Conditional alphas and realized betas

Valentina Corradi
University of Warwick

Walter Distaso
Imperial College London

Marcelo Fernandes
Queen Mary University of London

Abstract: This paper proposes a two-step procedure to back out the conditional alpha of a given stock from high-frequency returns. We first estimate the realized factor loadings of the stock, and then retrieve the conditional alpha by estimating the conditional expectation of the stock return in excess over the realized risk premia. The estimation method is fully nonparametric in stark contrast with the literature on conditional alphas and betas. Apart from the methodological contribution, we employ NYSE data to determine the main drivers of conditional alphas as well as to track mispricing over time. In addition, we assess economic relevance of our conditional alpha estimates by means of a market-neutral trading strategy that longs stocks with positive alphas and shorts stocks with negative alphas. The preliminary results are very promising.

Keywords: asset pricing, conditional CAPM, realized beta, risk-adjusted performance.

Acknowledgments: We thank the valuable comments by seminar participants at Rotterdam Erasmus University, Nonlinear and Financial Econometrics Conference: A Tribute to A. Ronald Gallant (Toulouse, May 2011), and Meeting of the Brazilian Finance Society (Rio de Janeiro, July 2011). We are also very grateful to Christian Brownlees for providing his realized beta data. The usual disclaimer applies.

1 Introduction

The unconditional CAPM does not provide a good description of equity markets. Apart from the well-known market anomalies, nonzero alpha may arise even if the CAPM holds period by period. Put differently, the absence of pricing error does not suffice to ensure a zero unconditional alpha. This happens because time variation in the betas may correlate with market volatility and/or with the risk premia. For instance, allowing for time-varying betas helps explain most of the (unconditional) value premium given that value stocks are riskiest precisely when risk premium is higher, namely, in recession times (Petkova and Zhang, 2005; Zhang, 2005).

The usual fix is to assume that the conditional alphas and betas are affine on pre-determined predictor variables, e.g., stock characteristics, interest rates and spreads as well as other business-cycle indicators. Ferson, Simin and Sarkissian (2008) discuss three stylized factors that emerge from this literature. First, market betas do vary over time in a significant manner (Shanken, 1990; Cochrane, 1996; Jagannathan and Wang, 2006; Lettau and Ludvigson, 2001; Santos and Veronesi, 2005). Second, the constant of the conditional alpha term is smaller than the unconditional alpha. This means that conditional asset pricing models entail on average smaller pricing errors than their unconditional versions. Third, despite of their better fit, conditional asset pricing models still fail in that conditional alphas are not only nonzero, but also time-varying. See, among others, Christopherson, Ferson and Glassman (1998), Wang (2002), Ang and Chen (2007), and Adrian and Franzoni (2005).

This paper goes in search for these pricing errors, but using a much more robust and flexible framework. We build on the realized beta technology (Barndorff-Nielsen and Shephard, 2004) to come up with a two-step procedure to estimate conditional alphas. In the first stage, we employ high-frequency data to retrieve a stock's realized beta or, in general, any other risk factor loadings. The second step then backs out the conditional alpha by estimating the conditional expectation of the stock risk-adjusted return (i.e., the return in excess over the realized risk premia) at the low frequency. The resulting estimator is nonparametric and hence as flexible and robust as it gets. It does not require conditional alphas and betas to depend linearly on the conditioning state variables, reducing misspecification risks in a substantial manner. Although we focus essentially on the conditional CAPM case, our framework is general enough to include any multifactor model for asset returns. The conditional CAPM with higher-order moments ensues as one includes additional powers of the S&P 500 index returns, whereas adding exchange-traded funds (ETFs) based on size and book-to-market considerations would entail a conditional Fama-French three-factor model. Alternatively, we could also think of continuous and discontinuous market betas as in Todorov and Bollerslev (2010).

Integrating observations at different sampling frequencies is now new to finance. Merton (1980) notes that one can accurately estimate the variance over a fixed interval of time by summing squared returns at a sufficiently high sampling frequency. French, Schwert and Stambaugh (1987) and Ghysels, Santa Clara and Valkanov (2005) thus exploit daily (squared) returns to estimate the monthly volatility in their search for the risk-return tradeoff, whereas the realized measure literature employs intraday returns to compute daily realized variances, covariances, and market betas (Barndorff-Nielsen and Shephard, 2004; Ait-Sahalia and Mykland, 2009; Andersen, Bollerslev and Diebold, 2009; Ait-Sahalia, Fan and Xiu, 2011) as well as to conduct statistical inference for *parametric* continuous-time stochastic volatility models (Bollerslev and Zhao, 2002; Corradi and Distaso, 2006; Todorov, 2009).

More recently, Chang, Kim and Park (2009) combine low- and high-frequency observations to estimate continuous-time factor pricing models with constant factor loadings. They argue that standard regression results are senseless for two reasons. First, asset returns are too volatile at the high frequency. Second, stochastic volatility processes are nonstationary and endogenous due to leverage effects. They show that a random sampling scheme based on the market volatility circumvents these issues as long as the endogenous nonstationarity is completely driven by the market volatility. They estimate the latter, as well as the other stochastic volatility components, using high frequency data, whereas they run the multifactor regressions at the low frequency to alleviate the excessive volatility at the high frequency.

There are also other nonparametric alternatives to estimate conditional alphas and betas in the literature. Lewellen and Nagel (2006) and Kristensen and Ang (2009) estimate time-varying alphas and betas by taking local averages in time either through rolling windows or through a kernel-based approach. We differ by assuming that both alphas and betas are measurable functions of conditioning state variables and hence predictable. Note however that the realized beta measures already suffice to estimate the conditional alpha and hence we make no attempt to pin down conditional betas as well. Although it is certainly tricky to nail down what drives the conditional alphas, it allows for a more-than-descriptive analysis of pricing errors. Apart from looking at the main features of the time-varying alphas (e.g., persistence), we can also ask all sort of interesting questions about alpha portability and cross-sectional variation in the impact of the predictor variables.

We first ask what are the main forces driving mispricings in the New York Stock Exchange (NYSE). In contrast to Welch and Goyal (2008), we work at the daily frequency, ruling out many of the usual suspects for the conditioning state variables. Accordingly, we employ as lagged predictors various interest rates and spreads, market liquidity measures as well as characteristic-based portfolios based on momentum, long-term reversal, short-term reversal, size, and value effects. We examine how their cross-sectional average partial effects on the conditional alphas change over time for a sample of 9 actively traded stocks. ANTICIPATE EMPIRICAL RESULTS.

We then investigate whether pricing errors are persistent. If stock markets are indeed near efficient (Grossman and Stiglitz, 1980), pricing errors should not persist for long. ANTICIPATE EMPIRICAL RESULTS.

Finally, we assess whether it is possible to exploit these pricing errors within a simple trading strategy. In particular, we take long positions in stocks with positive conditional alphas and shorts stocks with negative conditional alphas. The weights are such that the resulting portfolio is neutral with respect to the S&P 500 index. Note that, as both alphas and betas change daily, we have to update the portfolio weights every day. This raises the issue of whether the portfolio alpha is portable after controlling for transaction costs. Persistence in alphas and betas thus plays a major role, for otherwise portfolio rebalancing would cost too dearly. ANTICIPATE RESULTS: conditional and unconditional alphas vs realized beta of the portfolio! justify any difference using Ferson's (2009) argument that superior information not necessarily entail a portfolio with positive (conditional) alpha.

The rest of this paper is as follows. Section 2 spells out the assumptions we make on the continuous-time multivariate process that governs the dynamics of asset prices. Note that we do not start from the continuous-time version of the CAPM as in Mykland and Zhang (2006), for otherwise the CAPM would not hold in discrete time (Longstaff, 1989). Section 3 develops the asymptotic justification of the conditional alpha estimator, controlling for the fact that we only observe the realized beta and not the true conditional beta. Section 4 investigates whether there are persistent pricing errors in actively traded stocks on the NYSE. Section 5 offers some concluding remarks, whereas the Appendix collects all technical proofs.

2 Conditional factor model: From continuous to discrete time

Given that the estimation procedure rests on high-frequency data and infill asymptotic theory, we must think carefully about the underlying continuous-time process. In what follows, we show how a discrete-time conditional multifactor asset pricing model may ensue from the exact discretization of a conditional semimartingale process in continuous time. Our discretization results complement well those in Longstaff (1989) and Chang et al. (2009). The former shows that temporally aggregating the continuous-time CAPM results in a multifactor model in discrete time, whereas the latter paper considers a continuous-time model that is also consistent with a discrete-time multifactor model. The main difference is that our setting delivers conditional alphas and betas, whereas pricing errors and factor loadings are constant in theirs.

Let $P_i(s)$ and $\boldsymbol{F}(s)$ respectively denote the log-prices at time s of the i-th asset ($i = 1, \ldots, N$) and of k portfolios representing common risk factors that drive the assets' excess returns. For instance, the CAPM considers the market portfolio as a single factor, whereas Fama and French (1992) also include portfolios based on size and book-to-market effects. We assume that both $P_i(s)$ and $\boldsymbol{F}(s)$ follow continuous-time diffusion processes, with drift and volatility parameters evolving in discrete time as measurable functions of conditioning factors $\boldsymbol{C}_t$. One may think of the latter as conditioning state variables that reflect changes in the future investment opportunity set as in Merton's (1973) ICAPM, for example. More precisely, for any $t \leq s < t+1$ and $i \in \{1, \ldots, N\}$,

$$\mathrm{d}P_i(s) = \mu_{i,t}\,\mathrm{d}s + \boldsymbol{\Sigma}_{i,t}'\,\mathrm{d}\boldsymbol{W}_F(s) + \sigma_{i,t}\,\mathrm{d}W_i(s) \tag{1}$$

$$\mathrm{d}\boldsymbol{F}(s) = \boldsymbol{\mu}_{F,t}\,\mathrm{d}s + \boldsymbol{\Sigma}_{F,t}\,\mathrm{d}\boldsymbol{W}_F(s), \tag{2}$$

where $\mu_{i,t} \equiv \mu_i(\boldsymbol{C}_t)$ and $\boldsymbol{\mu}_{F,t} \equiv \boldsymbol{\mu}_F(\boldsymbol{C}_t)$ are drift parameters, $\boldsymbol{\Sigma}_{i,t} \equiv \boldsymbol{\Sigma}_i(\boldsymbol{C}_t)$ is a $k \times 1$ vector that determines the exposure of asset i to each risk factor, $\boldsymbol{\Sigma}_{F,t} \equiv \boldsymbol{\Sigma}_F(\boldsymbol{C}_t)$ is the $k \times k$ covariance matrix of the common risk factors, $\boldsymbol{W}_F(s)$ is a $k-$dimensional standard Brownian motion, and $W_i(s)$ is a standard Brownian motion independent of $\boldsymbol{W}_F(s)$. We next document under which conditions asset and factor prices are continuous-time semimartingale processes.

Lemma 1: Let $\boldsymbol{X}_i(s) = (P_i(s), \boldsymbol{F}(s))$ evolve as in (1) and (2). Let also $\boldsymbol{C}(s) = \boldsymbol{C}_t$ for any $s \in [t, t+1)$ and define the filtration $\mathcal{F}_C(s) = \sigma(\boldsymbol{C}(\tau),\ \tau \leq s)$ for $s > 0$. If $\boldsymbol{C}(s)$ is independent of both $W_i(s)$ and $\boldsymbol{W}_F(s)$, then $\boldsymbol{X}_i(s)$ is a conditional semimartingale with independent increments given $\mathcal{F}_C(s)$.

The assumption that $\boldsymbol{C}(s)$ is independent of $W_i(s)$ and $\boldsymbol{W}_F(s)$ implies that $\mathbb{E}\left[\boldsymbol{\Sigma}'_{i,t} W_i(s)\right] = 0$ and that $\mathbb{E}\left[\boldsymbol{\Sigma}_{F,t}\, \boldsymbol{W}_F(s)\right] = 0$. This does not imply however that $\boldsymbol{F}(s)$ and $\boldsymbol{C}(s)$ are independent. In fact, the common risk factors $\boldsymbol{F}(s)$ depend on the conditioning state variables $\boldsymbol{C}_t$ for any $s \in [t, t+1)$ through the drift and diffusion parameters. In addition, it follows from Lemma 1 that, given the value of $\boldsymbol{C}_t$, the continuous-time process $\boldsymbol{X}_i(s)$ has independent increments for any $s \in [t, t+1)$. This means that market microstructure effects are responsible for any autocorrelation pattern within the interval $[t, t+1)$, say, a day. We show nonetheless that, due to the dependence on the conditioning factors $\boldsymbol{C}_t$, there is genuine autocorrelation in the daily increments $\boldsymbol{x}_{i,t} = \int_t^{t+1} \mathrm{d}\boldsymbol{X}_i(s)$.

Suppose that we have M equidistant observations within a day: $P_{i,t+j/M}$ and $\boldsymbol{F}_{t+j/M}$ with $j = 0, \ldots, M-1$. We then define the vector of realized betas as

$$\widehat{\beta}^{(M)}_{i,t+1} = \left[\sum_{j=0}^{M-1} \left(\boldsymbol{F}_{t+\frac{j+1}{M}} - \boldsymbol{F}_{t+\frac{j}{M}}\right)\left(\boldsymbol{F}_{t+\frac{j+1}{M}} - \boldsymbol{F}_{t+\frac{j}{M}}\right)'\right]^{-1} \sum_{j=0}^{M-1} \left(\boldsymbol{F}_{i,t+\frac{j+1}{M}} - \boldsymbol{F}_{i,\frac{j}{M}}\right)\left(P_{t+\frac{j+1}{M}} - P_{t+\frac{j}{M}}\right). \quad (3)$$

Barndorff-Nielsen and Shephard (2004) show that, under very mild regularity conditions,

$$\operatorname{plim}_{M\to\infty} \widehat{\boldsymbol{\beta}}^{(M)}_{i,t+1} = \boldsymbol{\Sigma}^{-1}_{FF,t}\, \boldsymbol{\Sigma}_{F,t}\, \boldsymbol{\Sigma}_{i,t} \equiv \boldsymbol{\beta}_{i,t+1}, \quad (4)$$

where $\boldsymbol{\Sigma}_{FF,t} = \boldsymbol{\Sigma}_{F,t}\boldsymbol{\Sigma}'_{F,t}$. To simplify matters, we assume without any loss of generality that $\boldsymbol{F}_{t+j/M}$ are orthogonal risk factors and hence $\boldsymbol{\Sigma}_{FF,t}$ is diagonal. Note also that the conditioning state variables $\boldsymbol{C}_t$ are the only drivers of the daily factor loadings $\boldsymbol{\beta}_{i,t+1}$ in that $\boldsymbol{\beta}_{i,t+1} \equiv \boldsymbol{\beta}_i(\boldsymbol{C}_t)$.

We now move to the daily frequency by letting $r_{i,t+1} \equiv \int_t^{t+1} \mathrm{d}P_i(s)$ and $\boldsymbol{f}_{t+1} \equiv \int_t^{t+1} \mathrm{d}\boldsymbol{F}(s)$ denote the daily returns with continuous compounding on asset i and on the common factor portfolios over the time interval $[t, t+1)$, respectively. It then follows from (1) and (2) that

$$r_{i,t+1} = \mu_{i,t} + \sigma_{i,t} \int_t^{t+1} \mathrm{d}W_i(s) + \boldsymbol{\Sigma}_{i,t}' \int_t^{t+1} \mathrm{d}\boldsymbol{W}_F(s) \tag{5}$$

$$\boldsymbol{f}_{t+1} = \boldsymbol{\mu}_{F,t} + \boldsymbol{\Sigma}_{F,t} \int_t^{t+1} \mathrm{d}\boldsymbol{W}_F(s), \qquad t = 1, \ldots, T \tag{6}$$

This means that

$$\left(r_{i,t+1}, \boldsymbol{f}_{t+1}\right) | \boldsymbol{C}_t \sim \mathcal{N}\left(\begin{pmatrix} \mu_{i,t} \\ \boldsymbol{\mu}_{F,t} \end{pmatrix}, \begin{pmatrix} \sigma_{i,t}^2 + \boldsymbol{\Sigma}_{i,t}' \, \boldsymbol{\Sigma}_{i,t} & \boldsymbol{\Sigma}_{i,t}' \, \boldsymbol{\Sigma}_{F,t} \\ \boldsymbol{\Sigma}_{F,t} \, \boldsymbol{\Sigma}_{i,t} & \boldsymbol{\Sigma}_{FF,t} \end{pmatrix} \right)$$

and hence further conditioning on the risk factor portfolio returns entails

$$r_{i,t+1} | \boldsymbol{f}_{t+1}, \boldsymbol{C}_t \sim \mathcal{N}\left(\mu_{i,t} + (\boldsymbol{f}_t - \boldsymbol{\mu}_{F,t})' \boldsymbol{\Sigma}_{FF,t}^{-1} \boldsymbol{\Sigma}_{F,t} \, \boldsymbol{\Sigma}_{i,t}, \; \sigma_{i,t}^2 + \boldsymbol{\Sigma}_{i,t}' \boldsymbol{\Sigma}_{i,t} - \boldsymbol{\Sigma}_{i,t}' \boldsymbol{\Sigma}_{FF,t}^{-1} \boldsymbol{\Sigma}_{i,t} \right).$$

Using the definition of the true betas in the last equality of (4) then yields the following discrete-time factor model: $r_{i,t+1} = \alpha_{i,t} + \boldsymbol{f}_{t+1}' \boldsymbol{\beta}_{i,t} + \epsilon_{i,t+1}$, with $\alpha_{i,t} = \mu_{i,t} - \boldsymbol{\mu}_{F,t}' \boldsymbol{\Sigma}_{FF,t}^{-1} \boldsymbol{\Sigma}_{F,t} \boldsymbol{\Sigma}_{i,t} = \mu_{i,t} - \boldsymbol{\mu}_{F,t}' \beta_{i,t}$ and $\epsilon_{i,t+1} | \boldsymbol{f}_{t+1}, \boldsymbol{C}_t \sim \mathcal{N}\left(0, \sigma_{i,t}^2 + \boldsymbol{\Sigma}_{i,t}' \boldsymbol{\Sigma}_{i,t} - \boldsymbol{\Sigma}_{i,t}' \boldsymbol{\Sigma}_{FF,t}^{-1} \boldsymbol{\Sigma}_{i,t}\right)$.

Before showing how to estimate the conditional alphas, let us briefly stress the difference between the above setup and Mykland and Zhang's (2006) ANOVA for diffusions. The latter posits a single-factor model for asset prices in which

$$\mathrm{d}P_i(s) = \beta(s) \, \mathrm{d}F(s) + \mathrm{d}Z_i(s), \tag{7}$$

where $Z_i(s)$ is a residual process in continuous time. Given that beta is time-varying on a continuous scale in (7), Mykland and Zhang (2006) have to estimate it by taking a localized version of (3). In particular, they compute the ratio between the realized covariation and the realized variance over time intervals of order $M^{-1/2}$ (rather than over a single interval with M observations as in daily realized measures). They also estimate the residual process for the M equidistant intraday observations by

$$Z_{i,t+(j+1)/M} - Z_{i,t+j/M} = (P_{i,t+(j+1)/M} - P_{i,t+j/M}) - \widehat{\beta}_{j/M} \, (F_{i,t+(j+1)/M} - F_{i,t+j/M}).$$

Although this allows one to test whether the quadratic variation of the residual process is zero, it does not entail conditional alpha estimates due to the impossibility of consistently estimating nonzero drifts on a finite time span. The continuous-time CAPM formulation of Chang et al.'s (2009) is actually quite similar to (7), though with constant alphas and betas. In particular, they consider an increasing time span so as to regress $\int_{t_j}^{t_{j+1}} \mathrm{d}P_i(s)$ on $\int_{t_j}^{t_{j+1}} \mathrm{d}F(s)$, where $t_0, t_1, \ldots, t_n$ form a sequence of (possibly random) times. Checking whether the regression intercept is zero for every asset then entails a simple test for the validity of the unconditional CAPM.

3 Retrieving conditional alphas from realized betas

To estimate the realized factor loadings, we must first orthogonalize the factors $\boldsymbol{F}_t$ by taking linear combinations of the intraday returns on the risk factors, namely, $\breve{\boldsymbol{F}}_{m,t} = \boldsymbol{B}_t \boldsymbol{F}_{m,t}$ with $\breve{F}_{k,m,t} \perp \breve{F}_{\kappa,m,t}$ for all $1 \leq k \leq \kappa \leq K$ as well as for every instant m within day t (or week, whatever). Note that the rotation matrix $\boldsymbol{B}_t$ is known even if it changes every day and hence it is possible to recover the original (daily) factor loadings $\beta_{i,k,t}$ from the daily loadings of the orthogonal factors $\breve{\beta}_{i,k,t}$. We estimate the latter using the standard realized beta approach, yielding a realized loading for each orthogonal factor given by $\breve{\beta}_{i,k,t}^{(M)}$ and so a realized risk premium of $\sum_{k=1}^{K} \breve{\beta}_{i,k,t}^{(M)} \breve{F}_{k,t} = \sum_{k=1}^{K} \beta_{i,k,t}^{(M)} F_{k,t}$.

By subtracting the realized risk premia from the individual stock return, we find the realized counterpart of $Z_{i,t+1} = \alpha_{i,t} + \epsilon_{i,t+1}$, that is to say, $Z_{i,t+1}^{(M)} \equiv r_{i,t+1} - \sum_{k=1}^{K} \beta_{i,k,t}^{(M)} F_{k,t+1}$. Identification of the conditional alpha results from the fact that the conditional expectation of $\epsilon_{i,t+1}$ is zero, whereas $\alpha_{i,t}$ is measurable in the information set. It thus follows that $\alpha_{i,t} \equiv \mathbb{E}(Z_{i,t+1} | \boldsymbol{C}_t)$, where $\boldsymbol{C}_t$ is the vector of state variables. The second step of the procedure then amounts to estimating $\alpha_{i,t}^{(M)} = \mathbb{E}(Z_{i,t}^{(M)} | \boldsymbol{C}_t)$ using kernel methods. Note that there is an extensive list of state variables to include in $\boldsymbol{C}_t$ if we take the conditional alpha-beta literature seriously. This means that we should think about employing dimension-reduction techniques by imposing either an additive or a single-index dependence structure.

We use intraday observations on $P_i(s)$ and $\boldsymbol{F}_s$ to construct daily estimators $\beta_{i,t}$. Note that $\beta_{i,t}$ is not only time varying but it is also allowed to be a measurable function of $\boldsymbol{C}_t$. The scope for time varying betas, moving along market risk premia, expected dividend growth and fundamental risk, has been outlined by Santos and Veronesi (2004), within a general equilibrium framework. We then use the realized betas to risk adjust asset returns:

$$\begin{aligned}
\widehat{Z}_{i,t+1}^{(M)} &= r_{i,t+1} - \boldsymbol{f}_{t+1}' \widehat{\boldsymbol{\beta}}_{i,t}^{(M)} = r_{i,t+1} - \boldsymbol{f}_{t+1}' \boldsymbol{\beta}_{i,t} - \boldsymbol{f}_{t+1}' \left(\widehat{\boldsymbol{\beta}}_{i,t}^{(M)} - \boldsymbol{\beta}_{i,t} \right) \\
&= \alpha_{i,t} + \epsilon_{i,t+1} - \boldsymbol{f}_{t+1}' \left(\widehat{\boldsymbol{\beta}}_{i,t}^{(M)} - \boldsymbol{\beta}_{i,t} \right) = Z_{i,t+1} - \boldsymbol{f}_{t+1}' \left(\widehat{\boldsymbol{\beta}}_{i,t}^{(M)} - \boldsymbol{\beta}_{i,t} \right).
\end{aligned}$$

Our objective is to obtain a consistent estimator for $\alpha_{i,t}$ for all i, t. If we ignore the estimation error due to the realized betas, it is immediate to see that $\mathbb{E}\left(\epsilon_{i,t+1}|\boldsymbol{C}_t\right) = 0$. We therefore define $\alpha_{i,t}$ as the conditional expectation of the residual process $Z_{i,t+1} = r_{i,t+1} - \boldsymbol{f}'_{t+1}\boldsymbol{\beta}_{i,t}$ given $\boldsymbol{C}_t$, that is, $\alpha_{i,t} = \mathbb{E}\left(Z_{i,t+1}|\boldsymbol{C}t\right)$. We can then estimate $\alpha_{i,t}$ by standard Nadaraya-Watson kernel estimators. However, we do not observe $Z_{i,t+1}$, only observing $\widehat{Z}^{(M)}_{i,t+1} = r_{i,t+1} - \boldsymbol{f}'_{t+1}\widehat{\boldsymbol{\beta}}^{(M)}_{i,t}$. To retrieve $\alpha_{i,t}$, we thus construct a Nadaraya-Watson estimator using as a dependent variable $\widehat{Z}^{(M)}_{i,t}$. The next section provides the conditions on the rate of growth of the number of intraday observations under which the contribution of the estimation error is negligible.

Time varying alphas have been already considered. For example, Lewellen and Nagel (2006) estimate alphas and betas using short data windows. Kristensen and Ang (2009) estimate alphas and betas by kernel-weighted least squares. On the other hand, we explicitly model the alphas as a generic function of $\boldsymbol{C}_t$.

How we deal with data mining and spurious regressions due to persistent regressors (Ferson et al., 2008).

Typically, the dimension of the vector $\boldsymbol{C}_t$ is relatively high. In order, to circumvent the curse of dimensionality problem, as regressors we shall use the k largest principal components of $\boldsymbol{C}_t$. Indeed, there are several other methods for dimension reduction in a nonparametric setting, such as sliced inverse regression (Li, 1991), the single index model of Ichimura (1993) or group Lasso type criteria as in Huang, Horowitz and Wei (2010). However, identification of estimators based on these dimension reduction methods require that alpha is neither zero nor constant. While we conjecture that this is indeed the case for most stocks, we nevertheless want to allow for the possibility that some stock has a zero or constant alpha. For this reason, we simply rely on the use of principal components.

Let $C_t = (P_{1,t}, \ldots, P_{k,t})'$ be the k principal components of $\boldsymbol{C}_t$, $C_t = (P_{1,t}, ..., P_{k,t})'$ with $k < k_B$, say $k = 2, 3$. As $\mathbb{E}\left(\epsilon_{i,t+1}|\boldsymbol{C}_t\right) = 0$, then $\mathbb{E}\left(\epsilon_{i,t+1}|C_t\right) = 0$, and it is immediate to see that $\mathbb{E}\left(Z_{i,t+1}|C_t\right) = \alpha_{i,t+1}$. Define

$$\begin{aligned}
&\widehat{\alpha}_{i,T,t+1} = \widehat{m}_{i,T,M}(C_t) \\
&= \frac{\frac{1}{Th_T^k}\sum_{l=1}^{T-1} \widehat{Z}_{i,l+1,M}\boldsymbol{K}\left(\frac{C_\ell - C_t}{h}\right)}{\frac{1}{Th_T^k}\sum_{l=1}^{T-1} \boldsymbol{K}\left(\frac{C_\ell - C_t}{h_T}\right)} = \frac{\frac{1}{Th_T^k}\sum_{l=1}^{T-1} \widehat{Z}_{i,l+1,M}\boldsymbol{K}\left(\frac{C_\ell - C_t}{h}\right)}{\widehat{g}_T(C_t)}
\end{aligned} \tag{8}$$

As C_t is a random variable, we need a uniform result over its support. However, we cannot consistently estimate $m_i(c)$ over regions having very small density. Consider the following trimmed version of $\widehat{m}_{i,T,M}(c)$, (Section 3.2 in Andrews 1995)

$$tr_T\left(\widehat{m}_{i,T,M}(c)\right) = \widehat{m}_{i,T,M}(c)1\left\{c \in \widehat{G}_T(c)\right\}, \tag{9}$$

where $\widehat{G}_T(c) = \{c : \widehat{g}_T(c) > d_T\}$, where $d_T \to 0$ at an appropriate rate. In the sequel, we shall establish the uniform consistency of $\widehat{m}_{i,T,M}(C_t)$, by showing that

$$\left(\int_{\mathbb{R}^k} \left|tr_T\left(\widehat{m}_{i,T,M}(c)\right) - m_i(c)\right|^Q g(c)\mathrm{d}(c)\right)^{1/Q} = o_p(1), \tag{10}$$

where g is the density of C_t.

4 Main results

In the sequel, we rely on the following Assumption.

Assumption A:

(i) For $i = 1,..,N$, $\mu_{i,t}, \sigma_{i,t}$, and the elements of $\boldsymbol{\Sigma}_{iF^B,t}, \mu_{F^B,t}, \boldsymbol{\Sigma}_{F^B,t}$, as defined in (1)-(**??**), are $F^B_{t-1}-$measurable, and are $\beta-$dominated, with β defined in (ii).[1]

(ii) For $i = 1,..,N$, $\mathbb{E}\left(|Z_{i,t}|^\beta\right) < \infty$, with $\beta > 2$, $(Z_{i,t+1}, \boldsymbol{C}_t)$, $t = 1,...,T-1$ is strictly stationary and $\alpha-$mixing with mixing coefficients $\alpha(j)$, such that $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\beta-2}{\beta}} < \infty$.

(iii) $\boldsymbol{C}_t$ has a distribution which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{k_B}$, which density ϕ. $\phi(y)$ is bounded, and twice continuously differentiable, with bounded derivatives.

(iv) The kernel function $\boldsymbol{K}$ is a bounded density on $\mathbb{R}^k$, such that $\int_{\mathbb{R}^k} x\boldsymbol{K}(x)\mathrm{d}x = 0$, $\int_{\mathbb{R}^k} x^2\boldsymbol{K}(x)\mathrm{d}x < \infty$. Also $\boldsymbol{K}$ has absolutely integrable characteristic function $\Psi(u)$.[2]

(v) Let $m_i(c) = \mathbb{E}\left(Z_{i,t+1}|C_t = c\right)$ does not depend on t. $m_i(c)g(c)$ is bounded and twice continuously differentiable on $\mathbb{R}^k$, with bounded derivatives, for $i = 1,..,N,$.

(vi) $\sup_{c\in\mathbb{R}^k} |m_i(c)| < \infty$, for some $0 < a < 1$, and for $i = 1,..,N$, $\int_{\mathbb{R}^k} g(c)^{1-a}\mathrm{d}c < \infty$.

Assumption A(ii)-(iii) are stated in terms of the "usual suspects" $\boldsymbol{C}_t$; nevertheless the statistic in (10) is constructed using the first k principal components of $\boldsymbol{C}_t$. As each principal component

[1] $\mu_{i,t}$ is said to be $\beta-$dominated, if $|\mu_{i,t}| \leq D_t$, and $\mathbb{E}\left(D_t^\beta\right) < \infty$ (see e.g. Gallant and White 1988).

[2] Note that $\Psi(u) = \int_{\mathbb{R}^k} \exp(iux)\boldsymbol{K}(x)\mathrm{d}x$, and we require $\int_{\mathbb{R}^k} |\Psi(u)|\mathrm{d}u < \infty$.

is a linear combination of the k_B elements of $\boldsymbol{C}_t$, and as $k_B < \infty$, it follows straightforwardly that C_t is also $\alpha-$mixing, of the same size as $\boldsymbol{C}_t$. Hence, A(ii) holds also for C_t. The density of each individual principal component is obtained via the direct convolution of the k_B marginals and is therefore absolutely continuous on $\mathbb{R}$. As principal components are mutually orthogonal, their joint density is absolutely continuous on $\mathbb{R}^k$. Hence, A(iii) ensures that C_t has a density absolutely continuous on $\mathbb{R}^k$.

As pointed out by Bierens (1983), Assumption A1(iv) is satisfied by a multivariate standard normal kernel, or by the product of k univariate standard normal kernels. A(v)-A(vi) are conditions about the smoothness and the boundedness of the regression function, and on the thickness of the tails of the regressors. For example, in the case of $g(c)$ multivariate normal, the condition $\int_{\mathbb{R}^k} g(c)^{1-a}\mathrm{d}c < \infty$ is satisfied for any a arbitrary close to 1.

Define the infeasible estimator,

$$\left(\int_{\mathbb{R}^k} \left|tr_T\left(\widetilde{m}_{i,T}(c)\right) - m_i(c)\right|^Q g(c)\mathrm{d}(c)\right)^{1/Q},$$

where

$$\widetilde{\alpha}_{i,T,t+1} = \widetilde{m}_{i,T}(C_t)$$
$$= \frac{\frac{1}{Th_T^k}\sum_{l=1}^{T-1} Z_{i,l+1}\boldsymbol{K}\left(\frac{C_\ell - C_t}{h_T}\right)}{\frac{1}{Th_T^k}\sum_{l=1}^{T-1}\boldsymbol{K}\left(\frac{C_\ell - C_t}{h_T}\right)} = \frac{\frac{1}{Th_T^k}\sum_{l=1}^{T-1} Z_{i,l+1}\boldsymbol{K}\left(\frac{C_\ell - C_t}{h_T}\right)}{\widehat{g}_T(C_t)}$$

We first show that the difference between the feasible and the infeasible estimator is $o_p(1)$.

Proposition 1: Let the conditions of Lemma 1 hold, and let Assumption A(i)-(iv) hold. Then, if, as $T, M, h_T^{-1} \to \infty$, $d_T^{-1}M^{-1/2}h_T^{-k} \to 0$,

$$tr_T\left(\widetilde{m}_{i,T,M}(c)\right) - tr_T\left(\widehat{m}_{i,T,M}(c)\right) = o_p(1),$$

where the $o_p(1)$ terms holds uniformly in $G_T(c)$, as defined in (9).

It is immediate to see that M can grow at a slower rate than T. This is empirically important. It should be note that the conditions on the rate of growth of M are much weaker than in Theorem 1 in Corradi and Swanson (2009). This is due to the fact the the estimation error affects only dependent variable, and it does not enter in the kernel function.

Proposition 2: Let the conditions of Lemma 1 hold, and let Assumption A(ii)-(vi) hold. Then, for $0 < Q < \infty$,

$$\left(\int_{\mathbb{R}^k} |tr_T\left(\widehat{m}_{i,T,M}(c)\right) - m_i(c)|^Q g(c)\mathrm{d}c\right)^{1/Q}$$
$$= O_p\left(d_T^{-1}M^{-1/2}h_T^{-k}\right) + O_p\left(T^{-1/2}h_T^{-k}d_T^{-2}\right) + O\left(h_T^2 d_T^{-2}\right) + O_p\left(d_T^{a/Q}\right),$$

where a is defined in A(vi).

5 Information Ratio

Define the information ratio as

$$IR_{i,t+1} = IR_i(C_t) = \frac{\mathbb{E}\left(Z_{i,t+1}|C_t\right)}{\sqrt{\mathrm{Var}\left(Z_{i,t+1}|C_t\right)}} = \frac{\alpha_{i,t+1}}{\sqrt{\mathbb{E}\left(\epsilon_{i,t+1}^2|C_t\right)}},$$

where $\epsilon_{i,t} = Z_{i,t} - \alpha_{i,t}$, and define its estimator as

$$\widehat{IR}_{i,t+1,M} = \widehat{IR}_{i,T,M}(C_t) = \frac{\widehat{m}_{i,T,M}(C_t)}{\sqrt{\widehat{m}_{i,T,M}^{(2)}(C_t) - \widehat{m}_{i,T,M}^2(C_t)}},$$

where $\widehat{m}_{i,T,M}(C_t)$ is defined in (8), and

$$\widehat{m}_{i,T,M}^{(2)}(C_t) = \frac{\frac{1}{Th_T^k}\sum_{l=1}^{T-1}\widehat{Z}_{i,l+1,M}^2 \boldsymbol{K}\left(\frac{C_\ell - C_t}{h}\right)}{\widehat{g}_T(C_t)},$$

with $\widehat{g}_T(C_t)$ defined in (8). In the sequel, we need a slightly strengthened version of Assumption A.

Assumption A':
(i') For $i = 1,..,N$, $\mathbb{E}\left(|Z_{i,t}|^{2\beta}\right) < \infty$, with $\beta > 2$, $(Z_{i,t+1}, \boldsymbol{C}_t)$, $t = 1,...,T-1$ is strictly stationary and $\alpha-$mixing with mixing coefficients $\alpha(j)$, such that $\sum_{j=1}^{\infty}\alpha(j)^{\frac{\beta-2}{\beta}} < \infty$.
(ii') Let $m_i^{(2)}(c) = \mathbb{E}\left(Z_{i,t+1}^2|C_t = c\right)$ does not depend on t. $m_i^{(2)}(c)g(c)$ is bounded and twice continuously differentiable on $\mathbb{R}^k$, with bounded derivatives, for $i = 1,..,N$.
(iii') $\sup_{c\in\mathbb{R}^k}\left|m_i^{(2)}(c)\right| < \infty$, for some $0 < a < 1$, and for $i = 1,..,N$, $\int_{\mathbb{R}^k} g(c)^{1-a}\mathrm{d}c < \infty$.

Define

$$tr_T\left(\widehat{IR}_{i,T,M}(c)\right) = \widehat{IR}_{i,T,M}(c)1\left\{c \in \widehat{G}_T(c)\right\},$$

where $\widehat{G}_T(c)$ is defined as in (9).

Proposition 3: Let the conditions of Lemma 1 hold, and let Assumption A(ii)-(vi) and A'(ii')-(vi') hold. Then, for $0 < Q < \infty$,

$$\left(\int_{\mathbb{R}^k} \left| tr_T\left(\widehat{IR}_{i,T,M}(c)\right) - IR_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q}$$
$$= O_p\left(d_T^{-1}M^{-1/2}h_T^{-k}\right) + O_p\left(T^{-1/2}h_T^{-k}d_T^{-2}\right) + O\left(h_T^2 d_T^{-2}\right) + O_p\left(d_T^{a/Q}\right),$$

where a is defined in A(vi').

References

Adrian, T. and Franzoni, F. (2005). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM, *Journal of Empirical Finance* **16**: 537–556.

Ait-Sahalia, Y., Fan, J. and Xiu, D. (2011). High frequency covariance estimates with noisy and asynchronous financial data, *Journal of the American Statistical Association* **105**: 15041517.

Ait-Sahalia, Y. and Mykland, P. (2009). Estimating volatility in the presence of market microstructure noise: A review of the theory and practical considerations, *in* Thomas Mikosch et al. (ed.), *Handbook of Financial Time Series*, Springer-Verlag.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2009). Parametric and non-parametric volatility measurements, *in* Lars P. Hansen and Yacine Ait-Sahalia (ed.), *Handbook of Financial Econometrics*, North Holland, New York.

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models, *Econometric Theory* **11**: 560–596.

Ang, A. and Chen, J. (2007). CAPM over the long run: 1926–2001, *Journal of Empirical Finance* **14**: 1–40.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics, *Econometrica* **72**: 885–925.

Bierens, H. J. (1983). Uniform consistency of a regression function under generalized conditions, *Journal of the American Statistical Association* **78**: 699–707.

Bollerslev, T. and Zhao, H. (2002). Estimating stochastic volatility diusions using conditional moments of integrated volatility, *Journal of Econometrics* **109**: 33–65.

Chang, Y., Kim, H. and Park, J. Y. (2009). Evaluating factor pricing models using high frequency panels. working paper, Indiana University, Texas A&M University and Sungkyunkwan University.

Christopherson, J. A., Ferson, W. E. and Glassman, D. (1998). Conditioning manager alpha on economic information: Another look at the persistence of performance, *Review of Financial Studies* **11**: 111–142.

Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* **104**: 572–621.

Corradi, V. and Distaso, W. (2006). Semiparametric comparison of stochastic volatility models using realized measures, *Review of Economic Studies* **73**: 635–667.

Corradi, Valentina anf Distaso, W. and Swanson, N. R. (2009). Predictive density estimators for daily volatility based on realized measures, *Journal of Econometrics* **150**: 119–138.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns, *Journal of Finance* **47**: 427–465.

Ferson, W. (2009). Alpha and the problem of performance measurement. working paper, University of Southern California.

Ferson, W. E., Simin, T. and Sarkissian, S. (2008). Asset pricing models with conditional alphas and betas: The effects of data snooping and spurious regression, *Journal of Financial and Quantitative Analysis* **43**: 331–354.

French, K. R., Schwert, W. and Stambaugh, R. F. (1987). Expected stock returns and volatility, *Journal of Financial Economics* **19**: 3–29.

Ghysels, E., Santa Clara, P. and Valkanov, R. (2005). There is a risk-return tradeoff after all, *Journal of Financial Economics* **76**: 509–548.

Grossman, S. J. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets, *American Economic Review* **70**: 393408.

Huang, J., Horowitz, J. L. and Wei, F. (2010). Variable selection in nonparametric additive models, *Annals of Statistics* **38**: 2282–2313.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models, *Journal of Econometrics* **58**: 71–120.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable converge in law, *Seminaire de Probabilites XXXI*, Vol. 1635, Springer Verlag, New York.

Jagannathan, R. and Wang, W. (2006). The conditional CAPM and the cross-section of stock returns, *Journal of Finance* **51**: 3–53.

Kristensen, D. and Ang, A. (2009). Testing conditional factor models. working paper, Columbia University.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth and expected stock returns, *Journal of Finance* **56**: 815–849.

Lewellen, J. and Nagel, S. (2006). The conditional CAPM does not explain asset pricing anomalies, *Journal of Financial Economics* **82**: 289–314.

Li, K. C. (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association* **86**: 316–327.

Longstaff, F. A. (1989). Temporal aggregation and the continuous-time capital asset pricing model, *Journal of Finance* **44**: 871–887.

Merton, R. C. (1973). An intertemporal capital asset pricing model, *Econometrica* **41**: 867–887.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* **8**: 323–361.

Mykland, P. A. and Zhang, L. (2006). ANOVA for diffusions and Ito processes, *Annals of Statistics* **34**: 1931–1963.

Petkova, R. and Zhang, L. (2005). Is value riskier than growth?, *Journal of Financial Economics* **78**: 187–202.

Santos, T. and Veronesi, P. (2004). Conditional betas. working paper, University of Chicago.

Santos, T. and Veronesi, P. (2005). Labor income and predictable stock returns, *forthcoming in the Review of Financial Studies* .

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation, *Journal of Econometrics* **45**: 99–120.

Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data, *Journal of Econometrics* **148**: 131–148.

Todorov, V. and Bollerslev, T. (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks, *Journal of Econometrics* **157**: 220–235.

Wang, K. (2002). Asset pricing with conditioning information: A new test, *Journal of Finance* **58**: 161–196.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**: 1455–1508.

Zhang, L. (2005). The value premium, *Journal of Finance* **60**: 67–103.

Appendix

Proof of Lemma 1: Define the filtered probability space $\mathcal{B} = \left(\Omega^B, \mathcal{F}^B, \left(\mathcal{F}_s^B\right)_{s\geq 0}, \mathcal{P}^B\right)$, where $\mathcal{F}_s^B = \sigma\left(F_\tau^B,\ \tau \leq s\right)$, $s \in \mathbb{R}_+$, $F_\tau^B = \boldsymbol{C}_t$ for $\tau \in [t, t+1)$. Also, define the filtered probability space $\mathcal{A}^i = \left(\Omega^{A,i}, \mathcal{F}^{A,i}, \left(\mathcal{F}_s^{A,i}\right)_{s\geq 0}, \mathcal{P}^{A,i}\right)$ where $\mathcal{F}_s^{A,i} = \sigma\left(W_\tau,\quad \tau \leq s\right)$, with $W_\tau = \left(W_{i,\tau}, W_{F^A,\tau}\right)$. Given the independence of F_τ^B of W_τ, we can defined the enlarged filtered probability space $\widetilde{\mathcal{B}} = \left(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \left(\widetilde{\mathcal{F}}_s\right)_{s\geq 0}, \widetilde{\mathcal{P}}\right)$, where

$$\widetilde{\Omega} = \Omega^B \times \Omega^{A,i}, \quad \widetilde{\mathcal{F}} = \mathcal{F}^B \otimes \mathcal{F}^{A,i}$$

$$\widetilde{\mathcal{F}}_s = \cap_{\tau > s} \mathcal{F}_\tau^B \otimes \mathcal{F}_\tau^{A,i} \quad \textit{and} \quad \widetilde{\mathcal{P}}\left(\mathrm{d}\omega^B \mathrm{d}\omega^{A,i}\right) = \mathcal{P}^B\left(\mathrm{d}\omega^B\right) \mathcal{P}^{A,i}\left(\mathrm{d}\omega^{A,i}\right),$$

with $\omega^B \in \Omega^B$, $\omega^{A,i} \in \Omega^{A,i}$, and with $\otimes$ denoting the product measure. Now, $X_i(s) = (P_i(s), \boldsymbol{F}_A(s))$ can be defined on the enlarged filtered probability space $\widetilde{\mathcal{B}}$, and in fact $X_i(s)$ is $\widetilde{\mathcal{F}}_s-$measurable. Since, F_τ^B is independent of W_τ, it also follows that all measurable function of F_τ^B are independent of W_τ. Let $\boldsymbol{\Sigma}_{A,i,t} = \left(\mu_{i,t}, \mu_{F,t}, \sigma_{i,t}, \boldsymbol{\Sigma}_{iF^A,t}, \boldsymbol{\Sigma}_{F^A,t}\right)$, and define $\boldsymbol{\Sigma}_{A,i,\tau} = \boldsymbol{\Sigma}_{A,i,t}$ for $t \leq \tau < t+1$, and note that $\boldsymbol{\Sigma}_{A,i,\tau}$ is also independent of W_τ. Thus, for each $\omega^B \in \Omega^B$, except of a set of $\mathcal{P}^B-$zero probability, $X_i(s)$ is a conditional semimartingale with independent increments (Jacod, 1997). ■

Proof of Proposition 1: Given A(iv), $\boldsymbol{K}$ is a bounded density function, and so

$$\left|\mathrm{tr}\left(\widetilde{m}_{i,T,M}(c)\right) - \mathrm{tr}\left(\widehat{m}_{i,T,M}(c)\right)\right| \leq \Delta d_T^{-1} \left| \frac{1}{Th^k} \sum_{l=1}^{T-1} f_{l+1}^{A\prime} \left(\widehat{\beta}_{i,l+1,M} - \beta_{i,l+1}\right) \right|.$$

Let $f_{1,l}$ denotes the first component of f_ℓ^A, as f_ℓ^A are orthogonal factors, $\widehat{\beta}_{i,l+1,M}^{(1)}$, the first component of the vector $\widehat{\beta}_{i,l+1,M}$, is given by

$$\widehat{\beta}_{i,l+1,M}^{(1)} = \frac{\sum_{j=0}^{M-1} \left(P_{i,l+(j+1)/M} - P_{i,l+j/M}\right)\left(f_{1,l+(j+1)/M} - f_{1,l+j/M}\right)}{\sum_{j=0}^{M-1} \left(f_{1,l+(j+1)/M} - f_{1,l+j/M}\right)^2}, \tag{11}$$

and as $\boldsymbol{\Sigma}_{F_B,l}$ is a diagonal matrix,

$$\beta_{i,l+1}^{(1)} = \frac{\sigma_{i,1,F^B,l}\sigma_{F_B,11,l}}{\sigma_{F_B,11,l}^2}, \tag{12}$$

where $\sigma_{i,1,F^B,l}$ is the first element of $\boldsymbol{\Sigma}_{i,F^B,l}$, $\sigma_{F_B,11,l}$, $\sigma_{F_B,11,l}^2$ are the $1,1-$th element of $\boldsymbol{\Sigma}_{F_B,l}$ and $\boldsymbol{\Sigma}_{F_B,l} \times \boldsymbol{\Sigma}_{F_B,l}$, respectively. Thus, it is enough to show that $\frac{1}{Th^k}\sum_{l=1}^{T-1} f_{1,l+1}\left(\widehat{\beta}_{i,l+1,M}^{(1)} - \beta_{i,l+1}^{(1)}\right) = o_p\left(d_T\right)$.

$$\begin{aligned}
&\frac{1}{Th^k}\sum_{l=1}^{T-1} f_{1,l+1}\left(\widehat{\beta}_{i,l+1,M}^{(1)} - \beta_{i,l+1}^{(1)}\right) \\
&= \frac{1}{Th^k}\sum_{l=1}^{T-1} \mu_{f_1}\left(\widehat{\beta}_{i,l+1,M}^{(1)} - \beta_{i,l+1}^{(1)}\right) + \frac{1}{Th^k}\sum_{l=1}^{T-1} \left(f_{1,l+1} - \mu_{f_1}\right)\left(\widehat{\beta}_{i,l+1,M}^{(1)} - \beta_{i,l+1}^{(1)}\right) \\
&= I_{T,M,h} + II_{T,M,h},
\end{aligned} \tag{13}$$

where $\mu_{f_1} = \mathbb{E}(f_{1,l})$. Given (11) and (12), $\widehat{\beta}^{(1)}_{i,l+1,M} - \beta^{(1)}_{i,l+1}$ writes as,

$$
\begin{aligned}
&\widehat{\beta}^{(1)}_{i,l+1,M} - \beta^{(1)}_{i,l+1} \\
&= \frac{1}{\sigma^2_{F_B,11,l}} \left(\sum_{j=0}^{M-1} \Delta P_{i,l+(j+1)/M} \Delta f_{1,l+(j+1)/M} \right) \\
&\quad + \frac{\sum_{j=0}^{M-1} \Delta f^2_{1,l+(j+1)/M} - \sigma^2_{F_B,11,l}}{\sum_{j=0}^{M-1} \Delta f^2_{1,l+(j+1)/M} \sigma^2_{F_B,11,l}} \sum_{j=0}^{M-1} \Delta P_{i,l+(j+1)/M} \Delta f_{1,l+(j+1)/M} \\
&= A_{l,M} + B_{l,M},
\end{aligned} \tag{14}
$$

where $\Delta f_{1,l+(j+1)/M} = f_{1,l+(j+1)/M} - f_{1,l+j/M}$, $\Delta f^2_{1,l+(j+1)/M} = \left(f_{1,l+(j+1)/M} - f_{1,l+j/M}\right)^2$ and $\Delta P_{i,l+(j+1)/M} = P_{i,l+(j+1)/M} - P_{i,l+j/M}$. Recalling (1) and (**??**),

$$
\Delta P_{i,l+(j+1)/M} = \mu_{i,l} \frac{1}{M} + \sigma_{i,l} \Delta W_{i,l+(j+1)/M} + \boldsymbol{\Sigma}'_{iF^B,l} \Delta W_{F^A,l+(j+1)/M}
$$

and

$$
\Delta f_{1,l+(j+1)/M} = \mu_{1,F^B,l} \frac{1}{M} + \sigma_{F_B,11,l} \Delta W^{(1)}_{F^A,l+(j+1)/M},
$$

with $\Delta W^{(1)}_{F^A,l+(j+1)/M}$ denoting the first component of $\Delta W_{F^A,l+(j+1)/M}$.

Thus, given A(i), and recalling (14), it follows

$$
\begin{aligned}
&\frac{1}{Th^k} \sum_{l=1}^{T-1} A_{l,M} \\
&= \frac{1}{Th^k} \sum_{l=1}^{T-1} \frac{1}{\sigma^2_{F_B,11,l}} \sum_{j=0}^{M-1} \left(\sigma_{F_B,11,l} \sigma_{i,l} \int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W_i(s) \int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W^{(1)}_{F^B,s} \right. \\
&\quad \left. + \sigma_{i,1,F^B,l} \sigma_{F_B,11,l} \left(\left(\int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W^{(1)}_{F^B,s} \right)^2 \right) - 1 \right) + O_p\left(M^{-1/2} h^{-k}\right) \\
&= \frac{1}{Th^k} \sum_{l=1}^{T-1} \mu^{-1}_{\sigma^2_{F^B,11}} \sum_{j=0}^{M-1} \left(\sigma_{F_B,11,l} \sigma_{i,l} \int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W_i(s) \int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W^{(1)}_{F^B,s} \right. \\
&\quad \left. + \sigma_{i,1,F^B,l} \sigma_{F_B,11,l} \left(\left(\int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W^{(1)}_{F^B,s} \right)^2 - \frac{1}{M} \right) \right) (1 + o_p(1)) + O_p\left(M^{-1/2} h^{-k}\right) \\
&= \frac{1}{Th^k} \sum_{l=1}^{T-1} A_{l,M} (1 + o_p(1)) + O_p\left(M^{-1/2} h^{-k}\right),
\end{aligned}
$$

where $\mu_{\sigma^2_{F^A,11}} = \mathbb{E}\left(\sigma^2_{F_A,11,l}\right)$ and the $o_p(1)$ term holds as $T \to \infty$.

Now, define

$$u_{j,l,M} = \sigma_{F_A,11,l}\sigma_{i,l}\int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W_i(s)\int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W_A(s)^{(1)}$$
$$+\sigma_{i,1,F^A,l}\sigma_{F_A,11,l}\left(\left(\int_{l+j/M}^{l+(j+1)/M} \mathrm{d}W_A(s)^{(1)}\right)^2 - \frac{1}{M}\right).$$

It is immediate to note that $\mathbb{E}\left(A_{l,M}\right) = 0$. By Lemma 1, for $l \neq \iota$, and/or $k \neq j$, $\mathbb{E}\left(u_{j,l,M}u_{k,\iota,M}\right) = 0$, and, because of A(i),

$$\mathbb{E}\left(u_{j,l,M}^2\right) = \frac{1}{M^2}\left(\sigma_{F_A,11,l}\sigma_{i,l} + \sigma_{i,1,F^A,l}\sigma_{F_A,11,l}\right)^2.$$

Thus,

$$\mathrm{Var}\left(\frac{1}{Th^k}\sum_{l=1}^{T-1} A_{l,M}\right)$$
$$= \mu_{\sigma^2_{F^A,11}}^{-2}\frac{1}{T^2h^{2k}}\sum_{l=1}^{T-1}\sum_{\iota=1}^{T-1}\mathbb{E}\left(\sum_{j=0}^{M-1} u_{j,l,M}\sum_{j=0}^{M-1} u_{j,\iota,M}\right)$$
$$= \mu_{\sigma^2_{F^A,11}}^{-2}\frac{1}{T^2h^{2k}}\sum_{l=1}^{T-1}\mathbb{E}\left(\sum_{j=0}^{M-1} u_{j,l,M}^2\right)$$
$$= O\left(\frac{1}{TMh^{2k}}\right).$$

Hence, $\frac{1}{Th^k}\sum_{l=1}^{T-1} A_{l,M} = O_p\left(T^{-1/2}M^{-1/2}h^{-k}\right) + O_p\left(M^{-1/2}h^{-k}\right)$. By a similar argument, $\frac{1}{Th^k}\sum_{l=1}^{T-1} B_{l,M} = O_p\left(T^{-1/2}M^{-1/2}h^{-k}\right) + O_p\left(M^{-1/2}h^{-k}\right)$. Thus, the first term on the RHS of (13) is $O_p\left(T^{-1/2}M^{-1/2}h^{-k}\right)$.

Finally, as for $II_{T,M,h}$ in (13),

$$II_{T,M,h} \leq \left(\frac{1}{Th^k}\sum_{l=1}^{T-1}\left(f_{1,l} - \mu_{f_1}\right)^2\right)^{1/2} \times \left(\frac{1}{Th^k}\sum_{l=1}^{T-1}\left(\widehat{\beta}_{i,l,M}^{(1)} - \beta_{i,l}^{(1)}\right)^2\right)^{1/2}$$
$$= O_p\left(M^{-1/2}h^{-k}\right).$$

The statement in the Proposition then follows. ■

Proof of Proposition 2: Given Proposition 1,

$$\left(\int_{\mathbb{R}^k}\left|tr_T\left(\widehat{m}_{i,T,M}(c)\right) - m_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} - \left(\int_{\mathbb{R}^k}\left|tr_T\left(\widetilde{m}_{i,T}(c)\right) - m_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q}$$
$$= O_p\left(M^{-1/2}h_T^k d_T^{-1}\right).$$

It remains to show that

$$\left(\int_{\mathbb{R}^k}\left|tr_T\left(\widetilde{m}_{i,T}(c)\right) - m_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q}$$
$$= O_p\left(T^{-1/2}h_T^{-k}d_T^{-2}\right) + O\left(h_T^2 d_T^{-2}\right) + O_p\left(d_T^{a/Q}\right).$$

Note that, given (9),

$$
\begin{aligned}
&\left(\int_{\mathbb{R}^k} \left|tr_T\left(\widetilde{m}_{i,T}(c)\right) - m_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} \\
\leq &\left(\int_{\mathbb{R}^k} 1\left\{\widehat{g}_T(c) \geq d_T\right\} \left|\widetilde{m}_{i,T}(c) - m_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} \\
&+\left(\int_{\mathbb{R}^k} 1\left\{\widehat{g}_T(c) < d_T\right\} \left|m_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} \\
= &I_T + II_T
\end{aligned} \tag{15}
$$

The statement in the Proposition will follow from Theorem 1(b) and Corollary 1 in Andrews (1995) once we show that his assumptions NP1-NP7 are satisfied with $\eta = \infty$ and $d_{1,T} = d_{2,T} = d_T$. Now A(ii) implies NP1 hold with $\eta = \infty$, as we require $(Z_{i,t+1}, \boldsymbol{C}_t)$ to be $\alpha-$mixing, instead of being near epoch dependent (NED) on a mixing basis. A(iii) implies NP2. As we are not dealing with adaptive (random) bandwidths and we deal with strong mixing processes, rather than NED process, our A(iv) is equivalent to NP4 (see Andrews p.567). We do not need NP5, as we use the same deterministic bandwidth for the numerator and the denominator. Also bandwidth chosen by cross-validation of plug-in techniques straightforwardly satisfy NP5. A(v) is equivalent to his NP3 with $\lambda = 0$ and $\omega = 2$. Thus, A(ii)-A(v) ensures that the statement in Theorem 1(b) in Andrews (1995) hold, and so $I_T = O_p\left(T^{-1/2}h_T^{-k}d_T^{-2}\right) + O\left(h_T^2 d_T^{-2}\right)$. Now, the trimming device used in (9) ensure that Andrew NP6 is satisfied, Finally, our A(vi) is equivalent to his NP7. Recalling that, by Theorem 1(a) in Andrews (1995) $\sup_{c\in\mathbb{R}^k} \left|\widehat{g}_T(c) - g(c)\right| = o_p\left(d_T\right)$, it then follows that,

$$
\begin{aligned}
II_T &\leq \sup_{c\in\mathbb{R}^k} \left|m_i(c)\right|^Q \left(\int_{\mathbb{R}^k} d_T^a \frac{g(c)}{\widehat{g}_T(c)^a}\mathrm{d}c\right)^{1/Q} \\
&\leq \Delta\left(\int_{\mathbb{R}^k} d_T^a \frac{g(c)}{g(c)^a}\mathrm{d}c\right)^{1/Q} \left(1 + o_p(d_T)\right) = O_p\left(d_T^{a/Q}\right).
\end{aligned}
$$

The statement in the Proposition then follows. ■

Proof of Proposition 3: By the same argument as in Proposition 1, uniformly in c,

$$
tr_T\left(\widehat{IR}_{i,T,M}(c)\right) - tr_T\left(\widetilde{IR}_{i,T}(c)\right) = O_p(M^{-1/2}h_T^{-k}d_T^{-1}),
$$

where $\widetilde{IR}_{i,T}(c)$ is the infeasible counterpart of $\widehat{IR}_{i,T,M}(c)$, constructed using $Z_{i,t}$ instead of $\widehat{Z}_{i,l+1,M}$. Along the same lines used in the proof of Proposition 2,

$$
\begin{aligned}
&\left(\int_{\mathbb{R}^k} \left|tr_T\left(\widetilde{IR}_{i,T}(c)\right) - IR_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} \\
\leq &\left(\int_{\mathbb{R}^k} 1\left\{\widehat{g}_T(c) \geq d_T\right\} \left|\widetilde{IR}_{i,T}(c) - IR_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} \\
&+\left(\int_{\mathbb{R}^k} 1\left\{\widehat{g}_T(c) < d_T\right\} \left|IR_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q}
\end{aligned} \tag{16}
$$

Given A'(ii')(v'), by Theorem 1(b) in Andrews (1995),

$$\sup_{\{c:\widehat{g}_T(c)>d_T\}} \left| \widetilde{m}_{i,T}^{(2)}(c) - m(c)^{(2)} \right| = O_p\left(T^{-1/2}h_T^{-k}d_T^{-2}\right) + O\left(h_T^2 d_T^{-2}\right),$$

and so the first term on the RHS of (16) is $O_p\left(T^{-1/2}h_T^{-k}d_T^{-2}\right) + O\left(h_T^2 d_T^{-2}\right)$. As for the second term on the RHS of (16), given A(vi) and A'(vi'), and along the same lines as in the proof of Proposition 2,

$$\begin{aligned}
&\left(\int_{\mathbb{R}^k} 1\left\{\widehat{g}_T(c) < d_T\right\} \left|IR_i(c)\right|^Q g(c)\mathrm{d}c\right)^{1/Q} \\
&\leq \sup_{c\in\mathbb{R}^k} \left|IR_i(c)\right|^Q \left(\int_{\mathbb{R}^k} d_T^a \frac{g(c)}{\widehat{g}_T(c)^a}\mathrm{d}c\right)^{1/Q} \\
&\leq \Delta\left(\int_{\mathbb{R}^k} d_T^a \frac{g(c)}{g(c)^a}\mathrm{d}c\right)^{1/Q} \left(1 + o_p\left(d_T\right)\right) = O_p\left(d_T^{a/Q}\right).
\end{aligned}$$

This completes the proof. ■