CREATES Research Paper 2011-3

Forecasting Covariance Matrices: A Mixed Frequency Approach

Roxana Halbleib and Valeri Voev

Forecasting Covariance Matrices: A Mixed Frequency Approach

Roxana Halbleib[*]

Valeri Voev[†]

Université libre de Bruxelles

Aarhus University

ECARES, CoFE

CREATES

January 18, 2011

Abstract

This paper proposes a new method for forecasting covariance matrices of financial returns. The model mixes volatility forecasts from a dynamic model of daily realized volatilities estimated with high-frequency data with correlation forecasts based on daily data. This new approach allows for flexible dependence patterns for volatilities and correlations, and can be applied to covariance matrices of large dimensions. The separate modeling of volatility and correlation forecasts considerably reduces the estimation and measurement error implied by the joint estimation and modeling of covariance matrix dynamics. Our empirical results show that the new mixing approach provides superior forecasts compared to multivariate volatility specifications using single sources of information.

JEL classification: C32, C53, G11

Keywords: Volatility forecasting, High-frequency data, Realized variance

1 Introduction

Volatility modeling and forecasting have been of prime interest in financial econometrics since the seminal contributions of Engle (1982) and Bollerslev (1986). Recently, research developments in the field have been refueled by the availability of high-frequency financial data on various financial instruments.

In this paper, we propose a mixed-frequency approach for covariance matrix forecasting, which uses the decomposition of the covariance matrix into a diagonal volatility matrix and a correlation matrix. The same decomposition has been used by Engle (2002) and Tse & Tsui (2002) in their dynamic conditional correlation (DCC) models. Differently from these studies, we propose to forecast the volatility using a dynamic model for the univariate series of realized volatilities, which can be estimated by any of the below-mentioned techniques. The correlation matrix forecast is conceptually identical to the DCC specification, but with the important difference that we standardize (de-volatilize) the daily returns by realized volatilities rather than by GARCH volatilities. The forecasting improvement over daily-data models such as the standard DCC is thus driven by the improvement in volatility forecasts and by the less noisy standardized residuals used as an input to the correlation model. Comparing to pure high-frequency data approaches, our method only requires the estimation of realized volatility series, rather than realized covariance matrices, which, as discussed below, is more problematic. This gives the advantage that the mixed-frequency framework is better suited to handle matrices of large dimensions. Furthermore, model specifications for realized covariance/correlation matrices are only recently gaining more attention (see, e.g., Gourieroux et al. (2009), Bauer & Vorkink (2007), Chiriac & Voev (2010)) and there is still a lot of empirical work needed in order for these models to gain broader recognition. We derive the theoretical conditions under which the mixed-frequency model provides a smaller element-wise forecast mean squared error relative to a pure daily (low-frequency) and a pure high-frequency model, which we refer to collectively as single-frequency approaches. The empirical study of the paper finds evidence that confirms the validity of these conditions and thus not surprisingly, the mixed-frequency model outperforms the single-frequency specifications. The reason we resort to high-frequency data is that it contains information that allow for almost error-free measurement of volatility ex-post, based on the estimation of the quadratic variation of the price process. Early studies in the area (see e.g. Andersen, Bollerslev, Diebold & Labys (2001), Andersen, Bollerslev, Diebold & Ebens (2001), An-

dersen et al. (2003)) recognized that market microstructure effects can distort estimation at very high frequencies and proposed a sparse sampling approach, in which the available data is sampled every 5, 10 or 15 minutes to mitigate the impact of the market microstructure noise. More recently, techniques have been developed to use the data more efficiently by designing estimators that are noise-robust (see e.g. Barndorff-Nielsen et al. (2008), Barndorff-Nielsen et al. (2009), Jacod et al. (2009), Zhang (2006b), Zhang et al. (2005), Nolte & Voev (2009), etc.). Most of these approaches are applicable to univariate series, i.e., for volatility, rather than for covariance estimation. While multivariate extensions of the above mentioned approaches do exist (see e.g., Voev & Lunde (2007), Barndorff-Nielsen et al. (2010), Nolte & Voev (2008), Christensen et al. (2010)) they suffer from limitations especially when applied to many assets. In most empirical work, realized covariance estimation is still carried out using the sparse-sampling approach. The problem with the sparse-sampling method is that for dimensions higher than the number of observations on the sparse subgrid (e.g., at the typical 5-minute frequency there are 78 observations on a NYSE traded stock) the realized covariance matrices are of reduced rank and thus singular. Generally, it can be stated that covariance/correlation estimation with high-frequency data is much more challenging than volatility estimation due to issues of non-synchronicity of the raw multivariate series and parameter proliferation.

Beyond ex-post volatility measurement, high-frequency data has also proven very useful in forecasting future volatility. Currently, there are a number of methods, mostly univariate, which propose dynamic models for realized volatility time-series, or alternatively, ways to integrate realized volatility measures into standard GARCH-type specifications. Hansen & Lunde (2010) provide a review of this growing literature.

To the best of our knowledge, the paper of Bannouh et al. (2009) is the only other study that considers a mixed-frequency covariance model. The differences of their study to ours are very stark and it suffices to mention the two main points of departure. Firstly, their model uses a factor structure in which the factor covariance matrix is estimated with high-frequency data and the loadings on the factors are estimated with daily data. Our approach does not assume a factor structure of the covariance matrix, while it clearly does not exclude that there is one. Secondly, their model is a static one, in the sense that they focus on the issue of estimation of covariance matrices of very large dimension, rather than on forecasting. In fact, the only thing that the two papers have in common is that they both use in some way data at different frequencies.

The remainder of the paper is structured as follows: Section 2 introduces the dynamic mixed-frequency model, Section 3 contains the empirical study, and Section 4 concludes. The proofs to the three propositions in Section 2 are contained in Appendix A. Appendix B contains tables with descriptive statistics of the data and graphs related to the theoretical results from Section 2.

2 Dynamic Mixed-Frequency Model

Let r_t be a vector of daily log returns of dimension $n \times 1$, where n represents the number of assets considered. In this section we introduce a new approach for forecasting the time-varying variance-covariance matrix of the vector of daily returns, Σ_t, based on the following decomposition:

$$\Sigma_t = D_t R_t D_t, \tag{1}$$

where D_t is a diagonal matrix given by the conditional standard deviations of each stock and R_t is the correlation matrix. This decomposition has been used in Engle (2002) and Tse & Tsui (2002) in a dynamic conditional correlation (DCC) framework. Conditional forecasts of Σ_t are given by the conditional forecasts of D_t and R_t as follows:

$$\hat{\Sigma}_{t+1|t} = \hat{D}_{t+1|t} \hat{R}_{t+1|t} \hat{D}_{t+1|t}, \tag{2}$$

where $\hat{D}_{t+1|t} \equiv E[D_{t+1}|\mathcal{F}_t]$, $\hat{R}_{t+1|t} \equiv E[R_{t+1}|\mathcal{F}_t]$ and $\mathcal{F}_t$ is the information set at time t. The novelty of our approach is that it allows for the conditional forecasts of D_t and R_t to stem from different information sets and dynamic frameworks. We refer to this specification as a mixed frequency approach since we use high-frequency (intradaily) data in the model for volatilities (D_t) and daily data in the model for correlations (R_t). Equation (2) trivially implies:

$$
\begin{aligned}
\hat{\sigma}_{ii,t+1|t} &= \hat{d}^2_{i,t+1|t}, & \forall i = 1, \ldots, n & \tag{3} \\
\hat{\sigma}_{ij,t+1|t} &= \hat{d}_{i,t+1|t} \hat{\rho}_{ij,t+1|t} \hat{d}_{j,t+1|t}, & \forall i \neq j, \quad i,j = 1, \ldots, n & \tag{4}
\end{aligned}
$$

where $\hat{\sigma}_{ii,t+1|t}$ and $\hat{d}_{ii,t+1|t}$ are the i-th diagonal elements of $\hat{\Sigma}_{t+1|t}$ and $\hat{D}_{t+1|t}$, and $\hat{\sigma}_{ij,t+1|t}$ and $\hat{\rho}_{ij,t+1|t}$ are the ij-th off-diagonal elements of $\hat{\Sigma}_{t+1|t}$ and $\hat{R}_{t+1|t}$. In the sequel, we will differentiate between forecasts based on the information set containing high-frequency

data (time series of realized volatilities and correlations), $\mathcal{F}_t^H$, and forecasts based on the information set containing data at the low frequency (typically daily returns), $\mathcal{F}_t^L$. Thus, let $\hat{d}_{i,t+1|t}^H$ and $\hat{\rho}_{ij,t+1|t}^H$ be the i-th volatility and ij-th correlation forecast from a dynamic model for the daily series of realized measures such as the Autoregressive Fractionally Integrated Moving Average (ARFIMA) approach suggested by Andersen, Bollerslev, Diebold & Labys (2001) and Andersen, Bollerslev, Diebold & Ebens (2001) or the Heterogenous Autoregressive (HAR) model of Corsi (2009) and Corsi & Audrino (2010). Further, let $\hat{d}_{i,t+1|t}^L$ and $\hat{\rho}_{ij,t+1|t}^L$ be the i-th and ij-th volatility and correlation forecasts from conditional volatility and correlation models using daily data, such as the (Generalized) Autoregressive Conditional Heteroscedastic ((G)ARCH) model of Engle (1982) and Bollerslev (1986) and the Dynamic Conditional Correlation (DCC) approach of Engle (2002).

Below we define three approaches of forecasting Σ_t based on the decomposition given in Equation (2):

$$\hat{\Sigma}_{t+1|t}^{MF} = \hat{D}_{t+1|t}^H \hat{R}_{t+1|t}^L \hat{D}_{t+1|t}^H \tag{5}$$

$$\hat{\Sigma}_{t+1|t}^{LF} = \hat{D}_{t+1|t}^L \hat{R}_{t+1|t}^L \hat{D}_{t+1|t}^L \tag{6}$$

$$\hat{\Sigma}_{t+1|t}^{HF} = \hat{D}_{t+1|t}^H \hat{R}_{t+1|t}^H \hat{D}_{t+1|t}^H \tag{7}$$

Model (5) describes our mixed-frequency (MF) approach. To forecast volatilities we make use of high-frequency data in the form of time series of realized volatility measures, such as the ones obtained by the OLS approach of Nolte & Voev (2009) or an alternative noise-robust method.[1] The correlation forecasts are based on daily data in the spirit of the DCC model of Engle (2002). Models (6) and (7) describe single-frequency approaches based on daily (we will refer to this as the low-frequency (LF) model) and high-frequency data (the high-frequency (HF) model), respectively. It is important to note that for the MF model $\mathcal{F}_t^H$ contains only realized volatilities and not realized correlations, which are required for the HF model. Clearly, the mixed-frequency model $D_{t+1|t}^L \hat{R}_{t+1|t}^H \hat{D}_{t+1|t}^L$ is also conceivable, but not of practical interest. We note that the HF model (7) has been mentioned in Andersen et al. (2006) who also have a brief section on a version of the mixed-frequency model of Bannouh et al. (2009).

Before turning to the formal comparison of the three approaches above, we provide some intuition on why we believe that the mixed-frequency approach might be a valuable alter-

---

[1]Such methods have been developed by Barndorff-Nielsen et al. (2008), Zhang (2006a), Jacod et al. (2009), among others.

native to the single-frequency models. High-frequency data has proven to be extremely useful in the ex-post measurement of volatility. Nevertheless, multivariate approaches are not so well developed and suffer from difficulties associated with non-synchronous trading. This leads to data loss in approaches such as the multivariate kernels of Barndorff-Nielsen et al. (2010) who employ a synchronized sampling scheme or, alternatively, necessitates estimation of all covariances on an element-by-element basis (see, e.g., Nolte & Voev (2008) and Christensen et al. (2010)), which does not guarantee positive-definiteness of the matrix and involves an exponentially growing number of estimations as n increases. In this respect, correlations are much harder to estimate with high-frequency data compared to volatilities. Furthermore, dynamic model specifications for realized covariance matrices are only recently starting to get more attention (see, e.g., Gourieroux et al. (2009), Chiriac & Voev (2010) and Bauer & Vorkink (2007)) and estimation and forecasting with these models, especially with many assets, needs further empirical investigation. Consequently, we view the mixing approach developed in this paper as a method which allows us to extract the informational content of HF data in the estimation of volatilities and make use of the developed body of literature on modelling correlations with daily data. In terms of ease of implementation, the model is much more attractive compared to pure high frequency data models, since it only requires the estimation of n series of realized volatility measures (compared to a series of $n \times n$ realized covariance/correlation matrices).

In the following, we derive and discuss the conditions under which the MF approach provides smaller forecast mean squared errors (MSE) compared to the single-frequency models in equations (6) and (7). We focus on one-step ahead forecasts and simplify the notation in the following manner: $\hat{\sigma}_{ij} \equiv \hat{\sigma}_{ij,t+1|t}$, $\hat{d}_i \equiv \hat{d}_{i,t+1|t}$ and $\hat{\rho}_{ij} \equiv \hat{\rho}_{ij,t+1|t}$ for all $i, j = 1, \ldots, n$. We will use the representations:

$$\hat{\sigma}_{ij} = \sigma_{ij} + \varepsilon_{\sigma_{ij}}, \qquad \forall i, j = 1, \ldots, n \tag{8}$$

$$\hat{d}_i = d_i + \varepsilon_{d_i}, \qquad \forall i = 1, \ldots, n \tag{9}$$

$$\hat{\rho}_{ij} = \rho_{ij} + \varepsilon_{\rho_{ij}}, \qquad \forall i \neq j, \quad i, j = 1, \ldots, n \tag{10}$$

where the ε's represent forecast errors and σ_{ij}, d_i and ρ_{ij} are the true ex-post values of the variables at time $t + 1$. Based on this notation, we can rewrite equations (3) and (4)

as follows:

$$
\begin{aligned}
\hat{\sigma}_{ii} &= (d_i + \varepsilon_{d_i})^2 = d_i^2 + 2d_i\varepsilon_{d_i} + \varepsilon_{d_i}^2 \equiv \sigma_{ii} + \varepsilon_{\sigma_{ii}} && \forall i = 1, \ldots, n && (11) \\
\hat{\sigma}_{ij} &= (d_i + \varepsilon_{d_i})(\rho_{ij} + \varepsilon_{\rho_{ij}})(d_j + \varepsilon_{d_j}) \\
&= d_i\rho_{ij}d_j + d_i\rho_{ij}\varepsilon_{d_j} + d_j\rho_{ij}\varepsilon_{d_i} + d_id_j\varepsilon_{\rho_{ij}} + d_j\varepsilon_{d_i}\varepsilon_{\rho_{ij}} + d_i\varepsilon_{d_j}\varepsilon_{\rho_{ij}} + \rho_{ij}\varepsilon_{d_i}\varepsilon_{d_j} + \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j} \\
&\equiv \sigma_{ij} + \varepsilon_{\sigma_{ij}} && \forall i \neq j, \quad i, j = 1, \ldots, n, && (12)
\end{aligned}
$$

with $\varepsilon_{\sigma_{ii}} \equiv 2d_i\varepsilon_{d_i} + \varepsilon_{d_i}^2$ and $\varepsilon_{\sigma_{ij}} \equiv d_i\rho_{ij}\varepsilon_{d_j} + d_j\rho_{ij}\varepsilon_{d_i} + d_id_j\varepsilon_{\rho_{ij}} + d_j\varepsilon_{d_i}\varepsilon_{\rho_{ij}} + d_i\varepsilon_{d_j}\varepsilon_{\rho_{ij}} + \rho_{ij}\varepsilon_{d_i}\varepsilon_{d_j} + \varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}$. In the following, we compare the models based on their MSE where we make use of the decomposition:

$$
MSE(\hat{\sigma}_{ij}) = E[\varepsilon_{\sigma_{ij}}]^2 + V[\varepsilon_{\sigma_{ij}}]. \qquad (13)
$$

The mean and variance of $\varepsilon_{\sigma_{ij}}$ in the general case without assumptions on moments of the forecast errors are derived in Appendix A. We are now in a position to derive the conditions under which the forecast MSE of the mixing approach, $\hat{\sigma}_{ij}^{MF}$, is smaller than the forecast MSE of the single-frequency models, $\hat{\sigma}_{ij}^{LF}$ and $\hat{\sigma}_{ij}^{HF}$ for each $i, j = 1, \ldots, n$. Initially, we make restrictive assumptions on the dependence among the forecast errors in order to derive an easily interpretable result. We relax most of these assumptions and present a general result further in the paper. Note that we look at elementwise MSE. Alternatively, one can define a matrix error term as the discrepancy between the true covariance matrix and the forecast and consider as a loss function some norm of this error, e.g., the Frobenius norm. We opt for the elementwise MSE, since it is more conservative and more straightforward to interpret. Furthermore, loss functions based on the matrix error can have the undesirable feature that a very large error on one or more of the elements in the matrix can be compensated by very small errors on the other elements. In many applications (e.g., portfolio optimization), what is required is the inverse of the covariance matrix, which can be very badly behaved in such a scenario. Being able to demonstrate a uniform dominance of one model over another is clearly a much stronger statement than showing that the dominance only holds "on average" over the elements in the matrix.

Proposition 1: Variance elements

If, for a given $i \in \{1, 2, \ldots, n\}$, it holds that $E[\varepsilon_{d_i}^X] = 0$, for $X \in \{H, L\}$, then

$$
MSE(\hat{\sigma}_{ii}^{MF}) - MSE(\hat{\sigma}_{ii}^{LF})
$$
$$
= 4(V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L])d_i^2 + 4(E[(\varepsilon_{d_i}^H)^3] - E[(\varepsilon_{d_i}^L)^3])d_i + (E[(\varepsilon_{d_i}^H)^4] - E[(\varepsilon_{d_i}^L)^4]). \quad (14)
$$

A set of *sufficient (but not necessary) conditions* for $MSE(\hat{\sigma}_{ii}^{MF}) \leq MSE(\hat{\sigma}_{ii}^{LF})$ is that $V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L] \leq 0$, $E[(\varepsilon_{d_i}^H)^3] - E[(\varepsilon_{d_i}^L)^3] \leq 0$ and $E[(\varepsilon_{d_i}^H)^4] - E[(\varepsilon_{d_i}^L)^4] \leq 0$. The minimal sufficient conditions, i.e., the necessary and sufficient conditions for the inequality to hold are clearly weaker and are provided in Appendix A.

In Proposition 1, we assume that volatility forecasts are unbiased. The assumption of unbiasedness is rather natural and intuitive. Nevertheless, with daily data, it will be violated, at least in theory, if we use GARCH models to forecast variances and then take the square root of the forecast as the forecast of the volatility. One can either think of this bias as negligible (which is to be expected empirically) or as volatility forecasts stemming for a GARCH model for the standard deviation rather than for the variance (e.g., the threshold GARCH model of Zakoian (1994)). With high frequency data we have the flexibility of directly forecasting realized volatilities as opposed to realized variances so that the unbiasedness assumption is less problematic. In any case, if the unbiasedness assumption is clearly violated, then bias correction should be employed so that eventually unbiasedness is restored. The sufficient conditions we provide reveal that it is sufficient for the mixed-frequency approach to outperform the daily data model if the second, third and fourth moment of the volatility errors from the HF model are smaller than their counterparts from the LF model. We believe that the conditions on the second and fourth moment are likely to be satisfied (and show that they are empirically), since the basic motivation of using HF data is that it helps in measuring and forecasting volatility more precisely than with daily data. As for the third moment, there is no reason to assume that it should be theoretically different from zero for both models. The minimal sufficient conditions (which are also necessary) are clearly weaker and require that a quadratic polynomial in d_i is less then zero. These conditions are derived in Appendix A. Note that we do not compare variance forecast from the MF model to the HF model since they are identical by construction.

Proposition 2: Covariance elements I

If, for a given $i \neq j$, $i, j \in \{1, 2, \ldots, n\}$ it holds that

(i) $E[\varepsilon_x^X] = 0$, for all combinations of x and X where $x \in \{d_i, d_j, \rho_{ij}\}$ and $X \in \{H, L\}$

(ii) $\varepsilon_x^X \perp \varepsilon_y^Y$, for all combinations of x, y, X and Y, where $x, y \in \{d_i, d_j, \rho_{ij}\}$ and $X, Y \in \{H, L\}$,

then it follows that:

1. (mixed frequency to low frequency model comparison)

$$MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{LF}) = (\rho_{ij}^2 + V[\varepsilon_{\rho_{ij}}^L]) \left((V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_j}^L])d_j^2 \right.$$
$$\left. + (V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L])d_i^2 + (V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_i}^L]V[\varepsilon_{d_j}^L]) \right).$$

A set of *sufficient (but not necessary) conditions* for $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{LF})$ is that $V[\varepsilon_{d_i}^H] \leq V[\varepsilon_{d_i}^L]$ and $V[\varepsilon_{d_j}^H] \leq V[\varepsilon_{d_j}^L]$. The minimal sufficient conditions, i.e., the necessary and sufficient conditions for the inequality to hold are clearly weaker and are provided in Appendix A.

2. (mixed frequency to high frequency model comparison)

$$MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{HF}) = (V[\varepsilon_{\rho_{ij}}^L] - V[\varepsilon_{\rho_{ij}}^H])(d_i^2 d_j^2 + d_j^2 V[\varepsilon_{d_i}^H] + d_i^2 V[\varepsilon_{d_j}^H] + V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H]).$$

It follows that $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{HF})$ if and only if $V[\varepsilon_{\rho_{ij}}^L] \leq V[\varepsilon_{\rho_{ij}}^H]$.

In Proposition 2, beside assuming unbiased volatility forecasts, we also assume that correlation forecasts are unbiased and that all forecasting errors are mutually independent both in the volatility/correlation dimension and in the high-frequency/daily data dimension. This independence assumption is clearly too strong (and will be relaxed substantially in the following proposition), but provides a good starting point for the analysis. For the MSE comparison of the mixed frequency model to the low frequency model, the sufficient conditions described in the proposition are satisfied in our empirical work. For completeness, the weaker necessary and sufficient conditions for which a bivariate quadratic polynomial is less than zero while still observing the positivity of d_i and d_j are derived in Appendix A.

8

The second result states that a necessary and sufficient condition for the MF approach to have smaller elementwise MSE than the HF model is that the LF correlation model provides more precise correlation forecasts than the HF model. Whether this condition holds empirically depends on the particular HF model, on the input to the HF model (realized correlation series) and the quality of daily correlation model. We expect that especially in high-dimensional applications the daily model will be much easier to implement and will have an edge over the HF model. Ultimately, the choice of modeling correlations with daily data is not only motivated by the precision of the forecasts but also by better tractability and ease of implementation.

In the following proposition, we relax almost all independence assumptions and only require that volatility and correlation forecast errors from the HF and LF models are independent, which we believe is the least stringent among the assumptions in Proposition 2. It should be noted that this assumption can be relaxed as well. In the proposition below, we only compare MSE of covariance elements, as for the variances we only needed the unbiasedness assumption and therefore the result from Proposition 1 does not change.

Proposition 3: Covariance elements II

If, for a given $i \neq j$, $i, j \in \{1, 2, \ldots, n\}$ it holds that

(i) $E[\varepsilon_x^X] = 0$, for all combinations of x and X where $x \in \{d_i, d_j, \rho_{ij}\}$ and $X \in \{H, L\}$

(ii) $\varepsilon_x^H \perp \varepsilon_y^L$, for all combinations of x, y, where $x, y \in \{d_i, d_j, \rho_{ij}\}$,

then it follows that

1. (mixed frequency to low frequency model comparison)

$$MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{LF}) = F_1(d_i, d_j, \rho_{ij}), \tag{15}$$

where $F_1(d_i, d_j, \rho_{ij})$ is a fourth-order polynomial in d_i, d_j and ρ_{ij} as given in Equation (28) in Appendix A. If $\rho_{ij} \geq 0$, *sufficient (but not necessary)* conditions that $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{LF})$ are that all parameters of $F_1(d_i, d_j, \rho_{ij})$ are non-positive (see Proof A.3 in Appendix A).

2. (mixed frequency to high frequency model comparison)

$$MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{HF}) = F_2(d_i, d_j, \rho_{ij}), \tag{16}$$

where $F_2(d_i, d_j, \rho_{ij})$ is a fourth-order polynomial in d_i, d_j and ρ_{ij} as given in Equation (30) in Appendix A. If $\rho_{ij} \geq 0$, *sufficient (but not necessary)* conditions that $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{HF})$ are that all parameters of the $F_2(d_i, d_j, \rho_{ij})$ polynomial are non-positive (see Proof A.3 in Appendix A).

As can be inferred from the above results, relaxing most of the independence assumptions comes with a significant degree of complication. Sufficient conditions for the MSE of the mixed frequency model to be smaller than that of the single-frequency models cannot be summarized anymore in a few simple restrictions on moments of the forecasting errors as before. Nevertheless, if we assume that $\rho_{ij} \geq 0$ which is empirically relevant, some intuition can be provided. Loosely speaking, the MF model outperforms the LF model if, firstly, variances and certain cross-moments of volatility errors ε_{d_i} and ε_{d_j} up to fifth order are smaller with HF data than with LF data. Secondly, dependence (e.g., the covariance and the co-skewness) between volatility and correlation forecasts from LF data should be generally positive or zero. When compared to the HF approach, the sufficient conditions reduce to requiring that the dependence between volatility and correlation forecast errors stemming from HF data is generally non-negative and larger than the dependence between the volatility forecast errors stemming from HF data and the correlation forecast errors stemming from LF data. These conditions, however, can be too stringent in the sense that much weaker conditions can suffice to obtain the desired result.

The usefulness of Proposition 3, however, is not in providing intuition (that is why we had Propositions 1 and 2). More importantly, it gives us the exact form of the MSE difference as a function of the variables d_i, d_j and ρ_{ij}. Since the parameters of the polynomials F_1 and F_2 are easily estimated from data (simply as sample counterparts of the population moments), we can use the results in Proposition 3 to plot F_1 and F_2 against a range of values for d_i, d_j and ρ_{ij}. Since we cannot plot four-dimensional graphs, we view both F_1 and F_2 as functions of d_i and d_j and plot three-dimensional surfaces for a range of (positive and negative) values of ρ_{ij}. Whenever the surfaces lie below zero, the MF approach has a smaller MSE than the corresponding single-frequency model.

3   Empirical Application

In this section, we present the forecasting results for the mixing and single-frequency multivariate volatility forecasting approaches presented in Section 2. We measure the sta-

tistical precision of the forecasts by means of the MSE criterion discussed in the previous section. As volatility is not observable, we use a realized covariance proxy in our evaluation. We note that the MSE is a loss function which satisfies the conditions in Patton (2009) of being robust to the noise in the volatility proxy.

3.1 Data

The data consists of tick-by-tick bid and ask quotes from the NYSE Trade and Quotations (TAQ) database sampled from 9:30 until 16:00 for the period 01.01.2000 – 30.07.2008 ($T = 2156$ trading days).[2] For the current analysis, we select the following six highly liquid stocks: American Express Inc. (AXP), Citigroup (C), General Electric (GE), Home Depot Inc. (HD), International Business Machines (IBM) and JPMorgan Chase & Co (JPM). We employ the previous-tick interpolation method, described in Dacorogna et al. (2001) and obtain 78 intraday returns by sampling every 5 minutes and 1 open-to-close daily return. Table B.1 in Appendix B reports summary statistics for the 5-minute and daily return series.

For each $t = 1, \ldots, 2156$, a series of daily realized covariance matrices can be constructed as:

$$RCov_t = \sum_{j=1}^{M} r_{j,t} r'_{j,t} \tag{17}$$

where $M = 78$. The 5-minute returns, $r_{j,t}$, are computed as

$$r_{j,t} = p_{j\Delta,t} - p_{(j-1)\Delta,t}, \quad j = 1, \ldots, M$$

where $\Delta = 1/M$ and $p_{j\Delta,t}$ is the log midquote price at time $j\Delta$ in day t. The realized covariance matrices are symmetric by construction and, for $n < M$, positive definite almost surely. Since by sampling sparsely we disregard a lot of data, we refine the estimator by subsampling. With $\Delta = 300$ seconds, we construct 30 regularly Δ-spaced subgrids starting at seconds $1, 11, 21, \ldots, 291$, compute the realized covariance matrix on each subgrid and take the average. The resulting subsampled realized covariance is much more robust to the so called market microstructure noise than the simple 5-minute based one. Given the high liquidity of all the stocks and the very recent sample, we are confident that the effect of non-synchronicity is rather mild at the chosen frequency. In order to avoid the noise

---

[2]We are grateful to Asger Lunde for providing us with the data.

induced by measuring the overnight volatility as the squared overnight return we apply all models to open-to-close data and measure the volatility over the trading session. Table B.2 in Appendix B reports summary statistics of realized variances and covariances of the six stocks considered in the study. As already documented in Andersen, Bollerslev, Diebold & Ebens (2001), both realized variance and covariance distributions are extremely right skewed and leptokurtic.

It is important to emphasize that the sparse sampling realized covariance estimator can only lead to positive definite estimates for the covariance matrix as long as the number of intradaily returns (here $M = 78$) is larger than the number of assets (here $n = 6$). The HF data models presented below are thus only applicable if this condition is satisfied. The MF model does not involve this restriction making it much more suitable for larger dimensional systems.

The six daily realized variance series are given by the diagonal elements of the realized covariance ($RCov$) matrix defined above. Please note the distinction we make here between realized variance (RV) and its square root, for which we use the term realized volatilities ($RVol$). The series of daily realized correlation matrices $RCorr_t$ are computed from $RCov_t$ in the usual way.

3.2 Forecasting Models

In this section we elaborate on the implementation of the three forecasting models introduced in Section 2.

The mixed-frequency (MF) model

The covariance matrix forecast from the MF approach is given by:

$$\hat{\Sigma}_{t+1|t}^{MF} = \hat{D}_{t+1|t}^{H} \hat{R}_{t+1|t}^{L} \hat{D}_{t+1|t}^{H}, \tag{18}$$

where $\hat{D}_{t+1|t}^{H} = diag(RVol_{1,t+1|t}, RVol_{2,t+1|t}, RVol_{3,t+1|t}, RVol_{4,t+1|t}, RVol_{5,t+1|t}, RVol_{6,t+1|t})$ and $RVol_{i,t+1|t}$, $i = 1 \ldots, 6$ are volatility forecasts from the following ARFIMA(1,d,1) model:

$$(1 - \phi_i L)^{d_i} RVol_{i,t}^* = (1 - \theta_i L)\varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim N(0, \omega_i), \tag{19}$$

where $RVol_{i,t}^*$ are the demeaned series of daily realized volatilities, ϕ_i and θ_i are the AR and MA parameters and d_i is the parameter of fractional integration. The $RVol_{i,t}^*$ series

are stationary and invertible as long as $d_i < 0.5$, $-1 < \phi_i < 1$ and $-1 < \theta_i < 1$. $\hat{R}^L_{t+1|t}$ is the correlation matrix forecast derived from the dynamic correlation (DCC) approach of Engle (2002) estimated on daily data as follows:

$$
\begin{aligned}
R^L_t &= (diag(Q^L_t))^{-\frac{1}{2}} Q^L_t (diag(Q^L_t))^{-\frac{1}{2}} \\
Q^L_t &= (1 - \theta_1 - \theta_2)\bar{Q}^L + \theta_1 u_{t-1} u'_{t-1} + \theta_2 Q^L_{t-1},
\end{aligned}
\tag{20}
$$

where u_t is the vector of de-volatilized residuals with elements

$$
u_{i,t} = \frac{\epsilon_{i,t}}{RVol_{i,t}}, \quad i = 1, \ldots, 6,
$$

and $\bar{Q}^L$ is the unconditional covariance of u_t. Furthermore, we assume that the conditional mean of daily returns is constant, $r_{i,t} = E[r_{i,t}|\mathcal{F}_{t-1}] + \epsilon_{i,t} = \mu_i + \epsilon_{i,t}$ and estimate the model in Equation (20) on the demeaned series of daily returns. Note that we standardize the daily returns here by realized volatilities, rather than by GARCH volatilities as in the standard implementation of the DCC. In the theoretical section of the paper, we treated correlation errors from the MF and the LF model as identical. In fact, the standardization by $RVol$ is likely to improve upon the correlation model and is a secondary channel through which HF data leads to improvements. In this sense, the theoretical results on the conditions for the MF model to outperform the LF model are too conservative since they do not take into account these additional gains.

The low frequency (LF) model

The covariance matrix forecasts with daily data are obtained with the DCC model of Engle (2002):

$$
\hat{\Sigma}^{LF}_{t+1|t} = \hat{D}^L_{t+1|t} \hat{R}^L_{t+1|t} \hat{D}^L_{t+1|t}
\tag{21}
$$

where $\hat{D}^L_{t+1|t} = diag(h^{1/2}_{1,t+1|t} \ldots h^{1/2}_{6,t+1|t})$ and $h_{i,t+1|t}$ are forecasts from the GARCH(1,1) model

$$
h_{i,t} = w_i + \alpha_i \epsilon^2_{i,t-1} + \beta_i h_{i,t-1} \qquad \forall i = 1, \ldots, 6
\tag{22}
$$

with $w_i, \alpha_i, \beta_i \geq 0$ and $\alpha_i + \beta_i < 1$, $\forall i = 1, \ldots, 6$. The correlation forecast $\hat{R}^L_{t+1|t}$ is given in Equation (20), however, importantly, the standardized (de-volatilized) residuals are now

given by

$$u_{i,t} = \frac{\epsilon_{i,t}}{\sqrt{h_{i,t}}}, \quad i = 1, \ldots, 6.$$

The high frequency (HF) model

The covariance matrix forecasts stemming from high-frequency data model are given by:

$$\hat{\Sigma}_{t+1|t}^{HF} = \hat{D}_{t+1|t}^{H} \hat{R}_{t+1|t}^{H} \hat{D}_{t+1|t}^{H} \tag{23}$$

where $\hat{D}_{t+1|t}^{H}$ are obtained in the same manner as in the MF model and $\hat{R}_{t+1|t}^{H}$ is given by:

$$\hat{R}_{t+1|t}^{H} = \left(1 - \sum_{l=1}^{t} \hat{\lambda}_l\right) \bar{R}Corr + \sum_{l=1}^{t} \hat{\lambda}_l \tilde{R}Corr_{t-l+1}, \tag{24}$$

where $RCorr_t$ is the realized correlation matrix implied by $RCov_t$, $\bar{R}Corr = \frac{1}{t} \sum_{i=1}^{t} RCorr_i$, $\tilde{R}Corr_t = RCorr_t - \bar{R}Corr$ and λ_l is the sequence of coefficients of a pure AR-representation of the following vector ARFIMA(1,d,1) model:

$$(1 - \phi L)D(L)X_t = (1 - \theta L)\zeta_t, \qquad \zeta_t \sim N(0, \Omega), \tag{25}$$

where X_t is the vector obtained by stacking the lower triangular portion of $\tilde{R}Corr_t$ *without* the main diagonal and $D(L) = (1 - L)^d I_m$, where m is the number of correlation series $m = n(n-1)/2$. For comparison purposes and in order to integrate the new mixing approach in the current literature, we consider additionally the multivariate volatility model introduced by Chiriac & Voev (2010), who apply a VARFIMA model on the Cholesky factors of daily realized covariance matrices. This approach may be considered as an alternative to the HF model described in this section with a different model specification than the one in Equation (23). We will refer to this model as the HF2 model in the sequel.

3.3 Forecast Evaluation

We split the whole sample of data into an in-sample period from 01.01.2000 to 31.12.2005 (1508 days) and an out-of-sample period from 01.01.2006 to 30.07.2008 (648 days). The forecasts are carried out in a recursive manner, i.e., at each step the models are re-estimated with all of the available data.

Table 1 reports some statistics on the MSE of variance and covariance forecasts across the stocks and pairs of stocks for all four approaches. It is evident in the table that the MF approach leads to the smallest forecast MSE both for the variance and the covariance series. Of course the statistics of the MSE of variance forecast errors are identical for the MF and the HF model since both approaches use the same data and model specification to forecast daily volatilities. Due to the precise measurement of volatilities based on high frequency data and their long memory specification, the MF and the HF models provide a clear improvement in the variance forecast when compared to the DCC (the LF) model, which uses daily data and the GARCH variance specification. The improvements of the MF in forecasting covariances is also substantial compared to the LF model. It is reassuring that the MF model provides some improvements over the HF models even in such a small system with only six assets for which we expect that realized correlations are well estimated and modelled. Clearly, one can argue that the correlation dynamics are not well specified in the HF models. While this might be the case, the model specification in Equations (24) – (25) is clearly not inferior in terms of flexibility to the DCC model. Furthermore, the extant literature in the field does not provide many alternatives.

In order to have an idea of the matrix-wise error of the models, we also compute a MSE criterion based on the Frobenius norm of the matrix error term $\Sigma_{t+1} - \hat{\Sigma}_{t+1|t}$.[3] which is 2.885 for the MF model and thus much smaller than for the LF model (4.625) and the HF2 model (3.012), but only slightly smaller than for the HF model (2.896). These results should be viewed in the light of the considerable simplicity of the MF model compared to the HF models. Furthermore, as mentioned above, HF models use realized covariance matrices, which can become rank-deficient with many assets. The MF approach does not suffer from this limitation.

Model	Variances				Covariances			
	min	median	mean	max	min	median	mean	max
MF	0.915	4.576	5.577	12.040	0.259	1.132	1.808	6.483
LF	1.208	7.469	9.126	19.746	0.337	1.582	2.824	10.591
HF	0.915	4.576	5.577	12.040	0.256	1.180	1.823	6.524
HF2	0.947	4.800	5.970	12.866	0.259	1.197	1.828	6.572

Table 1: Descriptive statistics across stocks (for the variance) and pairs of stocks (for the covariance) of the MSE of variance and covariance forecast errors.

---

[3]The Frobenius norm of a real $m \times n$ matrix A is defined as $||A|| = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$.
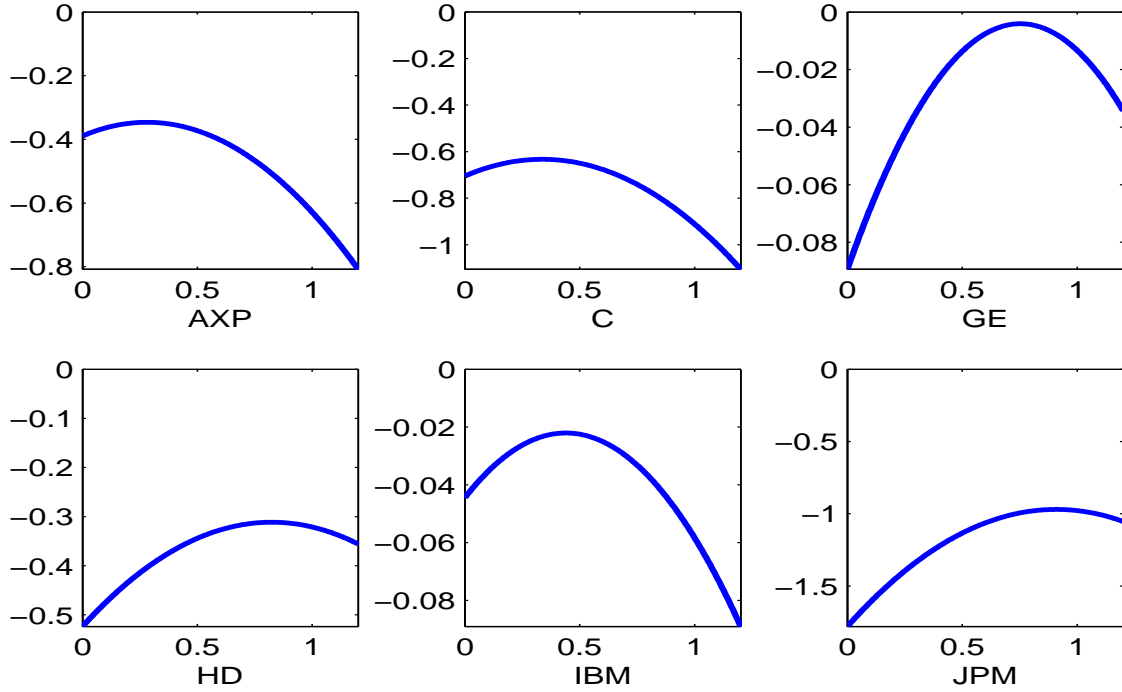
Figure 1: Estimates of the quadratic polynomial in Equation (14).

In the following, we report some empirical results on the necessary and sufficient conditions for the superiority of the MF approach, derived in Proposition 1. The average (across the stocks) estimates of the parameters $A_{1,i} \equiv V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L]$, $A_{2,i} \equiv E[(\varepsilon_{d_i}^H)^3] - E[(\varepsilon_{d_i}^L)^3]$ and $A_{3,i} \equiv E[(\varepsilon_{d_i}^H)^4] - E[(\varepsilon_{d_i}^L)^4]$ from Equation (14) are $\hat{\bar{A}}_1 = -0.136$, $\hat{\bar{A}}_2 = 0.140$ and $\hat{\bar{A}}_3 = -0.588$. Although on average the three parameters do not fulfill the sufficient conditions of Proposition 1 (to be negative), they fulfill the necessary conditions presented in Appendix A (see case 3 in the proof). These empirically validated conditions are very intuitive: on average, one expects a gain in the variance (efficiency) and in the kurtosis of volatility forecasts stemming from HF data compared to LF data, while there are no particular reasons for asymmetric forecast errors (skewness almost zero). The sign and relative magnitude of the estimated average parameters reported above confirm these expectations. Thus the MSE of the variance forecasts stemming from high frequency data is smaller than the one stemming from low frequency data due to a larger gain in the efficiency and in the kurtosis of the HF forecast errors.

In Figure 1 we plot the polynomial from Equation (14) for a wide range of values for d_i, i.e, $d_i = [0, 1.2]$, where the upper limit is chosen to correspond to an annualized standard deviation of approximately 20%. One may observe that for all values of d and all stocks,

the parabola lies under the zero line, which indicates that the HF data based models provide smaller variance mean squared forecast errors than the LF models.

Next, we verify empirically the conditions of Proposition 3 on the MSE inequalities of covariance forecasts. We choose to discuss Proposition 3 rather than Proposition 2 because the former allows for a more general dependence across forecast errors. We plot the estimated polynomials from Proposition 3, $\hat{F}_1(d_i, d_j, \rho_{ij})$ and $\hat{F}_2(d_i, d_j, \rho_{ij})$ for $d_i = [0, 1.2]$ and $d_j = [0, 1.2]$ and for different values of ρ_{ij}: e.g., for three of the 15 covariance pairs and $\rho_{ij} = 0.3$ in Figures 2 and 3, respectively. The plots for the rest of the covariance pairs as well as for other values of the correlation coefficient ($\rho_{ij} = -0.3$ and $\rho_{ij} = 0.9$) can be found in Appendix B, Figures B.1 – B.6. The choice of ρ_{ij} is motivated by the descriptive statistics of the realized correlations over the whole window: on average, the daily realized correlation among all stocks is around 0.3, the maximum is around 0.9 and the minimum is around -0.3.

Figure 2 depicts the behavior of $\hat{F}_1(d_i, d_j, \rho_{ij})$ for three of the covariance pairs for the average value of $\rho_{ij} = 0.3$. As one may notice, compared to the LF model, the MF model provides large MSE improvements in particular for large values of the volatilities. This indicates that the MF approach is especially attractive in a highly volatile environment. However, during volatile periods, the correlations also tend to be large. In this scenario, the MSE gains of the MF model are even stronger – Figure B.2 in Appendix B presents the case in focus with $\rho_{ij} = 0.9$. Generally, we observe that for positive correlations the polynomial surfaces lie under the zero surface for all pair of stocks in the whole range of volatilities. Interestingly, the slope of the MSE difference surfaces reverses for negative correlation. As figure B.3 shows, for $\rho_{ij} = -0.3$, the surface lies above zero for large values of volatilities.
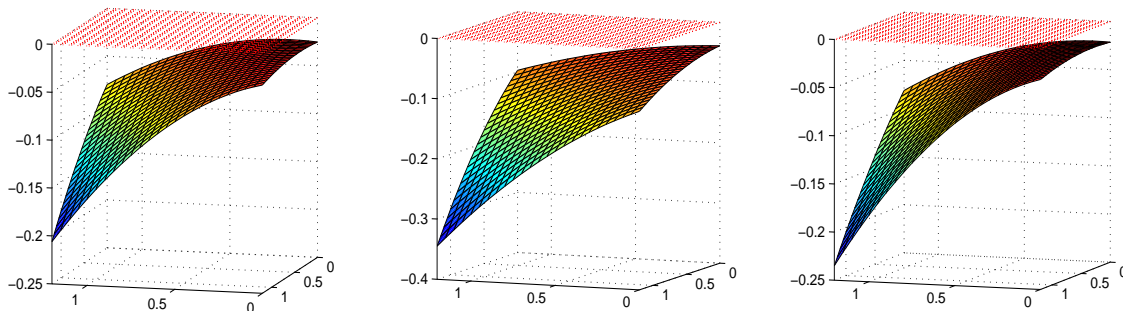

Figure 2: Estimates of the quadratic polynomial in Equation (15) for the stock pairs C-AXP (left), GE-AXP (middle) and GE-C (right) with $\rho = 0.3$. The red surface indicates the zero limit.

17

Although theoretically possible, negative correlations in the stock market are very unlikely; in our sample negative correlation days are on average (across stock pairs) about 1% of the total number of 2165 days. Given that the polynomial parameter estimates are based on 99% data with positive correlations, they are hardly representative of this scenario. Moreover, the empirical evidence shows that days of low and negative correlations correspond to days of low volatility, while high volatility drives high correlations. Therefore, the empirically relevant portion of Figure B.3 is in the lower right corner, indicating better performance of the MF model against the LF model.

Figure 3 features plots of $\hat{F}_2(d_i, d_j, \rho_{ij})$ for three of the covariance pairs for the average value of $\rho_{ij} = 0.3$. Similarly to the previous results, for an average correlation, the choice of the MF model over the HF models is particulary reasonable during periods of high volatility. This result is more pronounced during periods of high correlations which are also typically characterized by high volatilities (see Figure B.5 in Appendix B for $\rho_{ij} = 0.9$). As in the MF to LF model comparison, the behavior changes for negative correlations (see Figure B.6 in Appendix B for $\rho_{ij} = -0.3$). As argued above, however, the empirical irrelevance of negative stock correlations and the lack of representativeness of the estimated parameters in this case imply that these results are not very reliable.

A clear strength of the HF models considered in this study is that they model correlations as a long-memory process, which is empirically justified. The MF model on the contrary employs a standard DCC specification which cannot account for long-memory type of dynamics. In this respect, the HF models are more flexible and not directly comparable to the MF model. Arguably, a more versatile correlation specification can improve the performance of the MF model. We leave this research direction open for now.
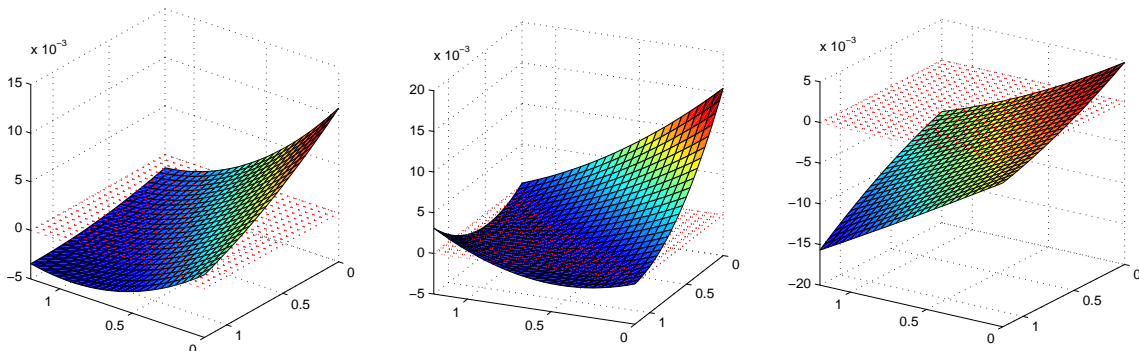

Figure 3: Estimates of the quadratic polynomial in Equation (16) for the stock pairs C-AXP (left), GE-AXP (middle) and GE-C (right) with $\rho = 0.3$. The red surface indicates the zero limit.

4 Conclusion

In this paper, we introduce a new methodology for forecasting multivariate volatility of possibly large dimensions by mixing forecasts stemming from daily and high frequency data. It consists of decomposing the covariance matrix of returns into a diagonal volatility matrix and a correlation matrix and predicting each of these matrices using data sampled at different frequencies. We forecast daily volatilities using univariate autoregressive fractionally integrated moving average models for the time series of daily realized volatilities. Correlations are modeled with the DCC model of Engle (2002) applied on daily returns standardized by realized volatility. This methodology provides an intuitive mixture of volatility and correlation forecasts by simultaneously exploiting the advantages of using high-frequency data to precisely measure daily volatilities and the advantages of using a DCC-type framework for forecasting correlation matrices. In terms of estimation, the new approach is easy to implement since it only requires estimation of univariate series of daily realized volatilities and a single estimation of the DCC model. In the theoretical section of the paper, we derive the conditions under which the new model outperforms the single-frequency forecasting approaches in terms of forecast mean squared error. Although seemingly cumbersome, the relevant theoretical conditions are easily verifiable in empirical work. In our application, we show that forecasting the covariance matrix of a portfolio of six highly liquid stocks traded on NYSE by means of the mixing approach provides smaller mean squared errors compared to the single-frequency models. Moreover, the empirical results show that the benefits of using the new method to forecast covariance matrices are particulary large during turbulent, highly volatile periods. In further work, we plan to consider a larger universe of assets in order to fully demonstrate the power of the proposed methodology.

References

Andersen, T., Bollerslev, T., Christoffersen, P. F. & Diebold, F. X. (2006), Practical volatility and correlation modeling for financial market risk management, *in* M. Carey & R. M. Stulz, eds, 'The Risks of Financial Institutions', The University of Chicago Press, chapter 11, pp. 513–544.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Bannouh, K., Martens, M., Oomen, R. & van Dijk, D. (2009), Realized mixed-frequency factor models for vast dimensional covariance estimation. Working Paper, Erasmus University Rotterdam.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2008), 'Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise', *Econometrica* **76**, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* . forthcoming.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2010), 'Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* . forthcoming.

Bauer, G. H. & Vorkink, K. (2007), Multivariate realized stock market volatility. Working Paper 2007-20, Bank of Canada.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Chiriac, R. & Voev, V. (2010), 'Modelling and forecasting multivariate realized volatility', *Journal of Applied Econometrics* . forthcoming.

Christensen, K., Kinnebrock, S. & Podolskij, M. (2010), 'Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data', *Journal of Econometrics* . forthcoming.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 174–196.

Corsi, F. & Audrino, F. (2010), 'Modeling tick-by-tick realized correlations', *Computational Statistics & Data Analysis* **54**, 2372–2382.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B. & Pictet, O. V. (2001), *An Introduction to High-Frequency Finance*, San Diego Academic Press.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models', *Journal of Business and Economic Statistics* **20**, 339–350.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation', *Econometrica* **50**, 987–1008.

Gourieroux, C., Jasiak, J. & Sufana, R. (2009), 'The wishart autoregressive process of multivariate stochastic volatility', *Journal of Econometrics* **150**, 167–181.

Hansen, P. & Lunde, A. (2010), Forecasting volatility using high frequency data, *in* M. P. Clements & D. F. Hendry, eds, 'Oxford Handbook of Economic Forecasting', Oxford University Press. forthcoming.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: The pre-averaging approach', *Stochastic Processes and their Applications* **119**(7), 2249 – 2276.

Nolte, I. & Voev, V. (2008), Estimating high-frequency based (co-) variances: A unified approach. CREATES Working paper 2008-31, Aarhus University.

Nolte, I. & Voev, V. (2009), Least squares inference on integrated volatility and the relationship between efficient prices and noise. Working paper, Warwick Business School.

Patton, A. (2009), 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics* . forthcoming.

Tse, Y. & Tsui, A. (2002), 'A multivariate generalized auto-regressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business and Economic Statistics* **20**, 351–362.

Voev, V. & Lunde, A. (2007), 'Integrated covariance estimation using high-frequency data in the presence of noise', *Journal of Financial Econometrics* **5**, 68–104.

Zakoian, J.-M. (1994), 'Threshold heteroskedastic models', *Journal of Economic Dynamics and Control* **18**(5), 931–955.

Zhang, L. (2006*a*), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* **12**, 1019–1043.

Zhang, L. (2006*b*), Estimating covariation: Epps effect and microstructure noise. Working Paper.

Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high frequency data', *Journal of the American Statistical Association* **100**, 1394–1411.

A   Appendix

In this appendix, in order to keep notational burden at a minimum, whenever we do not index variables by superscripts for models (MF, LF, HF) or data frequency (H, L), it means that the equation holds for any model/data frequency.

Preliminaries
We have that:

$$
E[\varepsilon_{\sigma_{ii}}] = 2d_i E[\varepsilon_{d_i}] + E[\varepsilon_{d_i}^2]
$$

$$
\begin{aligned}
E[\varepsilon_{\sigma_{ij}}] &= d_i \rho_{ij} E[\varepsilon_{d_j}] + d_j \rho_{ij} E[\varepsilon_{d_i}] + d_i d_j E[\varepsilon_{\rho_{ij}}] + d_i E[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + d_j E[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] \\
&+ \rho_{ij} E[\varepsilon_{d_i}\varepsilon_{d_j}] + E[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}]
\end{aligned}
$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ii}}] &= V[2d_i\varepsilon_{d_i} + \varepsilon_{d_i}^2] = 4d_i^2 V[\varepsilon_{d_i}] + V[\varepsilon_{d_i}^2] + 4d_i \mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{d_i}^2] \\
&= 4d_i^2 V[\varepsilon_{d_i}] + (E[\varepsilon_{d_i}^4] - E[\varepsilon_{d_i}^2]^2) + 4d_i(E[\varepsilon_{d_i}^3] - E[\varepsilon_{d_i}]E[\varepsilon_{d_i}^2])
\end{aligned}
$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}] &= d_i^2 \rho_{ij}^2 V[\varepsilon_{d_j}] + d_j^2 \rho_{ij}^2 V[\varepsilon_{d_i}] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}] + d_i^2 V[\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + d_j^2 V[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}\varepsilon_{d_j}] + V[\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] \\
&+ 2d_i d_j \rho_{ij}^2 \mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{d_j}] + 2d_i^2 d_j \rho_{ij}\mathrm{Cov}[\varepsilon_{d_j},\varepsilon_{\rho_{ij}}] + 2d_i d_j^2 \rho_{ij}\mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{\rho_{ij}}] \\
&+ 2d_i d_j \rho_{ij}\mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_j^2 \rho_{ij}\mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_j \rho_{ij}^2 \mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{d_i}\varepsilon_{d_j}] \\
&+ 2d_i d_j \rho_{ij}\mathrm{Cov}[\varepsilon_{d_j},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_i^2 \rho_{ij}\mathrm{Cov}[\varepsilon_{d_j},\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_i \rho_{ij}^2 \mathrm{Cov}[\varepsilon_{d_j},\varepsilon_{d_i}\varepsilon_{d_j}] \\
&+ 2d_i^2 d_j \mathrm{Cov}[\varepsilon_{\rho_{ij}},\varepsilon_{d_j}\varepsilon_{\rho_{ij}}] + 2d_i d_j^2 \mathrm{Cov}[\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_i d_j \rho_{ij}\mathrm{Cov}[\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{d_j}] \\
&+ 2d_i \rho_{ij}\mathrm{Cov}[\varepsilon_{d_j},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2d_j \rho_{ij}\mathrm{Cov}[\varepsilon_{d_i},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2d_i d_j \mathrm{Cov}[\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] \\
&+ 2d_i d_j \mathrm{Cov}[\varepsilon_{d_j}\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}] + 2d_i \rho_{ij}\mathrm{Cov}[\varepsilon_{d_j}\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{d_j}] + 2d_j \rho_{ij}\mathrm{Cov}[\varepsilon_{d_i}\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{d_j}] \\
&+ 2d_i \mathrm{Cov}[\varepsilon_{d_j}\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2d_j \mathrm{Cov}[\varepsilon_{d_i}\varepsilon_{\rho_{ij}},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}] + 2\rho_{ij}\mathrm{Cov}[\varepsilon_{d_i}\varepsilon_{d_j},\varepsilon_{d_i}\varepsilon_{\rho_{ij}}\varepsilon_{d_j}].
\end{aligned}
$$

Proof A.1 (Proposition 1):
Under the unbiasedness assumption $E[\varepsilon_{d_i}^X] = 0$, for $X \in \{H, L\}$, we have that

$$
E[\varepsilon_{\sigma_{ii}}] = 2d_i E[\varepsilon_{d_i}] + E[\varepsilon_{d_i}^2] = V[\varepsilon_{d_i}].
$$

Thus the MSE of $\hat{\sigma}_{ii}$ is given by

$$
\begin{aligned}
MSE(\hat{\sigma}_{ii}) &= E[\varepsilon_{\sigma_{ii}}]^2 + V[\varepsilon_{\sigma_{ii}}] = V[\varepsilon_{d_i}]^2 + 4d_i^2 V[\varepsilon_{d_i}] + E[\varepsilon_{d_i}^4] - V[\varepsilon_{d_i}]^2 + 4d_i E[\varepsilon_{d_i}^3] \\
&= 4d_i^2 V[\varepsilon_{d_i}] + E[\varepsilon_{d_i}^4] + 4d_i E[\varepsilon_{d_i}^3]
\end{aligned}
$$

It follows that

$$
MSE(\hat{\sigma}_{ii}^{MF}) - MSE(\hat{\sigma}_{ii}^{LF}) = 4d_i^2 \left(V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L]\right) + E[(\varepsilon_{d_i}^H)^4] - E[(\varepsilon_{d_i}^L)^4] + 4d_i \left(E[(\varepsilon_{d_i}^H)^3] - E[(\varepsilon_{d_i}^L)^3]\right).
$$

Let

$$
A_{1,i} \equiv V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L], \quad A_{2,i} \equiv E[(\varepsilon_{d_i}^H)^3] - E[(\varepsilon_{d_i}^L)^3], \quad A_{3,i} \equiv E[(\varepsilon_{d_i}^H)^4] - E[(\varepsilon_{d_i}^L)^4] \qquad (26)
$$

Then we have that

$$MSE(\hat{\sigma}_{ii}^{MF}) - MSE(\hat{\sigma}_{ii}^{LF}) \;=\; 4A_{1,i}d_i^2 + 4A_{2,i}d_i + A_{3,i}. \tag{27}$$

Clearly, $V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L] \leq 0$, $E[(\varepsilon_{d_i}^H)^3] - E[(\varepsilon_{d_i}^L)^3] \leq 0$ and $E[(\varepsilon_{d_i}^H)^4] - E[(\varepsilon_{d_i}^L)^4] \leq 0$ are sufficient conditions for $MSE(\hat{\sigma}_{ii}^{MF}) \leq MSE(\hat{\sigma}_{ii}^{LF})$ since $d_i \geq 0$. To verify the necessary and sufficient conditions, consider that $4A_{1,i}d_i^2 + 4d_iA_{2,i} + A_{3,i} \leq 0$ if and only if

1. $A_{1,i} = 0$, and $d_i \leq -\frac{A_{3,i}}{4A_{2,i}}$ or

2. $A_{1,i} > 0$, $A_{2,i}^2 - A_{1,i}A_{3,i} > 0$ and $d_i \in [d_{1,i}, d_{2,i}]$ where $d_{1/2,ij} = \frac{-A_{2,i}\mp\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$ or

3. $A_{1,i} < 0$ and if either

   3.1 $A_{2,i}^2 - A_{1,i}A_{3,i} \leq 0$ or

   3.2 $A_{2,i}^2 - A_{1,i}A_{3,i} > 0$ and $d_i \notin (d_{1,i}, d_{2,i})$ where $d_{1/2,ij} = \frac{-A_{2,i}\mp\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$.

The following table summarizes the necessary and sufficient conditions for $MSE(\hat{\sigma}_{ii}^{MF}) \leq MSE(\hat{\sigma}_{ii}^{LF})$ taking into account that $d_i \geq 0$:

$A_{1,i}$	$A_{2,i}$	$A_{3,i}$	d_i
≤ 0	≤ 0	≤ 0	$\forall d_i$
> 0	≤ 0	≤ 0	$d_i \leq \frac{-A_{2,i}+\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$
$\leq \frac{A_{2,i}^2}{A_{3,i}}$	> 0	≤ 0	$\forall d_i$
$0 > A_{1,i} > \frac{A_{2,i}^2}{A_{3,i}}$	> 0	≤ 0	$0 \leq d_i \leq \frac{-A_{2,i}-\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}},\; d_i \geq \frac{-A_{2,i}+\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$
$= 0$	> 0	≤ 0	$0 \leq d_i \leq -\frac{A_{3,j}}{4A_{2,i}}$
> 0	> 0	≤ 0	$0 \leq d_i \leq \frac{-A_{2,i}+\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$
≥ 0	≥ 0	$= 0$	$d_i = 0$
< 0	< 0	> 0	$d_i \geq \frac{-A_{2,i}+\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$
$= 0$	< 0	> 0	$d_i \geq -\frac{A_{3,i}}{4A_{2,i}}$
$0 < A_{1,i} \leq \frac{A_{2,i}^2}{A_{3,i}}$	< 0	> 0	$\frac{-A_{2,i}-\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}} \leq d_i \leq \frac{-A_{2,i}+\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$
< 0	≥ 0	≥ 0	$d_i \geq \frac{-A_{2,i}+\sqrt{A_{2,i}^2-A_{1,i}A_{3,i}}}{2A_{1,i}}$

$\square$

Proof A.2 (Proposition 2):

Under the assumptions of Proposition 2, $E[\varepsilon_{\sigma_{ij}}^{MF}] = E[\varepsilon_{\sigma_{ij}}^{LF}] = E[\varepsilon_{\sigma_{ij}}^{HF}] = 0$, and the variances of the forecasting errors of the three models are given by:

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}^{MF}] &= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^H] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^H] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^L] + d_i^2 V[\varepsilon_{d_j}^H \varepsilon_{\rho_{ij}}^L] + d_j^2 V[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^L] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] + V[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^H] \\
&= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^H] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^H] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^L] + d_i^2 V[\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] + d_j^2 V[\varepsilon_{d_i}^H]V[\varepsilon_{\rho_{ij}}^L] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H] + V[\varepsilon_{d_i}^H]V[\varepsilon_{\rho_{ij}}^L]V[\varepsilon_{d_j}^H]
\end{aligned}
$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}^{LF}] &= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^L] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^L] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^L] + d_i^2 V[\varepsilon_{d_j}^L \varepsilon_{\rho_{ij}}^L] + d_j^2 V[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}^L \varepsilon_{d_j}^L] + V[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L] \\
&= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^L] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^L] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^L] + d_i^2 V[\varepsilon_{d_j}^L]V[\varepsilon_{\rho_{ij}}^L] + d_j^2 V[\varepsilon_{d_i}^L]V[\varepsilon_{\rho_{ij}}^L] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}^L]V[\varepsilon_{d_j}^L] + V[\varepsilon_{d_i}^L]V[\varepsilon_{\rho_{ij}}^L]V[\varepsilon_{d_j}^L]
\end{aligned}
$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}^{HF}] &= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^H] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^H] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^H] + d_i^2 V[\varepsilon_{d_j}^H \varepsilon_{\rho_{ij}}^H] + d_j^2 V[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] + V[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H \varepsilon_{d_j}^H] \\
&= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^H] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^H] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^H] + d_i^2 V[\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^H] + d_j^2 V[\varepsilon_{d_i}^H]V[\varepsilon_{\rho_{ij}}^H] \\
&+ \rho_{ij}^2 V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H] + V[\varepsilon_{d_i}^H]V[\varepsilon_{\rho_{ij}}^H]V[\varepsilon_{d_j}^H]
\end{aligned}
$$

Let

$$
\begin{aligned}
B_{1,i} &\equiv V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L] \\
B_{1,j} &\equiv V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_j}^L] \\
B_{2,ij} &\equiv V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_i}^L]V[\varepsilon_{d_j}^L] \\
B_{3,ij} &\equiv d_j^2 B_{1,i} + d_i^2 B_{1,j} + B_{2,ij} \\
B_{4,ij} &\equiv d_i^2 d_j^2 + d_j^2 V[\varepsilon_{d_i}^H] + d_i^2 V[\varepsilon_{d_j}^H] + V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H]
\end{aligned}
$$

Then for a given $i \neq j$, we can write:

1. $MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{LF}) = \{\rho_{ij}^2 + V[\varepsilon_{\rho_{ij}}^L]\}B_{3,ij}$.

clearly, a set of sufficient conditions for $MSE(\hat{\sigma}_{ij}^1) \leq MSE(\hat{\sigma}_{ij}^2)$ is that $V[\varepsilon_{d_i}^H] \leq V[\varepsilon_{d_i}^L]$ and $V[\varepsilon_{d_j}^H] \leq V[\varepsilon_{d_j}^L]$. Considering necessary and sufficient conditions, since $\rho_{ij}^2 + V[\varepsilon_{\rho_{ij}}^L] \geq 0$, $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{LF})$ if and only if

$$
B_{3,ij} = d_j^2\left(V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L]\right) + d_i^2\left(V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_j}^L]\right) + \left(V[\varepsilon_{d_i}^H]V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_i}^L]V[\varepsilon_{d_j}^L]\right) \leq 0,
$$

which holds if and only if either:

1. $B_{1,i} = B_{1,j} = 0$ and $B_{2,ij} \leq 0$ or

2. $B_{1,i} = 0$ and $B_{1,j}d_i^2 + B_{2,ij} \leq 0$, which holds if either

    (a) $B_{1,j} > 0$ and $-B_{1,j}B_{2,ij} > 0 \Leftrightarrow B_{2,ij} < 0$ and $d_i \in [d_{1,i}, d_{2,i}]$ where $d_{1/2,i} = \mp\sqrt{\frac{-B_{2,ij}}{B_{1,j}}}$ or

    (b) $B_{1,j} < 0$ and $-B_{1,j}B_{2,ij} > 0 \Leftrightarrow B_{2,ij} > 0$ and $d_i \notin (d_{1,i}, d_{2,i})$ where $d_{1/2,i}$ are given in condition 2.(a) from above, or

    (c) $B_{1,j} < 0$ and $-B_{1,j}B_{2,ij} \leq 0 \Leftrightarrow B_{2,ij} \leq 0$

or

3. $B_{1,i} > 0$ and $-B_{1,i}(B_{1,j}d_i^2 + B_{2,ij}) > 0 \Leftrightarrow B_{1,j}d_i^2 + B_{2,ij} < 0$, which holds if either one of the conditions 2.(a)-2.(c) from above holds and $d_j \in [d_{1,j}, d_{2,j}]$ where $d_{1/2,j} = \mp\sqrt{\frac{-(B_{1,j}d_i^2 + B_{2,ij})}{B_{1,i}}}$ or

4. $B_{1,i} < 0$ and either

    (a) $-B_{1,i}(B_{1,j}d_i^2 + B_{2,ij}) \leq 0 \Leftrightarrow B_{1,j}d_i^2 + B_{2,ij} \leq 0$, which holds if either one of the conditions 2.(a)-2.(c) from above holds or

    (b) $-B_{1,i}(B_{1,j}d_i^2 + B_{2,ij}) > 0 \Leftrightarrow B_{1,j}d_i^2 + B_{2,ij} > 0$, which holds if neither of the conditions 2.(a)-2.(c) from above holds and $d_j \notin (d_{1,j}, d_{2,j})$ where $d_{1/2,j}$ are given in condition 3 from above.

The following table summarizes the necessary and sufficient conditions for $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{LF})$:

$B_{1,i}$	$B_{1,j}$	$B_{2,ij}$	d_i, d_j
≤ 0	≤ 0	≤ 0	$\forall d_i, \forall d_j$
0	> 0	≤ 0	$\forall d_i,\ d_j \leq \sqrt{-\frac{B_{2,ij}}{B_{1,j}}}$
> 0	0	≤ 0	$d_i \leq \sqrt{-\frac{B_{2,ij}}{B_{1,i}}},\ \forall d_j$
0	< 0	≥ 0	$\forall d_i,\ d_j \geq \sqrt{-\frac{B_{2,ij}}{B_{1,j}}}$
< 0	0	≥ 0	$d_i \geq \sqrt{-\frac{B_{2,ij}}{B_{1,i}}},\ \forall d_j$
> 0	> 0	≤ 0	$d_i \leq \sqrt{-\frac{(B_{1,j}d_j^2 + B_{2,ij})}{B_{1,i}}},\ d_j \leq \sqrt{-\frac{B_{2,ij}}{B_{1,j}}}$
> 0	< 0	≥ 0	$d_i \leq \sqrt{-\frac{(B_{1,j}d_j^2 + B_{2,ij})}{B_{1,i}}},\ d_j \geq \sqrt{-\frac{B_{2,ij}}{B_{1,j}}}$
< 0	> 0	≥ 0	$d_i \geq \sqrt{-\frac{B_{2,ij}}{B_{1,i}}},\ d_j \leq \sqrt{-\frac{(B_{1,i}d_i^2 + B_{2,ij})}{B_{1,j}}}$
> 0	< 0	≤ 0	$d_i \leq \sqrt{-\frac{(B_{1,j}d_j^2 + B_{2,ij})}{B_{1,i}}},\ \forall d_j$
< 0	> 0	≤ 0	$\forall d_i,\ d_j \leq \sqrt{-\frac{(B_{1,i}d_i^2 + B_{2,ij})}{B_{1,j}}}$
< 0	< 0	≥ 0	$\forall d_i,\ d_j \geq \sqrt{-\frac{B_{2,ij}}{B_{1,j}}}$ $d_i \geq \sqrt{-\frac{(B_{1,j}d_j^2 + B_{2,ij})}{B_{1,i}}},\ d_j < \sqrt{-\frac{B_{2,ij}}{B_{1,j}}}$

$$2. \quad MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{HF}) \;=\; \{V[\varepsilon_{\rho_{ij}}^{L}] - V[\varepsilon_{\rho_{ij}}^{H}]\}B_{4,ij}$$

Given that $B_{4,ij} > 0$, $\forall d_i, d_j$, $MSE(\hat{\sigma}_{ij}^{MF}) \leq MSE(\hat{\sigma}_{ij}^{HF})$ if and only if $V[\varepsilon_{\rho_{ij}}^{L}] \leq V[\varepsilon_{\rho_{ij}}^{H}]$. $\qquad\square$

Proof A.3 (Proposition 3):
Under the assumptions of Proposition 3, we have that:

$$E[\varepsilon_{\sigma_{ij}}^{MF}] = \rho_{ij}E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}]$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}^{MF}] &= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^{H}] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^{H}] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^{L}] + d_i^2 V[\varepsilon_{d_j}^{H}]V[\varepsilon_{\rho_{ij}}^{L}] + d_j^2 V[\varepsilon_{d_i}^{H}]V[\varepsilon_{\rho_{ij}}^{L}] + \rho_{ij}^2 V[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] \\
&+ V[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}]V[\varepsilon_{\rho_{ij}}^{L}] + 2d_i d_j \rho_{ij}^2 E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] + 2d_j\rho_{ij}^2 E[(\varepsilon_{d_i}^{H})^2\varepsilon_{d_j}^{H}] + 2d_i\rho_{ij}^2 E[\varepsilon_{d_i}^{H}(\varepsilon_{d_j}^{H})^2] \\
&+ 4d_i d_j E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}]V[\varepsilon_{\rho_{ij}}^{L}] + 2d_i E[\varepsilon_{d_i}^{H}(\varepsilon_{d_j}^{H})^2]V[\varepsilon_{\rho_{ij}}^{L}] + 2d_j E[(\varepsilon_{d_i}^{H})^2\varepsilon_{d_j}^{H}]V[\varepsilon_{\rho_{ij}}^{L}]
\end{aligned}
$$

$$E[\varepsilon_{\sigma_{ij}}^{LF}] = d_i E[\varepsilon_{d_j}^{L}\varepsilon_{\rho_{ij}}^{L}] + d_j E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}] + \rho_{ij}E[\varepsilon_{d_i}^{L}\varepsilon_{d_j}^{L}] + E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}]$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}^{LF}] &= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^{L}] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^{L}] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^{L}] + d_i^2 V[\varepsilon_{d_j}^{L}\varepsilon_{\rho_{ij}}^{L}] + d_j^2 V[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}] + \rho_{ij}^2 V[\varepsilon_{d_i}^{L}\varepsilon_{d_j}^{L}] \\
&+ V[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}] + 2d_i d_j\rho_{ij}^2 E[\varepsilon_{d_i}^{L}\varepsilon_{d_j}^{L}] + 2d_i^2 d_j\rho_{ij} E[\varepsilon_{d_j}^{L}\varepsilon_{\rho_{ij}}^{L}] + 2d_i d_j^2\rho_{ij} E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}] \\
&+ 6d_i d_j\rho_{ij} E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}] + 2d_j^2\rho_{ij} E[(\varepsilon_{d_i}^{L})^2\varepsilon_{\rho_{ij}}^{L}] + 2d_j\rho_{ij}^2 E[(\varepsilon_{d_i}^{L})^2\varepsilon_{d_j}^{L}] + 2d_i^2\rho_{ij} E[(\varepsilon_{d_j}^{L})^2\varepsilon_{\rho_{ij}}^{L}] \\
&+ 2d_i\rho_{ij}^2 E[\varepsilon_{d_i}^{L}(\varepsilon_{d_j}^{L})^2] + 2d_i^2 d_j E[(\varepsilon_{\rho_{ij}}^{L})^2\varepsilon_{d_j}^{L}] + 2d_i d_j^2 E[\varepsilon_{d_i}^{L}(\varepsilon_{\rho_{ij}}^{L})^2] + 4d_i\rho_{ij} E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}(\varepsilon_{d_j}^{L})^2] \\
&+ 4d_j\rho_{ij} E[(\varepsilon_{d_i}^{L})^2\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}] + 4d_i d_j E[\varepsilon_{d_i}^{L}(\varepsilon_{\rho_{ij}}^{L})^2\varepsilon_{d_j}^{L}] - 2d_i d_j E[\varepsilon_{d_j}^{L}\varepsilon_{\rho_{ij}}^{L}]E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}] \\
&- 2d_i\rho_{ij} E[\varepsilon_{d_j}^{L}\varepsilon_{\rho_{ij}}^{L}]E[\varepsilon_{d_i}^{L}\varepsilon_{d_j}^{L}] - 2d_j\rho_{ij} E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}]E[\varepsilon_{d_i}^{L}\varepsilon_{d_j}^{L}] + 2d_i E[\varepsilon_{d_i}^{L}(\varepsilon_{\rho_{ij}}^{L})^2(\varepsilon_{d_j}^{L})^2] \\
&- 2d_i E[\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}]E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}] + 2d_j E[(\varepsilon_{d_i}^{L})^2(\varepsilon_{\rho_{ij}}^{L})^2\varepsilon_{d_j}^{L}] - 2d_j E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}]E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}] \\
&+ 2\rho_{ij} E[(\varepsilon_{d_i}^{L})^2\varepsilon_{\rho_{ij}}^{L}(\varepsilon_{d_j}^{L})^2] - 2\rho_{ij} E[\varepsilon_{d_i}^{L}\varepsilon_{d_j}^{L}]E[\varepsilon_{d_i}^{L}\varepsilon_{\rho_{ij}}^{L}\varepsilon_{d_j}^{L}]
\end{aligned}
$$

$$E[\varepsilon_{\sigma_{ij}}^{HF}] = d_i E[\varepsilon_{d_j}^{H}\varepsilon_{\rho_{ij}}^{H}] + d_j E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}] + \rho_{ij}E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] + E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}]$$

$$
\begin{aligned}
V[\varepsilon_{\sigma_{ij}}^{HF}] &= d_i^2\rho_{ij}^2 V[\varepsilon_{d_j}^{H}] + d_j^2\rho_{ij}^2 V[\varepsilon_{d_i}^{H}] + d_i^2 d_j^2 V[\varepsilon_{\rho_{ij}}^{H}] + d_i^2 V[\varepsilon_{d_j}^{H}\varepsilon_{\rho_{ij}}^{H}] + d_j^2 V[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}] + \rho_{ij}^2 V[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] \\
&+ V[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}] + 2d_i d_j\rho_{ij}^2 E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] + 2d_i^2 d_j\rho_{ij} E[\varepsilon_{d_j}^{H}\varepsilon_{\rho_{ij}}^{H}] + 2d_i d_j^2\rho_{ij} E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}] \\
&+ 6d_i d_j\rho_{ij} E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}] + 2d_j^2\rho_{ij} E[(\varepsilon_{d_i}^{H})^2\varepsilon_{\rho_{ij}}^{H}] + 2d_j\rho_{ij}^2 E[(\varepsilon_{d_i}^{H})^2\varepsilon_{d_j}^{H}] + 2d_i^2\rho_{ij} E[(\varepsilon_{d_j}^{H})^2\varepsilon_{\rho_{ij}}^{H}] \\
&+ 2d_i\rho_{ij}^2 E[\varepsilon_{d_i}^{H}(\varepsilon_{d_j}^{H})^2] + 2d_i^2 d_j E[(\varepsilon_{\rho_{ij}}^{H})^2\varepsilon_{d_j}^{H}] + 2d_i d_j^2 E[\varepsilon_{d_i}^{H}(\varepsilon_{\rho_{ij}}^{H})^2] + 4d_i\rho_{ij} E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}(\varepsilon_{d_j}^{H})^2] \\
&+ 4d_j\rho_{ij} E[(\varepsilon_{d_i}^{H})^2\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}] + 4d_i d_j E[\varepsilon_{d_i}^{H}(\varepsilon_{\rho_{ij}}^{H})^2\varepsilon_{d_j}^{H}] - 2d_i d_j E[\varepsilon_{d_j}^{H}\varepsilon_{\rho_{ij}}^{H}]E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}] \\
&- 2d_i\rho_{ij} E[\varepsilon_{d_j}^{H}\varepsilon_{\rho_{ij}}^{H}]E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] - 2d_j\rho_{ij} E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}]E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}] + 2d_i E[\varepsilon_{d_i}^{H}(\varepsilon_{\rho_{ij}}^{H})^2(\varepsilon_{d_j}^{H})^2] \\
&- 2d_i E[\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}]E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}] + 2d_j E[(\varepsilon_{d_i}^{H})^2(\varepsilon_{\rho_{ij}}^{H})^2\varepsilon_{d_j}^{H}] - 2d_j E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}]E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}] \\
&+ 2\rho_{ij} E[(\varepsilon_{d_i}^{H})^2\varepsilon_{\rho_{ij}}^{H}(\varepsilon_{d_j}^{H})^2] - 2\rho_{ij} E[\varepsilon_{d_i}^{H}\varepsilon_{d_j}^{H}]E[\varepsilon_{d_i}^{H}\varepsilon_{\rho_{ij}}^{H}\varepsilon_{d_j}^{H}]
\end{aligned}
$$

In order to compute the corresponding MSE's we need the expression for the squared bias, which are given by

27

$$
\begin{aligned}
E[\varepsilon_{\sigma_{ij}}^{MF}]^2 &= \rho_{ij}^2 E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H]^2 \\
E[\varepsilon_{\sigma_{ij}}^{LF}]^2 &= d_i^2 E[\varepsilon_{d_j}^L \varepsilon_{\rho_{ij}}^L]^2 + d_j^2 E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L]^2 + \rho_{ij}^2 E[\varepsilon_{d_i}^L \varepsilon_{d_j}^L]^2 + E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L]^2 \\
&+ 2 d_i d_j E[\varepsilon_{d_j}^L \varepsilon_{\rho_{ij}}^L] E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L] + 2 d_i \rho_{ij} E[\varepsilon_{d_j}^L \varepsilon_{\rho_{ij}}^L] E[\varepsilon_{d_i}^L \varepsilon_{d_j}^L] + 2 d_i E[\varepsilon_{d_j}^L \varepsilon_{\rho_{ij}}^L] E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L] \\
&+ 2 d_j \rho_{ij} E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L] E[\varepsilon_{d_i}^L \varepsilon_{d_j}^L] + 2 d_j E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L] E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L] + 2 \rho_{ij} E[\varepsilon_{d_i}^L \varepsilon_{d_j}^L] E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L] \\
E[\varepsilon_{\sigma_{ij}}^{HF}]^2 &= d_i^2 E[\varepsilon_{d_j}^H \varepsilon_{\rho_{ij}}^H]^2 + d_j^2 E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H]^2 + \rho_{ij}^2 E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H]^2 + E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H \varepsilon_{d_j}^H]^2 \\
&+ 2 d_i d_j E[\varepsilon_{d_j}^H \varepsilon_{\rho_{ij}}^H] E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H] + 2 d_i \rho_{ij} E[\varepsilon_{d_j}^H \varepsilon_{\rho_{ij}}^H] E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] + 2 d_i E[\varepsilon_{d_j}^H \varepsilon_{\rho_{ij}}^H] E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H \varepsilon_{d_j}^H] \\
&+ 2 d_j \rho_{ij} E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H] E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] + 2 d_j E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H] E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H \varepsilon_{d_j}^H] + 2 \rho_{ij} E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] E[\varepsilon_{d_i}^H \varepsilon_{\rho_{ij}}^H \varepsilon_{d_j}^H].
\end{aligned}
$$

Combining terms of the same order, we obtain that:

$$
\begin{aligned}
MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{LF}) &= C_{1,j}\rho_{ij}^2 d_i^2 + C_{2,i}\rho_{ij}^2 d_j^2 + C_{3,ij}\rho_{ij}^2 d_i d_j + C_{4,ij}\rho_{ij} d_i^2 d_j + C_{5,ij}\rho_{ij} d_i d_j^2 \\
&+ C_{6,ij}\rho_{ij}^2 d_i + C_{7,ij}\rho_{ij}^2 d_j + C_{8,ij}\rho_{ij} d_i d_j + C_{9,ij}\rho_{ij} d_i^2 + C_{10,ij}\rho_{ij} d_j^2 + C_{11,ij} d_i^2 d_j \\
&+ C_{12,ij} d_i d_j^2 + C_{13,ij}\rho_{ij} d_i + C_{14,ij}\rho_{ij} d_j + C_{15,ij} d_i d_j + C_{16,ij}\rho_{ij}^2 + C_{17,ij} d_i^2 \\
&+ C_{18,ij} d_j^2 + C_{19,ij}\rho_{ij} + C_{20,ij} d_i + C_{21,ij} d_j + C_{22,ij} \qquad (28) \\
&\equiv F_1(d_i, d_j, \rho_{ij}), \qquad (29)
\end{aligned}
$$

where

$$
\begin{aligned}
C_{1,j} &= V[\varepsilon_{d_j}^H] - V[\varepsilon_{d_j}^L] \\
C_{2,i} &= V[\varepsilon_{d_i}^H] - V[\varepsilon_{d_i}^L] \\
C_{3,ij} &= 2\left(E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] - E[\varepsilon_{d_i}^L \varepsilon_{d_j}^L]\right) \\
C_{4,ij} &= -2E[\varepsilon_{d_j}^L \varepsilon_{\rho_{ij}}^L] \\
C_{5,ij} &= -2E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L] \\
C_{6,ij} &= 2\left(E[\varepsilon_{d_i}^H (\varepsilon_{d_j}^H)^2] - E[\varepsilon_{d_i}^L (\varepsilon_{d_j}^L)^2]\right) \\
C_{7,ij} &= 2\left(E[(\varepsilon_{d_i}^H)^2 \varepsilon_{d_j}^H] - E[(\varepsilon_{d_i}^L)^2 \varepsilon_{d_j}^L)]\right) \\
C_{8,ij} &= -6E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L] \\
C_{9,ij} &= -2E[(\varepsilon_{d_j}^L)^2 \varepsilon_{\rho_{ij}}^L] \\
C_{10,ij} &= -2E[(\varepsilon_{d_i}^L)^2 \varepsilon_{\rho_{ij}}^L] \\
C_{11,ij} &= -2E[\varepsilon_{d_j}^L (\varepsilon_{\rho_{ij}}^L)^2] \\
C_{12,ij} &= -2E[\varepsilon_{d_i}^L (\varepsilon_{\rho_{ij}}^L)^2] \\
C_{13,ij} &= -4E[\varepsilon_{d_i}^L \varepsilon_{\rho_{ij}}^L (\varepsilon_{d_j}^L)^2] \\
C_{14,ij} &= -4E[(\varepsilon_{d_i}^L)^2 \varepsilon_{\rho_{ij}}^L \varepsilon_{d_j}^L] \\
C_{15,ij} &= 4\left(E[\varepsilon_{d_i}^H \varepsilon_{d_j}^H] V[\varepsilon_{\rho_{ij}}^L] - E[\varepsilon_{d_i}^L (\varepsilon_{\rho_{ij}}^L)^2 \varepsilon_{d_j}^L]\right) \\
C_{16,ij} &= E[(\varepsilon_{d_i}^H \varepsilon_{d_j}^H)^2] - E[(\varepsilon_{d_i}^L \varepsilon_{d_j}^L)^2]
\end{aligned}
$$

$$
\begin{aligned}
C_{17,ij} &= V[\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_j}^L\varepsilon_{\rho_{ij}}^L)^2] \\
C_{18,ij} &= V[\varepsilon_{d_i}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_i}^L\varepsilon_{\rho_{ij}}^L)^2] \\
C_{19,ij} &= -2E[(\varepsilon_{d_i}^L)^2\varepsilon_{\rho_{ij}}^L(\varepsilon_{d_j}^L)^2] \\
C_{20,ij} &= 2\left(E[\varepsilon_{d_i}^H(\varepsilon_{d_j}^H)^2]V[\varepsilon_{\rho_{ij}}^L] - E[\varepsilon_{d_i}^L(\varepsilon_{\rho_{ij}}^L)^2(\varepsilon_{d_j}^L)^2]\right) \\
C_{21,ij} &= 2\left(E[(\varepsilon_{d_i}^H)^2\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_i}^L)^2(\varepsilon_{\rho_{ij}}^L)^2\varepsilon_{d_j}^L]\right) \\
C_{22,ij} &= V[\varepsilon_{d_i}^H\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_i}^L\varepsilon_{\rho_{ij}}^L\varepsilon_{d_j}^L)^2].
\end{aligned}
$$

Furthermore, the difference between the MSE's of the MF and the HF model is given by

$$
\begin{aligned}
MSE(\hat{\sigma}_{ij}^{MF}) - MSE(\hat{\sigma}_{ij}^{HF}) &= D_{1,ij}d_i^2d_j^2 + D_{2,ij}\rho_{ij}d_id_j^2 + D_{3,ij}\rho_{ij}d_i^2d_j + D_{4,ij}\rho_{ij}d_id_j + D_{5,ij}\rho_{ij}d_i^2 \\
&+ D_{6,ij}\rho_{ij}d_j^2 + D_{7,ij}d_i^2d_j + D_{8,ij}d_id_j^2 + D_{9,ij}\rho_{ij}d_i + D_{10,ij}\rho_{ij}d_j + D_{11,ij}d_id_j \\
&+ D_{12,ij}d_i^2 + D_{13,ij}d_j^2 + D_{14,ij}\rho_{ij} + D_{15,ij}d_i + D_{16,ij}d_j + D_{17,ij} \qquad (30) \\
&\equiv F_2(d_i, d_j, \rho_{ij}), \qquad (31)
\end{aligned}
$$

where

$$
\begin{aligned}
D_{1,ij} &= V[\varepsilon_{\rho_{ij}}^L] - V[\varepsilon_{\rho_{ij}}^H] \\
D_{2,ij} &= -2E[\varepsilon_{d_i}^H\varepsilon_{\rho_{ij}}^H] \\
D_{3,ij} &= -2E[\varepsilon_{d_j}^H\varepsilon_{\rho_{ij}}^H] \\
D_{4,ij} &= -6E[\varepsilon_{d_i}^H\varepsilon_{\rho_{ij}}^H\varepsilon_{d_j}^H] \\
D_{5,ij} &= -2E[(\varepsilon_{d_j}^H)^2\varepsilon_{\rho_{ij}}^H] \\
D_{6,ij} &= -2E[(\varepsilon_{d_i}^H)^2\varepsilon_{\rho_{ij}}^H] \\
D_{7,ij} &= -2E[\varepsilon_{d_j}^H(\varepsilon_{\rho_{ij}}^H)^2] \\
D_{8,ij} &= -2E[\varepsilon_{d_i}^H(\varepsilon_{\rho_{ij}}^H)^2] \\
D_{9,ij} &= -4E[\varepsilon_{d_i}^H\varepsilon_{\rho_{ij}}^H(\varepsilon_{d_j}^H)^2] \\
D_{10,ij} &= -4E[(\varepsilon_{d_i}^H)^2\varepsilon_{\rho_{ij}}^H\varepsilon_{d_j}^H] \\
D_{11,ij} &= 4\left(E[\varepsilon_{d_i}^H\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[\varepsilon_{d_i}^H(\varepsilon_{\rho_{ij}}^H)^2\varepsilon_{d_j}^H]\right) \\
D_{12,ij} &= V[\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_j}^H\varepsilon_{\rho_{ij}}^H)^2] \\
D_{13,ij} &= V[\varepsilon_{d_i}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_i}^H\varepsilon_{\rho_{ij}}^H)^2] \\
D_{14,ij} &= -2E[(\varepsilon_{d_i}^H)^2\varepsilon_{\rho_{ij}}^H(\varepsilon_{d_j}^H)^2] \\
D_{15,ij} &= 2\left(E[\varepsilon_{d_i}^H(\varepsilon_{d_j}^H)^2]V[\varepsilon_{\rho_{ij}}^L] - E[\varepsilon_{d_i}^H(\varepsilon_{\rho_{ij}}^H)^2(\varepsilon_{d_j}^H)^2]\right) \\
D_{16,ij} &= 2\left(E[(\varepsilon_{d_i}^H)^2\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_i}^H)^2(\varepsilon_{\rho_{ij}}^H)^2\varepsilon_{d_j}^H]\right) \\
D_{17,ij} &= V[\varepsilon_{d_i}^H\varepsilon_{d_j}^H]V[\varepsilon_{\rho_{ij}}^L] - E[(\varepsilon_{d_i}^H\varepsilon_{\rho_{ij}}^H\varepsilon_{d_j}^H)^2].
\end{aligned}
$$

$\square$

B Appendix

Stock	Mean	Max	Min	Std. dev.	Skewness	Kurtosis
			5-minute returns			
AXP	0.000	0.054	-0.034	0.0021	0.353	22.679
C	-0.100	0.060	-0.048	0.0022	0.235	24.834
GE	-0.050	0.039	-0.029	0.0018	0.257	15.937
HD	-0.061	0.033	-0.029	0.0021	0.078	12.680
IBM	0.080	0.074	-0.022	0.0017	0.849	36.369
JPM	0.039	0.056	-0.051	0.0023	0.198	22.785
			Daily returns			
AXP	0.004	0.094	-0.093	0.186	-0.049	5.917
C	-7.829	0.159	-0.130	0.179	0.312	10.732
GE	-3.911	0.099	-0.083	0.154	0.213	6.687
HD	-4.770	0.120	-0.086	0.184	0.466	6.369
IBM	6.225	0.123	-0.095	0.153	0.093	7.671
JPM	3.056	0.254	-0.164	0.203	1.258	20.051

Table B.1: Descriptive statistics of the 5-minute and daily returns over the period 01.01.2000 – 30.07.2008. The means are scaled by 10^4.

Stock	Mean	Max	Min	Std. dev.	Skewness	Kurtosis
			Realized Variance			
AXP	3.443	57.583	0.073	4.684	4.229	32.783
C	3.610	119.857	0.107	5.910	7.646	108.488
GE	2.429	51.402	0.103	3.174	4.904	46.966
HD	3.456	51.376	0.165	3.968	3.919	28.013
IBM	2.275	56.909	0.119	3.049	5.684	67.597
JPM	1.009	118.766	0.116	1.644	6.415	74.871
			Realized Covariance			
AXP-C	1.587	37.664	-0.548	2.775	5.320	46.132
AXP-GE	1.106	26.317	-1.467	1.845	5.899	58.081
AXP-HD	1.161	27.657	-2.452	1.967	5.327	47.599
AXP-IBM	0.917	23.434	-0.789	1.464	5.647	55.889
AXP-JPM	1.582	40.305	-0.975	2.812	6.098	62.128
C-GE	1.241	41.686	-0.583	2.117	7.019	91.587
C-HD	1.269	27.336	-0.928	2.175	5.019	39.512
C-IBM	1.028	36.726	-3.269	1.737	7.558	109.960
C-JPM	2.003	107.553	-0.472	3.995	11.349	245.086
GE-HD	1.040	26.852	-1.139	1.702	5.903	59.205
GE-IBM	0.901	24.054	-0.329	1.445	5.757	57.766
GE-JPM	1.197	49.23	-0.624	2.136	8.667	144.595
HD-IBM	0.875	18.318	-1.203	1.339	5.213	44.183
HD-JPM	1.233	26.399	-2.567	2.077	5.003	39.756
IBM-JPM	4.017	188.049	0.117	6.924	11.919	262.259

Table B.2: Descriptive statistics of realized variances of the six stocks. The realized variances are calculated from 5-minute intraday returns (measured in %), as described in the main text.
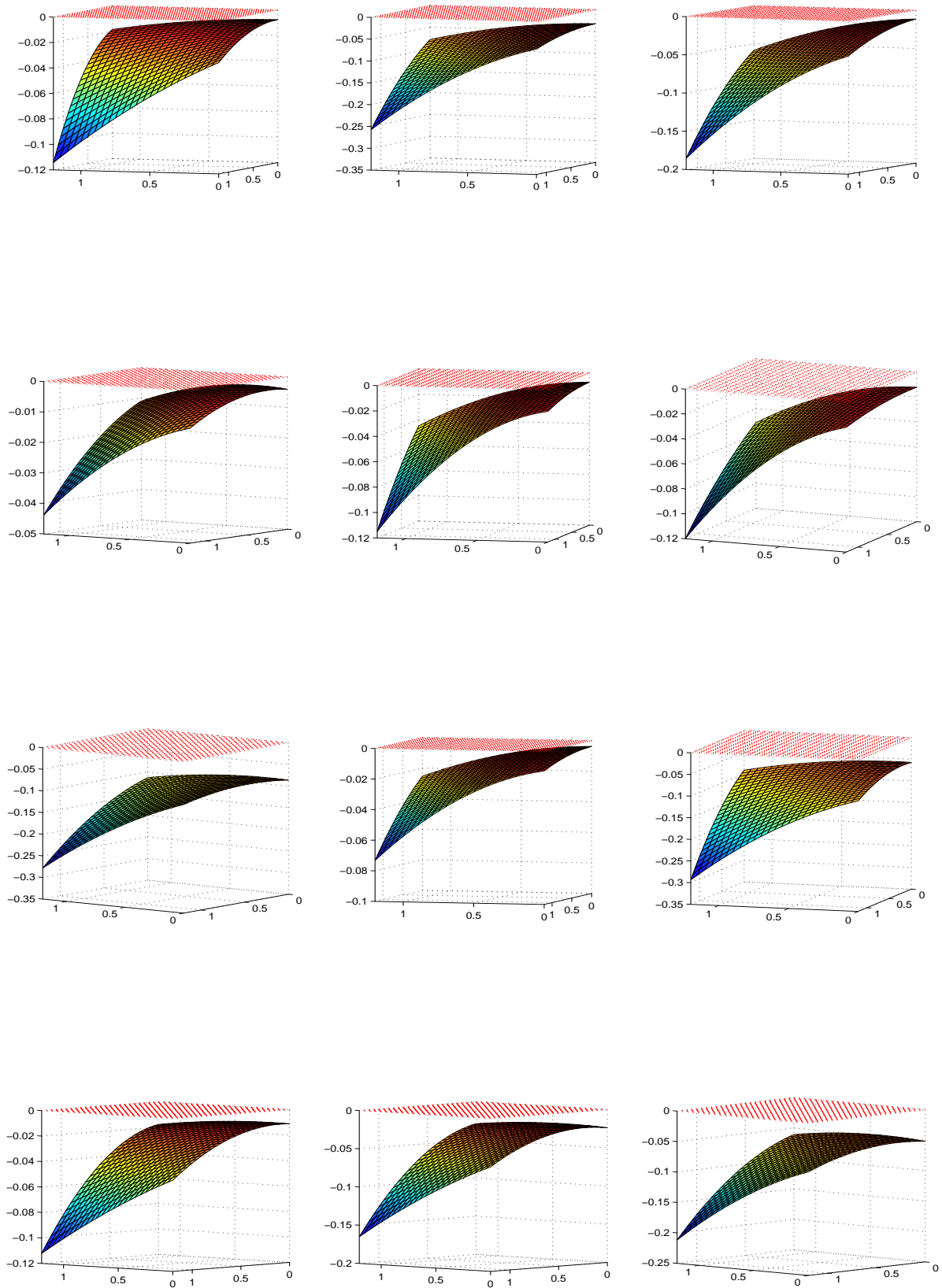
Figure B.1: Estimates of the quadratic polynomial in Equation (15) for the following stock pairs: first row – HD-AXP (left), HD-C (middle) and HD-GE (right), second row – IBM-AXP (left), IBM-C (middle) and IBM-GE (right), third row – IBM-HD (left), JPM-AXP (middle) and JPM-C (right), fourth row: JPM-GE (left), JPM-HD (middle) and JPM-IBM (right) and $\rho = 0.3$. The red surface indicates the zero limit.
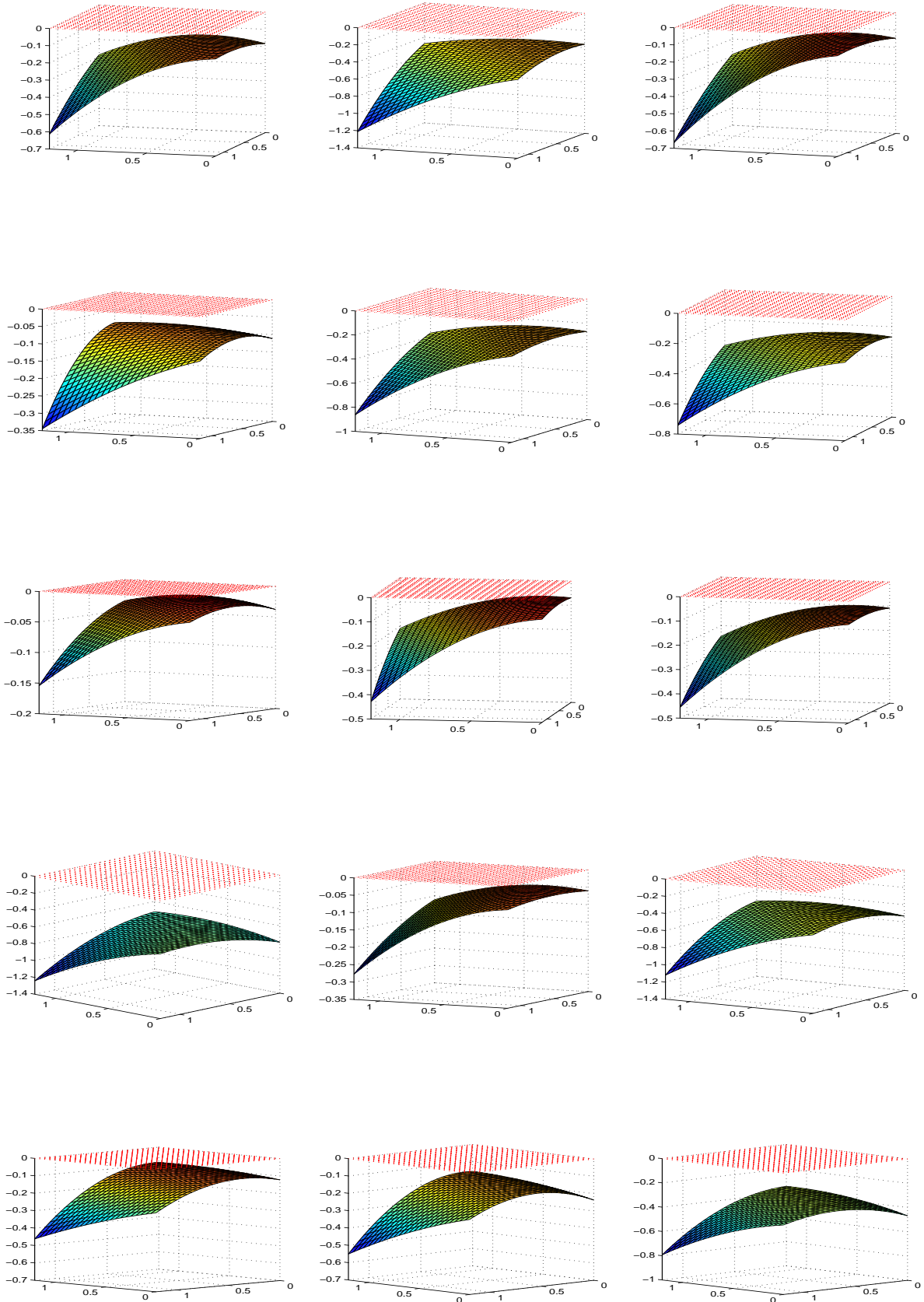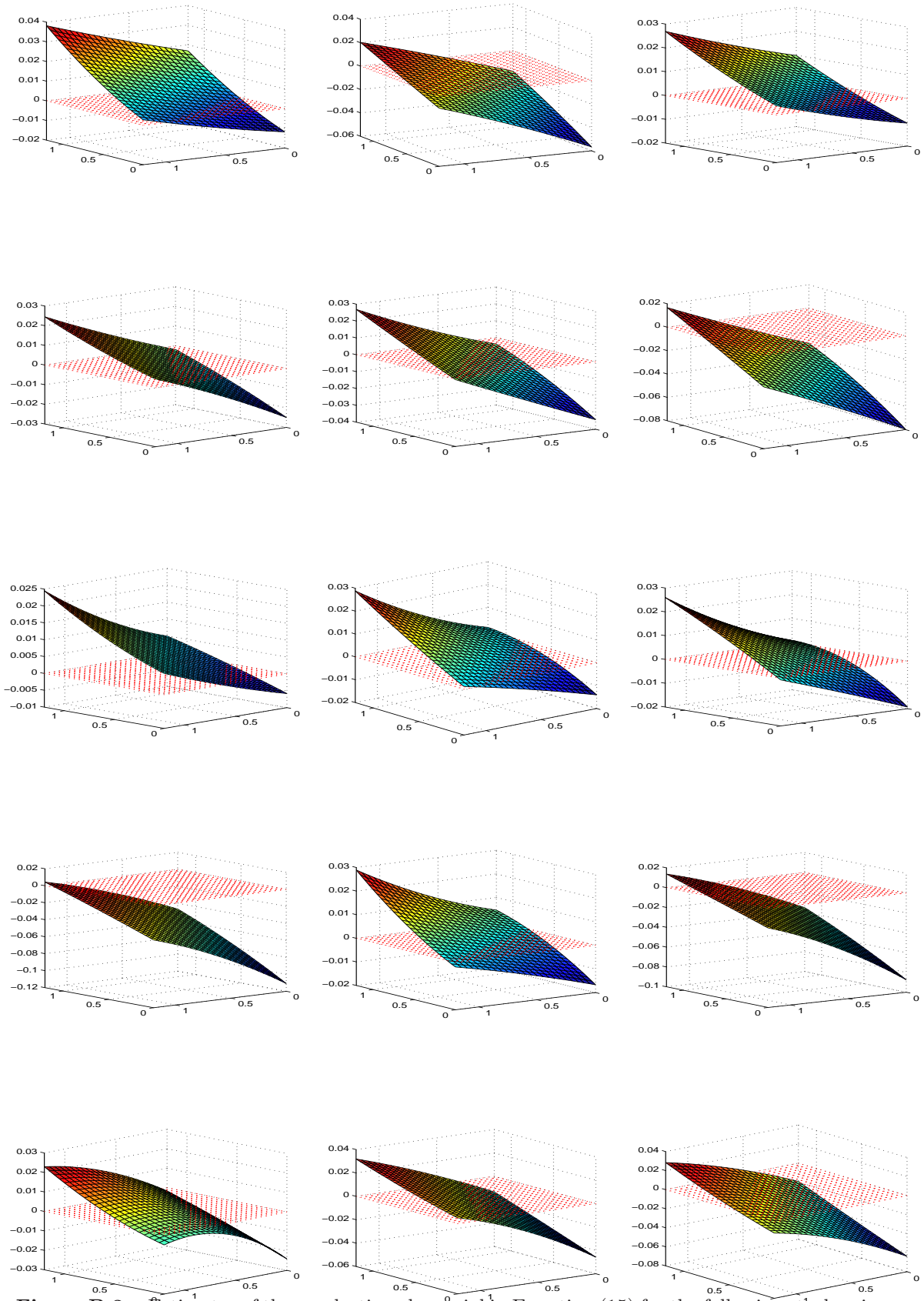
Figure B.2: Estimates of the quadratic polynomial in Equation (15) for the following stock pairs: first row – C-AXP (left), GE-AXP (middle) and GE-C (third plot), second row – HD-AXP (left), HD-C (middle) and HD-GE (right), third row – IBM-AXP (left), IBM-C (middle) and IBM-GE (right), fourth row – IBM-HD (left), JPM-AXP (middle) and JPM-C (right), fifth row: JPM-GE (left), JPM-HD (middle) and JPM-IBM (right) and $\rho = 0.9$. The red surface indicates the zero limit.

Figure B.3: Estimates of the quadratic polynomial in Equation (15) for the following stock pairs: first row – C-AXP (left), GE-AXP (middle) and GE-C (third plot), second row – HD-AXP (left), HD-C (middle) and HD-GE (right), third row – IBM-AXP (left), IBM-C (middle) and IBM-GE (right), fourth row – IBM-HD (left), JPM-AXP (middle) and JPM-C (right), fifth row: JPM-GE (left), JPM-HD (middle) and JPM-IBM (right) and $\rho = -0.3$. The red surface indicates the zero limit.
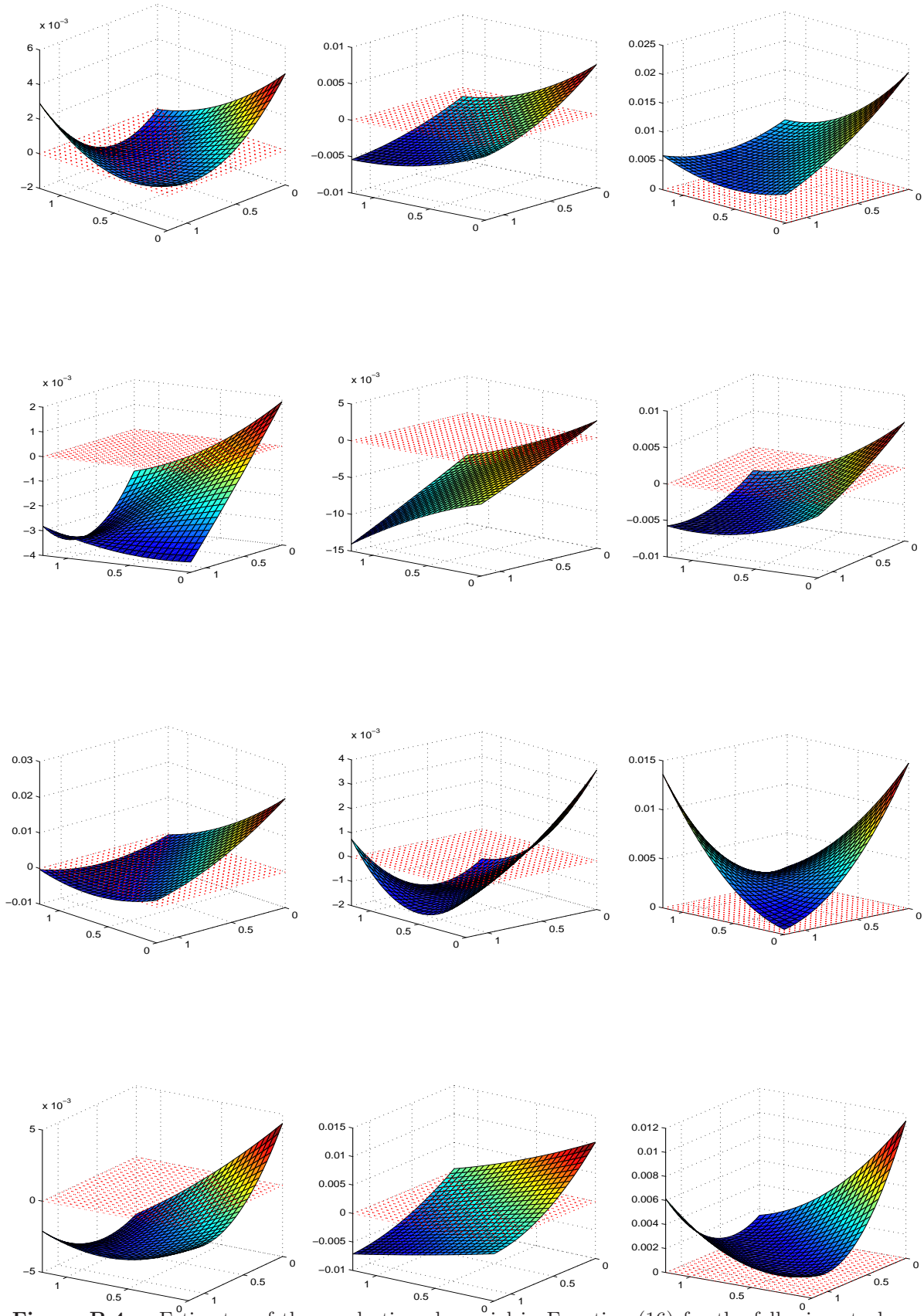
Figure B.4: Estimates of the quadratic polynomial in Equation (16) for the following stock pairs: first row – HD-AXP (left), HD-C (middle) and HD-GE (third plot), second row – IBM-AXP (left), IBM-C (middle) and IBM-GE (right), third row – IBM-HD (left), JPM-AXP (middle) and JPM-C (third plot), fourth row: JPM-GE (left), JPM-HD (middle) and JPM-IBM (right) and $\rho = 0.3$. The red surface indicates the zero limit.
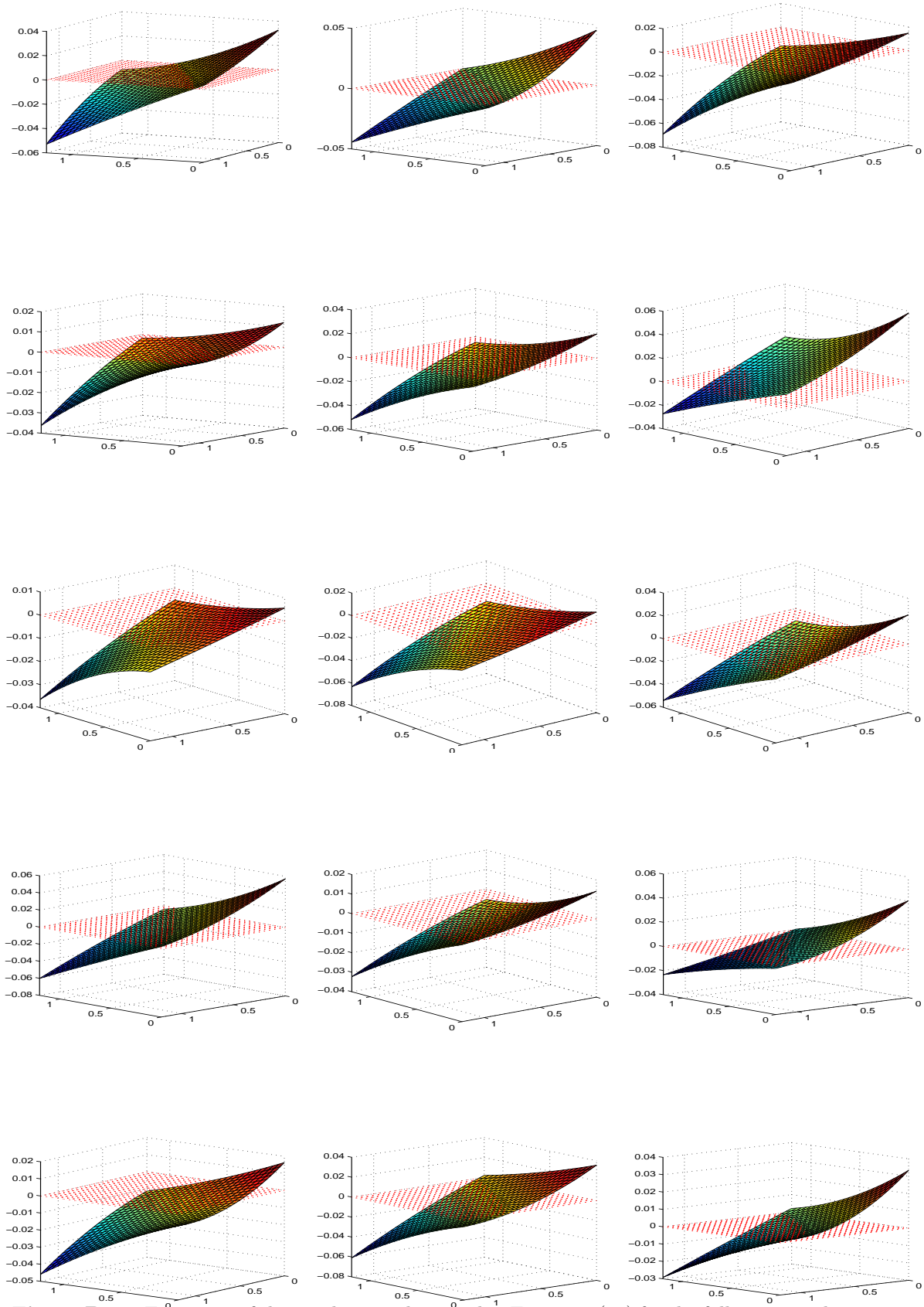
Figure B.5: Estimates of the quadratic polynomial in Equation (16) for the following stock pairs: first row – C-AXP (left), GE-AXP (middle) and GE-C (third plot), second row – HD-AXP (left), HD-C (middle) and HD-GE (right), third row – IBM-AXP (left), IBM-C (middle) and IBM-GE (right), fourth row – IBM-HD (left), JPM-AXP (middle) and JPM-C (right), fifth row: JPM-GE (left), JPM-HD (middle) and JPM-IBM (right) and $\rho = 0.9$. The red surface indicates the zero limit.
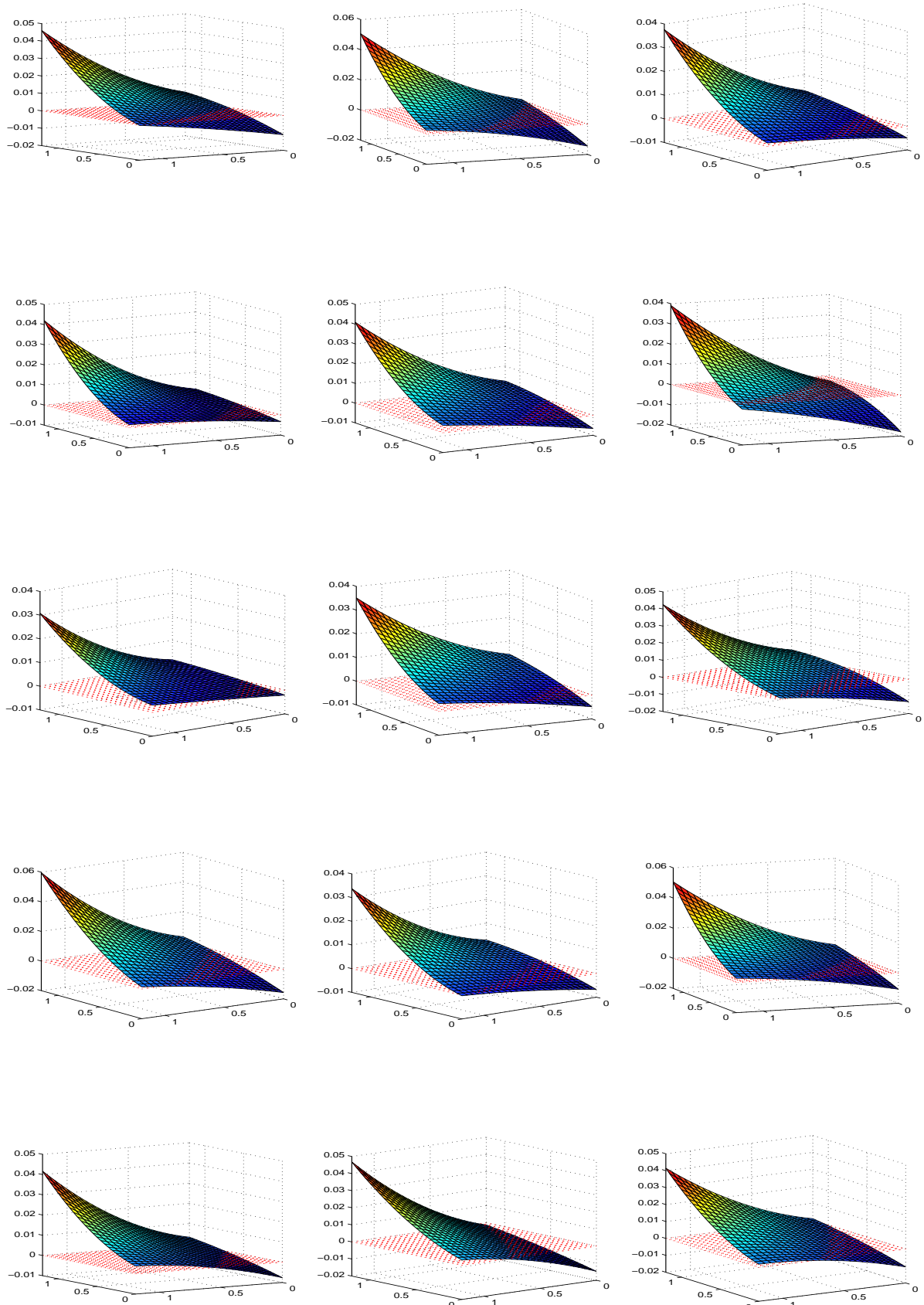
Figure B.6: Estimates of the quadratic polynomial in Equation (16) for the following stock pairs: first row – C-AXP (left), GE-AXP (middle) and GE-C (third plot), second row – HD-AXP (left), HD-C (middle) and HD-GE (right), third row – IBM-AXP (left), IBM-C (middle) and IBM-GE (right), fourth row – IBM-HD (left), JPM-AXP (middle) and JPM-C (right), fifth row: JPM-GE (left), JPM-HD (middle) and JPM-IBM (right) and $\rho = -0.3$. The red surface indicates the zero limit.

Research Papers
2011

2010-66:	Ole E. Barndorff-Nielsen, David G. Pollard and Neil Shephard: Integer-valued Lévy processes and low latency financial econometrics
2010-67:	Shin Kanaya and Dennis Kristensen: Estimation of Stochastic Volatility Models by Nonparametric Filtering
2010-68:	Dennis Kristensen and Anders Rahbek: Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models
2010-69:	Søren Johansen: The analysis of nonstationary time series using regression, correlation and cointegration –with an application to annual mean temperature and sea level
2010-70:	Søren Johansen and Morten Ørregaard Nielsen: A necessary moment condition for the fractional functional central limit theorem
2010-71:	Nektarios Aslanidis and Isabel Casas : Modelling asset correlations during the recent financial crisis: A semiparametric approach
2010-72:	Søren Johansen and Katarina Juselius: An invariance property of the common trends under linear transformations of the data
2010-73:	Peter Sandholt Jensen and Allan H. Würtz: Estimating the effect of a variable in a high-dimensional regression model
2010-74:	Peter R. Hansen, Asger Lunde and Valeri Voev: Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility and CoVolatility
2010-75:	Laurent A.F. Callot: A Bootstrap Cointegration Rank Test for Panels of VAR Models
2010-76:	Peter R. Hansen, Asger Lunde and James M. Nason: The Model Confidence Set
2011-01:	Cristina Amado and Timo Teräsvirta: Modelling Volatility by Variance Decomposition
2011-02:	Timo Teräsvirta: Nonlinear models for autoregressive conditional heteroskedasticity
2011-03:	Roxana Halbleib and Valeri Voev: Forecasting Covariance Matrices: A Mixed Frequency Approach