# Characteristic function-based factor modelling of affine jump diffusions using options*

H. Peter Boswijk
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Roger J. A. Laeven
Amsterdam School of Economics
University of Amsterdam, EURANDOM
and CentER

Niels Marijnen
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Evgenii Vladimirov
Econometric Institute
Erasmus University Rotterdam
and Tinbergen Institute

Preliminary version: February 29, 2024

**Abstract**

We develop a framework to analyse affine jump diffusions using factor modelling techniques, offering a novel method to study which and how many risk factors drive the price process of a single asset. We use information contained in options to construct observations on the characteristic function of the returns on the underlying asset, without having to specify a parametric model. We show how to form a linear factor model out of these observations. Our asymptotic framework is one in which the number of observed options, of varying strikes, tends to infinity. We prove asymptotic normality of the factor model, and provide a feasible central limit theorem which can be used for testing. In addition, we prove that diagnostic criteria based on the eigenvalues of the sample covariance matrix of the constructed factor model are able to consistently estimate the number of factors, and that principal component analysis is able to extract these factors. An empirical application suggests that the main factor driving the S&P 500 returns is a stochastic variance process, which explains 97% of the variation in the factor model.

*We are very grateful to Yi He, Kris Jacobs, Frank Kleibergen, Siem Jan Koopman, Olivier Scaillet, participants in the 2022 $(\mathrm{EC})^2$ Conference at ESSEC Business School in Paris, the 2023 Financial Econometrics Conference at Lancaster University, the 2023 Econometrics Workshop at the Erasmus University Rotterdam, the 2023 Aarhus Econometrics Workshop, the 2023 Netherlands Econometric Study Group, the 2023 QFFE conference at Aix-Marseille Université, the 2023 IAAE conference at BI Norwegian Business School in Oslo, the 2023 Tinbergen Institute Time Series Econometrics Theory Workshop at VU Amsterdam, the 2023 FinEML conference at the Erasmus University Rotterdam, and in seminars at the University of Amsterdam and the Tinbergen Institute for helpful comments. Niels Marijnen acknowledges financial support by the IAAE travel grant. *E-mail addresses:* H.P.Boswijk@uva.nl, R.J.A.Laeven@uva.nl, N.Marijnen@uva.nl, and Vladimirov@ese.eur.nl.

# 1 Introduction

The value of financial derivatives is determined by the stochastic behaviour of the price of the underlying asset. A natural, fundamental question is then to ask what kind of process underlies the price of such assets. The consensus is that asset pricing models need to be more involved than in the seminal work of Black and Scholes (1973), but the literature has not yet settled on what should be included in a "proper" model. The amount of research devoted to this question is vast, with many new models or extensions being proposed (see, among many others, Merton, 1976; Hull and White, 1987; Heston, 1993; Bates, 2000; Aït-Sahalia et al., 2015).

In this paper, we develop a model-free methodology to analyse what characteristics need to modelled in order to appropriately represent the stochastic process underlying an equity price. We propose a factor model of the characteristic function of the equity returns, exploiting information contained in option prices. In usual factor models in asset pricing, a small number of common risk factors drives the behaviour of a large set of returns. We, in contrast, are interested in the behaviour of the latent factors driving the distribution of a *single* asset price. These factors are best understood as the state variables that need to be included in a model, e.g., a stochastic volatility "factor". We develop methods to perform inference on both the number of factors, and their behaviour.

This question is not novel, nor is the use of options in answering it; the added value of our approach lies in the theoretical validity of the factor model representation, which is not present in the literature. Setting up a formal framework in which this factor analysis can be validly applied starts by appropriately transforming the option data: we construct observations on the characteristic function of the distribution of the returns on the underlying asset in a model-free way, using the Carr and Madan (2001) spanning result (see also Todorov, 2019). We perform this spanning for option panels differing in time-to-maturity, which allows us to study factors affecting the term-structure of the option prices as well as factors causing cross-sectional differences.

Existing research used principal component analysis (PCA) to describe implied volatility (IV) surfaces (Skiadopoulos et al., 2000; Cont and Da Fonseca, 2002; Andersen et al., 2015b). IVs are usually considered as a method of quoting option prices, which result from the dynamics in other state variables. However, by constructing a factor model directly for the IVs, the focal point is shifted to the IVs as the process of interest, away from the underlying state variables; an approach explicitly taken in Carr and Wu (2020). Another justification would be that IVs are linear functions of the actual state variables, but this approach is not consistent with most models, as usual option-pricing formulae are highly nonlinear.[1] This divergence is not present in our framework, but to ensure the validity of the linear factor model representation, a structural form for the underlying dynamics has to be assumed. In our option-based setting, this "necessary evil" presents itself as a restriction to the class of affine jump diffusions of Duffie et al. (2000). This broad and widely used class naturally lends itself to factor modelling, as its characteristic function is of the exponential affine form. By considering the characteristic function spanned by

[1] Volatility surface analysis is also the focal point of Aït-Sahalia et al. (2021a,b), though they take a Taylor expansion approach and examine derivatives, or "shape-characteristics", to summarize the surface. The *ad hoc* factor model assumption is then not required.

the option prices, instead of the IV surface directly, we are thus able to construct a factor model while maintaining the view that IVs are an outcome, and not the foundation, of a stochastic process. The unavoidable trade-off is that we have to limit the scope of our framework to the analysis of affine models. Fortunately, this affine class contains many popular option-pricing models, and it is often used in the financial literature, such as in Black and Scholes (1973), Heston (1993), Bates (2000), Pan (2002), and Aït-Sahalia et al. (2015).

Though it allows for a wide variety of popular models, the restriction to the affine class clearly limits the applicability of the developed framework. However, our characteristic function observations are generated in a model-free way; even if the true underlying process is not affine and the linear representation is invalid, there is some merit in the idea that PCA extracts factors that explain (part of) the dynamic behaviour of this characteristic function. The comparison to the aforementioned studies on PCA in the IV surface can easily be drawn; even if the volatility surface is not truly linearly dependent on some common factors, these studies still seek to learn some general behaviour through its principal components. In the same sense, one can read this current study as an exploratory analysis of what factors drive the dynamics of the characteristic function: an interesting topic on its own.

We derive the asymptotic theory for our factor model. Our asymptotic scheme is one in which we require the number of options, of varying strikes, to tend to infinity. Such an asymptotic scheme is also implemented in, among others, Andersen et al. (2021), and in Boswijk et al. (2022), of which our article can be seen as a non-parametric, more exploratory counterpart. The resulting framework can be regarded as one using so-called "small-sigma" asymptotics: the asymptotic set-up (indirectly) causes the errors to vanish, while the cross-sectional and time series dimensions of the factor model stay finite. It is important to stress this final point: as there is no requirement for the time series dimension, our method can be applied on any scale ranging from intradaily to, e.g., weekly observations, as long as there are observations on sufficiently many options. We provide a feasible CLT which can be used to construct tests, e.g., for the number of factors using the Kleibergen and Paap (2006) rank test. In addition, we prove the validity of the use of PCA to extract the state vector, and that inference based on the eigenvalues of the sample covariance matrix of the factor model can consistently estimate the number of factors.

We examine the performance of the developed methodology in a Monte Carlo study, considering a setting mimicking the S&P 500 using an empirically relevant model with a (multi-factor) stochastic volatility structure. Focusing on the aforementioned rank test for the number of factors, we find excellent power in all considered settings, regardless of tuning parameters. The same can be said about the size of the test, though finite sample biases cause it to slightly over-reject in the more involved models. A small empirical application highlights that the most important factor driving the S&P 500 index is related to a stochastic variance process, but that more factors are necessary to explain higher order moments.

The main contribution of this paper is the development of the characteristic function-based factor modelling approach to analyse option-pricing models. The framework is constructed to answer the fundamental question of what processes drive an asset price, but in contrast to the existing factor literature, we do not focus on a large panel of equity returns or option prices directly. To the best of our knowledge, our method is unique in its characteristic function-based factor modelling approach.

The rest of this article is organised as follows. Section 2 discusses the affine models, and the construction of the factor model. Section 3 contains the theoretical results. Section 4 presents the simulation results. Section 5 contains an empirical application. Section 6 concludes.

## 2 Methodology

This section explains how to construct the linear factor model in the affine class, starting with a short treatment of this affine class, and the option-based spanning of the conditional characteristic function.

### 2.1 The affine jump diffusion

Consider an arbitrage-free, dynamic financial market, defined on a filtered probability space $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, \left\{\mathcal{F}_t^{(0)}\right\}_t, \mathbb{P}^{(0)}\right)$. Absence of arbitrage implies the existence of a locally equivalent risk-neutral measure $\mathbb{Q}$. Consider a single asset price $S_t$, of which the dynamics under $\mathbb{Q}$ are driven by a $d$-dimensional state vector $X_t$, assumed to be a Markov process taking its values in $\mathcal{X} \subset \mathbb{R}^d$. As usual, we describe the dynamics of $\log S_t =: y_t$, instead of $S_t$ directly. The dynamics of $(y_t, X_t')'$ under the risk-neutral measure $\mathbb{Q}$ are described by the following jump diffusion:

$$\begin{aligned} \mathrm{d}y_t &= \mu_y(X_t)\,\mathrm{d}t + \sigma_y(X_t)\,\mathrm{d}W_t + Z_t^y\,\mathrm{d}N_t\,, \\ \mathrm{d}X_t &= \mu_X(X_t)\,\mathrm{d}t + \sigma_X(X_t)\,\mathrm{d}W_t + Z_t^X\,\mathrm{d}N_t\,, \end{aligned} \tag{2.1}$$

where $W_t$ is a $d+1$-dimensional standard Brownian motion, $N_t$ is a counting process with arrival rate $\lambda(X_t)$, and the associated jump-sizes are represented by $Z_t := (Z_t^y, Z_t^{X\prime})'$. Conditional on a jump-event at time $t$, the jump-size $Z_t$ is a random vector following the arbitrary, but fixed, distribution $\nu$. The size of the jumps is thereby independent of the current state.[2] Assume moreover the existence of a riskless asset offering, for simplicity, a constant rate of return of $r$. In addition, assume $\mu_y$ satisfies no-arbitrage restrictions such that the discounted asset price is a $\mathbb{Q}$-local martingale.

We follow Duffie et al. (2000) and impose that the functions $\mu(\cdot), \sigma(\cdot)\sigma(\cdot)'$ and $\lambda(\cdot)$ are affine, where $\mu(\cdot) := (\mu_y(\cdot), \mu_X(\cdot)')'$, and with $\sigma(\cdot)$ defined similarly. That is, we assume that they can be written as:

$$\begin{cases} \mu(x) &= K_0 + K_1 x, & \text{with } K_0 \in \mathbb{R}^{d+1}, K_1 \in \mathbb{R}^{d+1\times d} \\ \sigma(x)\sigma(x)' &= H_0 + \sum_{j=1}^d H_1^{(j)} x_j, & \text{with } H_0 \in \mathbb{R}^{d+1\times d+1}, H_1^{(j)} \in \mathbb{R}^{d+1\times d+1} \\ \lambda(x) &= l_0 + l_1' x, & \text{with } l_0 \in \mathbb{R}, l_1 \in \mathbb{R}^d \end{cases} \tag{2.2}$$

[2]It is possible to extend (2.1) to include multiple jump processes, or jumps specific to prices or volatility, but this is not a (direct) point of interest. Including multiple jump processes does allow for "changes" in the jump-size distribution, by changing the frequency of small relative to big jumps.

Herein, $H_0, H_1^{(j)}$ must be taken such that $\sigma(x)\sigma(x)'$ is positive semi-definite for all $x \in \mathcal{X}$, and $l_0, l_1$ must be such that $\lambda(x) \geq 0$ for all $x \in \mathcal{X}$, and such that it does not explode to infinity either. These so-called admissibility conditions, discussed in Duffie et al. (2003) and Singleton (2006) among others. These admissibility conditions lead to restrictions on the parameter space: a famous example is the Feller condition that ensures the stochastic volatility process stays positive in the Heston (1993) model. We do not discuss this further, but it must be clear that the applicability of our methodology is limited to models with admissible parameter values.

Assuming the state vector follows an affine jump diffusion seems restrictive, but it actually allows for a variety of popular models, as showcased by its use in Black and Scholes (1973), Heston (1993), Bates (2000), Duffie et al. (2000), Pan (2002), Aït-Sahalia et al. (2015), Andersen et al. (2015b), Boswijk et al. (2021), among others. The main appealing feature of the affine jump diffusion is the result of Duffie et al. (2000), who have shown that its conditional characteristic function (CCF) is of the exponential affine form. In particular, given the structure of (2.1), the CCF of the log return on a futures contract can be written as:

$$\phi_t(u,\tau) := \underbrace{\mathbb{E}_t^{\mathbb{Q}}\left[\exp\{iu\log(F_{t+\tau}/F_t)\}\right]}_{CCF} = \exp\big\{\alpha(u,\tau) + \beta(u,\tau)'X_t\big\}, \tag{2.3}$$

where $\tau \in \mathbb{R}_+$ and $u \in \mathbb{R}$. Both here and hereafter, $\mathbb{E}_t^{\mathbb{Q}}[\cdot] := \mathbb{E}^{\mathbb{Q}}[\cdot|\mathcal{F}_t^{(0)}]$ refers to the conditional expectation under the risk-neutral measure $\mathbb{Q}$ given the filtration $\mathcal{F}_t^{(0)}$. The coefficients $\alpha$ and $\beta$ solve a known system of ODEs, that deterministically depends on the CCF argument $u$, the parameters, and the jump distribution, but not on $\mathcal{F}_t^{(0)}$. There are some regularity conditions in order to guarantee a (unique) solution to this set of ODEs, for which we refer the reader to other work, such as Duffie et al. (2000).

Recall that the drift, diffusion, and intensity functions take only $X_t$ as argument, and not $y_t$. This implies the intuitive, and standard, assumption that returns are distributed independently of price levels. This does requires a simple adjustment to the standard definition of the ODEs, but this is not relevant for our exposition, and it is therefore omitted. Solving these ODEs is namely not required in our methodology, which is agnostic about the specific parametric structure of the model as long as it is part of the affine class. This exponential affine structure is the feature we aim to exploit, as it implies that the log[3] conditional characteristic function (CCF) of such a $d$-dimensional affine jump diffusion is affine in the state vector.

## 2.2 Spanning the characteristic function with an option-portfolio

Let $F_t$ be the price associated to a futures contract expiring at $T := t + \tau$, following some arbitrary martingale-dynamics under the risk-neutral measure $\mathbb{Q}$:

$$\frac{\mathrm{d}F_t}{F_{t-}} = v_t \mathrm{d}W_t + \mathrm{d}\tilde{J}_t\,, \tag{2.4}$$

[3]The logarithm of a complex number $z = re^{i\varphi}$ is not uniquely defined. The common solution, to focus on the "principal branch" and take $k \in \mathbb{Z}$ such that $\mathfrak{Im}(\log z) = \varphi + 2k\pi \in (-\pi, \pi]$, introduces discontinuities at the crossing of the interval's boundaries. As the log CCF is required to be continuous, we simply recall that $\phi(0) = 1$, or $\log\phi(0) = 0$, and take the aforementioned $k$ such that the function is continuous in $u$.

with $W_t$ a standard Brownian motion, $v_t$ an unspecified, but adapted and locally bounded volatility process, and $\tilde{J}_t$ a similarly unspecified compensated jump process of finite variation. Assume moreover, as before, the existence of a riskless asset with constant risk-free rate $r$. Note that these dynamics are not necessarily of the affine form as discussed before, but more general.

As shown by Carr and Madan (2001), any twice continuously differentiable function $f(\cdot)$ of the terminal price $F_{t+\tau}$ can be replicated by a portfolio of options as follows:

$$f(F_{t+\tau}) = f(x) + f'(x)(F_{t+\tau} - x) + \int_x^\infty f''(K)(F_{t+\tau} - K)^+ \,\mathrm{d}K + \int_0^x f''(K)(K - F_{t+\tau})^+ \,\mathrm{d}K\,, \tag{2.5}$$

with $x \in \mathbb{R}_+$ fixed. Denote by $O_t(\tau, K)$ the time-$t$ price of an out-of-the-money (OTM) European option which matures at $t+\tau$ and has strike $K$, that is, a put if $K \leq F_t$ and a call if $K > F_t$. Risk-neutral pricing dictates that this price is equal to the discounted $\mathcal{F}_t^{(0)}$-conditional expectation of the terminal payoff under the risk-neutral measure $\mathbb{Q}$. As such, taking $x = F_t$, we can write:

$$\mathbb{E}_t^{\mathbb{Q}}[f(F_{t+\tau})] = f(F_t) + e^{r\tau} \int_0^\infty f''(K) O_t(\tau, K)\,\mathrm{d}K\,. \tag{2.6}$$

Taking $f(x) = e^{iu\log(x/F_t)}$ for $u \in \mathbb{R}$, we find:

$$\begin{aligned}
\phi_t(u,\tau) &:= \mathbb{E}_t^{\mathbb{Q}}\left[\exp\{iu\log(F_{t+\tau}/F_t)\}\right] \\
&= 1 - e^{r\tau}(u^2 + iu)\int_0^\infty \frac{1}{K^2} e^{iu\log(K/F_t)} O_t(\tau, K)\,\mathrm{d}K \\
&= 1 - e^{r\tau}\frac{u^2 + iu}{F_t}\int_{\mathbb{R}} e^{(iu-1)m} O_t(\tau, m)\,\mathrm{d}m\,,
\end{aligned} \tag{2.7}$$

with $m := \log(K/F_t)$ the log-moneyness of an option with strike price $K$. Note that we use both $O_t(\tau, K)$ and $O_t(\tau, m)$ for the same option price; given the one-to-one relation between $K$ and $m$ there should be no confusion. The left-hand side of (2.7) is the conditional characteristic function of the log futures returns. As such, this spanning result allows us to construct functional observations on the CCF by using the current futures price and a portfolio of OTM options on a continuum of strikes. This idea is also used by Todorov (2019) to construct a spot volatility estimator using the Lévy-Khintchine formula. Note that these observations are model-free, and thereby do not rely on the assumption that the futures price follows an affine jump diffusion.

In practice, we must approximate the integral in (2.7): we do not have a continuum of options, and the options we do have are noisy. In our framework, we indeed allow for observation errors in the option prices:

$$\widehat{O}_t(\tau, m) := O_t(\tau, m) + \zeta_t(\tau, m), \tag{2.8}$$

where the observation errors can be factorized as $\zeta_t(\tau, m) := \sigma_t(\tau, m)\varkappa_t(\tau, m)$. Herein, $\sigma_t(\tau, m)$ is an $\mathcal{F}_t^{(0)}$-adapted random variable capturing the conditional heteroskedasticity, and $\{\varkappa_t(\tau, m_j)\}_j$ is a random sequence defined on the outcome space $\Omega^{(1)} = \bigtimes_t \bigtimes_\tau \mathbb{R}^{\mathbb{R}}$. Similar to, e.g., Andersen et al. (2021), we require a probability space of this form as the in-fill asymptotics imply the need to define an error for every value in the moneyness-grid. We complete the probability space with the product Borel $\sigma$-field $\mathcal{F}^{(1)}$ and transition probability $\mathbb{P}^{(1)}(\omega^{(0)}, \mathrm{d}\omega^{(1)})$. We join the two spaces by defining the product space:

$$\Omega = \Omega^{(0)} \times \Omega^{(1)},\ \mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)},\ \mathbb{P}(\mathrm{d}\omega^{(0)}, \mathrm{d}\omega^{(1)}) = \mathbb{P}^{(0)}(\mathrm{d}\omega^{(0)})\mathbb{P}^{(1)}(\omega^{(0)}, \mathrm{d}\omega^{(1)}).$$

To implement our approximation, we follow Todorov (2019) and employ a left Riemann sum. There is no particular reason to take a left Riemann sum instead of a right or middle Riemann sum, or even more involved techniques to approximate the integral. The employed approximation[4] is as follows:

$$\hat{\phi}_t(u,\tau) := 1 - e^{r\tau}\frac{u^2+iu}{F_t}\sum_{j=2}^{n} e^{(iu-1)m_{j-1}}\widehat{O}_t(\tau, m_{j-1})\Delta m_j, \tag{2.9}$$

where $n$ is the number of observed options, with associated log-moneyness $\{m_j\}_{j=1}^n$ and noisy prices $\{\widehat{O}_t(\tau, m_j)\}_{j=1}^n$, and $\Delta m_j := m_j - m_{j-1}$ is the difference in log-moneyness between the $j$-th and $(j-1)$-th OTM-option. In our exposition, $n \equiv n_{t,\tau}$ and $m_j \equiv m_{t,\tau}(j)$ for all $t, \tau$. This can be relaxed to a time- and tenor-varying, $\mathcal{F}_t^{(0)}$-adapted grid, but doing so clutters notation: asymptotic results would be denoted in terms of $\max_{t,\tau}\sup_j \Delta m_{t,\tau}(j) \downarrow 0$ as $\min_{t,\tau} n_{t,\tau} \to \infty$.

As already apparent from (2.7), we could approximate the integral in terms of the strike prices $K$ as well. Though the value of the integral is invariant to the choice of variable, this choice affects numerical approximations when only a finite number of strikes are available. It must be noted that the difference is likely small, at least with the current form of the approximation: the functions are evaluated in equivalent points, the only actual numerical difference is that we replace $\log\frac{K_j}{K_{j-1}}$ by $\frac{\Delta K_j}{K_{j-1}}$ in each term of the Riemann sum,[5] which are well known to be close if the steps in the strike grid are not too large. Nonetheless, it might be the case that the discrete approximation is more stable when using $m$, or *vice versa*.

## 2.3 Observing the affine CCF

Aligning the affine CCF of (2.3) with its spanned counterpart defined in (2.9), and taking the complex logarithm, we find the following observation equation:

$$\hat{\psi}_t(u,\tau) := \log\hat{\phi}_t(u,\tau) = \alpha(u,\tau) + \beta(u,\tau)'X_t + \xi_t(u,\tau), \tag{2.10}$$

where the complex-valued noise $\xi_t(u,\tau) := \hat{\psi}_t(u,\tau) - \psi_t(u,\tau)$ captures the observation errors in the option prices $\widehat{O}_t(\tau, m)$, and the errors caused by truncation and discretization of the integration interval in (2.9). These errors are discussed in more detail in Boswijk et al. (2022).

[4]Considering an inter-extrapolation scheme might lead to a better approximation (see Boswijk et al., 2022), and can be used to reduce the finite-sample bias of this approximation. This is discussed below in more detail.

[5]Using $K = F_t e^m$, this can be seen from the last equation (denoted by $(*)$) in the string of identities below

$$\int_0^\infty \frac{1}{K^2}e^{iu\log(K/F_t)}O_t(\tau,K)\,\mathrm{d}K = \frac{1}{F_t}\int_0^\infty e^{(iu-1)m}O_t(\tau,K)\frac{\mathrm{d}K}{K} \overset{(*)}{=} \frac{1}{F_t}\int_{\mathbb{R}} e^{(iu-1)m}O_t(\tau,m)\,\mathrm{d}m\,,$$

such that the only difference in the numerical approximation is the replacement of $\Delta m_j$ by $\frac{\Delta K_j}{K_{j-1}}$.

This is in fact the same observation-equation used by Boswijk et al. (2022), who use it to estimate parametric option pricing models using state space filtering techniques. We present to some extent a non-parametric counterpart to their paper, using the same observation equation but remaining more agnostic about the specific parametric structure. As such, our approach in analysing this equation is clearly also very different.

Note that the CCF under consideration is of the futures returns, and not of the full state vector. Even though the latter is more informative, we are not able to construct the required observations: to be able to span the characteristic function with a portfolio of options, we require it to be a function of the process on which the options are written. As options are written on the price of the underlying, we are restricted to the CCF of the price level, which is transformed to the CCF of returns. There are two main reasons to focus on returns. Firstly, the price process is observable, and thereby not interesting to analyse as part of the factor model. Secondly, the price process is generally assumed non-stationary, ensuring the log CCF is a factor model with a non-stationary factor, which will dominate other stationary factors, and complicate the analysis.

## 2.4 Evaluating the functional factor model

Note that the relation in (2.10) holds for any value of $(u, \tau)$, forming a functional linear factor model. Being interested not in this "direction" of the model, but in the state vector $X_t$, we reduce this to a more practical setting by evaluating the functional model in a given set of arguments for $(u, \tau)$.[6]

Assume at time $t$ we have observations on the futures price and on OTM options of a variety of strikes and of $k \in \mathbb{N}$ different maturities $\{\tau_i\}_{i=1}^k$. In our setting, we assume daily observations and let $t = 1, \ldots, T$, such that we consider $T$ trading days. The observed strike prices are allowed to change, but the elements of $\{\tau_i\}$ are assumed to be constant. A constant $\tau$ implies that the coefficients $\alpha(u, \tau)$ and $\beta(u, \tau)$ in, e.g., (2.10), are time-constant, such that they can indeed be interpreted as factor loadings. Given this information, we use (2.9) to construct the CCF for each $\tau_i$.

The spanned CCFs are then evaluated in $q \in \mathbb{N}$ different arguments, $0 < u_1 < \ldots < u_q$. The focus on the positive real half-line forms no restriction as all characteristic functions are Hermitian. How to optimally select these arguments $\{u_j\}_{j=1}^q$ is an open question, which is a point of interest in our simulations.

For a given maturity $\tau_i$, with $i = 1, \ldots, k$, we then stack our observations on the CCF evaluated in different $u_j$ to arrive at the vector-valued equation below:

$$\hat{\psi}_{t,i} = \alpha_i + \beta_i X_t + \xi_{t,i}, \tag{2.11}$$

with $\hat{\psi}_{t,i} \in \mathbb{C}^q$ with $j$-th element $\hat{\psi}_t(u_j, \tau_i)$, and similar for $\alpha_i$, $\beta_i$ and $\xi_{t,i}$. Recall that $\alpha_i$ and $\beta_i$ are constant in time if we consider the same set of tenors $\tau$ throughout.

[6]Our method thus resembles functional PCA using a discretization method, though we do not perform any interpolation afterwards. As mentioned, we do not interpolate, as the argument $(u, \tau)$ in which this is a functional model is not the "direction" of interest; we are after the state vector $X_t$.

### 2.4.1 "Realizing" the complex-valued log CCF

Through the complex-valued coefficients $\alpha$ and $\beta$, observations on the log CCF are complex-valued even if the state vector is real-valued. As we do not want to lose any information (which would be the case if we were, for instance, to only consider the real part of our factor model), while maintaining the tractable form of a real random vector, we decompose the observed CCF into its real and its imaginary part and stack these. That is, with $\hat{\psi}_{t,i}$ the observations related to a maturity $\tau_i$ as constructed in (2.11), we construct:

$$\tilde{\psi}_{t,i} := \begin{pmatrix} \mathfrak{Re}\left(\hat{\psi}_{t,i}\right) \\ \mathfrak{Im}\left(\hat{\psi}_{t,i}\right) \end{pmatrix} = \begin{pmatrix} \mathfrak{Re}(\alpha_i) \\ \mathfrak{Im}(\alpha_i) \end{pmatrix} + \begin{pmatrix} \mathfrak{Re}(\beta_i) \\ \mathfrak{Im}(\beta_i) \end{pmatrix} X_t + \begin{pmatrix} \mathfrak{Re}\left(\xi_{t,i}^{(1)}\right) \\ \mathfrak{Im}\left(\xi_{t,i}^{(1)}\right) \end{pmatrix} + r_{t,i}, \tag{2.12}$$

or simply

$$\tilde{\psi}_{t,i} = \boldsymbol{a}_i + \boldsymbol{B}_i X_t + \varepsilon_{t,i} + r_{t,i}, \tag{2.13}$$

which is a real-valued $2q$-dimensional factor model driven by the same $d$-dimensional $X_t$. Appendix A discusses the implications of this stacking on the further analysis. Note that we have separated $\xi_{t,i}$ into an error $\xi_{t,i}^{(1)}$, related to the observation error in the option prices, that turns out to be of first order importance, and a remainder term $r_{t,i}$, that can be shown to be asymptotically negligible; see Boswijk et al. (2022) for a more in-depth discussion.

## 2.5 The factor model

Our method straightforwardly allows to consider different maturities, simply by stacking the different maturities. This allows for detection of factors that affect the cross-section of options, as well as factors that affect the term structure of the volatility surface. Our next step indeed is to consider the model in this stacked form:

$$\tilde{\psi}_t := \begin{pmatrix} \tilde{\psi}_{t,1} \\ \vdots \\ \tilde{\psi}_{t,k} \end{pmatrix} = \boldsymbol{a} + \boldsymbol{B} X_t + \varepsilon_t + r_t, \tag{2.14}$$

which is a vector of length $p := 2kq$. By stacking in the time-dimension as well, setting $\tilde{\psi}_t$ as its $t$-th column, we obtain the $(p \times T)$ matrix $\tilde{\Psi}$:

$$\tilde{\Psi} = \boldsymbol{a}\iota_T' + \boldsymbol{B}X + \varepsilon + r, \tag{2.15}$$

where $X = (X_1, \ldots, X_T)$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)$, $r = (r_1, \ldots, r_T)$, and $\iota_T$ is the $T$-dimensional vector of ones. In the analysis that follows, the intercept is not of interest; we are after the behaviour of the state variable $X$. As such, we consider this equation after subtracting the mean over time. To this end, define the projection-matrix $M_{\iota_T} = I_T - T^{-1}\iota_T\iota_T'$, and right-multiply to obtain

$$\ddot{\tilde{\Psi}} = \boldsymbol{B}\ddot{X} + \ddot{\varepsilon} + \ddot{r}, \tag{2.16}$$

where $\ddot{X} := XM_{\iota_T}$, and $\ddot{\varepsilon}$ and $\ddot{r}$ are defined similarly.

In usual factor models, or with panel data in general, there are two dimensions in which it is possible, or necessary, to take the sample size to infinity. Both the number of features in the cross-section (i.e., $p$) can increase, as well as the number of time series observations $T$. However, as we show in the next section, neither is necessary in our factor model: the asymptotics are in terms of the observed number of options $n$.

# 3 Theoretical results

Recall that our interest in the state vector is twofold: we want to learn its unknown dimension $d$, and extract the sample paths of the latent processes in $X_t$. This section provides the required theoretical foundation. In the following, we require our state process to satisfy the following assumptions:

**Assumption 1.**

*(i) The state vector $(y_t, X_t')'$ follows, under the risk-neutral measure $\mathbb{Q}$, an affine jump diffusion, as specified in (2.1) and (2.2);*

*(ii) The $(d \times T)$-dimensional matrix $X$ has rank $d$;*

*(iii) The columns of $\boldsymbol{B}$ are linearly independent.*

Clearly, the first part of this assumption underlies the validity of considering a linear factor model for the log CCF, and is thereby central to our methodology. Note that the price $y_t$ does not feed back into the dynamics of the state vector. As such, returns are distributed independently of price levels. This is intuitive, and satisfied by most standard models. Implicitly, we also assume here that the true model satisfies the admissibility conditions as briefly mentioned in Section 2.1. The second part is straightforward, requiring that all state variables have positive sample variance, and that we observe a time window long enough for us to identify them. The last part states that none of the state variables are redundant, and consequently, not identifiable. This additionally implies that the grid of $(u, \tau)$ is appropriate; some factors might for instance drop out if the set of maturities is not sufficient. Note that we do not assume that the state variables are stationary; this is not necessary as we do not consider large $T$ asymptotics.

Moreover, in order to be able to derive limiting distributions, we make some assumptions on the observation error in our option prices, introduced in (2.8):

**Assumption 2.** *For all $t, \tau$, we have*

*(i) $\mathbb{E}^{\mathbb{P}}[\varkappa_t(\tau, m)|\mathcal{F}^{(0)}] = 0$, $\mathbb{E}^{\mathbb{P}}[\varkappa_t^2(\tau, m)|\mathcal{F}^{(0)}] = 1$, $\mathbb{E}^{\mathbb{P}}[\varkappa_t^4(\tau, m)|\mathcal{F}^{(0)}] < \infty$;*

*(ii) $\varkappa_t(\tau, m)$ are $\mathcal{F}^{(0)}$-conditionally independent along maturity $\tau$, moneyness $m$, and time $t$;*

*(iii) $\sigma_t(\tau, m)$ is $\mathcal{F}_t^{(0)}$-measurable, $0 < \inf_m \sigma_t(\tau, m) \leq \sup_m \sigma_t(\tau, m) < \infty$, $\sigma_t^4(\tau, m)$ is Lipschitz in $m$, and $\int_{\mathbb{R}} e^{-4m}\sigma_t^4(\tau, m)\mathrm{d}m + \int_{\mathbb{R}} |m|e^{-2m}\sigma_t^2(\tau, m)\mathrm{d}m < \infty$.*

This assumption decomposes the observation error in a stochastic component and an $\mathcal{F}_t^{(0)}$-adapted part capturing heteroskedasticity in the observation errors. The error is assumed to be mean zero. The moment condition is required for feasible inference. The second part is commonplace in the literature (see, e.g., Todorov (2019); Boswijk et al. (2022)). Andersen et al. (2021) show that dependence in observation errors is indeed limited, and decreasing in recent years, providing some empirical foundation. The last part is on the conditional variance of the errors, and presents some high-level conditions necessary for feasible inference. These high-level conditions are implied by standard parametrisations, such as a multiplicative pricing error in either the option price or in the implied volatility, combined with some standard moment conditions. These assumptions are all implied by those in Boswijk et al. (2022).

Lastly, we make some assumptions on the existence of certain moments, as well as the structure of the moneyness grid:

**Assumption 3.**

*(i) The true option price $O_t(\tau, m)$ is Lipschitz in $m$;*

*(ii) The forward price process satisfies $\mathbb{E}_t^{\mathbb{Q}} F_{t+\tau}^2 < \infty$ and $\mathbb{E}_t^{\mathbb{Q}} F_{t+\tau}^{-2} < \infty$;*

*(iii) For the log-moneyness grid $\underline{m} = m_1 < \ldots < m_n = \overline{m}$, there exists a deterministic sequence $\Delta m$ depending on $n$ such that $\Delta m \to 0$ as $n \to \infty$ and for some $\iota \in (0, 1]$*

$$\iota \Delta m \leq \inf_{j=2,\ldots,n} |m_j - m_{j-1}| \leq \sup_{j=2,\ldots,n} |m_j - m_{j-1}| \leq \Delta m. \tag{3.1}$$

*In addition, for all $t, \tau$, we have, as $\Delta m \to 0$,*

$$\sup_{j=2,\ldots,n} \left| \frac{m_j - m_{j-1}}{\Delta m} - \delta(m_j) \right| = o_p(\sqrt{\Delta m}), \tag{3.2}$$

*with $\delta : \mathbb{R} \to \mathbb{R}_+$ a continuous function.*

The Lipschitz continuity is satisfied whenever there are no atoms in the risk-neutral return distribution. The moment conditions are necessary to bound the truncation and discretization errors. The assumption on the shrinking moneyness grid forms the basis of the in-fill asymptotics, see also Andersen et al. (2015a) or Todorov (2019). The convergence of the relative step size to the function $\delta$, as in Todorov (2021) or Andersen et al. (2021), is nothing more than a useful vehicle allowing for a more clear and insightful presentation of the limiting results. To this end, we also assume it is the same for different $t, \tau$, though this can be relaxed. The function $\delta$ appears in the limiting results, but does not need to be observed or estimated.

Before we move to our results, we want to draw attention specifically to how the noise enters our factor model. Note that there is no idiosyncratic noise, as in usual factor models. After all, all elements of the factor model are constructed based on the same set of options. Apart from the observable price process, these options form the only input of the model, and they are the sole cause of the errors as in (2.10). As such, it is intuitive that the asymptotic scheme is one in the number of options, and specifically one that ensures the errors vanish in (2.10), and in extension, that the errors disappear from the model asymptotically. Clearly, this requires the CCF spanning to become exact, so the Riemann sum of (2.9) needs to converge to the integral of (2.7). Denoting $\underline{K} := \max_{t,\tau} F_t e^{\underline{m}}$ and $\overline{K} := \min_{t,\tau} F_t e^{\overline{m}}$, this intuition is summarized in the following Lemma:

**Lemma 1.** *Suppose Assumptions 1-3 hold. If* $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ *for some* $\underline{\alpha}, \overline{\alpha} > 0$, *and* $(p \wedge T - 1) > d$, *the first* $d$ *singular values of* $\hat{\ddot{\Psi}}$ *are bounded away from zero, while the* $d+1$*-th singular value tends to zero in probability.*

*Proof.* See Appendix B.3. □

This result combines the consistency of the spanned factor model with the fact that its limiting value is a rank-$d$ matrix. This result only requires an asymptotic scheme on the number of options, and not on the cross-sectional or the time series dimension. Note the resemblance with so-called "small-sigma" asymptotics, in which the errors disappear even though the model dimensions stay constant. We have phrased this result in terms of singular values, as this makes it more intuitive to extend it to selection criteria for the number of factors:

**Corollary 1.** *Suppose the assumptions of Lemma 1 hold. Denote* $M = (p \wedge T)$. *Denote by* $\sigma_1 \geq \ldots \geq \sigma_M$ *the singular values of* $\hat{\ddot{\Psi}}$. *Then,* $\hat{d} := \max\{j : \sigma_j > \gamma_n\}$ *consistently estimates* $d$ *for any deterministic sequence* $\gamma_n$ *such that* $\gamma_n \to 0$ *and* $\frac{\sqrt{n^{-1}\log n}}{\gamma_n} \to 0$.

Not any sequence $\gamma_n$ is allowed: its convergence rate needs to be slower than that of the maximal non-systematic eigenvalue. This rate can be found using Proposition 1 of Boswijk et al. (2022). Some standard options for $\gamma_n$ are, e.g., $n^{-1/3}$ or $(\log n)^{-1}$. This result essentially shows the validity of basing inference on the number of factors on the scree plot. This can be extended to include popular selection criteria, such as eigenvalue differences (Onatski, 2010) or eigenvalue ratios (Ahn and Horenstein, 2013), though the latter might require some form of perturbation to avoid potential division by zero, as pointed out by Pelger (2019). Clearly, other selection criteria can be thought of as well. At their core, all such criteria are transformations of the scree plot, which is commonly used in practice.

Additionally, as the spanned factor model consistently estimates the true $d$-dimensional factor model, the following well-known result follows, stating that PCA extracts the factors up to a rotation:

**Corollary 2.** *Suppose the assumptions of Lemma 1 hold. Denote the* $(d \times T)$*-dimensional matrix formed by the first* $d$ *right-singular vectors of* $\hat{\ddot{\Psi}}$ *by* $\hat{\ddot{X}}$. *Then, for some invertible matrix* $H$, *we have* $\max_t \left\| \hat{\ddot{X}}_t - H'\ddot{X}_t \right\| = o_p(1)$.

This result can straightforwardly be shown using the singular value decomposition of $\hat{\ddot{\Psi}}$. This result is also obtained, in some form, in Bai and Ng (2002) and Bai (2003), among others; consider for instance Proposition 2 in Bai (2003) and recall that our asymptotics are in terms of $n$. Replacing $d$ by a consistent estimator $\hat{d}$ satisfying the conditions of Corollary 1 does not alter this result.

In our atypical framework, we can do better than consistency. By Lemma 1, the rank of our consistent estimator $\hat{\ddot{\Psi}}$ is in the limit equal to the number of factors. As such, a natural approach is to appeal to the literature in rank testing, such as Kleibergen and Paap (2006). This contrasts with the conventional factor modelling literature, where such tests are unavailable as the non-systematic eigenvalues, though small, do not vanish. Clearly, testing does require a distributional result, which we present next.

In order to construct a feasible CLT for $\hat{\ddot{\Psi}}$, we need a consistent estimator for the variance matrix. Following Todorov (2021), we construct estimators of our observation error[7] by setting for $j = 2, \ldots, n-1$:

$$\hat{\zeta}_t(\tau_i, m_j) = \sqrt{\frac{2}{3}} \left( \widehat{O}_t(\tau_i, m_j) - \frac{1}{2} \left( \widehat{O}_t(\tau_i, m_{j-1}) + \widehat{O}_t(\tau_i, m_{j+1}) \right) \right), \tag{3.3}$$

and setting $\hat{\zeta}_t(\tau_i, m_1) = \hat{\zeta}_t(\tau_i, m_2)$ and $\hat{\zeta}_t(\tau_i, m_n) = \hat{\zeta}_t(\tau_i, m_{n-1})$. Lastly, let $j^*$ denote the index corresponding to the smallest absolute log-moneyness, i.e., $j^* = \operatorname{argmin}_j |m_j|$. Then, set

$$\hat{\zeta}_t(\tau_i, m_{j^*}) = \frac{1}{2} \left( |\hat{\zeta}_t(\tau_i, m_{j^*-1})| + |\hat{\zeta}_t(\tau_i, m_{j^*+1})| \right). \tag{3.4}$$

As explained by Todorov (2021), this latter adjustment is to incorporate the no-arbitrage restriction that prices are monotonic.

Define

$$\hat{\xi}_{t,i}(j, u) := -\frac{u^2 + iu}{\hat{\phi}_t(u, \tau_i)} e^{(iu-1)m_{j-1}} \hat{\zeta}_t(\tau_i, m_{j-1}) \Delta m_j. \tag{3.5}$$

Let $\hat{\xi}_{t,i}(j) = (\hat{\xi}_{t,i}(j, u_1), \ldots, \hat{\xi}_{t,i}(j, u_q))'$, and define $\hat{\varepsilon}_{t,i}(j) := \left( \mathfrak{Re}\left(\hat{\xi}_{t,i}(j)\right)' \ \mathfrak{Im}\left(\hat{\xi}_{t,i}(j)\right)' \right)'$. Then, define

$$\hat{H}_{t,i} := \frac{e^{2r\tau_i}}{F_t^2} \sum_{j=2}^{n} \hat{\varepsilon}_{t,i}(j) \hat{\varepsilon}_{t,i}(j)'. \tag{3.6}$$

Lastly, compose the variance estimator as follows:

$$\hat{\mathcal{H}} := (M_{\iota_T} \otimes I_p) \operatorname{blkdiag}\left\{ \left( \hat{H}_{t,i} \right)_{(t,i) \in \{1,\ldots,T\} \times \{1,\ldots,k\}} \right\} (M_{\iota_T} \otimes I_p), \tag{3.7}$$

**Theorem 1.** *Suppose Assumptions 1-3 hold. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some $\underline{\alpha}, \overline{\alpha} > 1/4$, then, with $\mathcal{L} - s$ denoting $\mathcal{F}^{(0)}$-stable convergence in distribution,*

$$(\Delta m)^{-1/2} \left( \operatorname{vec}\left( \hat{\ddot{\Psi}} \right) - \operatorname{vec}\left( \boldsymbol{B} \ddot{X} \right) \right) \xrightarrow{\mathcal{L}-s} \mathcal{N}(0, \mathcal{H}), \tag{3.8}$$

[7] Alternative estimators of $\hat{\zeta}_t(\tau_i, m_j)$ can be used as well. Sufficient conditions for the ensuing variance matrix to be consistent are asymptotic unbiasedness for $\sigma_t^2(\tau, m_j)$, except for maybe a finite number of indices $j$, and $\sum_{l=1}^{n} \mathbb{C}\mathrm{ov}_t^{\mathbb{P}} \left( \hat{\zeta}_t^2(\tau, m_j), \hat{\zeta}_t^2(\tau, m_l) \right) = O_p(1)$. We elect to employ iCOS-implied errors based on the techniques of Vladimirov (2023), which outperform comparable methods in the simulations with reasonable sample sizes. Those based on Todorov (2021) are presented in the text as they are a well-performing alternative, and are much simpler to explain.

*where $\mathcal{H}$ is some $\mathcal{F}^{(0)}$-measurable random matrix, and, additionally,*

$$\|(\Delta m)^{-1}\hat{\mathcal{H}} - \mathcal{H}\| = o_p(1). \tag{3.9}$$

*Proof.* See Appendix B.4. □

As the covariance matrix in the limiting distribution is random, our factor model is $\mathcal{F}^{(0)}$-mixing normally distributed. Our feasible variance estimator does allow for hypothesis testing. Though the mesh $\Delta m$ explicitly enters these expressions, this term does not appear in a standardized version which would be used for such tests. Note that the limiting variance $\mathcal{H}$ is singular, due to the demeaning, but potentially also due to the choice of $u$. This requires an adjustment to the rank test of Kleibergen and Paap (2006), which is presented in the next subsection.

A decomposed version can be found using the results of Bai and Ng (2020), and can for instance be used to construct confidence bounds around the extracted factors. A feasible variance matrix estimator is presented in Bai (2003), though we can base an estimator on the previously used $\hat{\mathcal{H}}$ as well.

**Proposition 1.** *Suppose the assumptions of Theorem 1 hold. Assume the nonzero singular values of $\ddot{\Psi}$ are distinct. Recall that $\widehat{\ddot{X}}$ are the first $d$ right-singular vectors of $\widehat{\ddot{\Psi}}$, and define $\widehat{B} := \widehat{\ddot{\Psi}}\widehat{\ddot{X}}'$. Then, for some invertible matrix $H \xrightarrow{p} Q^{-1}$,*

$$(i) \quad (\Delta m)^{-1/2}\left(\widehat{B}_j - H^{-1}B_j\right) \xrightarrow{\mathcal{L}-s} \mathcal{N}\left(0, Q'^{-1}\ddot{X}(I_T \otimes e_j^{(p)})'\mathcal{H}(I_T \otimes e_j^{(p)})\ddot{X}'Q^{-1}\right),$$

$$(ii) \quad (\Delta m)^{-1/2}\left(\widehat{\ddot{X}}_t - H'\ddot{X}_t\right) \xrightarrow{\mathcal{L}-s} \mathcal{N}\left(0, S^{-2}QB'(e_t^{(T)} \otimes I_p)'\mathcal{H}(e_t^{(T)} \otimes I_p)BQ'S^{-2}\right),$$

*where $e_i^{(j)}$ is the $j$-dimensional vector with a 1 at the $i$-th position and zeroes elsewhere, and $S$ is the diagonal matrix containing the $d$ nonzero singular values of $\ddot{\Psi}$ in descending order.*

*Proof.* See Appendix B.5 □

## 3.1 Testing for the number of factors

As briefly mentioned, the asymptotic rank of our factor model is equal to the number of factors, or the dimension of the state vector $X_t$. As such, we implement a rank test in order to perform inference on this quantity. We propose to use a restricted version of the Kleibergen and Paap (2006) rank test.

In our factor model, the standard test of Kleibergen and Paap (2006) jointly tests $(p-d)(T-1-d)$ restrictions on the singular values. Given that we look at a single equity, it is reasonable to assume that a proportion of these are zero without a need to test for them; i.e., we assume an upper bound for the number of factors, say $d_{\max}$. This implies our alternative hypothesis changes from, e.g., $H_1 : d > 2$ to $H_1 : 2 < d \leq d_{\max}$, reducing the power against the set $\{d : d > d_{\max}\}$. The main benefits of this added assumption are that the degrees of freedom in the test is much reduced, increasing the power of the test, and that the variance matrix that needs to be inverted is of a smaller dimension and is more likely to be non-singular, which benefits the size of the test.

Next, we briefly discuss the construction of the test, in relation to the Kleibergen and Paap (2006) test. However, we first have to deal with the singularity that arises when demeaning the factor model. As a solution, we elect to simply remove the last of its columns. That is, we post-multiply our factor model by the $(T \times (T-1))$-dimensional deletion-matrix $D_T := (I_{T-1} \vdots \mathbf{0}_{T-1})'$. Our proposed test is a restricted version of the rank test of Kleibergen and Paap (2006), superimposing that some of the smallest singular values are exactly zero. Write the SVD decomposition of our factor model as follows:

$$Y := \boldsymbol{B}\ddot{X}D_T = \begin{pmatrix} U_1 & U_2 & U_3 \end{pmatrix} \begin{pmatrix} S_1 & 0 & 0 \\ 0 & S_2 & 0 \\ 0 & 0 & \mathbf{0} \end{pmatrix} \begin{pmatrix} V_1' \\ V_2' \\ V_3' \end{pmatrix}.$$

In this decomposition, $S_1$ is a $d \times d$-dimensional matrix, $S_2$ is $(d_{\max}-d)\times(d_{\max}-d)$-dimensional, and $\mathbf{0}$ is a $(p-d_{\max})\times(T-1-d_{\max})$-dimensional matrix of zeroes. Superimposing that this lower-right block is identically zero implies that there is no need to test for it. With $\tilde{U}' = \begin{pmatrix} U_1 & U_2 \end{pmatrix}$ the $p \times d_{\max}$-dimensional submatrix containing the first $d_{\max}$ left singular values of the factor model, we consider instead:

$$\begin{aligned} \tilde{Y} := \tilde{U}Y = \begin{pmatrix} U_1' \\ U_2' \end{pmatrix} Y &= \begin{pmatrix} I_d & 0 & 0 \\ 0 & I_{d_{\max}-d} & 0 \end{pmatrix} \begin{pmatrix} S_1 & 0 & 0 \\ 0 & S_2 & 0 \\ 0 & 0 & \mathbf{0} \end{pmatrix} \begin{pmatrix} V_1' \\ V_2' \\ V_3' \end{pmatrix} \\ &= \begin{pmatrix} I_d & 0 \\ 0 & I_{d_{\max}-d} \end{pmatrix} \begin{pmatrix} S_1 & 0 & 0 \\ 0 & S_2 & 0 \end{pmatrix} \begin{pmatrix} V_{11}' & V_{21}' & V_{31}' \\ V_{12}' & V_{22}' & V_{32}' \\ V_{13}' & V_{23}' & V_{33}' \end{pmatrix}. \end{aligned}$$

This is simply a projection onto the lower-dimensional subspace spanned only by the first $d_{\max}$ left-singular vectors. We propose to perform the Kleibergen and Paap (2006) test on this reduced matrix instead. Note that we reduce the row dimension, not the column dimension. We do so as any singularity problem would likely be caused by the fact that the loadings can be very comparable for $(\tau, u)$ close, which forms the row dimension.

Following Kleibergen and Paap (2006), we write

$$\tilde{Y} = A_d C_d + A_{d,\perp}\Lambda_d C_{d,\perp}.$$

Applying the results of Kleibergen and Paap (2006) to our restricted setting, with

$$\tilde{V}' = \begin{pmatrix} V_{22}' & V_{32}' \\ V_{23}' & V_{33}' \end{pmatrix},$$

a $(T-1-d) \times (T-1-d)$-dimensional matrix, we have the following expressions:

$$\Lambda_d = S_2 \tilde{V}'(\tilde{V}\tilde{V}')^{-1/2},$$
$$A_{d,\perp} = \begin{pmatrix} 0 \\ I_{d_{\max}-d} \end{pmatrix},$$
$$C_{d,\perp} = (\tilde{V}\tilde{V}')^{1/2}(\tilde{V}')^{-1} \begin{pmatrix} V_{12}' & V_{22}' & V_{32}' \\ V_{13}' & V_{23}' & V_{33}' \end{pmatrix},$$

which are $(d_{\max}-d) \times (T-1-d)$, $d_{\max} \times (d_{\max}-d)$, and $(T-1-d) \times (T-1)-$dimensional matrices, respectively. We denote sample counterparts by a hat, e.g., $\hat{\Lambda}_d$.

**Assumption 4.** *The $(d_{\max}-d)(T-1-d) \times (d_{\max}-d)(T-1-d)$ covariance matrix*

$$\Omega_d = (C_{d,\perp} \otimes A_{d,\perp}')(D_T \otimes \tilde{U})\mathcal{H}(D_T' \otimes \tilde{U}')(C_{d,\perp}' \otimes A_{d,\perp}) \tag{3.10}$$

*is non-singular.*

It can straightforwardly be shown that a sufficient, though much stronger, condition is assuming that $H_{t,i}$ is full rank for all $(t,i)$. Though we by no means prefer to make this stronger assumption, it does provide the intuition that taking appropriate $u$ is highly important.

**Corollary 3.** *Suppose Assumptions 1-4 hold. Denote $\hat{\lambda}_d = \text{vec}\left(\hat{\Lambda}_d\right)$. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some $\underline{\alpha}, \overline{\alpha} > 1/4$, $p \geq d_{\max}$, and $T-1 > d$, then*

$$\hat{\lambda}_d' \hat{\Omega}_d^{-1} \hat{\lambda}_d \xrightarrow{\mathcal{L}} \chi^2\left((d_{\max}-d)(T-1-d)\right), \tag{3.11}$$

*where $\hat{\Omega}_d$ is the feasible counterpart of the variance matrix defined in Assumption 4*

*Proof.* This follows from Theorem 1 by the results of Kleibergen and Paap (2006). □

This is a test whether the elements of $\Lambda_d$ are zero, which is equivalent to the elements of $\Sigma_2$ being zero. This is true if the rank of the factor model is $d$. If the true rank is greater than $d$, this test-statistic diverges, such that a test based on this statistic has asymptotic power equal to one. Without Assumption 4, a similar, but less powerful, test can be based on Robin and Smith (2000). We propose a sequential test: start with $i=0$, and test $\mathcal{H}_0^{(i)} : d_0 = i$ against $\mathcal{H}_a^{(i)} : i < d_0 \leq d_{\max}$; if $\mathcal{H}_0^{(i)} : d_0 = i$ is rejected, set $i \mapsto i+1$, continuing until the null can no longer be rejected or $i = d_{\max}$; if the null is not rejected, conclude that $d = i$. This approach can be justified by appealing to Goeman and Solari (2010), as it clearly satisfies their monotonicity condition and the first, and only, true hypothesis being tested has asymptotically correct size, such that it also satisfies their familywise error control. The intuition behind the size control in this procedure is simple: as the test is consistent, the sequential testing asymptotically rejects almost surely until we reach $i = d$, at which point the test is size correct (cf. Johansen (1988)'s cointegration test). Of course, this presumes that for the true value $d$ we have $d \leq d_{\max}$.

### 3.2 Analysis in non-affine models

This section shortly discusses the possibilities for inference in non-affine models. Clearly, the assumption of the affine jump diffusion as the driver of price dynamics is key in the above analysis. Relaxing this assumption reduces the earlier derived techniques to approximations in the best case, and renders them invalid in the worst. The rank test, for instance, will overselect the number of factors, as the true model is no longer reduced rank. PCA however still delivers the best linear approximation to the dynamics in the CCF. As affine models are by far the most popular class in the literature, it might be interesting to address the question of how well such a linear representation can fit the data. To measure this goodness of fit, we investigate the correlations between the principal components and higher order risk-neutral moments. As the distribution of the log returns is essentially what is needed to accurately price options, we argue that the linear approximation to the CCF, by proxy of the PCs, needs to be well correlated with these objects.

Bakshi et al. (2003) use the Carr and Madan (2001) spanning result displayed in (2.6) to obtain a portfolio representation of the second, third, and fourth risk-neutral moment of log returns. In general, the $h$-th risk-neutral moment can be spanned as:

$$R_t^{\mathbb{Q}}(h,\tau) := \mathbb{E}_t^{\mathbb{Q}}\left[(\log(F_{t+\tau}/F_t))^h\right] = \frac{e^{r\tau}}{F_t}\int_{\mathbb{R}} h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)\mathrm{d}m.$$

We denote the feasible, Riemann sum-based counterpart by $\widehat{R}_{\mathbb{Q}}(h,\tau)$. Asymptotic normality of these spanned moments follows along the same lines as in Theorem 1:

**Proposition 2.** *Suppose Assumptions 2 and 3 hold. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some $\underline{\alpha}, \overline{\alpha} > 1/4$, then, with $\mathcal{L}-s$ denoting $\mathcal{F}^{(0)}$-stable convergence in distribution,*

$$(\Delta m)^{-1/2}\left(\widehat{R}_t^{\mathbb{Q}}(h,\tau) - R_t^{\mathbb{Q}}(h,\tau)\right) \xrightarrow{\mathcal{L}-s} \mathcal{N}(0,\mathcal{V}_t(h)), \tag{3.12}$$

*with*

$$\mathcal{V}_t(h) := \int_{\mathbb{R}} h^2(h-1-m)^2 m^{2h-4}e^{-2m}\sigma_t^2(\tau,m)\delta(m)\mathrm{d}m,$$

*for* $h \in \left\{k \in \mathbb{R} : |R_t^{\mathbb{Q}}(k,\tau)| < \infty, \exists\delta > 0 : \int_{\mathbb{R}}\left(|k-1-m|\,|m|^{k-2}e^{-m}\sigma_t(\tau,m)\right)^{2+\delta}\mathrm{d}m < \infty]\right\}$.

*Proof.* See Appendix B.6. □

A feasible version of the (scaled) variance can be found by replacing the integral by a Riemann sum, $\delta(m)$ by $\Delta m_j$, and the unknown $\{\sigma_t^2(\tau,m)\}$ by the estimators $\{\hat{\zeta}_t^2(\tau,m)\}$. This can easily be extended to the time-stacked $(T\times 1)$-vector $\widehat{R}_{h,\tau}^{\mathbb{Q}} := \left(\widehat{R}_1^{\mathbb{Q}}(h,\tau),\ldots,\widehat{R}_T^{\mathbb{Q}}(h,\tau)\right)'$ by making use of the time-independence of Assumption 2.

Canonical correlations and regression $R^2$s can be derived as eigenvalue problems. In particular, if we are interested in the canonical correlations of our extracted principal components $\hat{\ddot{X}}$ with a different set of mean-zero factors $G$, possibly observed with error, we can study the eigenvalues of the matrix $(\hat{\ddot{X}}'\hat{\ddot{X}})^{-1}\hat{\ddot{X}}'G(G'G)^{-1}G'\hat{\ddot{X}}$. Distributional results for the (sums of) eigenvalues of this matrix are established in Gaussian or other elliptical settings (see the discussion in Bai and Ng, 2006), in group factor models (Andreou et al., 2019), and in high-frequency settings (Pelger, 2019). Unfortunately, the assumptions in these papers do not align with our framework. Nonetheless, the general idea behind the proof of Pelger (2019) still works in our case, providing us with the following result:

**Proposition 3.** Suppose the conditions of Propositions 1 and 2 hold, and $\tau \notin \{\tau_i\}_{i=1}^k$. Then, with $\hat{\mathcal{R}}^2$ the R-squared of the regression of $\hat{R}_{h,\tau}^{\mathbb{Q}}$ onto $\hat{\ddot{X}}$ and a constant, and $\mathcal{R}^2 < 1$ its infeasible counterpart,

$$(\Delta m)^{-1/2}(\hat{\mathcal{R}}^2 - \mathcal{R}^2) \xrightarrow{\mathcal{L}-s} \mathcal{N}(0, \xi' D \Pi D' \xi), \tag{3.13}$$

where, with $\ddot{R}_{h,\tau}^{\mathbb{Q}} = M_{\iota_T} R_{h,\tau}^{\mathbb{Q}}$

$$\Pi = \begin{bmatrix} K^{(d,T)}(I_T \otimes S^{-2}QB')\mathcal{H}(I_T \otimes S^{-2}QB')'K^{(T,d)} & 0 \\ 0 & M_{\iota_T}\,\text{diag}\{\{\mathcal{V}_t(h)\}_{t=1,\dots,T}\}\,M_{\iota_T}' \end{bmatrix},$$
$$\xi = \text{vec}\left(\begin{bmatrix} -\left((\ddot{X}\ddot{X}')^{-1}\ddot{X}\ddot{R}_{h,\tau}^{\mathbb{Q}}(\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{R}_{h,\tau}^{\mathbb{Q}})^{-1}\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{X}'(\ddot{X}\ddot{X}')^{-1}\right)' & (\ddot{X}\ddot{X}')^{-1}\ddot{X}\ddot{R}_{h,\tau}^{\mathbb{Q}}(\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{R}_{h,\tau}^{\mathbb{Q}})^{-1} \\ (\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{R}_{h,\tau}^{\mathbb{Q}})^{-1}\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{X}'(\ddot{X}\ddot{X}')^{-1} & -\left((\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{R}_{h,\tau}^{\mathbb{Q}})^{-1}\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{X}'(\ddot{X}\ddot{X}')^{-1}\ddot{X}\ddot{R}_{h,\tau}^{\mathbb{Q}}(\ddot{R}_{h,\tau}^{\mathbb{Q}\prime}\ddot{R}_{h,\tau}^{\mathbb{Q}})^{-1}\right)' \end{bmatrix}\right),$$
$$D = \left(\begin{bmatrix} H'\ddot{X} \\ \ddot{R}_{h,\tau}^{\mathbb{Q}\prime} \end{bmatrix} \otimes I_{d+1}\right) K^{(T,d+1)} + \left(I_{d+1} \otimes \begin{bmatrix} H'\ddot{X} \\ \ddot{R}_{h,\tau}^{\mathbb{Q}\prime} \end{bmatrix}\right), \tag{3.14}$$

with $K$ the commutation matrix such that $K\,\text{vec}(A) = \text{vec}(A')$

*Proof.* See Appendix B.7. □

This result can directly be extended to the sum of the canonical correlations between $\hat{\ddot{X}}$ and a general set of factors $G$; the proof is completely identical, so long as the errors in $\hat{\ddot{X}}$ are independent[8] of those in $G$. To bring this into practice, we need consistent estimators of the components in the variance matrix. For the matrices $\xi$ and $D$, we can construct simple feasible analogues by plugging in $\hat{\ddot{X}}$ (note that $\hat{\ddot{X}}\hat{\ddot{X}}' = I$) and $\hat{\ddot{R}}_{h,\tau}^{\mathbb{Q}}$, which leads to consistent estimators by the continuous mapping theorem. As mentioned in the discussion after Proposition 2, a feasible estimator of $\mathcal{V}_t(h)$ can be constructed in a similar manner to how we estimate $\mathcal{H}$: a discretization of the integral and the replacement of $\{\sigma_t^2(\tau, m)\}$ by the estimators $\{\hat{\zeta}_t^2(\tau, m)\}$. Such an estimator is consistent under suitable integrability conditions. What remains is an estimator for the quantity $S^{-2}QB'$; based on Proposition 1, this role can be taken on by $\hat{S}^{-2}\hat{B}'$, with $\hat{S}$ the diagonal matrix with the singular values of $\hat{\ddot{\Psi}}$.

In contrast to the test of Andreou et al. (2019), this test is not applicable in the boundary case of $\mathcal{R}^2 = 1$ (or unit canonical correlations in a multivariate extension). This rules out the case where the PCs are an exact rotation of the other set of factors. As Pelger (2019) explains, this corner case leads to atypical behaviour, similar to unit-root tests.

[8]This can be relaxed, but a joint CLT such as in Lemma 8 needs to be satisfied.

# 4 Simulation study

In this section, we study the finite-sample performance of our factor modeling framework. For that, we consider three different AJD specifications: a one-factor, two-factor, and three-factor option pricing models.

## 4.1 One-factor model

As a starting point, we consider a one-factor option pricing model of Duffie et al. (2000) with a Gaussian jump size distribution in returns and co-jumps in volatility. The likelihood of jumps is additionally made stochastic and proportional to the stochastic variance. In particular, we assume the following process, referred to in shorthand as 'SVCJ', for the log forward price under both the $\mathbb{P}$ and $\mathbb{Q}$ probability measures:

$$\mathrm{d}\log F_t = (-\tfrac{1}{2}v_t - \mu\lambda_t)\,\mathrm{d}t + \sqrt{v_t}\mathrm{d}W_{1,t} + J_t\mathrm{d}N_t, \tag{4.1}$$

$$\mathrm{d}v_t = \kappa(\bar{v} - v_t)\,\mathrm{d}t + \sigma\sqrt{v_t}\mathrm{d}W_{2,t} + J_t^v\mathrm{d}N_t, \tag{4.2}$$

where two Brownian motions $W_{1,t}$ and $W_{2,t}$ are assumed to be correlated with the coefficient $\rho$; $N_t$ is a Poisson jump process with intensity $\lambda_t = \delta v_t$; $J^v \sim \exp(1/\mu_v)$ are jump sizes in volatility and $J \sim \mathcal{N}(\mu_j, \sigma_j^2)$ are jump sizes in prices. Given the jump size distributions, the expected relative jump size in returns $\mu = \exp\Big(\mu_j + \frac{1}{2}\sigma_j^2\Big) - 1$. In the simulation, the initial underlying price is set to $F_0 = 3000$, but the initial variance value $v_0$ is drawn for each iteration from the Gamma distribution with shape and scale parameters of 1.25 and 0.016. The parameters of the model are set as follows:

$$\kappa = 5.0,\ \bar{v} = 0.02,\ \rho = -0.95,\ \sigma = 0.4,\ \delta = 60,\ \mu_j = -0.08,\ \sigma_j = 0.04,\ \mu_v = 0.02.$$

Although the model in (4.1)-(4.2) is a one-factor AJD model, it exhibits all main features of option pricing models: stochastic volatility, (co-)jumps in returns and volatility, time-varying stochastic jump intensity and self-excitation feature. Furthermore, it embeds many popular one-factor option pricing models such as Heston (1993), Pan (2002), and Bates (1996).

The option price simulation roughly mimics the S&P 500 index option data. In particular, we fix the strike prices on an equidistant grid in increments of 5 and keep simulated option prices that exceed 0.1, which corresponds to a minimum ask price. We consider three fixed maturities of 10, 30, and 60 days, i.e., $k = 3$. The true option prices are simulated using the COS method of Fang and Oosterlee (2009) and then are distorted with the observation errors, which reflect Assumption 2:

$$\widehat{O}_t(\tau, m) = O_t(\tau, m) + \zeta_t(\tau, m),\ \text{ with } \zeta_t(\tau, m) = 0.001 \cdot \kappa_t(\tau, m)\nu_t(\tau, m) \cdot \frac{\epsilon}{\sqrt{\tau}}$$

where $\epsilon$ is an i.i.d. standard random normal variable, $\kappa_t(\tau, m)$ is the Black-Scholes implied volatility and $\nu_t(\tau, m)$ the corresponding vega. The division by $\sqrt{\tau}$ makes errors comparable along the maturity dimension.

Given the log futures price process and the observed noisy option prices, we span the log CCF following (2.9) at each time $t$, for each maturity $\tau$. To reduce the impact of truncation and discretization errors in finite samples, we additionally simulate option prices by increasing and decreasing strike prices up until the OTM call and put option values reach $10^{-8}$ and interpolate all option prices using a cubic spline on the implied volatility domain. Importantly, the interpolated and extrapolated option contracts are used only to estimate the CCF, i.e., these additional contracts are not used to estimate the covariance matrix. As discussed in Boswijk et al. (2022), the discretization and truncation errors are $\mathcal{F}_t$-measurable and, thus, affect only the mean, but not the variance of the CCF approximation. Furthermore, these errors (and, hence, biases they induce) are asymptotically negligible, but they affect the estimation in finite samples. Therefore, this interpolation-extrapolation scheme can be seen as a bias correction of the CCF estimators without an effect on the CLT.

In the simulation, we experiment with different time dimensions and different sets of argument values for the CCF to explore the limits of the novel test procedure. In particular, we consider $T = \{5, 10, 50, 100\}$ with the smaller time series being always the beginning of the larger one.

To experiment with different sets of CCF arguments, we consider the arguments of the form:

$$u_j = \frac{jL}{\sqrt{\tau}}, \; j = 1, \ldots, q,$$

where $q \in \mathbb{N}$ is the number of the CCF arguments in consideration and $L > 0$ is some constant that defines the grid of the argument set. The division by $\sqrt{\tau}$ again makes different maturities more comparable. Similar scaling is used by Todorov (2019). In simulations, we vary $L$ and $q$.

To test the dimension of the (latent) state vector we use the (adjustments of) rank test of Kleibergen and Paap (2006) as detailed in Section 4.1. For the SVCJ model $d = 1$, but in the simulations, we test the null hypotheses $d = 0$ and $d = 1$ against the alternatives $d \geq 1$ and $d \geq 2$, respectively, to analyze the power and the size of the test. We estimate the covariance matrix $\Omega_d$ using the observation-based feasible estimators $\widehat{\zeta}_t(\tau, m)$, which we obtain using the iCOS method following Vladimirov (2023). In the following tables, we display the rejection frequency of the corresponding null hypothesis. The number of Monte Carlo simulations is $N = 1000$.

Table 1: Monte Carlo results for the rank test size under one-factor model

| | $T=50$ | | | | | | $T=100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L=0.2$ | | | $L=1.0$ | | | $L=0.2$ | | | $L=1.0$ | | |
| $q\backslash\alpha$ | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 |
| 1.0 | 7.7 | 4.2 | 0.6 | 7.8 | 4.6 | 0.6 | 5.7 | 2.3 | 0.6 | 6.1 | 2.7 | 0.7 |
| 2.0 | 4.1 | 1.4 | 0.2 | 4.7 | 1.8 | 0.2 | 1.3 | 0.6 | 0.1 | 1.7 | 0.8 | 0.1 |
| 3.0 | 50.1 | 50.1 | 50.1 | 2.0 | 0.8 | 0.1 | 49.6 | 49.6 | 49.6 | 0.7 | 0.3 | 0.0 |
| 4.0 | 49.2 | 49.2 | 49.2 | 19.3 | 17.7 | 15.9 | 49.5 | 49.5 | 49.5 | 26.7 | 25.5 | 23.4 |

Note: This table provides Monte Carlo simulation results for the rejection frequencies of the rank test at 10%, 5% and 1% significance levels, based on 1000 replications from the SVCJ model. Each panel lists, the rejection frequencies for different null hypotheses, numbers of argument values, and grid sizes.

We start with the standard rank test of Kleibergen and Paap (2006) that jointly tests $(p-d)(T-1-d)$ restrictions on the singular values. Table 1 provides the Monte Carlo results for the one-factor SVCJ model for different values of $L$ and $q$. The results are indicative of problems with the standard rank test applied to our setting: the cross-sectional dependence between the log CCFs with different arguments leads to singularity issues in the constructed covariance matrix $\mathcal{H}$, which, in turn, distort the power and the size of the test. In fact, the size typically increases with the decrease of $L$, i.e., when the grid of arguments shrinks, and with the increase of the number of arguments $q$. The power and the size of the test for $q = 1$, on the other hand, are very close to their nominal level. This indicates that there is a clear separation between the first and the second eigenvalues that helps to identify the number of latent factors.

As mentioned in Section 2, we can set an upper bound for the number of factors $d_{max}$, which allows us to reduce the set of alternatives, e.g., to $H_1 : 2 \leq d \leq d_{max}$ when testing $H_0 : d = 1$. This in turn helps to reduce the dimension of the covariance matrix $\mathcal{H}$, limiting its singularity issues. In the simulations, we set $d_{max} = 6$ since this is the dimension of the covariance matrix $H_{t,i}$ for a fixed maturity with a single CCF argument. The number of factors is also unlikely to exceed $d_{max} = 6$ for option contracts on a single underlying. In fact, most of the parametric AJD option pricing models do not exceed 3 latent factors. Andersen et al. (2015b), for instance, formally show that their three-factor model captures most of the variation in option data[9].

Table 2: Monte Carlo results for the restricted rank test size under one-factor model

(a) $d_{max} = 6$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q\backslash\alpha$ | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 |
| 1.0 | 7.7 | 4.2 | 0.6 | 7.8 | 4.6 | 0.6 | 5.7 | 2.3 | 0.6 | 6.1 | 2.7 | 0.7 |
| 2.0 | 7.7 | 4.2 | 0.6 | 9.4 | 5.1 | 0.8 | 5.8 | 2.3 | 0.6 | 7.4 | 3.4 | 0.9 |
| 3.0 | 7.7 | 4.2 | 0.6 | 11.0 | 6.2 | 1.0 | 5.8 | 2.3 | 0.6 | 8.6 | 3.6 | 1.0 |
| 4.0 | 7.7 | 4.3 | 0.6 | 13.7 | 6.8 | 1.7 | 5.9 | 2.5 | 0.6 | 10.2 | 5.3 | 1.5 |

(b) $d_{max} = 8$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q\backslash\alpha$ | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 |
| 1.0 | 8.6 | 4.8 | 0.9 | 9.6 | 4.9 | 1.2 | 5.8 | 3.0 | 0.6 | 6.5 | 4.1 | 0.7 |
| 2.0 | 8.6 | 4.8 | 0.9 | 9.6 | 4.9 | 1.2 | 5.8 | 3.0 | 0.6 | 6.5 | 4.1 | 0.7 |
| 3.0 | 8.4 | 4.6 | 0.9 | 11.8 | 6.9 | 2.0 | 5.7 | 3.1 | 0.6 | 9.9 | 5.8 | 1.8 |
| 4.0 | 8.5 | 4.4 | 0.8 | 15.2 | 8.7 | 3.3 | 5.5 | 3.1 | 0.7 | 13.3 | 7.7 | 2.0 |

Note: This table provides Monte Carlo simulation results for the rejection frequencies of the adjusted rank test at 10%, 5% and 1% significance levels, based on 1000 replications from the SVCJ model. Each panel lists, the rejection frequencies for different null hypotheses, numbers of argument values, and grid sizes.

Table 2 provides the Monte Carlo results for the adjusted rank test with $d_{max} = 6$. We restrict our attention to the cases $T = 50$ and $T = 100$, since the dimension of the covariance matrix $\mathcal{H}$ cannot be further reduced for small $T$. With the adjusted rank test, the empirical rejection frequencies are very close to the chosen significance level for almost all cases of $L$ and $q$. The

[9] In practice, one could select $d_{max}$ by applying standard tests/information criteria for the number of factors on the implied volatility domain. By doing so, Andersen et al. (2015b) found seven to eight factors, which, as they mention, “likely reflects the failure of the linear approximation rather than the true number of underlying (linear) factors”.

size of the test is the largest for $L = 2$ with $q = 5$. This is likely due to the increase of the approximation errors in the CCF for large argument values.

## 4.2 Two-factor model

Next, we extend the SVCJ model to a two-factor option pricing model by adding the second stochastic volatility component. In particular, we consider the following model dynamics:

$$\begin{aligned}
\mathrm{d}\log F_t &= (-\tfrac{1}{2}v_{1,t} - \tfrac{1}{2}v_{2,t} - \mu\lambda_t)\,\mathrm{d}t + \sqrt{v_{1,t}}\mathrm{d}W_{1,t} + \sqrt{v_{2,t}}\mathrm{d}W_{3,t} + J_t \mathrm{d}N_t,\\
\mathrm{d}v_{1,t} &= \kappa_1(\bar{v}_1 - v_{1,t})\,\mathrm{d}t + \sigma_1\sqrt{v_{1,t}}\mathrm{d}W_{2,t} + J_t^v \mathrm{d}N_t,\\
\mathrm{d}v_{2,t} &= \kappa_2(\bar{v}_2 - v_{2,t})\,\mathrm{d}t + \sigma_2\sqrt{v_{2,t}}\mathrm{d}W_{4,t},
\end{aligned}$$

where $v_{1,t}$ is the same stochastic volatility as in the one-factor specification and $v_{2,t}$ is the second stochastic volatility component without jumps. Two additional Brownian motions $W_{3,t}$ and $W_{4,t}$ are assumed to be correlated with the coefficient $\rho_2$ but are independent of $W_{1,t}$ and $W_{2,t}$. The distributions of jump sizes are the same as in the one-factor specification, and the jump intensity is a linear function of two volatility factors, i.e., $\lambda_t = \delta_1 v_{1,t} + \delta_2 v_{2,t}$. The parameters of the model are set as follows:

$$\begin{aligned}
&\kappa_1 = 15,\ \bar{v}_1 = 0.02,\ \rho_1 = -0.95,\ \sigma_1 = 0.4,\ \delta_1 = 60,\ \mu_v = 0.02,\\
&\kappa_2 = 1,\ \bar{v}_2 = 0.01,\ \rho_2 = -0.5,\ \sigma_2 = 0.1,\ \delta_2 = 30,\ \mu_j = -0.08,\ \sigma_j = 0.04.
\end{aligned}$$

This two-factor model allows for short-term and long-term stochastic volatility components and is related to popular two-factor models considered in Bates (2000), Christoffersen et al. (2009) and Andersen et al. (2015b). Due to the co-jumps in the short-term volatility $v_{1,t}$ and the specification of the jump intensity $\lambda_t$, the first stochastic volatility also absorbs some self-excitation features with fast mean-reversion.

The simulation scheme of option prices and estimation of the CCF is exactly the same as for the one-factor model. The main difference is, of course, that $d = 2$ for the two-factor model, and in the simulation, we test the null hypotheses $d = 1$ and $d = 2$. Motivated by the simulation results for the one-factor model, here we provide the results of the adjusted rank test with $d_{max} = 6$ and two time periods of $T = 50$ and $T = 100$. The simulation results are provided in Table 3. We notice a very good power of the test in all settings and a very good size for $T = 50$. The size of the rank test for $T = 100$ is slightly larger than the nominal level.

Table 3: Monte Carlo results for the restricted rank test size under two-factor model

(a) $d_{max} = 6$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q \backslash \alpha$ | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 |
| 1.0 | 9.0 | 4.5 | 0.9 | 9.3 | 4.7 | 0.9 | 9.1 | 2.7 | 0.8 | 9.8 | 3.6 | 0.8 |
| 2.0 | 9.1 | 4.5 | 0.9 | 10.8 | 6.0 | 1.4 | 9.1 | 2.8 | 0.8 | 12.0 | 4.9 | 1.0 |
| 3.0 | 9.2 | 4.5 | 0.9 | 13.0 | 6.9 | 1.9 | 9.2 | 3.0 | 0.8 | 13.1 | 7.0 | 1.2 |
| 4.0 | 9.3 | 4.7 | 0.9 | 14.5 | 8.1 | 2.3 | 9.5 | 3.3 | 0.8 | 14.6 | 8.6 | 2.3 |

(b) $d_{max} = 8$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q \backslash \alpha$ | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 | 10 | 5 | 1 |
| 1.0 | 8.5 | 3.8 | 0.8 | 9.2 | 4.3 | 1.1 | 7.3 | 3.3 | 0.6 | 9.9 | 4.3 | 0.9 |
| 2.0 | 8.5 | 3.8 | 0.8 | 9.2 | 4.3 | 1.1 | 7.3 | 3.3 | 0.6 | 9.9 | 4.3 | 0.9 |
| 3.0 | 8.0 | 3.2 | 0.8 | 14.2 | 6.5 | 1.4 | 7.3 | 3.4 | 0.5 | 14.6 | 7.7 | 1.8 |
| 4.0 | 7.7 | 3.3 | 0.7 | 20.6 | 11.2 | 2.9 | 7.2 | 3.4 | 0.4 | 22.3 | 12.2 | 3.6 |

Note: This table provides Monte Carlo simulation results for the rejection frequencies of the adjusted rank test at 10%, 5% and 1% significance levels, based on 1000 replications from the two-factor model. Each panel lists, the rejection frequencies for different null hypotheses, numbers of argument values, and grid sizes.

# 5 Empirical application

This section contains a sketch of the empirical application that we are currently working on.

## 5.1 Data

In this paper, we consider 'weekly' options on the S&P 500 stock market index, obtained through OptionMetrics. In particular, we focus on 'weeklies' traded on Wednesdays and expiring in one, two, three, and four weeks. As emphasized in Andersen et al. (2017), weekly options have experienced a rapidly increased trading volume throughout the last decade and now represent the most actively traded segment of the option market. For now, we consider the option data spanning the period of 2019, resulting in $T = 47$ time series observations, although our methodology can be extended to larger time periods.

For each of the specified maturities, we select all option contracts that have positive bid prices and work with mid-quotes data as option observations. We define the moneyness of each contract with respect to the implied forward price, which we obtain as the median of the five forward prices implied from the put-call parity using the pairs closest to the at-the-money level. The risk-free rates are obtained by interpolating the LIBOR rates to each tenor.

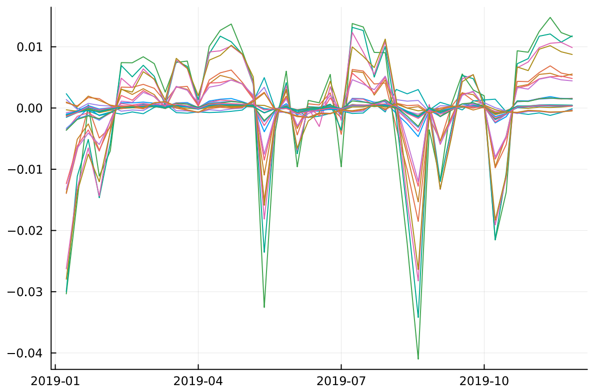
The construction of the option-implied CCF requires a wide coverage of moneyness range. Therefore, to reduce the impact of truncation and discretization errors in practice, we employ an interpolation-extrapolation scheme. In particular, following Boswijk et al. (2022), we interpolate option prices using cubic splines with carefully selected knot sequences and extrapolate beyond the observable range of strike prices based on a parametrization that satisfies the asymptotic results of Lee (2004). As discussed in Section 4, the interpolation-extrapolation scheme can be seen as a bias correction technique of the CCF estimators that does not affect the derived CLT results.

The calculation of the option-implied CCF then uses the Riemann sum approximation following (2.9). As in the simulations, we use the CCF arguments of the form:

$$u_j = \frac{jL}{\sqrt{\tau}}, \ j = 1, \ldots, q,$$

where we set $L = 0.5$ and $q = 3$. This results in $p = 2kq = 24$ cross-section dimension of the constructed dataset. The time series of the resulting factor model is displayed in Figure 1.

Figure 1: Time series plot of the factor model.



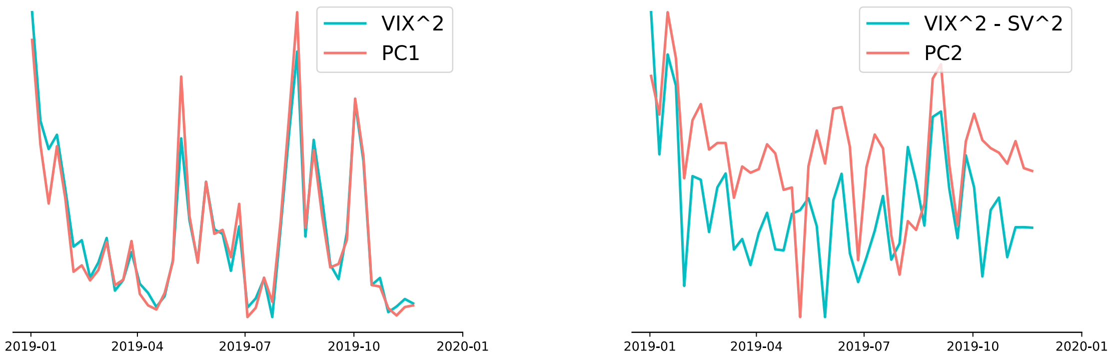

Note: This figure displays the time series of the factor model. Each line represents a row of the factor model and displays its behaviour over the sample period. The factor model is constructed as outlined in Section 2, using weekly observations on S&P 500 index options. The used options have expiration dates in one, two, three, and four weeks.

### 5.2 Results

Constructing the factor model as outlined in Section 2, we find that the first principal component explains 97.6% of the variation in the factor model. Applying the rank test as proposed in Section 3, we reject any level of reduced rank. A naive interpretation would be that at least $p = 2kq = 24$ factors are driving the dynamics of the options market. This is at odds with the literature and indeed seems highly unlikely: it might instead simply be a rejection of the affine model. Given the performance of the rank test in the simulations of Section 4, this is the most straightforward explanation.

With the potential failure of the affine model in mind, we take a different perspective; instead of asking how many factors are present in the data, we ask how many linear factors are required to provide a proper approximation to the underlying process. As a first, preliminary analysis, we consider the correlation of the set of principal components with the risk-neutral moments of Bakshi et al. (2003). They use the same spanning result of Carr and Madan (2001), presented in (2.6), to find portfolios that replicate the second, third, and fourth moment of log returns, and use this to find the risk-neutral variance, skewness, and kurtosis. The Bakshi et al. (2003) risk-neutral moments require options of a single tenor. We select the shortest, most liquid, seven-day tenor, but the results are robust to selecting a different tenor.

Figure 2: First principal components and selected proxies.




(a) First PC and the squared VIX.

(b) Second PC and jump variation.

Note: This figure displays the time series of first two principal component of the factor model, and two relevant proxies. In subfigure (a), the first principal component is plotted with the square of the CBOE VIX index (correlation $|\rho| = 0.972$). The VIX is a measure of risk-neutral expected volatility, which the first principal component clearly also tracks closely. In subfigure (b), the second principal component is displayed along with the difference between the square of the VIX and the square of the risk neutral spot volatility of Todorov (2019) ($|\rho| = 0.611$). This difference series is a crude measure of the jump variation.

Table 4 provides the results for the $R^2$ from regressing the time series of the risk-neutral moments onto the space spanned by up to the first ten principal components as extracted from the log CCF factor model. Clearly, the variance can be well explained by a single factor (see also Figure 2); it is in the higher order moments that the linear approximation struggles. If we use the raw moments, instead of the centered and scaled skewness and kurtosis, the performance increases by a fair amount, as displayed in the bottom half of the table. This is mainly due to the significantly better explanatory power of the first PC, and therefore seems to be driven by its high correlation with the variance, which is scaled out in the top half of the table.[10]

Table 4: $R^2$ of projection of risk-neutral moments onto the PCs.

| #PCs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Variance | 0.971 | 0.999 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Skewness | 0.299 | 0.327 | 0.646 | 0.661 | 0.745 | 0.792 | 0.798 | 0.803 | 0.812 | 0.813 |
| Kurtosis | 0.379 | 0.383 | 0.625 | 0.626 | 0.657 | 0.665 | 0.682 | 0.684 | 0.687 | 0.693 |
| $\mathbb{E}_t^{\mathbb{Q}}[\log(F_{t+7}/F_t)^2]$ | 0.965 | 0.999 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\mathbb{E}_t^{\mathbb{Q}}[\log(F_{t+7}/F_t)^3]$ | 0.740 | 0.847 | 0.924 | 0.940 | 0.975 | 0.989 | 0.992 | 0.993 | 0.994 | 0.998 |
| $\mathbb{E}_t^{\mathbb{Q}}[\log(F_{t+7}/F_t)^4]$ | 0.697 | 0.765 | 0.865 | 0.878 | 0.917 | 0.922 | 0.940 | 0.940 | 0.951 | 0.987 |

Note: This table provides the $R^2$ that results from regressing the time series of Bakshi et al. (2003) risk-neutral moments onto the space spanned by up to the first ten principal components as extracted from the characteristic function-based factor model. The top half of the table contains the scaled and centered variance, skewness and kurtosis, the bottom half uses the raw moments on which these are based.

[10] Proposition 3 can be used to analyse these results with more statistical rigour, but is not implemented in this version of the paper.

# 6 Conclusion

In this paper, we have developed a novel methodology to examine how many risk factors drive the price of a single asset, and what their behaviour is. We have proposed a linear factor model representation of the log discounted characteristic function of returns, on which we construct observations by exploiting information contained in options. The analysis further only uses standard and well-known factor modelling techniques, such as singular value decompositions. No specific parametric specification is necessary to perform this analysis, beyond the assumption that the underlying stochastic process is of the affine class. Our asymptotic framework is one in which the number of observed options, of varying strikes, tends to infinity. We have proved a feasible stable central limit theorem. In addition, we have proved that standard diagnostic criteria based on the sample eigenvalues of the covariance matrix can be used to consistently estimate the number of factors that drive the dynamics of the characteristic function over time. Moreover, we have shown the validity of the use of principal component analysis to extract these underlying factors from the constructed factor model.

# Appendices

# A Decomposing the complex-valued factor model

Let $Z_t = X_t + iY_t$, $t = 1, \dots, T$, be a complex, vector-valued time series of dimension $d$, and assume for simplicity it has mean zero. Let $\hat{\Gamma}_T = T^{-1}\sum_{t=1}^T Z_t Z_t^H$ be its sample covariance matrix, where $A^H$ denotes the conjugate transpose of any matrix $A$. Next, define the stacked process $\tilde{Z}_t = (X_t', Y_t')'$, with sample covariance matrix $\hat{\Sigma}_T = T^{-1}\sum_{t=1}^T \tilde{Z}_t \tilde{Z}_t'$.

The $k$-th (for $k = 1, \dots, d$) principal component (PC) of $Z_t$ is defined as the linear combination $\omega_k^H Z_t$, with $\omega_k = u_k + iv_k$, such that $\omega_k$ maximizes the (real-valued) sample variance $\hat{\lambda}_k$ of $\omega_k^H Z_t$:

$$\hat{\lambda}_k = \omega_k^H \hat{\Gamma} \omega_k = T^{-1}\sum_{t=1}^T \omega_k^H Z_t Z_t^H \omega_k = T^{-1}\sum_{t=1}^T (\omega_k^H Z_t)(\overline{\omega_k^H Z_t}), \tag{A.1}$$

under the constraint that $\omega_k^H \omega_i = \mathbb{1}_{\{i=k\}}$ for all $i = 1, \dots, k$. The linear combination itself can be written as

$$\omega_k^H Z_t = (u' - iv')(X_t + iY_t) = u'X_t + v'Y_t - i(u'Y_t - v'X_t), \tag{A.2}$$

which implies the $k$-th PC of $Z_t$ is a complex-valued time series unless $u'Y_t - v'X_t = 0$ for all $t$.

In contrast, the $k$-th (for $k = 1, \dots, 2d$) PC of $\tilde{Z}_t$ is defined as the linear combination $\tilde{\omega}_k' \tilde{Z}_t$, with $\tilde{\omega}_k = (\tilde{u}_k', \tilde{v}_k')'$, such that $\tilde{\omega}_k$ maximizes $\tilde{\lambda}_k = \tilde{\omega}_k' \hat{\Sigma} \tilde{\omega}_k = T^{-1}\sum_{t=1}^T \left(\tilde{\omega}_k' \tilde{Z}_t\right)^2$ under the constraint that $\tilde{\omega}_k' \tilde{\omega}_i = \mathbb{1}_{\{i=k\}}$ for all $i = 1, \dots, k$. It is obvious that the sets $\{\omega_k\}$ and $\{\tilde{\omega}_k\}$ solve different optimisation problems and are therefore generally not related. The exception to this is the case of circular symmetry, i.e., when the distribution of $e^{i\phi} Z_t$ is invariant to the choice of $\phi \in (-\pi, \pi]$, but this is a highly restrictive setting and not one we are likely to encounter. In any case, it is clear that

$$\tilde{\omega}_k' \tilde{Z}_t = \tilde{u}' X_t + \tilde{v}' Y_t, \tag{A.3}$$

which, by comparison to Equation (A.2), implies that the first optimisation problem does reduce to the second if we impose the restriction that the linear combination $\omega_k^H Z_t$ must be real-valued.

# B Theorems and proofs

This section contains the theorems, lemmas, and proofs omitted from the main text. The first subsections contain definitions and technical lemmas necessary for the proof of the stable CLT, which can be found in the subsection thereafter. Smaller results are collected in the last subsection.

## B.1 Definitions

Throughout the proofs, we denote by $C_t$ an $\mathcal{F}_t^{(0)}$-adapted random variable that is allowed to change from line to line. In addition, we use the following definitions:

$$\begin{aligned}
\chi_{j-1}(c) &:= (c_1 \cos(u m_{j-1}) + c_2 \sin(u m_{j-1})) e^{-m_{j-1}} \zeta_t(\tau, m_{j-1}) \Delta m_j, \\
s_{j-1}^2(c) &:= (c_1 \cos(u m_{j-1}) + c_2 \sin(u m_{j-1}))^2 e^{-2m_{j-1}} \sigma_t^2(\tau, m_{j-1}) (\Delta m_j)^2, \\
\tilde{\zeta}_t(u,\tau) &:= \sum_{j=2}^{n} e^{(iu-1)m_{j-1}} \zeta_t(\tau, m_{j-1}) \Delta m_j, \\
\varepsilon_t(u,\tau) &:= -\frac{e^{r\tau}}{F_t} \begin{pmatrix} \mathfrak{Re}\left(\frac{u^2+iu}{\phi_t(u,\tau)} \tilde{\zeta}_t(u,\tau)\right) \\ \mathfrak{Im}\left(\frac{u^2+iu}{\phi_t(u,\tau)} \tilde{\zeta}_t(u,\tau)\right) \end{pmatrix}, \\
\widehat{\Sigma}_{t,i}^{\zeta}(u,v) &:= \sum_{j=2}^{n} \begin{pmatrix} \cos(u m_{j-1}) \cos(v m_{j-1}) & \cos(u m_{j-1}) \sin(v m_{j-1}) \\ \sin(u m_{j-1}) \cos(v m_{j-1}) & \sin(u m_{j-1}) \sin(v m_{j-1}) \end{pmatrix} e^{-2m_{j-1}} \hat{\zeta}_t^2(\tau_i, m_{j-1}) (\Delta m_j)^2, \\
\Sigma_{t,i}^{\zeta}(u,v) &:= \int_{\mathbb{R}} \begin{pmatrix} \cos(um) \cos(vm) & \cos(um) \sin(vm) \\ \sin(um) \cos(vm) & \sin(um) \sin(vm) \end{pmatrix} e^{-2m} \sigma_t^2(\tau_i, m) \delta(m) \mathrm{d}m.
\end{aligned}$$

## B.2 Technical lemmas

**Lemma 2.** *Suppose Assumptions 2 and 3 hold. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some $\underline{\alpha}, \overline{\alpha} > 0$, then, for all $c \in \mathbb{R}^2 \setminus \{(0,0)'\}$,*

$$\frac{\sum_{j=2}^{n} \chi_{j-1}(c)}{\sqrt{\sum_{j=2}^{n} s_{j-1}^2(c)}} \xrightarrow{\mathcal{L}|\mathcal{F}^{(0)}} \mathcal{N}(0,1),$$

*with $\mathcal{L}|\mathcal{F}^{(0)}$ denoting $\mathcal{F}^{(0)}$-conditional convergence in distribution.*

*Proof.* Introduce the shorthand notation $\tilde{c}_{j-1} := c_1 \cos(u m_{j-1}) + c_2 \sin(u m_{j-1})$ and note that $\tilde{c}_{j-1} \leq |c_1| + |c_2|$. As the $\chi_j^2(c)$ are independent but heteroskedastic random variables, with $\mathbb{E}^{\mathbb{P}}\left[\chi_{j-1}^2(c) \middle| \mathcal{F}^{(0)}\right] \equiv s_{j-1}^2(c)$, we appeal to the Lyapunov central limit theorem. Write, for some

$\delta > 0$,

$$\frac{\sum_{j=2}^{n}\mathbb{E}^{\mathbb{P}}\left[|\chi_{j-1}(c)|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]}{\left(\sqrt{\sum_{j=2}^{n}s_{j-1}^{2}(c)}\right)^{2+\delta}} = \frac{\sum_{j=2}^{n}\mathbb{E}^{\mathbb{P}}\left[|\tilde{c}_{j-1}e^{-m_{j-1}}\zeta_t(\tau,m_{j-1})\Delta m_j|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]}{\left(\sqrt{\sum_{j=2}^{n}\tilde{c}_{j-1}^{2}e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})(\Delta m_j)^2}\right)^{2+\delta}}$$
$$\leq \frac{(\Delta m)^{1+\delta}\sum_{j=2}^{n}|\tilde{c}_{j-1}|^{2+\delta}e^{-(2+\delta)m_{j-1}}\mathbb{E}^{\mathbb{P}}\left[|\zeta_t(\tau,m_{j-1})|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\Delta m_j}{\left(\sqrt{\iota\Delta m\sum_{j=2}^{n}\tilde{c}_{j-1}^{2}e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})\Delta m_j}\right)^{2+\delta}}$$
$$= (\Delta m)^{\delta/2}\frac{\sum_{j=2}^{n}|\tilde{c}_{j-1}|^{2+\delta}e^{-(2+\delta)m_{j-1}}\mathbb{E}^{\mathbb{P}}\left[|\zeta_t(\tau,m_{j-1})|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\Delta m_j}{\iota^{1+\delta/2}\left(\sqrt{\sum_{j=2}^{n}\tilde{c}_{j-1}^{2}e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})\Delta m_j}\right)^{2+\delta}}$$
$$= C_t(\Delta m)^{\delta/2}.$$

The Lyapunov condition is satisfied if $C_t < \infty$. As the denominator is bounded away from zero, this is equivalent to the numerator being bounded. Using that the $\mathcal{F}^{(0)}$-conditional distribution of $\varkappa_t(\tau,m)$ does not depend on $m$, we have

$$\int_{\mathbb{R}}\mathbb{E}^{\mathbb{P}}\left[\left|e^{-m}\zeta_t(\tau,m)\right|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\mathrm{d}m = \mathbb{E}^{\mathbb{P}}\left[|\varkappa_t(\tau,m)|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\cdot\int_{\mathbb{R}}e^{-(2+\delta)m}\sigma_t^{2+\delta}(\tau,m)\mathrm{d}m < \infty,$$

by Assumption 2. As such, we have

$$\sum_{j=2}^{n}|\tilde{c}_{j-1}|^{2+\delta}e^{-(2+\delta)m_{j-1}}\mathbb{E}^{\mathbb{P}}\left[|\zeta_t(\tau,m_{j-1})|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\Delta m_k$$
$$\leq (|c_1|+|c_2|)^{2+\delta}\sum_{j=2}^{n}\mathbb{E}^{\mathbb{P}}\left[\left|e^{-m_{j-1}}\zeta_t(\tau,m_{j-1})\right|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\Delta m_j$$
$$\to (|c_1|+|c_2|)^{2+\delta}\int_{\mathbb{R}}\mathbb{E}^{\mathbb{P}}\left[\left|e^{-m}\zeta_t(\tau,m)\right|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\mathrm{d}m < \infty,$$

such that the Lyapunov condition is indeed satisfied. The result follows. □

**Lemma 3.** *Suppose Assumptions 2 and 3 hold. If $n\to\infty$, $\underline{K}\asymp n^{-\underline{\alpha}}$, $\overline{K}\asymp n^{\overline{\alpha}}$ for some $\underline{\alpha},\overline{\alpha}>0$, then, for any compact $\mathcal{U}\subset\mathbb{R}$,*

$$(\Delta m)^{-1/2}\begin{pmatrix}\mathfrak{Re}\left(\tilde{\zeta}_t(u,\tau)\right)\\ \mathfrak{Im}\left(\tilde{\zeta}_t(u,\tau)\right)\end{pmatrix}\overset{\mathcal{L}-s}{\to}\mathcal{N}(0,\Sigma_t^{\zeta}(u,\tau)),$$

*uniformly in $u\in\mathcal{U}$, where $(\Sigma_t^{\zeta}(u,\tau))_{ij} = \int_{\mathbb{R}}\varsigma_{ij}(u,m)e^{-2m}\sigma_t^2(\tau,m)\delta(m)\mathrm{d}m$, with $\varsigma_{11}(u,m) = \cos^2(um)$, $\varsigma_{22}(u,m)=\sin^2(um)$ and $\varsigma_{12}(u,m)=\cos(um)\sin(um)$, is $\mathcal{F}_t^{(0)}$-adapted.*

*Proof.* By the stable Cramér-Wold device, random vectors $X_n\in\mathbb{R}^p$ satisfy $X_n\overset{\mathcal{L}-s}{\to}X$ if and only if $c'X_n\overset{\mathcal{L}-s}{\to}c'X$ for all $c\in\mathbb{R}^p$. Note that the desired $\mathcal{F}^{(0)}$-stable convergence is implied by the stronger $\mathcal{F}^{(0)}$-conditional convergence; this trivially follows from their respective definitions and the Dominated Convergence Theorem. As

$$c_1\,\mathfrak{Re}\Big(\tilde{\zeta}_t(u,\tau)\Big)+c_2\,\mathfrak{Im}\Big(\tilde{\zeta}_t(u,\tau)\Big)\equiv\sum_{j=2}^{n}\chi_{j-1}(c),$$

pointwise convergence thus follows after an appropriate definition of the covariance matrix, on account of Lemma 2. This asymptotic covariance is clearly related to the $s_j(c)$. Note that, upon taking $c=(1,0)'$ or $c=(0,1)'$, we have a univariate CLT for the real and imaginary parts of $\tilde{\zeta}_t(u,\tau)$ separately. Specifically, we have

$$\begin{aligned}(\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left(\mathfrak{Re}\Big(\tilde{\zeta}_t(u,\tau)\Big)\Big|\mathcal{F}^{(0)}\right)&=(\Delta m)^{-1}\sum_{j=2}^{n}s_{j-1}^2((1,0)')\\&=\sum_{j=2}^{n}\cos^2(um_{j-1})e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})\frac{(\Delta m_j)^2}{\Delta m}\xrightarrow{p}\left(\Sigma_t^{\zeta}(u,\tau)\right)_{11},\\(\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left(\mathfrak{Im}\Big(\tilde{\zeta}_t(u,\tau)\Big)\Big|\mathcal{F}^{(0)}\right)&=(\Delta m)^{-1}\sum_{j=2}^{n}s_{j-1}^2((0,1)')\xrightarrow{p}\left(\Sigma_t^{\zeta}(u,\tau)\right)_{22},\end{aligned}$$

and lastly,

$$\begin{aligned}&(\Delta m)^{-1}\,\mathbb{C}\mathrm{ov}^{\mathbb{P}}\left(\mathfrak{Re}\Big(\tilde{\zeta}_t(u,\tau)\Big),\mathfrak{Im}\Big(\tilde{\zeta}_t(u,\tau)\Big)\Big|\mathcal{F}^{(0)}\right)\\&=\sum_{j=2}^{n}\cos(um_{j-1})\sin(um_{j-1})e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})\frac{(\Delta m_j)^2}{\Delta m}\xrightarrow{p}\left(\Sigma_t^{\zeta}(u,\tau)\right)_{12}.\end{aligned}$$

Note that all three terms are uniformly continuous in $u$. Upon developing the square in our expression for $s_{j-1}(c)$, it follows that the limiting variance of the linear combination corresponds to the variance of the limit of the linear combination. Therefore, the Cramér-Wold device can be applied. The limiting distribution for fixed $u\in\mathbb{R}$ directly follows from the properties of the multivariate normal distribution. The extension to the uniform variant is straightforward, noting that

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\left|\tilde{\zeta}_t(u,\tau)-\tilde{\zeta}_t(v,\tau)\right|^2\Big|\mathcal{F}^{(0)}\right]\\&=\mathbb{E}^{\mathbb{P}}\left[\left|\sum_{j=2}^{n}e^{(iu-1)m_{j-1}}\zeta_t(\tau,m_{j-1})\Delta m_j-\sum_{j=2}^{n}e^{(iv-1)m_{j-1}}\zeta_t(\tau,m_{j-1})\Delta m_j\right|^2\Big|\mathcal{F}^{(0)}\right]\\&\leq 3\sum_{j=2}^{n}\left|e^{ium_{j-1}}-e^{ivm_{j-1}}\right|^2\mathbb{E}^{\mathbb{P}}\left[\left|e^{-m_{j-1}}\zeta_t(\tau,m_{j-1})\right|^2(\Delta m_j)^2\Big|\mathcal{F}^{(0)}\right]\\&=3\sum_{j=2}^{n}\left|\int_v^u im_{j-1}e^{ixm_{j-1}}\mathrm{d}x\right|^2e^{-2m_{j-1}}\mathbb{E}^{\mathbb{P}}\left[\zeta_t^2(\tau,m_{j-1})(\Delta m_j)^2\Big|\mathcal{F}^{(0)}\right]\\&\leq 3\,(u-v)^2\sum_{j=2}^{n}|m_{j-1}|e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})(\Delta m_j)^2=C_t\,(u-v)^2\,\Delta m,\end{aligned}$$

with $C_t$ finite-valued by Assumption 2. The result follows from Theorems 8.1 and 12.3 of Billingsley (1968). □

**Lemma 4.** *Suppose Assumptions 2 and 3 hold. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some $\underline{\alpha}, \overline{\alpha} > 0$, then, for any $u, v \in \mathbb{R}$,*

$$\left\| (\Delta m)^{-1} \mathbb{E}^{\mathbb{P}} \left[ \widehat{\Sigma}^{\zeta}_{t,i}(u,v) \middle| \mathcal{F}^{(0)} \right] - \Sigma^{\zeta}_{t,i}(u,v) \right\| = o_p(1).$$

*Proof.* It suffices to show that $\mathbb{E}^{\mathbb{P}} \left[ \hat{\zeta}_t^2(\tau_i, m_j) \middle| \mathcal{F}^{(0)} \right] \xrightarrow{p} \sigma_t^2(\tau_i, m_j)$, as the result then follows from the convergence of the Riemann sum, which holds by Assumption 2. We focus on the "standard" values of $j$, as the finite number of adjusted terms have a negligible effect on the full sum. By the Lipschitz continuity of Assumption 3, we have

$$\hat{\zeta}_t(\tau_i, m_j) = \sqrt{\tfrac{2}{3}} \left( \zeta_t(\tau_i, m_j) - \tfrac{1}{2} \left( \zeta_t(\tau_i, m_{j-1}) + \zeta_t(\tau_i, m_{j+1}) \right) \right) + o_p(1).$$

Then,

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}} \left[ \hat{\zeta}_t^2(\tau_i, m_j) \middle| \mathcal{F}^{(0)} \right] &= \mathbb{E}^{\mathbb{P}} \left[ \left( \sqrt{\tfrac{2}{3}} \left( \zeta_t(\tau_i, m_j) - \tfrac{1}{2} \left( \zeta_t(\tau_i, m_{j-1}) + \zeta_t(\tau_i, m_{j+1}) \right) \right) + o_p(1) \right)^2 \middle| \mathcal{F}^{(0)} \right] \\
&= \tfrac{2}{3} \mathbb{E}^{\mathbb{P}} \left[ \left( \zeta_t^2(\tau_i, m_j) + \tfrac{1}{4} \left( \zeta_t^2(\tau_i, m_{j-1}) + \zeta_t^2(\tau_i, m_{j+1}) \right) \right) \middle| \mathcal{F}^{(0)} \right] + o_p(1) \\
&= \sigma_t^2(\tau, m_j) \mathbb{E}^{\mathbb{P}} \left[ \tfrac{2}{3} \left( \varkappa_t^2(\tau_i, m_j) + \tfrac{1}{4} \left( \varkappa_t^2(\tau_i, m_{j-1}) + \varkappa_t^2(\tau_i, m_{j+1}) \right) \right) \middle| \mathcal{F}^{(0)} \right] + o_p(1) \\
&= \sigma_t^2(\tau, m_k) + o_p(1),
\end{aligned}$$

using Assumption 2. □

**Lemma 5.** *Suppose Assumptions 2 and 3 hold. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some $\underline{\alpha}, \overline{\alpha} > 0$, then,*

$$\mathbb{V}\mathrm{ar}^{\mathbb{P}} \left( \hat{\zeta}_t^2(\tau_i, m_{j-1}) \middle| \mathcal{F}^{(0)} \right) = \tfrac{1}{2} \sigma_t^4(\tau, m_j) \left( \mathbb{E}^{\mathbb{P}} \left[ \varkappa_t^4 \middle| \mathcal{F}^{(0)} \right] + 1 \right) + o_p(1).$$

*Proof.* Using the Lipschitz continuity of Assumption 3, we write:

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}^{\mathbb{P}} \left( \hat{\zeta}_t^2(\tau_i, m_{j-1}) \middle| \mathcal{F}^{(0)} \right) \\
&= \sigma_t^4(\tau, m_j) \, \mathbb{V}\mathrm{ar}^{\mathbb{P}} \left( \tfrac{2}{3} \left( \varkappa_t(\tau_i, m_j) - \tfrac{1}{2} \varkappa_t(\tau_i, m_{j-1}) - \tfrac{1}{2} \varkappa_t(\tau_i, m_{j+1}) \right)^2 \middle| \mathcal{F}^{(0)} \right) + o_p(1) \\
&= \tfrac{4}{9} \sigma_t^4(\tau, m_j) \, \mathbb{V}\mathrm{ar}^{\mathbb{P}} \Big( \varkappa_t^2(\tau_i, m_j) + \tfrac{1}{4} \varkappa_t^2(\tau_i, m_{j-1}) + \tfrac{1}{4} \varkappa_t^2(\tau_i, m_{j-1}) \\
&\qquad - \varkappa_t(\tau_i, m_{j-1}) \varkappa_t(\tau_i, m_j) - \varkappa_t(\tau_i, m_j) \varkappa_t(\tau_i, m_{j+1}) + \tfrac{1}{2} \varkappa_t(\tau_i, m_{j-1}) \varkappa_t(\tau_i, m_{j+1}) \Big| \mathcal{F}^{(0)} \Big) + o_p(1) \\
&= \sigma_t^4(\tau, m_j) \left( \tfrac{1}{2} \, \mathbb{V}\mathrm{ar}^{\mathbb{P}} \left( \varkappa_t^2(\tau_i, m_j) \middle| \mathcal{F}^{(0)} \right) + \mathbb{V}\mathrm{ar}^{\mathbb{P}} \left( \varkappa_t(\tau_i, m_{j-1}) \varkappa_t(\tau_i, m_j) \middle| \mathcal{F}^{(0)} \right) \right) + o_p(1) \\
&= \tfrac{1}{2} \sigma_t^4(\tau, m_j) \left( \mathbb{E}^{\mathbb{P}} \left[ \varkappa_t^4 \middle| \mathcal{F}^{(0)} \right] + 1 \right) + o_p(1),
\end{aligned}$$

using Assumption 2. □

**Lemma 6.** *Suppose Assumptions 2 and 3 hold. If $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ for some*

$\underline{\alpha}, \overline{\alpha} > 0$, *then, for any* $u, v \in \mathbb{R}$,

$$\left\| \widehat{\Sigma}^{\zeta}_{t,i}(u,v) - \mathbb{E}^{\mathbb{P}}\left[ \widehat{\Sigma}^{\zeta}_{t,i}(u,v) \middle| \mathcal{F}^{(0)} \right] \right\| = o_p(\Delta m).$$

*Proof.* For the $\mathcal{F}^{(0)}$-conditional variance of the $(g,h)$-th element of $\widehat{\Sigma}^{\zeta}_{t,i}(u,v)$, we have

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left( (\Delta m)^{-1} \left( \widehat{\Sigma}^{\zeta}_{t,i}(u,v) \right)_{gh} \middle| \mathcal{F}^{(0)} \right) \\
&\leq \mathbb{V}\mathrm{ar}^{\mathbb{P}}\left( (\Delta m)^{-1} \sum_{j=2}^{n} e^{-2m_{j-1}} \hat{\zeta}^2_t(\tau_i, m_{j-1})(\Delta m_j)^2 \middle| \mathcal{F}^{(0)} \right) \\
&= (\Delta m)^{-2} \sum_{j=2}^{n} e^{-4m_{j-1}} \, \mathbb{V}\mathrm{ar}^{\mathbb{P}}\left( \hat{\zeta}^2_t(\tau_i, m_{j-1}) \middle| \mathcal{F}^{(0)} \right) (\Delta m_j)^4 \\
&\qquad + (\Delta m)^{-2} \sum_{|l-j|=1} e^{-2m_l - 2m_j} \, \mathbb{C}\mathrm{ov}^{\mathbb{P}}\left( \hat{\zeta}^2_t(\tau_i, m_l), \hat{\zeta}^2_t(\tau_i, m_j) \middle| \mathcal{F}^{(0)} \right) (\Delta m_j)^2 (\Delta m_l)^2 \\
&\leq 3 \sum_{j=2}^{n} e^{-4m_{j-1}} \, \mathbb{V}\mathrm{ar}^{\mathbb{P}}\left( \hat{\zeta}^2_t(\tau_i, m_{j-1}) \middle| \mathcal{F}^{(0)} \right) (\Delta m_j)^2 + o_p(1) \\
&\leq C_t \Delta m \sum_{j=2}^{n} e^{-4m_{j-1}} \sigma^4_t(\tau, m_j) \Delta m_j + o_p(1) = \mathcal{O}_p(\Delta m),
\end{aligned}$$

using Lemma 5 and Assumption 3. The result follows from Chebyshev's inequality. ☐

**Lemma 7.** *Suppose Assumptions 2 and 3 hold. If* $n \to \infty$, $\underline{K} \asymp n^{-\underline{\alpha}}$, $\overline{K} \asymp n^{\overline{\alpha}}$ *for some* $\underline{\alpha}, \overline{\alpha} > 0$, *then, for any compact* $\mathcal{U} \subset \mathbb{R}$,

$$\left\| (\Delta m)^{-1} \widehat{\Sigma}^{\zeta}_{t,i}(u,v) - \Sigma^{\zeta}_{t,i}(u,v) \right\| = o_p(1),$$

*uniformly in* $u, v \in \mathcal{U}$.

*Proof.* Pointwise convergence follows from Lemma 4 and 6, and the triangle inequality. As the dependence on the argument $(u,v)$ is through bounded and (uniformly) continuous trigonometric functions, uniform consistency directly follows. ☐

### B.3 Proof of Lemma 1

Denote $M = (p \wedge T)$, i.e., the smaller of the row and column dimension of the factor model. Denote by $\sigma_1(A) \geq \ldots \geq \sigma_M(A)$ the singular values of a matrix $A \in \mathbb{R}^{p \times T}$.

Given $(p \wedge T - 1) > d$, Assumption 1 trivially implies $\operatorname{rank}\left(\boldsymbol{B}\ddot{X}\right) = d$, such that $\sigma_j\left(\boldsymbol{B}\ddot{X}\right) > 0$ for $j = 1, \ldots, d$, while $\sigma_{d+1}\left(\boldsymbol{B}\ddot{X}\right) = 0$.

By Weyl's inequality for singular values (Theorem 3.3.16 in Horn and Johnson (1991)), we have for all $j = 1, ..., M$:

$$\begin{aligned}\sigma_j(\boldsymbol{B}\ddot{X}) &= \sigma_j\left(\widehat{\ddot{\Psi}} - \ddot{\varepsilon} - \ddot{r}\right) \\ &\leq \sigma_j\left(\widehat{\ddot{\Psi}}\right) + \sigma_1\left(-\ddot{\varepsilon}\right) + \sigma_1\left(-\ddot{r}\right) \\ &= \sigma_j\left(\widehat{\ddot{\Psi}}\right) + \sigma_1\left(\ddot{\varepsilon}\right) + \sigma_1(\ddot{r}),\end{aligned}$$

using that the singular values of $A$ and $-A$ are equivalent, being defined as the square roots of the eigenvalues of $A'A$. As for $j = 1, \ldots, d$ we have $\sigma_j(\boldsymbol{B}\ddot{X}) > 0$, so too is $\sigma_j\left(\widehat{\ddot{\Psi}}\right)$ bounded away from zero, provided that $\sigma_1\left(\ddot{\varepsilon}\right)$ and $\sigma_1(\ddot{r})$ tend to zero. This can easily be shown. Note that for any matrix $A$ with column dimension $T$, we have for all $j = 1, \ldots, M$:

$$\sigma_j\left(A\right) \geq \sigma_j\left(AM_{\iota_T}\right),$$

as the demeaning matrix $M_{\iota_T} = I_T - T^{-1}\iota_T\iota_T'$ only has eigenvalues 0 or 1, being a projection matrix. This implies that $\sigma_1\left(\ddot{\varepsilon}\right) \leq \sigma_1\left(\varepsilon\right)$. This tends to zero if the spanned CCF is consistent, which is a straightforward corollary to Proposition 1 of Boswijk et al. (2022). The same line of reasoning holds up for $\sigma_1(\ddot{r})$. As such, we indeed have that $\sigma_j\left(\widehat{\ddot{\Psi}}\right) \geq \sigma_j\left(\boldsymbol{B}\ddot{X}\right) - o_p(1) > 0$ for $n$ large, for all $j = 1, \ldots, d$. This proves the first statement of the lemma.

We continue by focusing on the $d+1$-th singular value specifically. Weyl's inequality can also be used to write:

$$\begin{aligned}\sigma_{d+1}(\widehat{\ddot{\Psi}}) &= \sigma_{d+1}\left(\boldsymbol{B}\ddot{X} + \ddot{\varepsilon} + \ddot{r}\right) \\ &\leq \sigma_{d+1}\left(\boldsymbol{B}\ddot{X}\right) + \sigma_1\left(\ddot{\varepsilon}\right) + \sigma_1(\ddot{r}) \\ &= \sigma_1\left(\ddot{\varepsilon}\right) + \sigma_1(\ddot{r}) = o_p(1),\end{aligned}$$

which proves the second part of the lemma. □

### B.4 Proof of Theorem 1.

We can write

$$\text{vec}\left(\widehat{\ddot{\Psi}}\right) - \text{vec}\left(\boldsymbol{B}\ddot{X}\right) = \text{vec}(\ddot{\varepsilon} + \ddot{r}) = \text{vec}(\ddot{\varepsilon}) + o_p(\sqrt{\Delta m}),$$

using Proposition 1 of Boswijk et al. (2022), who show that $r_t = o_p(\sqrt{\Delta m})$, while $\varepsilon_t = \mathcal{O}_p(\sqrt{\Delta m})$. As such, the object of interest is the $\varepsilon_{t,i}$, as defined in Equation (2.13), or indeed its functional counterpart $\varepsilon_t(u, \tau)$, which we can write as:

$$\varepsilon_t(u,\tau) = -\frac{e^{r\tau}}{F_t|\phi_t(u,\tau)|^2}\begin{pmatrix}u^2 & u \\ u & -u^2\end{pmatrix}\begin{pmatrix}\mathfrak{Re}(\phi_t(u,\tau)) & \mathfrak{Im}(\phi_t(u,\tau)) \\ \mathfrak{Im}(\phi_t(u,\tau)) & -\mathfrak{Re}(\phi_t(u,\tau))\end{pmatrix}\begin{pmatrix}\mathfrak{Re}\left(\tilde{\zeta}_t(u,\tau)\right) \\ \mathfrak{Im}\left(\tilde{\zeta}_t(u,\tau)\right)\end{pmatrix}.$$

This is an $\mathcal{F}^{(0)}$-measurable linear transformation of the $\tilde{\zeta}_t(u,\tau)$, for which we have derived a $\mathcal{F}^{(0)}$-stable CLT in Lemma 3. By the properties of the normal distribution, $(\Delta m)^{-1/2}\varepsilon_t(u,\tau)$ is also asymptotically $\mathcal{F}^{(0)}$-mixed normal, uniformly in $u \in \mathcal{U}$. Upon evaluation in the finite grid

$\{u_j\}_j$, this implies asymptotic normality for $(\Delta m)^{-1/2}\varepsilon_{t,i}$. The limiting covariance matrix $H_{t,i} := \operatorname{plim}\big((\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}(\varepsilon_{t,i}|\mathcal{F}^{(0)})\big)$ is simply a reordering of a block-matrix with the $\left\{\Sigma^{\zeta}_{t,i}(u_g,u_h)\right\}_{g,h}$ as blocks, combined with a pre- and post-multiplication by an $\mathcal{F}^{(0)}$-measurable matrix. In practice, it is far easier to use a more direct, algebraically equivalent expression for $H_{t,i}$ and its estimate, which we have used in the statement of this theorem.

The $\mathcal{F}^{(0)}$-conditional independence of the option pricing errors in both time $t$ and tenor $\tau$ imply that the limiting distributions can be "stacked": $(\Delta m)^{-1/2}\operatorname{vec}(\varepsilon)$ remains asymptotically normal, with a block-diagonal covariance matrix $\tilde{\mathcal{H}} := \operatorname{plim}\big((\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left(\operatorname{vec}(\varepsilon)\middle|\mathcal{F}^{(0)}\right)\big)$. Clearly, this is the case for the demeaned $\operatorname{vec}(\ddot{\varepsilon})$ as well, as this is a linear transformation:

$$\operatorname{vec}(\ddot{\varepsilon}) = \operatorname{vec}(\varepsilon M_{\iota_T}) = (M_{\iota_T}\otimes I_p)\operatorname{vec}(\varepsilon)\,.$$

This results in the (reduced rank) covariance matrix

$$\begin{aligned}\mathcal{H} &:= \operatorname{plim}\Big((\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left(\operatorname{vec}(\varepsilon M_{\iota_T})\middle|\mathcal{F}^{(0)}\right)\Big)\\ &= (M_{\iota_T}\otimes I_p)\operatorname{blkdiag}\{H_{1,1},\dots,H_{1,k},H_{2,1},\dots,H_{T,k}\}\,(M_{\iota_T}\otimes I_p)\,.\end{aligned}$$

We denote the feasible counterpart as $\hat{\mathcal{H}}$, replacing the $H_{t,i}$ by their unscaled estimators $\hat{H}_{t,i}$. Clearly, the scaled consistency of Lemma 7 carries over to this quantity. The result follows. □

### B.5 Proof of Proposition 1.

Define $H := (B'B)(\ddot{X}\hat{\ddot{X}}')S^{-2}$. Bai and Ng (2020) show the asymptotic equivalence of a set of alternative rotation matrices $H_l$ for $l = 1,2,3,4$. We use this result as well, though we do not normalize by $p$ or $T$ as the dimensions are fixed and finite in our setting.

*(i).* With $H_3 = (\hat{\ddot{X}}\ddot{X}')^{-1}$, we have

$$\begin{aligned}\widehat{B} = \hat{\ddot{\Psi}}\hat{\ddot{X}}' &= B\ddot{X}\hat{\ddot{X}}' + \ddot{\varepsilon}\hat{\ddot{X}}' + \ddot{r}\hat{\ddot{X}}'\\ &= BH_3'^{-1} + \ddot{\varepsilon}\ddot{X}'H_3 + \ddot{\varepsilon}\left(\hat{\ddot{X}} - H_3'\ddot{X}\right)' + o_p(\sqrt{\Delta m}).\end{aligned}$$

As $\ddot{\varepsilon}\left(\hat{\ddot{X}} - H_3'\ddot{X}\right)' = \mathcal{O}_p(\Delta m)$, it is asymptotically dominated by the $\mathcal{O}_p(\sqrt{\Delta m})$ term $\ddot{\varepsilon}\ddot{X}'H_3$. We find

$$(\Delta m)^{-1/2}\left(\widehat{B} - BH_3'^{-1}\right) = (\Delta m)^{-1/2}\ddot{\varepsilon}\ddot{X}'H_3 + o_p(1).$$

A mixed normal limit for $\varepsilon$ is derived in the proof of Theorem 1 as presented in Appendix B.4, and post-multiplication simply changes the variance matrix. This asymptotic variance can be found as in Bai and Ng (2020), using that

$$\operatorname{plim}\Big((\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left(\ddot{\varepsilon}'e^{(p)}_j\middle|\mathcal{F}^{(0)}\right)\Big) = (I_T\otimes e^{(p)}_j)'\mathcal{H}(I_T\otimes e^{(p)}_j).$$

*(ii).* With $H_4 = B'\widehat{B}(\widehat{B}'\widehat{B})^{-1}$, we can write

$$\widehat{\ddot{X}} = (\widehat{B}'\widehat{B})^{-1}\widehat{B}'\widehat{\ddot{\Psi}} = (\widehat{B}'\widehat{B})^{-1}\widehat{B}'B\ddot{X} + (\widehat{B}'\widehat{B})^{-1}\widehat{B}'\ddot{\varepsilon} + (\widehat{B}'\widehat{B})^{-1}\widehat{B}'\ddot{r}$$
$$= H_4'\ddot{X} + (\widehat{B}'\widehat{B})^{-1}H_4^{-1}B'\ddot{\varepsilon} + (\widehat{B}'\widehat{B})^{-1}\left(\widehat{B} - BH_4'^{-1}\right)'\ddot{\varepsilon} + o_p(\sqrt{\Delta m}).$$

The result follows analogously to part *(i)*. Straightforward algebra shows that

$$\operatorname{plim}\Big((\Delta m)^{-1}\,\mathbb{V}\mathrm{ar}^{\mathbb{P}}\left(\ddot{\varepsilon}e_t^{(T)}\middle|\mathcal{F}^{(0)}\right)\Big) = (e_t^{(T)} \otimes I_p)'\mathcal{H}(e_t^{(T)} \otimes I_p),$$

and the asymptotic variance follows. □

### B.6 Proof of Proposition 2.

We can write

$$\widehat{R}_t^{\mathbb{Q}}(h,\tau) - R_t^{\mathbb{Q}}(h,\tau) =$$
$$\underbrace{\frac{e^{r\tau}}{F_t}\sum_{j=2}^{n} h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}\zeta_t(\tau,m_{j-1})\Delta m_j}_{=:\nu^{(1)}(h,\tau)}$$
$$\underbrace{-\frac{e^{r\tau}}{F_t}\int_{-\infty}^{\underline{m}} h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)\mathrm{d}m - \frac{e^{r\tau}}{F_t}\int_{\overline{m}}^{\infty} h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)\mathrm{d}m}_{=:\nu^{(2)}(h,\tau)}$$
$$\underbrace{+\frac{e^{r\tau}}{F_t}\sum_{j=2}^{n}\int_{m_{j-1}}^{m_j}\Big[h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}O_t(\tau,m_{j-1}) - h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)\Big]\mathrm{d}m}_{=:\nu^{(3)}(h,\tau)},$$

which form the stochastic spanning error, and the $\mathcal{F}_t^{(0)}$-measurable truncation and discretization bias, respectively. This proof consists of three parts: finding a CLT for the sequence of random variables

$$\omega_{j-1} := h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}\zeta_t(\tau,m_{j-1})\Delta m_j,$$

to find the limit behaviour of $\nu^{(1)}(h,\tau)$, and the asymptotic order of the bias terms $\nu^{(2)}(h,\tau)$ and $\nu^{(3)}(h,\tau)$.

*Part 1: spanning error.* Define

$$s_\omega^2(j-1) :=\mathbb{E}^{\mathbb{P}}\left[\omega_{j-1}^2\middle|\mathcal{F}^{(0)}\right] = h^2(h-1-m_{j-1})^2 m_{j-1}^{2h-4}e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})(\Delta m_j)^2.$$

Write, for some $\delta > 0$:

$$\frac{\sum_{j=2}^n \mathbb{E}^{\mathbb{P}}\left[|\omega_{j-1}|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]}{\left(\sqrt{\sum_{j=2}^n s_\omega^2(j-1)}\right)^{2+\delta}} = \frac{\sum_{j=2}^n \mathbb{E}^{\mathbb{P}}\left[|h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}\zeta_t(\tau,m_{j-1})\Delta m_j|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]}{\left(\sqrt{\sum_{j=2}^n h^2(h-1-m_{j-1})^2 m_{j-1}^{2h-4}e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})(\Delta m_j)^2}\right)^{2+\delta}}$$
$$\leq (\Delta m)^{\delta/2}\frac{\sum_{j=2}^n |(h-1-m_{j-1})|^{2+\delta}|m_{j-1}|^{(2+\delta)(h-2)}e^{-(2+\delta)m_{j-1}}\mathbb{E}^{\mathbb{P}}\left[|\zeta_t(\tau,m_{j-1})|^{2+\delta}\middle|\mathcal{F}^{(0)}\right]\Delta m_j}{\iota^{1+\delta/2}\left(\sqrt{\sum_{j=2}^n (h-1-m_{j-1})^2 m_{j-1}^{2h-4}e^{-2m_{j-1}}\sigma_t^2(\tau,m_{j-1})\Delta m_j}\right)^{2+\delta}},$$

wherein the denominator is bounded away from zero. As such, for the fraction to tend to zero, what remains to be shown is that the numerator is bounded. Consider

$$\int_{\mathbb{R}} |h-1-m|^{2+\delta}|m|^{(2+\delta)(h-2)}e^{-(2+\delta)m}\sigma_t^{2+\delta}(\tau,m)\mathrm{d}m$$
$$= \int_{\mathbb{R}} \left(|h-1-m|\,|m|^{h-2}e^{-m}\sigma_t(\tau,m)\right)^{2+\delta}\mathrm{d}m,$$

which is finite by assumption. As such, the Riemann sum converges, and the Lyapunov condition is satisfied. The conditional convergence in distribution implies the desired stable convergence in distribution. The expression for the limiting variance is obvious.

*Part 2: truncation error.* As for the bias terms, analogous to Lemma 2 of Boswijk et al. (2022), we have

$$\left|\int_{\overline{m}}^{\infty} h(h-1-m)m^{h-2}e^{-m}\frac{O_t(\tau,m)}{F_t}\mathrm{d}m\right| \leq C_t\int_{\overline{m}}^{\infty}\left|(h-1-m)m^{h-2}e^{-(1+p)m}\right|\mathrm{d}m$$
$$= \mathcal{O}_p(\overline{m}^{h-1}e^{-(1+p)\overline{m}}),$$
$$\left|\int_{-\infty}^{\underline{m}} h(h-1-m)m^{h-2}e^{-m}\frac{O_t(\tau,m)}{F_t}\mathrm{d}m\right| \leq C_t\int_{-\infty}^{\underline{m}}\left|(h-1-m)m^{h-2}e^{qm}\right|\mathrm{d}m$$
$$= \mathcal{O}_p(\underline{m}^{h-1}e^{q\underline{m}}),$$

with $(p,q) = \sup_{(l,k)}\{l : \mathbb{E}_t^{\mathbb{Q}}[(F_{t+\tau}/F_t)^l] < \infty\} \times \{k : \mathbb{E}_t^{\mathbb{Q}}[(F_{t+\tau}/F_t)^{-k}] < \infty\}$. As such, the truncation-induced bias is of order

$$\mathcal{O}_p\left(|\overline{m}|^{h-1}e^{-(1+p)|\overline{m}|}\right) + \mathcal{O}_p\left(|\underline{m}|^{h-1}e^{-q|\underline{m}|}\right) = \mathcal{O}_p\left(\log(n)^{h-1}n^{-(q\underline{\alpha}\wedge(1+p)\overline{\alpha})}\right),$$

which is asymptotically negligible compared to the spanning error under the assumptions on $\underline{\alpha},\overline{\alpha}$ and the existing moments.

*Part 3: discretization error.* What remains to be shown is the order of the discretization error. We write

$$h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}O_t(\tau,m_{j-1}) - h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)$$
$$= h\left((h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}} - (h-1-m)m^{h-2}e^{-m}\right)O_t(\tau,m_{j-1})$$
$$+ h(h-1-m)m^{h-2}e^{-m}\left(O_t(\tau,m_{j-1}) - O_t(\tau,m)\right).$$

We first focus on the first term. Using Lemma 1 of Boswijk et al. (2022) and the mean value theorem, we find

$$\begin{aligned}
&\left|h\left((h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}-(h-1-m)m^{h-2}e^{-m}\right)O_t(\tau,m_{j-1})\right|\\
&\leq C_t\left|((m^*)^2+(2h-2)m^*+h^2-3h+2)(m^*)^{h-3}e^{-m^*}\right|e^{-pm_{j-1}\wedge(1+q)m_{j-1}}\Delta m_j\\
&\leq C_t\left|((m^*)^2+(2h-2)m^*+h^2-3h+2)(m^*)^{h-3}\right|e^{-(p+1)m_{j-1}\wedge qm_{j-1}}\Delta m_j\\
&\leq C_te^{-pm_{j-1}\wedge(q-1)m_{j-1}}\Delta m_j.
\end{aligned}$$

For the second term, we use the mean value theorem once more, along with the proof of Lemma 2 of Boswijk et al. (2022):

$$\begin{aligned}
&\left|h(h-1-m)m^{h-2}e^{-m}\left(O_t(\tau,m_{j-1})-O_t(\tau,m)\right)\right|\\
&\leq C_t\left|(h-1-m)m^{h-2}\right|e^{-(p+1)\tilde{m}\wedge q\tilde{m}}e^{\Delta m}\Delta m_j\\
&\leq C_t\left|(h-1-m)m^{h-2}\right|e^{-(p+1)m_{j-1}\wedge qm_{j-1}}e^{\Delta m}\Delta m_j\\
&\leq C_te^{-pm_{j-1}\wedge(q-1)m_{j-1}}\Delta m_j.
\end{aligned}$$

Hence, we conclude that

$$\begin{aligned}
&\left|h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}O_t(\tau,m_{j-1})-h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)\right|\\
&\leq C_te^{-pm_{j-1}\wedge(q-1)m_{j-1}}\Delta m_j,
\end{aligned}$$

and thus that

$$\begin{aligned}
&\left|\sum_{j=2}^{n}\int_{m_{j-1}}^{m_j}\left[h(h-1-m_{j-1})m_{j-1}^{h-2}e^{-m_{j-1}}O_t(\tau,m_{j-1})-h(h-1-m)m^{h-2}e^{-m}O_t(\tau,m)\right]\mathrm{d}m\right|\\
&\leq C_t\sum_{j=2}^{n}e^{-pm_{j-1}\wedge(q-1)m_{j-1}}(\Delta m_j)^2\\
&\leq C_t\Delta m\sum_{j=2}^{n}e^{-pm_{j-1}\wedge(q-1)m_{j-1}}\Delta m_j.
\end{aligned}$$

Note that this Riemann sum converges to a positive constant:

$$\begin{aligned}
\sum_{j=2}^{n}e^{-pm_{j-1}\wedge(q-1)m_{j-1}}\Delta m_j\to\int_{\underline{m}}^{\overline{m}}e^{-pm\wedge(q-1)m}\mathrm{d}m&=\int_{\underline{m}}^{0}e^{(q-1)m}\mathrm{d}m+\int_{0}^{\overline{m}}e^{-pm}\mathrm{d}m\\
&=(q-1)^{-1}\left(1-e^{(q-1)\underline{m}}\right)+p^{-1}\left(1-e^{-p\overline{m}}\right)\\
&=(q-1)^{-1}+p^{-1}+\mathcal{O}_p\left(n^{-((q-1)\underline{\alpha}\wedge p\overline{\alpha})}\right).
\end{aligned}$$

Finally, we can conclude that the discretization error is of order

$$\mathcal{O}_p(\Delta m)+\mathcal{O}_p\left(\Delta m\right)\mathcal{O}_p\left(n^{-((q-1)\underline{\alpha}\wedge p\overline{\alpha})}\right)=\mathcal{O}_p(\Delta m),$$

which is dominated by the spanning error as well. □

## B.7 Proof of Proposition 3

The proof of Pelger (2019)'s Theorem 7 assumes a useful but infeasible central limit theorem, then shows that sample quantities are able to approximate this central limit theorem, and then makes a clever appeal to the delta method. We can skip the middle step, by proving the required CLT directly. We start this approach with the following intermediate result:

**Lemma 8.** Suppose the conditions of Propositions 1 and 2 hold, and $\tau \notin \{\tau_i\}_{i=1}^{k}$, then

$$(\Delta m)^{-1/2} \operatorname{vec}\Big(\begin{bmatrix} \widehat{\ddot{X}}' & \widehat{\ddot{R}}^{\mathbb{Q}}_{h,\tau} \end{bmatrix} - \begin{bmatrix} \ddot{X}'H & \ddot{R}^{\mathbb{Q}}_{h,\tau} \end{bmatrix}\Big) \overset{\mathcal{L}-s}{\to} \mathcal{N}(0, \Pi), \tag{B.1}$$

with

$$\Pi = \begin{bmatrix} K^{(d,T)}(I_T \otimes S^{-2}QB')\mathcal{H}(I_T \otimes S^{-2}QB')'K^{(T,d)} & 0 \\ 0 & M_{\iota_T} \operatorname{diag}\{\mathcal{V}_t(h)\}_{t=1,\ldots,T}\, M'_{\iota_T} \end{bmatrix}, \tag{B.2}$$

with $K$ the commutation matrix such that $K \operatorname{vec}(A) = \operatorname{vec}(A')$.

*Proof.* The CLT for $\widehat{\ddot{X}}$ is proved as an intermediate result in Appendix B.5, the commutation matrix allows us to write a similar result in terms of its transpose. The CLT for $\widehat{\ddot{R}}^{\mathbb{Q}}_{h,\tau}$ follows from Proposition 2 by the time-independence of errors in Assumption 2. The joint result follows from the same assumption as different tenors are used for both series. □

This result can be used to show asymptotic normality of the sample covariances, on which canonical correlations are based.

**Lemma 9.** Suppose the conditions of Lemma 8 hold, then

$$(\Delta m)^{-1/2} \operatorname{vec}\left(\begin{bmatrix} \widehat{\ddot{X}}\widehat{\ddot{X}}' & \widehat{\ddot{X}}\widehat{\ddot{R}}^{\mathbb{Q}}_{h,\tau} \\ \widehat{\ddot{R}}^{\mathbb{Q}\,\prime}_{h,\tau}\widehat{\ddot{X}}' & \widehat{\ddot{R}}^{\mathbb{Q}\,\prime}_{h,\tau}\widehat{\ddot{R}}^{\mathbb{Q}}_{h,\tau} \end{bmatrix} - \begin{bmatrix} H'\ddot{X}\ddot{X}'H & H'\ddot{X}\ddot{R}^{\mathbb{Q}}_{h,\tau} \\ \ddot{R}^{\mathbb{Q}\,\prime}_{h,\tau}\ddot{X}'H & \ddot{R}^{\mathbb{Q}\,\prime}_{h,\tau}\ddot{R}^{\mathbb{Q}}_{h,\tau} \end{bmatrix}\right) \overset{\mathcal{L}-s}{\to} \mathcal{N}(0, D\Pi D'), \tag{B.3}$$

where

$$D = \left(\begin{bmatrix} H'\ddot{X} \\ \ddot{R}^{\mathbb{Q}\,\prime}_{h,\tau} \end{bmatrix} \otimes I_{d+1}\right) K^{(T,d+1)} + \left(I_{d+1} \otimes \begin{bmatrix} H'\ddot{X} \\ \ddot{R}^{\mathbb{Q}\,\prime}_{h,\tau} \end{bmatrix}\right). \tag{B.4}$$

*Proof.* This follows from the delta method using the mapping $\operatorname{vec}(A) \mapsto \operatorname{vec}(A'A)$. For an $(n \times m)$-dimensional matrix $A$, this has the derivative

$$\frac{\partial \operatorname{vec}(A'A)}{\partial \operatorname{vec}(A)'} = (A' \otimes I_m)\frac{\partial \operatorname{vec}(A')}{\partial \operatorname{vec}(A)'} + (I_m \otimes A')\frac{\partial \operatorname{vec}(A)}{\partial \operatorname{vec}(A)'} = (A' \otimes I_m)\,K^{(n,m)} + (I_m \otimes A').$$

□

Important to note is that $D\Pi D'$ is a reduced rank matrix by construction, as it is the covariance matrix of a vectorized *symmetric* matrix. A proper formulation uses the vech-operator, but the difference is immaterial for our purposes.

Recalling the Frisch-Waugh result that $M_{\iota_T}$ takes on the role of the constant, Proposition 3 now follows from Lemma 9 by the delta method, using the mapping

$$\mathrm{vec}\left(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]\right) \mapsto \mathrm{Tr}\left\{a^{-1}bd^{-1}c\right\}.$$

Its derivative, which leads to the expression for $\xi$, is computed in Pelger (2019). The matrix $H$ is irrelevant in the final results, as the canonical correlations are unaffected by this invertible transformation. $\square$

# C Additional simulation results

This section contains additional simulation results.

## C.1 Infeasible versions

We additionally illustrate the results of the rank test with the true simulated errors. The simulation scheme is exactly the same as described in Section 4, but instead of estimating the errors $\widehat{\zeta}_t(\tau, m)$, here we construct the covariance matrix estimator using the true errors $\zeta_t(\tau, m)$.

Tables 5 and 7 contain the results for the rank test for one- and two-factor models, respectively. Tables 6b and 8b provide the results for the restricted versions with $d_{max} = 6$.

Table 5: Monte Carlo results for one-factor model, infeasible version

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q\backslash\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 1.0 | 10.0 | 5.8 | 1.1 | 10.1 | 5.8 | 1.1 | 9.1 | 4.7 | 1.2 | 9.0 | 4.6 | 1.2 |
| 2.0 | 12.1 | 6.1 | 1.2 | 12.4 | 6.5 | 1.1 | 10.2 | 5.0 | 1.0 | 10.2 | 4.9 | 1.0 |
| 3.0 | 50.8 | 50.8 | 50.7 | 9.9 | 4.6 | 0.9 | 50.1 | 50.1 | 50.1 | 11.1 | 5.4 | 0.9 |
| 4.0 | 52.1 | 52.1 | 52.1 | 29.1 | 24.1 | 20.0 | 53.2 | 53.2 | 53.2 | 34.8 | 33.0 | 30.3 |

Table 6: Monte Carlo results for one-factor model, infeasible version

(a) $d_{max} = 6$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q \backslash \alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 1.0 | 10.0 | 5.8 | 1.1 | 10.1 | 5.8 | 1.1 | 9.1 | 4.7 | 1.2 | 9.0 | 4.6 | 1.2 |
| 2.0 | 10.1 | 5.8 | 1.1 | 9.8 | 6.3 | 1.0 | 9.0 | 4.7 | 1.2 | 9.2 | 4.5 | 1.2 |
| 3.0 | 10.2 | 5.9 | 1.2 | 10.6 | 6.1 | 1.4 | 9.1 | 4.6 | 1.2 | 9.4 | 5.2 | 1.1 |
| 4.0 | 10.1 | 5.8 | 1.1 | 12.0 | 6.7 | 1.2 | 9.1 | 4.6 | 1.2 | 10.3 | 5.3 | 1.2 |

(b) $d_{max} = 8$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q \backslash \alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 1.0 | 13.8 | 8.1 | 1.6 | 12.9 | 6.7 | 1.5 | 13.0 | 6.9 | 1.4 | 11.4 | 6.6 | 1.2 |
| 2.0 | 13.8 | 8.1 | 1.6 | 12.9 | 6.7 | 1.5 | 13.0 | 6.9 | 1.4 | 11.4 | 6.6 | 1.2 |
| 3.0 | 13.4 | 8.2 | 1.5 | 13.3 | 7.5 | 2.0 | 12.2 | 6.7 | 1.2 | 12.5 | 7.4 | 1.4 |
| 4.0 | 13.2 | 7.8 | 1.5 | 15.3 | 8.6 | 2.5 | 12.4 | 6.5 | 1.2 | 13.2 | 8.3 | 1.7 |

Note: This table provides Monte Carlo simulation results for the rejection frequencies of the adjusted rank test at a 5% significance level with $d_{max} = 6$ and $d_{max} = 8$, based on 1000 replications from the two-factor model. Each panel lists, the rejection frequencies for different null hypotheses, numbers of argument values, and grid sizes.

Table 7: Monte Carlo results for two-factor model, infeasible version

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q \backslash \alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 1.0 | 9.2 | 4.4 | 0.8 | 9.2 | 4.3 | 0.8 | 8.4 | 2.6 | 0.6 | 8.4 | 2.6 | 0.6 |
| 2.0 | 9.4 | 4.6 | 0.8 | 9.5 | 4.7 | 0.9 | 10.7 | 5.3 | 0.4 | 10.8 | 5.4 | 0.5 |
| 3.0 | 50.3 | 50.3 | 50.2 | 9.5 | 4.5 | 1.1 | 49.4 | 49.4 | 49.3 | 10.7 | 4.7 | 1.2 |
| 4.0 | 50.5 | 50.5 | 50.5 | 11.7 | 5.7 | 1.1 | 50.6 | 50.6 | 50.6 | 13.1 | 7.7 | 2.0 |

Table 8: Monte Carlo results for two-factor model, infeasible version

(a) $d_{max} = 6$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q\backslash\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 1.0 | 9.2 | 4.4 | 0.8 | 9.2 | 4.3 | 0.8 | 8.4 | 2.6 | 0.6 | 8.4 | 2.6 | 0.6 |
| 2.0 | 9.2 | 4.4 | 0.8 | 9.2 | 4.1 | 0.8 | 8.4 | 2.6 | 0.6 | 8.3 | 3.3 | 0.5 |
| 3.0 | 9.3 | 4.3 | 0.8 | 9.6 | 4.9 | 0.9 | 8.4 | 2.6 | 0.6 | 8.9 | 3.7 | 0.5 |
| 4.0 | 9.3 | 4.3 | 0.8 | 10.7 | 5.6 | 1.2 | 8.4 | 2.6 | 0.6 | 9.5 | 4.9 | 0.6 |

(b) $d_{max} = 8$

| | $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L = 0.2$ | | | $L = 1.0$ | | | $L = 0.2$ | | | $L = 1.0$ | | |
| $q\backslash\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 1.0 | 12.2 | 5.0 | 1.1 | 10.0 | 4.1 | 1.1 | 11.5 | 5.1 | 1.1 | 10.5 | 4.7 | 1.0 |
| 2.0 | 12.2 | 5.0 | 1.1 | 10.0 | 4.1 | 1.1 | 11.5 | 5.1 | 1.1 | 10.5 | 4.7 | 1.0 |
| 3.0 | 11.0 | 4.4 | 0.9 | 11.4 | 5.3 | 1.2 | 11.2 | 4.8 | 1.2 | 11.7 | 5.5 | 1.4 |
| 4.0 | 10.4 | 4.2 | 0.8 | 16.0 | 7.7 | 1.7 | 10.9 | 4.7 | 1.1 | 14.6 | 6.6 | 1.4 |

Note: This table provides Monte Carlo simulation results for the rejection frequencies of the adjusted rank test at a 5% significance level with $d_{max} = 6$ and $d_{max} = 8$, based on 1000 replications from the two-factor model. Each panel lists, the rejection frequencies for different null hypotheses, numbers of argument values, and grid sizes.

## References


Ahn, S. C. and A. R. Horenstein (2013): "Eigenvalue ratio test for the number of factors," *Econometrica*, 81, 1203–1227.

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. A. Laeven (2015): "Modeling financial contagion using mutually exciting jump processes," *Journal of Financial Economics*, 117, 585–606.

Aït-Sahalia, Y., C. Li, and C. X. Li (2021a): "Closed-form implied volatility surfaces for stochastic volatility models with jumps," *Journal of Econometrics*, 222, 364–392.

——— (2021b): "Implied stochastic volatility models," *The Review of Financial Studies*, 34, 394–450.

Andersen, T. G., N. Fusari, and V. Todorov (2015a): "Parametric Inference and Dynamic State Recovery From Option Panels," *Econometrica*, 83, 1081–1145.

——— (2015b): "The risk premia embedded in index options," *Journal of Financial Economics*, 117, 558–584.

——— (2017): "Short-term market risks implied by weekly options," *The Journal of Finance*, 72, 1335–1386.

Andersen, T. G., N. Fusari, V. Todorov, and R. T. Varneskov (2021): "Spatial dependence in option observation errors," *Econometric Theory*, 37, 205–247.

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2019): "Inference in Group Factor Models With an Application to Mixed-Frequency Data," *Econometrica*, 87, 1267–1305.

Bai, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

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

——— (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131, 507–537.

——— (2020): "Simpler proofs for approximate factor models of large dimensions," *arXiv preprint arXiv:2008.00254*.

Bakshi, G., N. Kapadia, and D. Madan (2003): "Stock return characteristics, skew laws, and the differential pricing of individual equity options," *The Review of Financial Studies*, 16, 101–143.

Bates, D. S. (1996): "Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options," *The Review of Financial Studies*, 9, 69–107.

——— (2000): "Post-'87 crash fears in the S&P 500 futures option market," *Journal of Econometrics*, 94, 181–238.

Billingsley, P. (1968): *Convergence of Probability Measures*, Wiley, New York.

Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy*, 81, 637–654.

Boswijk, H. P., R. J. A. Laeven, A. Lalu, and E. Vladimirov (2021): "Jump Contagion among Stock Market Indices: Evidence from Option Markets," *Tinbergen Institute Discussion Paper 2021-086/III.*

Boswijk, H. P., R. J. A. Laeven, and E. Vladimirov (2022): "Estimating Option Pricing Models Using a Characteristic Function-Based Linear State Space Representation," *Tinbergen Institute Discussion Paper 2022-075/III.*

Carr, P. and D. Madan (2001): "Optimal positioning in derivative securities," *Quantitative Finance*, 1, 19.

Carr, P. and L. Wu (2020): "Option profit and loss attribution and pricing: A new framework," *The Journal of Finance*, 75, 2271–2316.

Christoffersen, P., S. Heston, and K. Jacobs (2009): "The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well," *Management Science*, 55, 1914–1932.

Cont, R. and J. Da Fonseca (2002): "Dynamics of implied volatility surfaces," *Quantitative Finance*, 2, 45.

Duffie, D., D. Filipović, and W. Schachermayer (2003): "Affine Processes and Applications in Finance," *The Annals of Applied Probability*, 13, 984–1053.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-diffusions," *Econometrica*, 68, 1343–1376.

Fang, F. and C. W. Oosterlee (2009): "A novel pricing method for European options based on Fourier-cosine series expansions," *SIAM Journal on Scientific Computing*, 31, 826–848.

Goeman, J. J. and A. Solari (2010): "The sequential rejection principle of familywise error control," *The Annals of Statistics*, 3782–3810.

Heston, S. L. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *The Review of Financial Studies*, 6, 327–343.

Horn, R. A. and C. R. Johnson (1991): *Topics in Matrix Analysis*, Cambridge University Press.

Hull, J. and A. White (1987): "The pricing of options on assets with stochastic volatilities," *The Journal of Finance*, 42, 281–300.

Johansen, S. (1988): "Statistical analysis of cointegration vectors," *Journal of economic dynamics and control*, 12, 231–254.

Kleibergen, F. and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133, 97–126.

Lee, R. W. (2004): "The moment formula for implied volatility at extreme strikes," *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 14, 469–480.

Merton, R. C. (1976): "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 3, 125–144.

Onatski, A. (2010): "Determining the number of factors from empirical distribution of eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016.

Pan, J. (2002): "The jump-risk premia implicit in options: evidence from an integrated time-series study," *Journal of Financial Economics*, 63, 3–50.

Pelger, M. (2019): "Large-dimensional factor modeling based on high-frequency observations," *Journal of Econometrics*, 208, 23–42.

Robin, J.-M. and R. J. Smith (2000): "Tests of rank," *Econometric Theory*, 16, 151–175.

Singleton, K. J. (2006): *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*, Princeton University Press.

Skiadopoulos, G., S. Hodges, and L. Clewlow (2000): "The dynamics of the S&P 500 implied volatility surface," *Review of Derivatives Research*, 3, 263–282.

Todorov, V. (2019): "Nonparametric spot volatility from options," *The Annals of Applied Probability*, 29, 3590–3636.

——— (2021): "Higher-order small time asymptotic expansion of Itô semimartingale characteristic function with application to estimation of leverage from options," *Stochastic Processes and their Applications*, 142, 671–705.

Vladimirov, E. (2023): "iCOS: Option-Implied COS method," *Working Paper*.