Estimating Volatility Risk Factors Using Large Panels of Filtered or Realized Volatilities[*]

Elena Andreou[†]   **Eric Ghysels**[‡]

First Draft: July 2013

This Draft: September 29, 2013

Preliminary and Incomplete

Abstract

The paper considers estimating volatility risk factors using large panels of filtered or realized volatilities. The data structure involves three types of asymptotic expansions. There is the cross-section of volatilities estimates at each point in time, namely $i = 1, \ldots, N$ observed at dates $t = 1, \ldots, T$. In addition to expanding N and T, we also have the sampling frequency h of the data used to compute the volatility estimates where we rely on data collected at increasing frequency, $h \downarrow 0$. The continuous record or in-fill asymptotics ($h \downarrow 0$) allows us to control the cross-sectional and serial correlation among the idiosyncratic errors of the panel. A remarkable result emerges. Under suitable regularity conditions the traditional principal component analysis yields super-consistent estimates of the factors at each point in time. Namely, contrary to the root-N standard normal consistency we find N-consistency, also standard normal, due to the fact that the high frequency sampling scheme is tied to the size of the cross-section, boosting the rate of convergence.

1   Introduction

We consider situations where volatility is driven by a small set of common risk factors. Various theoretical asset pricing models, particularly the affine class of asset pricing models, yield such an environment. In this paper we consider the task of extracting the volatility risk factors from panels of filtered volatilities using ARCH-type models or realized volatilities. In case of the former, we know that ARCH models potentially feature specification errors, yet as argued by Nelson (1990), Nelson and Foster (1994), Nelson (1996), among others, they can be viewed as *filters* and deliver reliable estimates of spot volatility despite their mis-specification. To obtain such results we need some continuous record asymptotic arguments about the samples used to estimate the volatilities. On top of this, we need both large cross-sections (of dimension N) of such volatilities and we need a reasonably long time series (of length T). This entails some challenges regarding the large sample analysis used to estimate the factors, as the ARCH-type filters add one additional limiting argument - besides the large N and large T asymptotics of Stock and Watson (2002), Bai and Ng (2002), Bai (2003), among others. The data structure we have in mind involves three types of asymptotics. There is the cross-section of volatilities estimates at each point in time, namely $i = 1, \ldots, N$ observed at dates $t = 1, \ldots, T$. In addition to the expanding N and T, we also have the sampling frequency h used to compute volatility estimates or realized volatilities with data collected at increasingly finer frequency, $h \downarrow 0$. The continuous record (for ARCH filters) or in-fill (for realized volatilities) asymptotics with $h \downarrow 0$ allow us to control the cross-sectional and serial correlation among the idiosyncratic errors of the panel. Intuitively, a judicious choice of h will prevent the errors from becoming a pervasive factor in the panel. Hence, we create a well-behaved panel data set by controlling the sampling scheme used to construct filtered volatilities. A remarkable result emerges. Under suitable regularity conditions the traditional principal component analysis yields super-consistent estimates of the factor at each point in time. Namely, contrary to the root-N standard normal consistency we find N-consistency, also standard normal, due to the fact that the high frequency sampling scheme is tied to the size of the cross-section, boosting the rate of convergence.

The use of ARCH-type filters excludes jumps in the data generating process. We therefore also apply similar arguments to large panels of realized volatilities which can accommodate the presence of jumps. The latter is an idea explored notably by Anderson and Vahid (2007) and Barigozzi, Brownlees, Gallo, and Veredas (2011), although neither develop a comprehensive asymptotic theory, which is provided in this paper. In the case of realized measures, we consider various estimators, such as the realized variance estimators proposed by Barndorff-Nielsen and Shephard (2002b), Jacod and Protter (1998) and Zhang (2001a), the bi-power measures of Barndorff-Nielsen and Shephard (2004b) and Barndorff-Nielsen and Shephard (2004a), and the various measures which account for the presence of microstructure noise, see e.g. Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Gloter and Jacod (2000), among others. While the various realized measure estimators feature different convergence rates (involving different population objects) they yield similar asymptotic results for the estimation of the volatility factors, albeit for suitably tailored linkages between N and h, i.e. the expansion of the cross-

section and the high frequency data sampling scheme. This means that large panels of filtered or realized volatilities yield super-consistent estimation of volatility factors, where super-consistency refers to the rate at which the volatility factors can be recovered every period t in the panel.

In a first section of the paper we provide the asset pricing context for the volatility risk factor specification. The setting is one of a continuous time affine jump diffusion. The second section covers ARCH filters, the third the realized measure cases. Section four contains simulation evidence and section five concludes the paper.

2 Affine Asset Pricing Models

We start with the widely used class of continuous time affine jump diffusion (henceforth AJD) asset pricing models. To fix notation, we follow the presentation of Duffie, Pan, and Singleton (2000) and consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ where the filtration satisfies the usual conditions (see e.g. Protter (2004)) and $\mathbb{P}$ refers to the physical or historical probably measure. Moreover, we suppose that the k-dimensional $\mathcal{F}$-adapted process $\mathcal{X}^f$ of state variables or factors is Markov in some state space $D \subset \mathbb{R}^k$, solving the stochastic differential equation:

$$d\mathcal{X}_t^f = \mu^{\mathbb{P}}(\mathcal{X}_t^f)dt + \sigma(\mathcal{X}_t^f)dW_t^{\mathbb{P}} + dZ_t^{\mathbb{P}} \tag{2.1}$$

where $W_t^{\mathbb{P}}$ is an $\mathcal{F}_t$-adapted Brownian motion under $\mathbb{P}$ in $\mathbb{R}^k$, $\mu^{\mathbb{P}} : D \to \mathbb{R}^k$, $\sigma : D \to \mathbb{R}^{k \times k}$, and $Z_t^{\mathbb{P}}$ is a pure jump process whose jumps have a fixed probability distribution $\nu^{\mathbb{P}}$ on $\mathbb{R}^k$ with arrival intensity $\lambda^{\mathbb{P}}(\mathcal{X}_t^f) : t \geq 0$, for some $\lambda^{\mathbb{P}} : D \to [0, \infty)$.

Some of our analysis will require continuous path processes, i.e. exclude jumps. We therefore formulate two assumptions for the characterization of the state variable processes, namely:

Assumption 2.1 (AJD). *The distribution of $\mathcal{X}^f$, given an initial known $\mathcal{X}_0^f$ at $t = 0$, is completely characterized by a quadruplet $(K^{\mathbb{P}}, H, l^{\mathbb{P}}, \theta^{\mathbb{P}})$ of parameters determining the affine functions:*

$$
\begin{array}{rclcl}
\mu^{\mathbb{P}}(x) & = & K_0^{\mathbb{P}} + K_1^{\mathbb{P}}x, & K^{\mathbb{P}} \equiv (K_0^{\mathbb{P}}, K_1^{\mathbb{P}}) \in \mathbb{R}^k \times \mathbb{R}^{k \times k} \\
(\sigma(x)\sigma(x)^{\boldsymbol{T}})_{ij} & = & (H_0)_{ij} + (H_1)_{ij}^{\boldsymbol{T}}x & H \equiv (H_0, H_1) \in \mathbb{R}^{k \times k} \times \mathbb{R}^{k \times k \times k} \\
\lambda^{\mathbb{P}}(x) & = & l_0^{\mathbb{P}} + l_1^{\mathbb{P}}x, & l^{\mathbb{P}} \equiv (l_0^{\mathbb{P}}, l_1^{\mathbb{P}}) \in \mathbb{R} \times \mathbb{R}^k
\end{array}
\tag{2.2}
$$

and where the jump transform $\theta^{\mathbb{P}}$ determines the jump-size distribution, i.e. for $c \in \mathbb{C}^k$, the set of k-tuples of complex numbers, we let $\theta^{\mathbb{P}}(c) \equiv \int_{\mathbb{C}^k} \exp(cz)d\nu^{\mathbb{P}}(z)$ whenever the integral is well defined.

We also consider the more restricted class where either jumps do not occur and/or the jump size distribution is degenerate with unit mass at zero, yielding:

2

Assumption 2.2 (AD). *The $\mathcal{X}^f$ in equation (2.1) excludes the Z_t component and therefore, given an initial known $\mathcal{X}_0^f$ at $t = 0$, is completely characterized by the pair $(K^{\mathbb{P}}, H)$ of parameters determining the affine functions specified in Assumption 2.1.*

Finally, we also assume absence of arbitrage, which implies the existence of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ where $\mathbb{Q}$ is an equivalent martingale measure under the risk neutral world. Furthermore, we assume that $\mathcal{X}^f$ is also an AJD under $\mathbb{Q}$, and therefore:

Assumption 2.3 (AJD/Q). *Under the risk-neutral equivalent martingale measure $\mathbb{Q}$ the process X is described by:*

$$d\mathcal{X}_t^f = \mu^{\mathbb{Q}}(\mathcal{X}_t^f)dt + \sigma(\mathcal{X}_t^f)dW_t^{\mathbb{Q}} + dZ_t^{\mathbb{Q}} \tag{2.3}$$

completely characterized by a quadruplet $(K^{\mathbb{Q}}, H, l^{\mathbb{Q}}, \theta^{\mathbb{Q}})$ of parameters determining the affine functions:

$$
\begin{aligned}
\mu^{\mathbb{Q}}(x) &= K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}}x, & K^{\mathbb{Q}} &\equiv (K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}) \in \mathbb{R}^k \times \mathbb{R}^{k\times k} \\
(\sigma(x)\sigma(x)^{\mathbf{T}})_{ij} &= (H_0)_{ij} + (H_1)_{ij}^{\mathbf{T}}x, & H &\equiv (H_0, H_1) \in \mathbb{R}^{k\times k} \times \mathbb{R}^{k\times k\times k} \\
\lambda^{\mathbb{Q}}(x) &= l_0^{\mathbb{Q}} + l_1^{\mathbb{Q}}x, & l^{\mathbb{Q}} &\equiv (l_0^{\mathbb{Q}}, l_1^{\mathbb{Q}}) \in \mathbb{R} \times \mathbb{R}^k
\end{aligned}
\tag{2.4}
$$

and the analogue jump transform $\theta^{\mathbb{Q}}$.

Finally, we also impose a set of technical regularity conditions, namely:

Assumption 2.4. *The paths of $\sigma(\mathcal{X}_t^f)$ are locally bounded, i.e. with probability 1 the paths are of bounded variation on any compact subinterval of $[0, \infty)$, and are bounded away from zero under both probability measures. the affine jump diffusion satisfies:*

The main advantage in choosing the AJD stochastic environment lies in the fact that the Laplace transform of $\mathcal{X}_t^f$ will be exponentially affine. Under technical regularity conditions (see Duffie, Pan, and Singleton (2000) and next section), and given Assumptions 2.1 and 2.3 we have under either measure:

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}(\exp{(-\omega^{\mathbf{T}}\mathcal{X}_{t+\tau}^f)}|\mathcal{F}_t) \equiv \mathbb{E}_t^{\mathbb{P}}(\exp{(-\omega^{\mathbf{T}}\mathcal{X}_{t+\tau}^f)}) &= \exp{(-\alpha^{\mathbb{P}}(\tau, \omega)^{\mathbf{T}}\mathcal{X}_t^f + \beta^{\mathbb{P}}(\tau, \omega))} \\
\mathbb{E}^{\mathbb{Q}}(\exp{(-\omega^{\mathbf{T}}\mathcal{X}_{t+\tau}^f)}|\mathcal{F}_t) \equiv \mathbb{E}_t^{\mathbb{Q}}(\exp{(-\omega^{\mathbf{T}}\mathcal{X}_{t+\tau}^f)}) &= \exp{(-\alpha^{\mathbb{Q}}(\tau, \omega)^{\mathbf{T}}\mathcal{X}_t^f + \beta^{\mathbb{Q}}(\tau, \omega))}
\end{aligned}
\tag{2.5}
$$

where the coefficients (under both $\mathbb{P}$ and $\mathbb{Q}$) $\alpha(t, \omega) \in \mathbb{R}^n$ and $\beta(t, \omega) \in \mathbb{R}$ satisfy a Riccati ODE system (with $\mathbb{P}$ and $\mathbb{Q}$ superscripts for α, β, θ, K, and l omitted for simplicity):

$$
\begin{aligned}
\partial\alpha/\partial t &= (K_1)^{\mathbf{T}}\alpha - \tfrac{1}{2}(\alpha)^{\mathbf{T}}H_1\alpha - l_1(\theta(\alpha) - 1) \\
\partial\beta/\partial t &= (K_0)^{\mathbf{T}}\alpha - \tfrac{1}{2}(\alpha)^{\mathbf{T}}H_0\alpha - l_0(\theta(\alpha) - 1)
\end{aligned}
\tag{2.6}
$$

with boundary conditions (again with $\mathbb{P}$ and $\mathbb{Q}$ superscripts omitted for simplicity) $\alpha(0,\omega) = \omega$ and $\beta(0,\omega)$ $= 0$.

Again, following Duffie, Pan, and Singleton (2000), one can similarly derive characteristic functions as well as cumulant-generating functions, which allow us to compute conditional moments both under $\mathbb{P}$ and $\mathbb{Q}$.

In a generic AJD no-arbitrage asset price setting, Duffie, Pan, and Singleton (2000) show that the bond, equity and variance premia at different investment horizons are linear functions of the same risk factors - i.e. state variables $\mathcal{X}_t^f$. In the term structure literature it is common to rotate the factors such that they correspond to the commonly used level, slope and curvature factors. The equity premia literature has instead focused on factors driven by macroeconomic fundamentals, in particular consumption uncertainty in the context of long-run risk economies studied by Bansal and Yaron (2004), where agents have a preference for early resolution of uncertainty and therefore dislike increases in economic uncertainty.[1]

While we will consider a broad asset class, we start with equity returns. We start with a reduced form for expected excess log returns (on the market portfolio):

$$E_t^{\mathbb{P}}[r_{t,t+\tau}] = \gamma_{er}(\tau)\mathcal{X}_t^f, \qquad (2.7)$$

Following Britten-Jones and Neuberger (2000), Jiang and Tian (2005), and Carr and Wu (2009), we define the variance risk premium (VRP) as the difference between expected variance under the risk-neutral measure and expected variance under the objective measure.

The VRP has been studied extensively as it relates to variance swap contracts for which there is an active market particularly pertaining to the S&P 500 stock market index. One leg of the swap will pay an amount based upon the realized variance of the price changes of the underlying. Conventionally, these price changes will be daily log returns. The other leg of the swap will pay a fixed amount, which is the strike, quoted at the deal's inception. Thus the net payoff to the counter-parties will be the difference between these two and will be settled in cash at the expiration. Hence, at maturity, the payoff to the long side of the swap is equal to the difference between the realized variance over the life of the contract and a constant called the variance swap rate. No arbitrage dictates that the variance swap rate equals the risk-neutral expected value of the realized variance, i.e. $E_t^{\mathbb{Q}}[V_{t,t+\tau}^r]$, $V_{t,t+\tau}^r$ is the equity return forward integrated variance over the time interval t to $t + \tau$. The VRP is the difference between the time t expected equity returns variance under the historical ($\mathbb{P}$) and the risk-neutral ($\mathbb{Q}$) probability measures, over horizon τ can be written as:

$$VRP(t,\tau) = E_t^{\mathbb{Q}}[V_{t,t+\tau}^r] - E_t^{\mathbb{P}}[V_{t,t+\tau}^r] \qquad (2.8)$$

---

[1] See in particular Eraker and Shaliastovich (2008) for the linear pricing characterization of long-run risk equilibrium models.

With the linear affine risk premia setting, one obtains respectively:

$$VRP(t,\tau) = \delta_{vrp}(\tau) + \gamma_{vrp}(\tau)\mathcal{X}_t^f, \tag{2.9}$$

where δ_{vrp} and γ_{vrp} relate to the data generating process, i.e. relate to α^i and β^i for $i = \mathbb{Q}$ and $\mathbb{P}$ (see for example Todorov (2010), among others, for further details). Similarly, the pricing of future integrated volatility can be written as:

$$E_t^{\mathbb{Q}}[V_{t,t+\tau}^r] = \mu_{rv}^{\mathbb{Q}}(\tau) + \gamma_{rv}^{\mathbb{Q}}(\tau)\mathcal{X}_t^f, \qquad E_t^{\mathbb{P}}[V_{t,t+\tau}^r] = \mu_{rv}^{\mathbb{P}}(\tau) + \gamma_{rv}^{\mathbb{P}}(\tau)\mathcal{X}_t^f \tag{2.10}$$

The above linear pricing schemes have been used extensively, although with assumed factor representations - and their interpretation - that are different. See notably Carr and Wu (2009), Egloff, Leippold, and Wu (2010), Zhou (2010), Bakshi, Panayotov, and Skoulakis (2011), Mueller, Vedolin, and Zhou (2011), Wang, Zhou, and Zhou (2013) and Corradi, Distaso, and Mele (2013), among others. In the next section, we present a novel approach of estimating $\mathcal{X}_t^f$.

By analogy with the term structure of interest rates, one can estimate the principal components for a panel of variance risk premia with different maturities. Bühler (2006), Amengual (2009), Egloff, Leippold, and Wu (2010) and Aït-Sahalia, Karaman, and Mancini (2012) apply principal component analysis (PCA) to panels of variance swap rates and find that two factors - which can be interpreted as level and slope - explain close to 100 % of the variation in variance swap rates for the S&P 500. The similarities with term structure of interest rates also brings us to the topic of stochastic singularity - since we have potentially large cross-sections of variance swap rates driven - according to equation (2.9 - by just a few state variables. Adding measurement errors to the variance swap rates is the standard solution. The analysis in the next section will not require us to add an ad hoc measurement error as filtered volatilities have a natural error distribution which we know how to characterize asymptotically.[2]

3   Factor Analysis with Panels of ARCH Filters

Technically speaking we will consider for any asset i with (log) price x_t^i which has exposure to (some of) the risk factors, i.e.:

$$x_t^i \equiv \delta_0^i + \delta^i \mathcal{X}_t^f \qquad \delta^i \neq 0 \tag{3.1}$$

such as for example the log price of an equity claim, log price of a zero-coupon bond, a risk spread, etc., and we are interested in two objects: (1) spot volatility and (2) integrated volatility.

___
[2]Todorov (2010), Bollerslev and Todorov (2011), Aït-Sahalia, Karaman, and Mancini (2012) find that a large and time-varying jump risk component is embedded in variance swap rates. Regularity conditions underlying some of the factor filters discussed in the next section will exclude jump risk, the reason why we need the rely on the more restrictive Assumption 2.2.

We limit our attention to continuous path processes in this section, hence operate under Assumption 2.2. In particular, combining equation equations (2.1) and (3.2) implies that x_t^i satisfies the diffusion:

$$dx_t^i = \mu^i(\mathcal{X}_t^f)dt + \sigma^i(\mathcal{X}_t^f)dW_t^{\mathbb{P}} \tag{3.2}$$

where $\mathcal{X}_t^f$ satisfies Assumption 2.2, i.e. excluding jumps, and therefore by Itô's lemma: $\mu^i(\mathcal{X}_t^f) \equiv \delta^i \mu^{\mathbb{P}}(\mathcal{X}_t^f)$ and $\sigma^i(\mathcal{X}_t^f) \equiv \delta^i \sigma(\mathcal{X}_t^f)$.

We do not directly observe volatility, integrated or spot, (nor $\mathcal{X}_t^f$) but have at our disposable, for a large set of assets, some estimates for either type of volatility. We will start with the spot volatility case and consider integrated volatilities in the next section. We are working under the assumption that we can collect volatility data which span the space all the risk factors, formally defined later. Alternatively, we can think of estimating a sub-block of factors pertaining to volatility, which we still will denote by $\mathcal{X}_t^f$ to avoid further complicating notation.

When spot volatility is latent it is necessary to think about filtering. There are an abundant number of filtering schemes, many based on Markov Chain Monte Carlo simulation algorithms, inspired by the seminal work of Jacquier, Polson, and Rossi (2002). Such filtering schemes are analytically intractable and unattractive when we are dealing with potentially large cross-sections, like hundreds of individual asset volatilities. We therefore need to rely on computationally simple filters, convenient and analytically tractable. We therefore opt for ARCH-type models as *filters*, following the work of Nelson (1990), Nelson and Foster (1994), Nelson (1996), among others. Hence, we will rely on easy to estimate (univariate) ARCH-type models viewed as a filters through which one passes the data to produce an estimate of the conditional variance.

We lack continuous time observations x_t^i for asset i but have observations, denoted by $x_{[h:kh]}^i$ with k integer, at arbitrary small time intervals equally spaced by h_i and such data can be collected at an ever increasing frequency, $h_i \downarrow 0 \, \forall \, i$. Suppose we denote the filtered volatility by $\hat{V}_{[h:t]}^i$ and we define:

$$\begin{aligned} \hat{V}_{[h:t]}^i &\equiv \sigma^i(\mathcal{X}_t^f)^2 + \hat{\varepsilon}_{[h:t]}^i \\ &\equiv (\delta^i)^2 \sigma(\mathcal{X}_t^f)^2 + \hat{\varepsilon}_{[h:t]}^i \end{aligned} \tag{3.3}$$

where $\hat{\varepsilon}_{[h:t]}^i$ is a filtering error, i.e. the difference between the true spot volatility and the one obtained from the ARCH-filter. Nelson and Foster (1994) use continuous record asymptotics, i.e. using asset log asset price data at arbitrary small time intervals, denoted by $x_{[h:t]}^i$ for asset i in the cross-section, to characterize the distribution of the measurement error.

Suppose for the moment that $\mathcal{X}^f$ is a bivariate process and without loss of generality consists of two mutually
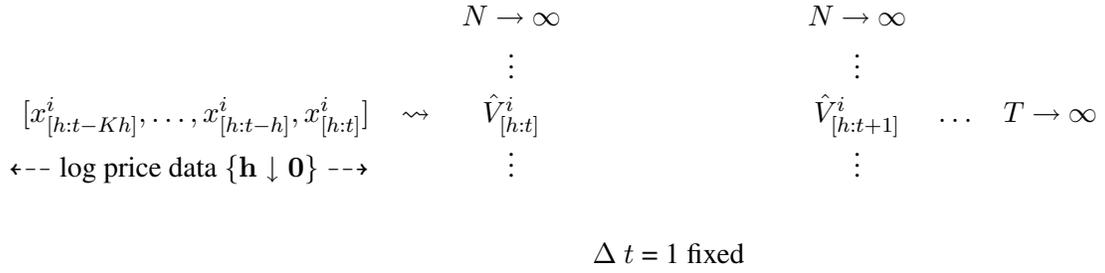
orthogonal processes, call them $\mathcal{X}^1$ and $\mathcal{X}^2$. Then we can rewrite (3.3) as:

$$
\begin{aligned}
\hat{V}^i_{[h:t]} &= (\delta^i_1)^2 \sigma(\bar{\mathcal{X}}^1_t)^2 + (\delta^i_2)^2 \sigma(\bar{\mathcal{X}}^2_t)^2 + \hat{\varepsilon}^i_{[h:t]} && (3.4) \\
&= (\delta^i_1)^2 F_{1t} + (\delta^i_2)^2 F_{2t} + \hat{\varepsilon}^i_{[h:t]} \quad \text{(say)} && (3.5)
\end{aligned}
$$

and therefore we are looking at a two-factor spot volatility model. It will be important to make this distinction with the integrated volatility factor representations we will study in the next section.

The data structure we have in mind involves three types of asymptotics. There is the cross-section of volatilities estimates at each point in time, namely $i = 1, \ldots, N$ observed at dates $t = 1, \ldots, T$. This setup is reminiscent of the large N and large T asymptotics for panel data models of Stock and Watson (2002), Bai and Ng (2002), Bai (2003), among others. In addition to the expanding N and T, we also have the sampling frequency h of data used to compute the volatility estimates using data collected at increasing frequency, $h \downarrow 0$, where $h \equiv \sup_{i=1,N} h_i$. The sampling scheme we therefore have in mind appears in the following diagram:

$$
\begin{array}{ccccccc}
& & N \to \infty & & N \to \infty & & \\
& & \vdots & & \vdots & & \\
[x^i_{[h:t-Kh]}, \ldots, x^i_{[h:t-h]}, x^i_{[h:t]}] & \rightsquigarrow & \hat{V}^i_{[h:t]} & & \hat{V}^i_{[h:t+1]} & \cdots & T \to \infty \\
\text{←-- log price data } \{\mathbf{h} \downarrow \mathbf{0}\} \text{ --→} & & \vdots & & \vdots & &
\end{array}
$$

$$\Delta\, t = 1 \text{ fixed}$$

Practically, we want to use principle component analysis (PCA) with panels of filtered volatilities to obtain estimates of $\mathcal{X}^f_t$. Since both N and T are allowed to grow and h to shrink, we need to prevent filtering errors to become a pervasive factor. The analysis in this section explains how to do this.

The time series of cross-sections is sampled at a fixed time interval, say $\Delta t = 1$. This means that we observe at best a discretization of $\mathcal{X}^f_t$, call it $\bar{\mathcal{X}}^f_t, \equiv \mathcal{X}^f_t$ for $t = k\Delta t\ k \in \mathbb{N}$. This may leave the impression that the underlying affine diffusion setting is detached from the large panel framework, and therefore irrelevant, for the purpose of our analysis, since in general there is no straightforward mapping from the continuous time process to a finite time grid discretization, except in a few special cases. However, the continuous time process is relevant because it provides: (a) the foundations for the volatility filters and their relationship to $\bar{\mathcal{X}}^f_t$, and (b) the stochastic properties of the idiosyncratic (i.e. measurement) errors in the panel data model, such that they satisfy the so called approximate (using the terminology of Chamberlain and Rothschild (1983)) panel structure.

We concentrate on univariate filters featuring the following updating scheme:

$$
\begin{aligned}
\hat{V}^i_{[h:t]} &= \hat{V}^i_{[h:t-h]} + h\hat{\kappa}^i_{t-h} + h^{1/2}g^i_t \\
\hat{\kappa}^i_{t-h} &\equiv \hat{\kappa}^i\left(x^i_{[h:t-h]}, \hat{V}^i_{[h:t-h]}, t, h\right) \\
g^i_t &\equiv g^i\left(\hat{\xi}^i_{x,t}, x^i_{[h:t-h]}, \hat{V}^i_{[h:t-h]}, t, h\right) \\
\hat{\mu}^i_{t-h} &\equiv \hat{\mu}^i(x^i_{[h:t-h]}, \hat{V}^i_{[h:t-h]}, t, h) \\
\hat{\xi}^i_{x,t} &\equiv h^{1/2}[x^i_{[h:t]} - x^i_{[h:t-h]} - h\hat{\mu}^i_{t-h}]
\end{aligned}
\tag{3.6}
$$

where $\hat{\kappa}^i$, $\hat{\xi}^i_{x,t}$, and g^i_t are (definitely misspecified) functions selected by the econometrician (see for example Corradi (2000) for a discussion of the weak GARCH(1,1) case). Moreover, $\hat{\kappa}^i$, $\hat{\xi}^i_{x,t}$, and g^i_t satisfy regularity conditions appearing in Assumption A.2 of Appendix A.1. Note that in principle our analysis extends to multivariate volatility filters - and therefore panels of both conditional variance as well as covariance filters. While in principle feasible - since Nelson (1996) establishes properties for filter errors which mirror the univariate ones in terms of asymptotic convergence rates - we prefer to focus our analysis on a collection of univariate volatility filters. In any practical application this will be to most convenient, computationally straightforward and appealing approach. Finally, it should also be noted that Nelson and Foster (1994) discuss at length how to design optimal filters, meaning in particular choices of g^i_t to minimize mean squared extraction errors. An econometrician could therefore opt for optimal design of the filters. However, the analysis in the rest of the paper does not depend on optimal filtering.

The idea that we estimate, say GARCH(1,1) models for individual assets may raise the question that we may have to deal with near-unit root type of models, i.e. IGARCH, and that this may adversely affect the PCA we about to discuss next. It is important to note, however, that we only skip-sample $\hat{V}^i_{[h:t]}$ at intervals Δt which are far apart in terms of the local time scale determined by h which is used to compute the filtered volatilities. Hence, as h shrinks, so does the serial dependence and among two subsequent panel data observations, i.e. $\hat{V}^i_{[h:t-1]}$ and $\hat{V}^i_{[h:t]}$ for any t. This issue will be further highlighted when we discuss the idiosyncratic error structure of the panel data models.

Having characterized both $\hat{V}^i_{[h:t]}$ and V^i_t, and therefore the filtering error $\hat{\varepsilon}^i_{[h:t]}$, we are ready to consider the statistical properties of principal component analysis. The typical route, taken by Stock and Watson (2002), Bai and Ng (2002), Bai (2003), among others, is to: (a) make assumptions about the common factors (existence of moments, etc.), (b) limit the cross-sectional and time series dependence across the idiosyncratic errors, (c) impose independence (or some weaker assumption) between the factors and the idiosyncratic errors, and finally (d) put restrictions on the behavior of the factor loadings. While we will obviously borrow some of the standard assumptions from the large N and T asymptotic analysis of panel data, we will need to deviate due to the specific nature of the (filtering) error process. In particular, the properties of the latter will be controlled by what we do with sampling high frequency data, i.e. how h shrinks. In the remainder of this section we discuss the regularity conditions which will guarantee that we

obtain the desired asymptotic properties of the principal component estimator, by desired we mean the same properties as those derived for large N and T.

To keep our analysis as close as possible to the standard large scale factor models in the literature, we adopt the commonly used notation with some modification and then discuss the mapping with the framework discussed so far. Namely, consider the vector form model representation for $X_{[h:t]} = (\hat{V}_{[h:t]}^i, i = 1, \ldots, N)'$:

$$X_{[h:t]} = \Lambda F_t + e_{[h:t]} \tag{3.7}$$

where the commonly used normalization $E(F_t) = 0$ and $E(F_t F_t')$ is a k-dimensional diagonal matrix ($k = \dim(\bar{\mathcal{X}}_t^f)$) is assumed. The latter implies that the factors we will uncover are some affine transformation: $F_t = \mathcal{H}\bar{\mathcal{X}}_t^f$, with $\mathcal{H}$ a nonsingular $k \times k$ matrix. Term structure applications of affine models in particular, involve anchoring factors to observable series - most notably the so called level, slope and curvature factors (see e.g. Dai and Singleton (2000) for a detailed discussion). In our analysis, there is no obvious way to calibrate the level and scale of F_t, and therefore we use a standard normalization. Moreover, we let the factor loadings $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ are non-random and, (b) $e_{[h:t]} = (\hat{\varepsilon}_{[h:t]}^i, i = 1, \ldots, N)'$. The matrix representation of the factor model is:

$$X_{[h]} = F\Lambda' + e_{[h]} \tag{3.8}$$

where $X_{[h]} = ((X_{[h:.]}^1)', \ldots, (X_{[h:.]}^N)')$ is a $T \times N$ matrix of observations on demeaned volatilities and $e_{[h]} = ((e_{[h:1]})', \ldots, (e_{[h:N]})')$ a $T \times N$ matrix of idiosyncratic errors. Henceforth, to simplify notation we will denote the individual elements as $x_{[h:t]}^i$, and $e_{[h:t]}^i$.

Our analysis shares some basic assumptions regarding F, but also differs substantially in terms of key assumptions regarding $e_{[h]}$. We devote our attention first to the latter. A typical assumption is that F and $e_{[h]}$ are mutually independent (with some notable exceptions - see e.g. Bai (2003, Assumption D)). In our setting they are not as $\hat{\varepsilon}_{[h:t]}^i$ features heteroskedasticity which depends on $\mathcal{X}_t^f$ (namely through the terms $B^i(x, y, t)$ and $C^i(x, y, t)$ appearing in (A.2), see Nelson and Foster (1994, Theorem 3.1)). Moreover, in the context of standard large scale factor models, some basic assumptions are also made about the time series and cross-section dependence of the idiosyncratic errors. In particular, the notion of an approximate factor model due to Chamberlain and Rothschild (1983) imposes regularity conditions such that the idiosyncratic errors are allowed to be mildly cross-sectionally correlated.

To discuss the stochastic properties of the idiosyncratic errors we start with stating an assumption we already mentioned earlier, namely:

Assumption 3.1. *The panel data are sampled at a fixed time interval Δt across time.*

The above assumption implies that the idiosyncratic errors $\hat{\varepsilon}_{[h:t]}^i$ are serially uncorrelated for h sufficiently small. Recall that the panel data set is sampled at a fixed time interval $\Delta t = 1$. Consequently, two discretely sampled consecutive filtering errors are sampled far apart in a continuous record sense and therefore become

serially uncorrelated (this feature was noted by Nelson and Foster (1994, Eq. (3.23))). Hence, discipline on the idiosyncratic errors is imposed as a natural by-product of the filtering asymptotic analysis. The next lemma essentially collects a number of key results, including the lack of serial correlation, derived by Nelson and Foster (1994) which we will use in the subsequent analysis:

Lemma 3.1. *Under Assumption 3.1, the sequences* $\hat{\varepsilon}^i_{[h:t]}$ $\forall i$ *are serially uncorrelated. Moreover, under Assumptions A.1-A.4 we have as* $h \downarrow 0$: $[h^{-1/4}\hat{\varepsilon}^i_{[h:t]}|x^i_{[h:t-\sqrt{h}]} = x, \mathcal{X}^f_{t-\sqrt{h}} = \mathcal{X}, \hat{\varepsilon}^i_{[h:t-\sqrt{h}]} = q] \xrightarrow{d} N(0, \sigma^2_{it})$, *uniformly on the set* $(x, \mathcal{X}, q)$.

We provide all the detailed proofs in the Appendix. Note that the conditioning information in the above lemma is the high frequency information filtration used to compute the filtered volatilities - which is different from the usual panel data filtration.

We also need to achieve discipline on the cross-sectional dependence by controlling the rate at which N grows vis-à-vis the rate h shrinks. Because Lemma 3.1 only deals with univariate properties, we actually need to control the expansion of the cross-section, namely:

Assumption 3.2. *Let* N *be the dimension of the cross-section of the large scale factor model appearing in (3.8) and let the sampling frequency of high frequency data to estimate filtered volatilities be* h. *It is assumed that* $h \sim O(N^{-2})$. *In addition, let* $E[\hat{\varepsilon}^i_{[h:t]}\hat{\varepsilon}^j_{[h:s]}] \equiv \tau_{ij,st}$, *then:*

$$(NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\tau_{ij,st}| \leq M,$$

with M *some finite positive constant.*

The above assumption guarantees that we sample at sufficiently small intervals, relative to N, to obtain well-behaved weak cross-sectional dependence of the idiosyncratic errors. The second part of the assumption guarantees there is only weak dependence between the filtering errors for two different assets i and j at different time periods t and s. The latter is a relative mild assumption commonly used in the literature.

The estimator we consider is standard, namely the method of asymptotic principal components, initiated by was first considered by Connor and Korajczyk ((1986) and (1988)) and refined by Stock and Watson (2002), Bai and Ng (2002), Bai (2003), as an estimator of the factors in a large N and T setup. For any given r not necessarily equal to the true number of factors k, the method of principal components (PC) constructs a $T \times r$ matrix of estimated factors and a corresponding $N \times r$ matrix of estimated loadings by solving the following optimization problem for a given h :

$$\min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (x^i_{[h:t]} - \lambda'_i F_t)^2 \tag{3.9}$$

subject to the normalization that $(\Lambda'_r \Lambda_r)/N = I_r$ and $(F'_r F_r)$ being diagonal. The estimated factor matrix $\tilde{F}_{[h:t]}$ is $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $X_{[h]} X'_{[h]}$. Moreover, $\tilde{\Lambda}' = (\tilde{F}'_{[h]} \tilde{F}_{[h]})^{-1} \tilde{F}'_{[h]} X_{[h]}$ are the corresponding factor loadings.

Next we state a number of standard assumptions which we share with the standard literature on large scale factor models involving stationary data. To do so, we define $||A|| = [\text{tr}(A'A)]^{1/2}$ as the norm of matrix A.

Assumption 3.3. *Let M again be a generic positive finite constant, then*

- $E||\bar{\mathcal{X}}_t^f||^4 < M$ *and* $\frac{1}{T} \sum_t \bar{\mathcal{X}}_t^f (\bar{\mathcal{X}}_t^f)' \to \Sigma_F > 0$, *a non-random $k \times k$ matrix.*

- *the factor loadings λ_i are non-random and satisfy $||\lambda_i|| < M \; \forall \, i$ and $||\Lambda^{0\prime} \Lambda^0 / N - \Sigma_\Lambda|| \to 0$ for some $k \times k$ positive definite matrix Σ_Λ.*

- *The eigenvalues of the $k \times k$ matrix $(\Sigma_\Lambda \cdot \Sigma_F)$ are distinct.*

The first part of the above assumption imposes moment conditions which are standard in the panel data literature. The process $\bar{\mathcal{X}}_t^f$ is a point sampled copy of the diffusion appearing in (2.1) (without jumps). The regularity conditions imposed by Nelson and Foster (1994) on the continuous process, involving a finite diffusion function and thus $E||\mathcal{X}_t^f||^2 < M$, are actually weaker. However, since we do not have a direct relationship between the continuous time affine diffusion and the moments of the discretely sampled snapshots $\bar{\mathcal{X}}_t^f$, we simply impose the necessary moment conditions. It is important to note that we do not require the stringent moment conditions typically imposed to study the large sample properties of GARCH models (see e.g. Lee and Hansen (1994), Lumsdaine (1996), and subsequent literature). What matters for our analysis is the local properties of the filtering errors, not the parameter estimates of the GARCH models and their asymptotic distribution, which may involve strong conditions on the moments of returns normalized by their conditional volatility (up to 32nd moment in the case of Lumsdaine (1996, Assumption 2)). The second part of the assumption is guaranteed in the context of asset pricing models where $\lambda_i = \gamma_{iv}(\mathcal{H}^{-1}$, which reflects finite risk exposure to factors.

Theorem 3.1. *Under Assumptions 3.1-3.3 and A.1-A.4, we have:*

(a) $\text{plim}_{T,N\to\infty, h\downarrow 0} \; \tilde{F}_{[h]} \bar{\mathcal{X}}^f / T = Q$ *an invertible matrix equal to* $V^{1/2} \Upsilon' \Sigma_\Lambda^{-1/2}$, *where* $V = diag(\nu_1, \nu_2, \ldots, \nu_k)$, $\nu_1 > \nu_2 \ldots, \nu_k$ *are the eigenvalues of* $\Sigma_\Lambda^{1/2} \Sigma_\Lambda \Sigma_\Lambda^{1/2}$, *and* Υ *is the corresponding eigenvector matrix such that* $\Upsilon' \Upsilon = I_k$.

(b) *if* $\sqrt{N}/T \to 0$, *then for each* $t : (\sqrt{N}/h^{1/4})(\tilde{F}_{[h:t]} - \mathcal{H} \bar{\mathcal{X}}_t^f) \xrightarrow{d} N(0, V^{-1} Q \Lambda_t Q' V^{-1})$,

The proof appears in Appendix B. The difference between Bai (2003, Theorem 1) and Theorem 3.1 is the rate of convergence for the estimation of the factors. Indeed, we note from the above theorem that the estimation of the factors has a rate of convergence $(\sqrt{N}/h^{1/4})$ instead of the standard $\sqrt{N}$ asymptotic results found

11

in Bai and Ng (2002), Bai (2003), among others. The reason for the difference is the combination of the continuous record asymptotics of Nelson and Foster (1994) and the large panel data analysis. In the latter case, standard central limit theorems are *assumed* for the error process of the panel data model (see e.g. Assumption F of Bai (2003)). In our analysis, the dependence structure of the errors follows directly from the assumed sampling scheme for the ARCH filters. A combination of Lemma 3.1 and Assumption 3.2 yields an intriguing result which is intuitively easy to explain. Because of Assumption 3.2, we need to sample the return data feeds for the ARCH filters more frequently, in fact let us take as example $h = N^{-2}$. Then, as we expand the cross-section, we simultaneously squeeze the ARCH filtering errors, and therefore the errors of the panel data. With $h = N^{-2}$, we actually obtain a rate of converge equal to N, instead of the usual $\sqrt{N}$. This actually implies that ARCH filter panel data models yield *super-consistent* estimators, borrowing a concept from the unit root literature (e.g. Dickey and Fuller (1981)), for the factor process $\bar{\mathcal{X}}_t^f$. In the next section, we will have different rates of convergence, when we replace the continuous record asymptotics with various in-fill asymptotic results for so called realized measures.

It should finally also be noted that the number of factors can be consistently estimated the number of factors k by analyzing the statistical properties of the minimand $V(\tilde{k})$ appearing in equation (3.9) as a function of $\tilde{k}$, where $\tilde{k}$ is not necessarily the true k. Bai and Ng (2002) showed that the number of factors can be estimated consistently by minimizing the following criterion:

$$IC(\tilde{k}) = \log\left(V(\tilde{k})\right) + \tilde{k}\left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right)$$

We show in Appendix B that this criterion remains valid for the selection of factors.

To conclude it should be noted that the analysis in this section is not limited to the ARCH-type filters discussed by Nelson and Foster (1994). A number of papers have recently proposed purely data-driven, typically kernel-based, estimators of spot volatility using high frequency data, see e.g. Zhang (2001b), Andreou and Ghysels (2002), Malliavin and Mancino (2005), Fan, Fan, and Jiang (2007), Fan and Wang (2008), Mykland and Zhang (2008), Zhao and Wu (2008), Kristensen (2010), among others. With suitable modifications of the regularity conditions our analysis can easily be extended to high frequency purely data-driven spot volatility estimators. With ARCH-type filters we required in Assumption 3.2 that $h \sim O(N^{-2})$. For example, in the case of kernel-based spot volatility estimators considered notably by Fan and Wang (2008) the required rate becomes $\sqrt{h}b \sim O(N^{-1})$ where b is the bandwidth used in the kernels. Hence, alternative estimators, with different convergence rates require different balancing of asymptotic rates with the cross-sectional dimension N. This is a theme which we will also explore further in the next section.

4 Factor Analysis with Panels of Realized Measures

We now consider a cross-section of asset driven by a state variable process which may feature jumps. Combining equation equations (2.1) and (3.2) implies that x_t^i satisfies the diffusion:

$$dx_t^i = \mu^i(\mathcal{X}_t^f)dt + \sigma^i(\mathcal{X}_t^f)dW_t^{\mathbb{P}} + dZ_t^i \tag{4.1}$$

where $\mathcal{X}_t^f$ satisfies Assumption 2.1, i.e. including jumps, and therefore by Itô's lemma: $\mu^i(\mathcal{X}_t^f) \equiv \delta^i \mu^{\mathbb{P}}(\mathcal{X}_t^f)$, $\sigma^i(\mathcal{X}_t^f) \equiv \delta^i \sigma(\mathcal{X}_t^f)$, and finally Z_t^i is a pure jump process whose jumps have a fixed probability distribution ν^i with $\nu^i(x) \equiv \nu^{\mathbb{P}}(\delta_0^i + \delta^i x)$ on $\mathbb{R}^k$ with the arrival intensity of $\mathcal{X}_t^f$. It will be useful to rewrite the above equation more explicitly with regards to the jump size realizations, namely:

$$dx_t^i = \mu^i(\mathcal{X}_t^f)dt + \sigma^i(\mathcal{X}_t^f)dW_t^{\mathbb{P}} + \theta_t^i dJ_t^i \tag{4.2}$$

where dJ_t is a counting process with $dJ_t = 1$ corresponding to a jump at t and $dJ_t = 0$ if no jump and θ_t^i is the jump size.

Instead of examining spot volatility we are interest in the realized volatility over some interval $[t-d, t]$, where $d \leq \Delta t$. This means that we could potentially look at realized volatilities over the entire interval Δt considered as sampling frequency of the panel data. This is therefore different from the analysis in the previous section where we were concerned about the spot volatility of the factor process (and therefore point sampled $\bar{\mathcal{X}}_t^f$) spaced at discrete distances Δt. Note that we can set d such that a monthly panel of (a) monthly, (b) last week of the month, (c) last day of the month, etc., realized volatilities are collected. The only constraint is that they are non-overlapping. As noted in the Introduction, we are not the first to suggest a large dimensional panel data approach to realized measures, since they were proposed by Anderson and Vahid (2007) and Barigozzi, Brownlees, Gallo, and Veredas (2011). Neither paper, however, deals with the asymptotic theory implications of estimation errors in realize measures and their impact on the implied approximate factor model structure. It is the purpose of this section to do so. To study realized volatilities, we are interested in measures such as the increments of quadratic variation:

$$QV_{[t-d,t]}^i = \int_{t-d}^t \sigma^i(\mathcal{X}_s^f)ds + \sum_{s \in [t-d,t]: J_s^i = 1} (\delta_0^i + \delta^i \theta_s^i)^2.$$

which can also be written as:

$$
\begin{aligned}
QV_{[t-d,t]}^i &= \delta^i \int_{t-d}^t \sigma(\mathcal{X}_s^f)ds + \sum_{s \in [t-d,t]: J_s^i = 1} (\theta_s^i)^2 \\
&= \delta^i QV_{[t-d,t]}^{\mathcal{X}^f} + \sum_{s \in [t-d,t]: J_s^i = 1} (\delta_0^i + \delta^i \theta_s^{\mathcal{X}^f})^2.
\end{aligned}
\tag{4.3}
$$

Suppose again for the moment that $\mathcal{X}^f$ is a bivariate process and without loss of generality consists of two mutually orthogonal processes and as in equation (3.4) call them $\mathcal{X}^1$ and $\mathcal{X}^2$. Then we can rewrite (4.3) as:

$$QV^i_{[t-d,t]} = (\delta^i_1)^2 QV^{\mathcal{X}^1}_{[t-d,t]} + (\delta^i_2)^2 QV^{\mathcal{X}^2}_{[t-d,t]} + \sum_{s\in[t-d,t]:J^i_s=1} (\delta^i_0 + \delta^i\theta^{\mathcal{X}^f}_s)^2 \tag{4.4}$$

which means we are looking at a three-factor model, with one being driven by the jump process. Note that the latter is not the usual factor process we typically encounter in panel data models, but the regularity conditions one imposes do not preclude this type of process. We could ignore the presence of jumps, and therefore the last term on the right hand side of (4.4) would disappear. Comparing equations (3.4) with (4.4) tells us that even in the absence of jumps, using spot volatility versus realized volatility estimates would not yield the same factors. The former yields factor representations in terms of point-sampled $\mathcal{X}^f$, whereas the latter yields integrated volatilities of the same process as factors. Note, however, that the *number* of factors should be the same if we operate under the assumption that no jumps are present (and thus the estimation of spot volatilities is legitimate).

The volatility measures appearing in equation (4.3) are not observable but can be estimated from data. It is possible to consistently estimate $QV_{[t-d,t]}$ in (4.3) by summing squared high frequency returns, yielding the so called realized variance, namely:

$$\overline{QV}^i_{[h:t-d,t]} = \sum_{j=1}^{d/h} \left(x^i_{[h:t-1+jh]} - x^i_{[h:t-1+(j-1)h]} \right)^2. \tag{4.5}$$

Barndorff-Nielsen and Shephard (2002b), Jacod and Protter (1998) and Zhang (2001a) show that, under mild regularity conditions satisfied by the class of affine diffusion with bounded volatility paths, the error of realized variance is asymptotically

$$h^{-1/\alpha}(\overline{QV}^i_{[h:t-d,t]} - QV^i_{[t-d,t]}) \xrightarrow{d} N\left(0, \sqrt{2Q^i_{[t-d,t]}}\right) \tag{4.6}$$

where $Q^i_{[t-d,t]}$ is the so called quarticity. Note that we parameterize the rates of convergence with α in order to formulate a generalization of Assumption 3.2 and being able to handle various realized measures. In equation (4.6) we have $\alpha = 2$. Note that this implies

$$\overline{QV}^i_{[h:t-d,t]} = \delta^i_1 QV^{\mathcal{X}^1}_{[t-d,t]} + \delta^i_2 QV^{\mathcal{X}^2}_{[t-d,t]} + \sum_{s\in[t-d,t]:J^i_s=1} (\delta^i_0 + \delta^i\theta^{\mathcal{X}^f}_s)^2 + \hat{\varepsilon}^i_{[h:t-d,t]} \tag{4.7}$$

where $h^{-1/\alpha}\hat{\varepsilon}^i_{[h:t-d,t]} \xrightarrow{d} N(0, \sqrt{2Q^i_{[t-d,t]}})$.

In order to separate the jump and continuous sample path components of $QV_{[t-d,t]}$ Barndorff-Nielsen and Shephard (2004a) and Barndorff-Nielsen and Shephard (2004b) introduce the concept of bi-power variation

defined as:

$$\overline{BPV}^i_{[h:t-d,t,,k]} = \mu_1^{-2} \sum_{j=k+1}^{1/h} \left| x^i_{[h:t-1+jh]} - x^i_{[h:t-1+(j-1)h]} \right| \left| x^i_{[h:t-1+jh]} - x^i_{[h:t-1+(j-1-k)h]} \right|, \qquad (4.8)$$

where $\mu_a = E|Z|^a$ and $Z \sim N(0,1)$, $a > 0$. Henceforth we will, without loss of generality, specialize our discussion the case $k = 1$, and therefore drop it to simplify notation. Barndorff-Nielsen and Shephard (2004b) establish the sampling behavior of $\overline{BPV}^i_{[h:t-d,t]}$ as $h \to 0$, and show that under suitable regularity conditions:

$$\lim_{h \to 0} \overline{BPV}^i_{[h:t-d,t,k]} = (\delta^i)^2 \int_{t-d}^t \sigma(\mathcal{X}_s^f)^2 ds \qquad (4.9)$$

Therefore, in the presence of jumps, $\overline{BPV}^i_{[h:t-d,t]}$ converges to the continuous path component of $QV_{[t-d,t]}$ and is not affected by jumps. The sampling error of the bi-power variation is

$$h^{-1/\alpha} \left(\overline{BPV}^i_{[h:t-d,t,,k]} - (\delta^i)^2 \int_{t-d}^t \sigma(\mathcal{X}_s^f)^2 \right) ds \sim N\left(0, \sqrt{\nu_{bb} Q_t}\right) \qquad (4.10)$$

where $\nu_{bb} = (\pi/4)^2 + \pi - 5 \approx 0.6090$ and α is again equal to two. Consequently, we also have:

$$\overline{BPV}^i_{[h:t-d,t]} = \delta_1^i QV_{[t-d,t]}^{\mathcal{X}1} + \delta_2^i QV_{[t-d,t]}^{\mathcal{X}2} + \hat{\varepsilon}^i_{[h:t-d,t,k]}. \qquad (4.11)$$

So far we have two types of panel data models, appearing respectively in equations (4.7) and (4.11). The former is a panel or realized volatilities and related to both volatility and jump risk factors. The latter is a panel linked only to volatility risk. The latter is closer to the panel of spot volatilities studied in the previous section, although as noted earlier the interpretation of the factors is not the same.

We will need to impose some discipline, by analogy with the analysis in the previous section, on the error process $\hat{\varepsilon}^i_{[h:t-d,t]}$ appearing respectively in equations (4.7) and (4.11). In particular, the generalization of Assumption 3.2 is as follows:

Assumption 4.1. *Let N be the dimension of the cross-section of the large scale factor model appearing in (3.8) and let the sampling frequency of high frequency data to compute realized volatilities be h. It is assumed that $h \sim O(N^{-(\alpha/2)})$. In addition, let $E[\hat{\varepsilon}^i_{[h:t-d,t]} \hat{\varepsilon}^j_{[h:t-d,s]}] \equiv \tau_{ij,st}$, then:*

$$(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,st}| \leq M,$$

with M some finite positive constant.

If a researcher decides to use panel data sets of realized measures, either $\overline{QV}^i_{[h:t-d,t]}$ or $\overline{BPV}^i_{[h:t-d,t,,k]}$, then the following theorem is proven in Appendix C:

Theorem 4.1. *Under Assumptions 3.1-3.3, with 3.2 replaced by 4.1 and A.1-A.4, we have:*

(a) $\text{plim}_{T,N\to\infty,h\downarrow 0}\ \tilde{F}_{[h]}\bar{\mathcal{X}}^f/T = Q$ *an invertible matrix equal to* $V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$, *where* $V = diag(\nu_1, \nu_2, \ldots, \nu_k)$, $\nu_1 > \nu_2 \ldots, \nu_k$ *are the eigenvalues of* $\Sigma_\Lambda^{1/2}\Sigma_\Lambda\Sigma_\Lambda^{1/2}$, *and* Υ *is the corresponding eigenvector matrix such that* $\Upsilon'\Upsilon = I_k$.

(b) if $\sqrt{N}/T \to 0$, *then for each* $t : (\sqrt{N}/h^{1/\alpha})(\tilde{F}_{[h:t]} - \mathcal{H}\bar{\mathcal{X}}_t^f) \xrightarrow{d} N(0, V^{-1}Q\Lambda_t Q'V^{-1})$,

The rate of convergence is now $(\sqrt{N}/h^{1/\alpha})$, which depends on the realized measure being used, the presence of microstructure noise, and the choice of feasible estimators. However, the relationship between the high frequency data sampling scheme and the cross-sectional dimension also changes accordingly, as specified in Assumption 4.1, achieving again superconsistency. For example, if we take the simple case of realized volatility in the absence of jumps, the case comparable to the analysis in the previous section, we note that the rate of convergence is again equal to N, namely with $\alpha = 2$, and $h = 1/N$.

The analysis in this section can be extended to various other realized measures. For example, when microstructure noise affects high frequency returns, we obtain estimators $\overline{QV}_t$ with different convergence rates α and different definitions of Q_t. In the case of the two scales realized volatility (Zhang, Mykland, and Aït-Sahalia (2005)) $\alpha = 12$, and in the case of the multi-scale estimator (Zhang (2006)), or the kernel estimators of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), $\alpha = 8$. The latter rate is efficient since it also occurs in the parametric case (see Gloter and Jacod (2000)). With a suitable change in links between high frequency sampling and the cross-section, as spelled out in Assumption 4.1 we obtain the same asymptotic results for large panel estimators applied to the modified realized measures.

5 Simulations

We consider 1,000 replications of samples each consisting of 500 'days' with in-sample (intra-daily sampling) sizes $n = 288$. These correspond to the use of five-minute returns in a 24-hour financial market. Every simulation has a 1000 days burn-in pre-sample period to eliminate starting value problems.

We will examine cases involve a single factor and two types of models involving two factors. Using the terminology of Chernov, Gallant, Ghysels, and Tauchen (2003) we will consider: SV1F, SV2F and SV1FJ type of models. The SV1F case consists of $\mathcal{X}_{1t}^f$ specified below, SV2F consists of $\mathcal{X}_{1t}^f$ and $\mathcal{X}_{2t}^f$, and finally SV1FJ will correspond to $\mathcal{X}_{1t}^f$ and $\mathcal{X}_{3t}^f$. The three processes are special cases of (2.1), namely:

$$\begin{bmatrix} d\mathcal{X}_{1t}^f \\ d\mathcal{X}_{2t}^f \\ d\mathcal{X}_{3t}^f \end{bmatrix} = \begin{bmatrix} \alpha_{01} + \alpha_{11}\mathcal{X}_{1t}^f \\ \alpha_{02} + \alpha_{12}\mathcal{X}_{2t}^f \\ 0 \end{bmatrix} dt + \begin{bmatrix} (\mathcal{X}_{1t}^f)^{1/2} \\ (\mathcal{X}_{2t}^f)^{1/2} \\ 0 \end{bmatrix} \begin{bmatrix} dW_{1t} \\ dW_{2t} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ dZ_t \end{bmatrix} \quad (5.1)$$

where: $dZ_t = J_t dq_t$, with $J_t \sim N(\mu_J, \sigma_J^2)$ $q_t \sim Poi(\lambda_J)$. Using equation (3.1) we generate a cross-section as follows:

- For SV1F: $x_t^i \equiv \delta_0^i + \delta_1^i \mathcal{X}_{1t}^f$, where the parameters are random draws $\delta_0^i \sim N(0, \sigma_0^2)$ and $\delta_1^i \sim N(1, \sigma_1^2)$, both i.i.d. across i.

- For SV2F: $x_t^i \equiv \delta_0^i + \delta_1^i \mathcal{X}_{1t}^f + \delta_2^i \mathcal{X}_{2t}^f$, where the parameters are again random draws $\delta_0^i \sim N(0, \sigma_0^2)$ and $\delta_j^i \sim N(1, \sigma_j^2)$, for $j = 1$ and 2 also i.i.d. across i.

- For SV1FJ: $x_t^i \equiv \delta_0^i + \delta_1^i \mathcal{X}_{1t}^f + \delta_3^i \mathcal{X}_{3t}^f$, where the parameters are again random draws $\delta_0^i \sim N(0, \sigma_0^2)$ and $\delta_j^i \sim N(1, \sigma_j^2)$, for $j = 1$ and 3 also i.i.d. across i.

Hence, the cross-section is created by a random draw of exposures to the factors.

<div style="text-align: center;">INCOMPLETE</div>

Technical Appendix

A   Regularity Conditions

In this section we list a number of regularity conditions and report a number of useful results reported elsewhere in the literature. We start with the case of ARCH filters and then cover realized volatilities.

A.1   Panel Data Analysis of ARCH Filters

Assumption A.1. *The continuous time process x_t^i satisfies for all $i = 1, \ldots, N$ the diffusion:*

$$dx_t^i = \mu^i(\mathcal{X}_t^f)dt + \sigma^i(\mathcal{X}_t^f)dW_t^{\mathbb{P}} \tag{A.1}$$

where $\mathcal{X}_t^f$ satisfies Assumption 2.2, i.e. excluding jumps, and therefore by Itô's lemma: $\mu^i(\mathcal{X}_t^f) \equiv \delta^i \mu^{\mathbb{P}}(\mathcal{X}_t^f)$ and $\sigma^i(\mathcal{X}_t^f) \equiv \sigma(\mathcal{X}_t^f)$, using equations (2.1) and (3.2). Moreover, for each $h > 0$ the discrete time Markov process $x_{[h:t]}^i$ satisfies for each $\Delta > 0$ and integer k,

- $\mu_{\Delta,h}^i(x) = h^{-\Delta} E[x_{[h:kh+1]}^i - x_{[h:kh]}^i | x_{[h:kh]}^i = x]$
- $\sigma_{\Delta,h}^i(x) = h^{-\Delta} Var[x_{[h:kh+1]}^i - x_{[h:kh]}^i | x_{[h:kh]}^i = x]$

where: $\mu_{\Delta,h}^i(x) \to \mu^i(x)$, $\sigma_{\Delta,h}^i(x) \to \sigma^i(x)$ and for some $d > 0$, $h^{-\Delta} E[(x_{[h:kh+1]}^i - x_{[h:kh]}^i)^{2=d} | x_{[h:kh]}^i = x] \to 0$, where $\to$ refers to convergence uniform on every bounded x set.

Then following Stroock and Varadhan (1979) and Nelson and Foster (1994), $x_{[h:t]}^i \Rightarrow x_t^i$, where $\Rightarrow$ denotes weak convergence. Moreover, let F^i be the cumulative distribution of the starting point x_0, and F_h^i be the cumulative distribution of the starting point $x_{[h:0]}^i$. If $x_{[h:0]}^i$ sets x_0^i with probability one for all h, then the weak convergence is uniform on every bounded x_0^i set.

The filter appearing in (3.6) satisfies:

Assumption A.2. *For all $i = 1, \ldots, N$ the functions $\hat{\kappa}^i$, $\hat{\xi}_{x,t}^i$, and g_t^i satisfy:*

- *$\hat{\kappa}^i$ and $\hat{\xi}_{x,t}^i$ are continuous in all arguments*
- *g_t^i is differentiable in $\hat{\xi}_{x,t}^i$, $x_{[h:t-h]}^i$, $\hat{V}_{[h:t-h]}^i$ and h almost everywhere and must possess one-sided derivatives everywhere.*

Moreover, let $\forall i : q_{[h:t]}^i \equiv h^{-1/4} \hat{V}_{[h:t]}^i$ - V_t^i, which satisfies:

Assumption A.3. *For all $i = 1, \ldots, N$, uniformly on every bounded (x, q, t) set:*

$$h^{-1/2} E[q_{[h:t+h]}^i - q_{[h:t]}^i | x_t^i = x, \mathcal{X}_t^f = \mathcal{X}, q_{[h:t]}^i = q] \to qB(x, y, t),$$

$$h^{-1/2} Var[q^i_{[h:t+h]} - q^i_{[h:t]}|x^i_t = x, \mathcal{X}^f_t = \mathcal{X}, q^i_{[h:t]} = q] \to C(x, y, t),$$

as $h \downarrow 0$ where:

$$
\begin{aligned}
B^i(x, y, t) &= \lim_{h \downarrow 0} E[\partial g^i\left(\hat{\xi}^i_{x,t}, x^i_{[h:t-h]}, \hat{V}^i_{[h:t-h]}, t, h\right) / \partial V^i | x^i_{[h:t-h]} = x, \hat{V}^i_{[h:t-h]} = y] \\
C^i(x, y, t) &= \lim_{h \downarrow 0} E[(g^i\left(\hat{\xi}^i_{x,t}, x^i_{[h:t-h]}, \hat{V}^i_{[h:t-h]}, t, h\right))^2 | x^i_{[h:t-h]} = x, \hat{V}^i_{[h:t-h]} = y]
\end{aligned}
\tag{A.2}
$$

Further $B^i(x, y, t)$ and $C^i(x, y, t)$ are twice continuously differentiable in x and y.

Assumption A.4. *For some $d > 0$ and for all $i = 1, \ldots, N$ then both:*

$$E[h^{-1/2}|x^i_{t+h} - x^i_t|^{2+d}|x^i_t = x, V^i_t = y]$$
$$E[h^{-1/2}|V^i_{t+h} - V^i_t|^{2+d}|x^i_t = x, V^i_t = y]$$

are bounded as $h \downarrow 0$, uniformly on every bounded (x, y, t) set, and

$$\limsup_{h \downarrow 0} E[g^i\left(\hat{\xi}^i_{x,t}, x^i_{[h:t-h]}, \hat{V}^i_{[h:t-h]}, t, h\right)^{2+d}|x^i_t = x, V^i_t = y, q_t = q]$$

is bounded uniformly on every bounded (x, y, q, t) set.

B   Proof of Theorem 3.1

We start by noting that the first part of Lemma 3.1 is a direct consequence of Theorem 3.1 in Nelson and Foster (1994), where σ^2_{it} equals:

$$\sigma^2_{it} = C^i(x, y, t)(1 - \exp\left(-B^i(x, y, t)\right))/(2B^i(x, y, t)) \tag{B.3}$$

using equation (3.22) in Nelson and Foster (1994) and the expressions in Assumption A.2 evaluated at $x^i_{t-\sqrt{h}} = x$, $\mathcal{X}^f_{t-\sqrt{h}} = \mathcal{X}$ and $t - \sqrt{(h)}$. Note also that the mean of the Gaussian limiting distribution is zero, a result also noted by Nelson and Foster (1994) in the case of diffusions - as is the case we consider. Moreover, the results of Nelson and Foster (1994) imply a finite variance across all i.

We first proof the following lemmas:

Lemma B.1. *Assuming the conditions in Lemma 3.1, $E[N^{-1}\sum_{i=1}^N (\hat{\varepsilon}^i_{[h:t]})^2] \leq M \ \forall \ t$ and all N.*

Proof: $E[N^{-1}\sum_{i=1}^N (\hat{\varepsilon}^i_{[h:t]})^2] = N^{-1}h^{1/2}\sum_{i=1}^N E[(h^{-1/2}\hat{\varepsilon}^i_{[h:t]})^2] = N^{-1}h^{1/2}\sum_{i=1}^N \sigma^2_{it} \leq N^{-1}h^{1/2}\sum_{i=1}^N (\sup_i \sigma^2_{it})$
$\leq M$, the latter following Lemma 3.1, namely the fact that $\forall \ i \ \sigma_{it}$ is finite. This holds for all t and N. ∎

The following two Lemmas hold straightfowardly because the sequences $\hat{\varepsilon}^i_{[h:t]} \ \forall i$ are serially uncorrelated.

Lemma B.2. *Assuming the conditions in Lemma 3.1, then $T^{-1}\sum_{s=1}^T \sum_{t=1}^T E[N^{-1}\sum_{i=1}^N (\hat{\varepsilon}^i_{[h:t]}\hat{\varepsilon}^i_{[h:s]})] \leq M,$ for all N and T.*

Lemma B.3. *Under Assumptions 3.2 and those in Lemma 3.1, for every t and s :*

$$E[N^{-1/2} \sum_{i=1}^{N} \hat{\varepsilon}_{[h:s]}^{i} \hat{\varepsilon}_{[h:t]}^{i} - E(\hat{\varepsilon}_{[h:s]}^{i} \hat{\varepsilon}_{[h:t]}^{i})]^4 \leq M$$

Lemma B.4. *Under Assumptions 3.2 and those in Lemma 3.1, then $N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} sup_t E[\hat{\varepsilon}_{[h:t]}^{i} \hat{\varepsilon}_{[h:t]}^{j}] \leq M$, for all N.*

Proof: Applying first the Cauchy-Schwarz and then applying the results of Lemma 3.1, namely the fact that $\forall~i~\sigma_{it}$ is finite yields: $N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} sup_t E[\hat{\varepsilon}_{[h:t]}^{i} \hat{\varepsilon}_{[h:t]}^{j}] = N^{-1} h^{1/2} \sum_{i=1}^{N} \sum_{j=1}^{N} sup_t E[h^{-1/2} \hat{\varepsilon}_{[h:t]}^{i} \hat{\varepsilon}_{[h:t]}^{j}] \leq N^{-1} h^{1/2} \sum_{i=1}^{N} \sum_{j=1}^{N} sup_t (\sigma_{it}^2)^{1/2} (\sigma_{jt}^2)^{1/2} \leq N^{-1} h^{1/2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sup_{i,t}(\sigma_{it}^2) \leq N h^{1/2} \sup_{i,t}(\sigma_{it}^2) \leq M$, the latter following Assumption 3.2, which balances the growth of N and the shrinkage of h. ∎

Lemma B.5. *Under Assumptions 3.2 and those in Lemma 3.1, for every t and s :*

$$E\left(N^{-1} \sum_{i=1}^{N} \|T^{-1/2} \sum_{t=1}^{T} \bar{\mathcal{X}}_t^f \hat{\varepsilon}_{[h:t]}^{i}\|^2 \right) \leq M$$

and for each t :

$$E\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \sum_{s=1}^{T} \bar{\mathcal{X}}_t^f [\hat{\varepsilon}_{[h:s]}^{k} \hat{\varepsilon}_{[h:t]}^{k} - E(\hat{\varepsilon}_{[h:s]}^{k} \hat{\varepsilon}_{[h:t]}^{k})]\|^2 \leq M$$

while for the $k \times k$ matrix:

$$E\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \sum_{t=1}^{T} \bar{\mathcal{X}}_t^f \lambda_k^0 \hat{\varepsilon}_{[h:t]}^{k}\|^2 \leq M$$

Proof: Regarding the first inequality, note that:

$$
\begin{aligned}
E(N^{-1} \sum_{i=1}^{N} \|T^{-1/2} \sum_{t=1}^{T} \bar{\mathcal{X}}_t^f \hat{\varepsilon}_{[h:t]}^{i}\|^2) &= E(N^{-1} \sum_{i=1}^{N} h^{1/2} \|T^{-1/2} \sum_{t=1}^{T} \bar{\mathcal{X}}_t^f h^{-1/4} \hat{\varepsilon}_{[h:t]}^{i}\|^2) \\
&\leq E(N^{-1} h^{1/2} \sum_{i=1}^{N} \|(T^{-1/2} \sum_{t=1}^{T} (h^{-1/2} \hat{\varepsilon}_{[h:t]}^{i})^2)^{1/2} (T^{-1/2} \sum_{t=1}^{T} (\|\bar{\mathcal{X}}_t^f\|^2)^{1/2}\|^2)) \\
&= E(N^{-1} h^{1/2} \sum_{i=1}^{N} (T^{-1/2} \sum_{t=1}^{T} (h^{-1/2} \hat{\varepsilon}_{[h:t]}^{i})^2) \|(T^{-1/2} \sum_{t=1}^{T} (\|\bar{\mathcal{X}}_t^f\|^2)^{1/2}\|^2)) \leq M
\end{aligned}
$$

where the latter is a result of Assumptions 3.3, 3.2 and those in Lemma 3.1. The second inequality is easy to obtain by applying the Cauchy-Schwarz combined with the fact that $\hat{\varepsilon}_{[h:t]}^{i}$ is serially uncorrelated. ∎

Lemma B.6. *Under Assumptions 3.2 and those in Lemma 3.1, for all T and N, and for every $t < T$, with $\sqrt{N}/T \to 0 : \sum_{s=1}^{T} E(N^{-1} \sum_{i=1}^{N} \hat{\varepsilon}_{[h:s]}^{i} \hat{\varepsilon}_{[h:t]}^{i}) < M$ and for every $i < N : \sum_{k=1}^{N} E(\hat{\varepsilon}_{[h:t]}^{k} \hat{\varepsilon}_{[h:t]}^{i}) < M$.*

Proof: The first inequality follows straightforwardly from the absence of serial dependence across time, while the proof of the second inequality is similar to that of Lemma **??** and therefore omitted.

Lemma B.7. *Under Assumptions 3.2 and those in Lemma 3.1, for each t as $N \to \infty$:*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i^0 h^{-1/4} \hat{\varepsilon}_{[h:t]}^i \overset{d}{\to} N(0, \Gamma_t),$$

where $\Gamma_t = \lim_{N \to \infty} (1/N) \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i^0 (\lambda_i^0)' E(h^{-1/2} \hat{\varepsilon}_{[h:t]}^i \hat{\varepsilon}_{[h:t]}^j);$

The sample average involves Gaussian random variables with dependence which decreases across the entire sample at a rate $h^{1/2}$ which is $O(1/N)$. Kolmogorov and Rozanov (1960) showed that stationary Gaussian sequences with such dependence are strong mixing, and therefore a CLT applies, see also Kuersteiner and Prucha (2013) for further discussion. ∎

Lemma B.8. *Under Assumptions 3.2, 3.3 and those in Lemma 3.1, for each i as $T \to \infty$:*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{\mathcal{X}}_t^f h^{-1/4} \hat{\varepsilon}_{[h:t]}^i \overset{d}{\to} N(0, \Phi_i),$$

where $\Phi_i = \lim_{T \to \infty} (1/T) \sum_{s=1}^{T} \sum_{t=1}^{T} E[\bar{\mathcal{X}}_t^f (\bar{\mathcal{X}}_t^f)' (h^{-1/2} \hat{\varepsilon}_{[h:t]}^i \hat{\varepsilon}_{[h:s]}^i)];$

Applying the Cauchy-Schwarz combined with the fact that $\hat{\varepsilon}_{[h:t]}^i$ is Gaussian and serially independent implies that a standard CLT applies. ∎

The result in part (a) of Theorem 3.1 follow from the fact that Assumptions 3.2, 3.3 and those in Lemma 3.1 imply that the conditions of Bai (2003, Proposition 1) hold. The results in Lemmas B.1 through B.8 imply that Bai (2003, Assumptions A-G) apply with the modification of the rates of convergence, namely:

$$(\sqrt{N}/h^{1/4})(\tilde{F}_{[h:t]} - \mathcal{H}\bar{\mathcal{X}}_t^f) = V_{NT}^{-1} T^{-1} \sum_{t=1}^{T} (\tilde{F}_{[h:t]} (\mathcal{H}\bar{\mathcal{X}}_t^f))' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i^0 h^{-1/4} \hat{\varepsilon}_{[h:t]}^i + o_p(1).$$

where the latter comes from the fact, as shown in Lemma B.7 that we need to adjust for the shrinkage in high frequency data, driven by $h^{-1/4}$, the rate of convergence in Nelson and Foster (1994). ∎

C   Proof of Theorem 4.1

We start with a result equivalent to Lemma 3.1, namely:

Lemma C.1. *Let Assumption 3.1 hold, then the sequences $\hat{\varepsilon}_{[h:t-d,t]}^i$ and $\hat{\varepsilon}_{[h:t-d,t,k]}^i$ defined in respectively equations (4.7) and (4.11) are serially uncorrelated when $d \leq \Delta t$. Moreover, under Assumptions 2.2 and 2.4 then as $h \downarrow 0$:*

$$h^{-1/\alpha} \hat{\varepsilon}_{[h:t-d,t]}^i \overset{d}{\to} N\left(0, \sqrt{2Q_{[t-d,t]}^i}\right) \tag{C.4}$$

and under Assumptions 2.1 and 2.4, again as $h \downarrow 0$:

$$h^{-1/2} \hat{\varepsilon}_{[h:t-d,t,k]}^i \overset{d}{\to} N\left(0, \sqrt{\nu_{bb} Q_t}\right) \tag{C.5}$$

where $\nu_{bb} = (\pi/4)^2 + \pi - 5 \approx 0.6090$.

Proof: Regarding the autocorrelation properties of the sequences $\hat{\varepsilon}^i_{[h:t-d,t]}$ and $\hat{\varepsilon}^i_{[h:t-d,t,k]}$ we rely on the results of Ghysels, Mykland, and Renault (2010) who show that the autocorrelations of $\hat{\varepsilon}^i_{[h:t-d,t]}$ and $\hat{\varepsilon}^i_{[h:t-d,t,k]}$ are $O(h)$, and there are asymptotically uncorrelated. The asymptotic distribution result in (C.4) follows from Barndorff-Nielsen and Shephard (2002a, Theorem 1). Finally, the distributional result in (**??**) follows from Assumption 2.4, which implies that the regularity conditions of Barndorff-Nielsen and Shephard (2004a, Theorem 1) are satisfied. ∎

We skip the details of the rest of the proof as the steps mimic those in the previous section. Namely, the proofs of Lemmas B.1 through B.8, suitably modified using $\hat{\varepsilon}^i_{[h:t-d,t]}$ and $\hat{\varepsilon}^i_{[h:t-d,t,k]}$ remain valid. The rest of the proof then follows straightforwardly using the same reasoning as in Appendix B.

References

AÏT-SAHALIA, Y., M. KARAMAN, AND L. MANCINI (2012): "The term structure of variance swaps, risk premia and the expectation hypothesis," Discussion Paper, Princeton University.

AMENGUAL, D. (2009): "The term structure of variance risk premia," Discussion Paper, Princeton University.

ANDERSON, H. M., AND F. VAHID (2007): "Forecasting the volatility of australian stock returns: Do common factors help?," *Journal of Business and Economic Statistics*, 25, 76–90.

ANDREOU, E., AND E. GHYSELS (2002): "Rolling sample volatility estimators: some new theoretical, simulation and empirical results," *Journal of Business and Economic Statistics*, 20, 363–376.

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*, pp. 191–221.

BAKSHI, G., G. PANAYOTOV, AND G. SKOULAKIS (2011): "Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios," *Journal of Financial Economics*, 100, 475–495.

BANSAL, R., AND A. YARON (2004): "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of Finance*, 59, 1481–1509.

BARIGOZZI, M., C. BROWNLEES, G. GALLO, AND D. VEREDAS (2011): "Disentangling systematic and idiosyncratic risk for large panels of assets," Available at SSRN 1618565.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002a): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 253–280.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002b): "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 17, 457–477.

——— (2004a): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.

——— (2004b): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–48.

BOLLERSLEV, T., AND V. TODOROV (2011): "Tails, fears, and risk premia," *Journal of Finance*, 66, 2165–2211.

BRITTEN-JONES, M., AND A. NEUBERGER (2000): "Option prices, implied price processes, and stochastic volatility," *Journal of Finance*, 55, 839–866.

BÜHLER, H. (2006): "Consistent variance curve models," *Finance and Stochastics*, 10, 178–203.

CARR, P., AND L. WU (2009): "Variance risk premiums," *Review of Financial Studies*, 22, 1311–1341.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, pp. 1281–1304.

CHERNOV, M., R. GALLANT, E. GHYSELS, AND G. TAUCHEN (2003): "Alternative models for stock price dynamics," *Journal of Econometrics*, 116, 225–257.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics*, 15, 373–394.

——— (1988): "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21, 255–289.

CORRADI, V. (2000): "Reconsidering the continuous time limit of the GARCH (1, 1) process," *Journal of Econometrics*, 96, 145–153.

CORRADI, V., W. DISTASO, AND A. MELE (2013): "Macroeconomic determinants of stock market volatility and volatility risk-premiums," *Journal of Monetary Economics* (forthcoming).

DAI, Q., AND K. J. SINGLETON (2000): "Specification analysis of affine term structure models," *Journal of Finance*, 55, 1943–1978.

DICKEY, D. A., AND W. A. FULLER (1981): "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49, 1057–1072.

DUFFIE, D., J. PAN, AND K. SINGLETON (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68, 1343–1376.

EGLOFF, D., M. LEIPPOLD, AND L. WU (2010): "The term structure of variance swap rates and optimal variance swap investments," *Journal of Financial and Quantitative Analysis*, 45, 1279–1310.

ERAKER, B., AND I. SHALIASTOVICH (2008): "An equilibrium guide to designing affine pricing models," *Mathematical Finance*, 18, 519–543.

FAN, J., Y. FAN, AND J. JIANG (2007): "Dynamic integration of time-and state-domain methods for volatility estimation," *Journal of the American Statistical Association*, 102, 618–631.

FAN, J., AND Y. WANG (2008): "Spot volatility estimation for high-frequency data," *Statistics and its Interface*, 1, 279–288.

GHYSELS, E., P. MYKLAND, AND E. RENAULT (2010): "In-sample asymptotics and across-sample efficiency gains for high frequency data statistics," Discussion Paper, University of Chicago and UNC.

GLOTER, A., AND J. JACOD (2000): "Diffusions with Measurement Errors: I - Local Asymptotic Normality and II - Optimal Estimators," Discussion paper, Université de Paris-6.

JACOD, J., AND P. PROTTER (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (2002): "Bayesian analysis of stochastic volatility models," *Journal of Business and Economic Statistics*, 20, 69–87.

JIANG, G., AND Y. TIAN (2005): "The model-free implied volatility and its information content," *Review of Financial Studies*, 18, 1305–1342.

KOLMOGOROV, A., AND Y. ROZANOV (1960): "On strong mixing conditions for stationary Gaussian processes," *Theory of Probability and Its Applications*, 5, 204–208.

KRISTENSEN, D. (2010): "Nonparametric filtering of the realized spot volatility: a kernel-based approach," *Econometric Theory*, 26, 60–93.

KUERSTEINER, G. M., AND I. R. PRUCHA (2013): "Limit theory for panel data models with cross sectional dependence and sequential exogeneity," *Journal of Econometrics*, (forthcoming).

LEE, S.-W., AND B. E. HANSEN (1994): "Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator," *Econometric theory*, 10, 29–29.

LUMSDAINE, R. L. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models," *Econometrica*, 64, 575–596.

MALLIAVIN, P., AND M. MANCINO (2005): "A Fourier transform method for nonparametric estimation of multivariate volatility," *Annals of Statistics*, 37, 1983–2010.

MUELLER, P., A. VEDOLIN, AND H. ZHOU (2011): "Short-run bond risk premia," Available at SSRN `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1851854`.

MYKLAND, P. A., AND L. ZHANG (2008): "Inference for volatility-type objects and implications for hedging," *Statistics and its Interface*, 1, 255–278.

NELSON, D. B. (1990): "ARCH models as diffusion approximations," *Journal of Econometrics*, 45, 7–38.

———— (1996): "Asymptotic filtering theory for multivariate ARCH models," *Journal of Econometrics*, 71, 1–47.

NELSON, D. B., AND D. P. FOSTER (1994): "Asymptotic filtering theory for univariate ARCH models," *Econometrica*, 62, 1–41.

PROTTER, P. (2004): *Stochastic Integration and Differential Equations*. Springer.

STOCK, J., AND M. WATSON (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

STROOCK, D., AND S. VARADHAN (1979): *Multidimensional diffusion processes*. Springer.

TODOROV, V. (2010): "Variance risk-premium dynamics: The role of jumps," *Review of Financial Studies*, 23, 345–383.

WANG, H., H. ZHOU, AND Y. ZHOU (2013): "Credit default swap spreads and variance risk premia," *Journal of Banking and Finance* (forthcoming).

ZHANG, L. (2001a): "From martingales to ANOVA: implied and realized volatility," Ph.D. thesis, Department of Statistics, University of Chicago.

———— (2001b): "From Martingales to ANOVA: Implied and Realized Volatility," The University of Chicago: Ph.D. dissertation, Department of Statistics.

ZHANG, L. (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043.

ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

ZHAO, Z., AND W. WU (2008): "Confidence bands in nonparametric time series regression," *Annals of Statistics*, 36, 1854–1878.

ZHOU, H. (2010): "Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty," Available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1400049`.