Encompassing Tests for Higher-Order Elicitable Functionals

Timo Dimitriadis[†] Julie Schnaitmann[‡]

February 10, 2019

First Version: June 18, 2018

Abstract

We introduce encompassing tests for forecasts of higher-order elicitable functionals such as the variance and the Expected Shortfall (ES). Encompassing tests rely on the existence of strictly consistent loss functions for the forecasted functionals under consideration, which do not exist for the variance and the ES. However, for these functionals, such loss functions exist for the pairs (mean,variance) and (quantile, ES). We utilize these joint loss functions in order to introduce joint encompassing tests for the quantile and the ES as well as stand-alone encompassing tests for the ES. These tests provide a theoretical justification for forecast combination of the ES when encompassing is rejected. We show through simulation studies that all proposed tests are reasonably sized and exhibit good power properties against general alternatives in typical financial applications. In the empirical application, we apply encompassing tests in order to demonstrate the superiority of forecast combination methods for the ES.

Keywords: Encompassing, Joint Elicitability, Expected Shortfall, Variance, Forecast Combination

1. Introduction

Evaluation and comparison of statistical forecasts is an essential issue which heavily relies on the existence of suitable loss and identification functions for the forecasts under consideration. Classically, point forecasts are issued for a central tendency of the variable under consideration, which are usually thought of as forecasts for the mean. An appropriate loss function for the evaluation of mean forecasts is the well-known squared loss function. In recent years, the emphasis was widened and also includes forecasting of other non-central functionals of the predictive distribution. E.g. in the financial world, attention has shifted towards forecasting of functionals which capture the risk involved in the financial products. Examples for these are the variance (volatility), quantiles (Value at Risk, VaR), expectiles and the Expected Shortfall (ES). For certain important functionals such as the ES and the variance (in the presence of a non-zero mean), there do not exist suitable loss functions for the evaluation of the forecasts (Gneiting, 2011b). We then say that these functionals are *non-elicitable*.

The ES currently attains a lot of attention due to its recent introduction into the Basel Accords as target risk measure for the regulation of the international financial markets (Basel Committee, 2016, 2017). The ES of some random variable at probability level $\alpha \in (0, 1)$ is defined as the mean, given that the variable is smaller than its α-quantile. It is favored over the VaR due to its ability to capture extreme financial risks

[†]University of Konstanz, Department of Economics, 78457 Konstanz, Germany, e-mail: timo.dimitriadis@uni-konstanz.de

[‡]University of Konstanz, Department of Economics, 78457 Konstanz, Germany, e-mail: julie.schnaitmann@uni-konstanz.de

beyond the VaR and as it overcomes certain deficiencies of the VaR such as sub-additivity (Artzner et al., 1999). The second candidate functional we consider in this article is the variance in presence of a non-zero mean. The variance is a risk measure which is often used in practice. In the simplified case of a zero mean which is often assumed for daily (or high frequency) returns, the variance equals the second moment of the return distribution and can be evaluated as an expectation (Gneiting, 2011a; Patton, 2011). However, in the presence of a non-zero mean, such as it is common for financial returns at lower frequencies or e.g. for monthly inflation rates, evaluation of the variance as a second moment is incorrect and one has to consider joint evaluation of mean and the variance forecasts.

So far, the non-elicitability of these functionals impeded forecast evaluation techniques such as encompassing tests to be applied correctly to these important functionals. Forecast encompassing tests for two competing forecasts A and B test the null hypothesis that forecast A *performs not worse* than any (linear) combination of forecast A and B (Hendry and Richard, 1982; Mizon and Richard, 1986; Diebold, 1989; Giacomini and Komunjer, 2005; Newbold and Harvey, 2007; Clements and Harvey, 2009). This null hypothesis allows for the convenient interpretation that forecast B does not add any information to forecast A and thus, forecast A is superior to forecast B. This is carried out by testing whether the combination weight of forecast B of the optimal forecast combination significantly deviates from zero, i.e. from simply choosing forecast A and ignoring forecast B in the optimal combination. For this, the existence of such a loss function is inevitable for two reasons. First, the *superior performance* is defined in the statistical sense by using strictly consistent loss functions. Second, loss (or identification) functions are utilized to set up a regression framework for the functional under consideration in order to estimate the optimal forecast combination weights.

For functionals such as the variance and the ES, the difficulty of non-elicitability can be overcome by considering multiple functionals at the same time, stacked as vectors and interpreted as vector-valued functionals (Lambert et al., 2008; Fissler and Ziegel, 2016, 2018). Then, we obtain that the pairs (mean, variance) and (quantile, ES) are *jointly elicitable* (or 2-elicitable). Thus, joint loss functions for the pairs of functionals do exist in these cases and we can utilize these loss functions for the definition of encompassing tests.

In this work, we propose encompassing tests for these higher-order elicitable funcitonals where we mainly focus on the two most prominent candidates, the variance and the ES. For this, we first introduce encompassing tests which test forecast encompassing of the joint vector of jointly elicitable functionals, i.e. for the vector (mean, variance) and the vector (quantile, ES). As the main objective of these tests is the second functional, i.e. the variance or the ES, we also propose encompassing tests which estimate the optimal combination weights for the joint vector, however, only test for the parameters of interest associated with the variance or the ES respectively. In the case of the ES, we also propose a third variant of the encompassing test which only needs ES forecasts and can consequently also be applied in situations where the person evaluating the forecasts only has forecasts for the ES at hand. This situation is particularly relevant due to the current set of rules imposed by the Basel Committee of Banking Supervision, which only impose the financial institutions to report ES forecasts (Basel Committee, 2016, 2017).

We implement the encompassing tests through GMM-estimation of the optimal combination weights, which has the following advantages: (1) we are agnostic about the origin of the investigated forecasts. We only impose that the forecasts are generated by a fixed procedure over time and that some moments of the forecasts exist. (2) we can test *conditional* encompassing at each time point instead of classical unconditional encompassing. This is achieved by using vectors of instruments in the GMM setting which can capture all the available information at the time the forecasts are issued. (3) our tests are derived in an environment with asymptotically nonvanishing estimation uncertainty in the estimation of the forecasting procedures which enables testing nested models. This links our test to the approach of Giacomini and White (2006) and Giacomini and Komunjer (2005).

Tests for forecast encompassing are fundamentally different from the literature on testing equal predictive ability following e.g. Diebold and Mariano (1995); West (1996); Giacomini and White (2006);

West (2006) in the following way. The null hypothesis of tests of equal predictive ability is that both competing forecasts perform equally well on average, i.e. a rejection of the test lets the researcher conclude that one forecast (or forecasting method) is significantly outperforming the other one. In contrast, the null hypothesis of encompassing tests is that one forecast encompasses the other, i.e. that the second forecast does not add any information to the first.

The classical idea of forecast encompassing goes back to Hendry and Richard (1982), Chong and Hendry (1986) and Mizon and Richard (1986) and is developed for mean forecasts and the squared loss function. Reviews on the technique of encompassing tests can be found in Newbold and Harvey (2007) and Clements and Harvey (2009). Harvey and Newbold (2000) extend the general encompassing technique which usually focuses on *two* competing forecasts to multiple forecast encompassing. Giacomini and Komunjer (2005) develop (conditional) encompassing of quantile forecasts and focus on encompassing tests for *methods* instead of *models*, which links this approach to the testing procedure of Giacomini and White (2006). Clements and Harvey (2010) generalize encompassing tests to probabilistic forecasts by relying on strictly consistent scoring rules. Giacomini and Komunjer (2005) and Clements and Harvey (2010) already investigate extensions of encompassing to more complicated functionals of the conditional distribution. Our work peruses this paths by developing encompassing tests for higher-order elicitable functionals where only joint loss functions for vector-valued functionals are available.

Tests for forecast encompassing are commonly used in order to establish a theoretical basis for forecast combination in cases when neither forecast encompasses its competitor (Clements and Harvey, 2009; Newbold and Harvey, 2007; Giacomini and Komunjer, 2005). For this, we test both encompassing hypotheses, i.e. that forecast A encompasses forecast B and that forecast B encompasses forecast A. In the case that both these hypotheses are rejected, this implies that for neither of the two forecasts, its respective competitor already incorporates the full information of both forecasts. Thus, we can conclude that a forecast combination incorporates significantly more information than both forecasts individually. In general, combining forecasts stemming from different models, estimation approaches, data or information sets has several advantages over stand-alone forecasts, where e.g. Timmermann (2006) provides three arguments in favor of forecast combination: first, there are diversification gains stemming from the combination of forecasts computed from different assumptions, specifications or information sets. Second, combined forecasts tend to be robust against structural breaks. Third, the influence of potential misspecification of the individual models is reduced due to averaging over a set of forecasts stemming from several models. Giacomini and Komunjer (2005) further argue that forecast combination is particularly advantageous for risk measures (quantiles) with small probability levels, as it is customary for the VaR and the ES. These extreme risk measure are very sensitive to the few observations in the tails of the empirical distribution of the sample, and thus, forecast combinations based on different information sets can be seen as a way to make the forecast performance more robust to the effects of sample-specific factors. Halbleib and Pohlmeier (2012) and Bayer (2018) provide further empirical evidence in favor of forecast combinations, especially for financial risk measures and in turbulent financial times.

We conduct a simulation study in order to assess the size and power properties of our ES encompassing tests in two realistic financial settings. The first simulation setup is based on two location-scale GARCH processes for which all three tests are correctly specified. The second setup is based on CARE models of Taylor (2017), where the data generating process is outside the class of location-scale models. Thus, the third strict ES encompassing test is subject to misspecification. As the data generating processes are calibrated to real financial data, we can consequently assess the degree of misspecification these tests face in financial applications. We find that all tests exhibit approximately the correct size, whereas the tests based on bootstrapped standard error perform better than the tests based on the asymptotic covariance. Furthermore, all tests exhibit satisfactory power against all types of misspecification we consider in the simulations. This also holds for the strict ES encompassing test which shows that the test is robust to misspecifications usually encountered in financial applications.

We apply the new encompassing tests for the ES to the problem of evaluating ES forecasts for financial returns, which is currently of high importance due to recent introduction of the ES into the regulatory framework of the Basel Committee of Banking Supervision Basel Committee (2016, 2017). We use daily returns from the IBM stock and show that combined forecasts for the ES outperform the stand-alone models in five (six) of the ten considered pairwise combinations of forecasting models. In contrast, by applying VaR encompassing tests (Giacomini and Komunjer, 2005), we find that only in two out of ten cases, the combined VaR forecasts significantly outperform the stand-alone models. Furthermore, the third variant of our test which can be applied when only ES forecasts are available (without their accompanying VaR forecasts), exhibits very similar results. Thus, we can draw the following conclusions. First, forecast combination methods for the ES significantly increase the forecast accuracy. Second, our results imply that the gains from forecast combination for the ES are even more pronounced than they are for the VaR.

The rest of the paper is organized as follows. In Section 2, we introduce the idea of encompassing tests for higher order elicitable functionals and derive the asymptotic distribution of the test statistics. Section 3 presents an extensive simulation study analyzing the size and power properties of our test. In Section 4, we apply the testing procedure to daily returns of the IBM stock and thereby illustrate the power of forecast combination techniques. Section 5 provides concluding remarks. The proofs are deferred to Appendix A.

2. Theory

Consider a stochastic process $Z = \left\{Z_t : \Omega \to \mathbb{R}^{k+1}, k \in \mathbb{N}, t = 1, \dots, T\right\}$, which is defined on some common complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F} = \{\mathcal{F}_t, t = 1, \dots, T\}$ and $\mathcal{F}_t = \sigma\{Z_s, s \le t\}$. We partition the stochastic process as $Z_t = (Y_t, X_t)$, where $Y_t : \Omega \to \mathbb{R}$ is an absolutely continuous random variable of interest and $X_t : \Omega \to \mathbb{R}^k$ is a vector of explanatory variables. We denote the conditional distribution of Y_{t+1} given the information set $\mathcal{F}_t$ by F_t. Accordingly, $\mathbb{E}_t$, Var_t and f_t denote the expectation, variance and density corresponding to F_t. We further denote the class of conditional distributions of Y_{t+1} given the information set $\mathcal{F}_t$ by $\mathcal{P}$.

Our goal is to construct encompassing tests for certain *statistical functionals* (also known as characteristics or properties) of the the predictive distribution F_t. We formally define an s-dimensional functional, $s \in \mathbb{N}$,

$$\Gamma : \mathcal{P} \to \mathbb{R}^s, \qquad F_t \mapsto \Gamma(F_t) \tag{2.1}$$

as a mapping from the space of predictive distributions to a real-valued s-dimensional vector. Many forecasts are issued as (vector-valued) point forecasts even though the forecaster has indeed a full predictive distribution F_t in mind (Gneiting, 2011a). Typical functionals which attract a lot of attention in practice are the mean, the variance, quantiles, expectiles and the ES. In many cases, the person or institution evaluating the forecasts is different from the forecaster and consequently only has access to the issued point forecasts and the full predictive distribution stays concealed. One prominent example for this are risk forecasts, where the Basel Committee requires the financial institutions to report forecasts for the ES of their forecasted return distributions (Basel Committee, 2016, 2017).

In this context, two important properties of statistical functionals are *elicitability* and *identifiability*. Elicitability means that there exists a strictly consistent loss function, i.e. a loss function $\rho_\Gamma(Y, x)$ depending on the random variable Y and the issued forecast x, whose expectation $\mathbb{E}\left[\rho_\Gamma(Y, \cdot)\right]$ is uniquely minimized by the true functional Γ. Using such a loss function, one can assess the quality of issued forecasts by comparing their average losses induced by the realized predicted variable of interest. Identifiability means that there exists an identification function $\varphi_\Gamma(Y, x)$, whose expectation $\mathbb{E}\left[\varphi_\Gamma(Y, x)\right]$ equals zero if and only if x equals the true functional Γ. Identification functions can usually be used in order to test for forecast rationality, i.e. for testing whether the average of $\varphi_\Gamma(Y, \cdot)$ equals zero (Diebold and Lopez, 1996; Elliott et al., 2005; Patton and Timmermann, 2007). As a direct consequence, almost all of the literature on tests for forecast comparison and forecast rationality evolves around the associated loss and identification functions (Gneiting, 2011a).

Unfortunately, there exist statistical functionals for which there are no such strictly consistent loss functions. Important examples for these functionals are the variance (with non-zero mean), the ES, the mode, the minimum and the maximum (Gneiting, 2011a; Heinrich, 2014; Fissler and Ziegel, 2016). Some of these can be made elicitable at some *higher-order*, by simply considering a pair of functionals such as e.g. the variance together with the mean, and the ES at level α jointly with the associated α-quantile (Lambert et al., 2008; Gneiting, 2011a; Fissler and Ziegel, 2016). In this work, we make use of joint elicitability which allows for two crucial ingredients of forecast encompassing test. First, we can jointly evaluate these forecasts using a strictly consistent loss function and thus define forecast superiority. Second, these loss and identification functions allow for M- and GMM-estimation of the associated semiparametric regression equations, which are a crucial ingredient of forecast encompassing tests.

In order to conduct the forecast evaluation in an out-of-sample fashion, we divide the sample size T in an in-sample part of size m and an out-of-sample part of size n such that $T = m + n$. The in-sample period is used to generate the forecasts, e.g. by estimating model parameters in parametric approaches. The out-of-sample period, $t = m + 1, \dots, T$ is used for the evaluation of the forecasts. Following Giacomini and Komunjer (2005), we pose little restrictions on how to generate the forecasts, where we allow for parametric, semiparametric or nonparametric techniques, and also allow for nested and non-nested forecasting procedures.

Let $\gamma_{t,m}$ denote the model parameters at time t (or alternatively the semi- or non-parametric estimator used in the construction of the forecasts), which are estimated using the previous m data points. We assume that the one-step ahead forecasts $\hat{f}_t = f\big(\hat{\gamma}_{t,m}, Z_t, Z_{t-1}, \dots\big)$, issued with the knowledge available at time t, are fixed functitons $f(\cdot)$ over time. This construction allows for both, fixed forecasting schemes, where the model parameters $\gamma_{t,m}$ are only estimated once, and also rolling window forecasting schemes, where the parameters $\gamma_{t,m}$ are re-estimated in each step.

In the following section, we formally introduce the concept of forecast encompassing in the classical case of one-dimensional, real-valued and 1-elicitable functionals. However, as e.g. the variance and the ES are not 1-elicitable, but 2-elicitable together with the mean and the quantile respectively, the main focus of this paper is to generalize encompassing tests to higher-order elicitable functionals, which is discussed in Section 2.2.

2.1. Encompassing Test for 1-Elicitable Functionals

Following Hendry and Richard (1982), Mizon and Richard (1986), Diebold (1989) and Giacomini and Komunjer (2005), we formally introduce the general concept of forecast encompassing in the following for one-dimensional, real-valued and elicitable functionals. We assume that two competing forecasters predict the variable of interest Y_{t+1} and issue point forecasts $\hat{f}_{t,1}$ and $\hat{f}_{t,2}$ for a given functional $\Gamma(F_t)$. Furthermore, let $\rho_\Gamma\big(Y_{t+1}, f\big)$ be a strictly consistent loss function for Γ, which means that

$$\mathbb{E}_t\left[\rho_\Gamma(Y_{t+1}, \Gamma(F_t))\right] \leq \mathbb{E}_t\left[\rho_\Gamma(Y_{t+1}, x)\right] \qquad \text{a.s.} \tag{2.2}$$

for all x in the domain of Y_{t+1}. Then, we say that the forecast $\hat{f}_{1,t}$ conditionally encompasses $\hat{f}_{2,t}$ when

$$\mathbb{E}_t\left[\rho_\Gamma\big(Y_{t+1}, \hat{f}_{1,t}\big)\right] \leq \mathbb{E}_t\left[\rho_\Gamma\big(Y_{t+1}, \theta_{1,t}\hat{f}_{1,t} + \theta_{2,t}\hat{f}_{2,t}\big)\right] \qquad \text{a.s.} \tag{2.3}$$

for all $\big(\theta_{1,t}, \theta_{2,t}\big) \in \Theta$ and for all $t = m, \dots, T-1$. Equation (2.3) implies that, in terms of the loss induced by ρ_Γ, forecast $\hat{f}_{1,t}$ is at least as good as any (linear) combination of $\hat{f}_{1,t}$ and $\hat{f}_{2,t}$. This implies that forecast $\hat{f}_{2,t}$ does not add any information on Y_{t+1} which is not already incorporated in $\hat{f}_{1,t}$. We define $\big(\theta^*_{1,t}, \theta^*_{2,t}\big)$ as the parameters which minimize the loss under the conditional expectation,

$$\big(\theta^*_{1,t}, \theta^*_{2,t}\big) = \underset{(\theta_{1,t}, \theta_{1,t}) \in \Theta}{\arg\min}\; \mathbb{E}_t\left[\rho_\Gamma\big(Y_{t+1}, \theta_{1,t}\hat{f}_{1,t} + \theta_{2,t}\hat{f}_{2,t}\big)\right] \qquad \text{a.s.,} \tag{2.4}$$

where $\Theta \subseteq \mathbb{R}^2$ is some compact, non-empty and convex parameter space. Then, it obviously holds that

$$\mathbb{E}_t \left[\rho_\Gamma\big(Y_{t+1}, \theta_{1,t}\hat{f}_{1,t} + \theta_{2,t}\hat{f}_{2,t}\big)\right] \geq \mathbb{E}_t \left[\rho_\Gamma\big(Y_{t+1}, \theta^*_{1,t}\hat{f}_{1,t} + \theta^*_{2,t}\hat{f}_{2,t}\big)\right] \qquad \text{a.s.}$$

for all $(\theta_{1,t}, \theta_{2,t}) \in \Theta$ and for all $t = m, \ldots, T-1$. In particular, this implies that

$$\mathbb{E}_t \left[\rho_\Gamma\big(Y_{t+1}, \hat{f}_{1,t}\big)\right] \geq \mathbb{E}_t \left[\rho_\Gamma\big(Y_{t+1}, \theta^*_{1,t}\hat{f}_{1,t} + \theta^*_{2,t}\hat{f}_{2,t}\big)\right] \qquad \text{a.s..} \tag{2.5}$$

Combining (2.3) and (2.5) yields the following definition of forecast encompassing

Definition 2.1 (Forecast Encompassing for Real-Valued and Elicitable Functionals). We say that forecast $\hat{f}_{1,t}$ encompasses $\hat{f}_{2,t}$ at time t for the real-valued and elicitable functional $\Gamma : \mathcal{P} \to \mathbb{R}$ if and only if

$$\mathbb{E}_t \left[\rho_\Gamma\big(Y_t, \hat{f}_{1,t}\big)\right] = \mathbb{E}_t \left[\rho_\Gamma\big(Y_t, \theta^*_{1,t}\hat{f}_{1,t} + \theta^*_{2,t}\hat{f}_{2,t}\big)\right], \tag{2.6}$$

with respect to the loss function ρ_Γ, which is equivalent to

$$\big(\theta^*_{1,t}, \theta^*_{2,t}\big) = (1, 0). \tag{2.7}$$

The preceding definition shows that one main ingredient for forecast encompassing is the specification of the underlying loss function. One could state Definition 2.1 using *any* loss function which has the appropriate dimensions. However, the inequality (2.3) only becomes meaningful if the loss function ρ_Γ is chosen appropriately for the functional Γ. It is possible to consider economically motivated loss functions, such as financial loss, utilities, etc. However, in this paper we only consider loss functions which are meaningful in the strong statistical sense of being strictly consistent for the underlying functional as defined in (2.2).

In fact, strict consistency of loss functions only implies that a correctly specified forecast exhibits the smallest loss in expectation. In reality however, competing forecasts are seldomly correctly specified, but rather somehow missspecified and we seek to know which one of competing forecasts is the one with a higher predictive power. Even though reasonable, the definition of strict consistency in (2.2) does not directly imply that the better of two misspecified forecasts receives the smaller loss. Holzmann and Eulert (2014) show that for competing forecasts which are based on increasing information sets and which are correctly specified given their underlying (but usually incomplete) information set (auto-calibrated), applying any strictly consistent loss function results in a correct ranking of the forecasts. In our case, we indeed have nested information sets as it obviously holds that $\sigma\left\{\hat{f}_{1,t}, \hat{f}_{2,t}\right\} \supseteq \sigma\left\{\hat{f}_{1,t}\right\}$. Thus, by further assuming that the issued forecasts are auto-calibrated given the forecasters information set, we can conclude that the ranking implies by (2.3) is indeed the correct one for any strictly consistent loss function.

2.2. Encompassing Test for Higher Order Elicitable Functionals

In this work, we consider statistical functionals such as the ES and the variance (in the presence of a non-zero mean) which are not elicitable, i.e. for which there does not exist such a strictly consistent loss function. Consequently, defining forecast encompassing such as in Definition 2.1 fails due to lack of an underlying and statistically meaningful loss function. However, the functionals we consider are elicitable jointly with some other functional, such as the variance is jointly elicitable with the mean and the α-ES is jointly elicitable with the α-quantile. Thus, Definition 2.1 can be generalized by using a candidate from these class of strictly consistent joint loss functions.

Definition 2.2 (Forecast Encompassing for Higher-Order Elicitable Functionals). We say that the vector of forecasts $\hat{\boldsymbol{f}}_{1,t} = \big(f^{(1)}_{1,t}, \ldots, f^{(s)}_{1,t}\big)$ encompasses $\hat{\boldsymbol{f}}_{2,t} = \big(\hat{f}^{(1)}_{2,t}, \ldots, \hat{f}^{(s)}_{2,t}\big)$ at time t for the vector-valued, elicitable functional $\Gamma : \mathcal{P} \to \mathbb{R}^s$ if and only if

$$\mathbb{E}_t \left[\rho_\Gamma\left(Y_{t+1}, \hat{f}^{(1)}_{1,t}, \ldots, \hat{f}^{(s)}_{1,t}\right)\right] = \mathbb{E}_t \left[\rho_\Gamma\left(Y_{t+1}, \theta^{*(1)}_{1,t}\hat{f}^{(1)}_{1,t} + \theta^{*(1)}_{2,t}\hat{f}^{(1)}_{2,t}, \ldots, \theta^{*(s)}_{1,t}\hat{f}^{(s)}_{1,t} + \theta^{*(s)}_{2,t}\hat{f}^{(s)}_{2,t}\right)\right], \tag{2.8}$$

with respect to the loss function ρ_Γ, which is equivalent to

$$\left(\theta_{1,t}^{*(1)}, \theta_{2,t}^{*(1)}\right) = \cdots = \left(\theta_{1,t}^{*(s)}, \theta_{2,t}^{*(s)}\right) = (1,0). \tag{2.9}$$

This definition allows for setting up joint encompassing tests for any vector-valued and elicitable functional of interest. Besides the maybe most interesting case of non-elicitable but higher-order elicitable functionals as the variance and the ES, this definition also enables us to consider forecast encompassing for combinations of 1-elicitable functoinals. Such an analysis might be interesting for the case of two quantiles at different probability levels such as the 1% and 5% VaR which are often considered (jointly) in financial applications. In the following section, we turn to the case of encompassing tests for the ES.

2.3. Joint Forecast Encompassing for the ES and the VaR

In this section, we turn to the first functional of interest, the ES which is defined as

$$ES_{t,\alpha}(Y_{t-1}) = \mathbb{E}_t\left[Y_{t+1} | Y_{t+1} \le Q_{t,\alpha}(Y_{t+1})\right], \tag{2.10}$$

where $Q_{t,\alpha}(Y_{t+1})$ denotes the conditional α-quantile of Y_{t+1} given $\mathcal{F}_t$, which is unique when the distribution F_t is absolutely continuous. As discussed above, the main ingredient of forecast encompassing testing is the specification of the underlying loss function, which has to be associated with the functional(s) we consider forecasts of. The crucial property the loss function has to fulfill is strict consistency. Weber (2006) and Gneiting (2011a) show that for the ES, there does not exist such a strictly consistent loss function. However, Fissler and Ziegel (2016) show that it is elicitable jointly with the quantile and they characterize the full class of strictly consistent loss functions for the pair consisting of the quantile and the ES,

$$\begin{aligned}\rho_e(Y, q_\alpha, e_\alpha) = &\left(\mathbb{1}_{\{Y \le q_\alpha\}} - \alpha\right) g(q_\alpha) - \mathbb{1}_{\{Y \le q_\alpha\}} g(Y) \\ &+ \phi'(e_\alpha)\left(e_\alpha - q_\alpha + \frac{(q_\alpha - Y)\mathbb{1}_{\{Y \le q_\alpha\}}}{\alpha}\right) - \phi(e_\alpha) + a(Y),\end{aligned} \tag{2.11}$$

where g and ϕ are smooth functions where g is non-negative and ϕ and ϕ' are strictly positive. This class of loss functions has the following interpretation. The first line is a generalized piecewise linear loss, which corresponds to the full class of strictly consistent loss functions for the quantile. The second line resembles the Bregman class of loss functions for the mean, where inside of the big bracket, instead of Y, there is a quantile-truncated version of Y. This form of the loss functions is not unexpected given that the ES is in fact a quantile-truncated expectation. Using this class of loss functions, we can now define the concept of joint forecast encompassing for the pair consisting of the quantile and the ES.

Definition 2.3 (Joint Quantile and ES Forecast Encompassing). Let $\left(\hat{q}_{1,t}, \hat{e}_{1,t}\right)$ and $\left(\hat{q}_{2,t}, \hat{e}_{2,t}\right)$ denote pair-wise competing forecasts for the pair consisting of the conditional quantile and ES of F_t. We say that $\left(\hat{q}_{1,t}, \hat{e}_{1,t}\right)$ encompasses $\left(\hat{q}_{2,t}, \hat{e}_{2,t}\right)$ at time t if and only if

$$\mathbb{E}_t\left[\rho_e\left(Y_{t+1}, \beta_{1,t}^*\hat{q}_{1,t} + \beta_{2,t}^*\hat{q}_{2,t}, \eta_{1,t}^*\hat{e}_{1,t} + \eta_{2,t}^*\hat{e}_{2,t}\right)\right] = \mathbb{E}_t\left[\rho_e\left(Y_{t+1}, \hat{q}_{1,t}, \hat{e}_{1,t}\right)\right], \tag{2.12}$$

where ρ_e is given in (2.11). This holds if and only if

$$\left(\beta_{1,t}^*, \beta_{2,t}^*, \eta_{1,t}^*, \eta_{2,t}^*\right) = (1,0,1,0). \tag{2.13}$$

We are interested in testing whether the pair $\left(\hat{q}_{1,t}, \hat{e}_{1,t}\right)$ conditionally encompasses $\left(\hat{q}_{2,t}, \hat{e}_{2,t}\right)$ over the whole out-of-sample period $t = m, \dots, T-1$. For this, we use the following notation $\theta = (\beta, \eta) = \left(\beta_1, \beta_2, \eta_1, \eta_2\right)$, $\hat{\boldsymbol{q}}_t = \left(\hat{q}_{1,t}, \hat{q}_{2,t}\right)$, $\hat{\boldsymbol{e}}_t = \left(\hat{e}_{1,t}, \hat{e}_{2,t}\right)$ and $\hat{\boldsymbol{f}}_t = \left(\hat{\boldsymbol{f}}_{1,t}, \hat{\boldsymbol{f}}_{2,t}\right) = \left(\hat{q}_{1,t}, \hat{e}_{1,t}, \hat{q}_{2,t}, \hat{e}_{2,t}\right)$.

In order to facilitate *conditional* encompassing, we rely on formulating our testing problem through GMM moment conditions. When an M-estimator is used for estimation of the combination weights, it is only feasible to condition on the explanatory variables used in the semiparametric model, in this case the issued forecasts. However, as shown in the following, formulating this as a GMM problem opens up the possibility to test encompassing conditional on any information set $\mathcal{G}_t \subseteq \mathcal{F}_t$.

In general, there exist infinitely many different identification functions (see Gneiting (2011a) and Fissler and Ziegel (2016)), which can be used for the formulation of conditional moment conditions in the GMM setting here. As Dimitriadis and Bayer (2017) find that the GMM-estimator is numerically unstable for many choices of these identification functions (obtained as the derivatives of a loss function), we rely on the following *conditional* moment conditions, which allows for an easy and numerically stable two-step estimation procedure.[1]

$$\varphi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) = \begin{pmatrix} \alpha - \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \\ \hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{1}{\alpha}(\hat{\boldsymbol{q}}_t^\top \beta - Y_t)\mathbb{1}_{\{Y_t \le \hat{\boldsymbol{q}}_t^\top \beta\}} \end{pmatrix}, \tag{2.14}$$

The following proposition shows that (2.14) is a conditional condition moment condition in the sense that its conditional expectation equals zero if and only if the parameters θ equal θ_t^*.

Proposition 2.4. Given that $\hat{\boldsymbol{q}}_t \neq 0$ and $\hat{\boldsymbol{e}}_t \neq 0$ almost surely for all t, we get that

$$\mathbb{E}_t \left[\varphi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) \right] = 0 \qquad \Longleftrightarrow \qquad \theta = \theta_t^*. \tag{2.15}$$

The proof is given in Appendix A.

The conditional encompassing condition in (2.15) is equivalent to

$$\left(\mathbb{E} \left[\varphi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) \otimes W_t \right] = 0 \quad \text{for all } \mathcal{F}_t\text{-measurable functions } W_t \right) \quad \Longleftrightarrow \quad \theta = \theta_t^*. \tag{2.16}$$

Testing whether something holds for all $\mathcal{F}_t$-measurable functions W_t is in general infeasible. In order to facilitate this, we choose a vector of instruments $\boldsymbol{W}_t = (\boldsymbol{W}_{q,t}, \boldsymbol{W}_{e,t}) : \Omega \to \mathbb{R}^l$, $l \in \mathbb{N}$, which captures *all the relevant information contained in* $\mathcal{F}_t$ and we define the unconditional moment conditions as

$$\psi_{QES}^{2step}(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) = \begin{pmatrix} \boldsymbol{W}_{q,t} \left(\alpha - \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \right) \\ \boldsymbol{W}_{e,t} \left(\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{1}{\alpha}(\hat{\boldsymbol{q}}_t^\top \beta - Y_{t+1})\mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \right) \end{pmatrix}. \tag{2.17}$$

For the estimation of the forecast combination weights, we choose to use the identification function ψ_{QES}^{2step} as this allows for a two-step estimation procedure in the sense that in the first step, we estimate the parameters of the semiparametric quantile model. In the second step, we use these pre-estimates in order to estimate the parameters of the ES model. This is desirable as Dimitriadis and Bayer (2017) find that the general one-step GMM-estimator is numerically unstable and requires vastly higher computation times. This is possible by the specifal form of ψ_{QES}^{2step}, as the first component is independent of the ES parameters and thus, has the following general form,

$$\psi_{QES}^{2step}(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) = \begin{pmatrix} \psi_{QES,1}^{2step}(Y_{t+1}, \boldsymbol{W}_{q,t}, \hat{\boldsymbol{q}}_t, \beta) \\ \psi_{QES,2}^{2step}(Y_{t+1}, \boldsymbol{W}_{e,t}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \beta, \eta) \end{pmatrix}, \tag{2.18}$$

where the first component only depends on the quantile-specific parameters. This estimator falls in the general class of two-step GMM estimators considered in Chapter 6 of Newey and McFadden (1994). In the following, we use the notation

$$\Psi_{QES,1}^{2step}(\beta) = \frac{1}{n} \sum_{t=m}^{T-1} \psi_{QES,1}^{2step}(Y_{t+1}, \boldsymbol{W}_{q,t}, \hat{\boldsymbol{q}}_t, \beta), \tag{2.19}$$

[1] A very similar two-step estimator is proposed by Barendse (2017) for the simplified case of $\boldsymbol{W}_{q,t} = \hat{\boldsymbol{q}}_t$ and $\boldsymbol{W}_{e,t} = \hat{\boldsymbol{e}}_t$ almost surely. Then, similar to the classical OLS estimator, we can find a closed form solution to the second moment condition in (2.14).

$$\Psi_{QES,2}^{2step}(\beta,\eta) = \frac{1}{n}\sum_{t=m}^{T-1}\psi_{QES,2}^{2step}(Y_{t+1}, \boldsymbol{W}_{e,t}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \beta, \eta), \qquad \text{and} \tag{2.20}$$

$$\Psi_{QES}^{2step}(\beta,\eta) = \left(\Psi_{QES,1}^{2step}(\beta)^\top, \Psi_{QES,2}^{2step}(\beta,\eta)^\top\right)^\top. \tag{2.21}$$

The first-step quantile estimator is defined by

$$\hat{\beta}_n := \arg\min_{\beta\in\Theta^\beta} \Psi_{QES,1}^{2step}(\beta)^\top \cdot \Psi_{QES,1}^{2step}(\beta), \tag{2.22}$$

Then, conditional on the pre-estimate $\hat{\beta}_n$, we define the estimator $\hat{\eta}_n$ as the minimizer of the inner product

$$\hat{\eta}_n := \arg\min_{\eta\in\Theta^\eta} \Psi_{QES,2}^{2step}(\eta,\hat{\beta}_n)^\top \cdot \Psi_{QES,2}^{2step}(\eta,\hat{\beta}_n), \tag{2.23}$$

The following proposition shows consistency of this two-step estimator and states the regularity conditions we need to impose for this.

Proposition 2.5. Assume that for every $t = m, \dots, T-1$, it holds that

(a) the process $(\boldsymbol{W}_t, Z_t)$ is α-mixing with α of size $-\tilde{r}/(\tilde{r}-1)$ for some $\tilde{r} > 1$,

(b) the distribution of Y_t given $\mathcal{F}_t$, denoted by F_t is absolutely continuous with continuous and strictly positive density f_t,

(c) $\hat{\boldsymbol{q}}_t$ and $\hat{\boldsymbol{e}}_t$ are nonzero almost surely and $\text{corr}(\hat{\boldsymbol{q}}_t) \neq 1$ and $\text{corr}(\hat{\boldsymbol{e}}_t) \neq 1$,

(d) $\mathbb{E}\left[\left\|\boldsymbol{W}_t\hat{\boldsymbol{q}}_t^\top\right\|^{\tilde{r}+\delta} + \left\|\boldsymbol{W}_t\hat{\boldsymbol{e}}_t^\top\right\|^{\tilde{r}+\delta} + \|\boldsymbol{W}_tY_t\|^{\tilde{r}+\delta}\right] < \infty$.

(e) $\theta^* \in \text{int}(\Theta)$, where the parameter space Θ is compact.

Then, it holds that $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*$.

The proof is given in Appendix A. The conditions of the proposition are similar to the standard conditions in the literature on asymptotic theory for semiparametric quantile and ES models, see e.g. Komunjer (2005), Giacomini and Komunjer (2005), Engle and Manganelli (2004) and Patton et al. (2017). The mixing condition (a) is a standard condition on the allowed dependence and heterogeneity of the underlying stochastic process such that a law of large numbers holds for the process $\Psi_{QES}^{2step}(\theta)$. Furthermore, existence of the moments in (d) is required for the law of large numbers. Condition (b) assures that the conditional quantile is unique and that there is positive mass at the conditional quantile in the form of a strictly positive density. This assumption is standard in the literature on conditional quantile estimation (Koenker and Bassett, 1978; Komunjer, 2005; Engle and Manganelli, 2004). Condition (c) rule out perfectly co-linear explanatory variables. Finally, assumption (e) is standard in the literature on time series parameter estimation inference. We now turn to asymptotic normality of the estimator, which is given in the following proposition.

Proposition 2.6. In addition to the conditions of Proposition 2.5, we assume that it holds that

(f) the process $(\boldsymbol{W}_t, Z_t)$ is α-mixing with α of size $-r/(r-2)$ for some $r > 2$,

(g) $\mathbb{E}\left[\left\|\boldsymbol{W}_t\hat{\boldsymbol{q}}_t^\top\right\|^{r+\delta} + \left\|\boldsymbol{W}_t\hat{\boldsymbol{e}}_t^\top\right\|^{r+\delta} + \|\boldsymbol{W}_tY_t\|^{r+\delta}\right] < \infty$,

(h) the matrix Σ_n, defined in (2.26) is positive definite with a determinant bounded away from zero for all n sufficiently large,

(i) the density f_t is bounded from above almost surely on the whole support of F_t,

(j) the matrices $\mathbb{E}\left[\boldsymbol{W}_{q,t}\hat{\boldsymbol{q}}_t^\top\right]$ and $\mathbb{E}\left[\boldsymbol{W}_{e,t}\hat{\boldsymbol{e}}_t^\top\right]$ have full row rank for all $t = 1, \dots, T$,

(k) for any n, the term $\sup_{\beta\in\Theta^\beta} \sum_{t=m}^{T-1} \mathbb{1}_{\{Y_{t+1}=\hat{\boldsymbol{q}}_t^\top\beta\}}$ is almost surely bounded from above,

(l) the sequence $\psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta\right)$ is stochastically equicontinuous.

Then, it holds that

$$\Omega_n^{-1/2}\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d} \mathcal{N}(0, I), \tag{2.24}$$

where

$$\Omega_n = \left(\Lambda_n^\top\Lambda_n\right)^{-1}\left(\Lambda_n^\top\Sigma_n\Lambda_n\right)\left(\Lambda_n^\top\Lambda_n\right)^{-1} \tag{2.25}$$

and

$$\Sigma_n = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta^*\right)\cdot\psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta^*\right)^\top\right]. \tag{2.26}$$

and

$$\Lambda_n = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\begin{pmatrix}\boldsymbol{W}_{q,t}\hat{\boldsymbol{q}}_t^\top f_t(\hat{\boldsymbol{q}}_t^\top\beta^*) & 0\\ 0 & \boldsymbol{W}_{e,t}\hat{\boldsymbol{e}}_t^\top\end{pmatrix}\right]. \tag{2.27}$$

The proof is given in Appendix A. The conditions imposed for Proposition 2.6 again resemble conditions usually imposed for asymptotic normality for quantile and ES models, see e.g. Komunjer (2005), Giacomini and Komunjer (2005), Engle and Manganelli (2004) and Patton et al. (2017). The strengthened condition for α-mixing in (f) is required such that a CLT holds for $\Psi_{QES}^{2step}(\theta)$. For this, we furthermore need condition (h) and the moment condition (g), which extends the moment condition (d), required for consistency. Equivalently, condition (i) strengthens condition (b) which requires that we have a sufficiently smooth conditional distribution, such that we can smoothen the indicator function in the moment conditions. In order to guarantee that the matrix $\Lambda_n^\top\Lambda_n$ has full rank and can consequently be inverted, we have to impose condition (j). Condition (k) limits the exact number of *hits* of Y_{t+1} in the conditional quantile model. For linear models and in the scenario of iid data, this condition is readily verified as in Ruppert and Carroll (1980). Eventually, condition (l) requires that the moment conditions are stochastically equicontinuous, where primitive conditions for this can be found e.g. in Engle and Manganelli (2004), Komunjer (2005) Giacomini and Komunjer (2005) and Patton et al. (2017).

We estimate the asymptotic covariance by the following estimators

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{t=m}^{T-1}\psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \hat{\theta}_n\right)\cdot\psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \hat{\theta}_n\right)^\top \quad \text{and} \tag{2.28}$$

$$\hat{\Lambda}_n = \frac{1}{n}\sum_{t=m}^{T-1}\begin{pmatrix}\boldsymbol{W}_{q,t}\hat{\boldsymbol{q}}_t^\top\frac{1}{2c_n}\mathbb{1}_{\{|Y_{t+1}-\hat{\boldsymbol{q}}_t^\top\hat{\beta}_n|<c_n\}} & 0\\ 0 & \boldsymbol{W}_{e,t}\hat{\boldsymbol{e}}_t^\top\end{pmatrix}, \tag{2.29}$$

and, $\hat{\Omega}_n = \left(\hat{\Lambda}_n^\top\hat{\Lambda}_n\right)^{-1}\left(\hat{\Lambda}_n^\top\hat{\Sigma}_n\hat{\Lambda}_n\right)\left(\hat{\Lambda}_n^\top\hat{\Lambda}_n\right)^{-1}$. The most problematic term to estimate is the conditional density which, following Engle and Manganelli (2004) and Patton et al. (2017), is estimated consistently by the estimator $\frac{1}{2c_n}\mathbb{1}_{\{|Y_{t+1}-\hat{\boldsymbol{q}}_t^\top\hat{\beta}_n|<c_n\}}$. We get consistency of the covariance estimators by the following proposition.

Proposition 2.7. There is a deterministic and positive sequence c_n that satisfies $c_n = o(1)$ and $c_n^{-1} = o(\sqrt{n})$. Furthermore, it holds that

(m) $\frac{1}{n}\sum_{t=m}^{T-1}\psi_{QES}^{2step}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\theta^*)\psi_{QES}^{2step}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\theta^*)^\top - \Sigma_n \xrightarrow{\mathbb{P}} 0$

(n) $\frac{1}{n}\sum_{t=m}^{T-1}\boldsymbol{W}_{e,t}\hat{\boldsymbol{e}}_t^\top - \mathbb{E}\left[\boldsymbol{W}_{e,t}\hat{\boldsymbol{e}}_t^\top\right] \xrightarrow{\mathbb{P}} 0$

(o) $\frac{1}{n}\sum_{t=m}^{T-1}\boldsymbol{W}_{q,t}\hat{\boldsymbol{q}}_t^\top f_t(\hat{\boldsymbol{q}}_t^\top\beta^*) - \mathbb{E}\left[\boldsymbol{W}_{q,t}\hat{\boldsymbol{q}}_t^\top f_t(\hat{\boldsymbol{q}}_t^\top\beta^*)\right] \xrightarrow{\mathbb{P}} 0$

Then, it holds that $\hat{\Sigma}_n - \Sigma_n \xrightarrow{\mathbb{P}} 0$, $\hat{\Lambda}_n - \Lambda_n \xrightarrow{\mathbb{P}} 0$ and $\hat{\Omega}_n - \Omega_n \xrightarrow{\mathbb{P}} 0$.

The proof is given in Appendix A.

Using the three previously established propositions, we can now derive the asymptotic distribution of the test statistic.

Theorem 2.8 (Joint Encompassing Test). Under $\mathbb{H}_0$, it holds that

$$Z_n^{Joint-Enc} = n\big(\hat{\theta}_n - (1,0,1,0)\big)\hat{\Omega}_{n,QES}^{-1}\big(\hat{\theta}_n - (1,0,1,0)\big)^\top \xrightarrow{d} \chi_4^2. \tag{2.30}$$

The proof is given in Appendix A. Thus, we establish a joint encompassing test for the VaR and the ES and derive the asymptotic distribution of the test statistic.

2.4. Forecast Encompassing Tests for the ES Stand-Alone

The primary objective of this paper is to setup encompassing tests for high-order elicitable functionals as e.g. the ES. So far, we manage to develop a joint test for the VaR and ES, which is reasonable given the joint elicitability property of the VaR and ES. However, in this section we also develop encompassing tests for the ES stand-alone. This is one feature where encompassing tests are superior to the literature on testing for superior predictive ability in the sense of Diebold and Mariano (1995), Giacomini and White (2006) and West (2006). As these tests are based directly on the average loss differential, they can only test for joint predictive ability of the VaR and ES.

In contrast, encompassing tests are based on the regression coefficients of the semiparametric models, which are of course estimated jointly but still are separate coefficients. Thus, the joint encompassing test defined in Definition 2.3 depends on the associated loss functions only *indirectly*. This fundamental difference allows for only testing the hypothesis that the ES-specific regression parameters equal zero and one. In this approach, we do not test the quantile specific parameters and consequently, under the null hypothesis, we do not impose that the underlying quantile forecasts also have to encompass its competitor. We call this test *Auxiliary ES Encompassing Test* as it still depends on the auxiliary VaR forecasts which are used for the estimation. The test is formally defined in the following.

Definition 2.9 (Auxiliary ES Forecast Encompassing). Let $(\hat{q}_{1,t},\hat{e}_{1,t})$ and $(\hat{q}_{2,t},\hat{e}_{2,t})$ denote competing forecasts for the pair consisting of the conditional quantile and ES of F_t. We say that $\hat{e}_{1,t}$ encompasses $\hat{e}_{2,t}$ at time t if and only if

$$(\eta_{1,t}^*,\eta_{2,t}^*) = (1,0). \tag{2.31}$$

As we utilize the same estimation technique as for the joint encompassing test, the asymptotic test distribution can be derived by using the same asymptotic theory as derived in Theorem 2.5 and Theorem 2.6, which we formulate in the following.

Theorem 2.10 (Auxiliary ES Encompassing Test). Under $\mathbb{H}_0$, it holds that

$$Z_n^{ES-Enc} = n\big(\hat{\eta}_n - (1,0)\big)\hat{\Omega}_{n,ES}^{-1}\big(\hat{\eta}_n - (1,0)\big)^\top \xrightarrow{d} \chi_2^2. \tag{2.32}$$

The proof is straight-forward and follows the proof of Theorem 2.8.

Even though the emphasis of this test is on the ES, we still need quantile forecasts for the implementation of the parameter estimation. This is problematic for two reasons. First, the quantile forecasts are still utilized in the estimation procedure and thus have an indirect effect on the parameter estimates of the ES specific parameters. Second, the test is only applicable in the setup where the person who is interested in applying the tests actually have the pair of competing forecasts at hand. In the current implementation of the regulatory framework of the Basel Committee (Basel Committee, 2016, 2017), the banks are only obligated to report their ES forecasts at probability level 2.5%. Thus, the accompanying VaR forecast, which the ES forecast is based on internally, is in general not available to the regulator who has to decide on an adequate risk management of the financial institution at hand.

This motivates a third design of the encompassing test, which works without the accompanying VaR forecasts. The idea is that in location-scale models where $Y_{t+1} = \sigma_{t+1} u_t$, $u_t \sim F(0,1)$, which are still the most frequently used class of models for risk management, the VaR and ES are linear (or affine) transformations of themselves,

$$\hat{e}_t = \frac{\xi_\alpha}{z_\alpha} \hat{q}_t, \tag{2.33}$$

where z_α and ξ_α are the α-quantile and ES of the distribution $F(0,1)$. Thus, we estimate the following joint model

$$Y_{t+1} = \beta_{1,t} \hat{e}_{1,t} + \beta_{2,t} \hat{e}_{2,t} + u_t^q \tag{2.34}$$

$$Y_{t+1} = \eta_{1,t} \hat{e}_{1,t} + \eta_{2,t} \hat{e}_{2,t} + u_t^e. \tag{2.35}$$

Given that the underlying data stems from a location-scale model and under $\mathbb{H}_0$: the forecasts $\hat{e}_{1,t}$ encompass the forecasts $\hat{e}_{2,t}$, we get that $\left(\beta_{1,t}^*, \beta_{2,t}^*, \eta_{1,t}^*, \eta_{2,t}^*\right) = \left(z_\alpha/\xi_\alpha, 0, 1, 0\right)$. As we are generally agnostic about encompassing of VaR forecasts and as the ratio z_α/ξ_α is generally unknown, we only test for the ES specific parameters and set up the following definition of the strict ES encompassing test.

Definition 2.11 (Strict ES Forecast Encompassing). Let $\hat{e}_{1,t}$ and $\hat{e}_{2,t}$ denote competing ES forecasts of the underlying predictive distribution F_t. We say that $\hat{e}_{1,t}$ encompasses $\hat{e}_{2,t}$ at time t if and only if

$$\left(\eta_{1,t}^*, \eta_{2,t}^*\right) = \left(1, 0\right). \tag{2.36}$$

Thus, we test whether $\left(\eta_{1,t}^*, \eta_{2,t}^*\right) = \left(1, 0\right)$ for all $t = m, \dots, T-1$, i.e. throughout the entire out-of-sample period and we can establish the distribution of the test statistic in the following theorem.

Theorem 2.12 (Strict ES Forecast Encompassing). Assume that the conditions from Proposition 2.5, 2.6 and 2.7 hold. Then, under $\mathbb{H}_0$ and given that our data is from a location-scale model, it holds that

$$Z_n^{ES-Enc} = n\left(\hat{\eta}_n - (1,0)\right)\hat{\Omega}_{n,ES}^{-1}\left(\hat{\eta}_n - (1,0)\right)^\top \xrightarrow{d} \chi_2^2. \tag{2.37}$$

The proof is straight-forward and follows the proof of Theorem 2.8. Derivation of the asymptotic distribution in the case where the underlying forecasts do not stem from an exact location-scale model is still an open problem. However, parameter estimates of financial data of non-location-scale models implies that the deviations from location-scale are not huge, and consequently, this test can be seen as a good approximation. Furthermore, the simulation study in Section 3 shows that this test is still well-sized and has good power properties. Thus, we propose to use this test in the scenario where the user of this test does not have the accompanying VaR forecasts at hand.

One important application of these ES encompassing tests is in the context of selecting the best-performing forecast, i.e. selecting at time $\tilde{t}$ a best forecasting method for time $\tilde{t}+1$. This is particularly relevant as the ES is recently introduced into the Basel regulations without having proper forecast selection procedures at hand. Consequently, following Giacomini and Komunjer (2005), we propose the following decision rule for ES forecasts. Perform the two encompassing tests of $\mathbb{H}_{0,1}$: $\hat{e}_{1,t}$ encompasses $\hat{e}_{2,t}$ and $\mathbb{H}_{0,2}$: $\hat{e}_{2,t}$ encompasses $\hat{e}_{1,t}$ on data up to time $\tilde{t}$. Then, there are four possible scenarios: (1) If neither $\mathbb{H}_{0,1}$ nor $\mathbb{H}_{0,2}$ are rejected, then the test is not helpful for forecast selection. (2) if $\mathbb{H}_{0,1}$ is rejected while $\mathbb{H}_{0,2}$ is not rejected, then we can conclude that forecast $\hat{e}_{2,t}$ did not add any information to forecast $\hat{e}_{1,t}$ and consequently, we decide to use the forecasting method of $\hat{e}_{1,t}$. (3) if $\mathbb{H}_{0,2}$ is rejected while $\mathbb{H}_{0,1}$ is not rejected, then the same logic applies and we would use the forecasting method of $\hat{e}_{2,t}$. (4) eventually, if both $\mathbb{H}_{0,1}$ and $\mathbb{H}_{0,2}$ are rejected, then the test delivers statistical evidence that both forecasts contain information and that a forecast combination performs significantly better. Consequently, one would choose a combined forecast $\hat{e}_{c,t} = \hat{\eta}_{n,1}\hat{e}_{1,t} + \hat{\eta}_{n,2}\hat{e}_{2,t}$ where the combination weights are estimated by the GMM estimator proposed in this paper.

3. Simulation Study

In this section, we evaluate the empirical performance of our three proposed ES encompassing tests in two different scenarios. The first, presented in Section 3.1 analyses the test properties for two competing forecasts, both stemming from locations scale GARCH-type models. In Section 3.2, we present a simulation setup based on CARE models of Taylor (2017), which are outside of the class of location-scale models. This simulation setup also enables us to evaluate the performance of the third strict ES encompassing test in a situation of misspecification. We report and discuss the results of the simulation study in Section 3.3.

3.1. GARCH setting

In this section, we consider two model specifications from the location-scale GARCH family with zero mean. The first is a standard GARCH(1,1) model, calibrated to IBM data, which leads to the following model specification, $Y_{t+1} = \sigma_{1,t+1}u_{t+1}$, where $u_{t+1} \sim \mathcal{N}(0,1)$ and

$$\sigma_{1,t}^2 = 0.042 + 0.053Y_t^2 + 0.925\sigma_{1,t}^2. \tag{3.1}$$

We obtain forecasts for the VaR and ES by the following formulas

$$\hat{q}_{1,t+1} = z_\alpha \sigma_{1,t+1} \qquad \text{and} \qquad \hat{e}_{1,t+1} = \xi_\alpha \sigma_{1,t+1}, \tag{3.2}$$

where z_α and ξ_α are the α-quantile and α-ES of the standard normal distribution.

The second model is a GJR-GARCH(1,1) model with Gaussian returns, which we calibrate to the same data, and which leads to the following model specification, $Y_{t+1} = \sigma_{2,t+1}u_{t+1}$, where $u_{t+1} \sim \mathcal{N}(0,1)$ and

$$\sigma_{2,t+1}^2 = 0.044 + \big(0.024 + 0.058 \cdot \mathbb{1}_{\{Y_t \le 0\}}\big)Y_t^2 + 0.923. \tag{3.3}$$

This model specification allows for a leverage term by including the indicator function into the squared return part of the GARCH equation. This corrects for the stylized fact that negative returns lead to higher future volatilities than positive returns of the same magnitude. Forecasts for the VaR and ES of the GJR-GARCH model are again given by

$$\hat{q}_{2,t+1} = z_\alpha \sigma_{2,t+1} \qquad \text{and} \qquad \hat{e}_{2,t+1} = \xi_\alpha \sigma_{2,t+1}. \tag{3.4}$$

Using these two relatively realistic volatility models for daily return data, we simulate data from the following process

$$Y_t = \big((1-\rho)\sigma_{1,t} + \rho\sigma_{2,t}\big)u_t, \tag{3.5}$$

where $u_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and where $\rho \in [0,1]$ is chosen on a linear grid between zero and one. For $\rho = 0$, the simulation setup described in (3.5) implies that the simulated data follows the GARCH model specified in

(3.1) and consequently, the true VaR and ES forecasts are given by (3.2). Consequently, the GJR-GARCH forecasts in (3.4) are misspecified and cannot add any valuable information to the forecast combination. In other words, the GARCH forecasts are already correctly specified and generate the lowest possible loss (in expectation or alternatively in the limit by applying a weak law of large numbers). Consequently, the setup $\rho = 0$ corresponds to the null hyptothesis.

$$\mathbb{H}_0^{(1)}: \text{ the forecasts} \big(\hat{q}_{1,t}, \hat{e}_{1,t}\big) \text{ encompass the forecasts } \big(\hat{q}_{2,t}, \hat{e}_{2,t}\big). \tag{3.6}$$

Conversely, the choice $\rho = 1$ equals to the inverted hypothesis

$$\mathbb{H}_0^{(2)}: \text{ the forecasts} \big(\hat{q}_{2,t}, \hat{e}_{2,t}\big) \text{ encompass the forecasts } \big(\hat{q}_{1,t}, \hat{e}_{1,t}\big). \tag{3.7}$$

Furthermore, all choices of $\rho \in (0, 1)$ correspond to the alternatives and the size of ρ determines the degree of misspecification for each hypothesis.

3.2. CARE setting

We also use a different simulation setup in order to generate data which does not follow a location-scale process. This is particularly important as the underlying theory of the strict ES test imposes data stemming from location-scale models. Thus, we conduct this simulation experiment in order to demonstrate robustness of this test against typical deviations from location-scale models. For this, we adapt Conditional Autoregressive Expected Shortfall (CARE) models introduced by Taylor (2017) in order to generalize the classical CAViaR quantile model of Engle and Manganelli (2004). The asymmetric slope (AS) CARE model, calibrated to the same daily IBM data as used in Section 3.1 is given by

$$\hat{q}_{1,t+1} = -0.0003 - 0.05|Y_t|\mathbb{1}_{\{Y_t \geq 0\}} - 0.15|Y_t|\mathbb{1}_{\{Y_t < 0\}} + 0.8\hat{q}_{1,t}, \qquad \text{and} \tag{3.8}$$

$$\hat{e}_{1,t+1} = \hat{q}_{1,t+1} - x_{t+1}, \qquad \text{where} \tag{3.9}$$

$$x_{t+1} = \begin{cases} 0.00017 + 0.125(\hat{q}_{1,t} - Y_t) + 0.84\hat{q}_{1,t} & \text{if } \hat{q}_{1,t} \leq Y_t \\ x_t & \text{if } \hat{q}_{1,t} > Y_t. \end{cases} \tag{3.10}$$

Using this model, we can directly generate quantile and ES forecasts. The second model variant we consider is the symmetric absolute value (SAV) CARE model of Taylor (2017), given by

$$\hat{q}_{2,t+1} = -0.0003 - 0.1|Y_t| + 0.8\hat{q}_{2,t} \qquad \text{and} \tag{3.11}$$

$$\hat{e}_{2,t+1} = \hat{q}_{2,t+1} - x_{t+1} \qquad \text{where} \tag{3.12}$$

$$x_{t+1} = \begin{cases} 0.00017 + 0.125(\hat{q}_{2,t} - Y_t) + 0.84\hat{q}_{2,t} & \text{if } \hat{q}_{2,t} \leq Y_t \\ x_t & \text{if } \hat{q}_{2,t} > Y_t. \end{cases} \tag{3.13}$$

In this simulation setup, we simulate data according to the following additive model

$$Y_t = \big((1-\rho)\hat{e}_{1,t} + \rho\hat{e}_{2,t}\big) + \varepsilon_t, \qquad \text{where} \qquad \varepsilon_t \sim \mathcal{N}\big(\sigma\xi_\alpha, \sigma^2\big). \tag{3.14}$$

The underlying reason for this additive simulation design is that for $\rho = 0$, it holds that

$$ES_\alpha(Y_{t+1}|\mathcal{F}_t) = \hat{e}_{1,t} + ES_\alpha(\varepsilon_{t+1}|\mathcal{F}_t)) = \hat{e}_{1,t} \tag{3.15}$$

$$Q_\alpha(Y_{t+1}|\mathcal{F}_t) = \hat{e}_{1,t} + Q_\alpha(\varepsilon_{t+1}|\mathcal{F}_t)) = \hat{e}_{1,t} + \sigma(\xi_\alpha\alpha - z_\alpha) \neq \hat{q}_{1,t}. \tag{3.16}$$

The same holds inversely for $\rho = 1$. This implies that we generate data for which the ES forecasts $\hat{e}_{1,t}$ equal the true conditional ES, whereas this does not hold for the respective quantile forecasts $\hat{q}_{1,t}$. Consequently, the ES forecasts should be able to encompass any other misspecified forecasting scheme without the associated quantile forecasts. This establishes a simulation setup which justifies the second and third variant of our encompassing tests which test encompassing of ES forecasts stand-alone.

3.3. Simulation Results

Section 3.3 reports the empirical sizes of the three different ES encompassing tests introduced in Section 2 at a 5% nominal significance level based on 2000 Monte Carlo replications. The respective null hypothesis being tested is that the model given in the line $\mathbb{H}_0$ encompasses the respective competitor. The upper panel depicts results for the GARCH simulation design of Section 3.1 whereas the lower panel depicts results for the CARE simulation design of Section 3.2. We show results for a test versions based on estimation of the asymptotic covariance for the sample sizes $n = 1000$ and $n = 2500$ and a bootstrap test versions for $n = 1000$.

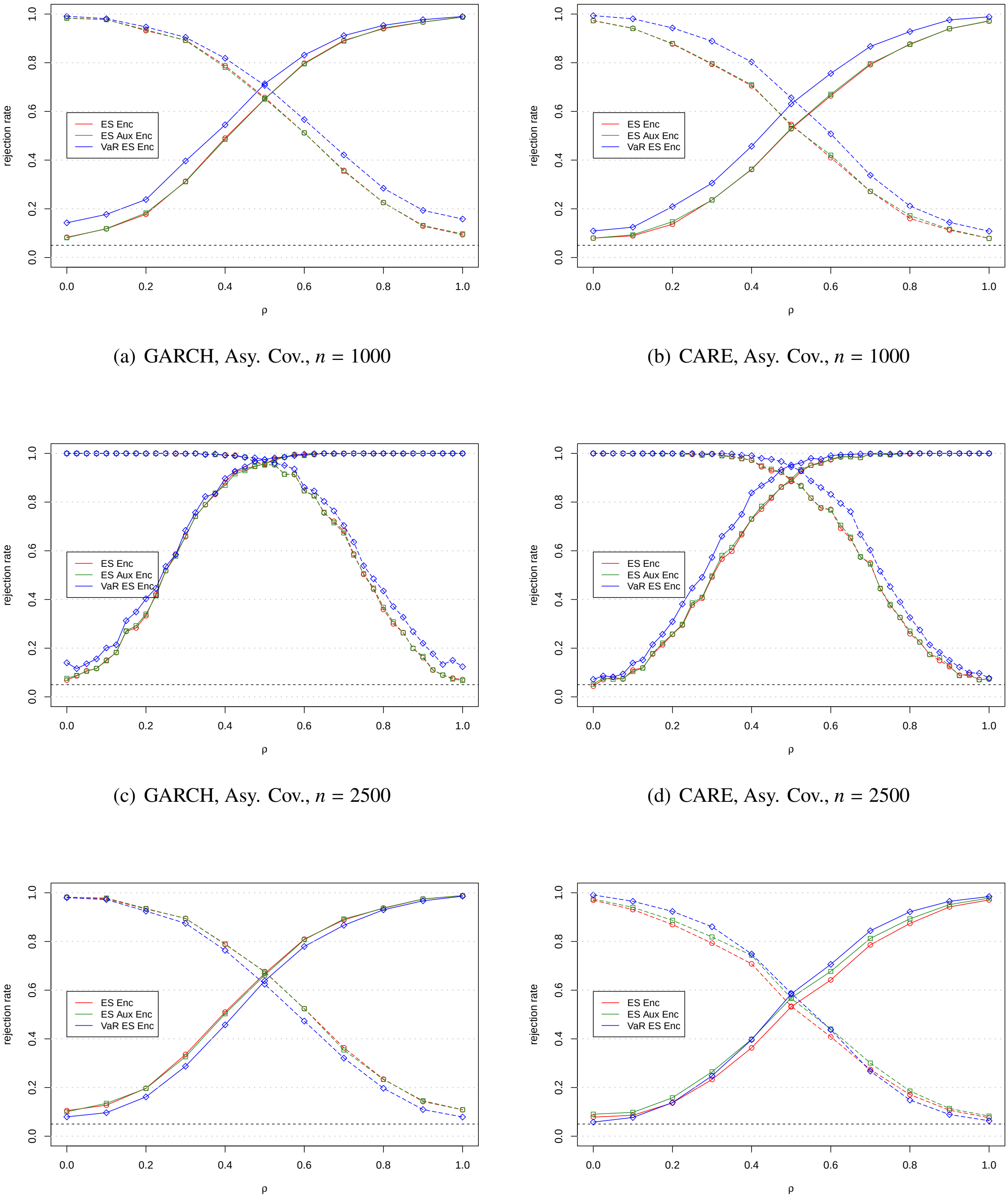
We find that our tests are well-sized, especially for the increased sample size $n = 2500$ and for the bootstrap estimator. The second and third test which only test the ES forecasts exhibit overall better size properties than the joint encompassing test. This observation can be explained as the joint test is based on an estimate of a four by four covariance which includes the density quantile function as a known nuisance quantity in the covariance estimation (Koenker and Bassett, 1978; Giacomini and Komunjer, 2005; Dimitriadis and Bayer, 2017). As a consequence, this test benefits strongly from applying a bootstrap procedure which circumvents this source of inaccuracy. In contrast, the ES versions of the test do not rely on covariance submatrices containing this quantity and consequently perform better. Furthermore, we find that the ES encompassing test, which is subject to misspecification in the quantile equation for the CARE simulation setup exhibits good size properties for both simulation setups. As we simulate with calibrated data, this shows that this test is robust against misspecifications usually encountered in financial data and delivers robust results.

Table 1: This table reports the empirical sizes [%] of the three different ES encompassing tests introduced in Section 2 at a 5% nominal significance level. The upper panel depicts results for the GARCH simulation design of Section 3.1 whereas the lower panel depicts results for the CARE simulation design of Section 3.2. The null hypothesis being tested is that the model given in the line $\mathbb{H}_0$ encompasses the respective competitor. We show results for a test versions based on estimation of the asymptotic covariance for the sample sizes $n = 1000$ and $n = 2500$ and a bootstrap test versions for $n = 1000$.

	Asy. Cov. $n = 1000$		Asy. Cov. $n = 2500$		Bootstrap $n = 1000$	
$\mathbb{H}_0$:	GARCH	GJR-GARCH	GARCH	GJR-GARCH	GARCH	GJR-GARCH
VaRES Encmp	14.25	15.80	14.00	12.40	7.95	7.90
Aux ES Encmp	8.10	9.70	7.60	6.70	10.05	10.90
ES Encmp	8.30	9.35	6.90	7.10	10.50	10.85
	Asy. Cov. $n = 1000$		Asy. Cov. $n = 2500$		Bootstrap $n = 1000$	
$\mathbb{H}_0$:	AS-CARE	SAV-CARE	AS-CARE	SAV-CARE	AS-CARE	SAV-CARE
VaRES Encmp	10.90	10.80	7.20	7.70	5.80	6.35
Aux ES Encmp	7.95	7.85	5.20	7.50	9.05	8.25
ES Encmp	7.95	7.85	4.40	7.30	7.85	7.60

Figure 1 displays power curves (empirical rejection rates) for the both simulation setups, both covariance estimator and the two different sample sizes for the tests at a nominal significance level of 5% and for 2000 Monte Carlo replications. Solid lines show rejection rates for the test with the null hypothesis that the GARCH (AS-CARE) model encompasses the GJR-GARCH (SAV-CARE) model. The dashed lines depict the rejection rates of the opposing null hypothesis of encompassing. We analyze the power of our tests using the DGPs specified in (3.5) and (3.14) for different values of ρ on a regularly spaced grid of values between zero and one. This plot includes the respective empirical sizes of Section 3.3 for $\rho = 0$ and $\rho = 1$. Increasing (decreasing) values of ρ feature continuously increasing degrees of misspecification for the respective tests.

We find that overall, all tests exhibit strong power properties against the alternatives, even for mediocre values of ρ and as expected, the test power increases by increasing the sample size from $n = 1000$ to $n = 2500$. As in the size analysis, we notice that the joint VaR and ES test benefits the most from a bootstrapped covariance as the joint test rejects too often in the version based on asymptotic covariances but not as often in the bootstrapped test versions. However, the power of all three tests does not change dramatically by applying the bootstrap which implies that the testing procedures benefit from the bootstrap especially in terms of being well-sized. Furthermore, the tests seems to perform equally well for both opposing hypotheses, which implies that the test size and power are not influenced by considering misspecifications which result either from over- or underspecification of the true model. Eventually, the results imply that all three ES encompassing tests exhibit very similar power patterns, which implies that it is relatively unimportant which test we apply in practice. This results especially emphasizes the applicability of the third strict ES test in situations where only the ES is available, even under more general specifications than location-scale models.



(a) GARCH, Asy. Cov., $n = 1000$

(b) CARE, Asy. Cov., $n = 1000$

(c) GARCH, Asy. Cov., $n = 2500$

(d) CARE, Asy. Cov., $n = 2500$

(e) GARCH, Bootstrap, $n = 1000$

(f) CARE, Bootstrap, $n = 1000$

Figure 1: This figure shows the empirical rejection rates of the three different ES encompassing tests introduced in Section 2 at a 5% nominal significance level. The first column of plots shows results for the GARCH simulation setup described in Section 3.1 whereas the second column shows plots for the CARE simulation setup described in Section 3.2. The tests used in the first line of plots is based on the estimator for the asymptotic covariance for $n = 1000$ whereas the second line shows results for $n = 2500$. The third line shows plots for a bootstrapped based test for $n = 1000$. In each of the plots, the case $\rho = 0$ corresponds to the null hypothesis that the GARCH (AS-CARE) model encompasses its competitor and the case $\rho = 1$ corresponds to the null hypothesis that the GJR-GARCH (SAV-CARE) model encompasses its competitor. The cases $\rho \in (0, 1)$ represent misspecified cases with continuously increasing degree of misspecification.

4. Empirical Application

In this section, we illustrate the application of our encompassing tests for the ES (jointly with the VaR). The Basel Committee of Banking supervision recently decided to shift from the VaR to the ES as standard market risk measure for the international banking regulation (Basel Committee, 2016, 2017). Thus, it becomes obvious that tests on superior forecast performance for the ES are of uttermost importance.

We use daily open-close returns from the IBM stock from January 1st, 2001 until October, 1st, 2018, which amounts to a total of $T = 4417$ daily observations. We use the first $m = 1000$ observations to estimate models in order to generate the forecasts, while the remaining $n = 3417$ observations are used for the forecast evaluation in an out-of-sample fashion. In the following, we use a fixed forecasting scheme, i.e. (for the parametric models) the model parameters are estimated once on the first $m = 1000$ in-sample observation. These parameter estimates are then used for generating the forecasts of the VaR and ES for the remaining days in a rolling-window fashion. In the following, we focus on one-day ahead forecasts, whereas multi-day ahead forecasting can easily be applied in the same fashion. As the Basel Committee specified that the ES has to reported for the probability level $\alpha = 2.5\%$, we use this probability level for the VaR and ES throughout the empirical application.

In the analysis, we consider the following competing forecasting models. First, we use the GJR-GARCH(1,1)-n model of Glosten et al. (1993) based on Gaussian returns, which is estimated by maximum likelihood. This model is based on the assumption of a location-scale framework, i.e. $Y_{t+1} = \hat{\sigma}_t u_{t+1}$, where $u_{t+1} \sim \mathcal{N}(0,1)$ and

$$\hat{\sigma}_t^2 = 0.044 + \big(0.024 + 0.058 \cdot \mathbb{1}_{\{Y_t \le 0\}}\big) Y_t^2 + 0.923 \hat{\sigma}_{t-1}^2. \tag{4.1}$$

We use this model in order to generate volatility forecasts $\hat{\sigma}_t$ and then obtain forecasts for the VaR and ES from the model by using the formulas $\hat{q}_t = z_\alpha \hat{\sigma}_t$ and $\hat{e}_t = \xi_\alpha \hat{\sigma}_t$, where z_α and ξ_α are the α-quantile and α-ES of the standard normal distribution. The second model is the RiskMetrics model, which is also in the location-scale family but models the conditional volatility as a fixed function (without estimated parameters) and without leverage effects,

$$\hat{\sigma}_t^2 = 0.94 \hat{\sigma}_{t-1}^2 + 0.06 Y_t^2. \tag{4.2}$$

The forecasts for the VaR and ES are obtained by applying the same location-scale formulas as for the GJR-GARCH model. Third, we employ the Historical Simulation model which computes VaR forecasts by computing the empirical α-quantile of the past 250 trading days. Equivalently, ES forecasts are generated by computing the empirical α-ES of the past 250 trading days. The forth and fifth model are GAS models (Creal et al., 2013; Harvey, 2013) for the VaR and ES proposed by Patton et al. (2017), which are estimated by minimizing the strictly consistent loss function for the VaR and ES given in (2.11). The 2-factor GAS model together with the estimated parameters is given by

$$\begin{pmatrix} \hat{q}_t \\ \hat{e}_t \end{pmatrix} = \begin{pmatrix} -0.031 \\ -0.072 \end{pmatrix} + \begin{pmatrix} 0.987 \hat{q}_{t-1} \\ 0.981 \hat{e}_{t-1} \end{pmatrix} + \begin{pmatrix} -0.2414 & -0.1593 \\ -0.0002 & 0.0032 \end{pmatrix} \begin{pmatrix} \hat{\lambda}_{q,t} \\ \hat{\lambda}_{e,t} \end{pmatrix}, \tag{4.3}$$

where the *forcing variables* of the system are the identification functions for the quantile and ES,

$$\hat{\lambda}_{q,t} = \hat{q}_{t-1} \big(\alpha - \mathbb{1}_{\{Y_t \le \hat{q}_{t-1}\}}\big) \qquad \text{and} \qquad \hat{\lambda}_{e,t} = \frac{1}{\alpha} \mathbb{1}_{\{Y_t \le \hat{q}_{t-1}\}} Y_t - \hat{e}_{t-1}. \tag{4.4}$$

Patton et al. (2017) also introduce a one factor GAS model for the VaR and ES. For this model, the VaR and ES are driven by one forcing variable, which is similar to the classical GARCH framework. However, in order to focus on the VaR and ES instead of the volatility, this model is again estimated by minimizing the strictly consistent loss function for pair VaR and ES given in (2.11). The model together with the estimated parameters is given by

$$\hat{q}_t = -2.4717 \exp(\hat{\kappa}_t) \tag{4.5}$$

$$\hat{e}_t = -3.9250 \exp(\hat{\kappa}_t) \tag{4.6}$$

$$\hat{\kappa}_t = 0.9904\hat{\kappa}_{t-1} + \frac{0.005}{\hat{e}_{t-1}} \left(\frac{1}{\alpha} \mathbb{1}_{\{Y_t \le \hat{q}_t\}} Y_t - \hat{e}_t \right). \tag{4.7}$$

The parameter values of the GJR-GARCH and the two GAS models are estimated. In contrast, the RiskMetrics and Historical Simulation models do not require parameter estimation and can directly be applied to generate the forecasts.

Table 2 shows the correlations of the respective forecasts. We can see that no pair of forecasts is perfectly correlated, which is crucial for the applicability of the encompassing tests as statetd in Theorem 2.5. We run pair-wise encompassing tests comparing all five described forecasting methods.

Table 2: This table shows the correlation of the respective quantile and ES forecasts obtained from the five forecasting models described in Section 4.

	Correlations VaR Forecasts					Correlations ES Forecasts				
	GJR-G	RM	HS	G-1F	G-2F	GJR-G	RM	HS	G-1F	G-2F
GJR-GARCH	1	0.93	0.66	0.95	0.90	1	0.93	0.69	0.95	0.87
RiskMetrics		1	0.57	0.91	0.91		1	0.57	0.91	0.90
Historical Simulation			1	0.60	0.62			1	0.64	0.49
VaR/ES GAS-1Factor				1	0.96				1	0.95
VaR/ES GAS-2Factor					1					1

I.e. for each pair, we run encompassing test of the hypothesis that forecast A encompasses forecast B and vice versa that forecast B encompasses forecast A. For this, we choose the vector of instruments $\boldsymbol{W}_{q,t} = \left(1, \hat{q}_{1,t}, \hat{q}_{2,t}\right)$ and $\boldsymbol{W}_{e,t} = \left(1, \hat{e}_{1,t}, \hat{e}_{2,t}\right)$. This implies that we test forecast encompassing conditional on the information set the respective forecasters incorporate in their issued forecasts. It is important to notice that this might be a subset of the full information set the respective forecaster had at hand.

Table 3: This table reports the p-values of the three different encompassing tests introduced in Section 2 and of the quantile encompassing test of Giacomini and Komunjer (2005). The p-value in the i-th row and j-th column of the respective matrices corresponds to testing for $\mathbb{H}_0$: the forecasts from model i encompass the forecasts from model j. * denote model pairs, where both encompassing tests (test whether model i encompasses model j and vice versa) are significant at the 10% level.

	Joint VaR ES Encmp					VaR Encmp				
	GJR-G	RM	HS	G-1F	G-2F	GJR-G	RM	HS	G-1F	G-2F
GJR-GARCH		0.00*	0.08*	0.21	0.00		0.00	0.00*	0.00*	0.00
RiskMetrics	0.04*		0.03*	0.00*	0.09*	0.14		0.14	0.11	0.21
Historical Simulation	0.00*	0.00*		0.00*	0.00	0.00*	0.00		0.00	0.00
VaR/ES GAS-1Factor	0.00	0.00*	0.03*		0.00	0.00*	0.01	0.82		0.01
VaR/ES GAS-2Factor	0.22	0.00*	0.86	0.23		0.14	0.00	0.98	0.38	
	Aux ES Encmp					**Strict ES Encmp**				
	GJR-G	RM	HS	G-1F	G-2F	GJR-G	RM	HS	G-1F	G-2F
GJR-GARCH		0.00*	0.03*	0.11	0.00		0.00*	0.03*	0.10	0.00
RiskMetrics	0.06*		0.06*	0.03*	0.14	0.05*		0.06*	0.04*	0.16
Historical Simulation	0.00*	0.00*		0.00*	0.00	0.00*	0.00*		0.00*	0.00
VaR/ES GAS-1Factor	0.08	0.00*	0.04*		0.00	0.08	0.00*	0.05*		0.00
VaR/ES GAS-2Factor	0.78	0.46	0.52	0.12		0.76	0.36	0.50	0.10	

Table 3 reports the respective p-values of the three ES-specific encompassing tests introduced in this paper, together with the quantile encompassing test of Giacomini and Komunjer (2005), based on the same vector of instruments. The p-value in the i-th row and j-th column of the respective matrices corresponds to testing for $\mathbb{H}_0$: the forecasts from model i encompass the forecasts from model j. Furthermore, * denote model pairs, where both encompassing tests (test whether model i encompasses model j and vice versa) are significant at the 10% level.

A rejection of both null hypotheses implies that neither forecast can encompass its competitor. Thus, both forecasts add some information and a forecast combination is superior to the stand-alone forecasting models. We find that out of the ten pairwise comparisons, the VaR encompassing test rejects both null hypotheses for two pairs. In comparison, the Joint VaR and ES encompassing tests rejects for six pairs and both, the Auxiliary ES and the strict ES encompassing reject for five pairs. These results imply the following conclusions. First, we find that all three ES encompassing tests reject fairly often and thus, our results support the theoretical advantages of forecast combination for the ES. Second, the three ES-specific encompassing tests jointly reject both hypotheses in more cases than the VaR test and thus, forecast combination is even more promising for the ES (and the pair VaR and ES) than it already is for the VaR. Third, the two tests which only focus on testing regression parameters for the ES perform very similar to the joint VaR and ES encompassing test. Thus, one can apply tests for forecast encompassing of the ES in cases where one does not have VaR forecasts at hand, such as it is currently imposed by the Basel Committee of Banking Supervision Basel Committee (2016, 2017).

5. Conclusion

This paper provides theory for tests on conditional forecast encompassing when the competing forecasts are issued for functionals which are non-elicitable, which means that there does not exist an adequate loss function for the evaluation of these forecasts. Important examples for these functionals are the Expected Shortfall (ES) and the variance in presence of a non-zero mean. We overcome the problem of non-elicitability by considering pairs of functionals such as the pair (mean, variance) and (quantile, ES) in order to facilitate the elicitation of these vector-valued functionals. We focus on proposing encompassing tests for the ES, where we propose three test variants. The first tests joint encompassing of the quantile and the ES, whereas the second and third consider encompassing of the ES stand-alone. We show through simulation studies that all proposed tests are reasonably sized in typical financial applications. Furthermore, all tests exhibit good power properties against general alternatives.

Tests for forecast encompassing establish a theoretical foundation for forecast combination of two competing forecasts when both opposing hypotheses of forecast encompassing are rejected. This situation corresponds to the case when neither forecast encompasses its competitor. Generally, applying forecast combinations can be highly beneficial through the diversification gains stemming from combining different model specifications and underlying information sets. This benefit can be particularly pronounced for extreme risk measures such as the ES as the stand-alone models are very sensitive to the very little observations in the tails of the return distributions. Thus, forecast combination can be seen as a robustification of the forecasts.

We apply the new encompassing tests for the ES to the problem of evaluating ES forecasts for financial returns, which is paramount at presence due to recent introduction of the ES into the regulatory framework of the Basel Committee of Banking Supervision (Basel Committee, 2016, 2017). We use daily returns from the IBM stock and show that combined forecasts for the ES at level 2.5% outperform the stand-alone models in five or six of the ten considered cases. Furthermore, our results imply that the gains from forecast combination for the ES are even more pronounced than they are for the VaR (Giacomini and Komunjer, 2005). The third variant of our test, which can also be applied when only ES forecasts (without their accompanying VaR forecasts) are available exhibits very similar results. This test variant can be seen as particularly relevant as by the current set of implied rules, the regulatory authorities only obtain ES forecasts and thus, can only apply this test.

Appendix A Proofs

Proof of Proposition 2.4. Recall that we define

$$\theta_t^* = (\beta_{1,t}^*, \beta_{2,t}^*, \eta_{1,t}^*, \eta_{2,t}^*) := \operatorname*{argmin}_{(\beta,\eta)\in\Theta} \mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right], \tag{A.1}$$

where we assume that this minimum is unique. We first show that

$$\nabla_\theta \mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right] = \mathbb{E}_t \left[\tilde{\psi}_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) \right], \tag{A.2}$$

almost surely, where

$$\tilde{\psi}_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) = \begin{pmatrix} \hat{\boldsymbol{q}}_t \left(\alpha - \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}\right) \left(g'(\hat{\boldsymbol{q}}_t^\top \beta) + \frac{\phi'(\hat{\boldsymbol{e}}_t^\top \eta)}{\alpha}\right) \\ \hat{\boldsymbol{e}}_t \phi''(\hat{\boldsymbol{e}}_t^\top \eta) \left(\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{(\hat{\boldsymbol{q}}_t^\top \beta - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}}{\alpha}\right) \end{pmatrix}. \tag{A.3}$$

For this, we first notice that it holds that

$$\mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right] = \left(F_t(\hat{\boldsymbol{q}}_t^\top \beta) - \alpha\right) g(\hat{\boldsymbol{q}}_t^\top \beta) - \mathbb{E}_t \left[g(Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \right] \tag{A.4}$$

$$+\phi'(\hat{\boldsymbol{e}}_t^\top \eta) \left(\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{\hat{\boldsymbol{q}}_t^\top \beta F_t(\hat{\boldsymbol{q}}_t^\top \beta)}{\alpha} - \frac{1}{\alpha} \mathbb{E}_t \left[Y_{t+1} \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \right] \right) - \phi(\hat{\boldsymbol{e}}_t^\top \eta), \tag{A.5}$$

and $\nabla_\beta \mathbb{E}_t \left[g(Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \right] = \hat{\boldsymbol{q}}_t g(\hat{\boldsymbol{q}}_t^\top \beta) f_t(\hat{\boldsymbol{q}}_t^\top \beta)$ as the distribution F_t is absolutely continuous. Thus,

$$\nabla_\beta \mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right] = \hat{\boldsymbol{q}}_t \Big\{ f_t(\hat{\boldsymbol{q}}_t^\top \beta) g(\hat{\boldsymbol{q}}_t^\top \beta) + g'(\hat{\boldsymbol{q}}_t^\top \beta)\left(F_t(\hat{\boldsymbol{q}}_t^\top \beta) - \alpha\right) - g(\hat{\boldsymbol{q}}_t^\top \beta) f_t(\hat{\boldsymbol{q}}_t^\top \beta) \tag{A.6}$$

$$+ \phi'(\hat{\boldsymbol{e}}_t^\top \eta) \left(1 + \frac{\hat{\boldsymbol{q}}_t^\top \beta f_t(\hat{\boldsymbol{q}}_t^\top \beta)}{\alpha} + \frac{F_t(\hat{\boldsymbol{q}}_t^\top \beta)}{\alpha} - \frac{\hat{\boldsymbol{q}}_t^\top \beta f_t(\hat{\boldsymbol{q}}_t^\top \beta)}{\alpha} \right) \Big\} \tag{A.7}$$

$$= \hat{\boldsymbol{q}}_t \left(F_t(\hat{\boldsymbol{q}}_t^\top \beta) - \alpha\right) \left(g'(\hat{\boldsymbol{q}}_t^\top \beta) + \frac{\phi'(\hat{\boldsymbol{e}}_t^\top \eta)}{\alpha} \right). \tag{A.8}$$

Equivalently, we get that

$$\nabla_\eta \mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right] = \hat{\boldsymbol{e}}_t \phi''(\hat{\boldsymbol{e}}_t^\top \eta) \left(\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{\hat{\boldsymbol{q}}_t^\top \beta F_t(\hat{\boldsymbol{q}}_t^\top \beta)}{\alpha} - \frac{1}{\alpha} \mathbb{E}_t \left[Y_{t+1} \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}} \right] \right). \tag{A.9}$$

Consequently, (A.2) follows.

As θ_t^* uniquely minimizes $\mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right]$ almost surely, which is a continuously differentiable function, we get that $\nabla_\theta \mathbb{E}_t \left[\rho_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right]_{\theta=\theta_t^*} = 0$ almost surely. As the $\mathcal{F}_t$-measureable vectors $\hat{\boldsymbol{q}}_t$ and $\hat{\boldsymbol{e}}_t$ are nonzero almost surely, we also get that

$$\mathbb{E}_t \left[\varphi_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right] = \mathbb{E}_t \left[\begin{pmatrix} \left(\alpha - \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}\right) \left(g'(\hat{\boldsymbol{q}}_t^\top \beta) + \frac{\phi'(\hat{\boldsymbol{e}}_t^\top \eta)}{\alpha}\right) \\ \phi''(\hat{\boldsymbol{e}}_t^\top \eta) \left(\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{(\hat{\boldsymbol{q}}_t^\top \beta - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}}{\alpha}\right) \end{pmatrix} \right] = 0 \tag{A.10}$$

almost surely. Furthermore, it holds that

$$\mathbb{E}_t \left[\varphi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right] = A(\hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \times \mathbb{E}_t \left[\varphi_{QES}^{g,\phi}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta) \right], \tag{A.11}$$

for all $\theta = (\beta, \eta) \in \Theta$, where

$$A\big(\hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta\big) = \begin{pmatrix} \Big(g'(\hat{\boldsymbol{q}}_t^\top \beta) + \frac{\phi'(\hat{\boldsymbol{e}}_t^\top \eta)}{\alpha}\Big) & 0 \\ 0 & \phi''(\hat{\boldsymbol{e}}_t^\top \eta) \end{pmatrix}^{-1}, \tag{A.12}$$

which has full rank almost surely as ϕ' and ϕ'' are strictly positive and g' is non-negative. Thus, we can conclude that $\mathbb{E}_t\left[\varphi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta_t^*)\right] = 0$ almost surely.

$$\psi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) = \begin{pmatrix} \hat{\boldsymbol{q}}_t\left(\alpha - \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}\right) \\ \hat{\boldsymbol{e}}_t\left(\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta + \frac{(\hat{\boldsymbol{q}}_t^\top \beta - Y_{t+1})\mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}}{\alpha}\right) \end{pmatrix}. \tag{A.13}$$

and

$$\mathbb{E}_t\left[\rho_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta)\right] = A(\hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta) \times \mathbb{E}_t\left[\rho_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t^\top \beta, \hat{\boldsymbol{e}}_t^\top \eta)\right], \tag{A.14}$$

We further have to show that $\mathbb{E}_t\left[\varphi_{QES}^{2step}(Y_{t+1}, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \theta)\right] = 0$ almost surely (for some $\theta \in \Theta$) implies that $\theta = \theta_t^*$. For this, we first consider the first component of φ_{QES}^{2step}, where $0 = \mathbb{E}_t\left[\alpha - \mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta\}}\right] = \alpha - F_t(\hat{\boldsymbol{q}}_t^\top \beta)$, implies that $F_t(\hat{\boldsymbol{q}}_t^\top \beta) = \alpha$ almost surely. As F_t is absolutely continuous, this implies that $\hat{\boldsymbol{q}}_t^\top \beta = \hat{\boldsymbol{q}}_t^\top \beta_t^*$ almost surely and consequently $\beta = \beta_t^*$. Now, we consider the second component of φ_{QES}^{2step}, where we obtain $\beta = \beta_t^*$ from above. This yields that

$$\begin{aligned} 0 &= \mathbb{E}_t\left[\hat{\boldsymbol{e}}_t^\top \eta - \hat{\boldsymbol{q}}_t^\top \beta_t^* + \frac{(\hat{\boldsymbol{q}}_t^\top \beta_t^* - Y_{t+1})\mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta_t^*\}}}{\alpha}\right] && \text{(A.15)} \\ &= \hat{\boldsymbol{q}}_t \beta_t^* \frac{F_t(\hat{\boldsymbol{q}}_t^\top \beta_t^*) - \alpha}{\alpha} + \hat{\boldsymbol{e}}_t^\top \eta - \frac{1}{\alpha}\mathbb{E}_t\left[Y_{t+1}\mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta_t^*\}}\right] && \text{(A.16)} \\ &= \hat{\boldsymbol{e}}_t^\top \eta - \frac{1}{\alpha}\mathbb{E}_t\left[Y_{t+1}\mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta_t^*\}}\right], && \text{(A.17)} \end{aligned}$$

where the third equality follows from $F_t(\hat{\boldsymbol{q}}_t^\top \beta_t^*) = \alpha$ almost surely. An equivalent calculation yields that $\hat{\boldsymbol{e}}_t^\top \eta_t^* = \frac{1}{\alpha}\mathbb{E}_t\left[Y_{t+1}\mathbb{1}_{\{Y_{t+1} \le \hat{\boldsymbol{q}}_t^\top \beta_t^*\}}\right]$ almost surely. Thus, we can conclude that $\hat{\boldsymbol{e}}_t^\top \eta_t^* = \hat{\boldsymbol{e}}_t^\top \eta_t$ almost surely which implies that $\eta = \eta_t^*$.

□

Proof of Proposition 2.5. The consistency proof of this two-step estimate $(\hat{\beta}_n, \hat{\eta}_n)$ follows the same idea as Theorem 2.6 of Newey and McFadden (1994). The only difference ist that we have to show uniform convergence in probability for both moment conditions $\Psi_{QES,1}^{2step}$ and $\Psi_{QES,2}^{2step}$. The reasoning for the second step is equivalent as in the proof of Theorem 2.6 of Newey and McFadden (1994). One has to apply the same inequality as in this proof for the second moment condition in order to be able to apply Theorem 2.1 of Newey and McFadden (1994) again.

The consistency proof of the first-step estimate $\hat{\beta}_n$ follows Theorem 2.1 of Newey and McFadden (1994). Condition (i) directly follows from Proposition 2.4 and condition (ii) is simply assumed. Condition (iii) follows as

$$\mathbb{E}\left[\psi_{QES,1}^{2step}(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \beta^*)\right] = \mathbb{E}\left[\boldsymbol{W}_t\left(\alpha - F_t(\hat{\boldsymbol{q}}_t \beta)\right)\right], \tag{A.18}$$

which is continuous in β as F_t is assumed to be continous.

Next, we show the uniform convergence condition (iv) by applying the uniform weak law of large numbers given in Theorem A.2.5. in White (1994). For that, we have to show that

(A) the map $\beta \mapsto \psi^{2step}_{QES,1}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \beta\right)$ is Lipschitz-L_1 almost surely on Θ^β (see Definition A.2.3 in White (1994),

(B) For all $\beta^o \in \Theta^\beta$, there exists $\delta^o > 0$, such that for all $\delta, 0 < \delta \leq \delta^o$, the sequences

$$\bar{\psi}_{1,t}(\beta^o, \delta) := \sup_{\beta \in \Theta^\beta} \left\{ \psi^{2step}_{QES,1}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \beta\right), \middle| \, ||\beta - \beta^o|| < \delta \right\} \qquad \text{and} \tag{A.19}$$

$$\underline{\psi}_{1,t}(\beta^o, \delta) := \inf_{\beta \in \Theta^\beta} \left\{ \psi^{2step}_{QES,1}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \beta\right), \middle| \, ||\beta - \beta^o|| < \delta \right\} \tag{A.20}$$

obey a weak law of large numbers.

The L_1-Lipschitz continuity condition (A) of $\psi^{2step}_{QES,1}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \beta\right)$ on Θ^β is obvious as the function is a constant times $\boldsymbol{W}_t$ apart from the point where $Y_{t+1} = \hat{\boldsymbol{q}}_t^\top \beta$, which forms a nullset with respect to the distribution F_t. We next show condition (B). The sequences $\bar{\psi}_{1,t}(\beta^o, \delta)$ and $\underline{\psi}_{1,t}(\beta^o, \delta)$ are strong mixing of size $-\tilde{r}/(\tilde{r}-1)$ for some $r > 1$ by condition (a) and by applying Theorem 3.49 in White (2001), p. 50 as the functions $\psi^{2step}_{QES,1}$ and the supremum and infimum functions are measureable. Furthermore, we get that $\mathbb{E}\left[|\bar{\psi}_t(\beta^o, \delta)|^{\tilde{r}+\delta}\right] \leq \sup_{1 \leq t \leq T} < \infty$ for all $t, 1 \leq t \leq T, T \geq 1$ and for some $\delta > 0$ and the same inequality holds for $|\underline{\psi}_{1,t}(\theta^o, \delta)|$ by condition (d).

Thus, we can apply the weak law of large numbers for strong mixing sequences in Corollary 3.48 in White (2001), p. 49 in order to conclude that for all $\beta^o \in \Theta^\beta$ such that $||\beta^o - \beta|| \leq \delta$, it holds that $\frac{1}{T}\sum_{t=1}^T \left(\bar{\psi}_{1,t}(\beta^o, \delta) - \mathbb{E}\left[\bar{\psi}_{1,t}(\beta^o, \delta)\right]\right) \xrightarrow{\mathbb{P}} 0$ and $\frac{1}{T}\sum_{t=1}^T \left(\underline{\psi}_{1,t}(\theta^o, \delta) - \mathbb{E}\left[\underline{\psi}_{1,t}(\beta^o, \delta)\right]\right) \xrightarrow{\mathbb{P}} 0$, which shows condition (B). Consequently, the uniform convergence condition (iv) holds by applying the uniform weak law of large numbers given in Theorem A.2.5. in White (1994).

The proof of the uniform convergence condition of the second identification function goes along these lines by again applying Theorem A.2.5. in White (1994). First, the map $(\beta, \eta) \mapsto \psi^{2step}_{QES,2}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \beta, \eta\right)$ is Lipschitz-L_1 almost surely on Θ. The map $\psi^{2step}_{QES,2}$ consists of a piecewise linear function multiplied by $\boldsymbol{W}_t$ and thus, it is Lipschitz-L_1 as the sequences $\frac{1}{n}\sum_{t=m+1}^T \mathbb{E}\left[||\boldsymbol{W}_t \hat{\boldsymbol{q}}_t^\top||\right]$ and $\frac{1}{n}\sum_{t=m+1}^T \mathbb{E}\left[||\boldsymbol{W}_t \hat{\boldsymbol{e}}_t^\top||\right]$ are bounded for all $\theta^o \in \Theta$.

The functions $\bar{\psi}_{2,t}(\beta^o, \delta)$ and $\underline{\psi}_{2,t}(\beta^o, \delta)$ which are defined as above obey a weak law of large numbers as they are strong mixing by the same arguments. Furthermore, $\mathbb{E}\left[|\bar{\psi}_{2,t}(\beta^o, \delta)|^{\tilde{r}+\delta}\right] \leq \sup_{1 \leq t \leq T} \mathbb{E}\left[\sup_{\theta \in \Theta} \left\|\boldsymbol{W}_t \hat{\boldsymbol{q}}_t^\top \beta\right\|^{\tilde{r}+\delta} + \left\|\boldsymbol{W}_t \hat{\boldsymbol{e}}_t^\top \eta\right\|^{\tilde{r}+\delta} + ||\boldsymbol{W}_t Y_{t+1}||^{\tilde{r}+\delta}\right] < \infty$. The result then follows by applying Theorem 2.6 of Newey and McFadden (1994), which completes the proof of consistency of the two-step estimator. □

Proof of Proposition 2.6. For this proof, we apply Theorem 1 of Bartalotti (2013), who shows that the asymptotic normality results for the two-step GMM estimator given in Theorem 6.1 of Newey and McFadden (1994) also holds in the case of nonsmooth objective functions. In this case, instead of checking the standard regularity conditions of Theorem 3.4 of Newey and McFadden (1994), one has to verify the conditions of Theorem 7.2, which is the counterpart for nonsmooth objective functions.

In this case, the weighting matrix is the identity matrix in our case and can consequently be ignored in the proof. Condition (i) of Theorem 7.2 follows directly as in the proof of Proposition 2.5 and condition (iii) follows directly from assumption (e). For condition (ii), we have to show continuity of the two components $\mathbb{E}\left[\Psi^{2step}_{QES,1}(\beta)\right]$ and $\mathbb{E}\left[\Psi^{2step}_{QES,2}(\beta, \eta)\right]$. For this,

$$\mathbb{E}\left[\Psi^{2step}_{QES}(\beta)\right] = \mathbb{E}\left[\boldsymbol{W}_t\left(\alpha - F_t(\hat{\boldsymbol{q}}_t^\top \beta)\right)\right], \tag{A.21}$$

which is continously differentibale in β as F_t is continuously differentibale with derivative

$$\nabla_\theta \mathbb{E}\left[\Psi^{2step}_{QES,1}(\beta)\right] = \left(\mathbb{E}\left[\boldsymbol{W}_t \hat{\boldsymbol{q}}_t^\top f_t(\hat{\boldsymbol{q}}_t^\top \beta)\right] \quad 0\right). \tag{A.22}$$

For the second component, we get that

$$\mathbb{E}\left[\Psi^{2step}_{QES,2}(\beta,\eta)\right] = \mathbb{E}\left[\boldsymbol{W}_t\left(\hat{\boldsymbol{e}}_t^\top\eta - \hat{\boldsymbol{q}}_t^\top\beta + \frac{1}{\alpha}(\hat{\boldsymbol{q}}_t^\top\beta - Y_{t+1})\mathbb{1}_{\{Y_{t+1}\leq\hat{\boldsymbol{q}}_t^\top\beta\}}\right)\right] \tag{A.23}$$

$$= \mathbb{E}\left[\boldsymbol{W}_t\left(\hat{\boldsymbol{q}}_t^\top\beta\left(\frac{F_t(\hat{\boldsymbol{q}}_t^\top\beta)-\alpha}{\alpha}\right) + \hat{\boldsymbol{e}}_t^\top\eta - \frac{1}{\alpha}\mathbb{E}_t\left[Y_{t+1}\mathbb{1}_{\{Y_{t+1}\leq\hat{\boldsymbol{q}}_t^\top\beta\}}\right]\right)\right]. \tag{A.24}$$

It holds that the term $\mathbb{E}_t\left[Y_{t+1}\mathbb{1}_{\{Y_{t+1}\leq\hat{\boldsymbol{q}}_t^\top\beta\}}\right]$ is continuously differentiable as F_t is continuously differentiable and thus

$$\nabla_\beta\mathbb{E}_t\left[Y_{t+1}\mathbb{1}_{\{Y_{t+1}\leq\hat{\boldsymbol{q}}_t^\top\beta\}}\right] = \nabla_\beta\int y\mathbb{1}_{\{y\leq\hat{\boldsymbol{q}}_t^\top\beta\}}\mathrm{d}F_t(y) = \nabla_\beta\int_{-\infty}^{\hat{\boldsymbol{q}}_t^\top\beta} y\,\mathrm{d}F_t(y) = \hat{\boldsymbol{q}}_t(\hat{\boldsymbol{q}}_t^\top)\beta f_t(\hat{\boldsymbol{q}}_t^\top). \tag{A.25}$$

Consequently, we get that

$$\nabla_\beta\mathbb{E}\left[\Psi^{2step}_{QES,2}(\beta^*,\eta^*)\right] = \boldsymbol{W}_t\hat{\boldsymbol{q}}_t^\top\left(\frac{F_t(\hat{\boldsymbol{q}}_t^\top\beta^*)-\alpha}{\alpha}\right) + \boldsymbol{W}_t(\hat{\boldsymbol{q}}_t^\top\beta^*)\frac{f_t(\hat{\boldsymbol{q}}_t^\top\beta^*)}{\alpha} - \boldsymbol{W}_t(\hat{\boldsymbol{q}}_t^\top\beta^*)\frac{f_t(\hat{\boldsymbol{q}}_t^\top\beta^*)}{\alpha} \tag{A.26}$$

$$= \boldsymbol{W}_t\hat{\boldsymbol{q}}_t^\top\left(\frac{F_t(\hat{\boldsymbol{q}}_t^\top\beta^*)-\alpha}{\alpha}\right) \tag{A.27}$$

and $\nabla_\eta\mathbb{E}\left[\Psi^{2step}_{QES,2}(\beta^*,\eta^*)\right] = \boldsymbol{W}_t\hat{\boldsymbol{e}}_t^\top$. Together, this yields that

$$\Lambda_n = \frac{1}{n}\sum_{t=m+1}^{T}\nabla_\theta\mathbb{E}\left[\Psi^{2step}_{QES,2}(\beta^*,\eta^*)\right] = \frac{1}{n}\sum_{t=m+1}^{T}\mathbb{E}\left[\begin{pmatrix}\boldsymbol{W}_t\hat{\boldsymbol{q}}_t^\top f_t(\hat{\boldsymbol{q}}_t^\top\beta^*) & 0\\ 0 & \boldsymbol{W}_t\hat{\boldsymbol{e}}_t^\top\end{pmatrix}\right]. \tag{A.28}$$

From condition (j), we get that the matrix Λ_n has full column rank and thus, $\Lambda_n^\top\Lambda_n$ is positive definite.

We now show condition (iv), i.e. the asymptotic normality of $\mathbb{E}\left[\Psi^{2step}_{QES,2}(\beta^*,\eta^*)\right]$ in the following. We show this (multivariate result) by applying the Cramér-Wold theorem, i.e. by showing that the conditions for the univariate CLT for α-mixing sequences given in Theorem 5.20 in White (2001), p.130 hold for all linear combinations $u^\top\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta,\eta)$ for all $u\in\mathbb{R}^k$ such that $||u||=1$. By assumption (f) and by applying Theorem 3.49 in White (2001) p.50, we get that the sequences $\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta,\eta)$ and $u^\top\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta,\eta)$ are strong mixing of size $-r/(r-2)$ for some $r>2$. Furthermore, by condition (g), it holds that for some $\delta>0$,

$$\mathbb{E}\left[\left|u^\top\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta,\eta)\right|^{r+\delta}\right] < \infty,$$

for all $t\in\mathbb{N}$. As $\mathbb{E}_t\left[\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta^*,\eta^*)\right]=0$ for all $t\in\mathbb{N}$, we first notice that for all $s,t\in\mathbb{N}$, $s<t$,

$$\mathrm{Cov}\left(\psi^{2step}_{QES}(Y_{s+1},\boldsymbol{W}_s,\hat{\boldsymbol{q}}_s,\hat{\boldsymbol{e}}_s,\beta^*,\eta^*),\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta^*,\eta^*)\right) \tag{A.29}$$

$$=\mathbb{E}\left[\psi^{2step}_{QES}(Y_{s+1},\boldsymbol{W}_s,\hat{\boldsymbol{q}}_s,\hat{\boldsymbol{e}}_s,\beta^*,\eta^*)\cdot\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta^*,\eta^*)^\top\right] \tag{A.30}$$

$$=\mathbb{E}\left[\psi^{2step}_{QES}(Y_{s+1},\boldsymbol{W}_s,\hat{\boldsymbol{q}}_s,\hat{\boldsymbol{e}}_s,\beta^*,\eta^*)\cdot\mathbb{E}_t\left[\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta^*,\eta^*)^\top\right]\right]=0. \tag{A.31}$$

Thus, for all $n\geq 1$,

$$\Sigma_n = \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=m+1}^{T}\psi^{2step}_{QES}(Y_{t+1},\boldsymbol{W}_t,\hat{\boldsymbol{q}}_t,\hat{\boldsymbol{e}}_t,\beta^*,\eta^*)\right) \tag{A.32}$$

$$= \frac{1}{n} \sum_{t=m+1}^{T} \mathbb{E}\left[\psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \beta^*, \eta^*\right) \cdot \psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \beta^*, \eta^*\right)^\top\right]. \tag{A.33}$$

As Σ_n is real and symmetric and positive definite for all $n \in \mathbb{N}$, it can be diagonalized with a real orthogonal matrix S_n, i.e. $S_n^\top \Sigma_n S_n = D_n$, where D_n is a diagonal matrix containing the Eigenvalues of Σ_n, denoted by $\{\lambda_{1,n}, \dots, \lambda_{k,n}\}$. Consequently, for any $u \in \mathbb{R}^k$,

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=m+1}^{T} u^\top \psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \beta^*, \eta^*\right)\right) = u^\top \Sigma_n u = u^\top S_n^\top D_n S_n u = v_n^\top D_n v_n > \min_{i=1,\dots,k} \lambda_{i,n}, \tag{A.34}$$

where $v_n = S_n u$, i.e. $||v|| = 1$ as S_n is orthogonal and where the Eigenvalues $\{\lambda_{1,n}, \dots, \lambda_{k,n}\}$ are bounded away from zero for n sufficiently large. Thus, we can apply Theorem 5.20 in White (2001) p. 130 for asymptotic normality of the sequences $u^\top \psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \beta^*, \eta^*\right)$ for all $u \in \mathbb{R}^k$ such that $||u|| = 1$. Applying the Cramér-Wold theorem concludes the argument.

We conclude that

$$\frac{1}{\sqrt{n}} \sum_{t=m+1}^{T} \psi_{QES}^{2step}\left(Y_{t+1}, \boldsymbol{W}_t, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t, \hat{\theta}_n\right) \xrightarrow{\mathbb{P}} 0 \tag{A.35}$$

as $n \to \infty$ as in the proof of Lemma 2 in the online supplement of Patton et al. (2017), which is a primitive condition for $\Psi_{QES}^{2step}\left(\hat{\theta}_n\right)^\top \Psi_{QES}^{2step}\left(\hat{\theta}_n\right) \le \inf_{\theta \in \Theta} \Psi_{QES}^{2step}(\theta)^\top \Psi_{QES}^{2step}(\theta) + o_P(n^{-1})$ (see the discussion on p.2187 in Newey and McFadden, 1994). Eventually, the equicontinuity condition (v) directly follows from assumption (l).

Thus, we have verified the conditions of Theorem 7.2 of Newey and McFadden (1994) and can consequently apply Theorem 1 of Bartalotti (2013). In the notation of Bartalotti (2013), it holds that

$$G_{12} = \nabla_\beta \mathbb{E}\left[\Psi_{QES,2}^{2step}(\beta, \eta)\right] = 0 \tag{A.36}$$

and consequently, we obtain that

$$B = \begin{pmatrix} -G_{11}^{-1} & G_{11}^{-1} G_{12} G_{22}^{-1} \\ 0 & -G_{22}^{-1} \end{pmatrix} = \begin{pmatrix} -G_{11}^{-1} & 0 \\ 0 & -G_{22}^{-1} \end{pmatrix}. \tag{A.37}$$

Thus, the result of the Proposition follows. □

Proof of Proposition 2.7. The proof of $\hat{\Sigma}_n - \Sigma_n \xrightarrow{\mathbb{P}} 0$ is straight-forward given condition (m) and will be omitted here. The proof of $\hat{\Lambda}_n - \Lambda_n \xrightarrow{\mathbb{P}} 0$ follows the idea of Engle and Manganelli (2004). For this, we define

$$\tilde{\Lambda}_n = \frac{1}{n} \sum_{t=m+1}^{T} \begin{pmatrix} \boldsymbol{W}_{q,t} \hat{\boldsymbol{q}}_t^\top \frac{1}{2c_n} \mathbb{1}_{\{|Y_{t+1} - \hat{\boldsymbol{q}}_t^\top \beta^*| < c_n\}} & 0 \\ 0 & \boldsymbol{W}_{e,t} \hat{\boldsymbol{e}}_t^\top \end{pmatrix}. \tag{A.38}$$

We first consider

$$\|\tilde{\Lambda}_n - \hat{\Lambda}_n\| = \left\| \frac{1}{2nc_n} \sum_{t=m+1}^{T} \boldsymbol{W}_{q,t} \hat{\boldsymbol{q}}_t^\top \left(\mathbb{1}_{\{|Y_{t+1} - \hat{\boldsymbol{q}}_t^\top \beta^*| < c_n\}} - \mathbb{1}_{\{|Y_{t+1} - \hat{\boldsymbol{q}}_t^\top \hat{\beta}_n| < c_n\}} \right) \right\| \tag{A.39}$$

$$\tag{A.40}$$

which is $o_P(1)$ as in the proof of Theorem 3 in Engle and Manganelli (2004). We now turn to

$$\|\tilde{\Lambda}_n - \hat{\Lambda}_n\| = \left\| \frac{1}{2nc_n} \sum_{t=m+1}^{T} \left(\boldsymbol{W}_{q,t} \hat{\boldsymbol{q}}_t^\top \mathbb{1}_{\{|Y_{t+1} - \hat{\boldsymbol{q}}_t^\top \beta^*| < c_n\}} - \mathbb{E}\left[\boldsymbol{W}_{q,t} \hat{\boldsymbol{q}}_t^\top \mathbb{1}_{\{|Y_{t+1} - \hat{\boldsymbol{q}}_t^\top \hat{\beta}_n| < c_n\}} \right] \right) \right\| \tag{A.41}$$

$$= \dots \tag{A.42}$$

which is again $o_P(1)$ as in Engle and Manganelli (2004). Thus, the result of the Theorem follows. □

Proof of Theorem 2.8. From Proposition 2.6, it follows that

$$\left(\Lambda_n^\top \Sigma_n^{-1} \Lambda_n\right)^{-1/2} \sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d} \mathcal{N}(0, I), \tag{A.43}$$

and consistency of the covariance estimator follows from Proposition 2.7. Thus, the results of the Theorem follows by applying Theorem 4.30 of White (2001). □

References

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228.

Barendse, S. (2017). Interquantile Expectation Regression. Available at https://ssrn.com/abstract=2937665.

Bartalotti, O. (2013). GMM Efficiency and IPW Estimation for Nonsmooth Functions. Working Paper.

Basel Committee (2016). Minimum capital requirements for Market Risk. Technical report, Bank for International Settlements. Available at http://www.bis.org/bcbs/publ/d352.pdf.

Basel Committee (2017). Pillar 3 disclosure requirements – consolidated and enhanced framework. Technical report, Basel Committee on Banking Supervision. Available at http://www.bis.org/bcbs/publ/d400.pdf.

Bayer, S. (2018). Combining value-at-risk forecasts using penalized quantile regressions. *Econometrics and Statistics*, 8:56 – 77.

Chong, Y. Y. and Hendry, D. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, 53(4):671–690.

Clements, M. P. and Harvey, D. I. (2009). *Forecast Combination and Encompassing*, pages 169–198. Palgrave Macmillan UK, London.

Clements, M. P. and Harvey, D. I. (2010). Forecast encompassing tests and probability forecasts. *Journal of Applied Econometrics*, 25(6):1028–1062.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Diebold, F. and Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Diebold, F. X. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*, 5(4):589 – 592.

Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination. Working Paper 192, National Bureau of Economic Research.

Dimitriadis, T. and Bayer, S. (2017). A Joint Quantile and Expected Shortfall Regression Framework. arXiv:1704.02213 [math.ST].

Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies*, 72(4):1107–1125.

Engle, R. and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4):1680–1707.

Fissler, T. and Ziegel, J. F. (2018). Elicitability of interquantile expectation.

Giacomini, R. and Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*, 23:416–431.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Gneiting, T. (2011a). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gneiting, T. (2011b). Quantiles as optimal point forecasts. *International Journal of Forecasting*, 27(2):197–207.

Halbleib, R. and Pohlmeier, W. (2012). Improving the value at risk forecasts: Theory and evidence from the financial crisis. *Journal of Economic Dynamics and Control*, 36(8):1212–1228.

Harvey, A. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.

Harvey, D. and Newbold, P. (2000). Tests for multiple forecast encompassing. *Journal of Applied Econometrics*, 15(5):471–482.

Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, 101(1):245–251.

Hendry, D. and Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20(1):3–33.

Holzmann, H. and Eulert, M. (2014). The role of the information set for forecasting—with applications to risk management. *Ann. Appl. Stat.*, 8(1):595–621.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, 128(1):137–164.

Lambert, N. S., Pennock, D. M., and Shoham, Y. (2008). Eliciting Properties of Probability Distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138. ACM.

Mizon, G. and Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica*, 54(3):657–78.

Newbold, P. and Harvey, D. I. (2007). *Forecast Combination and Encompassing*, chapter 12, pages 268–283. John Wiley and Sons, Ltd.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.

Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Patton, A. J. and Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, 102(480):1172–1184.

Patton, A. J., Ziegel, J. F., and Chen, R. (2017). Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk). arXiv:1707.05108 [q-fin.EC].

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares esimation in the linear model. 75:828–838.

Taylor, J. W. (2017). Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution. *Forthcoming in Journal of Business & Economic Statistics*. DOI: 10.1080/07350015.2017.1281815.

Timmermann, A. (2006). Forecast combinations. volume 1, chapter 04, pages 135–196. Elsevier, 1 edition.

Weber, S. (2006). Distribution Invariant Risk Measures, Information, and Dynamic Consistency. *Mathematical Finance*, 16(2):419–441.

West, K. (2006). Forecast evaluation. volume 1, chapter 03, pages 99–134. Elsevier, 1 edition.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press, San Diego.