - S|| \cdot ||\frac{\boldsymbol{f}_t}{\sqrt{T}}|| + ||\frac{\hat{\Lambda}\boldsymbol{v}_t}{n}|| \\
&= O_p\big(\max\big(T^{-\delta}, n^{-1/2}\big)\big).
\end{aligned}$$

Proof of Theorem 1

We will prove this theorem only for the MAR model with a stochastic trend as the other specifications are similar and simpler. The proof of this theorem requires a number of steps. First of all we need uniform invertibility and moments conditions for the filter. Then we can focus on the Maximum Likelihood estimation properties. The proof follows Blasques et al. (2023) in showing consistency of the estimator taking into account that the filter is applied to an estimate of the underlying process. Recall that the true process $\{f_t\}_{t\in\mathbb{Z}}$ is defined as a MAR(r,s) process with a stochastic trend: The filtered process can be defined as:

$$f_t = \mu_t + \psi(L^{-1})^{-1}\phi(L)^{-1}\varepsilon_t, \quad \varepsilon_t \sim t_\nu$$
$$\mu_t = \mu_{t-1} + \alpha\varepsilon_{t-s}.$$

Define the unfeasible prediction error:

$$\tilde{\boldsymbol{g}}_t(\boldsymbol{\theta}) = f_t - \tilde{\mu}_t(\boldsymbol{\theta}).$$

By Proposition 2 of Blasques et al. (2023) we know that:

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\tilde{\boldsymbol{g}}_t(\boldsymbol{\theta}) - g_t(\boldsymbol{\theta})\right| \xrightarrow{e.a.s.} 0.$$

We also know that such a prediction error can be unfolded to obtain, in vector form:

$$\tilde{\boldsymbol{g}}_{t+1}(\boldsymbol{\theta}) = \sum_{r=0}^{t-k} A(\boldsymbol{\theta})^r C(\boldsymbol{\theta}) + A(\boldsymbol{\theta})^{t-k}\tilde{\boldsymbol{g}}_k(\boldsymbol{\theta}) + \sum_{r=0}^{t-k} A(\boldsymbol{\theta})^r B_{t+1}. \tag{11}$$

Define now the feasible prediction error:

$$\hat{g}_t(\boldsymbol{\theta}) = \hat{f}_t - \hat{\mu}_t(\boldsymbol{\theta}),$$

where $\hat{f}_t$ is the first stage estimate of the common factor and $\hat{\mu}_t(\boldsymbol{\theta})$ is the non-stationary filter that depends on this estimate. Note that also this process can be unfolded as:

$$\hat{\boldsymbol{g}}_{t+1}(\boldsymbol{\theta}) = \sum_{r=0}^{t-k} A(\boldsymbol{\theta})^r C(\boldsymbol{\theta}) + A(\boldsymbol{\theta})^{t-k}\hat{\boldsymbol{g}}_k(\boldsymbol{\theta}) + \sum_{r=0}^{t-k} A(\boldsymbol{\theta})^r \hat{B}_{t+1}, \tag{12}$$

where the only non zero component of $\hat{B}_{t+1}$ and B_{t+1} are respectively $\Delta\tilde{f}_t$ and Δf_t, with $\tilde{\boldsymbol{g}}_k(\boldsymbol{\theta}), \hat{\boldsymbol{g}}_k(\boldsymbol{\theta})$ depending on the initialization that can be treated as constant.

Then:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\tilde{\boldsymbol{g}}_{t+1}(\boldsymbol{\theta})-\hat{\boldsymbol{g}}_{t+1}(\boldsymbol{\theta})\right|\right| \leq K\sum_{r=0}^{t-k}\rho^r \sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\hat{B}_{t+1}-B_{t+1}\right|\right| + K\rho^{t-k}\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\tilde{\boldsymbol{g}}_k(\boldsymbol{\theta})-\hat{\boldsymbol{g}}_k(\boldsymbol{\theta})\right|\right|.$$

Now note that by Proposition 2 we have:

$$\begin{aligned}\left|\left|\Delta\hat{\boldsymbol{f}}_t - S\Delta\boldsymbol{f}_t\right|\right| &= \left|\left|\frac{\hat{\Lambda}\Lambda}{n}\Delta\boldsymbol{f}_t - S\Delta\boldsymbol{f}_t + \frac{\hat{\Lambda}\Delta\boldsymbol{v}_t}{n}\right|\right| \\ &\leq \left|\left|\frac{\hat{\Lambda}\Lambda}{n} - S\right|\right|\cdot\left|\left|\Delta\boldsymbol{f}_t\right|\right| + \left|\left|\frac{\hat{\Lambda}\boldsymbol{v}_t}{n}\right|\right| \\ &= O_p\big(\max\left(T^{-\delta}, n^{-1/2}\right)\big).\end{aligned}$$

Then as $\theta \to \infty$ where $\theta = \max\left(T^{-\delta}, n^{-1/2}\right)$ we have:

$$|\Delta\hat{f}_t - \Delta f_t| \xrightarrow{p} 0.$$

This implies that:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\tilde{\boldsymbol{g}}_t(\boldsymbol{\theta}) - \hat{\boldsymbol{g}}_t(\boldsymbol{\theta})\right|\right| \xrightarrow{p} 0, \quad \text{as } \delta \to \infty.$$

Moreover define the residuals based on the prediction errors as:

$$\varepsilon_t(\boldsymbol{\theta}) = \phi(L)\psi(L^{-1})g_t(\boldsymbol{\theta}).$$

Then we have that:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{\varepsilon}_t(\boldsymbol{\theta}) - \tilde{\varepsilon}_t(\boldsymbol{\theta})\right| \xrightarrow{p} 0, \quad \text{as } \delta \to \infty.$$

This concludes the uniform invertibility section. From Proposition 3 of Blasques et al. (2023) we have that the under Assumption 1.1:

$$\mathbb{E}|g_t(\boldsymbol{\theta})|^n < \infty.$$

Now we can follow the approach in Blasques et al. (2022) building on the classical consistency argument using uniform convergence of the criterion function and the identifiable uniqueness of the true parameter vector $\boldsymbol{\theta}_{i,0}$. We note first that the log-likelihood takes the form,

$$\hat{L}_T(\boldsymbol{\theta}) = \frac{1}{T-k}\sum_{t=r}^{T-s}\hat{l}_t(\boldsymbol{\theta}) = \frac{1}{T-k}\sum_{t=r}^{T-s}\log f(\psi(L^{-1})\phi(L)\hat{g}_t(\boldsymbol{\theta})),$$

where we have $k = r + s$ and $\hat{l}_t(\boldsymbol{\theta})$ is the log-likelihood contribution of the observation at time t and $\hat{g}_t(\boldsymbol{\theta}) = \hat{f}_t - \hat{\mu}_t(\boldsymbol{\theta})$ as defined before. Recall also:

$$\log f(\hat{\varepsilon}_t(\boldsymbol{\theta})) = \log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left(1 + \frac{\hat{\varepsilon}_t(\boldsymbol{\theta})^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right), \tag{13}$$

and filtered residuals are defined as $\hat{\varepsilon}_t(\boldsymbol{\theta}) = \psi(L^{-1})\phi(L)\hat{g}_t(\boldsymbol{\theta})$. We further let $L_T(\boldsymbol{\theta})$ denote the log-likelihood with the limit sequence $\varepsilon_t(\boldsymbol{\theta})$,

$$L_T(\boldsymbol{\theta}) = \frac{1}{T-k}\sum_{t=r}^{T-s} l_t(\boldsymbol{\theta}) = \sum_{t=r}^{T-s} l(\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\gamma}),$$

let $\tilde{L}_T(\boldsymbol{\theta})$ denote the log-likelihood with the filtered sequence $\tilde{\varepsilon}_t(\boldsymbol{\theta})$,

$$\tilde{L}_T(\boldsymbol{\theta}) = \frac{1}{T-k}\sum_{t=r}^{T-s} \tilde{l}_t(\boldsymbol{\theta}) = \sum_{t=r}^{T-s} l(\tilde{\varepsilon}_t(\boldsymbol{\theta}), \boldsymbol{\gamma}),$$

To prove uniform convergence of the criterion function, note that we have two types of approximation involved. The filtered unfeasible likelihood that approximates the underlying true process, with filter converging *e.a.s.* to the true process with $t \to \infty$, and the feasible likelihood approximating the unfeasible filter likelihood as $\delta \to \infty$. Using the triangle inequality,

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{L}_T(\boldsymbol{\theta}) - L_\infty(\boldsymbol{\theta})\right| \le \sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{L}_T(\boldsymbol{\theta}) - \tilde{L}_T(\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\in\Theta}\left|\tilde{L}_T(\boldsymbol{\theta}) - L_T(\boldsymbol{\theta})\right| \\ + \sup_{\boldsymbol{\theta}\in\Theta}\left|L_T(\boldsymbol{\theta}) - L_\infty(\boldsymbol{\theta})\right|.\end{aligned} \tag{14}$$

Now consider,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{L}_T(\boldsymbol{\theta}) - \tilde{L}_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{l}_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})\right|.$$

Then by the mean value theorem:

$$\begin{aligned}\hat{l}_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta}) &= \frac{\nu+1}{2}\left[\log\left(\nu + \hat{\varepsilon}_t(\boldsymbol{\theta})^2\right) - \log\left(\nu + \tilde{\varepsilon}_t(\boldsymbol{\theta})^2\right)\right] \\ &= \frac{\nu+1}{2(\nu + \varepsilon_t^*(\boldsymbol{\theta})^2)}\left(\hat{\varepsilon}_t(\boldsymbol{\theta})^2 - \tilde{\varepsilon}_t(\boldsymbol{\theta})^2\right),\end{aligned}$$

where ε^* is a point between $\hat{\varepsilon}$ and $\tilde{\varepsilon}$. Since $\tilde{\varepsilon}_t(\boldsymbol{\theta})^2$ is always positive and we assumed $\nu \ge 1$ we have $\frac{\nu+1}{2(\nu+\varepsilon_t^*(\boldsymbol{\theta})^2)} \le 1$. Hence,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{l}_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})\right| \le \sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{\varepsilon}_t(\boldsymbol{\theta})^2 - \tilde{\varepsilon}_t(\boldsymbol{\theta})^2\right| \xrightarrow{p} 0, \quad \text{as } \delta \to \infty.$$

Then,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{L}_T(\boldsymbol{\theta})-\tilde{L}_T(\boldsymbol{\theta})\right| \xrightarrow{p} 0, \qquad \text{as } \delta\to\infty.$$

A different approach must be used for the second term,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{L}_T(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{l}_t(\boldsymbol{\theta})-l_t(\boldsymbol{\theta})\right|.$$

Here the convergence of the terms if over t, due to uniform invertibility, not T. This means that for this average to go to zero we need the individual terms to converge to zero fast enough such that the whole sum converges. Using a similar argument to what discussed before we have:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{l}_t(\boldsymbol{\theta})-l_t(\boldsymbol{\theta})\right| \leq \sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{\varepsilon}_t(\boldsymbol{\theta})^2-\varepsilon_t(\boldsymbol{\theta})^2\right|.$$

Since $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is SE, $\mathbb{E}\log|\varepsilon_t(\boldsymbol{\theta})|<\infty$ and $\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{\varepsilon}_t(\boldsymbol{\theta})-\varepsilon_t(\boldsymbol{\theta})\right| \xrightarrow{e.a.s.} 0$, by Lemma TA.17 of Blasques et al. (2022) we have that $\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{\varepsilon}_t(\boldsymbol{\theta})^2-\varepsilon_t(\boldsymbol{\theta})^2\right| \xrightarrow{e.a.s.} 0$. It follows then from Lemma 2.1 of Straumann and Mikosch (2006) that the sum of these terms converges, this implies that the average converges to zero, that is,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{L}_T(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0, \quad \text{as } T\to\infty.$$

For what concerns the last term in (14) we have that $\{l_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ defined as $l_t(\boldsymbol{\theta}) = \log f(\varepsilon_t(\boldsymbol{\theta}))$ is a stationary and ergodic sequence by Proposition 4.3 Krengel (1985) being a measurable function of $\{\varepsilon_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that is an SE sequence. Moreover:

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}|l(\varepsilon_t(\boldsymbol{\theta}),\gamma)| &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left(1+\frac{\varepsilon_t(\boldsymbol{\theta})^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right)\right| \\
&\leq c_0+\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(\left(1+\frac{\varepsilon(\boldsymbol{\theta})^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right)\right| \\
&\leq c_0+c_1\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\varepsilon_t(\boldsymbol{\theta})^2}{\nu}\right)\right| \\
&= c_0+c_1\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\varepsilon_t(\boldsymbol{\theta})\right|^{\delta}<\infty.
\end{aligned}$$

for some $\delta<1$. Then we can apply a uniform law of large numbers such that:

$$\sup_{\boldsymbol{\theta}\in\Theta} \big| L_T(\boldsymbol{\theta}) - L_\infty(\boldsymbol{\theta}) \big| \xrightarrow{a.s.} 0, \quad \text{as } T \to \infty. \tag{15}$$

For what concerns identifiable uniqueness we have follows by noting that $L(\boldsymbol{\theta})$ exists for every $\boldsymbol{\theta} \in \Theta$, by C2. To show uniqueness of the maximizer $\boldsymbol{\theta}_0$ we need that for any $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ we have $L(\boldsymbol{\theta}) < L(\boldsymbol{\theta}_0)$. We first show that $l\big(\varepsilon_t(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0\big) = l\big(\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\gamma}\big)$ almost surely if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. We know that $\varepsilon_t(\boldsymbol{\theta}_0) = \varepsilon_t$ almost surely for all t. We also know ε_t is Student's t distributed so it has a non-zero density on all $\mathbb{R}$. Hence it is enough to show that $l(h + \varepsilon; \boldsymbol{\gamma}) = l(\varepsilon; \boldsymbol{\gamma}_0)$ can hold with probability 1 if and only if $h = 0$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. By the definition of $l(\cdot)$, for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$, this requires,

$$\log\left(\frac{\Gamma(\frac{\nu_1+1}{2})}{\Gamma(\frac{\nu_1}{2})\sqrt{\pi\nu_1\sigma_1^2}} \left(1 + \frac{(x+h)^2}{\sigma_1^2\nu_1}\right)^{-\frac{\nu_1+1}{2}} \right) = \log\left(\frac{\Gamma(\frac{\nu_2+1}{2})}{\Gamma(\frac{\nu_2}{2})\sqrt{\pi\nu_2\sigma_2^2}} \left(1 + \frac{x^2}{\sigma_2^2\nu_2}\right)^{-\frac{\nu_2+1}{2}} \right).$$

for all $x \in \mathbb{R}$. Clearly $l(h + \varepsilon; \boldsymbol{\gamma}) = l(\varepsilon; \boldsymbol{\gamma}_0)$ almost surely for all t requires $h = 0$ and $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$.

We now need to prove that given that $\boldsymbol{\theta} = (\alpha, \omega, \Psi, \boldsymbol{\gamma})$ is such that $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ we can conclude that $g_t(\boldsymbol{\theta}) = g_t(\boldsymbol{\theta}_0) = v_t$ almost surely if and only if $(\alpha, \omega, \Psi) = (\alpha_0, \omega_0, \Psi_0)$. Suppose this is not the case and that $g_t(\boldsymbol{\theta}) = v_t$ almost surely for some t, than it must hold for all $t \in \mathbb{Z}$. Then we would have,

$$\begin{aligned}
g_{t+1}(\boldsymbol{\theta}) &= g_t(\boldsymbol{\theta}) - \omega - \alpha\phi(L)\psi(L^{-1})v_{t-s} + \Delta y_{t+1} \\
&= g_t(\boldsymbol{\theta}) - v_t - \big(\omega - \omega_0\big) - \alpha \sum_{h=-\infty}^{\infty} \rho_h \varepsilon_{t+h} + \alpha_0 \varepsilon_{t-s} + v_{t+1} \\
&= g_t(\boldsymbol{\theta}) - v_t + \omega_0 - \omega + \alpha_0 \varepsilon_{t-s} - \alpha \sum_{h=-\infty}^{\infty} \rho_h \varepsilon_{t+h} + v_{t+1}.
\end{aligned}$$

Since by hypothesis $g_t(\boldsymbol{\theta}) = v_t$ for all t then we must have,

$$\omega_0 - \omega = \alpha_0 \varepsilon_{t-s} - \alpha \sum_{h=-\infty}^{\infty} \rho_h \varepsilon_{t+h}, \quad \text{almost surely for all } t.$$

Now if $\omega \neq \omega_0$ it means that the right-hand side must be a non-zero constant. But the right-hand side expression is a non degenerate function of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ that is $\neq 0$ almost surely for all t for all $\boldsymbol{\theta} \in \Theta$ if $\alpha \neq \alpha_0$ and $\Psi \neq \Psi_0$. This means that it must be that $\omega = \omega_0$. Then since the right-hand side is non zero with probability one we can have $g_{t+1}(\boldsymbol{\theta}) = v_{t+1}$ if and only if $\alpha = \alpha_0$ and $\Psi \neq \Psi_0$.

Now that we showed that $l\big(\varepsilon_t(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0\big) = l\big(\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\gamma}\big)$ almost surely if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we can use an argument on the lines of the one used in Blasques et al. (2022) using some of the arguments from Breid et al. (1991) to conclude the proof of C3. We will rely on a mean value expansion around $\boldsymbol{\theta}_0$. Recall that unfolding our limit prediction error process we have:

$$g_t(\boldsymbol{\theta}) = \sum_{i=-\infty}^{\infty} \gamma_i z_t, \tag{16}$$

with $z_t = \delta + \Delta y_t$. Moreover recall ζ_i the coefficient of the i-th element of the polynomial $\psi(L^{-1})\phi(L)$. Then consider Θ as a compact set satisfying Assumption 1-4 such that:

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta} |\zeta_i - \zeta_{i,0}| \leq C\epsilon \\
&\sup_{\boldsymbol{\theta}\in\Theta} |\gamma_i| \leq C|d|^i \\
&\sup_{\boldsymbol{\theta}\in\Theta} |\gamma_i - \gamma_{0,i}| \leq C\epsilon|d|^i \\
&\sup_{\boldsymbol{\theta}\in\Theta} |\delta - \delta_0| \leq C\epsilon,
\end{aligned}$$

with $|d| < 1$. This allows us to conclude that:

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta} \big|g_t(\boldsymbol{\theta}) - g_t(\boldsymbol{\theta}_0)\big| &\leq \sum_{i=-\infty}^{\infty} \sup_{\boldsymbol{\theta}\in\Theta} \big|\gamma_i - \gamma_{0,i}\big| \cdot \big(\big|\Delta y_t\big| + \big|\delta_0\big|\big) + \sum_{i=-\infty}^{\infty} \sup_{\boldsymbol{\theta}\in\Theta} \big|\gamma_i\big| \cdot \sup_{\boldsymbol{\theta}\in\Theta} \big|\delta - \delta_0\big| \\
&\leq \epsilon\Big(C_0 + C_1 \sum_{i=-\infty}^{\infty} |d|^i \big|z_t\big|\Big) \\
\sup_{\boldsymbol{\theta}\in\Theta} \big|\varepsilon_t(\boldsymbol{\theta}) - \varepsilon_t(\boldsymbol{\theta}_0)\big| &= \sup_{\boldsymbol{\theta}\in\Theta} \big|\phi(L)\psi(L^{-1})g_t(\boldsymbol{\theta}) - \phi_0(L)\psi_0(L^{-1})g_t(\boldsymbol{\theta}_0)\big| \\
&\leq \epsilon\Big(C_0 + C_1 \sum_{i=1}^{k} \big|g_{t-i}(\boldsymbol{\theta}_0)\big| + C_2 \sum_{i=-\infty}^{\infty} |d|^i \big|z_t\big|\Big).
\end{aligned}$$

Moreover following Breid et al. (1991) we can write:

$$\varepsilon_t(\boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\theta}) + \varepsilon_t(\boldsymbol{\theta}_0) - \varepsilon_t(\boldsymbol{\theta}_0),$$

with:

$$\varepsilon_t(\boldsymbol{\theta}_0) - \epsilon K_t \leq \varepsilon_t(\boldsymbol{\theta}) \leq \varepsilon_t(\boldsymbol{\theta}_0) + \epsilon K_t.$$

Note that the second derivatives of the log likelihood function, avoiding the repetitions in the cross derivatives, will be:

$$\frac{\partial l^2(\varepsilon(\boldsymbol{\theta}^*))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} = \begin{cases} \frac{\partial^2 \varepsilon_t(\boldsymbol{\theta})}{\partial \phi_i \partial \phi_j} h(\varepsilon_t(\boldsymbol{\theta})) + \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \phi_i} \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \phi_j} h'(\varepsilon_t(\boldsymbol{\theta})) \\ \frac{\partial^2 \varepsilon_t(\boldsymbol{\theta})}{\partial \phi_i \partial \psi_j} h(\varepsilon_t(\boldsymbol{\theta})) + \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \phi_i} \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \psi_j} h'(\varepsilon_t(\boldsymbol{\theta})) \\ \frac{\partial^2 \varepsilon_t(\boldsymbol{\theta})}{\partial \Pi_i \partial \Pi_j} h(\varepsilon_t(\boldsymbol{\theta})) + \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \Pi_i} \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \Pi_j} h'(\varepsilon_t(\boldsymbol{\theta})) \\ \frac{\partial^2 \varepsilon_t(\boldsymbol{\theta})}{\partial \Pi_i \partial \phi_j} h(\varepsilon_t(\boldsymbol{\theta})) + \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \Pi_i} \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \phi_j} h'(\varepsilon_t(\boldsymbol{\theta})) \\ \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \Psi_i} h(\varepsilon_t(\boldsymbol{\theta})) + \sigma^{-1} \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \Psi_i} h'(\varepsilon_t(\boldsymbol{\theta})) \\ \sigma^{-1} \varepsilon_t(\boldsymbol{\theta}) h(\varepsilon_t(\boldsymbol{\theta})) + \sigma^{-2} \varepsilon_t(\boldsymbol{\theta})^2 h'(\varepsilon_t(\boldsymbol{\theta})) + 1. \end{cases}$$

Note that similarly to what has been done at the beginning of the Proof for Theorem 1 for the filtered prediction error, all the first and second derivatives of $\varepsilon_t(\boldsymbol{\theta})$ can be written as unfoldable and converging SREs. Unfolding these expression it is possible to show these expressions as infinite sums of the underlying z_t as in (16) with the same sequence of coefficients $\{\gamma_i\}_{t \in \mathbb{Z}}$.

Then we have:

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{\partial \varepsilon_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} - \frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} \right| \leq \epsilon C Z_t$$
$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{\partial^2 \varepsilon_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} - \frac{\partial^2 \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right| \leq \epsilon C Z_t,$$

where Z_t is such that $\mathbb{E}|Z_t|^n < \infty$ with n such that $\mathbb{E}|\varepsilon_t|^n < \infty$. Now that we defined bounds on these given quantities we can use the same approach as Breid et al. (1991) to conclude the proof. Here we define a mean value expansion in $\boldsymbol{\theta}_0$ of our expected likelihood difference.

$$
\begin{aligned}
&\mathbb{E}\big[l(\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\theta}) - l(\varepsilon_t(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)\big] \\
&= \mathbb{E}\Bigg[\sum_{i=1}^{k}\frac{\partial l(\varepsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i}\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big) + \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial l^2(\varepsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)\big(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j,0}\big) \\
&+ \sum_{i=1}^{k}\sum_{j=1}^{k}\left(\frac{\partial l^2(\varepsilon(\boldsymbol{\theta}^*))}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j} - \frac{\partial l^2(\varepsilon(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right)\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)\big(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j,0}\big)\Bigg].
\end{aligned}
\tag{17}
$$

From now on we will provide an argument for the derivative taken with respect to $i \leq r$ but the same argument holds for the non-causal part. Note that using $\varepsilon_t(\boldsymbol{\theta}_0) = \varepsilon_t$ we have:

$$
\mathbb{E}\left[\frac{\partial l(\varepsilon(\boldsymbol{\theta}_0))}{\partial\boldsymbol{\theta}_i}\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\partial l(\varepsilon(\boldsymbol{\theta}_0))}{\partial\boldsymbol{\theta}_i}\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)\middle|\mathcal{F}_{t-1}\right]\right] = 0.
$$

For what concerns the third term we have:

$$
\mathbb{E}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial l^2(\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)\big(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j,0}\big)\right] = -\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)'\mathcal{I}(\boldsymbol{\theta}_0)\big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big).
$$

Finally for the last term we can apply a similar reasoning as what it is done in Breid et al. (1991). We have:

$$
\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\theta}}\bigg|\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}^*)) + \frac{\partial^2\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i\partial\phi_j}h(\varepsilon_t(\boldsymbol{\theta}^*)) \\
&\quad - \frac{\partial\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}_0)) - \frac{\partial^2\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_i\partial\phi_j}h(\varepsilon_t(\boldsymbol{\theta}_0))\bigg| \\
&\quad\leq\mathbb{E}\sup_{\boldsymbol{\theta}}\left|\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}_0)) - \frac{\partial\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}_0))\right| \\
&\quad + \mathbb{E}\sup_{\boldsymbol{\theta}}\left|\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}_0)) - \frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}_0))\right| \\
&\quad + \mathbb{E}\sup_{\boldsymbol{\theta}}\left|\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}^*)) - \frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i}\frac{\partial\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_j}h'(\varepsilon_t(\boldsymbol{\theta}_0))\right| \\
&\quad + \mathbb{E}\sup_{\boldsymbol{\theta}}\left|\frac{\partial^2\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i\partial\phi_j}h(\varepsilon_t(\boldsymbol{\theta}_0)) - \frac{\partial^2\varepsilon_t(\boldsymbol{\theta}_0)}{\partial\phi_i\partial\phi_j}h(\varepsilon_t(\boldsymbol{\theta}_0))\right| \\
&\quad + \mathbb{E}\sup_{\boldsymbol{\theta}}\left|\frac{\partial^2\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i\partial\phi_j}h(\varepsilon_t(\boldsymbol{\theta}^*)) - \frac{\partial^2\varepsilon_t(\boldsymbol{\theta}^*)}{\partial\phi_i\partial\phi_j}h(\varepsilon_t(\boldsymbol{\theta}_0))\right| \\
&\quad = c_1 + c_2 + c_3 + c_4 + c_5.
\end{aligned}
$$

Then,

$$c_1 \leq \epsilon C \mathbb{E} \left| Z_t \frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \phi_j} h'(\varepsilon_t(\boldsymbol{\theta}_0)) \right| \to 0, \quad \text{as } \epsilon \to 0.$$

Note that $\mathbb{E}\left|Z_t \frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \phi_j} h'(\varepsilon_t(\boldsymbol{\theta}_0))\right| < \infty$ as $\frac{\partial \varepsilon_t(\boldsymbol{\theta}_0)}{\partial \phi_j}$ and $h'(\varepsilon_t(\boldsymbol{\theta}_0))$ are independent and it is possible to split the infinite past and future elements in Z_t such that all the elements in the expectation are bounded by $\mathbb{E}|\varepsilon_t|^2 < \infty$.
By a similar argument also $c_2 \to 0$ as $\epsilon \to 0$. Moreover as in Breid et al. (1991) we can split,

$$h'(x) = h_1(x) - h_2(x),$$

with $h_i(\cdot)$ non-decreasing functions such that:

$$h_i(x) = O(|x|^k), \quad \text{as } |x| \to \infty$$

with k such that $\mathbb{E}|\varepsilon_t|^{2+k} < \infty$. Note also that the same operation is possible for $h(x)$. With this definition we can define:

$$X_{i,t} = \begin{cases} h_i\left(\frac{\varepsilon_t(\boldsymbol{\theta}_0) - \varepsilon C K_t}{\sigma_0 - \epsilon}\right) - h_i\left(\frac{\varepsilon_t(\boldsymbol{\theta}_0) + \varepsilon C K_t}{\sigma_0 + \epsilon}\right), & \text{if } \varepsilon_t(\boldsymbol{\theta}_0) + \varepsilon C K_t ¿0 \\ h_i\left(\frac{\varepsilon_t(\boldsymbol{\theta}_0) - \varepsilon C K_t}{\sigma_0 + \epsilon}\right) - h_i\left(\frac{\varepsilon_t(\boldsymbol{\theta}_0) + \varepsilon C K_t}{\sigma_0 - \epsilon}\right), & \text{if } \varepsilon_t(\boldsymbol{\theta}_0) + \varepsilon C K_t ¡0 \\ h_i\left(\frac{\varepsilon_t(\boldsymbol{\theta}_0) - \varepsilon C K_t}{\sigma_0 - \epsilon}\right) - h_i\left(\frac{\varepsilon_t(\boldsymbol{\theta}_0) + \varepsilon C K_t}{\sigma_0 - \epsilon}\right), & \text{otherwise.} \end{cases}$$

Then we can bound:

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} c_3 \leq \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \frac{\partial \varepsilon_t(\boldsymbol{\theta}^*)}{\partial \phi_i} \frac{\partial \varepsilon_t(\boldsymbol{\theta}^*)}{\partial \phi_j} \Big(X_{1,t} + X_{2,t} \Big).$$

Using the moment bounds it is possible to show that this expected value is finite, then by dominated convergence we have that $c_3 \to 0$ as $\varepsilon \to 0$. We can apply the same approach to c_4 and c_5 so that we showed that the difference between the second derivatives in the last term of (17) goes to zero with ϵ for $i, j \leq k$. The reasoning is similar for other elements of the second derivative of the score as argued in Breid et al. (1991), hence we have that:

$$\mathbb{E}\left[\sum_{i=1}^{k} \sum_{j=1}^{k} \left(\frac{\partial l^2(\varepsilon(\boldsymbol{\theta}^*))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} - \frac{\partial l^2(\varepsilon(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right) \big(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0}\big)\big(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j,0}\big) \right] \to 0, \quad \text{for } \epsilon \to 0.$$

so that there is a $\varepsilon > 0$ such that for all $\boldsymbol{\theta} \in \Theta$ such that $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ we have:

$$L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0) = \mathbb{E}\big[l(\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\gamma}) - l(\varepsilon_t(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0)\big] < 0.$$

Moreover assumption 3 ensures that $\boldsymbol{\theta}$ is a compact set and the uniform convergence result showed in part **i)** implies the continuity of the limit criterion function $L(\boldsymbol{\theta})$. These two results combined with the uniqueness of the maximizer imply the result. □

B Lemmas

Lemma 1 *Define $\nu_i^{\Delta y}, \nu_i^{\Delta \zeta}, \nu_i^{\Delta v}$ the i-th eigenvalue of the covariance matrices of respectively the first differences processes $\Delta \boldsymbol{y}_t, \Delta \boldsymbol{\zeta}_t = \Lambda \Delta \boldsymbol{f}_t$ and $\Delta \boldsymbol{v}_t$. Then:*

1. $0 < \underline{c}_1 \leq \frac{\nu_i^{\Delta \zeta}}{n} \leq \overline{c}_1 < \infty$*, for* $i = 1, ..., r$ *and* $n \geq \overline{n}$.
2. $\nu_1^{\Delta v} \leq \overline{c}_2 < \infty$.
3. $\underline{c}_3 \leq \frac{\nu_j^{\Delta y}}{n} \leq \overline{c}_3$ *for* $i = 1, ..., r$ *and* $n \geq \overline{n}$ *and* $\nu_{r+1}^{\Delta y} \leq \overline{c}_4 < \infty$ *for all* $n \in \mathbb{N}$.

Proof
The proof of this Lemma follows the proof of Lemma D.2 in Barigozzi et al. (2021). Define $\Sigma_{\Delta f} = \mathbb{E}\big[\Delta \boldsymbol{f}_t \Delta \boldsymbol{f}_t'\big]$, $\Sigma_{\Delta \zeta} = \mathbb{E}\big[\Delta \boldsymbol{\zeta}_t \Delta \boldsymbol{\zeta}_t'\big]$, $\Sigma_{\Delta v} = \mathbb{E}\big[\Delta \boldsymbol{v}_t \Delta \boldsymbol{v}_t'\big]$ and $\Sigma_{\Delta y} = \mathbb{E}\big[\Delta \boldsymbol{y}_t \Delta \boldsymbol{y}_t'\big]$.

1) Note that we can write $\Sigma_{\Delta f} = Q^{\Delta f} N^{\Delta f} Q^{\Delta f'}$ with $N^{\Delta f}$ the matrix of eigenvalues and $Q^{\Delta f}$ the $r \times r$ matrix of normalized eigenvectors. We can also define a matrix $n \times r$ as $L = \Lambda Q^{\Delta f} \big(N^{\Delta f}\big)^{1/2}$. Under Assumption 3 we have that there is an $\overline{n}$ big enough such that for all $n > \overline{n}$ we have:

$$\frac{L'L}{n} = N_{\Delta f}, \tag{18}$$

as we have $\Lambda' \Lambda / n = I_r$ while $Q^{\Delta f'} Q^{\Delta f} = I_r$ by definition. Now by Assumption 1 we have that all eigenvalues of $\Sigma_{\Delta f}$ are finite and positive, in other words:

$$0 < \underline{c}_1 \leq \nu_i^{\Delta_f} \leq \overline{c}_1 < \infty. \tag{19}$$

Then for all $n > \overline{n}$:

$$\frac{\Sigma_{\Delta\zeta}}{n} = \frac{\Lambda'\Sigma_{\Delta f}\Lambda}{n} = \frac{\Lambda' Q^{\Delta f} N^{\Delta f} Q^{\Delta f'}\Lambda}{n} = \frac{LL'}{n}.$$

Then the non-zero eigenvalues of the covariance matrix of $\frac{\Sigma_{\Delta\zeta}}{n}$ are the same as the ones of $\Sigma_{\Delta f}$ by (18). Then the lemma follows by (19).

2) Note that we can write in vector form the model for the idiosyncratic processes in 6 as:

$$\Delta \boldsymbol{v}_t = \sum_{h=-\infty}^{\infty} C_h \boldsymbol{\eta}_{i,t+h}.$$

Then,

$$\nu_1^{\Delta v} = \big|\big|\Sigma^{\Delta v}\big|\big| \leq \sum_{h=-\infty}^{\infty} \big|\big|C_h\big|\big|^2 \big|\big|\Sigma^{\eta}\big|\big|.$$

Under Assumption 4.3 we have that:

$$\sup_{i=1,\dots,n} \sum_{h=-\infty}^{\infty} \psi_{ih}^2 \leq K.$$

Moreover we have that by 4.2:

$$\frac{1}{n}\sum_{j,i=1}^{n} \big|\mathbb{E}\big[\eta_{it}\eta_{jt}\big]\big| \leq \max_j \sum_{i=1}^{n} \big|\mathbb{E}\big[\eta_{it}\eta_{jt}\big]\big| = \big|\big|\Sigma^{\eta}\big|\big|_1 \leq M.$$

The Lemma statement follows from the inequality $\big|\big|\Sigma^{\eta}\big|\big| \leq \big|\big|\Sigma^{\eta}\big|\big|_1$, such that,

$$\nu_1^{\Delta v} \leq \sum_{h=-\infty}^{\infty} \big|\big|C_h\big|\big|^2 \big|\big|\Sigma^{\eta}\big|\big| \leq K\big|\big|\Sigma^{\eta}\big|\big|_1 \leq KM = \bar{c}_2.$$

3) To prove the final part note that under the independence assumption between the innovations of the common factors and of the idiosyncratic components (Assumption 4.4) we have,

$$\Sigma^{\Delta y} = \Sigma^{\Delta\zeta} + \Sigma^{\Delta v}.$$

Then using Weyl's inequality and the previous results we have that for $n \geq \overline{n}$, for $j = 1, ..., r$ we have an upper bound:

$$\frac{\nu_i^{\Delta y}}{n} \leq \frac{\nu_i^{\Delta \zeta}}{n} + \frac{\nu_1^{\Delta v}}{n} \leq \bar{c}_1 + \frac{\nu_1^{\Delta v}}{n} \leq \bar{c}_1 + \frac{\bar{c}_2}{n},$$

and a lower bound,

$$\frac{\nu_i^{\Delta y}}{n} \geq \frac{\nu_i^{\Delta \zeta}}{n} + \frac{\nu_n^{\Delta v}}{n} \geq \underline{c}_1 + \frac{\nu_n^{\Delta v}}{n} \geq \underline{c}_1.$$

On the other hand for $j = r+1$ we have that $rank(\Sigma^{\Delta\zeta}) = r$, then:

$$\nu_{r+1}^{\Delta y} \leq \nu_{r+1}^{\Delta \zeta} + \nu_1^{\Delta v} \leq \nu_1^{\Delta v} \leq \bar{c}_2.$$

Lemma 2 *Define $\hat{\Sigma}_{ij}^{\Delta y}$ the generic element on the $i-$th row, $j-$th column of $\hat{\Sigma}_{\Delta y}$ the sample covariance matrix of Δy_t, then we have:*

$$\left|\hat{\Sigma}_{ij}^{\Delta y} - \Sigma_{ij}^{\Delta y}\right| = O_p\left(T^{-\delta}\right), \quad \textit{for } \delta \in (0, 1/2]$$

Proof The proof of this lemma follows the proof of Lemma D.3 in Barigozzi et al. (2021), using the fact that the convergence of the covariance of infinite MA processes with regularly varying tail distributions has a known convergence rate by Davis and Resnick (1986). Under Assumption 1.2 we can represent the process $\Delta \boldsymbol{f}_t$ as the infinite two sided MA process:

$$\Delta \boldsymbol{f}_t = \sum_{h=-\infty}^{\infty} \Gamma_h \boldsymbol{\varepsilon}_{t+h},$$

with $||\Gamma_h||$ decaying at a geometric rate as $h \to \infty$. Then we have,

$$\begin{aligned}
\mathbb{E}\left[\Delta \boldsymbol{f}_t \Delta \boldsymbol{f}_t'\right] &= \mathbb{E}\left[\left(\sum_{h=-\infty}^{\infty} \Gamma_h \boldsymbol{\varepsilon}_{t+h}\right)\left(\sum_{h=-\infty}^{\infty} \Gamma_h \boldsymbol{\varepsilon}_{t+h}\right)'\right] \\
&= \sum_{h=-\infty}^{\infty} \Gamma_h \mathbb{E}\left[\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t}'\right]\Gamma_h' \\
&= \sum_{h=-\infty}^{\infty} \Gamma_h \Gamma_h' = \Sigma_f,
\end{aligned}$$

where we used 1 for the independence over time of the innovations of the common factors, the fact that $\mathbb{E}\left[\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t}'\right] = \mathbb{I}_r$ and the square summability of the coefficient matrices, then Σ_f is well defined. Now we have,

$$\left|\left|\hat{\Sigma}_{\Delta f} - \Sigma_{\Delta f}\right|\right| \leq \max_j \sum_{i=1}^r \left|\hat{\Sigma}_{ij}^{\Delta f} - \Sigma_{ij}^{\Delta f}\right|.$$

Note that for $i \neq j$ we have:

$$\begin{aligned}
\left(\hat{\Sigma}_{ij}^{\Delta f} - \Sigma_{ij}^{\Delta f}\right)^2 &= \frac{1}{T^2}\sum_{t,s}^T \left(\Delta f_{j,t}\Delta f_{i,t} - \Sigma_{ij}^{\Delta f}\right)\left(\Delta f_{j,s}\Delta f_{i,s} - \Sigma_{ij}^{\Delta f}\right) \\
&= \frac{1}{T^2}\sum_{t,s}^T \left(\Delta f_{j,t}\Delta f_{i,t}\Delta f_{j,s}\Delta f_{i,s} - \left(\Sigma_{ij}^{\Delta f}\right)^2\right) \\
&= \frac{1}{T^2}\sum_{t,s}^T \sum_{d,d',l,l'}^q \left(\sum_{h,m,n,r=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id',m}\zeta_{jl,n}\zeta_{il',r}\varepsilon_{d,t+h}\varepsilon_{d',t+m}\varepsilon_{l,s+n}\varepsilon_{l',s+r}\right),
\end{aligned}$$

as by Assumption 1.4, $\Sigma^{\Delta f}$ is a diagonal matrix. Then,

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t,s}^T \sum_{d,d',l,l'}^q \left(\sum_{h,m,n,r=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id',m}\zeta_{jl,n}\zeta_{il',r}\varepsilon_{d,t+h}\varepsilon_{d',t+m}\varepsilon_{l,s+n}\varepsilon_{l',s+r}\right) \\
&= \frac{1}{T^2}\sum_{t=1}^T \sum_{d=1}^q \left(\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id,h}\zeta_{jd,r}\zeta_{id,r}\varepsilon_{d,t+h}^4\right) \\
&+ \frac{1}{T^2}\sum_{t=1}^T \sum_{d,l=1}^q \left(\sum_{h=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id,h}\zeta_{jd,m}\zeta_{il,r}\varepsilon_{d,t+h}^3\varepsilon_{l,t+r}\right) \\
&+ \frac{1}{T^2}\sum_{t=1}^T \sum_{d,l=1}^q \left(\sum_{h,m,r,s=-\infty}^{\infty} \zeta_{jd,h}\zeta_{jd,m}\zeta_{il,r}\zeta_{il,s}\varepsilon_{d,t+h}^2\varepsilon_{l,t+r}^2\right) \\
&+ \frac{1}{T^2}\sum_{t=1}^T \sum_{d,l=1}^q \left(\sum_{h,m,r,s=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id,m}\zeta_{jl,r}\zeta_{il,s}\varepsilon_{d,t+h}^2\varepsilon_{l,t+r}^2\right) \\
&+ \frac{1}{T^2}\sum_{t,s=1}^T \sum_{d,l=1}^q \left(\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id,h}\zeta_{jl,r}\zeta_{il,r}\varepsilon_{d,t+h}^2\varepsilon_{l,s+r}^2\right) \\
&+ \frac{1}{T^2}\sum_{t,s}^T \sum_{d\neq d'\neq l\neq l'}^q \left(\sum_{h,r,m,n=-\infty}^{\infty} \zeta_{jd,h}\zeta_{id',p}\zeta_{jd,q}\zeta_{id',r}\varepsilon_{d,t+h}\varepsilon_{d',t+m}\varepsilon_{d,s+n}\varepsilon_{d',s+r}\right) \\
&+ G \\
&= A + B + C + D + E + F + G,
\end{aligned}$$

with H a remainder component of combinations of terms that go to zero by independence between the errors. Note that,

$$\begin{aligned}
\mathbb{E}\big[\Delta f_{it}\Delta f_{jt}\big] &= \mathbb{E}\Big[\sum_{d,l}^{q}\sum_{h,k}^{\infty}\zeta_{id,h}\zeta_{jl,k}\varepsilon_{d,t-h}\varepsilon_{l,t-k}\Big] \\
&= \mathbb{E}\Big[\sum_{d=1}^{q}\sum_{h=-\infty}^{\infty}\zeta_{id,h}\zeta_{jd,h}\varepsilon^2_{d,t-h}\Big] \\
&= \sum_{d=1}^{q}\sum_{h=-\infty}^{\infty}\zeta_{id,h}\zeta_{jd,h}\mathbb{E}\big[\varepsilon^2_{d,t-h}\big] \\
&= \sum_{d=1}^{q}\sum_{h=-\infty}^{\infty}\zeta_{id,h}\zeta_{jd,h} = 0,
\end{aligned}$$

due to Assumption 3.2, for all $i, j = 1, ..., r$. This means that $\zeta_{id,h}\zeta_{jd,h} = 0$ for all $i, j = 1, ..., r$, $d = 1, ..., q$, $h = -\infty, \infty$.

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{d=1}^{q}\Bigg(\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\zeta_{jd,h}\zeta_{id,h}\zeta_{jd,r}\zeta_{id,r}\varepsilon^4_{d,t+h}\Bigg) = 0,$$

that is $A = 0$. If Δf_{it} and Δf_{jt} would depend by the same underlying shock, then $\Sigma^{\Delta f}_{ij}$ would not be diagonal. Note also that,

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{d,l=1}^{q}\Bigg(\sum_{h=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\zeta_{jd,h}\zeta_{id,h}\zeta_{jd,m}\zeta_{il,r}\varepsilon^3_{d,t+h}\varepsilon_{l,s+r}\Bigg) = 0 \\
&\frac{1}{T^2}\sum_{t=1,s\neq t}^{T}\sum_{d,l=1}^{q}\Bigg(\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\zeta_{jd,h}\zeta_{id,h}\zeta_{jl,r}\zeta_{il,r}\varepsilon^2_{d,t+h}\varepsilon^2_{l,s+r}\Bigg) = 0,
\end{aligned}$$

for the same reason. For the rest of the proof we can start from the remainder term, since $\varepsilon_{i,t}$ is an i.i.d sequence with a bounded second moment and independent from $\varepsilon_{j,s}$ for all $j \neq i$ and all s then we have that:

$$\frac{1}{T^2}\sum_{t,s}^{T}\sum_{d,d'}^{q}\Bigg(\sum_{h,r=-\infty}^{\infty}\sum_{m\neq q,n\neq p}\zeta_{jd,h}\zeta_{id',p}\zeta_{jd,q}\zeta_{id',r}\varepsilon_{d,t+h}\varepsilon_{d',t+m}\varepsilon_{d,s+n}\varepsilon_{d',s+r}\Bigg) \xrightarrow{p} 0.$$

Then we have that $F = o_p(1)$. The same holds for component G. For the same reasoning, using the square summability of the coefficients of the inifinite MA representation of the process we have that:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{d,l=1}^{q}\left(\sum_{h,m,r,s=-\infty}^{\infty}\zeta_{jd,h}\zeta_{jd,m}\zeta_{il,r}\zeta_{il,s}\varepsilon_{d,t+h}^{2}\varepsilon_{l,t+r}^{2}\right)\xrightarrow{p} q^2K\mathbb{E}\big[\varepsilon_{d,t}^2\big]\mathbb{E}\big[\varepsilon_{l,t}^2\big]=M<\infty.$$

Then we have that $C = O_p(1/T)$ and the same is true for D. Then for $i \neq j$ $\big|\hat{\Sigma}_{ij}^{\Delta f}-\Sigma_{ij}^{\Delta f}\big| = O_p\big(1/\sqrt{T}\big)$. These passages are simplified without taking into account the sum of the q errors, as the convergence results does not change. Now we take into account the terms such that $i = j$. We know that by Theorem 2.2 in Davis and Resnick (1986) under Assumption 1.1, we have:

$$\frac{T}{a_T^2}\big(\hat{\Sigma}_{ii}^{\Delta f}-\Sigma_{ii}^{\Delta f}\big)\xrightarrow{d}\left(S-\frac{\alpha}{\alpha-2}\right)\Big|\sum_{h=-\infty}^{\infty}\zeta_h^2\Big|^{\alpha/2},$$

where S is a stable random variable with index $\alpha/2$ and a_T is a regularly varying sequence with index $1/\alpha$. This means that:

$$P\left(\frac{T}{a_T^2}\big|\hat{\Sigma}_{ii}^{\Delta f}-\Sigma_{ii}^{\Delta f}\big|>x\right)=x^{-\alpha/2}L(x).$$

and $a_T = T^{1/\alpha}L(T)$, with $L(\cdot)$ a slowly varying function. Then $\big|\hat{\Sigma}_{ii}^{\Delta f}-\Sigma_{ii}^{\Delta f}\big| \approx O_p\big(T^{-\frac{2-\alpha}{\alpha}}\big)$. This means that:

$$\big|\big|\hat{\Sigma}_{\Delta f}-\Sigma_{\Delta f}\big|\big| = O_p\big(T^{-\delta}\big),\quad \text{with } 0<\delta\le 1/2 \text{ defined as } \delta\le(2-\alpha)/\alpha.$$

In a similar way we have that $\Sigma_{\Delta v}$ is well defined, moreover defining $\Sigma_{ij}^{\Delta v}$ its i,j-th element, we have:

$$\begin{aligned}
&\mathbb{E}\Big|\Big|T^{-1}\sum_{t=1}^{T}\Delta v_{i,t}\Delta v_{j,t}-\Sigma_{ij}^{\Delta v}\Big|\Big|^2\\
&\le\frac{1}{T^2}\sum_{i,j=1}^{r}\mathbb{E}\Big[\sum_{t,s=1}^{\infty}\Big(\Delta v_{i,t}\Delta v_{j,t}-\Sigma_{ij}^{\Delta v}\Big)\Big(\Delta v_{i,s}\Delta v_{j,s}-\Sigma_{ij}^{\Delta v}\Big)\Big]\\
&\le\frac{1}{T^2}\mathbb{E}\Big[\sum_{t,s=1}^{\infty}\Big(\Delta v_{i,t}\Delta v_{j,t}\Delta v_{i,s}\Delta v_{j,s}-\big(\Sigma_{ij}^{\Delta v}\big)^2\Big)\Big].
\end{aligned}$$

Now recalling that $\Delta v_{i,t}=\sum_{h=-\infty}^{\infty}\xi_{i,h}\eta_{i,t+h}$, we have:

$$
\begin{aligned}
\sum_{s,t=1}^{\infty} \mathbb{E}\big[\Delta v_{i,t}\Delta v_{j,t}\Delta v_{i,s}\Delta v_{j,s}\big] &= \sum_{s,t=1}^{\infty} \sum_{h,m,n,q=-\infty}^{\infty} \mathbb{E}\big[\xi_{i,h}\eta_{i,t+h}\xi_{j,m}\eta_{j,t+m}\xi_{i,n}\eta_{i,s+n}\xi_{j,q}\eta_{i,s+q}\big] \\
&\leq K^4 \sum_{s,t=1}^{\infty} \mathbb{E}\big[\eta_{i,t}\eta_{j,t}\eta_{i,s}\eta_{i,s}\big] \\
&= K^4 \sum_{s,t=1}^{\infty} \Big(\mathbb{E}\big[\eta_{i,t}^2\eta_{j,t}^2\big] + \mathbb{E}\big[\eta_{i,t}^4\big] + \mathbb{E}\big[\eta_{i,t}^2\big]\mathbb{E}\big[\eta_{j,t}^2\big]\big]\Big).
\end{aligned}
$$

Then:

$$
\begin{aligned}
\mathbb{E}\bigg|\bigg|T^{-1}\sum_{t=1}^{T}\Delta v_{i,t}\Delta v_{j,t} - \Sigma_{ij}^{\Delta v}\bigg|\bigg|^2 &\leq \frac{K^4}{T^2}\sum_{t,s=1}^{\infty}\mathbb{E}\big[\eta_{i,t}^2\eta_{j,t}^2\big] + \frac{K^4}{T^2}\sum_{t,s=1}^{\infty}\mathbb{E}\big[\eta_{i,t}^4\big] - \frac{K^4}{T^2}\sum_{t,s=1}^{\infty}\mathbb{E}\big[\eta_{i,t}^2\big]^2 \\
&\leq \frac{K^4}{T^2}\sum_{t,s=1}^{\infty}\mathbb{E}\big[\eta_{i,t+h}^2\eta_{j,t+h}^2\big] \leq \frac{cK^4}{T} = O\big(\frac{1}{T}\big),
\end{aligned}
$$

where we used the Assumption 1 of the independence of the innovation of the idiosyncratic components, the existence of their fourth moment and the square summability of the coefficients. Combining the previous results and the fact that we have bounded loadings we have that:

$$
\big|\hat{\Sigma}_{ij}^{\Delta y} - \Sigma_{ij}^{\Delta y}\big| = O_p\big(T^{-\delta}\big),
$$

uniformly over i, j.

C Testing Procedure for Multimodal Predictive Densities

We use the Brier score, from Brier (1950), is computed as,

$$
BS = \sum_{t=1}^{T}(p_t - o_t)^2,
$$

where p_t is the probability of our event and o_t is the realization of that event (1 if it happens, 0 otherwise). The range of this score is between 0 and 1. We now consider the multicategory Brier score defined as,

$$
BS = \frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{R}(p_{t,r} - o_{t,r})^2,
$$

where the r represents the different events and they must be such that $\sum_{r=1}^{R} p_{t,r} = 1$ for all t and $o_{t,r} = 1$ only for one r and it is 0 for the others. The range of this score is between 0 and 2. This multicategory score allows us to compare interval forecast. Since we are interested in prediction during a bubble (so we want to correctly address sharp increases and crashes) we consider as category movements that are within or outside the range of one standard deviation of a baseline Gaussian random walk. Our categories will then be,

$$p_{t,r} = \begin{cases} \mathbb{1}_{\Delta y_t < -\sigma_{rw}} & \text{if } r = 1 \\ \mathbb{1}_{|\Delta y_t| < \sigma_{rw}} & \text{if } r = 2 \\ \mathbb{1}_{\Delta y_t > \sigma_{rw}} & \text{if } r = 3. \end{cases} \tag{20}$$

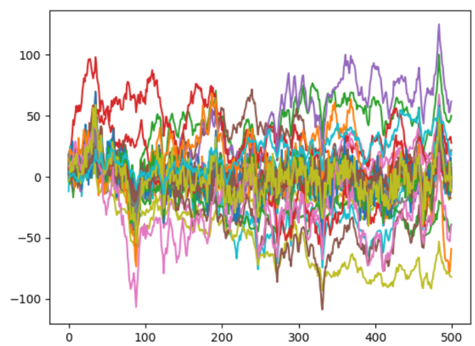
With these scores we can create a Diebold Mariano test statistic. The test statistic for the multicategory Brier score will be,

$$d_t = \sum_{r=1}^{R} (p_{m,rt} - o_{m,rt})^2 - \sum_{r=1}^{R} (p_{i,rt} - o_{i,rt})^2$$

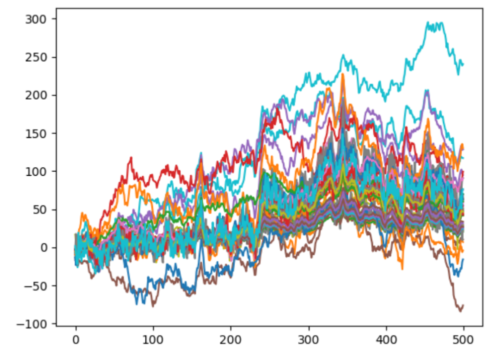
$$DM = \sqrt{T} \frac{\bar{d}}{\sigma_d}.$$

where $\sigma_d = \sqrt{\hat{\gamma}(0) + 2 \sum_{i=1}^{k} w_i \hat{\gamma}(i)}$, with k is of the same order as the square root of the test sample size and $w_i = 1 - i/k$.

D Simulation Plots

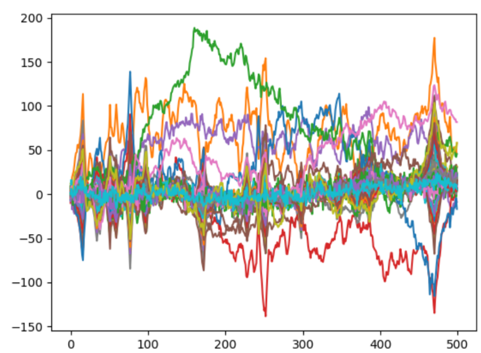


(a) Single stationary explosive common factor.

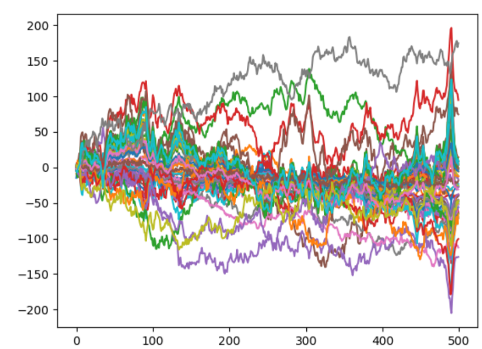


(b) Single non-stationary explosive common factor.

Figure 9: Example of simulated samples from a data generating process with a single common factor.



(a) Stationary explosive common factor.



(b) Non-stationary explosive common factor.

Figure 10: Example of simulated samples from a data generating process with two common factors, one of the two with explosive dynamics and the other as a random walk.