Inference on the Asymmetry of Loss Functions with Persistent Instruments

Matei Demetrescu — University of Kiel*

Christoph Roling — University of Bonn†

Preliminary version: February 1, 2017

Abstract

Based on observed forecast errors and a set of instruments, Elliott et al. (2005, Review of Economic Studies 72, pp. 1107-1125) propose a procedure for estimation of the degree of asymmetry of the loss function employed in forecasting. The resulting GMM estimator of the parameter characterizing the asymmetry of the loss function possesses a standard normal limiting distribution under stationarity assumptions for the instruments. In practice, instruments may not always be seen as stationary, though. The present note discusses the asymptotics of the GMM estimator and of the corresponding t statistic when some of the instruments may be highly persistent. It turns out that neither the estimator nor the t ratio possess limiting normal distributions in general, but normality is recovered in some interesting particular cases which are relevant in practice.

Key words: General cost-of-error function, Unknown persistence, Nonstandard distribution

JEL classification: C12 (Hypothesis Testing), C22 (Time-Series Models)

1 Introduction

The evaluation of forecasts is one of the important feed-back loops in applied econometrics that allow applied researchers to double-check the validity of their methodologies. For instance, it has become routine to test realized forecast error series for unbiasedness

*Corresponding author: Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, email: mdeme@stat-econ.uni-kiel.de.

†Bonn Graduate School of Economics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany.

or orthogonality. But such tests rely more or less explicitly on the assumption that the optimality criterion used to derive the forecasts is minimal MSFE. Under other loss functions, and in particular asymmetric ones, it is actually quite rational to produce biased forecasts. So it is of interest to know what shape the loss function has, that underlies some sequence of forecasts.

Elliott et al. (2005) propose a flexible class of loss functions characterized by two parameters, the asymmetry and the tail weight parameter. GMM estimation of the asymmetry parameter is possible (for simplicity often assuming known tail weight parameter), provided that variables are available, which would improve forecasts if the latter were not rational; such variables are for instance the original predictors. Assuming strict stationarity of the instrument variables, Elliott et al. show the limiting distribution of the GMM estimator to be normal, such that confidence intervals are easily computed and test decisions straightforward to make. Building on this method, Christodoulakis and Mamatzakis (2009, 2008) find asymmetric preferences in series of GDP growth forecasts of EU institutions and countries. See also Clements et al. (2007) and Capistrán (2008). More recently, Pierdzioch et al. (2012) find evidence of asymmetry in the loss function of the Bank of Canada, and Komunjer and Owyang (2011) extend the work of Elliott et al. to a multivariate setting.

One important limitation of Elliott et al. (2005) is that the instruments are assumed to be stationary. It may seem a benign technical assumption; there is however evidence that typical instruments are not stationary, but rather quite persistent. For instance, Christodoulakis and Mamatzakis (2013) use among others the lagged spot exchange rate as instrument, whose behavior can well be approximated by a random walk.

The note shows that the limiting distribution of the GMM estimator is not normal in general in the presence of persistent instruments. But we also find that there are particular cases where the limiting distribution of the t-statistic is normal such that one can still use standard normal critical values are appropriate. This is for instance the case when a persistent instrument is combined with an intercept only, or when a homoskedasticity restriction is fulfilled.

2 Estimation of asymmetry

The one-step ahead optimal predictor of a series y_t is given under a general loss function by

$$\hat{y}_t = \arg\min_{y^*} \mathrm{E}_{t-1}\left(\mathcal{L}\left(y_t - y^*\right)\right),$$

where $\mathcal{L}$ denotes the relevant loss function—which should be quasi-convex (see e.g. Granger, 1999)—and E_{t-1} denotes the expectation taken w.r.t. the conditional forecast density given the available information. The class of functions proposed by Elliott et al. (2005) is given for a forecast error u by

$$\mathcal{L}(u) = (\alpha + (1-2\alpha)\,\mathbf{1}(u<0))\,|u|^p\,. \tag{1}$$

Let $u_t = y_t - \hat{y}_t$ denote the realized forecast error at time t and $\mathcal{L}^{(1)}$ the first-order derivative of $\mathcal{L}$. Given optimality of the forecasts, information available at the time of the forecast can't reduce forecasting risk, and the so-called generalized forecast error $\mathcal{L}^{(1)}(u_t) \equiv \tilde{u}_t$ is a martingale difference sequence, $\mathrm{E}(\tilde{u}_t|\,\tilde{u}_{t-1},\ldots) = 0$ (Granger, 1999; Patton and Timmermann, 2007a).

To estimate the asymmetry parameter α given a series of forecast errors u_t, $t = 2,\ldots,T$, one employs for given p (typically $p = 1$ or $p = 2$) a set of K instrument variables gathered in the vector $\boldsymbol{w}_{t-1}$. The instruments may be, but are not restricted to, predictors from the original forecasting model. The instruments cannot improve the forecasts when these are rational, so the sequence $\boldsymbol{w}_t$ must belong to the information set and $\mathrm{E}(\tilde{u}_t|\,\tilde{u}_{t-1},\ldots,\boldsymbol{w}_{t-1},\ldots) = 0$ under rationality. This implies K moment restrictions,

$$\mathrm{E}\left(\boldsymbol{w}_{t-1}\mathcal{L}^{(1)}(u_t)\right) = \mathrm{E}(\boldsymbol{w}_{t-1}\tilde{u}_t) = 0,$$

leading for loss functions from (1) to the GMM estimator

$$\hat{\alpha} = \frac{\hat{\boldsymbol{h}}'\hat{S}^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\mathbf{1}(u_t<0)|u_t|^{p-1}\right)}{\hat{\boldsymbol{h}}'\hat{S}^{-1}\hat{\boldsymbol{h}}} = \alpha - \frac{\hat{\boldsymbol{h}}'\hat{S}^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\tilde{u}_t\right)}{\hat{\boldsymbol{h}}'\hat{S}^{-1}\hat{\boldsymbol{h}}},$$

where

$$\hat{\boldsymbol{h}} = \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\,|u_t|^{p-1}$$

and

$$\hat{S} = \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\boldsymbol{w}_{t-1}'\left(\mathbf{1}(u_t<0)-\hat{\alpha}\right)^2|u_t|^{2p-2}\,.$$

Note that, for estimation, an iterative procedure is required since $\hat{S}$ depends on α. Under the assumptions of Elliott et al. (2005), a limiting normal distribution holds for $\hat{\alpha}$ as $T \to \infty$,

$$\mathcal{T} = \sqrt{T}\,\frac{\hat{\alpha}-\alpha}{\hat{V}^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)\,,$$

where the standard error is given by

$$\hat{V}^{1/2} = \frac{1}{\sqrt{\hat{\boldsymbol{h}}'\hat{S}^{-1}\hat{\boldsymbol{h}}}}.$$

Thus, null hypotheses on α can easily be checked using the t ratio $\mathcal{T}$.[1] Confidence intervals with asymptotic coverage $1-\gamma$ are easily built as $\hat{\alpha} \pm z_{1-\gamma/2}\frac{\hat{V}^{1/2}}{\sqrt{T}}$ with $z_{1-\gamma/2}$ the $1-\gamma/2$ quantile of the standard normal distribution.

The normality of the t-type ratio $\mathcal{T}$ hinges on whether a central limit theorem may be applied to the term $\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\tilde{u}_t$. The summands $\boldsymbol{w}_{t-1}\tilde{u}_t$ form by construction a martingale difference sequence, but a "classical" central limit theorem does not apply when elements of $\boldsymbol{w}_{t-1}$ exhibit stochastic trends (e.g. random walks) and is thus nonstationary. The following section analyzes the limiting distribution of $\mathcal{T}$ when some of the instruments are allowed to be highly persistent in the sense that they possess a generic stochastic trend.

3 Asymptotics under persistence

To cover the case of both stationary and strongly persistent instruments, partition $\boldsymbol{w}_{t-1} = \left(\boldsymbol{w}_{0,t-1}', 1, \boldsymbol{w}_{1,t-1}'\right)'$ where the stationary instruments $\boldsymbol{w}_{0,t}$ satisfy the following

Assumption 1 *Let* $\left(\tilde{u}_t, \boldsymbol{w}_{0,t-1}'\right)' \in \mathbb{R}^{K_0+1}$ *be a zero-mean stationary, ergodic, uniformly* $L_{4+\delta}$*-bounded sequence with* $\mathrm{E}\left(\tilde{u}_t \middle| \tilde{u}_{t-1}, \ldots, \boldsymbol{w}_{t-1}, \ldots\right) = 0$.

Theorem 27.14 in Davidson (1994) then implies

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{[sT]}\begin{pmatrix} \tilde{u}_t \\ \boldsymbol{w}_{0,t-1}\tilde{u}_t \end{pmatrix} \Rightarrow \begin{pmatrix} \tilde{W}(s) \\ \bar{\boldsymbol{W}}(s) \end{pmatrix}, \tag{2}$$

with "$\Rightarrow$" standing for weak convergence in a space of càdlàg functions endowed with a suitable norm, and $\tilde{W}(s)$ and $\bar{\boldsymbol{W}}(s)$ denote Brownian motions. In terms of notation, let also also

$$\mathrm{Cov}\begin{pmatrix} \tilde{W}(1) \\ \bar{\boldsymbol{W}}(1) \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(\tilde{u}_t) & \mathrm{E}\left(\boldsymbol{w}_{0,t-1}'\tilde{u}_t^2\right) \\ \mathrm{E}\left(\boldsymbol{w}_{0,t-1}\tilde{u}_t^2\right) & \mathrm{E}\left(\boldsymbol{w}_{0,t-1}\boldsymbol{w}_{0,t-1}'\tilde{u}_t^2\right) \end{pmatrix} = \begin{pmatrix} \sigma_{\tilde{u}}^2 & \Gamma_0' \\ \Gamma_0 & \bar{\Omega}_0 \end{pmatrix}.$$

In case of conditional homoskedasticity of $\tilde{u}$, $\bar{\Omega}_0 = \sigma_{\tilde{u}}^2\,\mathrm{Cov}\left(\boldsymbol{w}_{0,t-1}\right)$ and $\Gamma_0 = \mathbf{0}$.

[1] When testing rather than estimating, one may compute $\hat{S}$ under the null – i.e. replace $\hat{\alpha}$ with α_0 – to reduce computational requirements.

Having a constant in the vector of instruments is on the one hand common in practice; on the other hand, the constant stands in for I(0) processes with nonzero mean since the purely stochastic component would be dominated in the limit; see the proof of Proposition 1. Hence requiring that $\boldsymbol{w}_{0,t-1}$ have zero mean does not imply any loss of generality, and one essentially has what one may call weakly persistent instruments.[2] In contrast, the stochastically trending instruments are taken to satisfy

Assumption 2 *Let N_T be a diagonal matrix with elements going to infinity, and assume that there exists a continuous-time vector process $\boldsymbol{X}(s)$ such that*

$$N_T^{-1}\boldsymbol{w}_{1,[sT]} \Rightarrow \boldsymbol{X}(s)$$

jointly with convergence in (2)*.*

The assumption allows e.g. for near-integrated modelling of persistent predictors (see e.g. Campbell and Yogo, 2006, and the references therein) but not exclusively. E.g. $X(s)$ may be an Ornstein-Uhlenbeck [OU] process with non-zero starting value, corresponding in discrete time to a near-integrated process with initial condition drawn from the unconditional distribution (as employed by Müller and Elliott, 2003). Also, $X(s)$ may be a fractional Brownian motion (see e.g. Marinucci and Robinson, 1999, and the references therein).

We are now in the position to discuss the asymptotics when strongly persistent regressors are present in the instrument set.

Proposition 1 *Under Assumptions 1 and 2, it holds as $T \to \infty$ that*

$$\sqrt{T}(\hat{\alpha}-\alpha) \Rightarrow \frac{\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{U}}{\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{H}} \qquad \textit{and} \qquad \mathcal{T} \Rightarrow \frac{\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{U}}{\sqrt{\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{H}}}$$

where

$$\boldsymbol{H} \equiv \begin{pmatrix} \mathrm{E}\left(\boldsymbol{w}_{0,t-1}\,|u_t|^{p-1}\right) \\ \mathrm{E}\left(|u_t|^{p-1}\right) \\ \mathrm{E}\left(|u_t|^{p-1}\right)\int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s \end{pmatrix},$$

$$\boldsymbol{S} \equiv \begin{pmatrix} \bar{\Omega}_0 & \Gamma_0 & \Gamma_0 \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \Gamma_0' & \sigma_{\tilde{u}}^2 & \sigma_{\tilde{u}}^2 \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s\,\Gamma_0' & \sigma_{\tilde{u}}^2 \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s & \sigma_{\tilde{u}}^2 \int_0^1 \boldsymbol{X}(s)\,\boldsymbol{X}'(s)\,\mathrm{d}s \end{pmatrix}$$

[2] This terminology may conflict with the persistence notion associated with long memory processes: stationary long memory is allowed for $\boldsymbol{w}_{0,t}$, uniform $L_{4+\delta}$-boundedness provided. We stick to it, though, since it complements strongly persistent processes which are not stationary or ergodic.

and

$$\boldsymbol{U} \equiv \begin{pmatrix} \bar{\boldsymbol{W}}(1) \\ \tilde{W}(1) \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}\tilde{W}(s) \end{pmatrix}.$$

Proof: *See the Appendix.*

Given the presence of the Ito-type integral in $\boldsymbol{U}$, the result implies a nonstandard distribution in general, for both $\hat{\alpha}$ and for its t-statistic. Take as an extreme example the case where there is exactly one persistent instrument such that, under the null,

$$\mathcal{T} \Rightarrow \operatorname{sgn}\left(\int_0^1 X(s)\,\mathrm{d}s\right) \frac{\int_0^1 X(s)\,\mathrm{d}\tilde{W}(s)}{\sqrt{\int_0^1 X^2(s)\,\mathrm{d}s}}.$$

Note that, in spite of the nonstandard distribution of $\hat{\alpha}$, the GMM estimator is $\sqrt{T}$-consistent only, and not superconsistent as might have been expected given stochastically trending instruments.

Worse yet, the distributions depend on the distributional properties of the process $\boldsymbol{X}(s)$, whose parameters may not always be consistently estimated (e.g. the mean reversion parameter of an OU process). Moreover, other types of persistence than near-integration generate similar behavior, and distinguishing among them to take the right corrective action—e.g. deciding between fractional integration and near integration—is difficult; see Müller and Watson (2008).

In some cases limiting normality is recovered for the statistic $\mathcal{T}$ in spite of the nonstandard limiting distribution of $\hat{\alpha}$. The first, and in a sense obvious, such case is when the instruments $\boldsymbol{w}_{1,t}$ are "exogenous" in such a way that mixed Gaussianity of $\boldsymbol{U}$ is provided for:

Corollary 1 *Let $\boldsymbol{X}(s)$ be independent of $\left(\tilde{W}(s), \bar{\boldsymbol{W}}'(s)\right)'$. Then*

$$\mathcal{T} \stackrel{d}{\to} \mathcal{N}(0,1).$$

Proof: *Obvious and omitted.*

But the interesting aspect of the studied inferential problem is that normality for $\mathcal{T}$ is not given exclusively when $\boldsymbol{U}$ is mixed Gaussian. If the forecast errors fulfil some restrictions on the serial dependence of the conditional higher-order moments, the following corollary shows that normality of $\mathcal{T}$ is recovered even in cases where Corollary 1 does not apply.

Corollary 2 *Let* $\mathrm{E}\left(\left|u_t\right|^{p-1}\middle|\,\tilde{u}_{t-1},\ldots,\boldsymbol{w}_{t-1},\ldots\right)=\mu_{p-1}$ *and* $\mathrm{E}\left(\tilde{u}_t^2\middle|\,\tilde{u}_{t-1},\ldots,\boldsymbol{w}_{t-1},\ldots\right)=\sigma_{\tilde{u}}^2$ *be constant. Then*

$$\mathcal{T}\xrightarrow{d}\mathcal{N}\left(0,1\right).$$

Proof: *See the Appendix.*

The corollary essentially requires constant conditional scale of u_t. The presence of a constant instrument is paramount for the result. To understand why, consider the simple bivariate case with one weakly and one strongly persistent instrument; let also the weakly persistent instrument have non-zero mean. Then

$$\begin{aligned}\boldsymbol{H}'\boldsymbol{S}^{-1} \equiv & \left(\begin{array}{cc}\mathrm{E}\left(w_{0,t-1}\left|u_t\right|^{p-1}\right); & \mathrm{E}\left(\left|u_t\right|^{p-1}\right)\int_0^1 X\left(s\right)\mathrm{d}s\end{array}\right)\\ & \times\left(\begin{array}{cc}\bar{\omega}_0 & \gamma_0\int_0^1 X\left(s\right)\mathrm{d}s\\ \gamma_0\int_0^1 X\left(s\right)\mathrm{d}s & \sigma_{\tilde{u}}^2\int_0^1 X^2\left(s\right)\mathrm{d}s\end{array}\right)^{-1}\end{aligned}$$

while

$$\boldsymbol{U}\equiv\left(\begin{array}{c}\int_0^1\mathrm{d}\bar{W}\left(s\right)\\ \int_0^1 X\left(s\right)\mathrm{d}\tilde{W}\left(s\right)\end{array}\right).$$

The coefficient of $\int_0^1 X\left(s\right)\mathrm{d}\tilde{W}\left(s\right)$ in $\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{U}$ should be zero for normality to be recovered in general. Some algebra indicates this to be the case when

$$-\mathrm{E}\left(w_{0,t-1}\left|u_t\right|^{p-1}\right)\gamma_0\int_0^1 X\left(s\right)\mathrm{d}s+\mathrm{E}\left(\left|u_t\right|^{p-1}\right)\int_0^1 X\left(s\right)\mathrm{d}s\,\bar{\omega}_0=0,$$

or

$$\frac{\mathrm{E}\left(w_{0,t-1}\left|u_t\right|^{p-1}\right)}{\mathrm{E}\left(\left|u_t\right|^{p-1}\right)}=\frac{\mathrm{E}\left(w_{0,t-1}^2\tilde{u}_t^2\right)}{\mathrm{E}\left(w_{0,t-1}\tilde{u}_t^2\right)}.$$

For constant conditional scale this reduces to $\mathrm{E}\left(w_{0,t-1}^2\right)=\mathrm{E}\left(w_{0,t-1}\right)^2$ and the weakly persistent instrument must thus be constant.

Moreover, the constant alone may also eliminate nonstandard distribution components from $\mathcal{T}$. The corresponding requirement is that only a constant and persistent instruments are employed, but no stationary ones, as shown by the following

Corollary 3 *Let* $\boldsymbol{w}_{t-1}=\left(1,\boldsymbol{w}_{1,t-1}'\right)'$. *Then*

$$\mathcal{T}\xrightarrow{d}\mathcal{N}\left(0,1\right).$$

Proof: *See the Appendix.*

The corollaries are illustrated in finite samples in the following section.

4 Monte Carlo experiment

We illustrate the above asymptotic findings by means of a Monte Carlo experiment. The purpose of this exercise is to examine inference on the asymmetry parameter α in (1) and to highlight the main results presented in section 3. We combine the frameworks of Engle et al. (1987), Bollerslev (1990) and Gospodinov (2009) to obtain the following data generating process:

$$\begin{bmatrix} s_t \\ f_t \end{bmatrix} = \begin{bmatrix} f_{t-1} + \delta\sqrt{h_{1,t}} \\ f_{t-1} \end{bmatrix} + \Omega_t^{1/2}\boldsymbol{\epsilon}_t \tag{3}$$

for $t = 2, \ldots, T$ and $f_0 = 0$. Here, $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{1,t})'$ is an i.i.d. Gaussian process with zero mean and identity covariance matrix. Furthermore,

$$\Omega_t = \begin{bmatrix} \sqrt{h_{1,t}} & 0 \\ 0 & \sqrt{h_{2,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_{1,t}} & 0 \\ 0 & \sqrt{h_{2,t}} \end{bmatrix} \tag{4}$$

and

$$h_{i,t} = \gamma_{0,i} + \gamma_{1,i}\epsilon_{i,t-1}^2 \qquad \text{for } i = 1, 2.$$

To fix ideas, s_t may denote the logarithm of the weekly spot rate of a given currency (relative to USD, say) and f_t may denote the corresponding forward rate formed at time $t-1$ for time t. Hence we adopt the ARCH in mean (ARCH-M) model for the series of excess returns $s_t - f_{t-1}$. Following Engle et al. (1987), the specification in (3) implies that the excess return is is proportional to the conditional standard deviation of the spot rate.[3]

To provide some context for this design, recall that the efficient markets hypothesis can be tested by running a regression of excess returns on predictors available at time t, say,

$$s_{t+1} - f_t = \hat{\beta}_0 + \hat{\beta}_1 (f_t - s_t) + \hat{\epsilon}_{t+1},$$

in which the joint hypothesis $\hat{\beta}_0 = \hat{\beta}_1 = 0$ is tested. Equivalently, the rational expectations hypothesis amounts to estimating a linear regression as in

$$s_{t+1} - s_t = \hat{\gamma}_0 + \hat{\gamma}_1 (f_t - s_t) + \hat{\varepsilon}_{t+1},$$

[3] See also Baillie and Bollerslev (2000) who show that in a consumption-based asset pricing model, the discrepancy between the (expected) spot and forward rate is a function of the conditional variance deviation of the spot rate.

and testing $\hat{\gamma}_0 = 0$ and $\hat{\gamma}_1 = 1$. The latter hypothesis tends to be rejected in a number of empirical applications. In these studies it is usually assumed that the statistical loss function adopted by market forecasters is symmetric. Christodoulakis and Mamatzakis (2013) investigate the role of asymmetric loss and estimate the asymmetry parameter α in (1) for G7 exchange rate markets. They find that the estimated asymmetry paramater is substantially and significantly smaller than 0.5 for most exchange rate series, suggesting that overprediction is more costly for market participants.

We do not consider estimation and inference in the predictive regression above, however, and rather focus on inference on the asymmetry parameter in this simplified framework using the test statistic $\mathcal{T}$ and forecast errors $\{s_t - f_{t-1}\}_{t=3}^{T}$. To estimate the loss function parameter, we employ the GMM estimator described above and choose the instruments as

$$(w_{0,t-2}, w_{1,t-2})' = (1, f_{t-2})'.$$

for $t = 3, \ldots, T$. Notice that, by design, the instrument f_{t-2} is highly persistent, which is in line with empirical evidence.[4]

We set $T \in \{100, 1000\}$, while the (constant) conditional correlation is $\rho = 0.5$, the ARCH parameters are given by $\gamma_{0,2} = \gamma_{0,1} = 0.01$, $\gamma_{1,2} = 0$, and $\delta = 0.5$. Furthermore, we consider the cases $\gamma_{1,1} = 0.95$ and $\gamma_{1,1} = 0$ separately below. In each case, we draw 20,000 samples from this data generating process and estimate the loss function parameter $\hat{\alpha}$ and the variance parameter $\hat{V}$ in each step, assuming $p = 2$ for the loss function.

By Theorem 1 in Patton and Timmermann (2007b), the optimal forecast of $s_{t-1} - f_t$ in this model satisfies $\delta\sqrt{h_{1,t}} + C\sqrt{h_{1,t}}$ and is thus zero if $\delta = -C$, where C is a constant that depends only on the distribution of the idiosyncratic error and the loss function $\mathcal{L}$. Under normality and given that $\delta = 0.5$, we can then select the asymmetry parameter α consistent with $s_{t-1} - f_t$ being unforecastable under the asymmetric quadratic loss function. This value turns out to be $\alpha = 0.22$.[5]

Figure 1 shows the kernel density estimates for $T = 100$ (red) and $T = 1000$ (blue) of the distribution of $\mathcal{T}$ when $\gamma_{1,1} = 0.95$, so that ARCH effects are present. The standard normal density is also depicted (black). Figure 2 presents the density estimates in the

[4] See Liu and Maynard (2005) or Gospodinov (2009), for instance

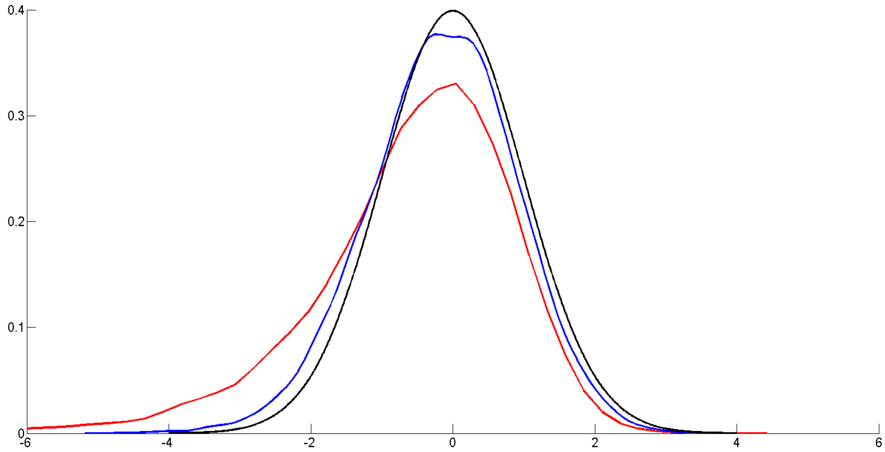
[5] The constant C satisfies $E\left[\mathcal{L}'\left(\epsilon_{1,t} - C\right)\right] = 0$, see Patton and Timmermann (2007b). From this condition and the normality assumption, we obtain

$$\Phi\left(C\right)\left(\frac{E\left(\epsilon_{1,t}\middle|\epsilon_{1,t} < C\right)}{C} - 1\right) = \frac{\alpha}{1-2\alpha}$$

with $E\left(\epsilon_{1,t}\middle|\epsilon_{1,t} < C\right) = \frac{1}{\Phi(C)}\int_{-\infty}^{C} x\phi\left(x\right)dx$, where ϕ and Φ are the standard normal p.d.f and c.d.f..

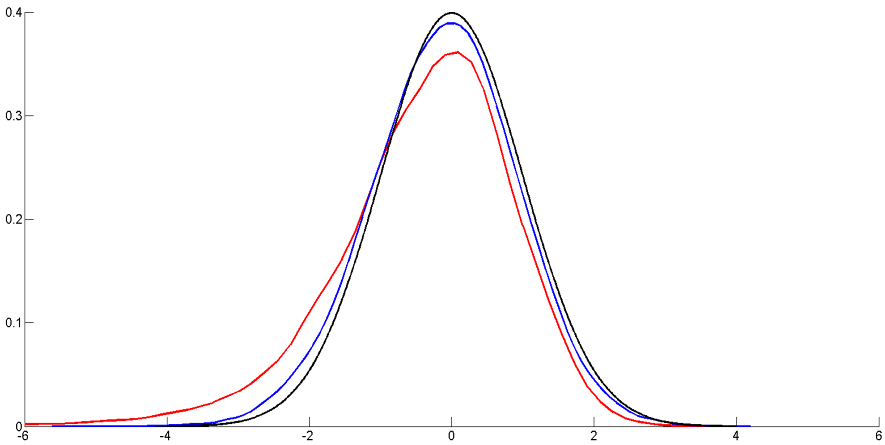
special case of conditional homoskedasticity when $\gamma_{1,1} = 0$.

Figure 1: Density estimates of $\mathcal{T}$ when $\alpha = 0.22$ and $\gamma_{1,1} = 0.95$



Note: The figures shows the density of the standard normal distribution (black) and the kernel density estimates of the distribution of $\mathcal{T}$ when $T = 100$ (red) and $T = 1000$ (blue). The underlying DGP is given by (3) - (4) with $\delta = \rho = 0.5$, $\gamma_{0,1} = \gamma_{0,2} = 0.01$, and $\gamma_{1,2} = 0$.

Figure 2: Density estimates of $\mathcal{T}$ when $\alpha = 0.22$ and $\gamma_{1,1} = 0$



Note: The figures shows the density of the standard normal distribution (black) and the kernel density estimates of the distribution of $\mathcal{T}$ when $T = 100$ (red) and $T = 1000$ (blue). The underlying DGP is given by (3) - (4) with $\delta = \rho = 0.5$, $\gamma_{0,1} = \gamma_{0,2} = 0.01$, and $\gamma_{1,2} = 0$.

Consistent with Corollary 3, the test statistic $\mathcal{T}$ is asymptotically normally distributed, even if one of the instruments is persistent. Although in small samples the approximation of the distribution by the normal distribution is not perfect, the estimated density of the statistic resembles the standard normal density more closely as the sample size increases from $T = 100$ to $T = 1000$, illustrating that inference using the standard normal approximation is valid asymptotically.[6]

[6]The skewness is likely induced by the iterative nature of the estimator $\hat{\alpha}$; should one use the null

5 Concluding remarks

We investigate the limiting distribution of the t-type test statistic to conduct inference on the asymmetry parameter of an asymmetric loss function considered by Elliott et al. (2005). In particular, we study the empirically relevant case when some of the instruments employed in the GMM estimation procedure are persistent.

Unless a case can be made that the (generalised) forecast error is conditionally homoskedastic, approximating the distribution of the GMM-based t-statistic by the normal distribution is generally invalid when some of the instruments are persistent. Practitioners could then conduct a test of the null hypothesis of constant conditional moments of $\tilde{u}_t$ (e.g. in the spirit of Bierens, 1982, but a parametric test for no ARCH effects might also be used) to be able to justify χ^2 critical values via Corollary 2.

Alternatively, the set of instruments may be split to separate weakly from strongly persistent instrument and run separate tests for the two sets of instruments. This would allow to exploit 3 without worrying about the conditional homoskedasticity requirement, but would have to take the multiple testing nature of the situation into account, say via a Bonferroni correction.

Finally, it may be more convenient to ensure mixed Gaussianity from the beginning by suitably choosing the instrumental variables $\boldsymbol{w}_{1,t}$. To this end, one may use persistent exogenous variables such as functions of time instead of the instruments suspected of generating nonnormality. As pointed out by Breitung and Demetrescu (2015), this is a nice exploit of the spurious correlation effect between trending variables: any persistent variable will be instrumented by another trending one, and all one has to do is choose the instrument in such a way that Gaussianity is recovered. Deterministic functions of time satisfy this; they propose $w_{1,t,k} = \sin\left(k\frac{\pi}{2}\frac{t}{T}\right)$ based on the the Loève-Karhunen representation of $\boldsymbol{X}$, should it exist (mean-square integrability is a sufficient condition).

Appendix

Before proving the main result, we state and prove a useful lemma.

value α_0 when testing, the finite-sample distribution is indeed closer to the standard normal in the given experiments.

Lemma 1 *For any strictly stationary ergodic process z_t, uniformly $L_{1+\delta}$-bounded for some $\delta > 0$ and any and w_t strongly persistent in the sense of Assumption 2, it holds that*

$$\frac{1}{Tn_T}\sum_{t=2}^{T} w_{t-1}z_t \xrightarrow{d} \mathrm{E}\left(z_t\right)\int_0^1 X\left(s\right)\mathrm{d}s.$$

Note that the lemma, applied elementwise, implies under the assumptions of Proposition 1 that

1. $\frac{1}{T}N_T^{-1}\sum_{t=2}^{T}\boldsymbol{w}_{1,t-1}\left|u_t\right|^{p-1} \xrightarrow{d} \mathrm{E}\left(\left|u_t\right|^{p-1}\right)\int_0^1 \boldsymbol{X}\left(s\right)\mathrm{d}s$
2. $\frac{1}{T}N_T^{-1}\sum \boldsymbol{w}_{0,t-1}\boldsymbol{w}'_{1,t-1}\tilde{u}_t^2 \xrightarrow{p} \Gamma_0\int_0^1 \boldsymbol{X}'\left(s\right)\mathrm{d}s$
3. $\frac{1}{T}N_T^{-1}\sum_{t=2}^{T}\boldsymbol{w}_{1,t-1}\tilde{u}_t^2 \xrightarrow{d} \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}\left(s\right)\mathrm{d}s$
4. $\frac{1}{T}N_T^{-1}\left(\sum_{t=2}^{T}\boldsymbol{w}_{1,t-1}\boldsymbol{w}'_{1,t-1}\tilde{u}_t^2\right)N_T^{-1} \xrightarrow{d} \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}\left(s\right)\boldsymbol{X}'\left(s\right)\mathrm{d}s.$

Proof of Lemma 1

Write

$$\frac{1}{Tn_T}\sum_{t=1}^{T-1} w_t z_t = \frac{1}{Tn_T}\sum_{t=1}^{T-1} w_t\left(z_t - \mathrm{E}\left(z_t\right)\right) + \mathrm{E}\left(z_t\right)\frac{1}{Tn_T}\sum_{t=1}^{T-1} w_t.$$

Should the first term vanish as $T \to \infty$, the desired result follows directly from Assumption 2 with the continuous mapping theorem. Let then $\tilde{z}_t = z_t - \mathrm{E}\left(z_t\right)$ and note that, since z_t is ergodic, $\mathrm{E}\left(\tilde{z}_t|\tilde{z}_{t-m},\tilde{z}_{t-m-1},\ldots\right) \xrightarrow{p} 0$ as $m \to \infty$. Furthermore, since $\tilde{z}_t$ is uniformly $L_{1+\delta}$-bounded for some $\delta > 0$, it is uniformly integrable and thus $\mathrm{E}\left(\left|\mathrm{E}\left(\tilde{z}_t|\tilde{z}_{t-m},\tilde{z}_{t-m-1},\ldots\right)\right|\right) \to 0$. Then, Theorem 3.3 of Hansen (1992) applies, such that

$$\left|\frac{1}{Tn_T}\sum_{t=1}^{T-1} w_t\tilde{z}_t\right| \leq \sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{[sT]}\frac{w_t}{n_T}\tilde{z}_t\right| \xrightarrow{p} 0$$

as required for the result.

Proof of Proposition 1

It holds that

$$\sqrt{T}\left(\hat{\alpha}-\alpha\right) = \frac{\hat{\boldsymbol{h}}'\hat{S}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\tilde{u}_t\right)}{\hat{\boldsymbol{h}}'\hat{S}^{-1}\hat{\boldsymbol{h}}}$$

and

$$t_{\hat{\alpha}} = \frac{\hat{\boldsymbol{h}}'\hat{S}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\tilde{u}_t\right)}{\sqrt{\hat{\boldsymbol{h}}'\hat{S}^{-1}\hat{\boldsymbol{h}}}}.$$

We have, regularity conditions assumed, that

$$D_T^{-1}\hat{\boldsymbol{h}} \Rightarrow \begin{pmatrix} \mathrm{E}\left(\boldsymbol{w}_{0,t-1}\,|u_t|^{p-1}\right) \\ \mathrm{E}\left(|u_t|^{p-1}\right) \\ \mathrm{E}\left(|u_t|^{p-1}\right)\int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s \end{pmatrix}$$

with

$$D_T = \begin{pmatrix} I & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & N_T \end{pmatrix}$$

and

$$D_T^{-1}\hat{\boldsymbol{S}}D_T^{-1} \Rightarrow \begin{pmatrix} \bar{\Omega}_0 & \Gamma_0 & \Gamma_0\int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \Gamma_0' & \sigma_{\tilde{u}}^2 & \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \left(\int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s\right)\Gamma_0' & \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s & \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}(s)\,\boldsymbol{X}'(s)\,\mathrm{d}s \end{pmatrix}.$$

Moreover,

$$D_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{w}_{t-1}\tilde{u}_t \Rightarrow \begin{pmatrix} \bar{\boldsymbol{W}}(1) \\ \tilde{W}(1) \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}\tilde{W}(s) \end{pmatrix};$$

see Hansen (1992, Theorem 2.1). The result follows with the continuous mapping theorem.

Proof of Corollary 2

From the assumptions of the corollary it follows that $\mathrm{E}\left(\boldsymbol{w}_{0,t-1}\,\left|u_t\right|^{p-1}\right) = \boldsymbol{0}$ just like Γ_0 and, with $\mathrm{E}\left(\widetilde{u}_t^2\right) = \sigma_{\widetilde{u}}^2$, $\bar{\Omega}_0 = \sigma_{\widetilde{u}}^2\,\mathrm{E}\left(\boldsymbol{w}_{0,t-1}\boldsymbol{w}_{0,t-1}'\right)$ one obtains

$$\begin{aligned} \boldsymbol{H}\boldsymbol{S}^{-1} &\equiv \mathrm{E}\left(|u_t|^{p-1}\right)\left(\boldsymbol{0}', 1, \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s\right)\begin{pmatrix} \frac{1}{\sigma_{\widetilde{u}}^2}\mathrm{E}\left(\boldsymbol{w}_{0,t-1}\boldsymbol{w}_{0,t-1}'\right)^{-1} & \boldsymbol{0}' \\ \boldsymbol{0} & Q^{-1} \end{pmatrix} \\ &\equiv \mathrm{E}\left(|u_t|^{p-1}\right)\left(\boldsymbol{0}', \left(1, \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s\right)Q^{-1}\right) \end{aligned}$$

with

$$Q = \sigma_{\widetilde{u}}^2\begin{pmatrix} 1 & \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s & \int_0^1 \boldsymbol{X}(s)\,\boldsymbol{X}'(s)\,\mathrm{d}s \end{pmatrix}.$$

Now, $\left(1, \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s\right)'$ is the first column of $Q/\sigma_{\widetilde{u}}^2$ so its transpose, postmultiplied with the inverse of Q, gives $\sigma_{\widetilde{u}}^{-2}\left(1, \boldsymbol{0}\right)'$ where there are exactly as many zeros as elements of $\boldsymbol{w}_{1,t}$. Hence

$$\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{U} \equiv \frac{\mathrm{E}\left(|u_t|^{p-1}\right)}{\sigma_{\widetilde{u}}}\tilde{W}(1).$$

The same reasoning indicates that

$$\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{H} \equiv \frac{\left(\mathrm{E}\left(|u_t|^{p-1}\right)\right)^2}{\sigma^2_{\tilde{u}}}$$

such that

$$\mathcal{T} \Rightarrow \mathcal{N}(0,1)$$

whenever $\tilde{u}_t$ has constant conditional scale in the sense that the conditional expectation of the relevant powers of $|\tilde{u}_t|$ are constant.

Proof of Corollary 3

The result follows by noting that, without stationary instruments $\boldsymbol{w}_{0,t-1}$, $\boldsymbol{H}$ is proportional to the first row of $\boldsymbol{S}$, such that $\boldsymbol{H}'\boldsymbol{S}^{-1}$ is proportional to the first column of the identity matrix, which then cancels out all nonstandard terms in $\mathcal{T}$ and the result follows.

References

Baillie, R. and T. Bollerslev (2000). The forward premium anomaly is not as bad as you think. *Journal International Money and Finance 19*(4), 471–488.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics 20*(1), 105–134.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics 72*, 498–505.

Breitung, J. and M. Demetrescu (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics 187*(1), 358–375.

Campbell, J. Y. and M. Yogo (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics 81*(1), 27–60.

Capistrán, C. (2008). Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious? *Journal of Monetary Economics 55*(8), 1415–1427.

Christodoulakis, G. and E. Mamatzakis (2008). An assessment of the EU growth forecasts under asymmetric preferences. *Journal of Forecasting 27*(6), 483–492.

Christodoulakis, G. and E. Mamatzakis (2009). Assessing the prudence of economic forecasts in the EU. *Journal of Applied Econometrics 24*(4), 583–606.

Christodoulakis, G. and E. Mamatzakis (2013). Behavioural asymmetries in the G7 foreign exchange market. *International Review of Financial Analysis 29*, 261–270.

Clements, M., F. Joutz, and H. Stekler (2007). An evaluation of the forecasts of the Federal Reserve: A pooled approach. *Journal of Applied Econometrics 22*(1), 121–136.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

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

Engle, R., D. Lilien, and R. Robins (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica 55*(2), 391–407.

Gospodinov, N. (2009). A new look at the forward premium puzzle. *Journal of Financial Econometrics 7*(3), 312–338.

Granger, C. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review 1*(2), 161–173.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory 8*(4), 489–500.

Komunjer, I. and M. T. Owyang (2011). Multivariate forecast evaluation and rationality testing. *The Review of Economics and Statistics 94*(4), 1066–1080.

Liu, W. and A. Maynard (2005). Testing forward rate unbiasedness allowing for persistent regressors. *Journal of Empirical Finance 12*(5), 613–628.

Marinucci, D. and P. M. Robinson (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference 80*(1), 111–122.

Müller, U. K. and G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica 71*(4), 1269–1286.

Müller, U. K. and M. W. Watson (2008). Testing models of low-frequency variability. *Econometrica 76*(5), 979–1016.

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

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

Pierdzioch, C., J. Rülke, and G. Stadtmann (2012). On the loss function of the Bank of Canada: A note. *Economics Letters 115*(2), 155–159.