MULTIVARIATE FORECAST EVALUATION AND RATIONALITY TESTING

IVANA KOMUNJER AND MICHAEL T. OWYANG

Abstract. In this paper, we propose a new family of multivariate loss functions that can be used to test the rationality of vector forecasts without assuming independence across individual variables. When only one variable is of interest, the loss function reduces to the flexible asymmetric family recently proposed by Elliott, Komunjer, and Timmermann (2005). Following their methodology, we derive a GMM test for multivariate forecast rationality that allows the forecast errors to be dependent, and takes into account forecast estimation uncertainty. We apply our test to the study of rationality of macroeconomic vector forecasts in: the growth rate in nominal output, the CPI inflation rate, and a short-term interest rate.

1. Introduction

Since the seminal works of Muth (1961) and Lucas (1973), rationality in expectation formation has been the cornerstone of economic models. The notion that agents form expectations rationally is found, for example, in the early work on monetary policy (Friedman, 1968), the natural rate hypothesis (Sargent, 1973; Shiller, 1978), and bond markets (Modigliani and Shiller, 1973; Poole, 1976), among others. Consequently, a large body of empirical work has been devoted to testing the rational expectations hypothesis

Affiliations and Contact information. Komunjer: Department of Economics, University of California, San Diego [komunjer@ucsd.edu]. Owyang: Federal Reserve Bank of St. Louis [Michael.T.Owyang@stls.frb.org].

Acknowledgments:

(REH). Market-based studies typically consider the REH in financial markets.[1] Survey-based studies, on the other hand, exploit cross-sections (and often panels) of data collected from forecasters to test rationality.[2]

While survey-based tests have typically rejected rationality for groups of professional forecasters, these tests tend to treat expectations formation as independent across variables. That is, forecast rationality is tested for each variable individually under the assumption that the forecast errors are independent. Most economic theories, on the other hand, are contrived from multivariate models with comoving variables. It therefore seems contrary to test rationality in either univariate or independent multivariate frameworks as both treatments belie Muth's (1961) characterization of rational expectations as "distributed, for the same information set, about the prediction of the theory (or the "objective" probability distributions of outcomes)".

In fact, the general tenor of the forecasting literature in recent years has been a move toward the incorporation of interaction among many individual variables. For example, the VAR forecasting models used to investigate the effects of monetary policy build in an interdependence of key macroeconomic variables such as output, prices, employment, and interest rates; see Sims (1986) and, more recently, Christiano, Eichenbaum, and Evans (1999). While these models have evolved from forecasting applications into devices for explaining the policy dynamics, the underlying belief that macrovariables are interconnected no doubt extends back through to the construction of forecasts. Employing even more data, forecasting models recently-popularized by Stock and Watson (1999) and Boivin and Ng (2006) are constructed around the notion that the state of the economy is influenced by multiple comoving data series through any number of dynamic factors.[3] Even

[1]In financial markets, rationality has implications for the realization of interest rate spreads (Fama, 1990; Campbell and Shiller, 1987; Bekaert and Hodrick, 2001) and exchange rates (Engel and West, 2005).

[2]These papers examine either the consensus, i.e., the distributions of expectations (see Pesaran (1987) for a survey), or pooled agents (Figlewski and Wachtel, 1981). Bonham and Cohen (2001) examine conditions under which either of these methodologies are valid.

[3]Perhaps the most telling evidence for multivariate forecasting is FRBUS, the model used by the Federal Reserve Board of Governors to construct forecasts for monetary policymaking (Brayton, Levin, Tryon, and Williams, 1997).

the belief in the most elementary macroeconomic relationships – such as the Phillips curve or the Fisher hypothesis – introduces an interdependence in the forecasted variables.[4]

It is, therefore, not unusual for the survey forecasts to involve two or more correlated variables. A prominent example is the Survey of Professional Forecasters (SPF) which reports forecasts of both output and inflation. When testing the rationality of such multivariate forecasts, a multivariate loss function is needed. Desirable properties of such loss include: (1) it does not treat the components of the forecast vectors as independent; (2) it allows for asymmetry in the treatment of over and under-prediction of the individual variables being forecast. Most of literature on forecast rationality testing implicitly or explicitly assumes the losses for individual variables are independent; see Kirchgässner and Müller (2006), for example. Under independence, multivariate losses reduce to sums of univariate losses in each of the variables taken separately. In presence of dependence across variables being forecast, this decomposition no longer necessarily holds.

In addition, rationality tests typically assume that the forecaster's loss function is quadratic. Thus, the forecaster's objective is simply to minimize the magnitude of the forecast error, regardless of its directionality. Conclusions drawn using tests bearing this assumption have, for the most part, revealed a lack of forecast rationality. Recently, however, Elliott, Komunjer, and Timmermann (2005) (EKT hereafter) argued that a simple quadratic loss may not be sufficiently flexible for evaluating forecast rationality. They argue that asymmetric loss, in which positive and negative forecast errors may be weighted differently (Zellner, 1986; Christoffersen and Diebold, 1997; Batchelor and Peel, 1998; Elliott, Komunjer, and Timmermann, 2005, 2006), might better represent the forecaster's objective function. In particular, Elliott, Komunjer, and Timmermann (2006) find evidence for asymmetric loss in the output and inflation SPF forecasts.

In light of these facts, we propose a test of multivariate forecast rationality that accounts for both interdependence of the forecast errors and directional asymmetries. The test is based on a novel multivariate objective (loss) function that explicitly models codependence in the forecasted variables. If agents have symmetric preferences, our multivariate

[4]The viability of these relationships for forecasting is discussed at length in Stock and Watson (1999) and Barsky (1987), respectively.

loss function reduces to the sum of univariate losses. In this case, our test is equivalent to a joint test of univariate rationality. However, if agents have directional preferences, assuming independence across forecasted variables produces two biases. First, the independence assumption can alter the result of rationality tests. Second, the econometrician may incorrectly infer a greater degree of directional preference for the forecaster. In this sense, the assumption of independence amounts to a misspecification of the forecaster's loss function. Accounting for potential correlation across forecasted variables can, in some cases, lessen the degree of asymmetry found in the panel of forecasters.

Empirically, the misspecification due to the assumption of independence is highlighted if the correlations of variables differs across subperiods. One such event occurred in early 1994 when the Federal Reserve began to announce the federal funds target, altering the informational environment for forecasters. Under the assumption of independence, one might not expect a change in Fed policy to affect the forecaster's loss associated with, say, output or inflation. Neglecting the codependence between variables can, thus, bias the estimation of the forecaster's directional preferences. Under the multivariate framework, on the other hand, we can properly account for the codependence of output and inflation with the policy variable.

The remainder of the paper is organized as follows: Section 2 develops the theoretical foundation for our multivariate approach. Here, we review the notation and assumptions, propose a new family of multivariate loss functions, and derive their properties. Where appropriate, we emphasize the differences between the univariate and multivariate loss functions. Section 3 outlines the testing procedures for multivariate forecast rationality. Section 4 describes the data used in our empirical application and presents the results. In particular, we focus on the results of the rationality tests and the values of the asymmetry coefficients in the estimated loss functions obtained in the pre- and post-1994 subsamples. Section 5 concludes.

2. Multivariate Forecasts and Loss Function

2.1. **Setup and Notation.** Consider a stochastic process $Z \equiv \{\mathbf{z}_t : \Omega \longrightarrow \mathbb{R}^{n+m}, (n,m) \in \mathbb{N}^2, t = 1, 2, \ldots\}$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where

$\mathcal{F} = \{\mathcal{F}_t, t = 1, 2, \ldots\}$ and $\mathcal{F}_t$ is the σ-field $\mathcal{F}_t \equiv \sigma\{\mathbf{z}_s, s \leq t\}$. In what follows, we let $\mathbf{z}_t \equiv (\mathbf{y}_t', \mathbf{w}_t')'$, where $\mathbf{y}_t$ denotes the subvector of interest of the observed vector $\mathbf{z}_t$—$\mathbf{y}_t \in \mathbb{R}^n$—and where the remaining subvector $\mathbf{w}_t$—$\mathbf{w}_t \in \mathbb{R}^m$—stands for other variables.[5] We denote by $F_t^0(\cdot)$ the distribution of $\mathbf{y}_t$ conditional on $\mathcal{F}_t$, i.e. $F_t^0(\mathbf{y}) = \mathcal{P}(\mathbf{y}_t \leqslant \mathbf{y})$ for any $\mathbf{y} \in \mathbb{R}^n$ where $\leqslant$ denotes the usual partial order on $\mathbb{R}^n$.[6] We shall assume that:

A1. *For all $t = 1, 2, \ldots$ the conditional distribution $F_t^0(\cdot)$ is continuously differentiable and the corresponding conditional density $f_t^0(\cdot) > 0$ on $\mathbb{R}^n$.*

In the forecasting problem considered here, we let $\mathbf{f}_{t+s,t}$ denote the time-t forecast of the n-vector $\mathbf{y}_{t+s}$, where s is the prediction horizon of interest, $s \geq 1$. The forecast vector $\mathbf{f}_{t+s,t}$ contains all the information comprised in $\mathcal{F}_t$ which is informative for $\mathbf{y}_{t+s}$, including lagged values of $\mathbf{y}_t$ in addition to other variables $\mathbf{w}_t$ used to predict $\mathbf{y}_{t+s}$. For simplicity, we focus on the one-step-ahead predictions of $\mathbf{y}_{t+1}$—denoted $\mathbf{f}_{t+1,t}$—knowing that all results developed in this case can readily be generalized to any $s > 1$. Using the standard notation, we let $\mathbf{e}_{t+1}$ denote the time-$t+1$ forecast error n-vector, $\mathbf{e}_{t+1} = \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}$.

Hereafter, for any scalar u, $u \in \mathbb{R}$, we let $\mathbb{1} : \mathbb{R} \to [0, 1]$ be the Heaviside (or indicator) function, i.e. $\mathbb{1}(u) = 0$ if $u < 0$, $\mathbb{1}(u) = 1$ if $u > 0$, and $\mathbb{1}(0) = \frac{1}{2}$ (Bracewell, 2000). Similarly, we use $\text{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ to denote the sign function: $\text{sgn}(u) = \mathbb{1}(u) - \mathbb{1}(-u) = 2\mathbb{1}(u) - 1$, and let $\delta : \mathbb{R} \to \mathbb{R}$ be the Dirac delta function. Note that the Heaviside function is the indefinite integral of the Dirac function, i.e. $\mathbb{1}(u) = \int_a^u d\delta$, where a is an arbitrary (possibly infinite) negative constant, $a \leqslant 0$. For any real function $f : \mathbb{R}^n \to \mathbb{R}$ that is continuously differentiable to order $R \geqslant 2$ on $\mathbb{R}^n$, we let $\nabla_{\mathbf{u}} f(\mathbf{u})$ denote the gradient of $f(\cdot)$ with respect to $\mathbf{u}$, $\nabla_{\mathbf{u}} f(\mathbf{u}) \equiv (\partial f(\mathbf{u})/\partial u_i, \ldots, \partial f(\mathbf{u})/\partial u_n)'$, and use $\Delta_{\mathbf{uu}} f(\mathbf{u})$ to denote its Hessian matrix, $\Delta_{\mathbf{uu}} f(\mathbf{u}) \equiv (\partial^2 f(\mathbf{u})/\partial u_i \partial u_j)_{1 \leqslant i,j \leqslant n}$.

For any n-vector $\mathbf{u}$, $\mathbf{u} = (u_1, \ldots, u_n)' \in \mathbb{R}^n$, we denote by $\|\mathbf{u}\|_p$ its l_p-norm, i.e. $\|\mathbf{u}\|_p = (|u_1|^p + \ldots + |u_n|^p)^{1/p}$ for $1 \leqslant p < \infty$, and $\|\mathbf{u}\|_\infty = \max_{1 \leqslant i \leqslant n}(|u_i|)$. We define the open unit ball $\mathcal{B}_p^n$ in $\mathbb{R}^n$ as $\mathcal{B}_p^n = \{u \in \mathbb{R}^n : \|\mathbf{u}\|_p < 1\}$. $\boldsymbol{\nu}_p(\mathbf{u})$, $\mathbf{V}_p(\mathbf{u})$ and $\mathbf{W}_p(\mathbf{u})$ are an n-vector and two $n \times n$-diagonal matrices defined as: $\boldsymbol{\nu}_p(\mathbf{u}) \equiv (\text{sgn}(u_1)|u_1|^{p-1}, \ldots, \text{sgn}(u_n)|u_n|^{p-1})'$,

[5]Following the standard convention, we use bold letters for vectors (e.g., $\mathbf{z}_t$) and matrices (e.g., $\mathbf{B}_0$).

[6]For any $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{2n}$ with $\mathbf{a} = (a_1, \ldots, a_n)'$, $\mathbf{b} = (b_1, \ldots, b_n)'$, $\mathbf{a} \leqslant \mathbf{b}$ means $a_i \leqslant b_i$ for all $1 \leqslant i \leqslant n$.

$\mathbf{V}_p(\mathbf{u}) \equiv \text{diag}(\delta(u_1)|u_1|^{p-1}, ..., \delta(u_n)|u_n|^{p-1})$, and $\mathbf{W}_p(\mathbf{u}) \equiv \text{diag}(|u_1|^{p-2}, ..., |u_n|^{p-2})$, respectively. Then $\nabla_{\mathbf{u}} \|\mathbf{u}\|_p = \|\mathbf{u}\|_p^{1-p} \boldsymbol{\nu}_p(\mathbf{u})$ and $\Delta_{\mathbf{uu}} \|\mathbf{u}\|_p = \|\mathbf{u}\|_p^{1-p} \{2\mathbf{V}_p(\mathbf{u}) + (p-1)\cdot \left[\mathbf{W}_p(\mathbf{u}) - \|\mathbf{u}\|_p^{-p} \boldsymbol{\nu}_p(\mathbf{u})\boldsymbol{\nu}_p'(\mathbf{u})\right]\}$, which we shall often be using in what follows.

Finally, for any $m \times n$-matrix $\mathbf{A} = (a_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$, we let $\|\mathbf{A}\|_\infty = \max_{1\leqslant i,j\leqslant n}(|a_{ij}|)$. Moreover, if $\mathbf{B} = (b_{kl})_{1\leqslant k\leqslant p, 1\leqslant l\leqslant q}$ is a $p \times q$-matrix, we let $\mathbf{A} \otimes \mathbf{B}$, be the direct product of $\mathbf{A}$ and $\mathbf{B}$ (also called their Kronecker product), i.e. $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ is an $(mp) \times (nq)$-matrix with elements defined by $c_{\alpha\beta} = a_{ij}b_{kl}$ where $\alpha = p(i-1) + k$ and $\beta = q(j-1) + l$.

2.2. **Multivariate Loss Function.** In this paper, we generalize the flexible family of loss functions introduced by EKT to n–variate forecasts. In the univariate case, EKT map an exponent p, $1 \leqslant p < \infty$, and an asymmetry parameter α, $0 \leqslant \alpha \leqslant 1$, into a non-negative function of an error $e \in \mathbb{R}$; the resulting family of losses is flexible enough to accommodate the absolute value or quadratic loss, yet allows the latter to become asymmetric. We now extend their definition to a vector-valued argument $\mathbf{e} \in \mathbb{R}^n$. Fix a scalar p, $1 \leqslant p < \infty$, and let $\boldsymbol{\tau}$ be an n-vector with l_q-norm less than unity, i.e. $\boldsymbol{\tau} \in \mathcal{B}_q^n$, where $1/p + 1/q = 1$ with the convention that $q = \infty$ when $p = 1$. For any $\mathbf{e} \in \mathbb{R}^n$, we then define an n-variate loss function as follows:

Definition 1 (n**-variate Loss**). *The n-variate loss function $L_n(p, \boldsymbol{\tau}, \mathbf{e}) : [1, +\infty) \times \mathcal{B}_q^n \times \mathbb{R}^n \to \mathbb{R}$ (with $1/p + 1/q = 1$) is defined as:*

$$L_n(p, \boldsymbol{\tau}, \mathbf{e}) \equiv \left(\|\mathbf{e}\|_p + \boldsymbol{\tau}'\mathbf{e}\right) \|\mathbf{e}\|_p^{p-1}. \tag{1}$$

When $p = 1$ the multivariate loss $L_n(1, \boldsymbol{\tau}, \cdot)$ can be used to define the geometric quantile of the forecast n-vector error $\mathbf{e}$, as proposed in Chaudhuri (1996), for example. In a sense, $L_n(1, \boldsymbol{\tau}, \cdot)$ is a multivariate extension of the univariate loss $2[1 - \alpha + \tau \mathbb{1}(e)]|e|^p$ well-known in the literature on quantile estimation (Koenker and Bassett, 1978). When $p > 1$, the expression of the n-variate loss $L_n(p, \boldsymbol{\tau}, \cdot)$ is entirely novel and not yet seen in the literature. We start by establishing some of its useful properties.

Proposition 1. *Let $L_n(p, \boldsymbol{\tau}, \mathbf{e})$ be the n-variate loss in Definition 1. Then, the following properties hold: (i) $L_n(p, \boldsymbol{\tau}, \cdot)$ is continous and non-negative on $\mathbb{R}^n$; (ii) $L_n(p, \boldsymbol{\tau}, \mathbf{e}) = 0$ if and only if $\mathbf{e} = \mathbf{0}$, and $\lim_{\|\mathbf{e}\|_p \to \infty} L_n(p, \boldsymbol{\tau}, \mathbf{e}) = \infty$; (iii) $L_n(p, \boldsymbol{\tau}, \cdot)$ is convex on $\mathbb{R}^n$.*

The shape of the n-variate loss $L_n(p, \boldsymbol{\tau}, \cdot)$ is characterized by the exponent p, $1 \leqslant p < \infty$, and the n-vector $\boldsymbol{\tau}$ that quantifies the extent of asymmetry in $L_n(p, \boldsymbol{\tau}, \cdot)$. When $\boldsymbol{\tau} = \mathbf{0}$, the n-variate loss in Equation (1) reduces to $\|\mathbf{e}\|_p^p$, which is perfectly symmetric. On the other hand, for a nonzero $\boldsymbol{\tau}$, its magnitude $\|\boldsymbol{\tau}\|_q$ measures the extent of deviation of the n-variate loss from the perfectly symmetric case; the direction of this deviation is determined by the direction of $\boldsymbol{\tau}$. In a sense, both the direction and the magnitude of the n-vector $\boldsymbol{\tau}$ influence the degree of asymmetry in the forecaster's loss (see left and middle panels in Figure 1).

When the variable of interest is of dimension $n = 1$ and the forecasts are univariate, the loss in Equation (1) reduces to $L_1(p, \tau, e) = [|e| + \tau e]\,|e|^{p-1} = 2[1 - \alpha + \tau \mathbb{1}(e)]|e|^p$, where $\tau = 2\alpha - 1$, $\alpha \in (0, 1)$, and $p \geqslant 1$ as previously (Elliott, Komunjer, and Timmermann, 2005, 2006).[7] In the univariate case, this flexible loss family includes: (i) squared loss function $L_1(2, 0, e) = e^2$, (ii) absolute deviation loss function $L_1(1, 0, e) = |e|$, as well as their asymmetrical counterparts obtained when $\tau \neq 0$ (i.e. $\alpha \neq 1/2$) called (iii) quad-quad loss, $L_1(2, \tau, e)$, and (iv) lin-lin loss, $L_1(1, \tau, e)$.

2.3. **Asymmetry and Dependence Properties.** In order to gain further insight into the features of the loss $L_n(p, \boldsymbol{\tau}, \mathbf{e})$ in Equation (1), we consider in more details the case $n = 3$. In this trivariate case, it is assumed that the forecaster cares about the magnitude and the sign of her errors e_1, e_2 and e_3 committed when forecasting jointly the three variables of interest y_1, y_2 and y_3. The iso-loss curves corresponding to $L_3(p, \boldsymbol{\tau}, \mathbf{e}) =$ constant, where $\mathbf{e} = (e_1, e_2, e_3)'$ and $\boldsymbol{\tau} = (2\alpha_1 - 1, 2\alpha_2 - 1, 2\alpha_3 - 1)'$, are then as plotted in the left and middle panels of Figure 1.

For example, when $p = 1$, we have $L_3(1, \boldsymbol{\tau}, \mathbf{e}) = |e_1| + |e_2| + |e_3| + \tau_1 e_1 + \tau_2 e_2 + \tau_3 e_3$ and the loss corresponding to the trivariate error $\mathbf{e} = (e_1, e_2, e_3)'$ equals the sum of individual lin-lin losses corresponding to e_1, e_2 and e_3: $L_3(1, \boldsymbol{\tau}, \mathbf{e}) = L_1(1, \tau_1, e_1) + L_1(1, \tau_2, e_2) + L_1(1, \tau_3, e_3)$. In other words, when the shape parameter $p = 1$, the forecaster behaves as if the variables of interest y_1, y_2 and y_3 were independent. However, her loss in each of the two forecasts taken separately is still allowed to be asymmetric.

[7] Note that we have the following useful identity: $1 + \tau\,\mathrm{sgn}(x) = 2[1 - \alpha + \tau \mathbb{1}(x)]$, for all $x \in \mathbb{R}$.

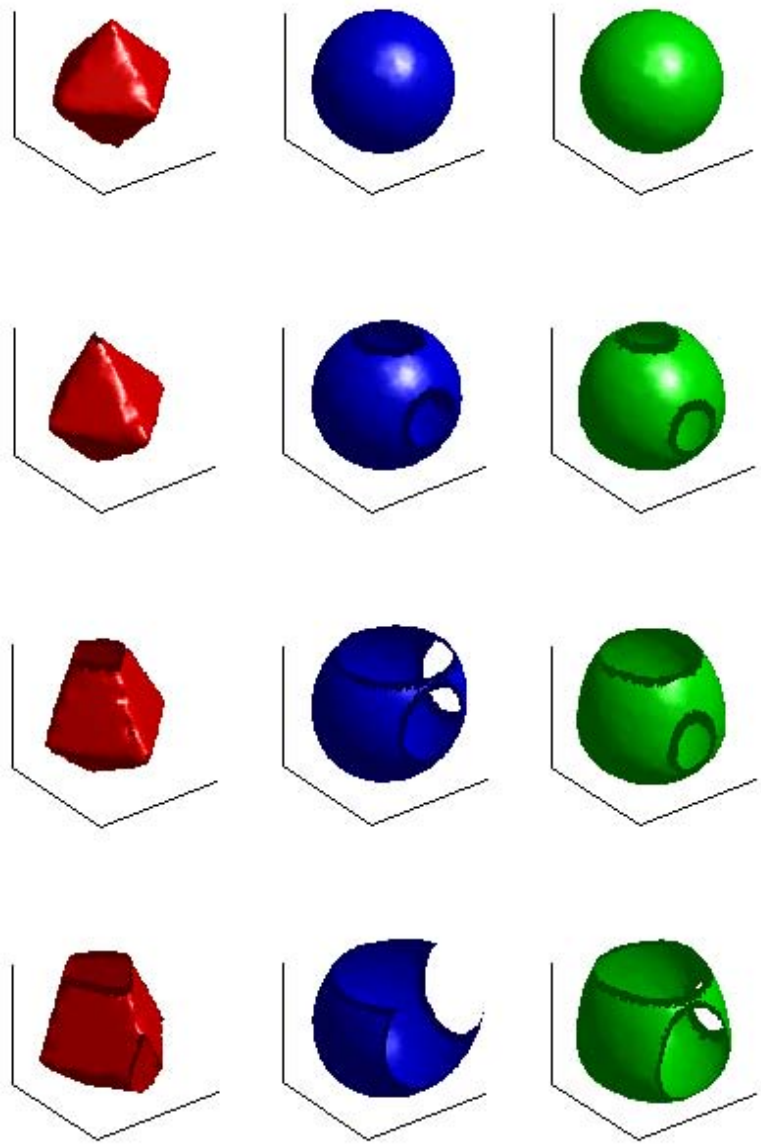

FIGURE 1. Contour plots of: $L_3(1, \boldsymbol{\tau}, \cdot)$ (left), $L_3(2, \boldsymbol{\tau}, \cdot)$ (middle), and $L_1(2, \tau_1, \cdot) + L_1(2, \tau_2, \cdot) + L_1(2, \tau_3, \cdot)$ (right) when $\boldsymbol{\alpha} = (.5, .5, .5)', (.4, .6, .4)', (.4, .6, .25)', (.4, .7, .2)'$ (top to bottom).

When on the other hand $p = 2$, we have $L_3(2, \boldsymbol{\tau}, \mathbf{e}) = |e_1|^2 + |e_2|^2 + |e_3|^2 + (\tau_1 e_1 + \tau_2 e_2 + \tau_3 e_3)\left(|e_1|^2 + |e_2|^2 + |e_3|^2\right)^{1/2}$, which is no longer additive separable in individual losses of either e_1, e_2 or e_3 alone: unless $\tau_1 = \tau_2 = \tau_3 = 0$, we have

$L_3(2, \boldsymbol{\tau}, \mathbf{e}) \neq L_1(2, \tau_1, e_1) + L_1(2, \tau_2, e_2) + L_1(2, \tau_3, e_3)$ for general values of the forecast errors. In this case, the trivariate loss differs from a simple sum of the individual quad-quad losses, as demonstrated by the middle and right panels in Figure 1.

When the shape parameter p of the n-variate loss in Equation (1) is strictly greater than one, $L_n(p, \boldsymbol{\tau}, \mathbf{e})$ will in general differ from the sum of coordinate–wise univariate losses $L_1(p, \tau_1, e_1) + ... + L_1(p, \tau_n, e_n)$. Hence, minimizing the n-variate loss $L_n(p, \boldsymbol{\tau}, \mathbf{e})$ will in general produce an optimal n-vector $\mathbf{e}^*$ whose coordinates e_i^* do not necessarily each minimize $L_1(p, \tau_i, e_i)$. In other words, $L_n(p, \boldsymbol{\tau}, \mathbf{e})$ captures not only the asymmetry but also dependence between different coordinates of $\mathbf{e}$.

3. Multivariate Forecast Rationality: Estimation and Testing

We now define multivariate forecast rationality. Intuitively, the n-variate forecasts shall be called rational with respect to the n-variate loss $L_n(p, \boldsymbol{\tau}, \cdot)$ defined in Equation (1), if they minimize its expected value. Since the information sets available to the forecasters change in time, the expectation of the loss is conditional on $\mathcal{F}_t$; hence, any forecast in the sequence necessarily satisfies an orthogonality condition We shall use this condition as a starting point of our estimation and multivariate forecast rationality testing procedures.

3.1. Rationality Condition.

Throughout the paper we assume that the forecaster's n-vector optimal forecasts of $\mathbf{y}_{t+1}$, forecasts which we denote $\mathbf{f}^*_{t+1,t}$, satisfy the following:

A2. *For all $t = 1, 2, ...$ we have: $\mathbf{f}^*_{t+1,t} = \arg\min_{\{\mathbf{f}_{t+1,t}\}} E\left[L_n(p_0, \boldsymbol{\tau}_0, \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}) | \mathcal{F}_t\right]$, where $L_n(p_0, \boldsymbol{\tau}_0, \cdot)$ is the n-variate loss function with parameters p_0, $1 \leqslant p_0 < \infty$, and $\boldsymbol{\tau}_0 \in \mathcal{B}^n_{q_0}$, $1/p_0 + 1/q_0 = 1$, as defined in Equation (1).*

Implicit in Assumption A2 are two properties: (1) when constructing her optimal forecasts, the forecaster has in mind a loss function whose argument is the forecast error n-vector $\mathbf{e}_{t+1}$ alone; (2) the forecaster's loss is parametrized by p_0 and $\boldsymbol{\tau}_0$ as introduced in Equation (1).

We now derive a necessary and sufficient condition for multivariate forecast rationality, which provides the basis of our test for multivariate forecast rationality (Section 3). We need the following property:

A3. *Given* $p_0 \in [1, +\infty)$*, and for all* $t = 1, 2, ...$ *we have:* $E\big(\|\mathbf{y}_t\|_1^{p_0-1} \big| \mathcal{F}_t\big) < \infty$ *a.s.-P and* $\left\|\mathbf{f}^*_{t+1,t}\right\|_1^{p_0-1} < \infty$ *a.s.-P.*

The conditions in Assumption A3 combined with the convexity of the n-variate loss in Equation (1) together ensure—by Lebesgue's dominated convergence theorem—that we can safely differentiate the loss L_n with respect to the error $\mathbf{e}_{t+1}$ inside the conditional expectation operator in Assumption A2. This yields the following necessary and sufficient condition of multivariate forecast optimality.

Proposition 2. *Let Assumptions A1 and A3 hold. Then the optimality of* $\{\mathbf{f}^*_{t+1,t}\}$ *in A2 holds if and only if for all* $t = 1, 2, \ldots$ *we have:*

$$E\left[p_0 \boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) + \boldsymbol{\tau_0} \left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} + (p_0-1)\tau_0' \mathbf{e}^*_{t+1} \frac{\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})}{\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}} \middle| \mathcal{F}_t\right] = \mathbf{0}, a.s. - P. \quad (2)$$

Note that while the necessity of the above first order condition is obvious, the sufficiency part of the above result heavily relies on the convexity of the loss $L_n(p_0, \boldsymbol{\tau}_0, \cdot)$ established in Proposition 1.

3.2. **Identification of Multivariate Loss Function Parameters.** Identification of the true multivariate loss parameters used by the forecasters exploits the orthogonality condition derived in Proposition (2). Consider an $\mathcal{F}_t$-measurable d-vector $\mathbf{x}_t$, and denote by $\mathbf{g}(\cdot, \cdot; \mathbf{e}^*_{t+1}, \mathbf{x}_t)$ an nd-vector-valued function $\mathbf{g}(\cdot, \cdot; \mathbf{e}^*_{t+1}, \mathbf{x}_t) : [1, +\infty) \times \mathcal{B}_q^n \to \mathbb{R}^{nd}$ such that

$$\mathbf{g}(p, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t) \equiv \Big(p\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1}) + \boldsymbol{\tau} \left\|\mathbf{e}^*_{t+1}\right\|_p^{p-1} + (p-1)\boldsymbol{\tau}' \mathbf{e}^*_{t+1} \left\|\mathbf{e}^*_{t+1}\right\|_p^{-1} \boldsymbol{\nu}_p(\mathbf{e}^*_{t+1})\Big) \otimes \mathbf{x}_t. \quad (3)$$

The key element of our identification strategy is the following: under rationality, we have $E[\mathbf{g}(p_0, \tau_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = \mathbf{0}$ for all $t = 1, 2, \ldots$. If for a given $p_0 \in [1, +\infty)$, τ_0 is the unique value of the n-variate asymmetry parameter $\tau \in \mathcal{B}_{q_0}^n$ (with $1/p_0 + 1/q_0 = 1$) that satisfies those nd orthogonality conditions, then the latter can be utilized by the forecast evaluator (econometrician) to estimate $\boldsymbol{\tau}_0$ using Hansen's (1982) GMM approach. For this, we shall first make the assumption that all of the variables appearing in Equations (9) and (3) come from a process that is stationary:

A4. *The process* $\{(\mathbf{e}^{*\prime}_{t+1}, \mathbf{x}'_t)'\}$ *is strictly stationary.*

We further restrict the d-vector of instruments $\mathbf{x}_t$ to satisfy the following properties:

A5. *Given* $p_0 \in [1, +\infty)$, *(i)* $E[\left\|\mathbf{e}^*_{t+1}\right\|_1^{p_0-1} \|\mathbf{x}_t\|_1] < \infty$; *(ii)* $\operatorname{rank} E\big[\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0-1)\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1} (\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}\big] = n$.

We can then characterize the true forecaster's asymmetry parameter $\boldsymbol{\tau}_0$ as follows.

Lemma 3. *Let Assumptions A1 through A5 hold. Given* $p_0 \in [1, +\infty)$ *and for any* $\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}$, *let* $Q(\boldsymbol{\tau}) \equiv E[\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$, *with* $\mathbf{S}$ *positive definite. Then* $\boldsymbol{\tau}_0$ *is the unique minimum of* $Q(\boldsymbol{\tau})$ *on* $\mathcal{B}^n_{q_0}$.

The weighting matrix $\mathbf{S}$ in Lemma 3 is usually set to be equal to: $\mathbf{S} \equiv E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)']$. In order to ensure that $\mathbf{S}$ is positive definite, we need that the covariance matrix of d-vector of instruments $\mathbf{x}_t$ be of full rank.

A6. $\operatorname{rank} E[\mathbf{x}_t\mathbf{x}_t'] = d$.

We then have the following result:

Lemma 4. *Let Assumptions A1 through A6 hold. Given* $p_0 \in [1, +\infty)$ *and for any* $\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}$, *let* $\mathbf{S}(\boldsymbol{\tau}) \equiv E\left[\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)'\right]$. *Then* $\mathbf{S}(\boldsymbol{\tau})$ *is positive definite.*

3.3. **GMM Estimation.** Now, given $p_0 \in [1, +\infty)$ and given the observations $((\mathbf{x}'_\tau, \hat{\mathbf{e}}'_{\tau+1})', ..., (\mathbf{x}'_{\tau+T-1}, \hat{\mathbf{e}}'_{\tau+T})')'$, the GMM estimator of the n-variate loss asymmetry parameter $\boldsymbol{\tau}_0$, denoted $\hat{\boldsymbol{\tau}}_T$, can be defined as a solution to the minimization problem:

$$\min_{\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}} \left[T^{-1} \sum_{t=R}^{T+R-1} \mathbf{g}(p_0, \boldsymbol{\tau}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)\right]' \hat{\mathbf{S}}^{-1} \left[T^{-1} \sum_{t=R}^{T+R-1} \mathbf{g}(p_0, \boldsymbol{\tau}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)\right], \tag{4}$$

where $\hat{\mathbf{S}}$ is a consistent estimator of $\mathbf{S} = E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)']$.

It is worth pointing out that the optimal forecast errors $\mathbf{e}^*_{t+1}$ are unobservable in reality. Instead, for every t, $R \leqslant t \leqslant T+R-1$, the forecast evaluator observes $\hat{\mathbf{e}}_{t+1,t} = \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}$, which implicitly incorporates all of the forecast estimation uncertainty embodied in $\mathbf{f}_{t+1,t}$. In order to make sure that this uncertainty does not interfere with our rationality test we need to impose a set of restrictions on how the observed forecaster's n-vector errors $\{\mathbf{e}_{t+1,t}\}_{t=R}^{T+R-1}$ differ from their optimal counterparts $\{\mathbf{e}^*_{t+1}\}_{t=R}^{T+R-1}$.

A7. *For every t, $R \leq t \leq T+R-1$, and any $\varepsilon > 0$,* $\lim_{R,T\to\infty} \Pr\left(\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 > \varepsilon\right) = 0$.

In addition, we need to ensure that appropriate sample averages converge to their expected values. Recall that Assumption A4 restricts the heterogeneity of the process $\{(\mathbf{e}^{*\prime}_{t+1}, \mathbf{y}'_t)'\}$ by guaranteeing that the latter is strictly stationary. We now impose a similar condition on $\{(\hat{\mathbf{e}}'_{t+1}, \mathbf{x}'_t)'\}$ and further restrict its dependence structure.

A8. *The process $\{(\hat{\mathbf{e}}'_{t+1}, \mathbf{x}'_t)'\}$ is strictly stationary and α-mixing with mixing coefficient α of size $-r/(r-2)$, $r > 2$, and, given $p_0 \in [1, +\infty)$, there exist some $\varepsilon > 0$, $\Delta_1 > 0$ and $\Delta_2 > 0$ such that $E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p_0-1)(2r+\varepsilon)}] \leqslant \Delta_1 < \infty$ and $E[\|\mathbf{x}_t\|_1^{2r+\varepsilon}] \leqslant \Delta_2 < \infty$.*

In particular, using the fact that $\{\mathbf{g}(p_0, \tau_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t), \mathcal{F}_t\}$ is a martingale difference sequence, as shown in Equation (2), a consistent estimator of $\mathbf{S}$ is given by

$$\hat{\mathbf{S}}(\tilde{\boldsymbol{\tau}}) \equiv T^{-1} \sum_{t=R}^{T+R-1} \mathbf{g}(p_0, \widetilde{\boldsymbol{\tau}}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)\mathbf{g}(p_0, \widetilde{\boldsymbol{\tau}}; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)', \tag{5}$$

where $\widetilde{\boldsymbol{\tau}}$ is some initial consistent estimate of $\boldsymbol{\tau}_0$. As already pointed out, the optimal sequence $\{\mathbf{e}^*_{t+1}\}$ is in reality unobservable; what the forecast evaluator (econometrician) observes instead are the forecaster's n-vector errors $\{\hat{\mathbf{e}}_{t+1,t}\}_{t=R}^{T+R-1}$. Given that the forecaster produces forecasts which are "close" to optimal as quantified in Assumption A7, the consistency of $\hat{\mathbf{S}}$ in Equation (5) holds, despite the presence of the forecast estimation uncertainty.

We are now able to show that our GMM estimator $\hat{\boldsymbol{\tau}}_T$ of the asymmetry parameter $\boldsymbol{\tau}_0$ is consistent:

Theorem 5. *Let Assumptions A1 through A8 hold. Then, given $p_0 \in [1, +\infty)$ we have $\hat{\boldsymbol{\tau}}_T \xrightarrow{p} \boldsymbol{\tau}_0$ as $(R, T) \to \infty$.*

3.4. **Forecast Rationality Test Statistic.** Our test for multivariate forecast rationality comes in a form of a J-test. Hence, it necessitates the derivation of the asymptotic distribution of our GMM estimator $\hat{\boldsymbol{\tau}}_T$ which we derive next. We start by strengthening our stationarity assumption A4 as follows:

A4'. *The process* $\{(\mathbf{e}_{t+1}^{*\prime}, \mathbf{x}_t')'\}$ *is strictly stationary and* α*-mixing with mixing coefficient* α *of size* $-r/(r-2)$, $r > 2$, *and, given* $p_0 \in [1, +\infty)$, *there exist some* $\epsilon > 0$, $\Delta_3 > 0$ *such that* $E[\left\|\mathbf{e}_{t+1}^*\right\|_1^{(p_0-1)(2r+\epsilon)}] \leqslant \Delta_3 < \infty$.

Above conditions, similar to those stated in Assumption A8, ensure that appropriate laws of large numbers and central limit theorems apply. We shall also strengthen our assumption A7 by requiring the following:

A7'. *(i) For some small* ε *in* $(0,1)$ $R^{1-2\varepsilon}/T \to \infty$ *as* $R \to \infty$ *and* $T \to \infty$*; (ii) for any* $\delta > 0$ *we have:* $\lim_{R,T\to\infty} \Pr\left(\sup_{R\leqslant t\leqslant T+R-1} \left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 > \delta\right) = 0$.

The above condition ensures that the forecast estimation uncertainty, embodied in $\hat{\mathbf{e}}$, does not affect the asymptotic distribution of our GMM estimator $\hat{\boldsymbol{\tau}}_T$. Note that A7'(i) imposes a condition on the relative growth of sample sizes R and T. Assumption A7'(ii), on the other hand, strengthens the requirement in A7 by making it uniform in t. Finally, we need two additional new assumptions:

A9. *Given* $p_0 \in [1, +\infty)$, *we have:* $E\left(\sup_{c\in(0,1)} \left\|c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}_{t+1}^*\right\|_1^{p_0-2}\right) < \infty$ *and* $E\left(\left\|\mathbf{x}_t\right\|_1 \sup_{c\in(0,1)} \left\|c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}_{t+1}^*\right\|_1^{p_0-2}\right) < \infty$.

A10. *The marginal densities* $f_{it}^0(\cdot)$ *are such that* $\max_{1\leq i\leq n} f_{it}^0(y) \leq M$ *for any* $y \in \mathbb{R}$.

We are now ready to state our asymptotic distribution result for $\hat{\boldsymbol{\tau}}_T$.

Theorem 6. *Let Assumptions A1-A3, A4', A5-A6, A7', A8-A10 hold. Then, given* $p_0 \in [1, +\infty)$ *we have:* $\sqrt{T}(\hat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1})$, *as* $R, T \to \infty$, *where* $\mathbf{S} = E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t)\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t)']$ *and* $\mathbf{B}^* \equiv E[\left\|\mathbf{e}_{t+1}^*\right\|_{p_0}^{p_0-1}(\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0 - 1)\left\|\mathbf{e}_{t+1}^*\right\|_{p_0}^{-1}(\boldsymbol{\nu}_{p_0}(\mathbf{e}_{t+1}^*) \otimes \mathbf{x}_t)\mathbf{e}_{t+1}^{*\prime}]$.

The asymptotic normality result of Theorem 6 is the basis for our forecast rationality test. When the dimension of the d-vector $\mathbf{x}_t$ used in Equation (3) is large enough, $d > 1$, then a test for overidentification provides a joint test of rationality of the n-vector forecasts $\{\hat{\mathbf{f}}_{t+1,1}\}$ under the n-variate loss $L(p_0, \hat{\boldsymbol{\tau}}_T, \cdot)$. More formally, we have the following Corollary to our Theorem 6:

Corollary 7. *Let the assumptions of Theorem 6 hold. Then a joint test of n-vector forecast rationality under the n-variate loss function $L(p_0, \hat{\boldsymbol{\tau}}_T, \cdot)$ can be conducted with $d > 1$ instruments $\mathbf{x}_t$ through the test statistic*

$$\hat{J}_T \equiv T^{-1} \left[\sum_{t=R}^{R+T-1} \mathbf{g}(p_0, \hat{\boldsymbol{\tau}}_T; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) \right]' \hat{\mathbf{S}}^{-1} \left[\sum_{t=R}^{R+T-1} \mathbf{g}(p_0, \hat{\boldsymbol{\tau}}_T; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) \right] \sim \chi^2_{n(d-1)},$$

where $\hat{\mathbf{S}}$ is as defined in Equation (5).

4. Empirical Application

We illustrate the performance of our multivariate forecast rationality test in a situation in which the forecasters have reason to believe that the forecasts are codependent. We focus on three macro variables: the growth rate in output y, the CPI inflation rate π, and a short-term interest rate r. Examples of models using these variables include Taylor's (1993) interest rate targeting rule, monetary VARs (Christiano, Eichenbaum, and Evans, 1999), optimizing ISLM models (McCallum and Nelson, 1999), and reduced-form New Keynesian models (Clarida, Gali, and Gertler, 2000). Common to these models is a relationship—either estimated or imposed—between output and prices combined with the Federal Reserve's control of short-term interest rates. We would therefore expect the forecasters to account for the covariation of output, prices, and interest rates when constructing their optimal forecasts.

4.1. **Data.** Forecast data are taken from the Blue Chip Economic Indicators (BCEI), a compilation of industry forecasts of a number of economic variables. Each month, participating firms report forecasts of the current- or next-year growth rate in output and prices and the current- or next-year average short-term interest rate.[8] The BCEI began collecting data in 1976:08. Our sample includes forecasts through 2004:12.[9]

[8]Prior to 1984, firms reported current-year forecasts for the first five or six months of the year. In later months, they reported next-year forecasts. Starting in 1984, both current- and next-year forecasts were reported each month.

[9]The sample of output forecasts is split between GNP (1976:08 through 1991:12) and GDP (1992:01 through 2004:12). The BCEI began collecting CPI inflation forecasts in 1979:01 through the end of our sample in 2004:12. The short-term interest rate forecasts are split between the 3-month commercial paper

We assume that the forecaster's objective is to predict true values and that revisions to the realizations are a more accurate reflection of the true values. Thus, in constructing the forecast errors, we use the latest revision of the variable in question. The realizations are yearly growth rates of GDP, GNP, and CPI inflation.[10] Short-term interest rate realizations are the yearly average.

Over time, some forecasters leave the sample while others are added. In addition, firms occasionally fail to report forecasts for any given month. We therefore omit any observation in which forecasts for all three variables are not reported.[11] Finally, forecasters with fewer than 100 valid observations are dropped from the sample. For the full sample, this leaves 57 firms with an average of 171 valid observations per firm.

Table 1 (Instrument Sets)

As we have shown above, rationality depends on the set of variables included in the forecaster's information set. The set of instruments includes combinations of the lagged growth rates of output, inflation, the unemployment rate, and the short-term interest rate. Instruments are, for each month, a snapshot of the real-time data available at that time.[12] The change in the forecast is also included as a possible instrument. For each forecaster, we conduct tests employing different information sets, described in Table 1. As a baseline for comparison, we repeat each test under the assumption of independence and the joint assumption of independence and symmetry.

(1976:08 through 1980:06), the 6-month commercial paper (1980:07 through 1981:12), and the 3-month T-bill (1982:01 through 2004:12) rates.

[10]For output and inflation, the target variable is the rate of change between the average of the levels for that year. This method is described by the BCEI in their monthly newsletter.

[11]These observations may affect both the period in which the forecast is made and the information set of the forecaster. In these cases, both observations are omitted.

[12]These data are taken from the Federal Reserve Bank of St. Louis's archival dataset Archival Federal Reserve Economic Data (ALFRED), available at www.stlsfrb.org. The short-term interest rate, which is not typically revised, was taken from the Federal Reserve Board of Governors.

4.2. **Multivariate Rationality Test Results.** The multivariate rationality test outlined in the previous sections is essentially a test of overidentifying restrictions. It examines whether the series of forecast errors can be reconciled with rationality for some set of asymmetry parameters.

Table 2: Results of Rationality Tests, include the baseline univariate case with symmetry

Table 2 illustrates the effect of testing rationality jointly. We report the percentage of forecasts for which rationality is not rejected. Results are reported for three confidence levels—90, 95 and 99 percent—for each set of instruments. In addition, each instrument set is estimated for fixed values of $p = 1, 2$.[13] The salient result is that for each instrument set, both the univariate asymmetric and multivariate asymmetric loss functions accept rationality at a much higher rate than the univariate symmetric baseline reported in the last three columns. The rate at which rationality is accepted under multivariate loss is nearly identical to that under univariate asymmetric loss. For most instrument sets, the difference in acceptances between flexible loss methods is smaller than 10 percent.[14]

4.3. **Asymmetry Coefficients.** For a given specification of the forecaster's loss function, our procedure delivers estimates of the asymmetry parameters $(\alpha^y, \alpha^\pi, \alpha^r)$ most consistent with the orthogonality conditions implied by rationality of joint forecasts of y, π and r.

Table 3 (Loss Function Coefficients)

EKT found that the addition of symmetric loss alone can increase the rate at which rationality is confirmed in forecasters. However, this finding often requires substantial directional asymmetry in the forecasters' loss functions. Accounting for the codependence of the forecasted variables may mitigate this problem. Recall that interpretation of the asymmetry parameters $(\alpha^y, \alpha^\pi, \alpha^r)$ depends on their values relative to the baseline

[13]Recall from the discussion above that for $p = 1$, the univariate and multivariate cases are equivalent. Thus, results of both the rationality tests and the sets of asymmetry parameters are identical.

[14]Differences in the rate of acceptance of rationality between the univariate and multivariate approaches when varying p were minor.

0.5. Values greater (less) than 0.5 indicate greater losses for negative (positive) forecast errors. Table 3 provides summary statistics for the cross-forecaster distributions of the estimated asymmetry parameters for different instrument sets. Figure 2 provides graphical representations of these distributions for a subset of instruments.

The joint directionality in preferences appears pervasive across forecasters. More than half of the forecasters exhibit higher loss when jointly overpredicting output, overpredicting the short-term interest rate, and underpredicting inflation. Figure 3 presents a three-dimensional representation of the forecasters' joint preference parameters. These directionality preferences are each associated with an unexpectedly worse economic outcome, i.e., lower-than-expected output growth, looser-than-expected monetary policy, and higher-than-expected inflation.

In addition, the asymmetry for each forecaster is typically preserved when the loss is estimated jointly rather than independently. That is, if the asymmetry parameters indicate that a forecaster has a preference for overpredicting GDP under independence, joint estimation of her loss function does not tend to reverse this preference. For each forecasted variable, only about 10 percent of the total number of forecasters experience preference reversals, with the majority of these being statistically-indistinguishable from symmetric loss.

The salient result for multivariate rationality lies in the difference between the estimated loss function parameters. We find that the degree of directional asymmetry is reduced once independence is relaxed. Figure 4 plots the ratio of the absolute deviation from symmetry ($\alpha^i = 0.5$) for the multivariate case to the univariate case. With few exceptions, the distribution of this ratio across forecasters lies below 1, indicating the decline in the estimated asymmetry when accounting for the comovement of variables. Neglecting the comovement of variables then leads the econometrician to assume more directional asymmetry than may actually be warranted.

4.4. **Pre- vs. post-1994 dependence structure.** We have observed that accounting for the comovement of variables may lead to a decline in the estimated asymmetric preferences of forecasters. However, the main advantage of a multivariate approach may be

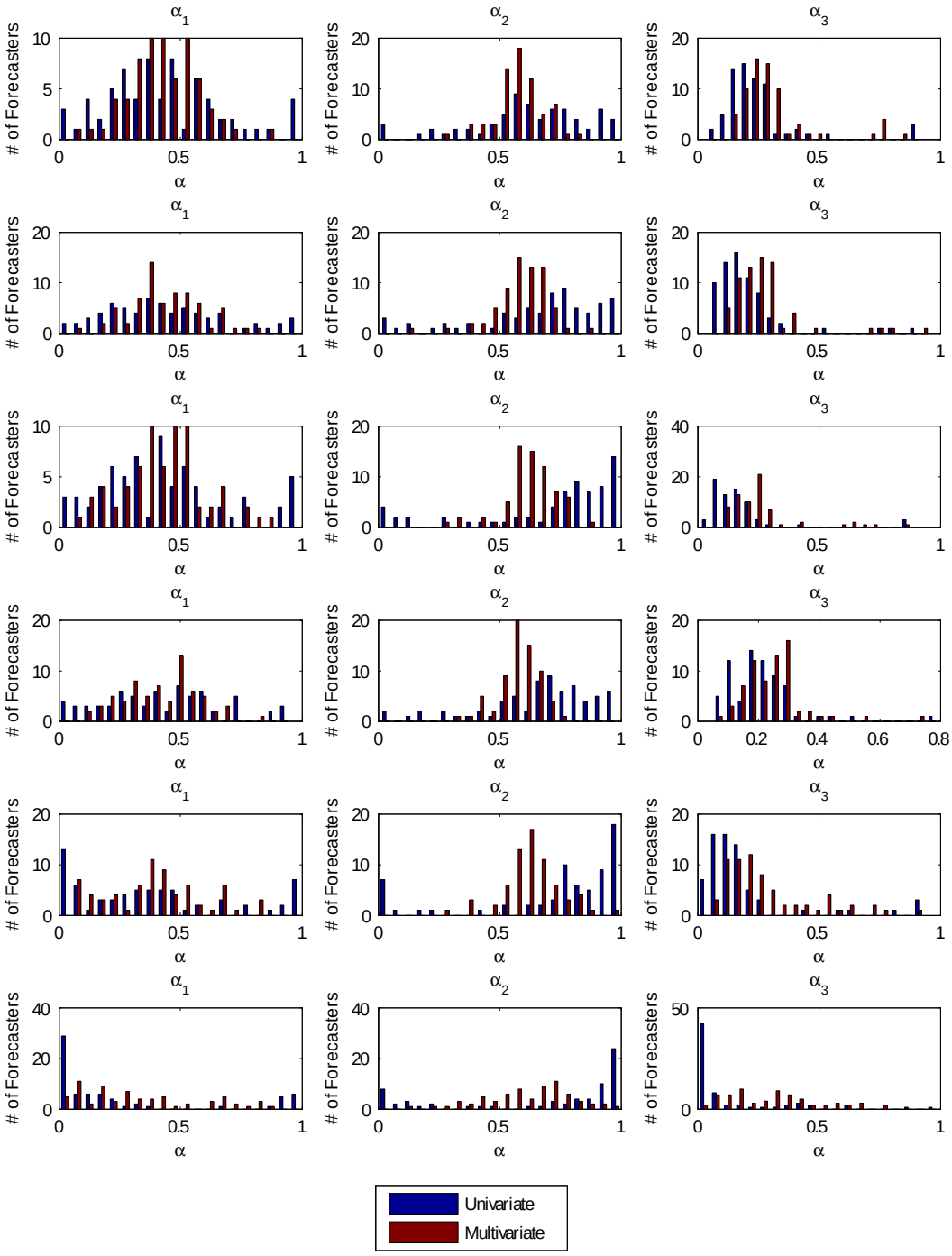

FIGURE 2. Histograms

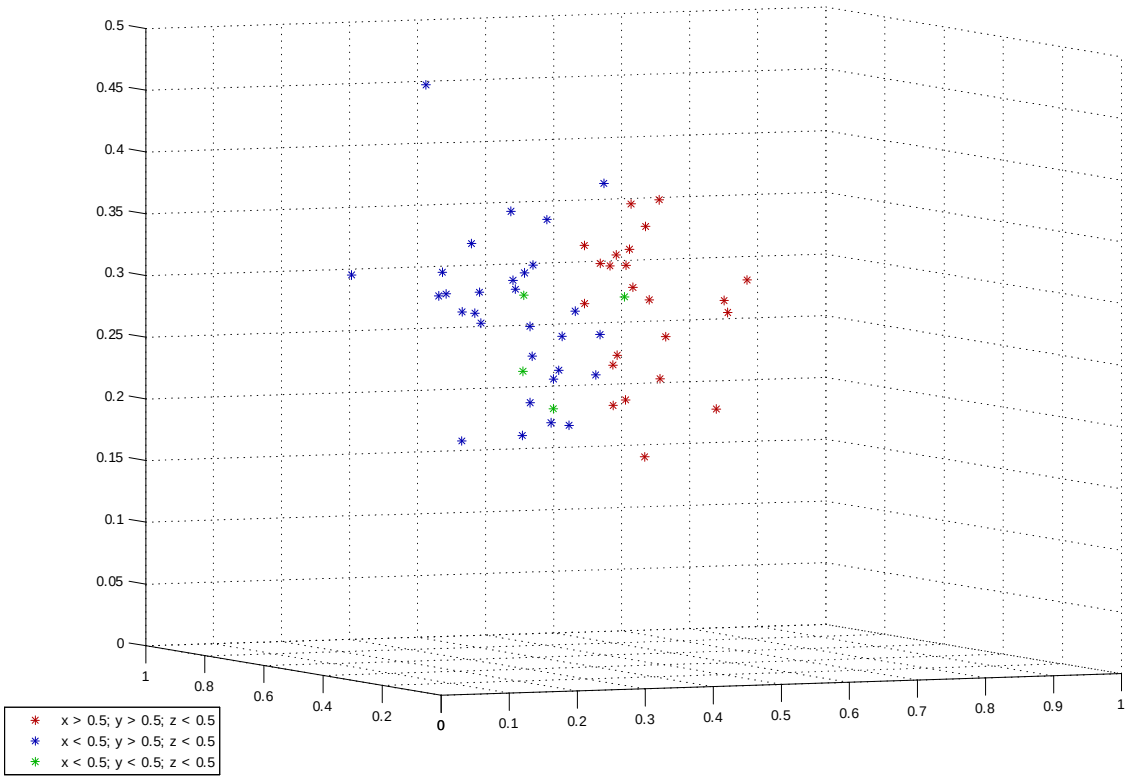

FIGURE 3. 3D Plot

revealed by changes in the relationships between correlated variables. An example of such a change occurred when the Federal Reserve began releasing statements describing policy actions in early 1994.[15] This shift in policy has been thought to cause changes in the behavior of forecasters, possibly increasing the information of private forecasters.[16] Here, we consider whether accounting for this innovation in Fed policy alters the rationality results and/or directional preferences of forecasters. To accomplish this, we reestimate the forecasters' loss functions over the split sample periods up to and beginning in 1994.

Table 4 (Subsample Estimates)

[15]Poole and Rasche (2003) and Eijffinger, Geraats, and van der Cruijsen (2006) documented this change in the Fed's transparency dubbed "the announcement period".

[16]Swanson (2006) suggests an increase in forecasting accuracy of the private sector during the announcement period.

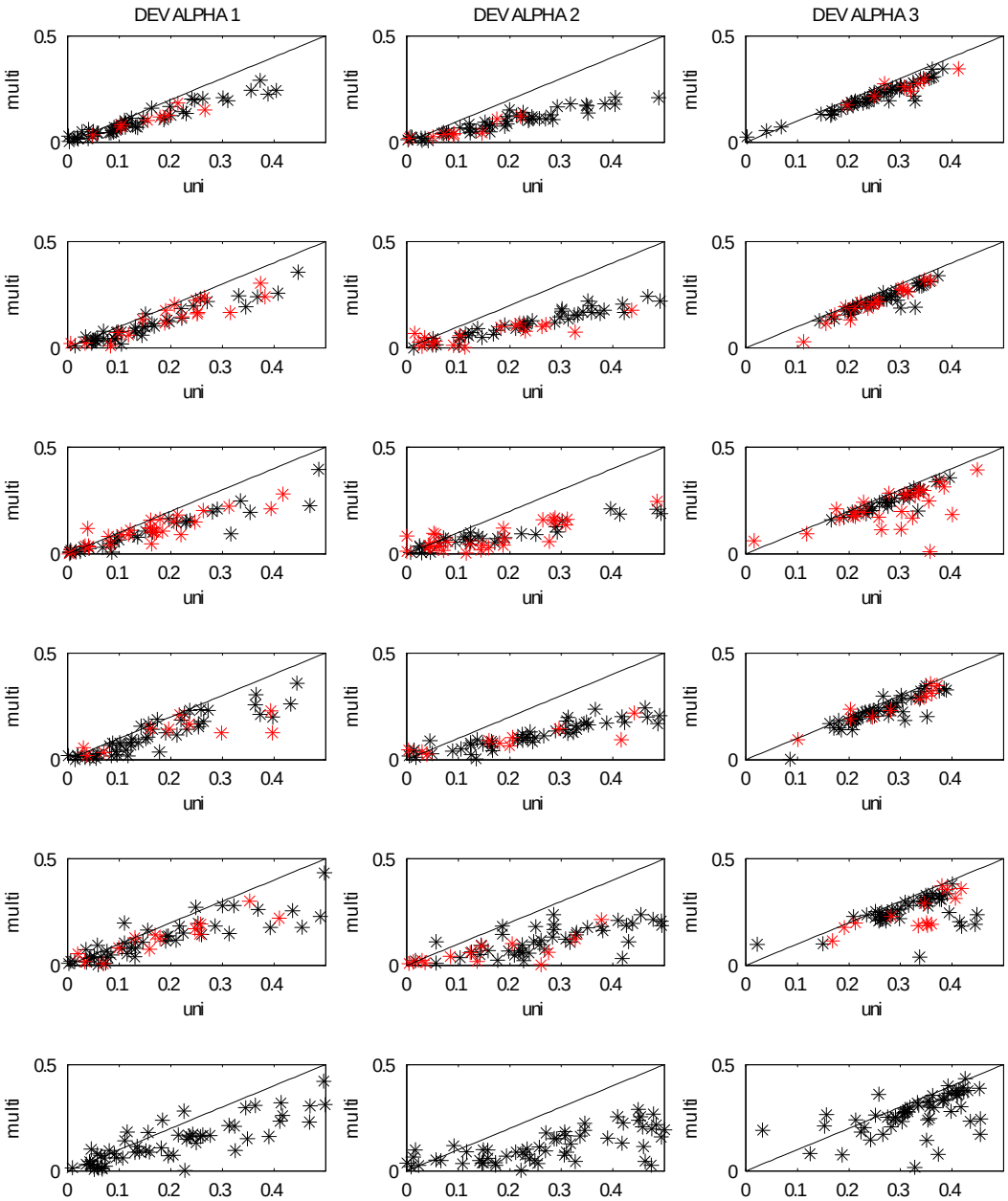

FIGURE 4. Proximity to Symmetry

Table 4 reports the asymmetry coefficients from the announcement subsample alongside results for the pre-1994 sample and the full sample.[17] Two results are readily apparent.

[17]Reducing the sample size reduces the number of forecasters eligible for consideration. Prior to 1994, 34 forecasters had at least 100 valid observations. For the post-1994 sample period, 32 forecasters had

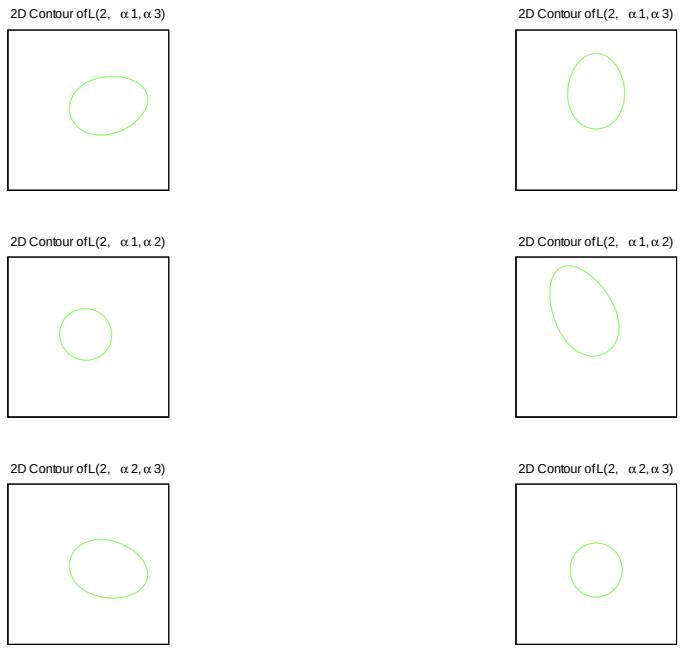

FIGURE 5. Iso-loss contours

First, cross-period results suggest a change in forecaster behavior at the onset of the announcement period. Prior to 1994, forecasters appeared to exhibit more sensitivity to overshooting interest rates. During the same period, forecasters appeared virtually symmetric to output forecast errors. On the other hand, during the announcement period, forecasters receive signals from the Fed about future policy actions. This appears to shift their preferences toward higher losses for overshooting output.

Comparing across frameworks, however, reveals that the multivariate approach assigns less asymmetry to short-term interest rate errors than if the errors were assumed independent. During the announcement period, the forecasting environment and, thus, the relationship between these three variables changed substantially. The differences across methodologies are depicted in Figure 5, which shows cross-sections of the iso-loss contours. In the first panel, forecasters have symmetric preferences as to output and inflation but

at least 100 valid observations. These sets of forecasters overlap but are not identical. In this section, results for the full sample are for the same 32 forecasters with 100 valid observations in the announcement period.

have extreme directional preference for interest rate errors. The second panel shows the change in the iso-loss contours as preferences become more symmetric for interest rates but less symmetric for output.

If the forecaster's preferences are assumed to be determined by the sum of independent losses, the interaction between output, inflation, and interest rates is neglected. In this case, the econometrician would account for biases in the short-term interest rate forecasts by assigning more asymmetric loss to these forecasts. However, accounting for the comovements reduces the estimated asymmetry in forecaster losses, particularly for the short-term interest rate. For the 32 forecasters in our sample, the reduction in asymmetry for overshooting the short rate is up to 50 percent. The multivariate approach takes into account the fact that overpredicting the short rate may often be associated with overpredicting output and underpredicting inflation. Once this type of systematic covariation is controlled for, forecaster losses may appear dramatically less asymmetric.

Table 5 about here

As a final test, Table 5 presents the results of tests against symmetry for various combinations of the forecaster's preference parameters. The table shows the percentage of forecasters for which the Wald test cannot reject the given null hypothesis. While many of the columns indicate the differences across methodologies are small, we focus on the restriction $\alpha^r = 0.5$. Here, we can clearly see that accounting for covariation substantially increases the number of forecasters for which interest rate symmetry cannot be rejected.

5. Conclusions and Implications for Rationality

The results of the preceding tests have important implications for the prospects of rational expectations in macroeconomic models. In univariate tests, EKT argue that rationality requires the econometrician to allow forecasters to have asymmetric loss across directional errors for output and inflation. These conclusions are drawn from a model considering the forecasted series in isolation. Our multivariate tests indicate that asymmetric loss for output and inflation may an aberration rather than the norm. These findings show that symmetric rationality over output and inflation is the predominant finding once

the econometrician accounts for forecast errors in the short-term interest rates. In other words, imposing zero correlation between the three variables leads to a misspecification that biases the result toward asymmetry.

From a macroeconomic point of view, the preceding argument amounts to the following conclusion: agents account for monetary policy when establishing their forecasts for output and inflation.[18] Neglecting the correlations in the forecast errors for these variables is akin to the assumption that output, inflation, and monetary policy are independent. Our findings suggest that, in light of the forecasters' expectation of future monetary policy, their predictions for output and inflation appear rational with less directional asymmetry. One final concern, however, is the rate at which directional asymmetry for short-term interest rates is rejected even in the multivariate framework. A number of alternatives to true directional asymmetry can be posited. For example, the loss function may still be misspecified if key correlations are omitted. A second possibility is that the asymmetry is produced by the process by which monetary policy is conducted, i.e., monetary policy tightenings are more predictable than easings.[19]

Appendix A. Assumptions

A1. *For all* $t = 1, 2, \ldots$ *the conditional distribution* $F_t^0(\cdot)$ *is continuously differentiable and the corresponding conditional density* $f_t^0(\cdot) > 0$ *on* $\mathbb{R}^n$.

A2. *For all* $t = 1, 2, \ldots$ *we have:* $\mathbf{f}_{t+1,t}^* = \arg\min_{\{\mathbf{f}_{t+1,t}\}} E\left[L_n\left(p_0, \boldsymbol{\tau}_0, \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}\right) \middle| \mathcal{F}_t\right]$, *where* $L_n\left(p_0, \boldsymbol{\tau}_0, \cdot\right)$ *is the* n*-variate loss function with parameters* p_0, $1 \leqslant p_0 < \infty$, *and* $\boldsymbol{\tau}_0 \in \mathcal{B}_{q_0}^n$, $1/p_0 + 1/q_0 = 1$, *as defined in Equation* (1).

A3. *Given* $p_0 \in [1, +\infty)$, *and for all* $t = 1, 2, \ldots$ *we have:* $E\left(\|\mathbf{y}_t\|_1^{p_0-1} \middle| \mathcal{F}_t\right) < \infty$ *a.s.-P and* $\left\|\mathbf{f}_{t+1,t}^*\right\|_1^{p_0-1} < \infty$ *a.s.-P.*

A4. *The process* $\{(\mathbf{e}_{t+1}^{*\prime}, \mathbf{x}_t')'\}$ *is strictly stationary.*

---

[18] In this, we treat the short-term interest rate as a proxy for the policy instrument.

[19] We conducted a back-of-the-envelope test of this hypothesis by omitting from the estimation months in which the FOMC made an intermeeting move. We found no significant differences in either the rates of rationality rejections or the distributions of the asymmetry parameters.

A4'. *The process* $\{(\mathbf{e}_{t+1}^{*\prime}, \mathbf{x}_t')'\}$ *is strictly stationary and* α*-mixing with mixing coefficient* α *of size* $-r/(r-2)$, $r > 2$, *and, given* $p_0 \in [1, +\infty)$, *there exist some* $\epsilon > 0$, $\Delta_3 > 0$ *such that* $E[\left\|\mathbf{e}_{t+1}^*\right\|_1^{(p_0-1)(2r+\epsilon)}] \leqslant \Delta_3 < \infty$.

A5. *Given* $p_0 \in [1, +\infty)$, *(i)* $E[\left\|\mathbf{e}_{t+1}^*\right\|_1^{p_0-1} \|\mathbf{x}_t\|_1] < \infty$; *(ii)* $\operatorname{rank} E[\left\|\mathbf{e}_{t+1}^*\right\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0-1) \left\|\mathbf{e}_{t+1}^*\right\|_{p_0}^{-1} (\boldsymbol{\nu}_{p_0}(\mathbf{e}_{t+1}^*) \otimes \mathbf{x}_t)\mathbf{e}_{t+1}^{*\prime}] = n$.

A6. $\operatorname{rank} E[\mathbf{x}_t\mathbf{x}_t'] = d$.

A7. *For every* t, $R \leq t \leq T+R-1$, *and any* $\varepsilon > 0$, $\lim_{R,T\to\infty} \Pr\left(\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 > \varepsilon\right) = 0$.

A7'. *(i) For some small* ε *in* $(0,1)$ $R^{1-2\varepsilon}/T \to \infty$ *as* $R \to \infty$ *and* $T \to \infty$*; (ii) for any* $\delta > 0$ *we have:* $\lim_{R,T\to\infty} \Pr\left(\sup_{R\leqslant t\leqslant T+R-1} \left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 > \delta\right) = 0$.

A8. *The process* $\{(\hat{\mathbf{e}}_{t+1}', \mathbf{x}_t')'\}$ *is strictly stationary and* α*-mixing with mixing coefficient* α *of size* $-r/(r-2)$, $r > 2$, *and, given* $p_0 \in [1, +\infty)$, *there exist some* $\varepsilon > 0$, $\Delta_1 > 0$ *and* $\Delta_2 > 0$ *such that* $E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p_0-1)(2r+\varepsilon)}] \leqslant \Delta_1 < \infty$ *and* $E[\|\mathbf{x}_t\|_1^{2r+\varepsilon}] \leqslant \Delta_2 < \infty$.

A9. *Given* $p_0 \in [1, +\infty)$, *we have:* $E\left(\sup_{c\in(0,1)} \left\|c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}_{t+1}^*\right\|_1^{p_0-2}\right) < \infty$ *and* $E\left(\|\mathbf{x}_t\|_1 \sup_{c\in(0,1)} \left\|c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}_{t+1}^*\right\|_1^{p_0-2}\right) < \infty$.

A10. *The marginal densities* $f_{it}^0(\cdot)$ *are such that* $\max_{1\leq i\leq n} f_{it}^0(y) \leq M$ *for any* $y \in \mathbb{R}$.

Appendix B. Proofs

Proof of Proposition 1. Fix p, $1 \leqslant p < \infty$, $\boldsymbol{\tau} \in \mathcal{B}_q^n$ $(1/p + 1/q = 1)$, and consider the n-variate loss function $L(p, \boldsymbol{\tau}, \cdot) : \mathbb{R}^n \to \mathbb{R}$ as in Definition 1. That $L(p, \boldsymbol{\tau}, \cdot)$ is continuous on $\mathbb{R}^n$ follows by the continuity of the p-norm $\mathbf{e} \mapsto \|\mathbf{e}\|_p$ and the Euclidean inner product $\mathbf{e} \mapsto \boldsymbol{\tau}'\mathbf{e}$ on $\mathbb{R}^n$. We now establish that $L(p, \boldsymbol{\tau}, \mathbf{e}) \geqslant 0$ for every $\mathbf{e} \in \mathbb{R}^n$ with equality if and only if $\mathbf{e} = \mathbf{0}$. By Hölder's inequality, we have: $|\boldsymbol{\tau}'\mathbf{e}| \leqslant \|\boldsymbol{\tau}\|_q \|\mathbf{e}\|_p < \|\mathbf{e}\|_p$ where the second inequality uses the fact that $\boldsymbol{\tau} \in \mathcal{B}_q^n$ so that $\|\boldsymbol{\tau}\|_q < 1$. Hence $\|\mathbf{e}\|_p + \boldsymbol{\tau}'\mathbf{e} > 0$ for every $\mathbf{e} \in \mathbb{R}^n$. This implies that $L(p, \boldsymbol{\tau}, \mathbf{e}) = \left(\|\mathbf{e}\|_p + \boldsymbol{\tau}'\mathbf{e}\right) \|\mathbf{e}\|_p^{p-1} \geqslant 0$ for every $\mathbf{e} \in \mathbb{R}^n$ with equality if and only if $\|\mathbf{e}\|_p^{p-1} = 0$, which holds if and only if $\mathbf{e} = \mathbf{0}$. Since $x \mapsto x^p$ $(p \geqslant 1)$ is a strictly increasing function on $\mathbb{R}_+$, we moreover have $\lim_{\|\mathbf{e}\|_p\to\infty} L(p, \boldsymbol{\tau}, \mathbf{e}) = \infty$. This establishes (i) and (ii) of Proposition 1. We now show (iii) that $L_n(p, \boldsymbol{\tau}, \cdot)$ is a convex function on $\mathbb{R}^n$: i.e. that

$$L_n(p, \boldsymbol{\tau}, (1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2) \leqslant (1-\lambda)L_n(p, \boldsymbol{\tau}, \mathbf{e}_1) + \lambda L_n(p, \boldsymbol{\tau}, \mathbf{e}_2), \qquad 0 < \lambda < 1,$$

for every $(\mathbf{e}_1, \mathbf{e}_2) \in \mathbb{R}^{2n}$ (see, e.g. Theorem 4.1 in Rockafellar (1970)). We have

$$\begin{aligned}
&L_n\left(p, \boldsymbol{\tau}, (1-\lambda)\mathbf{e}_1 + \lambda \mathbf{e}_2\right) \\
&= \left[\|(1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\|_p + \boldsymbol{\tau}'\left((1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\right)\right] \|(1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\|_p^{p-1} \\
&\leqslant \left[(1-\lambda)\left(\|\mathbf{e}_1\|_p + \boldsymbol{\tau}'\mathbf{e}_1\right) + \lambda\left(\|\mathbf{e}_2\|_p + \boldsymbol{\tau}'\mathbf{e}_2\right)\right] \|(1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\|_p^{p-1}, \qquad (6)
\end{aligned}$$

where the last inequality uses the convexity of $\mathbf{e} \mapsto \|\mathbf{e}\|_p$ when $p \geqslant 1$ and the linearity of $\mathbf{e} \mapsto \boldsymbol{\tau}'\mathbf{e}$ on $\mathbb{R}^n$. We now show that $\|(1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\|_p^{p-1} \leqslant \|\mathbf{e}_1\|_p^{p-1} + \|\mathbf{e}_2\|_p^{p-1}$. First consider the case $1 \leqslant p < 2$: we have

$$\begin{aligned}
\|(1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\|_p^{p-1} &\leqslant \left[(1-\lambda)\|\mathbf{e}_1\|_p + \lambda\|\mathbf{e}_2\|_p\right]^{p-1} \\
&\leqslant \left[(1-\lambda)\|\mathbf{e}_1\|_p\right]^{p-1} + \left[\lambda\|\mathbf{e}_2\|_p\right]^{p-1} \\
&\leqslant \|\mathbf{e}_1\|_p^{p-1} + \|\mathbf{e}_2\|_p^{p-1}, \qquad (7)
\end{aligned}$$

where the first inequality uses triangular inequality, the second follows from Theorem 19 in Hardy, Littlewood, and Pólya (1952) applied with $r \equiv p-1$ and $s \equiv 1$, and the last inequality uses $0 < \lambda < 1$.[20] When $p \geqslant 2$, we have:

$$\begin{aligned}
\|(1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\|_p^{p-1} &\leqslant \left[(1-\lambda)\|\mathbf{e}_1\|_p + \lambda\|\mathbf{e}_2\|_p\right]^{p-1} \\
&\leqslant (1-\lambda)\|\mathbf{e}_1\|_p^{p-1} + \lambda\|\mathbf{e}_2\|_p^{p-1} \\
&\leqslant \|\mathbf{e}_1\|_p^{p-1} + \|\mathbf{e}_2\|_p^{p-1}, \qquad (8)
\end{aligned}$$

where the first inequality again uses triangular inequality, the second uses the convexity of $x \mapsto x^\rho$ $(\rho \geqslant 1)$ on $\mathbb{R}_+$, and the third inequality follows from $0 < \lambda < 1$. Combining the inequalities (6) – (8) then yields:

$$\begin{aligned}
&L_n\left(p, \boldsymbol{\tau}, (1-\lambda)\mathbf{e}_1 + \lambda\mathbf{e}_2\right) \\
&\leqslant \left[(1-\lambda)\left(\|\mathbf{e}_1\|_p + \boldsymbol{\tau}'\mathbf{e}_1\right) + \lambda\left(\|\mathbf{e}_2\|_p + \boldsymbol{\tau}'\mathbf{e}_2\right)\right]\left[\|\mathbf{e}_1\|_p^{p-1} + \|\mathbf{e}_2\|_p^{p-1}\right] \\
&\leqslant (1-\lambda)\left(\|\mathbf{e}_1\|_p + \boldsymbol{\tau}'\mathbf{e}_1\right)\|\mathbf{e}_1\|_p^{p-1} + \lambda\left(\|\mathbf{e}_2\|_p + \boldsymbol{\tau}'\mathbf{e}_2\right)\|\mathbf{e}_2\|_p^{p-1} \\
&= (1-\lambda)L_n\left(p, \boldsymbol{\tau}, \mathbf{e}_1\right) + \lambda L_n\left(p, \boldsymbol{\tau}, \mathbf{e}_2\right),
\end{aligned}$$

[20] Theorem 19 in Hardy, Littlewood, and Pólya (1952) shows that for every $(a_1, a_2) \in \mathbb{R}_+^2$ and $0 < r < s$ we have $(a_1^s + a_2^s)^{1/s} \leqslant (a_1^r + a_2^r)^{1/r}$.

where the second inequality uses the non-negativity of $\|\mathbf{e}_1\|_p + \boldsymbol{\tau}'\mathbf{e}_1$ and $\|\mathbf{e}_2\|_p + \boldsymbol{\tau}'\mathbf{e}_2$ (established in item (i) of the Proposition). This shows (iii), and thus completes the proof of Proposition 1. □

Proof of Proposition 2. Fix p_0, $1 \leqslant p_0 < \infty$, and $\boldsymbol{\tau}_0 \in \mathcal{B}^n_{q_0}$, where $1/p_0 + 1/q_0 = 1$. Differentiating the loss $L(p_0, \boldsymbol{\tau}_0, \cdot)$ in Equation (1) we have

$$\nabla_{\mathbf{e}} L(p_0, \boldsymbol{\tau}_0, \mathbf{e}) = p_0 \boldsymbol{\nu}_{p_0}(\mathbf{e}) + \boldsymbol{\tau}_0 \|\mathbf{e}\|_{p_0}^{p_0-1} + (p_0-1)\boldsymbol{\tau}'\mathbf{e}\frac{\boldsymbol{\nu}_{p_0}(\mathbf{e})}{\|\mathbf{e}\|_{p_0}}, \tag{9}$$

for all $\mathbf{e} \in \mathbb{R}^n$. Note that in the univariate case $n = 1$, the expression in Equation (9) reduces to $\nabla_e L(p_0, \tau_0, e) = [\tau_0 + \operatorname{sgn}(e)]|e|^{p_0-1}$ (see Equation (8) in Elliott, Komunjer, and Timmermann (2005), p.1121). By triangular inequality and norm equivalence, $\|\nabla_{\mathbf{e}} L(p_0, \boldsymbol{\tau}_0, \mathbf{e})\|_1 \leqslant p_0 \|\mathbf{e}\|_{p_0-1}^{p_0-1} + n\|\mathbf{e}\|_{p_0-1}^{p_0-1} + (p_0-1)n\|\mathbf{e}\|_1 \|\mathbf{e}\|_{p_0-1}^{p_0-1} / \|\mathbf{e}\|_{p_0} \leqslant C_1 \|\mathbf{e}\|_1^{p_0-1}$ with $C_1 < \infty$ when $\mathbf{e} \neq \mathbf{0}$, and $\|\nabla_{\mathbf{e}} L(p_0, \boldsymbol{\tau}_0, \mathbf{0})\|_1 \leqslant C_2 < \infty$. By assumption A3, we have $E(\|\mathbf{y}_t\|_1^{p_0-1} | \mathcal{F}_t) < \infty$ a.s.-P and $\left\|\mathbf{f}^*_{t+1,1}\right\|_1^{p_0-1} < \infty$ a.s.-P, which together with the fact that $\left\|\mathbf{e}^*_{t+1}\right\|_1^{p_0-1} \leqslant C_3 \left(\|\mathbf{y}_t\|_1^{p_0-1} + \left\|\mathbf{f}^*_{t+1,1}\right\|_1^{p_0-1}\right)$ a.s.-P then ensure $E[\nabla_{\mathbf{e}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) | \mathcal{F}_t] < \infty$ a.s.-P. This last condition combined with the convexity of $L(p_0, \boldsymbol{\tau}_0, \cdot)$, which implies that $L(p_0, \boldsymbol{\tau}_0, \cdot)$ is locally Lipschitz, allows us to interchange the order of differentiation and expectation to get: $\nabla_{\mathbf{e}} E\left[L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) | \mathcal{F}_t\right] = E[\nabla_{\mathbf{e}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) | \mathcal{F}_t]$. This, combined with the gradient expression in Equation (9) and with the convexity of the loss $L(p_0, \boldsymbol{\tau}_0, \cdot)$ shows that the first order condition in Equation (2) is necessary and sufficient for A2 to hold. □

Proof of Lemma 3. Given that $\mathbf{S}$ (and hence $\mathbf{S}^{-1}$) is positive definite, then for any $\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}$ we have $Q(\boldsymbol{\tau}) \geq 0$ with equality if and only if $E[\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = 0$. Now, the optimality condition derived in Proposition 2 implies that $E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = \mathbf{0}$. Hence, $\boldsymbol{\tau}_0$ is a minimum of $Q(\boldsymbol{\tau})$ on $\mathcal{B}^n_{q_0}$. Given $p_0 \in [1, +\infty)$ and for any $\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}$ we can write:

$$\begin{aligned}
&\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t) \\
&= p_0(\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t) + \left(\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0-1)\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1} (\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}\right)\boldsymbol{\tau} \\
&= \mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t) + \mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)\boldsymbol{\tau}
\end{aligned} \tag{10}$$

where $\mathbf{Id}_n$ denotes an $n \times n$ identity matrix, and we define the $nd \times 1$ vector

$$\mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t) \equiv p_0(\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t) \tag{11}$$

and the $nd \times n$ matrix

$$\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t) \equiv \left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0 - 1) \left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1} (\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1} \tag{12}$$

A necessary condition for $Q(\boldsymbol{\tau})$ to be minimized at $\boldsymbol{\tau}_0$ is that the latter solves the first order condition: $\nabla_{\boldsymbol{\tau}} Q(\boldsymbol{\tau}_0) = \mathbf{0}$, i.e. $\nabla_{\boldsymbol{\tau}} E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\mathbf{S}^{-1}E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = \mathbf{0}$. We first show that $\nabla_{\boldsymbol{\tau}} E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = E[\nabla_{\boldsymbol{\tau}} \mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$ where $\nabla_{\boldsymbol{\tau}} \mathbf{g}$ denotes the $n \times nd$-matrix of partial derivatives $[\partial g_j / \partial \tau_i, 1 \leq i \leq n, 1 \leq j \leq nd]$. Using the equality in 10, given $p_0 \in [1, +\infty)$ and for any $\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}$ we have:

$$\begin{aligned}
\nabla_{\boldsymbol{\tau}} \mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t) &= \mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)' \\
&= \left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t)' + (p_0 - 1) \left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1} \mathbf{e}^*_{t+1} \left(\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t\right)'
\end{aligned}$$

Note that $\sup_{\boldsymbol{\tau} \in (-1,1)^n} \|\nabla_{\boldsymbol{\tau}} \mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}_{t+1}, \mathbf{x}_t)\|_1 \leq \left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} \|\mathbf{x}_t\|_1 + (p_0 - 1) \|\mathbf{x}_t\|_1 \left\|\mathbf{e}^*_{t+1}\right\|_{p_0-1}^{p_0-1}$ and by assumption A5(i) the expected value of the latter is finite; hence we can apply Lebesgue's dominated convergence theorem to show that $\nabla_{\boldsymbol{\tau}} E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = E[\nabla_{\boldsymbol{\tau}} \mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$. The first order condition can then be written as:

$$\begin{aligned}
\mathbf{0} &= E[\nabla_{\boldsymbol{\tau}} \mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\mathbf{S}^{-1}E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] \\
&= E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}\{E[\mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)] + E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\boldsymbol{\tau}_0\}
\end{aligned}$$

so

$$-E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\boldsymbol{\tau}_0 \tag{13}$$

Now note that the $n \times n$ matrix $\mathbf{H} \equiv E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$ is positive semidefinite. In addition, for any $\xi \in \mathbb{R}^n$ we have $\xi'\mathbf{H}\xi = E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)\xi]'\mathbf{S}^{-1}E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)\xi]$ and since $\mathbf{S}^{-1}$ is positive definite, $\xi'\mathbf{H}\xi = 0$ only if $E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)\xi] = \mathbf{0}$. The latter implies $\xi = \mathbf{0}$ since the rank of the $n \times n$ matrix $E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$ is the same as that of $E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$, which

under assumption A5(ii) equals n. So $\mathbf{H}$ is positive definite, $Q(\boldsymbol{\tau})$ is strictly concave and minimized at the unique solution to the equation 13 given by:

$$\boldsymbol{\tau}_0 = \tag{14}$$
$$- \{E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\}^{-1}E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$$

□

Proof of Lemma 4. Given $p_0 \in [1, +\infty)$ and for any $\boldsymbol{\tau} \in \mathcal{B}^n_{q_0}$, let $\mathbf{S}(\boldsymbol{\tau}) \equiv E\left[\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)'\right]$. For any $\chi \in \mathbb{R}^{nd}$, $\chi = (\chi'_1, \ldots, \chi'_n)'$ (with $\chi_i \in \mathbb{R}^d$, $1 \leq i \leq n$) we have $\chi'\mathbf{S}(\boldsymbol{\tau})\chi \equiv E\left[\left(\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right)\left(\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right)'\right]$, where

$$\begin{aligned}
&\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t) \\
&= \chi'\left(p_0\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) + \boldsymbol{\tau}\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} + (p_0-1)\boldsymbol{\tau}'\mathbf{e}^*_{t+1}\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1}\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})\right) \otimes \mathbf{x}_t \\
&= (\Upsilon'\mathbf{x}_t)'\left(p_0\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1}) + \boldsymbol{\tau}\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1} + (p_0-1)\boldsymbol{\tau}'\mathbf{e}^*_{t+1}\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1}\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})\right) \\
&= (\Upsilon'\mathbf{x}_t)'\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})
\end{aligned}$$

where Υ is an $d \times n$ matrix with columns χ_i, $1 \leq i \leq n$. Note that $\left(\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right)\left(\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right)' \geq 0$ a.s. so $\mathbf{S}(\boldsymbol{\tau})$ is positive semidefinite. In addition $\chi'\mathbf{S}(\boldsymbol{\tau})\chi = 0$ only if $\left(\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right)\left(\chi'\mathbf{g}(p_0, \boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right)' = 0$ a.s. which is equivalent to $(\Upsilon'\mathbf{x}_t)'\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})'(\Upsilon'\mathbf{x}_t) = 0$ a.s. Now, given that the n-variate loss is strictly convex and such that $L_n(p_0, \tau, \mathbf{0}) = 0$, we have that $\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})' \geq 0$ with equality only if $\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1}) = 0$, i.e. only if $\mathbf{e}^*_{t+1} = 0$. Since by assumption A1 $\mathbf{y}_t$ is continuously distributed, we have that $\Pr(\mathbf{e}^*_{t+1} = 0|\mathcal{F}_t) = 0$ so $\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})' > 0$ a.s.. Then, $(\Upsilon'\mathbf{x}_t)'\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})\nabla_{\mathbf{e}}L_n(p_0, \boldsymbol{\tau}, \mathbf{e}^*_{t+1})'(\Upsilon'\mathbf{x}_t) = 0$ a.s. implies $\Upsilon'\mathbf{x}_t = \mathbf{0}$ a.s. which in turn implies $\Upsilon'E(\mathbf{x}_t\mathbf{x}'_t) = \mathbf{0}$. If $E(\mathbf{x}_t\mathbf{x}'_t)$ is of full rank as assumed in A6, then it is invertible and $\Upsilon'E(\mathbf{x}_t\mathbf{x}'_t) = \mathbf{0}$ only holds if $\Upsilon = \mathbf{0}$, i.e. $\chi = \mathbf{0}$. Hence, $\mathbf{S}(\boldsymbol{\tau})$ is positive definite. □

Proof of Theorem 5. From equation (4) we have $\hat{\boldsymbol{\tau}}_T \equiv -[\hat{\mathbf{B}}_T'\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_T]^{-1}\hat{\mathbf{B}}_T'\hat{\mathbf{S}}^{-1}\hat{\mathbf{a}}_T$ with the $nd \times 1$ vector

$$\hat{\mathbf{a}}_T \equiv T^{-1} \sum_{t=R}^{T+R-1} p_0(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t) \tag{15}$$

and the $nd \times n$ matrix

$$\hat{\mathbf{B}}_T \equiv T^{-1} \sum_{t=R}^{T+R-1} \|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0 - 1) \|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-1} (\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\hat{\mathbf{e}}_{t+1}' \tag{16}$$

To show $\hat{\boldsymbol{\tau}}_T \xrightarrow{p} \boldsymbol{\tau}_0$ it is sufficient to show that: *(i)* $\hat{\mathbf{a}}_T - E[\mathbf{a}(p_0, \mathbf{e}_{t+1}^*, \mathbf{x}_t)] \xrightarrow{p} \mathbf{0}$, and *(ii)* $\hat{\mathbf{B}}_T - E[\mathbf{B}(p_0, \mathbf{e}_{t+1}^*, \mathbf{x}_t)] \xrightarrow{p} \mathbf{0}$. Then, by using Lemma 3, the consistency of $\hat{\mathbf{S}}$, $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$, the positive definiteness of $\mathbf{S}$ (and thus of $\mathbf{S}^{-1}$) established in Lemma 4, and the continuity of the inverse function (away from zero), we have that $\hat{\boldsymbol{\tau}}_T \xrightarrow{p} \boldsymbol{\tau}_0$. By the triangle inequality we have $\left\|\hat{\mathbf{a}}_T - E[\mathbf{a}(p_0, \mathbf{e}_{t+1}^*, \mathbf{x}_t)]\right\|_1 \leqslant \|\hat{\mathbf{a}}_T - E[\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\|_1 + \left\|E[\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{a}(p_0, \mathbf{e}_{t+1}^*, \mathbf{x}_t)]\right\|_1$ and $\left\|\hat{\mathbf{B}}_T - E[\mathbf{B}(p_0, \mathbf{e}_{t+1}^*, \mathbf{x}_t)]\right\| \leqslant \left\|\hat{\mathbf{B}}_T - E[\mathbf{B}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\right\| + \left\|E[\mathbf{B}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{B}(p_0, \mathbf{e}_{t+1}^*, \mathbf{x}_t)]\right\|$, where by norm of the $nd \times n$ matrix $\mathbf{B}$ we mean: $\|\mathbf{B}\| = \max |b_{ij}|_{1 \leq i \leq nd, 1 \leq j \leq n}$. We first show that as $T \to \infty$, $\|\hat{\mathbf{a}}_T - E[\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\|_1 \xrightarrow{p} \mathbf{0}$ and $\left\|\hat{\mathbf{B}}_T - E[\mathbf{B}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\right\| \xrightarrow{p} \mathbf{0}$ by using a law of large numbers (LLN) for α-mixing sequences [e.g., Corollary 3.48 in White (2001)]. From Theorems 3.35 and 3.49 in White (2001) measurable functions of strictly stationary and mixing processes are strictly stationary and mixing of the same size. Hence, by A8 we have $\{p_0(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\}$ and $\{\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{p_0-1} (\mathbf{Id}_n \otimes \mathbf{x}_t) + (p_0 - 1) \|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-1} (\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\hat{\mathbf{e}}_{t+1}'\}$ strictly stationary and α-mixing of size $-r/(r-2)$ with $r > 2$. Now let $\delta = \varepsilon/2 > 0$; we have

$$\begin{aligned} & E[\|(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\|_1^{r+\delta}] \\ & \leq nE[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p_0-1)(r+\delta)} \|\mathbf{x}_t\|_1^{r+\delta}] \\ & \leq n \left\{E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p_0-1)(2r+2\delta)}]E[\|\mathbf{x}_t\|_1^{2r+2\delta}]\right\}^{1/2} \\ & \leq n \{\Delta_1 \Delta_2\}^{1/2} < \infty \end{aligned} \tag{17}$$

where the second inequality follows by Cauchy-Schwartz inequality, and the third uses assumption A8. Hence, $\hat{\mathbf{a}}_T$ in equation (15) satisfies the LLN and

$\|\hat{\mathbf{a}}_T - E[\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\|_1 \xrightarrow{p} \mathbf{0}$ as $T \to \infty$. Similarly, we have

$$\begin{aligned}
&E[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{(p_0-1)(r+\delta)} \|(\mathbf{Id}_n \otimes \mathbf{x}_t)\|^{r+\delta}] \\
&\leq E[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{(p_0-1)(r+\delta)} \|\mathbf{x}_t\|_1^{r+\delta}] \\
&\leq c^{(p_0-1)(r+\delta)} \left\{ E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p_0-1)(2r+2\delta)}] E[\|\mathbf{x}_t\|_1^{2r+2\delta}] \right\}^{1/2} \\
&\leq c \left\{ \Delta_1 \Delta_2 \right\}^{1/2} < \infty
\end{aligned} \tag{18}$$

where the second inequality uses the norm equivalence, i.e. there exists some $(c, d) > 0$ such that $d \|\hat{\mathbf{e}}_{t+1}\|_1 \leq \|\hat{\mathbf{e}}_{t+1}\|_{p_0} \leq c \|\hat{\mathbf{e}}_{t+1}\|_1$, and Cauchy-Schwartz inequality, and the third uses assumption A8. In addition,

$$\begin{aligned}
&E[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-(r+\delta)} \left\|(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\hat{\mathbf{e}}'_{t+1}\right\|^{r+\delta}] \\
&\leq E[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-(r+\delta)} \|(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\|_1^{r+\delta} \|\hat{\mathbf{e}}_{t+1}\|_1^{(r+\delta)}] \\
&\leq (1/d)^{r+\delta} E[\|(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\|_1^{r+\delta}] < \infty
\end{aligned} \tag{19}$$

where the second inequality uses again the norm equivalence, and the third follows from equation (17). Combining equations (18)−(19) with triangular inequality and the fact that for any $(a, b) \in \mathbb{R}$, there exists some $n_{r+\delta} > 0$ such that $|a + b|^{r+\delta} \leq n_{r+\delta}[|a|^{r+\delta} + |b|^{r+\delta}]$, shows that $\hat{\mathbf{B}}_T$ in equation (16) satisfies the LLN and so $\left\|\hat{\mathbf{B}}_T - E[\mathbf{B}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\right\| \xrightarrow{p} \mathbf{0}$ as $T \to \infty$. Next we need to show that $\left\|E[\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_1 \to 0$ and $\left\|E[\mathbf{B}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\| \to 0$ as $T \to \infty$. We have

$$\begin{aligned}
&\left\|E[\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) - \mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_1 \\
&\leq E[\left\|\mathbf{a}(p_0, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) - \mathbf{a}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right\|_1] \\
&= p_0 E\{\left\|[\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1}) - \boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})] \otimes \mathbf{x}_t\right\|_1\} \\
&\leq p_0 n E[\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1^{(p_0-1)} \|\mathbf{x}_t\|_1] \\
&\leq p_0 n \{E[\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1^{2(p_0-1)}] E[\|\mathbf{x}_t\|_1^2]\}^{1/2} \to 0 \text{ as } t \to \infty
\end{aligned}$$

where the last statement follows by assumptions A7 and A8. Similarly,

$$
\begin{aligned}
&\left\|E[\mathbf{B}(p_0,\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)-\mathbf{B}(p_0,\mathbf{e}^*_{t+1},\mathbf{x}_t)]\right\| \\
&\leq E[\left\|\mathbf{B}(p_0,\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)-\mathbf{B}(p_0,\mathbf{e}^*_{t+1},\mathbf{x}_t)\right\|] \\
&= E\left[\left\|[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{p_0-1}-\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1}](\mathbf{Id}_n\otimes\mathbf{x}_t)\right.\right. \\
&\left.\left.+(p_0-1)[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-1}(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1})\otimes\mathbf{x}_t)\hat{\mathbf{e}}'_{t+1}-\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1}(\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})\otimes\mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}]\right\|\right] \\
&\leq E\left[\left|\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{p_0-1}-\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{p_0-1}\right|\|\mathbf{x}_t\|_1\right] \\
&+(p_0-1)E\left[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-1}\left\|(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1})\otimes\mathbf{x}_t)(\hat{\mathbf{e}}'_{t+1}-\mathbf{e}^{*\prime}_{t+1})\right\|\right] \\
&+(p_0-1)E\left[\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-1}\left\|\{[(\boldsymbol{\nu}_{p_0}(\hat{\mathbf{e}}_{t+1})-\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})]\otimes\mathbf{x}_t\}\mathbf{e}^{*\prime}_{t+1}\right\|\right] \\
&+(p_0-1)E\left[\left(\|\hat{\mathbf{e}}_{t+1}\|_{p_0}^{-1}-\left\|\mathbf{e}^*_{t+1}\right\|_{p_0}^{-1}\right)\left\|(\boldsymbol{\nu}_{p_0}(\mathbf{e}^*_{t+1})\otimes\mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}\right\|\right]\to 0 \text{ as } t\to\infty
\end{aligned}
$$

Hence, as $R\to\infty$ we have $\left\|E[\mathbf{a}(p_0,\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)-\mathbf{a}(p_0,\mathbf{e}^*_{t+1},\mathbf{x}_t)]\right\|_1\to 0$ and $\left\|E[\mathbf{B}(p_0,\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)-\mathbf{B}(p_0,\mathbf{e}^*_{t+1},\mathbf{x}_t)]\right\|\to 0$, so $\hat{\boldsymbol{\tau}}_T\xrightarrow{p}\boldsymbol{\tau}_0$ as $(R,T)\to\infty$. □

Proof of Theorem 6. To show that $T^{1/2}(\hat{\boldsymbol{\tau}}_T-\boldsymbol{\tau}_0)$ is asymptotically normal, note that we have

$$
\begin{aligned}
\sqrt{T}(\hat{\boldsymbol{\tau}}_T-\boldsymbol{\tau}_0) &= -[\hat{\mathbf{B}}'_T\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_T]^{-1}\hat{\mathbf{B}}'_T\hat{\mathbf{S}}^{-1}[\sqrt{T}(\hat{\mathbf{a}}_T+\hat{\mathbf{B}}_T\boldsymbol{\tau}_0)] \qquad (20)\\
&= -[\hat{\mathbf{B}}'_T\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_T]^{-1}\hat{\mathbf{B}}'_T\hat{\mathbf{S}}^{-1}[\sqrt{T}\hat{\mathbf{m}}^*_T+\sqrt{T}\hat{\mathbf{m}}+\sqrt{T}(\hat{\mathbf{m}}_T-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_T)]
\end{aligned}
$$

where we have let $\hat{\mathbf{m}}\equiv E[\mathbf{a}(p_0,\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)]+E[\mathbf{B}(p_0,\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)]\boldsymbol{\tau}_0$, and

$$
\hat{\mathbf{m}}_T\equiv T^{-1}\sum_{t=R}^{T+R-1}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)=\hat{\mathbf{a}}_T+\hat{\mathbf{B}}_T\boldsymbol{\tau}_0,\ \text{and}\ \hat{\mathbf{m}}^*_T\equiv T^{-1}\sum_{t=R}^{T+R-1}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\mathbf{e}^*_{t+1},\mathbf{x}_t) \qquad (21)
$$

The idea then is to show that the terms $\sqrt{T}\hat{\mathbf{m}}$ and $\sqrt{T}(\hat{\mathbf{m}}_T-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_T)$ on the right hand side of equation (20) are $o_p(\mathbf{1})$. We start by showing that the first term is $o(\mathbf{1})$. Let $\mathbf{m}^*\equiv E[\mathbf{a}(p_0,\mathbf{e}^*_{t+1},\mathbf{x}_t)]+E[\mathbf{B}(p_0,\mathbf{e}^*_{t+1},\mathbf{x}_t)]\boldsymbol{\tau}_0$. First, we show that $\nabla_{\mathbf{e}}E[\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)]=E[\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)]$ for every $\bar{\mathbf{e}}_{t+1}\equiv c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1}$ with $c\in(0,1)$. Differentiating

$\nabla_{\mathbf{e}} L\left(p_0, \boldsymbol{\tau}_0, \cdot\right)$ in Equation (9) we get

$$\begin{aligned}\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}\right) &= 2p_0 \mathbf{V}_{p_0}(\mathbf{e}) + p_0(p_0-1)\mathbf{W}_{p_0}(\mathbf{e}) \\ &\quad +(p_0-1)\left[2\frac{\boldsymbol{\tau}_0 \boldsymbol{\nu}'_{p_0}(\mathbf{e})}{\|\mathbf{e}\|_{p_0}} + \frac{\boldsymbol{\tau}'_0 \mathbf{e}}{\|\mathbf{e}\|_{p_0}}\left((p_0-1)\mathbf{W}_{p_0}(\mathbf{e}) - \frac{\boldsymbol{\nu}_{p_0}(\mathbf{e})\boldsymbol{\nu}'_{p_0}(\mathbf{e})}{\|\mathbf{e}\|_{p_0}^{p_0}}\right)\right],\end{aligned} \tag{22}$$

where we have used the fact that for any $1 \leqslant p_0 < \infty$, $\dfrac{\boldsymbol{\tau}'_0 \mathbf{e}}{\|\mathbf{e}\|_{p_0}} \mathbf{V}_{p_0}(\mathbf{e}) = \mathbf{0}$ for all $\mathbf{e} \in \mathbb{R}^n$. Note that in the univariate case $n = 1$, the Hessian in Equation (22) reduces to $\Delta_{ee} L\left(p_0, \tau_0, e\right) = 2\{p_0 \delta(e)|e|^{p_0-1} + p_0(p_0-1)[1+\tau_0 \operatorname{sgn}(e)]|e|^{p_0-2}\}$ [see Equation (9) in Elliott, Komunjer, and Timmermann (2005), p.1121]. Hence

$$\begin{aligned}&\|\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \bar{\mathbf{e}}_{t+1}\right)\| \\ &\leq 2p_0 \|\mathbf{V}_{p_0}(\bar{\mathbf{e}}_{t+1})\| + p_0(p_0-1)c_3 \|\bar{\mathbf{e}}_{t+1}\|_1^{p_0-2} \\ &+ (p_0-1)\left[2d_3 \|\bar{\mathbf{e}}_{t+1}\|_1^{p_0-2} + (p_0-1)c_3 \|\bar{\mathbf{e}}_{t+1}\|_1^{p_0-2} + c_3 \|\bar{\mathbf{e}}_{t+1}\|_1^{p_0-2}\right] \\ &= 2p_0 \|\mathbf{V}_{p_0}(\bar{\mathbf{e}}_{t+1})\| + 2(p_0-1)\left(p_0 c_3 + d_3\right) \|\bar{\mathbf{e}}_{t+1}\|_1^{p_0-2}\end{aligned} \tag{23}$$

where we have used the norm equivalences: $c_1 \|\bar{\mathbf{e}}_{t+1}\|_1 \leq \|\bar{\mathbf{e}}_{t+1}\|_{p_0-2} \leq c_2 \|\bar{\mathbf{e}}_{t+1}\|_1$ for some $(c_1, c_2) > 0$ and $c_3 = c_2^{p_0-1}$ if $p_0 \geq 2$ and $= c_1^{2-p_0}$ otherwise, and similarly, $d_1 \|\bar{\mathbf{e}}_{t+1}\|_1 \leq \|\bar{\mathbf{e}}_{t+1}\|_{p_0} \leq d_2 \|\bar{\mathbf{e}}_{t+1}\|_1$ for some $(d_1, d_2) > 0$ and $d_3 = d_2^{p_0-1}$ if $p_0 \geq 2$ and $= d_1^{2-p_0}$ otherwise. Under A9, we have that $E[\sup_{c\in(0,1)} \left\|c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}^*_{t+1}\right\|_1^{p_0-2}] < \infty$. Moreover, under A10, when $p_0 = 1$ we have $E[\|\mathbf{V}_1(\bar{\mathbf{e}}_{t+1})\|] \leq M$ and when $p_0 > 1$ we have $E[\|\mathbf{V}_1(\bar{\mathbf{e}}_{t+1})\|] = 0$, so the right hand side of equation (23) is bounded above by a quantity that is integrable; hence we can apply Lebesgue's dominated convergence theorem to interchange the derivation and integration in:

$$\begin{aligned}\nabla_{\mathbf{e}} E[\mathbf{g}(p_0, \boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)] &= \nabla_{\mathbf{e}} E[\nabla_{\mathbf{e}} L\left(p_0, \boldsymbol{\tau}_0, \bar{\mathbf{e}}_{t+1}\right) \otimes \mathbf{x}_t] \\ &= E[\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \bar{\mathbf{e}}_{t+1}\right) \otimes \mathbf{x}_t] = E[\nabla_{\mathbf{e}} \mathbf{g}(p_0, \boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\end{aligned}$$

Second, we can use a mean value expansion around $\mathbf{e}^*_{t+1}$ which yields $\mathbf{0} = \sqrt{T}\mathbf{m}^* = \sqrt{T}\hat{\mathbf{m}} - E[T^{-1}\sum_{t=R}^{T+R-1} \nabla_{\mathbf{e}} \mathbf{g}(p_0, \boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)'\sqrt{T}(\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1})]$, where for every t, $R \leqslant t \leqslant T+R-1$, we have $\bar{\mathbf{e}}_{t+1} \equiv c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}^*_{t+1}$ with $c \in (0,1)$. We now show that

$T^{-1/2}\sum_{t=R}^{T+R-1}\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\xrightarrow{p}\mathbf{0}$ as $R\to\infty$ and $T\to\infty$: we have

$$\begin{aligned}
&\left\|T^{-1/2}\textstyle\sum_{t=R}^{T+R-1}\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1\\
&=\left\|T^{-1/2}\textstyle\sum_{t=R}^{T+R-1}t^{-1/2+\varepsilon}\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'t^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1\\
&\leqslant\sup_{R\leqslant t\leqslant T+R-1}\left\|t^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1 T^{-1/2}\textstyle\sum_{t=R}^{T+R-1}\left\|\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|t^{-1/2+\varepsilon}
\end{aligned}$$

Moreover, $\|\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\|\leq\|\Delta_{\mathbf{ee}}L\left(p_0,\boldsymbol{\tau}_0,\bar{\mathbf{e}}_{t+1}\right)\|\cdot\|\mathbf{x}_t\|_1$ so that under A9

$$\begin{aligned}
&E\left(\sup_{c\in(0,1)}\left\|\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1},\mathbf{x}_t)\right\|\right)\\
&\leq 2(p_0-1)\left(p_0c_3+d_3\right)E\left(\|\mathbf{x}_t\|_1\sup_{c\in(0,1)}\left\|c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1}\right\|_1^{p_0-2}\right)<\infty
\end{aligned}$$

Now, for any given $\nu>0$, by Chebyshev's inequality we have

$$\begin{aligned}
&\Pr\left(T^{-1/2}\textstyle\sum_{t=R}^{T+R-1}\left\|\nabla_{\mathbf{e}}\mathbf{g}(p_0,\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|t^{-1/2+\varepsilon}>\nu\right)\\
&\leq\nu^{-1}E\left(\|\mathbf{x}_t\|_1\sup_{c\in(0,1)}\left\|c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1}\right\|_1^{p_0-2}\right)T^{-1/2}\textstyle\sum_{t=R}^{T+R-1}t^{-1/2+\varepsilon}\\
&\leq\nu^{-1}E\left(\|\mathbf{x}_t\|_1\sup_{c\in(0,1)}\left\|c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1}\right\|_1^{p_0-2}\right)\left(\frac{T}{R^{1-2\varepsilon}}\right)^{1/2}\to 0
\end{aligned}$$

as $R\to\infty$ and $T\to\infty$, where the last limit results uses assumptions A9 and A7'(i). Hence $\sqrt{T}\hat{\mathbf{m}}\to\mathbf{0}$ as $R\to\infty$ and $T\to\infty$. The term $\sqrt{T}(\hat{\mathbf{m}}_T-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_T)$ on the right hand side of equation (20) is $o_p(\mathbf{1})$ provided that $\mathbf{g}$ satisfies a certain Lipshitz condition (given below) and we have, for any $\varepsilon>0$, $\Pr\left(\sup_{R\leqslant t\leqslant T+R-1}\left\|\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1}\right\|_1>\varepsilon\right)\to 0$ as $R\to\infty$. This follows because for any $\eta>0$ and $\delta_R>0$ we have:

$$\begin{aligned}
&\lim_{R,T\to\infty}\Pr\left(\sqrt{T}\left\|\hat{\mathbf{m}}_T-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_T\right\|_1>\eta\right)\\
&\leq\lim_{R,T\to\infty}\Pr\left(\sqrt{T}\left\|\hat{\mathbf{m}}_T-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_T\right\|_1>\eta,\sup_{R\leqslant t\leqslant T+R-1}\left\|\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1}\right\|_1\leq\delta_R\right)\\
&+\lim_{R,T\to\infty}\Pr\left(\sup_{R\leqslant t\leqslant T+R-1}\left\|\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1}\right\|_1>\delta_R\right)\\
&\leq\Pr\left(\sqrt{T}\left\|\hat{\mathbf{m}}_T-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_T\right\|_1>\eta,\sup_{R\leqslant t\leqslant T+R-1}\left\|\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1}\right\|_1\leq\delta_R\right)
\end{aligned}$$

where the last inequality uses A7'(ii). Now, let $r_T(\delta_R) \equiv \sup\{r_{t+1}(\hat{\mathbf{e}}_{t+1}) : \left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 \leq \delta_R, R \leqslant t \leqslant T+R-1\}$ where we let

$$
\begin{aligned}
& r_{t+1}(\hat{\mathbf{e}}_{t+1}) \qquad (24)\\
& \equiv \frac{\left\|\mathbf{g}(p_0, \boldsymbol{\tau}_0; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) - \mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t) - [\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t](\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1})\right\|_1}{\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1}
\end{aligned}
$$

where $\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right)$ is as defined in equation (22). Then, by the definition of $r_{t+1}(\hat{\mathbf{e}}_{t+1})$

$$
\begin{aligned}
& \sqrt{T}\left\|\hat{\mathbf{m}}_T - \hat{\mathbf{m}} - \hat{\mathbf{m}}^*_T\right\|_1 \\
& \leq \sqrt{T}\left\{\left\|\frac{1}{T}\sum_{t=R}^{T+R-1}[\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t](\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1})\right.\right. \\
& \left. - E\{[\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t](\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1})\}_1\right\|_1 \\
& \left. + \frac{1}{T}\sum_{t=R}^{T+R-1} r_{t+1}(\hat{\mathbf{e}}_{t+1})\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 + E\left(r_{t+1}(\hat{\mathbf{e}}_{t+1})\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1\right)\right\} \\
& \leq \sqrt{T}\left\{\frac{1}{T}\sum_{t=R}^{T+R-1}\left\|[\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t] - E\{[\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t]\}\right\|_1 \cdot\right. \\
& \left. \sup_{R \leqslant t \leqslant T+R-1}\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 + [r_T(\delta_R) + E(r_T(\delta_R))]\sup_{R \leqslant t \leqslant T+R-1}\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1\right\}
\end{aligned}
$$

Using standard arguments for stochastic equicontinuity such as those given in Andrews (1994), we can show that $r_{t+1}(\hat{\mathbf{e}}_{t+1}) \to 0$ as $\Pr(\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 > \varepsilon) \to 0$ for any $\varepsilon > 0$, so that $r_T(\delta_R) \to 0$ with probability one, which by the dominated convergence theorem ensures $E(r_T(\delta_R)) \to 0$ as $\delta_R \to 0$. Next, we show that the sample mean of $\{\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t\}$ converges in probability to its expected value. By assumption A4' we know that $\{\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t\}$ is strictly stationary and α-mixing with α of size $-r/(r-2)$ with $r > 2$ [see Theorems 3.35 and 3.49 in White (2001)]. Moreover, for

$\delta = \min\{\varepsilon/2, \epsilon/2\} > 0$ in assumptions A4' and A8, we have

$$\begin{aligned}
&E[\left\|\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t\right\|^{r+\delta}] \\
&\leq \{E[\left\|\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right)\right\|^{2r+2\delta}] E[\|\mathbf{x}_t\|_1^{2r+2\delta}]\}^{1/2} \\
&\leq \left(\max\{E[\left\|\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right)\right\|^{2r+\epsilon}], 1\}\right)^{1/2} \left(\max\{E[\|\mathbf{x}_t\|_1^{2r+\varepsilon}], 1\}\right)^{1/2} < \infty
\end{aligned}$$

since from equation (23) we know

$$\begin{aligned}
&\left\|\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right)\right\|^{2r+\epsilon} \\
&\leq n_r\{[2p_0]^{2r+\epsilon} \left\|\mathbf{V}_{p_0}(\mathbf{e}^*_{t+1})\right\|^{2r+\epsilon} + [2(p_0-1)\left(p_0 c_3 + d_3\right)]^{2r+\epsilon} \left\|\mathbf{e}^*_{t+1}\right\|_1^{(p_0-2)(2r+\epsilon)}\},
\end{aligned}$$

where again n_r is such that for any $(a, b) > 0$ we have $(a+b)^{2r+\epsilon} \leq n_r(a^{2r+\epsilon} + b^{2r+\epsilon})$; and A10 and A4' imply that $E[\left\|\mathbf{V}_1(\mathbf{e}^*_{t+1})\right\|^{2r+\epsilon}] \leq M$, $E[\left\|\mathbf{V}_{p_0}(\mathbf{e}^*_{t+1})\right\|^{2r+\epsilon}] = 0$ and $E[\left\|\mathbf{e}^*_{t+1}\right\|_1^{(p_0-2)(2r+\epsilon)}] < \infty$. Using the weak LLN for α-mixing sequences [e.g., Corollary 3.48 in White (2001)] then gives

$$T^{-1} \sum_{t=R}^{T+R-1} \Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t \xrightarrow{p} E[\Delta_{\mathbf{ee}} L\left(p_0, \boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t]$$

as $T \to \infty$. Then, by using the Markov inequality

$$\lim_{T\to\infty} \Pr\left(\sqrt{T}\left\|\hat{\mathbf{m}}_T - \hat{\mathbf{m}} - \hat{\mathbf{m}}^*_T\right\|_1 > \eta, \sup_{R\leqslant t\leqslant T+R-1} \left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 \leq \delta_R\right) = 0$$

and the term $\sqrt{T}(\hat{\mathbf{m}}_T - \hat{\mathbf{m}} - \hat{\mathbf{m}}^*_T)$ on the right hand side of equation (20) is $o_p(\mathbf{1})$ as $R \to \infty$ and $T \to \infty$. Finally, we use the central limit theorem (CLT) for strictly stationary and α-mixing sequences [e.g., Theorem 5.20 in White (2001)] to show that $\sqrt{T}\hat{\mathbf{m}}^*_T \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S})$. Using Theorems 3.35 and 3.49 in White (2001), which together show that time-invariant measurable functions of strictly stationary and mixing sequences are strictly stationary and mixing of the same size, we know by A4' that $\{\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\}$ is strictly stationary and α-mixing with mixing coefficient of size $-r/(r-2)$, $r > 2$. In the proof of Theorem 5 we have moreover shown that $E[\left\|\mathbf{g}(p_0, \boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right\|_1^{r+\varepsilon}] < \infty$. The CLT [e.g., Theorem 5.20 in White (2001)] then ensures

$$\sqrt{T}\hat{\mathbf{m}}^*_T \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}). \tag{25}$$

The remainder of the asymptotic normality proof is similar to the standard case: the positive definiteness of $\mathbf{S}^{-1}$, $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$ and $\hat{\mathbf{B}}_T \xrightarrow{p} \mathbf{B}^* \equiv E[\mathbf{B}(p_0, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$ as $R \to \infty$ and $T \to \infty$ ($\mathbf{B}$ was defined in equation (12)) together with A5(ii) ensure that $(\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1}$ exists, so by using $\sqrt{T}(\hat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) = -[\hat{\mathbf{B}}'_T\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_T]^{-1}\hat{\mathbf{B}}'_T\hat{\mathbf{S}}^{-1}[\sqrt{T}\hat{\mathbf{m}}^*_T + o_p(\mathbf{1})]$, the limit result in (25) and the Slutsky theorem we have $\sqrt{T}(\hat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1})$, which completes the proof of asymptotic normality. □

References

Andrews, D. W. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62(1), 43–72.

Barsky, R. B. (1987): "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19(1), 3–24.

Batchelor, R., and D. A. Peel (1998): "Rationality Testing under Asymmetric Loss," *Economics Letters*, 61, 49–54.

Bekaert, G., and R. J. Hodrick (2001): "Expectations Hypotheses Tests," *Journal of Finance*, 56(4), 1357–1394.

Boivin, J., and S. Ng (2006): "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 132(1), 169–194.

Bonham, C. S., and R. H. Cohen (2001): "To Aggregate, Pool, or Neither: Testing the Rational-Expectations Hypothesis Using Survey Data," *Journal of Business and Economic Statistics*, 19(3), 278–291.

Bracewell, R. N. (2000): *The Fourier Transform and Its Applications*. McGraw–Hill, 3rd edn.

Brayton, F., A. Levin, R. Tryon, and J. C. Williams (1997): "The Evolution of Macro Models at the Federal Reserve Board," *Carnegie-Rochester Conference Series on Public Policy*, 47, 43–81.

Campbell, J. Y., and R. J. Shiller (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95(5), 1062–1088.

Chaudhuri, P. (1996): "On a Geometric Notion of Quantiles for Multivariate Data," *Journal of the American Statistical Association*, 91, 862–872.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and To What End?," in *Handbook of Macroeconomics, Volume 1A*, ed. by J. B. Taylor, and M. Woodford, Amsterdam. Elsevier.

Christoffersen, P. F., and F. X. Diebold (1997): "Optimal Prediction under Asymmetric Loss," *Econometric Theory*, 13, 808–817.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macreconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115(1), 147–180.

Eijffinger, S. C., P. M. Geraats, and C. A. van der Cruijsen (2006): "Does Central Bank Transparency Reduce Interest Rates?," CEPR Discussion Paper 5526.

Elliott, G., I. Komunjer, and A. Timmermann (2005): "Estimation and Testing of Forecast Rationality under Flexible Loss," *The Review of Economic Studies*, 72, 1107–1125.

——— (2006): "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?," mimeo, University of California San Diego.

Engel, C., and K. D. West (2005): "Exchange Rates and Fundamentals," *Journal of Political Economy*, 113(3), 485–517.

Fama, E. F. (1990): "Term-Structure Forecasts of Interest Rates, Inflation, and Real Returns," *Journal of Monetary Economics*, 25(1), 59–76.

Figlewski, S., and P. Wachtel (1981): "The Formation of Inflationary Expectations," *Review of Economics and Statistics*, 63(1), 1–10.

Friedman, M. (1968): "The Role of Monetary Policy," *American Economic Review*, 58(1), 1–17.

Hansen, L. P. (1982): "Large-Sample Properties of Generalized Method-of- Moments Estimators," *Econometrica*, 50, 1029–1054.

Hardy, G. H., J. E. Littlewood, and G. Pólya (1952): *Inequalities*. Cambridge University Press, United Kingdom.

Kirchgässner, G., and U. K. Müller (2006): "Are Forecasters Reluctant to Revise their Predictions? Some German Evidence," *Journal of Forecasting*, 25(6), 401–413.

Koenker, R., and G. Bassett, Jr. (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Lucas, Jr., R. E. (1973): "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 63(3), 326–334.

McCallum, B. T., and E. Nelson (1999): "An Optimizing IS-LM Specification for Monetary Policy and Business Cycle Analysis," *Journal of Money, Credit, and Banking*, 31(3), 296–316.

Modigliani, F., and R. J. Shiller (1973): "Inflation, Rational Expectations, and the Term Structure of Interest Rates," *Economica*, 40(157), 12–43.

Muth, J. F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29(3), 315–335.

Pesaran, M. H. (1987): *The Limits to Rational Expectations*. Basil Blackwell, Oxford.

Poole, W. (1976): "Rational Expectations in the Macro Model," *Brookings Papers on Economic Activity*, 2, 463–505.

Poole, W., and R. H. Rasche (2003): "The Impact of Changes in FOMC Disclosure Practices on the Transparency of Monetary Policy: Are Markets and the FOMC Better 'Synched'?," *Federal Reserve Bank of St. Louis Review*, 85(1), 1–10.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

Sargent, T. J. (1973): "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," *Brookings Papers on Economic Activity*, 2, 429–472.

Shiller, R. J. (1978): "Rational Expectations and the Dynamic Structure of Macroeconomic Models," *Journal of Monetary Economics*, 4(1), 1–44.

Sims, C. A. (1986): "Are Forecasting Models Usable for Policy Analysis?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 10(1), 2–16.

Stock, J. H., and M. W. Watson (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44(2), 293–335.

SWANSON, E. T. (2006): "Have Increases in Federal Reserve Transparency Improved Private Sector Interest Rate Forecasts?," *Journal of Money, Credit, and Banking*, 38(3), 791–819.

TAYLOR, J. B. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

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

ZELLNER, A. (1986): "Biased Predictors, Rationality and the Evaluation of Forecasts," *Economics Letters*, 21, 45–48.

Table 1: Information Sets for Regressions

Info Set	C	GDP/GNP	CPI	UR	SR	F Err
1	1	n/a	n/a	n/a	n/a	1
2	1	1	n/a	n/a	n/a	n/a
3	1	n/a	1	n/a	n/a	n/a
4	1	n/a	n/a	n/a	1	n/a
5	1	1	1	n/a	n/a	n/a
6	1	n/a	1	n/a	1	n/a
7	1	1	n/a	n/a	1	n/a
8	1	n/a	n/a	n/a	3	n/a
9	1	1	1	n/a	1	n/a
10	1	1	1	n/a	1	1
11	1	1	1	1	1	1
12	1	1	1	1	1	n/a
13	1	1	3	n/a	n/a	n/a
14	1	1	3	n/a	n/a	1

Table 2: Rationality Tests

	Multivariate						Univariate						Symmetry and Independence		
	Acceptances of Rationality						Acceptances of Rationality						Acceptances of Rationality		
Info	P = 1			P = 2			P = 1			P = 2					
Set	J90	J95	J99	J90	J95	J99	J90	J95	J99	J90	J95	J99	J90	J95	J99
1	1	1	1	0.9825	1	1	1	1	1	0.9649	0.9825	0.9825	0.1930	0.3860	0.8246
2	0.8421	0.9298	1	0.8596	0.9825	1	0.8421	0.9298	1	0.8596	0.9649	1	0.0175	0.0526	0.1579
3	0.4035	0.5789	0.9825	0.7018	0.8246	1	0.4035	0.5789	0.9825	0.7018	0.8246	0.9649	0.0175	0.0175	0.0526
4	0.6491	0.7719	0.9298	0.5263	0.7193	0.9298	0.6491	0.7719	0.9298	0.4386	0.6667	0.9298	0	0.0175	0.0877
5	0.5439	0.7895	1	0.8246	0.9298	1	0.5439	0.7895	1	0.8772	0.9123	1	0	0.0702	0.1404
6	0.4386	0.7018	0.9649	0.4561	0.6842	0.9649	0.4386	0.7018	0.9649	0.5965	0.8246	0.9649	0	0	0.1228
7	0.7018	0.8246	1	0.6842	0.9123	1	0.7018	0.8246	1	0.6316	0.8596	1	0.0175	0.0526	0.2807
8	0.9474	0.9825	1	0.8947	1	1	0.9474	0.9825	1	0.9123	0.9649	1	0.0877	0.2105	0.6140
9	0.5614	0.8246	0.9825	0.7719	0.9298	0.9825	0.5614	0.8246	0.9825	0.8070	0.8947	0.9825	0	0.0702	0.3860
10	0.9649	1	1	1	1	1	0.9649	1	1	1	1	1	0.5088	0.6842	0.9649
11	0.9298	0.9825	1	0.9649	1	1	0.9298	0.9825	1	0.9825	1	1	0.5439	0.7544	0.9649
12	0.4561	0.6842	0.9649	0.5789	0.7895	1	0.4561	0.6842	0.9649	0.6316	0.8772	1	0	0.0526	0.4561
13	0.9649	0.9825	1	1	1	1	0.9649	0.9825	1	0.9825	1	1	0.1754	0.3684	0.7368
14	0.9825	1	1	1	1	1	0.9825	1	1	0.9825	1	1	0.7895	0.9298	1

Table 3: Asymmetry Coefficients

Multivariate

Info Set	Means						Medians					
	P = 1			P = 2			P = 1			P = 2		
	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR
1	0.4133	0.6379	0.4485	0.4508	0.5706	0.2675	0.4011	0.6411	0.4392	0.4542	0.5664	0.2689
2	0.4177	0.6339	0.4427	0.4653	0.5763	0.2827	0.4128	0.6407	0.4410	0.4590	0.5793	0.2822
3	0.4257	0.6339	0.4215	0.4435	0.5895	0.2791	0.4243	0.6196	0.4251	0.4302	0.5968	0.2850
4	0.4163	0.6477	0.4480	0.4352	0.5646	0.2741	0.4071	0.6487	0.4349	0.4208	0.5633	0.2603
5	0.4313	0.6377	0.4165	0.4311	0.5975	0.2678	0.4268	0.6327	0.4203	0.4225	0.5922	0.2730
6	0.4287	0.6681	0.4312	0.4285	0.5898	0.2564	0.4212	0.6670	0.4297	0.4280	0.5860	0.2559
7	0.4109	0.6471	0.4451	0.4268	0.5723	0.2740	0.4051	0.6583	0.4401	0.4267	0.5722	0.2556
8	0.4156	0.6440	0.4602	0.4291	0.5787	0.2762	0.4033	0.6407	0.4527	0.4320	0.5685	0.2595
9	0.4336	0.6675	0.4285	0.4232	0.5950	0.2565	0.4295	0.6703	0.4197	0.4188	0.5968	0.2533
10	0.4445	0.6849	0.4356	0.4288	0.6014	0.2395	0.4318	0.6867	0.4462	0.4210	0.5950	0.2232
11	0.4286	0.7414	0.4232	0.4183	0.6305	0.2314	0.3946	0.7241	0.4084	0.4129	0.6197	0.1994
12	0.4346	0.7282	0.4218	0.4189	0.6264	0.2452	0.4276	0.7173	0.3999	0.4132	0.6214	0.2375
13	0.4304	0.6514	0.4064	0.4191	0.5869	0.2452	0.4176	0.6435	0.4109	0.4211	0.5856	0.2467
14	0.4321	0.6677	0.4140	0.4159	0.5891	0.2314	0.4153	0.6659	0.4188	0.4264	0.5977	0.2214

Univariate

Info Set	Means						Medians					
	P = 1			P = 2			P = 1			P = 2		
	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR
1	0.4133	0.6379	0.4485	0.4404	0.6453	0.2312	0.4011	0.6411	0.4392	0.3967	0.6533	0.2335
2	0.4177	0.6339	0.4427	0.4551	0.6444	0.2442	0.4128	0.6407	0.4410	0.4500	0.6563	0.2462
3	0.4257	0.6339	0.4215	0.4357	0.6709	0.2424	0.4243	0.6196	0.4251	0.4054	0.7114	0.2473
4	0.4163	0.6477	0.4480	0.4251	0.6209	0.2266	0.4071	0.6487	0.4349	0.4044	0.6221	0.2113
5	0.4313	0.6377	0.4165	0.4336	0.6965	0.2343	0.4268	0.6327	0.4203	0.3992	0.7251	0.2332
6	0.4287	0.6681	0.4312	0.4187	0.6736	0.2000	0.4212	0.6670	0.4297	0.4022	0.7279	0.1795
7	0.4109	0.6471	0.4451	0.4221	0.6261	0.2210	0.4051	0.6583	0.4401	0.3906	0.6475	0.2066
8	0.4156	0.6440	0.4602	0.4273	0.6384	0.2283	0.4033	0.6407	0.4527	0.3951	0.6350	0.2116
9	0.4336	0.6675	0.4285	0.4223	0.6815	0.1923	0.4295	0.6703	0.4197	0.4015	0.7245	0.1626
10	0.4445	0.6849	0.4356	0.4230	0.6996	0.1763	0.4318	0.6867	0.4462	0.4062	0.7492	0.1489
11	0.4286	0.7414	0.4232	0.3812	0.7449	0.1535	0.3946	0.7241	0.4084	0.3590	0.8511	0.1372
12	0.4346	0.7282	0.4218	0.4090	0.7557	0.1684	0.4276	0.7173	0.3999	0.4033	0.8229	0.1434
13	0.4304	0.6514	0.4064	0.4108	0.6843	0.2061	0.4176	0.6435	0.4109	0.4052	0.7081	0.1994
14	0.4321	0.6677	0.4140	0.4252	0.7058	0.1956	0.4153	0.6659	0.4188	0.4241	0.7292	0.1880

Table 4: Asymmetry Coefficients, Subsample Estimates

Multivariate

Info Set	Post-1994						Pre-1994						Full Sample*					
	Means (P=2)			Medians (P=2)			Means (P=2)			Medians (P=2)			Means (P=2)			Medians (P=2)		
	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR
1	0.3314	0.5376	0.3823	0.3069	0.5426	0.3644	0.5377	0.6008	0.1901	0.5200	0.6023	0.1810	0.4196	0.5515	0.2906	0.4074	0.5487	0.2759
2	0.3710	0.5473	0.3974	0.3535	0.5598	0.3817	0.5502	0.5987	0.2268	0.5488	0.6091	0.2240	0.4351	0.5571	0.2972	0.4227	0.5648	0.3009
3	0.2938	0.5489	0.3360	0.2706	0.5620	0.3193	0.5422	0.6281	0.2420	0.5484	0.6430	0.2420	0.4033	0.5611	0.2856	0.3941	0.5691	0.2822
4	0.3829	0.5312	0.3454	0.3607	0.5370	0.3732	0.5012	0.5862	0.2194	0.5012	0.5809	0.2220	0.4219	0.5425	0.2822	0.3967	0.5491	0.2594
5	0.2952	0.5691	0.3368	0.2659	0.5771	0.3271	0.5213	0.6315	0.2181	0.5222	0.6401	0.2025	0.3982	0.5693	0.2778	0.3986	0.5779	0.2748
6	0.3345	0.5411	0.2894	0.2969	0.5430	0.2569	0.4977	0.6237	0.2103	0.4953	0.6155	0.1886	0.4103	0.5591	0.2616	0.4024	0.5647	0.2520
7	0.3724	0.5500	0.3869	0.3470	0.5736	0.4061	0.5009	0.5910	0.2136	0.4884	0.5870	0.2153	0.4117	0.5547	0.2869	0.3941	0.5637	0.2516
8	0.3463	0.5679	0.3983	0.3155	0.6008	0.4226	0.4977	0.5914	0.2127	0.5012	0.5873	0.2139	0.4121	0.5584	0.2871	0.4020	0.5605	0.2639
9	0.3389	0.5537	0.2918	0.3074	0.5388	0.2373	0.4974	0.6302	0.2043	0.4772	0.6215	0.1904	0.4047	0.5642	0.2611	0.3935	0.5619	0.2444
10	0.2933	0.5575	0.3139	0.2590	0.5540	0.2681	0.5227	0.6326	0.1730	0.5377	0.6307	0.1508	0.3978	0.5780	0.2551	0.4010	0.5748	0.2162
11	0.2109	0.6201	0.5401	0.1579	0.6387	0.6200	0.5388	0.6681	0.1867	0.5393	0.6516	0.1883	0.3727	0.6086	0.2539	0.3856	0.6088	0.2021
12	0.2854	0.6383	0.4976	0.2526	0.6625	0.5789	0.5120	0.6500	0.1911	0.5054	0.6472	0.1727	0.3857	0.6051	0.2656	0.3876	0.6165	0.2335
13	0.2616	0.5744	0.3345	0.2338	0.5823	0.3126	0.5091	0.6087	0.1804	0.5088	0.6199	0.1702	0.3851	0.5703	0.2707	0.3872	0.5791	0.2619
14	0.2278	0.5763	0.3477	0.2163	0.5757	0.3274	0.5048	0.6100	0.1495	0.4854	0.6144	0.1365	0.3810	0.5778	0.2658	0.3822	0.5884	0.2374

Univariate

Info Set	Post-1994						Pre-1994						Full Sample*					
	Means (P=2)			Medians (P=2)			Means (P=2)			Medians (P=2)			Means (P=2)			Medians (P=2)		
	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR	Alpha GNP	Alpha INF	Alpha SR
1	0.2877	0.5489	0.3522	0.2620	0.5655	0.3302	0.5520	0.7199	0.1609	0.5479	0.7594	0.1551	0.3990	0.5936	0.2507	0.3702	0.5962	0.2434
2	0.3402	0.5914	0.3738	0.3148	0.6176	0.3681	0.5779	0.6948	0.1919	0.5669	0.7008	0.1880	0.4121	0.6120	0.2606	0.3948	0.6340	0.2552
3	0.2763	0.5808	0.2894	0.2434	0.5969	0.2787	0.5656	0.7698	0.2104	0.5911	0.8120	0.2074	0.3816	0.6052	0.2425	0.3528	0.6228	0.2415
4	0.3549	0.5483	0.2423	0.3335	0.5696	0.1743	0.5250	0.6605	0.1867	0.5176	0.6377	0.1655	0.3949	0.5838	0.2180	0.3753	0.6082	0.1957
5	0.2893	0.6373	0.2890	0.2658	0.6801	0.2907	0.5770	0.7817	0.2003	0.5563	0.8159	0.2032	0.3816	0.6318	0.2359	0.3532	0.6625	0.2209
6	0.3135	0.5497	0.1833	0.3046	0.6670	0.1390	0.5086	0.7495	0.1570	0.4874	0.8049	0.1350	0.3946	0.6131	0.1946	0.3839	0.6460	0.1724
7	0.3432	0.5589	0.2809	0.3181	0.6217	0.2038	0.5576	0.6575	0.1729	0.5077	0.6610	0.1531	0.3850	0.5985	0.2240	0.3761	0.6412	0.1963
8	0.3380	0.6151	0.2688	0.3085	0.7045	0.2118	0.5740	0.6624	0.1903	0.5474	0.6745	0.1853	0.3849	0.6109	0.2273	0.3636	0.6315	0.2000
9	0.3068	0.5649	0.1855	0.3030	0.7045	0.1416	0.5491	0.7563	0.1434	0.5379	0.7766	0.1354	0.3884	0.6153	0.1896	0.3831	0.6513	0.1640
10	0.2484	0.5703	0.1806	0.2527	0.7068	0.1114	0.5837	0.7764	0.1402	0.6046	0.8294	0.1164	0.3711	0.6456	0.1674	0.3633	0.6824	0.1471
11	0.2178	0.6574	0.3325	0.1791	0.8907	0.1448	0.5559	0.8395	0.1125	0.5680	0.9142	0.0922	0.3336	0.6907	0.1686	0.3410	0.8045	0.1257
12	0.2794	0.6165	0.2320	0.2759	0.8860	0.1447	0.5681	0.8679	0.1256	0.5690	0.8945	0.1170	0.3614	0.6859	0.1839	0.3727	0.7998	0.1471
13	0.2336	0.6489	0.2781	0.1996	0.6774	0.2614	0.5593	0.7532	0.1515	0.5627	0.8013	0.1309	0.3592	0.6416	0.2275	0.3461	0.6758	0.2192
14	0.2065	0.6541	0.2763	0.1863	0.7009	0.2512	0.5771	0.7757	0.1436	0.5647	0.8157	0.1215	0.3662	0.6651	0.2167	0.3360	0.7034	0.2007

* Indicates same forecasters as in the post-break sample

Table 5

Pre-1994 (P = 2)

Info Set	Univariate				Multivariate			
	a1 = a2 = 0.5	a1 = 0.5	a2 = 0.5	a3 = 0.5	a1 = a2 = 0.5	a1 = 0.5	a2 = 0.5	a3 = 0.5
1	0.2353	0.5588	0.2647	0	0.2647	0.5588	0.2941	0
2	0.3235	0.4706	0.3529	0	0.2647	0.5294	0.2647	0
3	0.1176	0.5294	0.1176	0	0.0882	0.5882	0.1176	0
4	0.2647	0.5882	0.4706	0.0294	0.2647	0.6471	0.3529	0.0294
5	0.0588	0.5	0.0882	0	0.0588	0.6471	0.0588	0
6	0.0588	0.5882	0.0588	0.0294	0.0294	0.6471	0.0588	0
7	0.2353	0.3529	0.4412	0	0.2059	0.5	0.3824	0
8	0.2059	0.4118	0.4118	0.0294	0.1765	0.5882	0.3235	0.0294
9	0.0588	0.4118	0.1176	0	0.0588	0.5294	0.1471	0
10	0.0294	0.2647	0.0588	0	0.0294	0.4412	0.0882	0
11	0	0.3235	0.0294	0	0.0294	0.2941	0.0882	0.0294
12	0	0.4412	0	0	0	0.4118	0.0588	0
13	0.0294	0.5294	0.0588	0	0.0588	0.5882	0.0882	0
14	0.0294	0.4118	0.0588	0	0.0588	0.4706	0.0882	0

Post-1994 (P = 2)

Info Set	Univariate				Multivariate			
	a1 = a2 = 0.5	a1 = 0.5	a2 = 0.5	a3 = 0.5	a1 = a2 = 0.5	a1 = 0.5	a2 = 0.5	a3 = 0.5
1	0.125	0.1562	0.6562	0.3438	0.0938	0.1562	0.625	0.3438
2	0.1562	0.4062	0.5	0.5625	0.1875	0.3438	0.5	0.5
3	0.0938	0.2188	0.5312	0.1875	0.0312	0.0625	0.5312	0.2188
4	0.1562	0.3438	0.4062	0.125	0.0938	0.3125	0.375	0.4062
5	0.0625	0.25	0.375	0.1875	0.0312	0.0625	0.3125	0.1875
6	0.0312	0.2812	0.125	0.125	0.0625	0.1875	0.4375	0.1875
7	0.0938	0.2812	0.2812	0.2188	0	0.25	0.25	0.5625
8	0.0312	0.2812	0.1875	0.1875	0	0.1562	0.1562	0.5
9	0.0312	0.25	0.125	0.125	0.0312	0.1562	0.375	0.2188
10	0	0.0625	0.0938	0.0625	0	0.125	0.3125	0.2812
11	0	0.0625	0.0312	0	0	0	0.0938	0.0625
12	0	0.125	0	0	0	0.0312	0.0312	0.1875
13	0.0312	0.0625	0.375	0.125	0.0312	0.0312	0.2812	0.125
14	0	0.0625	0.3125	0.2188	0	0	0.3125	0.125