Mincer-Zarnovitz Quantile and Expectile Regressions for Forecast Evaluations under Aysmmetric Loss Functions

Kemal Guler *

Anadolu University

Pin T. Ng†

Northern Arizona University

Zhijie Xiao‡

Boston College

Abstract

Forecasts are pervasive in all areas of applications in business and daily life. Hence, evaluating the accuracy of a forecast is important for both the generators and consumers of forecasts. There are two aspects in forecast evaluation: (1) measuring the accuracy of past forecasts using some summary statistics and (2) testing the optimality properties of the forecasts through some diagnostic tests. On measuring the accuracy of a past forecast, this paper illustrates that the summary statistics used should match the loss function that was used to generate the forecast. If there is strong evidence that an asymmetric loss function has been used in the generation of a forecast, then a summary statistic that corresponds to that asymmetric loss function should be used in assessing the accuracy of the forecast instead of the popular root -mean-square error or mean-absolute error. On testing the optimality of the forecasts, it is demonstrated how the quantile regressions set in the prediction-realization framework of Mincer and Zarnowitz (1969) can be used to recover the unknown parameter that controls the potentially asymmetric loss function used in generating the past forecasts. Finally, the prediction-realization framework is applied to the Federal Reserve's economic growth forecast and forecast sharing in a PC manufacturing supply chain. It is found that the Federal Reserves values overprediction approximately 1.5 times more costly than underprediction. It is also found that the PC manufacturer weighs positive forecast errors (under forecasts) about four times as costly as negative forecast errors (over forecasts).

KEY WORDS: Asymmetric loss; Expectile regression; Forecast evaluation; Quantile regression

1 Introduction

*Department of Economics, Faculty of Economic & Administrative Sciences, Anadolu University, Eskisehir, Turkey.

†Franke College of Business, Northern Arizona University, Flagstaff, AZ 86011-5066, USA and School of Economics, Anhui University, Hefei, China.

‡Department of Economics, Boston College, Chestnut Hill, MA 02467, USA.

Forecasts are pervasive in all areas of business and daily life. Weather forecasts are important for planning day-to-day activities. Farmers rely on them for the planting and harvesting of crops while airline and cruise industries need them to make decisions that maintain safety in the sky and the sea. The insurance industry relies on them to form informed pricing and capital decisions. Corporations use forecasting to predict their future financial needs, production planning, human resource planning, etc. Forecasts are used by investors to value companies and their securities. The startup of a new business requires forecasts of the demand for the product, the expected shares in the market, the capacity of competitor, and the amount and sources for funds, etc. In supply chain management, businesses have to synchronize the ordering of supplies to meet the forecasted demand of their customers. In government policy decisions, economic forecasts are important for determining the appropriate monetary/fiscal policies. In the healthcare industry, forecasts can be used to target disease management or device personalized health care based on predicted risk.

As a result, evaluating the accuracy of a forecast is important for both the generators and consumers of forecasts. However, there is abundant evidence that many of the forecasts being generated are inconsistent with the realizations of the forecasted values. Silver (2012) discussed the weather industry's bias toward forecasting more precipitation than would actually occur, what meteorologists call "wet bias". Using 121 responses to a 26-question mail questionnaire sent to the highest ranking financial officers in 500 firms on the Fortune 500 listing, Pruitt and Gitman (1987) found that capital budgeting forecasts were optimistically biased by people with work experience. Ali et al. (1991) found that analysts set overly optimistic forecasts of the next period's annual earnings per shares. Lee et al. (1997) and Cohen et al. (2003) provided ample evidence of overoptimistic forecasts across industries ranging from electronics and semiconductors to medical equipment and commercial aircraft in the supply chains. In terms of economic variables forecasts, Capistrán (2008) provided evidence that the Federal Reserve's inflation forecasts systematically underpredicted before Paul Volcker appointment as Chairman and systematically overpredicted afterwards until the second quarter of 1998.

Do these forecast biases signify suboptimal forecast performance? In the traditional sense, the overprediction or underprediction bias are indications of suboptimal forecasts. However, the traditional tests for forecast optimality rely, typically, on the assumption of the symmetric square error loss function. Under this square error loss, overprediction and underprediction are weighted equally, and optimal forecasts imply that the observed forecast errors will have a zero bias and are uncorrelated with variables in the forecasters' information set.

However, strong arguments can be provided for the rationale that forecasters might not have adopted a symmetric error loss function. For example, in firms' forecasting of sales, overpredictions will result in over inventory and increased insurance costs, and tied up capital while underpredictions will lead to loss of goodwill, reputation, and current and future sales. Firms may decide that the cost of loss of goodwill is much higher than increased insurance costs. Therefore, they weigh the underprediction errors more than the overprediction errors. For money managers of banks, overpredicting the value-at-risk ties up more capital than necessary while underpredicting leads to regulatory penalties and the need for increased capital provisions. They may conclude that the cost of increased capital provisions is higher than the cost of tied up capital and decide to weigh the underprediction errors more than the overprediction errors. It might be particularly costly for the Federal Reserves to overpredict GDP growth when growth is already slow, signaling a false recovery, which could lead to an overly tight monetary policy at exactly the wrong time. The cost of over forecasting is not always the same as that of under forecasting. The dissatisfaction that people have when the weatherman forecasts a sunny day but it turns out to be a rainy day and, hence, ruin a picnic party, is higher than when it is forecasted to be rainy but turns out to be sunny. This can be the explanation for the wet bias and illustrate the asymmetric loss function used by the weatherman when performing their forecasts.

Keane and Runkle (1990) argued that:

> If forecasters have differential costs of over- and underprediction, it could be rational for them to produce biased forecasts. If we were to find that forecasts are biased, it could still be claimed that forecasters were rational if it could be shown that they had such differential costs. (719)

Varian (1974), Waud (1976), Zellner (1986), Christoffersen and Diebold (1997), and Patton and Timmermann (2007) all argued that the presence of forecast bias is not necessary an indication of suboptimal forecast. Rostek (2010) provided a foundation for the practical and theoretical justifications for the assignment of different weights for overprediction and underprediction by a forecaster. Inspired by prior works of Manski (1988) and Chambers (2007), Rostek (2010) formalized the concept of quantile maximization in choice-theoretic language in choice theory and demonstrated the characteristics of robustness and ordinality being its advantages when compared to the traditional moments based decision criteria. The specification of a τ quantile maximizer provides a systematic definition of riskiness in terms of downside risk and upside chance (losses and gains) in a forecaster's asymmetric preference towards overprediction and underprediction. In an attempt to improve the forecast performance of predicting state tax revenues in Iowa, Lewis and Whiteman (2015) provided example that the Institute for Economic Research at the University of Iowa had used an asymmetric loss function that treated forecasted revenue shortfalls $d = 1, 2, \ldots, 10$ times as costly as equal-sized surpluses.

Numerous articles have argued for the likelihood of and addressed the issues related to an asymmetric loss functions being used by forecasters (see e.g., Granger, 1969 & 1999; Varian, 1974; Granger and Newbold, 1986; Zellner, 1986; Ito, 1990; West et al., 1993; Weiss, 1996; Christoffersen and Diebold, 1997; Batchelor and Peel, 1998; Granger and Pesaran, 2000; Artis and Marcellino, 2001; Pesaran and Skouras, 2002; Carpistran, 2006; Patton and Timmermann, 2007; and Elliott and Timmermann, 2008).

Elliott and Timmermann (2008) summarized that forecast evaluation usually comprised of two separate, but related, tasks: (1) measuring the accuracy of past forecasts using some summary statistics and (2) testing the optimality properties of the forecasts through some diagnostic tests.

On the first task of assessing the accuracy of a forecast, the most popular summary measures have been the sample mean-square error (MSE), the sample root-mean-square error ($RMSE$) and the sample mean-absolute error (MAE). However, the MSE and $RMSE$ metrics are appropriate only if the loss function used in the forecaster's decision making process is that of the symmetric mean-square error. The MAE is appropriate if the loss function is the symmetric mean-absolute error. Elliott and Timmermann (2008) demonstrated that a forecast that was considered as good using one measure might not be good according to a different measure. Gneiting (2011a) provided detailed theoretical justifications for such recommendation and concluded that:

> If point forecasts are to be issued and evaluated, it is essential that either the scoring function be specified ex ante, or an elicitable target functional be named, such as the mean or a quantile of the predictive distribution, and scoring functions be used that are consistent for the target functional. (757)

The *scoring function* that Gneiting alluded to is the *summary measure* to be used to assess the accuracy of a forecast that Elliott and Timmermann (2008) referred to above while the elicitable *target functional* is determined by the loss function adopted by the forecaster. Hence, the chosen summary measure used to assess the accuracy of a forecast should depend on the loss function adopted by the forecaster when performing the forecast.

In Section 2, we demonstrate that if there is strong evidence that an asymmetric loss function has been used in the generation of a forecast, then a summary metric that corresponds to that asymmetric loss function, in particular the sample root-mean-weighted-square error ($RMWSE$) or the sample mean-weighted-absolute error ($MWAE$), should be used in assessing the accuracy of the forecast instead of the popular $RMSE$ or MAE, which are appropriate only if the symmetric loss function has been used to generate the forecast.

The second task of testing the optimality properties of a forecast also relies on the knowledge of the loss function being used in generating the forecast as well as the underlying data generating process (DGP) that generates the future values of the predicted variable, both of which are, unfortunately, unknown in typical situations. There are a few families of popular loss function specifications being used in the literature. Elliott et al. (2005) proposed a flexible family of loss functions which subsumed the asymmetric lin-lin piecewise linear loss function, in which the MAE loss was a special case, and the asymmetric quad-quad loss function, which nested the MSE loss. Varian (1974), Zellner (1986), and Christoffersen and Diebold (1997) used the linex loss while Christoffersen and Diebold (2006) investigated the sign loss function.

It can be readily shown that the optimal forecasts for the lin-lin loss function are the conditional quantiles while those for the quad-quad loss functions are the conditional expectiles. See, e.g., Raiffa and Schlaifer (1961), Ferguson (1967) and Gneiting (2011a, 2011b) for conditional quantiles, and Gneiting(2011a) for conditional expectiles. Increasingly, forecasts are being generated using the conditional quantiles or expectiles. Kokic et al. (2000) illustrated how expectile regression can be used to forecast farm's income. Bremnes (2004) and Nielsen et al. (2006) used quantile regression to forecast wind powers. Friederichs and Hense (2007) used censored quantile regression to forecast extreme precipitation in Germany. Taylor (2007) forecasted daily sales of various items from an outlet of a large UK supermarket chain using exponentially weighted quantile regression. Weerts et al. (2011) used quantile regression to forecast and estimate hydrological uncertainty in England and Wales. Soyiri and Reidpath (2013) applied quantile regression to forecast respiratory deaths in New York City. Soyiri et al. (2013) forecasted peak asthma admissions in London using quantile regression. Yu (2013) forecasted the development of information and communication technology using quantile regression. Bastianin et al. (2014) employed expectile regression to forecast food prices. Voudouris et al., (2014) forecasted plausible trajectories of natural gas production in the world using both quantile and expectile regressions.

With the increasing popularity of quantile regression (which corresponds to the adoption of the lin-lin loss function) in forecasting, an important issue that becomes eminent in testing the optimality properties of a forecast is the determination of the parameter that dictates the degree of asymmetry in the asymmetric loss function being used. Elliott and Timmermann (2008) suggested that one can estimate the unknown parameter that controls the degree of asymmetry of the lin-lin and quad-quad loss functions through the first order condition of the risk adopted by the forecasters. In Section 3, we demonstrate how the estimation of the unknown parameter can be accomplished using the quantile regressions and expectile regressions that are set naturally in the prediction-realization framework of Mincer and Zarnowitz (1969) depending on whether a lin-lin or quad-quad loss function is assumed to have been employed in generating the forecasts. After having estimated the asymmetric parameter that controls the asymmetry of the loss functions, one can then choose between the $RMSE$ and $RMWSE$, or the MAE and $MWAE$ as the appropriate metric in evaluating forecast accuracy. The major advantage of the Mincer-Zarnovitz quantile and expectile regression approach proposed here is its extremely easy implementation via existing statistical software that is capable of performing quantile and expectile regression estimations.

The rest of the paper is organized as follows. In Section 2, arguments and evidence are provided to illustrate that the metric used in measuring the accuracy of a forecast should be chosen according to the loss function being used when a forecast was performed. Section 3 demonstrates how the unknown parameter that controls the degree of asymmetry in a lin-lin loss function used by a forecaster can be recovered using the Mincer-Zarnowitz prediction-realization framework. It also presents the asymptotic of the Mincer-Zarnovitz quantile regression analysis, illustrates the implementation of the Mincer-Zarnovitz quantile regression forecast optimality test, and provides simulation results of the optimality test. Section 6 illustrates applications of the Mincer-Zarnowitz prediction-realization approach to assess the Federal Reserve's *Greenbook* forecasts and the sales forecasts of an electronic component manufacturer to try to recover the parameter of the possibly asymmetric loss function used in their forecast decisions.

2 Matching the Accuracy Measurements for Forecasts with the Loss Functions

We first introduce some basic notations before illustrating why the summary measures used for evaluating forecast accuracy should be determined by the loss functions used by the forecasters. Let Y be the random variable to be forecasted, $\mathcal{F}_t$ be the information set available at time t, $Z = \{Z_t\}_{t=1}^{T} = \{Y_t, X_t\}_{t=1}^{T}$ be the vector of relevant dependent and independent variables (data) which is part of $\mathcal{F}_t$ used in the 1 to n-period ahead forecast ($\left\{\hat{Y}_t\right\}_{t=T+1}^{T+n}$) of Y, T be the time when the forecast is performed, $\hat{Y} = g\left(Z, \theta\right)$ be the point forecast, in which θ is the unknown vector of parameters in the underlying forecast model, and $\rho\left(Y - g(Z, \theta)\right)$ be the loss function that maps the forecast ($\hat{Y}$), outcome of the forecast (Y) and data (Z) into the real line. The forecast decision is to choose $\hat{Y} = g\left(Z, \theta\right)$ that minimizes the risk

$$\begin{aligned} R\left(\theta, g\right) &= \mathrm{E}_{Y,Z}\left[\rho\left(Y - g\left(Z, \theta\right)\right)\right] \\ &= \int_z \int_y \rho\left(y - g\left(z, \theta\right)\right) f_Y\left(y|z, \theta\right) f_Z\left(z|\theta\right) dy dz \\ &= \int_z \mathrm{E}_Y\left[\rho\left(Y - g\left(Z, \theta\right)\right)|Z, \theta\right] f_Z\left(z|\theta\right) dz \end{aligned}$$

where $f_Y\left(y|z, \theta\right)$ and $f_Z\left(z|\theta\right)$ are the conditional probability density functions. The classical forecast minimizes

$$\mathrm{E}_Y\left[\rho\left(Y - g\left(Z, \theta\right)\right)|Z, \theta\right] = \int_y \rho\left(y - g\left(z, \theta\right)\right) f_Y\left(y|z, \theta\right) dy \tag{1}$$

given Z and θ. We define the forecast error as $e = Y - \hat{Y} = Y - g\left(z, \theta\right)$.

A few of the popular and commonly used loss functions are

1. Square loss: $\rho(e) = e^2$

2. Absolute loss: $\rho(e) = |e|$

3. Lin-lin loss: $\rho(e) = 2[\tau + (1-2\tau) I(e<0)]|e|$ for $0 < \tau < 1$ of which the absolute loss is a special case with $\tau = 0.5$

4. Quad-quad loss: $\rho(e) = 2[\omega + (1-2\omega) I(e<0)]\left(e^2\right)$ for $0 < \omega < 1$ of which the squared loss is a special case with $\omega = 0.5$

The most popular forecasts accuracy measures are the sample $MSE = \frac{\Sigma_{i=1}^{K} e_i^2}{K}$ and $MAE = \frac{\Sigma_{i=1}^{K}|e_i|}{K}$, which are the sample counterparts of the expected square loss and the expected absolute loss, respectively. Hence, it is natural to use the sample MSE, or root-mean-square error, $RMSE = \sqrt{\frac{\Sigma_{i=1}^{K} e_i^2}{K}}$, as a measure of forecast accuracy when the forecasts are generated using the symmetric square loss function and use the sample mean-absolute error, MAE, for the symmetric absolute loss function. However, we have provided arguments in the Introduction that one should use the sample mean-weighted-absolute error,

$$MWAE = \frac{\Sigma_{i=1}^{K} 2\left[(w) + (1-2w) I(e_i < 0)\right]|e_i|}{K}$$

as the metric to measure forecast accuracy for asymmetric lin-lin loss function, and use the sample mean-weighted-square error,

$$MWSE = \frac{\Sigma_{i=1}^{K} 2\left[(w) + (1-2w) I(e_i < 0)\right]\left(e_i^2\right)}{K}$$

or root mean-weighted-square error,

$$RMWSE = \sqrt{\frac{\Sigma_{i=1}^{K} 2\left[(w) + (1-2w) I(e_i < 0)\right](e_i^2)}{K}}$$

as the metric for the quad-quad loss function since these metrics are the corresponding sample counterparts of the respective expected loss functions.

2.1 Simulated Examples

A few simulations were performed to illustrate that the summary measures used for forecasts accuracy should match the loss functions used in generating the forecasts. Realizations of the forecasted variable Y were generated by the following data generating processes. In particular, the first three processes are traditional AR processes with different values of autocorrelation, the other two processes are QAR processes (Koenker and Xiao, 2006.).

DGP1 $Q_{y_t}(\tau|y_{t-1}) = \alpha_0(\tau) + \alpha_1 y_{t-1} = \left(2 + 3\Phi^{-1}(\tau)\right) + 0.6y_{t-1}$

where $Q_{y_t}(\tau|y_{t-1})$ was the conditional quantile function of Y given y_{t-1}. The realizations were generated using

$$y_t = \left(2 + 3\Phi^{-1}(u_t)\right) + 0.6y_{t-1} \text{ where } u_t \sim \text{i.i.d. } U[0,1]$$

with u_t being generated from an independently and identically distributed uniform distribution, $U[0,1]$.

DGP2

$$Q_{y_t}(\tau|y_{t-1}) = \alpha_0(\tau) + \alpha_1 y_{t-1} = \left(2 + 3\Phi^{-1}(\tau)\right) + 0.95y_{t-1},$$

$$y_t = \left(2 + 3\Phi^{-1}(u_t)\right) + 0.95y_{t-1} \text{ where } u_t \sim \text{i.i.d. } U[0,1]$$

DGP3

$$Q_{y_t}(\tau|y_{t-1}) = \alpha_0(\tau) + \alpha_1 y_{t-1} = \left(2 + 3\Phi^{-1}(\tau)\right) + 0.6y_{t-1} + 0.2y_{t-2},$$

$$y_t = \left(2 + 3\Phi^{-1}(u_t)\right) + 0.6y_{t-1} + 0.2y_{t-2} \text{ where } u_t \sim \text{i.i.d. } U[0,1]$$

DGP4

$$\begin{aligned} Q_{y_t}(\tau|y_{t-1}) &= \alpha_0(\tau) + \alpha_1(\tau) y_{t-1} \\ &= \left(2 + 3\Phi^{-1}(\tau)\right) + \min\left\{\frac{1}{4} + \tau, \frac{3}{4}\right\} y_{t-1} \end{aligned}$$

$$\begin{aligned} y_t &= \left(2 + 3\Phi^{-1}(u_t)\right) + \min\left\{\frac{1}{4} + u_t, \frac{3}{4}\right\} y_{t-1} \\ &\text{where } u_t \sim \text{i.i.d. } U[0,1] \end{aligned}$$

DGP5

$$\begin{aligned} Q_{y_t}(\tau|y_{t-1}) &= \alpha_0(\tau) + \alpha_1(\tau) y_{t-1} \\ &= \left(2 + 3\Phi^{-1}(\tau)\right) + \left[0.2I\left(2 + 3\Phi^{-1}(\tau) \geq 0\right)\right. \\ &\left. + 0.8I\left(2 + 3\Phi^{-1}(\tau) < 0\right)\right] y_{t-1} \end{aligned}$$

$$
\begin{aligned}
y_t &= \left(2 + 3\Phi^{-1}(u_t)\right) + \left[0.2I\left(2 + 3\Phi^{-1}(u_t) \geq 0\right)\right. \\
&\quad \left. + 0.8I\left(2 + 3\Phi^{-1}(u_t) < 0\right)\right] y_{t-1} \\
&\quad \text{where } u_t \sim \text{i.i.d. } U[0,1]
\end{aligned}
$$

The simulations were performed with an in-sample size of $n = 100$ with the number of 1-period ahead rolling forecasts used in the hold-out sample set at $K = 400$. A random sample of $N = n + K$ realizations from one of the above DGPs was generated. Starting from $T = n, \cdots, N-1$, 1-period ahead forecast on Y_{T+1} was performed using rolling quantile regression of Koenker and Bassett (1978) with $\tau \in (0.1, 0.2, \cdots, 0.8, 0.9)$ or expectile regression in Newey and Powell (1987) with $\omega \in (0.1, 0.2, \cdots, 0.8, 0.9)$ as well as arima (p, d, q) based on the data $Z = \{Z_t\}_{t=T-n+1}^{T} = \{Y_t\}_{t=T-n+1}^{T}$. The quantile regression minimizes the sample $MWAE$ of the residuals as its loss function while the expectile regression minimizes the sample $RMWSE$. Hence, the $MWAE$ should be the appropriate metric to use in measuring the accuracy of forecasts performed using the quantile regressions while $RMWSE$ should be used as the metric when forecasts are performed using expectile regressions.

Table 1 presents the sample $RMSE$ and $MWAE$ for the various DGPs when the quantile regressions and the arima are used to generate the forecast with the asymmetric lin-lin loss function. It can be seen from Table 1 that the smallest $MWAE$ (highlighted as the boxed numbers) of the quantile regression forecasts occurs at the weight w that matches the corresponding τ of the quantile regressions used in generating the forecasts in general with the exception of the near unit root DGP3, DGP4 and DGP5 where the smallest $MWAE$ occurs at $w = 0.3$ for $\tau = 0.4$. Hence, if MAE ($MWAE$ with $w = 0.5$) is used in assessing the accuracy of the forecasts, only the forecasts performed using the quantile regression with $\tau = 0.5$ which corresponds to the symmetric lin-lin loss with $\tau = 0.5$ will be deemed as being optimal. When the forecasters use any of the asymmetric loss function with $\tau \neq 0.5$, their forecast performance will be deemed as suboptimal using the MAE metric. However, if the $MWAE$ metric with the weight w that corresponds to the asymmetric parameter τ of the lin-lin loss function is used, instead, all the quantile regression forecasts performed using the different degrees of asymmetry determined by τ will be deemed as optimal. The smallest $RMSE$ occurs at $\tau = 0.5$ for the two stationary DGPs and at $\tau = 0.4$ for the three near unit root DGPs. Similarly, only the quantile regression forecasts with τ close to 0.5 are considered as optimal using the $RMSE$ metric. The forecasts performed using the arima model is comparable to those performed using the quantile regression with τ close to 0.5 when measured by $RMSE$.

Table 2 presents the sample $RMSE$ and $MWSE$ when the expectile regressions are used to generate the forecasts with the asymmetric quad-quad loss function. Again, the smallest $MWSE$ of the expectile regression forecasts occurs at the weight w that matches the corresponding ω used in generating the forecasts.

Results from both Table 1 and Table 2 support the recommendation that the metrics being used to measure forecasts accuracy should match the loss functions used in generating the forecasts.

However, one will need to know the weights of the asymmetric loss function used by a forecaster, which are typically unknown, in order to incorporate the correct weights, w, in the $MWAE$ or $MWSE$ metrics. Fortunately, this asymmetric parameter of the loss function can be recovered through the Mincer-Zarnowitz quantile/regression approach that will be introduced next.

3 Recovering the Loss Functions via the Mincer-Zarnowitz Prediction-Realization Analysis

In this section, we illustrate how the loss function that a forecaster has adopted can be recovered (backed-out) via the quantile regression or expectile regression that is set in the framework introduced in Mincer and Zarnowitz (1969). If the risk expressed in Equation (1) is differentiable, the first order condition for the minimization of the risk becomes

$$E_Y\left[\rho'\left(Y-g\left(Z,\theta\right)\right)|Z,\theta\right]=\int_y \rho'\left(y-g\left(z,\theta\right)\right)f_Y\left(y|z,\theta\right)dy=0. \tag{2}$$

(In the case of quantile regression, the loss function $\rho_\tau\left(u\right)$ is differentiable everywhere except at $u=0$.) As presented in Elliott and Timmermann (2008), the generalized forecast errors, $\rho^{'}\left(Y-g\left(Z,\theta\right)\right)$, should be unpredictable and follow a martingale difference sequence given all the information utilized to generate the forecasts. This leads to the following sample analog of the orthogonality condition for a one-period ahead forecast:

$$\frac{1}{K}\sum_{t=T}^{T+K-1}\rho'\left(y_{t+1}-g\left(z_{t+1},\theta\right)\right)=0 \tag{3}$$

In addition, $\rho'\left(y_{t+1}-g\left(z_{t+1},\theta\right)\right)$ should be uncorrelated with any information in the formation set, $\mathcal{F}_t$, used in the forecast at time t. Hence, another common orthogonality condition being adopted is

$$\frac{1}{K}\sum_{t=T}^{T+K-1}\rho'\left(y_{t+1}-g\left(z_{t+1},\theta\right)\right)v_t=0 \tag{4}$$

where v_t is any function of the data, $\{z_s\}_{s=1}^{t+1}$, available at time $t+1$.

It is well know from the orthogonality condition in Equation (2) that the optimal forecast when minimizing the risk in Equation (1) for the symmetric square loss function is the conditional mean function,

$g(z,\theta) = E(Y|Z=z,\theta)$, while the conditional median is the optimal forecast for the symmetric absolute loss function. Likewise, the conditional quantile function, $g(z,\theta,\tau) = F_Y^{-1}(\tau|Z=z,\theta) = Q_\tau(Y|Z=z,\theta)$, is the optimal forecast for the asymmetric lin-lin loss (see, e.g., Koenker, 2005, pp. 5-6 and Gneiting, 2011b) and the conditional expectile function, $g(z,\theta) = \mu_\omega(Y|Z=z,\theta)$, is the optimal forecast for the quad-quad loss function; see, e.g. Newey and Powell (1987, p. 823) and Gneiting (2011a). Hence, the sample orthogonality condition in Equation (3) yields the sample least-squares regressions for the symmetric square loss function, the sample least-absolute-deviation regressions for the symmetric absolute loss function, the quantile regressions for the lin-lin loss and the expectile regressions for the quad-quad loss function.

Mincer and Zarnowitz (1969) used the forecast, $\hat{y} = g(z,\theta)$, as v_t in the orthogonality condition in Equation (4) for the square-error loss and this boiled down to performing an ordinary least squared regression on the following linear model:

$$y_{t+1} = \alpha + \beta\hat{y}_{t+1} + \varepsilon_{t+1} \tag{5}$$

where ε_{t+1} was an error term satisfying $E(\varepsilon_{t+1}|z_t) = 0$. The unbiasedness and efficiency of the forecast were evaluated by testing the intercept and slope through the joint hypothesis,

$$H_0 : \alpha = 0 \cap \beta = 1 \tag{6}$$

Optimal forecast was characterized by the upholding (non-rejection) of H_0.

Since the conditional quantile is the optimal forecast for the lin-lin loss function, one can perform the τ quantile regression for the model in Equation (5) with $0 < \tau < 1$, and ε_{t+1} such that $F_{e_{t+1}}^{-1}(\tau) = 0$; the optimality of the forecast can then be evaluated using the Wald-type test on the joint hypothesis:

$$H_0 : \alpha(\tau) = 0 \cap \beta(\tau) = 1 \tag{7}$$

where $\alpha(\tau)$ and $\beta(\tau)$ are regression quantile estimates of the intercept and slope coefficients.

Similarly, for the quad-quad loss function, the optimality of the forecast can be tested through the similar joint hypothesis

$$H_0 : \alpha(\omega) = 0 \cap \beta(\omega) = 1 \tag{8}$$

using the ω-expectile regression with $0 < \omega < 1$. Assuming the loss function used by a forecaster belongs to one of the flexible lin-lin or quad-quad family, one can recover (back-out) the parameter of asymmetry by selecting the τ or ω which fails to reject the joint hypothesis in Equation (7) or Equation (8), respectively.

This is coined as the *MZ quantile/expectile regression approach to forecast optimality test* in this paper.

We first provide the asymptotics for the Mincer-Zarnovitz (MZ) regression for completeness. Then, the Mincer-Zarnovitz (MZ) quantile regression approach is presented in the following two subsections; similar results can be obtained for the Mincer-Zarnovitz expectile regression approach.

3.1 The Original Mincer-Zarnovitz Regression

The original Mincer-Zarnovitz Regression considers linear models and uses a quardratic loss function, i.e. the forecast decision is to choose $\rho(\cdot) = (\cdot)^2$ with $g(Z,\theta) = X'\theta$ that minimizes the risk [1]

$$E_{Y|X}\left[(Y - X'\theta)^2\right].$$

Given $\{Z_t\}_{t=1}^T = \{Y_t, X_t\}_{t=1}^T$, supposed that $\widehat{\theta}$ is estimated by

$$\widehat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} (Y_t - X_t'\theta)^2 \tag{9}$$

and the n forecasts of $\{Y_t\}_{t=T+1}^{T+n}$ are constructed as

$$\widehat{Y}_t = X_t'\widehat{\theta},\ t = T+1, \cdots, T+n. \tag{10}$$

The original Mincer-Zarnovitz regression approach involves performing an OLS regression of Y_{T+i} on $\widehat{Y}_{T+i}$ for $i = 1, \cdots, n$, i.e.

$$\left(\widehat{\alpha}, \widehat{\beta}\right) = \arg\min_{\alpha,\beta} \sum_{t=T+1}^{T+n} \left(Y_t - \alpha - \beta\widehat{Y}_t\right)^2.$$

Theorem 1 below gives the asymptotic result for the OLS based Mincer-Zarnovitz regression estimator based on a general forecastor $g(Z,\theta)$.

For convenience of asymptotic analysis, the following assumptions are made:

Assumption O1: The data $\{Y_t, X_t\}$ is generated by the model $Y_t = g(X_t,\theta) + \varepsilon_t$.

Assumption O2: The n forecasts of $\{Y_t\}_{t=T+1}^{T+n}$ are constructed using Equation (10) where $\widehat{\theta}$ is estimated by Equation (9).

Assumption O3. $\{X_t, \varepsilon_t\}$ is stationary β-mixing with mixing decay rate $\beta_t = O(b^{-t})$ for some $b > 1$. $\mathrm{E}\|Z_t\|^{2+\delta} < \infty$, for some $\delta > 0$. $\lim_{m\to\infty} \mathrm{Var}\left(m^{-1/2}\sum_{t=1}^m \varepsilon_t\right) = \sigma_\varepsilon^2 < \infty$

Assumption O4. Both n and T approach ∞, and $n/T \to 0$.

[1] The Z as defined in Section 2 consists of the historical realized values of the dependent and independent variables, $Z = \{Z_t\}_{t=1}^T = \{Y_t, X_t\}_{t=1}^T$. In the regression setting of Mincer-Zarnovitz, the values of the dependent variable Y_t in the data Z_t are used in the estimation of θ.

For notational convenience, the following are denoted:

$$\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \widehat{\gamma} = \begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix}, \gamma^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Denote

$$\Sigma_0 = \begin{bmatrix} 1 & \mathrm{E}g\left(X_t, \theta\right) \\ \mathrm{E}g\left(X_t, \theta\right) & \mathrm{E}g\left(X_t, \theta\right)^2 \end{bmatrix}$$

Theorem 1: Under Assumptions O1-O4, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\gamma} - \gamma\right) \Rightarrow N\left(0, \sigma_\varepsilon^2 \Sigma_0^{-1}\right).$$

Result of Theorem 1 facilitate the OLS based Mincer-Zarnovitz inference.

3.2 The Mincer-Zarnovitz Quantile Regression

Suppose the forecaster uses an asymmetric lin-lin loss function. More specifically, the forecast decision is to choose $g\left(Z, \theta\right)$ that minimizes the risk

$$E_Y\left[\rho_\tau\left(Y - g(X, \theta)\right)\right]$$

where ρ_τ is the lin-lin loss function as specified in Section 2 or better known as the "check" loss function in the quantile regression literature.

Given $\{Z_t\}_{t=1}^T = \{Y_t, X_t\}_{t=1}^T$, it is supposed that $\widehat{\theta} = \widehat{\theta}(\tau)$ is estimated by

$$\widehat{\theta}(\tau) = \arg\min_\theta \sum_{t=1}^T \rho_\tau\left(Y_t - g(X_t, \theta)\right) \tag{11}$$

and the n forecasts of $\{Y_t\}_{t=T+1}^{T+n}$ are constructed as

$$\widehat{Y}_t = g(X_t, \widehat{\theta}(\tau)),\ t = T+1, \cdots, T+n. \tag{12}$$

The MZ quantile regression approach involves performing a quantile regression of Y_{T+i} on $\widehat{Y}_{T+i}$ for $i = 1, \cdots, n$, i.e.

$$\left(\widehat{\alpha}(\tau), \widehat{\beta}\left(\tau\right)\right) = \arg\min_{\alpha,\beta} \sum_{t=T+1}^{T+n} \rho_\tau\left(Y_t - \alpha - \beta\widehat{Y}_t\right).$$

The Mincer-Zarnovitz type analysis on the hypothesis that the forecasting is performed using an asymmetric lin-lin loss function can then be tested based on the hypothesis stated in Equation (7). The theorem in the next section provides a theoretical foundation for the Wald-type test.

3.3 Asymptotics of The Mincer-Zarnovitz Quantile Regression Analysis

For convenience of asymptotic analysis, the following assumptions are made:

Assumption 1: Denote the conditional quantile function of Y_t, given existing information $\mathcal{F}_t$, by $Q_{Y_t}(\tau|\mathcal{F}_t)$, and assume that $Q_{Y_t}(\tau|\mathcal{F}_t) = Q_{Y_t}(\tau|X_t) = g(X_t, \theta(\tau))$.

Assumption 2: The n forecasts of $\{Y_t\}_{t=T+1}^{T+n}$ are constructed using Equation (12) where $\widehat{\theta}(\tau)$ is estimated by Equation (11).

Assumption 3. $\{Z_t\}$ is a stationary β-mixing with mixing decay rate $\beta_t = O(b^{-t})$ for some $b > 1$.

Assumption 4. Let $F_t(\cdot)$ and $f_t(\cdot)$ be the conditional distribution function and density function of Y_t given $\mathcal{F}_t$, respectively, and $F_t(\cdot)$ and $f_t(\cdot)$ are continuously differentiable while $0 < f_t(\cdot) < \infty$ on its support.

Assumption 5. Both n and T approach ∞, and $n/T \to 0$.

Assumptions 1 and 2 assume that the forecaster uses an asymmetric lin-lin loss function in the forecast, and the true conditional quantile function is in the class of functions under consideration. Under Assumption 1, $\Pr(Y_t \leq g(X_t, \theta(\tau))) = \tau$. Assumption 3 ensures that the appropriate LLN and CLT apply to the sample average. This is assumed for convenience of the asymptotic analysis and the identification of the weakest conditions is not attempted here. Assumption 5 assumes that the sample used for the estimation is larger than the number of periods for prediction. Under this assumption, the preliminary estimation error is of smaller order of magnitude than $n^{-1/2}$ and, thus, not affecting the limiting distribution of the MZ quantile regression estimator.

For notational convenience, the following are denoted:

$$\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \widehat{\gamma}(\tau) = \begin{bmatrix} \widehat{\alpha}(\tau) \\ \widehat{\beta}(\tau) \end{bmatrix}, \gamma(\tau) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$Z_t(\theta) = \begin{bmatrix} Z_{1t}(\theta) \\ Z_{2t}(\theta) \end{bmatrix} = \begin{bmatrix} 1 \\ g(X_t, \theta) \end{bmatrix}.$$

Then the corresponding MZ quantile regression can be written as

$$\widehat{\gamma}(\tau) = \arg\min_{\gamma} \sum_{t=T+1}^{T+n} \rho_\tau \left(Y_t - \gamma^\top Z_t(\widehat{\theta}(\tau)) \right). \tag{13}$$

Theorem 2: Under Assumptions 1-5, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\gamma}(\tau) - \gamma(\tau)\right) \Rightarrow N\left(0, \Omega_1^{-1}\Omega_0\Omega_1^{-1}\right),$$

where

$$\Omega_0 = E\begin{bmatrix} 1 & Q_{Y_t}(\tau|X_t) \\ Q_{Y_t}(\tau|X_t) & Q_{Y_t}(\tau|X_t)^2 \end{bmatrix},$$

and

$$\Omega_1 = E\begin{bmatrix} f_t(Q_{Y_t}(\tau|X_t)) & f_t(Q_{Y_t}(\tau|X_t))Q_{Y_t}(\tau|X_t) \\ f_t(Q_{Y_t}(\tau|X_t))Q_{Y_t}(\tau|X_t) & f_t(Q_{Y_t}(\tau|X_t))Q_{Y_t}(\tau|X_t)^2 \end{bmatrix}.$$

While Theorem 1 facilitates the hypothesis expressed in Equation (6), Theorem 2 allows testing of the hypothesis expressed in Equation (7).

4 Illustration of the Mincer-Zarnowitz Quantile/Expectile Regressions Optimality Test

To illustrate the MZ quantile/expectile regression approach, $n + K = 300$ observations were first generated using DGP1 of Section 2.1. The first $n = 100$ of these observations were used as the in-sample data which served as the basis for future forecasts constructed by Equation (12) using quantile regression estimates obtained through Equation (11) to simulate the lin-lin loss function adopted by a forecaster. The remaining $K = 200$ observations were used as the 200 realized observations $\{y_{t+1}\}_{t=1}^{K}$ in the hold–out sample for which the forecasts were targeting. Specifically, nine sets of $K = 200$ forecasts $\{\hat{y}_{t+1}\}_{t=1}^{K}$ were constructed using Equation (12) through nine rolling quantile regression fits of Equation (11) to the in-sample data, one for each of $\tau_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$, to simulate nine different lin-lin lost functions used by the forecaster, each parameterized by the asymmetry parameter τ_f. Hence, there were 200 pairs of predictions and realizations $\{y_{t+1}, \hat{y}_{t+1}\}_{t=1}^{K}$ for each of the nine $\tau_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$. To try to recover the actual τ_f used in one of the nine asymmetric lin-lin loss functions that generated the forecast,$\{\hat{y}_{t+1}\}_{t=1}^{K}$, nine MZ quantile regressions of Equation (5) with $\tau_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ were performed using each of the nine sets of predictions and realizations pairs $\{y_{t+1}, \hat{y}_{t+1}\}_{t=1}^{K}$. It was expected that the MZ quantile regression with τ_{MZ} that was equal to the corresponding τ_f in Equation (11) that generated the forecasts to be closest to the 45^o line and to result in not rejecting the null hypothesis in Equation (7).

Figure 1 shows the results of the nine $\tau_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ-quantile regression fits of Equation (5) for four selected $\tau_f \in \{0.2, 0.4, 0.6, 0.8\}$ quantile regressions of Equation (11) which were subsequently used in generating the forecasts in Equation (12) for DGP1. The red line represents the 45^o line, the green line is the ordinary least-squares regression line, and the blue line is the $\tau_{MZ} = \tau_f$ quantile regression line when the recovered τ_{MZ} matches exactly the τ_f that is used in generating the forecast while the grey lines are the τ_{MZ} quantile regression fits when $\tau_{MZ} \neq \tau_f$. Figure 2 shows similar results of expectile regressions for Equation (5) for $\omega_f \in \{0.2, 0.4, 0.6, 0.8\}$ to illustrate the case when the quad-quad loss function instead of the lin-lin loss function was used in generating the forecasts in Equation (12) for DGP1. Again the red line represents the 45^o line, the green line is the ordinary least-squares regression line, and the blue line is the $\omega_{MZ} = \omega_f$ expectile regression line while the grey lines are the ω_{MZ} expectile regression fits when $\omega_{MZ} \neq \omega_f$. Figure 3 and Figure 4 show the results for the quantile regressions and expectile regressions, respectively, for DGP4.

It can be seen that both the τ_{MZ}-quantile regressions and ω_{MZ}-expectile regressions of y_{t+1} on $\hat{y}_{t+1}$ track the 45^o line closely when $\tau_{MZ} = \tau_f$ or $\omega_{MZ} = \omega_f$ while the ordinary least-squares regression line deviates from the 45^o line in general except when $\tau_f = 0.5$ or $\omega_f = 0.5$ which are not shown in the figures.

5 Simulations of Mincer-Zarnowitz Quantile/Expectile Regression Optimality Test

To see how the MZ quantile/expectile regression approach performs when applied to the lin-lin and quad-quad loss function, simulations were performed for the various DGPs depicted in Section 2.1. The size of the in-sample data used in the first-stage estimation of the various $\tau_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ quantile regressions in Equation (11) was $n \in \{100, 200, 400\}$. These estimated τ_f-quantile regressions were subsequently used to perform forecasts for the $K = N - n \in \{100, 200, 400\}$ hold-out periods through Equation (12).

Each of the subsequent plots in Figure 5 to Figure 9 show the percentage of the $NMC = 100$ Monte Carlo replications in which the H_0 in (7) was not rejected for the various combinations of n, K, τ_{MZ}, τ_f, and DGP1 through DGP5 for the MZ quantile regression of Equation (5).

In general, the percentage of non-rejection is the highest when $\tau_{MZ} = \tau_f$ as one will expect. The sample size n used in the first-stage in-sample estimation does not have as significant an impact as the sample size K in the second hold-out stage on the percentage of "correct" guess of the weight, τ_f, used in the forecasts in the lin-lin loss function. This is to be expected for the consistency of the test in the second-stage relies on the sample size K in the second-stage. Results are qualitatively similar for the MZ expectile regression approach. Due to space constraint, the detailed results are not presented here but they are available from the authors upon request.

Hence, the MZ quantile/expectile approach that finds the τ_{MZ} (ω_{MZ}) quantile (expectile) regression for which H_0 in (7) ((8)) is not being rejected provides a reliable way to recover the weight, τ_f (ω_f), used in the lin-lin (quad-quad) loss function of the forecaster.

5.1 Robustness Against Model Misspecifications

To study how robust is the MZ quantile/expectile regression approach to forecast optimality test against model misspecifications in Equation (11), two simulations were performed based on two augmentations of the DGPs introduced in Section 2.1: (1) The DGPs were augmented with an additional linear term γx_t where $\gamma = 1$ and $x_t \sim \chi^2(10)$ was i.i.d. from a chi-square distribution with ten degrees of freedom and (2) the DGPs were augmented with an additional nonlinear term γx_t^2 where $\gamma = 1$ and $x_t \sim \chi^2(10)$.

The percentages of non-rejection of H_0 in (7) for both the quantile regression model in Equation (11) used in the forecasts without the augmented x_t (misspecified model) and for the model with the first augmented linear term γx_t (correctly specified model) for the augmented DGP1 are presented in Figure 14 and Figure 15 in Appendix 1, respectively while those for the augmented DGP3 are presented in Figure 16 and Figure 17, respectively. The results for the augmented DGP2, DGP4 and DGP5 are similar and, hence, are not presented in this paper. The results for the expectile regressions are qualitatively similar and are not reported here either.

For the second simulation where DGP1 was augmented with the additional non-linear term γx^2, the results are presented in Figure 18 and Figure 19, respectively, in Appendix 1 for the misspecified and correctly specified stage 1 quantile regressions while Figure 20 and Figure 21, respectively, in Appendix 1 present those for DGP3.

One can see that the MZ quantile/expectile regression approach is quite robust to misspecifications in the regression equation in terms of missing variables for the high percentages of non-rejection cluster around the diagonals when $\tau_{MZ} = \tau_f$ and $\omega_{MZ} = \omega_f$.

5.2 Robustness Against Misspecification in the Loss Functions

Even though the majority of the forecasts in applied research have been generated using the quantile regression rather the expectile regression as summarized in Section 1, one still has to assume the class of loss function that has been employed by a forecaster to recover (estimate) the parameter which controls the degree of asymmetric in the loss function employed by the forecaster. We have demonstrated how effective the MZ quantile/expectile regression approach is in recovering the asymmetric parameter when the class of loss function used is assumed to be known. However, there is still the question: “How robust is the

MZ quantile/expectile approach in recovering the asymmetric parameter when the loss function used by a forecaster is misspecified?"

To investigate this, we modified the simulations described in Section 5 in the following ways: (1) Forecasts were generated using the $\tau_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ quantile regressions for the DGPs specified in Section 2.1 but the asymmetric parameter τ_f was incorrectly estimated using the $\omega_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ expectile regressions of (5); (2) Forecasts were generated using the $\omega_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ expectile regressions but the asymmetric parameter ω_f was mistakenly estimated using the $\tau_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ quantile regressions. Again, due to space constraint, only the results for the first set of simulations for DGP1 and DGP3 are presented in Figure 22 to Figure 23 in Appendix 1 while those for the second set are presented in Figure 24 to Figure 25 also in Appendix 1. The results for the other DGPs are available upon request from the authors.

It is obvious that the misspecifications of the loss function introduce bias in the recovered asymmetric parameters for the assumed loss functions in both sets of simulations. For the first set of simulations when the quantile regressions were used to generate the forecasts but the ω_{MZ} expectile regressions were used to estimate the asymmetric parameters τ_f, the ω_{MZ} expectile regressions overestimated τ_f for $\tau_f < 0.5$ but underestimated for $\tau_f > 0.5$ for DGP1 and DGP3 as shown in Figure 22 and 23. On the other hand, for the second set of simulations when the ω_f expectile regressions were used to generate the forecasts but the τ_{MZ} quantile regressions were used to estimate the asymmetric parameters ω_f, the τ_{MZ} quantile regression underestimated ω_f for $\omega_f < 0.5$ but overestimated for $\omega_f > 0.5$ for DGP1 and DGP3 as shown in Figure 24 and Figure 25. Results are qualitatively similar for DGP2, DGP4 and DGP5.

There is yet another question: "How robust is the MZ quantile/expectile approach when the DGPs in Section 2.1 are characterized by distributions other than the Gaussian?"We did not perform simulations on this aspect. However, we conjecture that the MZ quantile approach will be robust for any distribution while the expectile approach should still be somewhat reliable as long as the distribution does not have overly heavy tails so that the first moment can be estimated consistently and efficiently based on what we have learned from robust statistics.

6 Applications

In this section, the MZ quantile/expectile regression approach was applied to try to recover the asymmetric parameter in the potentially asymmetric loss functions used in the Federal Reserves' forecasting of economic variables and the forecasting of an electronic component manufacturer's demand in its supply chain.

6.1 The Federal Reserve's Greenbook Forecast

The report "Current Economical and Financial Conditions: Summary and Outlook" prepared by the Board of Governors of the Federal Reserve System for the Federal Open Market Committee Meeting is viewed by many as the authoritative forecast of the important economic variables such as GDP, inflation rate, unemployment rate, etc. It is also known as the *Greenbook*. Many studies have been done using the *Greenbook* forecasts and found that the forecasts were not optimal. The question though is whether the *Greenbook* forecasts are really suboptimal or they are in fact optimal for an asymmetric loss function used by the Federal Reserves.

Similar to Patton and Timmermman (2007), the real GDP and the Federal Reserves' forecast from the first quarter of 1969 to the first quarter of 2000 were used here. Figure 10 presents the recovered asymmetric parameter τ used in the Federal Reserve's *Greenbook* forecasts using the quantile regression approach while Figure 11 shows the asymmetric parameter ω recovered using the expectile regression. The ordinary least squares regression rejects the null hypothesis in (6) and, hence, the *Greenbook*'s forecast is deemed not optimal if one assumes that the Federal Reserves actually used the square loss decision function. The MZ quantile regression approach fails to reject the null in (7) for $\tau \in \{.30, .35, .40, .45\}$ while the expectile regression fails to reject the null in (8) for $\omega \in \{.35, .40, .45\}$. Using the quantile regression estimates, the ratio of optimal loss on negative errors to positive errors, $\mathcal{L}(-e)/\mathcal{L}(e) = (1-\tau)/\tau$, is between 1.22 and 2.33 while the ratio derived from the expectile regression estimates falls between 1.22 and 1.86. Patton and Timmermman's (2007) average estimated ratio was 1.44 with a minimum of 0.52 and a maximum of 2.76 when the loss function did not depend on the forecasts, and the average was 3.48, 1.97, and 1.26 for the 0.25, 0.5, and 0.75 quantiles of the real GDP growth, respectively, when the loss function depended on the forecasts. So the general picture of the inference using the MZ quantile/expectile regression approach is consistent with what Patton and Timmermman (2007) found: the Federal Reserves appeared to value overprediction roughly 1.5 times more costly than underprediction.

6.2 An Electronic Component Manufacturer Supply Chain

Data of an electronic component manufacturer in the U.S. were available on the actual shipment (y_{t+h}) and one-month through twelve-month ahead ($h = 1, \cdots, 12$) forecasts ($\hat{y}_{t+h}$) for 106 stock-keeping units (SKU) in 19 product families (FAM) processed at 4 global business units (GBU) and 6 distribution hubs (HUB) across 3 regions (REG) over the globe. Only the combinations of SKU, FAM, GBU, HUB and REG with at least 30 usable observations were analyzed. There were altogether 90 such combinations. Among these combinations, the incomplete observations with zero actual shipment were dropped from the analysis even though there were positive forecasts.

The recovered asymmetric parameter of the lin-lin loss function and quad-quad loss function at the one-month through twelve-month forecast horizons using the MZ quantile/expectile approach are presented in the histograms in Figure 12 and Figure 13, respectively. One can see from both figures that the recovered asymmetric parameters for both loss functions have values scattered around $\tau = 0.8$ and $\omega = 0.8$ for all the different forecast horizons. The global median of the median τ at each of the twelve forecast horizons is 0.8 and so is the global median of the median ω across the various horizons. The global mean of the mean τ at the various forecast horizons is 0.78 while the global mean of the mean ω is 0.76. This suggests that the manufacturer weighs positive forecast errors (under forecasts) about four times ($\mathcal{L}(e)/\mathcal{L}(-e) = \tau/(1-\tau) = 0.8/0.2 = 4$) as costly as negative forecast errors (over forecasts). Under forecasts will lead to unfulfilled order, which could lead to loss of goodwill, reputations, and loss of current and future sales while over forecasts will result in over inventory, higher insurance costs, and tied up capital. In this case, the manufacturer views under forecasts more costly than over forecasts.

7 Conclusions

Forecasts are ubiquitous in all areas of daily life. Typical summary metrics used in measuring forecast performance are the sample root-mean-square error and the mean-absolute error. However, these popular summary metrics are appropriate only for the symmetric mean-square error and mean-absolute error loss functions used by the forecasters. This paper has argued and demonstrated that when forecasters use an asymmetric loss function such as the lin-lin or quad-quad lost functions, the appropriate summary metrics for measuring forecast performance should, instead, be the sample mean-weighted-absolute error and mean-weighted-square error, respectively, that reflect the different weights assigned to over and underprediction.

However, to correctly utilize the sample mean-weighted-absolute error and mean-weighted-square error, one will need to know the relative weights for over and underprediction that a forecaster assigned when performing the forecasts. These weights that characterize the asymmetric loss functions can be recovered in the recommended MZ quantile/expectile regression approach.

Regarding the task of evaluation of forecast optimality, theoretical justification is provided for extending Mincer and Zarnowitz (1969) prediction-realization framework that is based on the ordinary least-squares regression to one that uses the quantile regressions and expectile regressions. Simulation results demonstrating the efficacy of the proposed MZ quantile/expectile regression approach are provided and show that this approach is robust to specific forms of model misspecification in the data generating process. This approach is not robust to misspecification in the loss functions, especially for the near unit root data generating processes. However, Gneiting (2011a) argued that:

> effective point forecasting depends on ‘guidance’ or ‘directives,’ which can be given in one of two complementary ways, namely, by disclosing the *scoring function* ex ante to the forecaster, or by requesting a specific *functional* of the forecaster’s predictive distribution, such as the mean or a quantile.... An alternative to disclosing the scoring function is to request a specific functional of the forecaster’s predictive distribution, such as the mean or a quantile, and to apply any scoring function that is consistent with the functional, roughly in the following sense. (748-9)

Hence, if one is to evaluate the optimality of a forecast but has no information about the form of the loss function being used, one should request for such information and use a summary metric (scoring function) that is consistent with the elicited loss function instead of cavalierly using the traditional tests for forecast optimality that rely on the symmetric loss functions and concluding that the forecast is suboptimal.

The MZ quantile/expectile regression approach is then applied to Federal Reserve’s *Greenbook* forecast and the forecast of an electronic component manufacturer’s demand in its supply chain. It has been found that the Federal Reserve appeared to value overprediction roughly 1.5 times more costly than underprediction. It has also been found that the electronic component manufacturer appeared to weigh under forecast about four times as costly as over forecast.

Finally, we feel that even though we have introduced and demonstrated how both the quantile and expectile approahces work in recovering the unknown parameter that controls the potentially asymmetric loss function used in generating forecasts, we have the obligation to warn the readers about the potential issues that could be encountered when using the expectile approach. In this respect, no one has done a better job than Koenker (2013) in his discussion of the article ”Beyond mean regression” by Kneib (2013):

> Many aspects of the case against expectiles are familiar: they are slippery, although they seek to describe a local property of a distribution, they depend on global properties of that distribution; they are inherently nonrobust, by manipulating the tails of the distribution one can make the expectiles dance at your will; and they are not equivariant to monotone transformations as are the quantiles. I could rest my case here, but why?-when we are having fun. (327)

Advocating the values of expectile, Waltrup et al. (2015) argued that:

> Apparently, referring again to the comparison of expectiles and quantiles to David and Goliath is undissolved. There is no final fight, and research on both ends continues. It is certainly true that quantiles are dominant in the literature but we wanted to show that expectiles are an

interesting alternative to quantiles and that their combined use is helpful, in particular, for the estimation of the ES ....

All in all, we hope to have convinced the reader that expectiles do not immediately ¿belong in the spittoon¿ as Koenker (2013a) provocatively postulates. We think that expectiles provide an interesting and worthwhile alternative to the well-established quantile regression. (452-53)

Whether the expectile approach belongs in the spittoon, we will leave that decision to the readers.

Acknowledgement The research supporting the final revision of this paper is undertaken during Kemal Guler's visit at Bilkent University supported by a TUBITAK BIDEP 2236 Co-Circulation fellowship. He thanks TUBITAK for financial support, colleagues at Bilkent University Industrial Engineering Department for their hospitality, and Baris, Ali, Betul, Elfe and Sertug of Sundogs for their big hearts and warm Ankara memories.

8 References

Ali A, Klein A, Rosenfeld J. 1992. "Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS." *The Accounting Review* 67(1),183-198.

Artis M, Marcellino M. 2001. "Fiscal forecasting: the track record of the IMF, OECD and EC." *The Econometrics Journal* 4(1), 20–36.

Bastianin A, Galeottic M, Manera M. 2014. "Causality and predictability in distribution: the ethanol–food price relation revisited." *Energy Economics* 42, 152-160.

Batchelor R, Peel DA. 1998. "Rationality testing under asymmetric loss." *Economics Letters* 61(1), 49–54 .

Capistran CC. 2006. "Bias in Federal Reserve inflation forecasts: is the Federal Reserve irrational or just cautious?" Banco de Mexico Working Paper, no.2006-14.

Chambers CP. (2007). "Ordinal aggregation and quantiles." *Journal of Economic Theory* 137, 416–431.

Christoffersen P F, Diebold FX. 1997. "Optimal prediction under asymmetric loss." *Econometric Theory* 13(6), 808–17.

Cohen M, Ho T, Ren Z, Terwiesch C. 2003. "Measuring imputed cost in the semiconductor equipment supply chain." *Management Science* 49(12), 1653–1670.

Elliott G, Komunjer I, Timmermann A. 2005. "Estimation and testing of forecast rationality under flexible loss." *Review of Economic Studies* 72(4), 1107–25.

Elliott G, Timmermann A. 2008. "Economic forecasting." *Journal of Economic Literature* 46(1), 3-56.

Friederichs P, Hense A. 2007. "Statistical downscaling of extreme precipitation events using censored quantile regression." *Monthly Weather Review* 135, 2365–2378.

Ferguson TS. 1967. *Mathematical Statistics: A Decision-Theoretic Approach*. New York: Academic.

Gneiting T. 2011a. "Making and evaluating point forecasts." *Journal of the American Statistical Association* 106, 746-762.

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

Granger CWJ. 1969. "Prediction with a generalized cost of error function." *Operations Research* 20(2), 199–207.

Granger CWJ. 1999. "Outline of forecast theory using generalized cost functions." *Spanish Economic Review* 1(2), 161–73.

Granger CWJ, Newbold P. 1986. *Forecasting Economic Time Series*. Second ed., Orlando, Fla.; London; Sydney and Toronto: Harcourt, Brace, Jovanovich; Academic Press.

Granger CWJ, Pesaran MH. 2000. "Economic and statistical measures of forecast accuracy." *Journal of Forecasting* 19(7), 53 7–60.

Ito T. 1990. "Foreign exchange rate expectations: micro survey data." *American Economic Review* 80(3), 434-49.

Kneib T. 2013. "Beyong mean regression (with discussion and rejoinder). " *Statistical Modelling* 13(4), 273-303.

Koenker R. 2013. "Discussion of 'beyond mean regression' by T. Kneib. " *Statistical Modelling* 13(4), 323-33.

Koenker R, Bassett G. 1978. "Regression Quantiles." *Econometrica*, 46(1), 33-50.

Koenker R, Xiao Z. 2006. "Quantile autoregression." *Journal of the American Statistical Association* 101, 980-990.

Kokic P, Chambers R, Beare S. 2000. "Microsimulation of business performance." *International Statistical Review* 68(3), 259-275.

Lee H, Padmanabhan V, Whang S. 1997. "Information distortion in a supply chain: the bullwhip effect." *Management Science* 43(4), 546-588.

Lewis KF, Whiteman CH. 2015. Empirical Bayesian density forecasting in Iowa and shringkage for the monte carlo era. *Journal of Forecasting* 34, 15-35.

Manski CF. 1988. "Ordinal utility models of decision making under uncertainty." *Theory and Decision* 25 (1), 79-104.

Mincer J, Zarnowitz V. 1969. "The evaluation of economic forecasts." In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. Ed. J. Mincer. New York: National Bureau of Economic Research, 14–20.

Newey WK, Powell JL. 1987. "Asymmetric least squares estimation and testing." *Econometrica* 55, 819-847.

Nielsen HA, Madsen H, Nielsen TS. 2006. "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts." *Wind Energy* 9, 95-108.

Patton AJ, Timmermann A. 2007. "Properties of optimal forecasts under asymmetric loss and nonlinearity." *Journal of Econometrics* 140(2), 884–918.

Pesaran MH, Skouras S. 2002. "Decision-based methods for forecast evaluation." in *A Companion to Economic Forecasting*. Eds. M. Clements and D. Hendry. Malden, Mass. and Oxford: Blackwell, 241–67.

Pruitt SW, Gitman LJ. 1987. "Capital budgeting forecast biases: evidence from the Fortune 500." *Financial Management* 16(1), 46-51.

Raiffa H, Schlaifer R. 1961. *Applied Statistical Decision Theory*. Clinton: Colonial Press.

Rostek M. 2010. "Quantile maximization in decision theory." *Review of Economic Studies* 77, 339-71.

Silver N. 2012. *The Signal and the Noise: Why So Many Predictions Fail - but Some Don't*, Penguin Books.

Soyiri IN, Reidpath DD, Sarran C. 2013. "Forecasting peak asthma admissions in London: an application of quantile regression models." *International Journal of Biometeorology* 57(4), 569-78.

Soyiri IN, Reidpath DD. 2013. "The Use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000." *PLoS ONE* 8(10), e78215.

Taylor JW. 2007. "Forecasting daily supermarket sales using exponentially weighted quantile regression." *European Journal of Operational Research* 178, 154-167.

Varian HR. 1974. "A Bayesian approach to real estate assessment." in *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*. Eds. S. E. Fienberg and A. Zellner, Amsterdam: North-Holland, 195–208.

Voudouris V, Matsumoto K, Sedgwick J, Rigby R, Stasinopoulos D, Jefferson M. 2014. "Exploring the production of natural gas through the lenses of the ACEGES model." *Energy Policy* 64, 124-133.

Waltrup LS, Sobotka F, Kneib T, Gorän K. 2015. "Expectile and quatnile regression-David and Goliath? " *Statistical Modelling* 15(5), 433-456.

Weerts AH, Winsemius HC, Verkade JS. 2011. "Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales)." *Hydrology and Earth System Sciences* 15, 255-265.

Figure 1: A single realization of the MZ quantile regressions for DGP1. The red line represents the 45^o line, the green line is the ordinary least-squares regression line, and the blue line is the $\tau_{MZ} = \tau_f$ quantile regression line when the recovered τ_{MZ} matches exactly the τ_f that is used in generating the forecast while the grey lines are the τ_{MZ} quantile regression fits when $\tau_{MZ} \neq \tau_f$.

Figure 2: A single realization of the MZ expectile regressions for DGP1. The red line represents the 45^o line, the green line is the ordinary least-squares regression line, and the blue line is the $\omega_{MZ} = \omega_f$ expectile regression line while the grey lines are the ω_{MZ} expectile regression fits when $\omega_{MZ} \neq \omega_f$.

Weiss AA. 1996. "Estimating time series models using the relevant cost function." *Journal of Applied Econometrics* 11(5), 53 9–60.

West KD, Edison HJ, Cho D. 1993. "A utility-based comparison of some models of exchange rate volatility." *Journal of International Economics* 35 (1-2), 23–45 .

Yu HT. 2013. "A quantile regression model to forecast ICT development." *International Journal of Innovative Management, Information & Production* 4, 34-38.

Zellner A. 1986. "Bayesian estimation and prediction using asymmetric loss functions." *Journal of the American Statistical Association* 81(394), 44 6–51.

Figure 3: A single realization of the MZ quantile regressions for DGP4. The red line represents the 45^o line, the green line is the ordinary least-squares regression line, and the blue line is the $\tau_{MZ} = \tau_f$ quantile regression line when the recovered τ_{MZ} matches exactly the τ_f that is used in generating the forecast while the grey lines are the τ_{MZ} quantile regression fits when $\tau_{MZ} \neq \tau_f$.

Table 1: Sample $MWAE$ and $RMSE$ for the quantile regression and arima fits under the various DGPs. The boxed numbers correspond to the w that yields the smallest $MWAE$ for the various τ in the quantile regression fits.

DGPs		MWAE									RMSE
		w									
	τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
DGP1	0.1	$\boxed{1.16}$	1.92	2.67	3.43	4.18	4.93	5.69	6.44	7.20	4.93
	0.2	1.29	$\boxed{1.80}$	2.31	2.83	3.34	3.85	4.36	4.88	5.39	4.07
	0.3	1.60	1.90	$\boxed{2.20}$	2.50	2.79	3.09	3.39	3.69	3.98	3.45
	0.4	2.01	2.14	2.28	$\boxed{2.41}$	2.55	2.68	2.82	2.95	3.09	3.18
	0.5	2.58	2.55	2.51	2.48	$\boxed{2.45}$	2.42	2.39	2.36	2.32	3.12
	0.6	3.25	3.07	2.88	2.70	2.52	$\boxed{2.34}$	2.16	1.98	1.79	3.24
	0.7	4.07	3.74	3.41	3.08	2.75	2.42	$\boxed{2.09}$	1.76	1.43	3.50
	0.8	5.31	4.79	4.27	3.75	3.24	2.72	2.20	$\boxed{1.68}$	1.17	4.04
	0.9	6.96	6.22	5.48	4.75	4.01	3.27	2.53	1.79	$\boxed{1.05}$	4.84
	arima	2.49	2.49	2.48	2.48	2.47	2.47	2.46	2.46	2.45	3.13
DGP2	0.1	$\boxed{1.15}$	1.89	2.62	3.36	4.10	4.84	5.57	6.31	7.05	4.84
	0.2	1.27	$\boxed{1.76}$	2.26	2.75	3.25	3.74	4.23	4.73	5.22	3.97
	0.3	1.53	1.86	$\boxed{2.19}$	2.53	2.86	3.19	3.52	3.85	4.19	3.53
	0.4	1.93	2.10	2.26	$\boxed{2.42}$	2.58	2.75	2.91	3.07	3.24	3.20
	0.5	2.46	2.47	2.49	2.51	$\boxed{2.53}$	2.54	2.56	2.58	2.60	3.12
	0.6	3.17	3.03	2.89	2.75	2.61	$\boxed{2.47}$	2.33	2.19	2.05	3.22
	0.7	4.18	3.86	3.53	3.20	2.87	2.55	$\boxed{2.22}$	1.89	1.56	3.54
	0.8	5.50	4.97	4.43	3.89	3.36	2.82	2.28	$\boxed{1.75}$	1.21	4.13
	0.9	7.37	6.58	5.79	5.00	4.20	3.41	2.62	1.83	$\boxed{1.04}$	5.04
	arima	2.51	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	3.13
DGP3	0.1	$\boxed{1.15}$	1.88	2.61	3.34	4.08	4.81	5.54	6.27	7.00	4.85
	0.2	1.29	$\boxed{1.79}$	2.29	2.80	3.30	3.81	4.31	4.82	5.32	4.08
	0.3	1.61	1.90	$\boxed{2.20}$	2.49	2.78	3.07	3.36	3.65	3.95	3.49
	0.4	1.96	2.10	2.25	$\boxed{2.39}$	2.54	2.68	2.83	2.97	3.12	3.26
	0.5	2.46	2.46	2.46	2.45	$\boxed{2.45}$	2.45	2.45	2.44	2.44	3.17
	0.6	3.13	2.98	2.82	2.67	2.52	$\boxed{2.36}$	2.21	2.06	1.90	3.24
	0.7	3.92	3.62	3.32	3.02	2.73	2.43	$\boxed{2.13}$	1.83	1.53	3.48
	0.8	5.08	4.61	4.13	3.65	3.18	2.70	2.22	$\boxed{1.75}$	1.27	3.96
	0.9	7.37	6.59	5.81	5.03	4.25	3.47	2.69	1.90	$\boxed{1.12}$	5.06
	arima	2.46	2.46	2.46	2.46	2.46	2.47	2.47	2.47	2.47	3.18
DGP4	0.1	$\boxed{1.43}$	2.44	3.45	4.46	5.47	6.47	7.48	8.49	9.50	6.46
	0.2	1.61	$\boxed{2.25}$	2.89	3.53	4.17	4.81	5.45	6.09	6.73	5.06
	0.3	2.06	2.43	$\boxed{2.80}$	3.16	3.53	3.90	4.27	4.64	5.00	4.29
	0.4	2.75	2.87	2.99	$\boxed{3.12}$	3.24	3.36	3.48	3.61	3.73	4.05
	0.5	3.57	3.46	3.36	3.26	$\boxed{3.15}$	3.05	2.94	2.84	2.73	4.03
	0.6	4.23	3.97	3.70	3.43	3.16	$\boxed{2.90}$	2.63	2.36	2.10	4.14
	0.7	5.12	4.68	4.25	3.82	3.39	2.96	$\boxed{2.52}$	2.09	1.66	4.48
	0.8	6.32	5.70	5.08	4.46	3.84	3.22	2.60	$\boxed{1.98}$	1.36	5.00
	0.9	8.32	7.44	6.55	5.67	4.78	3.90	3.01	2.13	$\boxed{1.24}$	5.91
	arima	3.16	3.16	3.16	3.15	3.15	3.15	3.15	3.14	3.14	3.93
DGP5	0.1	$\boxed{0.78}$	1.33	1.87	2.42	2.97	3.51	4.06	4.61	5.15	3.73
	0.2	0.83	$\boxed{1.24}$	1.66	2.07	2.49	2.91	3.32	3.74	4.15	3.25
	0.3	0.99	1.30	$\boxed{1.61}$	1.92	2.24	2.55	2.86	3.17	3.49	2.96
	0.4	1.25	1.45	1.64	$\boxed{1.84}$	2.04	2.24	2.44	2.64	2.83	2.68
	0.5	1.59	1.68	1.78	1.87	$\boxed{1.97}$	2.06	2.16	2.25	2.35	2.53
	0.6	2.21	2.16	2.11	2.07	2.02	$\boxed{1.97}$	1.93	1.88	1.84	2.52
	0.7	2.96	2.77	2.58	2.39	2.20	2.01	$\boxed{1.82}$	1.63	1.44	2.68
	0.8	4.13	3.76	3.38	3.01	2.64	2.26	1.89	$\boxed{1.51}$	1.14	3.14
	0.9	6.22	5.56	4.91	4.25	3.60	2.94	2.28	1.63	$\boxed{0.97}$	4.16
	arima	1.98	1.99	2.00	2.01	2.02	2.03	2.04	2.05	2.06	2.54

Table 2: Sample $MWSE$ and $RMSE$ for the expectile regression and arima fits under the various DGPs. The boxed numbers correspond to the w that yields the smallest $MWSE$ for the various ω in th expectile regression fits.

DGPs		MWSE									RMSE
		w									
	ω	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
DGP1	0.1	$\boxed{5.33}$	8.10	10.87	13.64	16.42	19.19	21.96	24.73	27.50	4.05
	0.2	5.97	$\boxed{7.55}$	9.13	10.71	12.29	13.87	15.45	17.03	18.61	3.51
	0.3	7.07	7.94	$\boxed{8.81}$	9.68	10.55	11.42	12.29	13.16	14.03	3.25
	0.4	8.47	8.79	9.12	$\boxed{9.44}$	9.76	10.09	10.41	10.73	11.05	3.12
	0.5	10.18	10.02	9.87	9.71	$\boxed{9.55}$	9.40	9.24	9.08	8.93	3.09
	0.6	12.31	11.69	11.06	10.44	9.81	$\boxed{9.19}$	8.56	7.94	7.31	3.13
	0.7	15.14	14.00	12.87	11.74	10.60	9.47	$\boxed{8.34}$	7.20	6.07	3.26
	0.8	19.34	17.57	15.79	14.02	12.24	10.46	8.69	$\boxed{6.91}$	5.14	3.50
	0.9	27.14	24.33	21.52	18.71	15.90	13.09	10.28	7.47	$\boxed{4.66}$	3.99
	arima	10.48	10.31	10.13	9.95	9.77	9.60	9.42	9.24	9.06	3.13
DGP2	0.1	$\boxed{5.16}$	7.96	10.76	13.56	16.35	19.15	21.95	24.75	27.55	4.04
	0.2	5.76	$\boxed{7.42}$	9.08	10.74	12.39	14.05	15.71	17.36	19.02	3.52
	0.3	6.83	7.80	$\boxed{8.77}$	9.74	10.71	11.67	12.64	13.61	14.58	3.27
	0.4	8.19	8.62	9.05	$\boxed{9.48}$	9.91	10.34	10.77	11.19	11.62	3.15
	0.5	9.90	9.84	9.78	9.72	$\boxed{9.67}$	9.61	9.55	9.49	9.43	3.11
	0.6	12.10	11.55	11.00	10.45	9.90	$\boxed{9.35}$	8.80	8.25	7.70	3.15
	0.7	15.10	14.00	12.90	11.80	10.70	9.60	$\boxed{8.50}$	7.40	6.29	3.27
	0.8	19.57	17.77	15.98	14.18	12.39	10.59	8.80	$\boxed{7.00}$	5.21	3.52
	0.9	27.76	24.87	21.97	19.08	16.18	13.29	10.39	7.50	$\boxed{4.60}$	4.02
	arima	10.06	10.00	9.93	9.87	9.80	9.74	9.67	9.61	9.54	3.13
DGP3	0.1	$\boxed{5.45}$	8.29	11.12	13.96	16.79	19.63	22.46	25.30	28.13	4.10
	0.2	6.04	$\boxed{7.70}$	9.36	11.02	12.68	14.34	16.00	17.66	19.32	3.56
	0.3	7.10	8.06	$\boxed{9.01}$	9.96	10.91	11.87	12.82	13.77	14.73	3.30
	0.4	8.43	8.85	9.27	$\boxed{9.68}$	10.10	10.52	10.93	11.35	11.76	3.18
	0.5	10.08	10.02	9.96	9.90	$\boxed{9.85}$	9.79	9.73	9.67	9.61	3.14
	0.6	12.19	11.66	11.13	10.60	10.06	$\boxed{9.53}$	9.00	8.47	7.93	3.17
	0.7	15.05	14.00	12.94	11.88	10.83	9.77	$\boxed{8.72}$	7.66	6.61	3.29
	0.8	19.41	17.68	15.95	14.23	12.50	10.77	9.04	$\boxed{7.31}$	5.58	3.54
	0.9	28.08	25.19	22.31	19.43	16.55	13.67	10.79	7.90	$\boxed{5.02}$	4.07
	arima	10.27	10.23	10.19	10.16	10.12	10.08	10.04	10.01	9.97	3.18
DGP4	0.1	$\boxed{8.25}$	12.88	17.51	22.13	26.76	31.39	36.02	40.65	45.28	5.17
	0.2	9.26	$\boxed{11.86}$	14.45	17.05	19.64	22.24	24.84	27.43	30.03	4.43
	0.3	11.13	12.52	$\boxed{13.91}$	15.29	16.68	18.07	19.45	20.84	22.23	4.08
	0.4	13.53	13.98	14.44	$\boxed{14.89}$	15.35	15.80	16.26	16.72	17.17	3.92
	0.5	16.42	16.06	15.70	15.34	$\boxed{14.98}$	14.62	14.26	13.90	13.54	3.87
	0.6	19.87	18.74	17.62	16.49	15.37	$\boxed{14.24}$	13.11	11.99	10.86	3.92
	0.7	24.14	22.24	20.33	18.42	16.51	14.61	$\boxed{12.70}$	10.79	8.89	4.06
	0.8	30.08	27.25	24.43	21.60	18.77	15.95	13.12	$\boxed{10.30}$	7.47	4.33
	0.9	40.74	36.49	32.25	28.00	23.75	19.50	15.25	11.00	$\boxed{6.76}$	4.87
	arima	16.88	16.52	16.16	15.79	15.43	15.07	14.71	14.35	13.99	3.93
DGP5	0.1	$\boxed{2.85}$	4.68	6.50	8.33	10.16	11.98	13.81	15.64	17.46	3.19
	0.2	3.04	$\boxed{4.30}$	5.55	6.81	8.06	9.32	10.57	11.83	13.08	2.84
	0.3	3.53	4.41	$\boxed{5.28}$	6.16	7.03	7.91	8.78	9.66	10.54	2.65
	0.4	4.25	4.80	5.36	$\boxed{5.92}$	6.48	7.04	7.59	8.15	8.71	2.55
	0.5	5.22	5.48	5.73	5.99	$\boxed{6.24}$	6.50	6.75	7.01	7.26	2.50
	0.6	6.57	6.50	6.44	6.37	6.31	$\boxed{6.24}$	6.18	6.11	6.05	2.51
	0.7	8.50	8.06	7.63	7.19	6.76	6.33	$\boxed{5.89}$	5.46	5.02	2.60
	0.8	11.61	10.68	9.75	8.81	7.88	6.95	6.02	$\boxed{5.09}$	4.16	2.81
	0.9	17.85	16.07	14.29	12.50	10.72	8.94	7.16	5.38	$\boxed{3.60}$	3.27
	arima	5.63	5.83	6.04	6.24	6.44	6.65	6.85	7.06	7.26	2.54

Figure 4: A single realization of the MZ expectile regressions for DGP4. The red line represents the 45^o line, the green line is the ordinary least-squares regression line, and the blue line is the $\omega_{MZ} = \omega_f$ expectile regression line while the grey lines are the ω_{MZ} expectile regression fits when $\omega_{MZ} \neq \omega_f$.

Figure 5: Percentage of the total number of replications (NMC) that fails to reject H_0 for DGP1. The in-sample data sizes are $n \in [100, 200, 400]$ while the hold-out sample sizes are $K \in [100, 200, 400]$. The quantile regressions for Equation (11) are performed for $\tau_f \in [0.1, 0.2, \cdots, 0.8, 0.9]$ which are subsequently used to perform forecast using Equation (12) while the MZ quantile regressions for Equation (5) are performed for $\tau_{MZ} \in [0.1, 0.2, \cdots, 0.8, 0.9]$.

Figure 6: Percentage of the total number of replications (NMC) that fails to reject H_0 for DGP2. The in-sample data sizes are $n \in [100, 200, 400]$ while the hold-out sample sizes are $K \in [100, 200, 400]$. The quantile regressions for Equation (11) are performed for $\tau_f \in [0.1, 0.2, \cdots, 0.8, 0.9]$ which are subsequently used to perform forecasts using Equation (12) while the MZ quantile regressions for Equation (5) are performed for $\tau_{MZ} \in [0.1, 0.2, \cdots, 0.8, 0.9]$.

Figure 7: Percentage of the total number of replications (NMC) that fails to reject H_0 for DGP3. The in-sample data sizes are $n \in [100, 200, 400]$ while the hold-out sample sizes are $K \in [100, 200, 400]$. The quantile regressions for Equation (11) are performed for $\tau_f \in [0.1, 0.2, \cdots, 0.8, 0.9]$ which are subsequently used to perform forecasts using Equation (12) while the MZ quantile regressions for Equation (5) are performed for $\tau_{MZ} \in [0.1, 0.2, \cdots, 0.8, 0.9]$.

Figure 8: Percentage of the total number of replications (NMC) that fails to reject H_0 for DGP4. The in-sample data sizes are $n \in [100, 200, 400]$ while the hold-out sample sizes are $K \in [100, 200, 400]$. The quantile regressions for Equation (11) are performed for $\tau_f \in [0.1, 0.2, \cdots, 0.8, 0.9]$ which are subsequently used to perform forecasts using Equation (12) while the MZ quantile regressions for Equation (5) are performed for $\tau_{MZ} \in [0.1, 0.2, \cdots, 0.8, 0.9]$.

Figure 9: Percentage of the total number of replications (NMC) that fails to reject H_0 for DGP5. The in-sample data sizes are $n \in [100, 200, 400]$ while the hold-out sample sizes are $K \in [100, 200, 400]$. The quantile regressions for Equation (11) are performed for $\tau_f \in [0.1, 0.2, \cdots, 0.8, 0.9]$ which are subsequently used to perform forecasts using Equation (12) while the MZ quantile regressions for Equation (5) are performed for $\tau_{MZ} \in [0.1, 0.2, \cdots, 0.8, 0.9]$.

Figure 10: Mincer-Zarnowitz quantile regression estimates of the asymmetric parameter used in the Federal Reserve's Greenbook forecast.

Figure 11: Mincer-Zarnowitz expectile regression estimates of the asymmetric parameter used in the Federal Reserve's Greenbook forecast.

Figure 12: Histogram of the asymmetric parameter of the lin-lin loss function recovered by the MZ quantile regression approach.

Figure 13: Histogram of recovered asymmetric parameter of the quad-quad loss function recovered by the MZ expectile regression approach.

Appendix 1

Figure 14: Percentage of non-rejections of H_0 for the misspecified quantile regressions used in the forecast in Equation (11) for DGP1 augmented with an linear term γx_t.

Figure 15: Percentage of non-rejections of H_0 for the correctly specified quantile regressions used in the forecast in Equation (11) for DGP1 augmented with an linear term γx_t.

Figure 16: Percentage of non-rejections of H_0 for the incorrectly specified quantile regressions used in the forecast in Equation (11) for DGP3 augmented with an linear term γx_t.

Figure 17: Percentage of non-rejections of H_0 for the correctly specified quantile regressions used in the forecast in Equation (11) for DGP3 augmented with an linear term γx_t.

Figure 18: Percentage of non-rejections of H_0 for the misspecified quantile regressions used in the forecast in Equation (11) for DGP1 augmented with a nonlinear term γx_t^2.

Figure 19: Percentage of non-rejections of H_0 for the correctly specified quantile regressions used in the forecast in Equation (11) for DGP1 augmented with a nonlinear term γx_t^2.

Figure 20: Percentage of non-rejections of H_0 for the misspecified quantile regressions used in the forecast in Equation (11) for DGP3 augmented with a nonlinear term γx_t^2.

Figure 21: Percentage of non-rejections of H_0 for the correctly specified quantile regressions used in the forecast in Equation (11) for DGP3 augmented with a nonlinear term γx_t^2.

Figure 22: Percentage of non-rejections of H_0 when the forecasts were generated using the $\tau_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ quantile regressions for DGP1 while the asymmetric parameter τ_f was estimated using the $\omega_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ expectile regressions.

Figure 23: Percentage of non-rejections of H_0 when the forecasts were generated using the $\tau_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ quantile regressions for DGP3 while the asymmetric parameter τ_f was estimated using the $\omega_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ expectile regressions.

Figure 24: Percentage of non-rejections of H_0 when the forecasts were generated using the $\omega_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ expectile regressions for DGP1 while the asymmetric parameter ω_f was estimated using the $\tau_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ quantile regressions.

Figure 25: Percentage of non-rejections of H_0 when the forecasts were generated using the $\omega_f \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ expectile regressions for DGP3 while the asymmetric parameter ω_f was estimated using the $\tau_{MZ} \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ MZ quantile regressions.

Appendix 2

Proof of Theorem 2

Denote

$$G_n(\gamma,\theta) = \frac{1}{n}\sum_{t=T+1}^{T+n} \psi_\tau(Y_t - \gamma^\top Z_t(\theta)) Z_t(\theta),$$

where $\psi_\tau(u) = \tau - I(u < 0)$, and let

$$G(\gamma,\theta) = \mathrm{E}\left[\psi_\tau(Y_t - \gamma^\top Z_t(\theta)) Z_t(\theta)\right].$$

By the Law of Iterated Expectations

$$G(\gamma,\theta) = \mathrm{E}\left[\left\{\tau - F_t(\gamma^\top Z_t(\theta))\right\} Z_t(\theta)\right]$$

where $F_t(\cdot)$ and $f_t(\cdot)$ are the conditional distribution function and density function of Y_t given $\mathcal{F}_t$.

Under our conditions, the asymptotic behavior of the MZ quantile regression estimator $\widehat{\gamma}(\tau)$ is the same as that of $\arg\min_\gamma \|G_n(\gamma, \widehat{\theta}(\tau))\|$, and $\gamma(\tau)$ solves $\min_\gamma \|G(\gamma, \theta(\tau))\|$.

We first establish $\sqrt{n}$-consistency of $\widehat{\gamma}(\tau)$ to $\gamma(\tau)$. Let

$$\Gamma_1(\gamma,\theta) = \frac{\partial G(\gamma,\theta)}{\partial \gamma} = -E\left[f_t(\gamma^\top Z_t(\theta)) Z_t(\theta) Z_t(\theta)^\top\right],$$

and

$$\Gamma_{10} = \Gamma_1(\gamma,\theta)|_{\gamma=\gamma(\tau),\theta=\theta(\tau)}.$$

Under our regularity assumptions, $\Gamma_1(\gamma,\theta)$ is continuous at $\gamma = \gamma(\tau)$ and Γ_{10} is nonsingular; thus, there exists a constant $C > 0$ such that $C\|\widehat{\gamma}(\tau) - \gamma(\tau)\|$ is bounded by $\|G(\widehat{\gamma}(\tau), \theta(\tau))\|$ with probability going to 1.

Define

$$\Gamma_2(\gamma,\theta) = \frac{\partial G(\gamma,\theta)}{\partial \theta^\top}.$$

Notice that $\|G(\gamma(\tau), \theta(\tau))\| = 0$ and $\|G_n(\gamma(\tau), \theta(\tau))\| = O_p(n^{-1/2})$. By the triangle inequality we have

$$
\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \theta(\tau))\| \\
&\leq \|G(\widehat{\gamma}(\tau), \theta(\tau)) - G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| && (14)\\
&+ \|G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau)) - G(\gamma(\tau), \theta(\tau)) - G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau)) + G_n(\gamma(\tau), \theta(\tau))\| && (15)\\
&+ \|G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| && (16)\\
&+ O_p(n^{-1/2}).
\end{aligned}
$$

We now analyze the terms in Equations (14), (15) and (16) sequentially. First, for Equation (14), again, by triangle inequality

$$
\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \theta(\tau)) - G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| \\
&\leq \|G(\widehat{\gamma}(\tau), \theta(\tau)) - G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau)) - \Gamma_2(\widehat{\gamma}(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\| \\
&+ \|\Gamma_2(\widehat{\gamma}(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau)) - \Gamma_2(\gamma(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\| \\
&+ \|\Gamma_2(\gamma(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\|.
\end{aligned}
$$

Under the regularity assumptions, we have

$$
\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \theta(\tau)) - G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau)) - \Gamma_2(\widehat{\gamma}(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\| \\
&= O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|^2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\|\Gamma_2(\widehat{\gamma}(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau)) - \Gamma_2(\gamma(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\| \\
&= O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|\|\widehat{\gamma}(\tau) - \gamma(\tau)\|\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \theta(\tau)) - G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| \\
&\leq O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|^2\right) + O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|\|\widehat{\gamma}(\tau) - \gamma(\tau)\|\right) && (17)\\
&+ \|\Gamma_2(\gamma(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\|. && (18)
\end{aligned}
$$

In addition,

$$\Gamma_2(\gamma(\tau), \theta(\tau)) = -E\left[f_t(\gamma(\tau)^\top Z_t(\theta(\tau)))Z_t(\theta(\tau))\frac{\partial g(X_t, \theta(\tau))}{\partial\theta}\right]$$

Thus,

$$\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \theta(\tau)) - G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| \\
&\leq \|\Gamma_2(\gamma(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\|(1 + o_p(1)) \\
&= o_p(1)
\end{aligned}$$

and, as a result,

$$\|G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| \leq \|G(\widehat{\gamma}(\tau), \theta(\tau))\|(1 + o_p(1)).$$

For Equation (15), we need to verify stochastic equicontinuity. Using the fact that $I(Y_t < \cdot)$ is a monotonic function, and under our smoothness assumption on $F_t(\cdot)$ and the moment condition on Y, we have

$$\sup_{\|\gamma-\gamma(\tau)\|\leq\delta, \|\theta-\theta(\tau)\|\leq\delta} \frac{\sqrt{n}\|G_n(\gamma,\theta) - G(\gamma,\theta) - G_n(\gamma(\tau), \theta(\tau)) + G(\gamma(\tau), \theta(\tau))\|}{1 + \sqrt{n}\{\|G_n(\gamma,\theta)\| + \|G(\gamma,\theta)\|\}} = o_p(1),$$

by Lemma 4.2 of Chen (2006), and, consequently,

$$\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau)) - G(\gamma(\tau), \theta(\tau)) - G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau)) + G_n(\gamma(\tau), \theta(\tau))\| \\
&\leq o_p(1) \times \left\{\|G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| + \|G(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\|\right\} \\
&\leq o_p(1) \times \left\{\|G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| + \|G(\widehat{\gamma}(\tau), \theta(\tau))\|(1 + o_p(1)\right\}
\end{aligned}$$

where the last inequality comes from (18). Thus,

$$\begin{aligned}
&\|G(\widehat{\gamma}(\tau), \theta(\tau))\| \\
&\leq \|\Gamma_2(\gamma(\tau), \theta(\tau))(\widehat{\theta}(\tau) - \theta(\tau))\| + O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|^2\right) + O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|\|\widehat{\gamma}(\tau) - \gamma(\tau)\|\right) \\
&+ o_p(1) \times \left\{\|G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\| + \|G(\widehat{\gamma}(\tau), \theta(\tau))\|(1 + o_p(1)\right\} \\
&+ \|G_n(\widehat{\gamma}(\tau), \widehat{\theta}(\tau))\|,
\end{aligned}$$

and

$$\begin{aligned}\|G(\widehat{\gamma}(\tau),\theta(\tau))\|(1-o_p(1)) &\le \|G_n(\widehat{\gamma}(\tau),\widehat{\theta}(\tau))\|(1+o_p(1))+O_p(n^{-1/2})\\ &= \inf_{\gamma}\|G_n(\gamma,\widehat{\theta}(\tau))\|+O_p(n^{-1/2}).\end{aligned}$$

We only need to show that

$$\inf_{\gamma}\|G_n(\gamma,\widehat{\theta}(\tau))\|=O_p(n^{-1/2}),$$

which is true since

$$\begin{aligned}\|G_n(\gamma,\widehat{\theta}(\tau))\| &\le \|G_n(\gamma,\widehat{\theta}(\tau))-G(\gamma,\widehat{\theta}(\tau))-G_n(\gamma(\tau),\theta(\tau))\|\\ &\quad+\|G(\gamma,\widehat{\theta}(\tau))-G(\gamma,\theta(\tau))\|+\|G(\gamma,\theta(\tau))\|+\|G_n(\gamma(\tau),\theta(\tau))\|\\ &\le o_p(1)\times\left\{\|G_n(\gamma,\widehat{\theta}(\tau))\|+\|G(\gamma,\widehat{\theta}(\tau))\|\right\}+\|G(\gamma,\theta(\tau))\|+O_p(n^{-1/2}).\end{aligned}$$

Thus,

$$\|G_n(\gamma,\widehat{\theta}(\tau))\|(1-o_p(1))\le o_p(1)\times\left\{\|G(\gamma,\widehat{\theta}(\tau))\|\right\}+\|G(\gamma,\theta(\tau))\|+O_p(n^{-1/2}),$$

and

$$\inf_{\gamma}\|G_n(\gamma,\widehat{\theta}(\tau))\|=O_p(n^{-1/2}),$$

since $\|G(\gamma(\tau),\theta(\tau))\|=0$ and

$$\|G(\gamma,\widehat{\theta}(\tau))\|\le\|G(\gamma,\theta(\tau))\|+\|\Gamma_2(\gamma(\tau),\theta(\tau))(\widehat{\theta}(\tau)-\theta(\tau))\|(1+o_p(1)).$$

And, consequently,

$$C\|\widehat{\gamma}(\tau)-\gamma(\tau)\|\le\|G(\widehat{\gamma}(\tau),\theta(\tau))\|=O_p(n^{-1/2}).$$

Next we show asymptotic normality. Define the linearization

$$L_n(\gamma,\widehat{\theta}(\tau))=G_n(\gamma(\tau),\theta(\tau))+\Gamma_1(\gamma-\gamma(\tau))+\Gamma_2(\gamma(\tau),\theta(\tau))(\widehat{\theta}(\tau)-\theta(\tau)),$$

and note that

$$
\begin{aligned}
G_n(\gamma,\widehat{\theta}(\tau)) &= G_n(\gamma\,(\tau)\,,\theta\,(\tau)) + \Gamma_1(\gamma-\gamma\,(\tau)) + \Gamma_2(\gamma\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau)) \\
&\quad + G(\gamma,\theta\,(\tau)) - G(\gamma\,(\tau)\,,\theta\,(\tau)) - \Gamma_1(\gamma-\gamma\,(\tau)) \\
&\quad + \Gamma_2(\gamma,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau)) - \Gamma_2(\gamma\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau)) \\
&\quad + G(\gamma,\widehat{\theta}(\tau)) - G(\gamma,\theta\,(\tau)) - \Gamma_2(\gamma,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau)) \\
&\quad + G_n(\gamma,\widehat{\theta}(\tau)) - G(\gamma,\widehat{\theta}(\tau)) - G_n(\gamma\,(\tau)\,,\theta\,(\tau)) + G(\gamma\,(\tau)\,,\theta\,(\tau)) \\
&\quad - G(\gamma\,(\tau)\,,\theta\,(\tau)).
\end{aligned}
$$

Under our assumptions,

$$
\begin{aligned}
&\|G_n(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - L_n(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau))\| \\
&\leq \|G(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau)) - G(\gamma\,(\tau)\,,\theta\,(\tau)) - \Gamma_1(\widehat{\gamma}\,(\tau)-\gamma\,(\tau))\| \\
&+ \|\Gamma_2(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau)) - \Gamma_2(\gamma\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau))\| \\
&+ \|G(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - G(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau)) - \Gamma_2(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau))\| \\
&+ \|G_n(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - G(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - G_n(\gamma\,(\tau)\,,\theta\,(\tau)) + G(\gamma\,(\tau)\,,\theta\,(\tau))\| \\
&+ \|G(\gamma\,(\tau)\,,\theta\,(\tau))\| \\
&= o_p(n^{-1/2}),
\end{aligned}
$$

because

$$
\|G(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau)) - G(\gamma\,(\tau)\,,\theta\,(\tau)) - \Gamma_1(\widehat{\gamma}\,(\tau)-\gamma\,(\tau))\| = O_p(\|\widehat{\gamma}\,(\tau)-\gamma\,(\tau)\,\|^2) = o_p(n^{-1/2}),
$$

$$
\|\Gamma_2(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau)) - \Gamma_2(\gamma\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau))\| = o_p(1)\|\widehat{\gamma}\,(\tau)-\gamma\,(\tau)\,\| = o_p(n^{-1/2}),
$$

by root-n consistency;

$$
\begin{aligned}
\|G(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - G(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau)) - \Gamma_2(\widehat{\gamma}\,(\tau)\,,\theta\,(\tau))(\widehat{\theta}(\tau)-\theta\,(\tau))\| &\leq C\left(\|\widehat{\theta}(\tau)-\theta\,(\tau)\,\|^2\right) \\
&= o_p(n^{-1/2}), \text{ since } T > n
\end{aligned}
$$

$$
\|G_n(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - G(\widehat{\gamma}\,(\tau)\,,\widehat{\theta}(\tau)) - G_n(\gamma\,(\tau)\,,\theta\,(\tau)) + G(\gamma\,(\tau)\,,\theta\,(\tau))\| = o_p(n^{-1/2}),
$$

by stochastic equicontinuity, and

$$\|G(\gamma\left(\tau\right),\theta\left(\tau\right))\| = o_p(n^{-1/2}),$$

by definition. Thus

$$\min_{\gamma} \|G_n(\gamma, \widehat{\theta}(\tau))\| = \min_{\gamma} \|L_n(\gamma, \widehat{\theta}(\tau))\| + o_p(n^{-1/2}), \tag{19}$$

and

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\gamma}(\tau) - \gamma(\tau)\right) \\
&= -\left(\Gamma_1^\top \Gamma_1\right)^{-1} \Gamma_1^\top \sqrt{n}\left[G_n(\gamma\left(\tau\right),\theta\left(\tau\right)) + \Gamma_2(\gamma\left(\tau\right),\theta\left(\tau\right))(\widehat{\theta}(\tau) - \theta\left(\tau\right))\right] \\
&= -\Gamma_1^{-1}\sqrt{n}\left[G_n(\gamma\left(\tau\right),\theta\left(\tau\right)) + \Gamma_2(\gamma\left(\tau\right),\theta\left(\tau\right))(\widehat{\theta}(\tau) - \theta\left(\tau\right))\right] \\
&= -\Gamma_1^{-1}\sqrt{n}G_n(\gamma\left(\tau\right),\theta\left(\tau\right)) - \Gamma_1^{-1}\Gamma_{20} \cdot \frac{\sqrt{n}}{\sqrt{T}}\sqrt{T}(\widehat{\theta}(\tau) - \theta\left(\tau\right))
\end{aligned}$$

The first term,

$$-\Gamma_1^{-1}\sqrt{n}G_n(\gamma\left(\tau\right),\theta\left(\tau\right)) \Rightarrow N\left(0, \Omega_1^{-1}\Omega_0\Omega_1^{-1}\right)$$

The second term will converge to 0 under Assumption 5.