# Multivariate Forecast Evaluation and Rationality Testing

| | |
|---|---|
| **Authors** | Ivana Komunjer, and Michael T. Owyang |
| **Working Paper Number** | 2007-047D |
| **Revision Date** | October 2011 |
| **Citable Link** | https://doi.org/10.20955/wp.2007.047 |
| **Suggested Citation** | Komunjer, I., Owyang, M.T., 2011; Multivariate Forecast Evaluation and Rationality Testing, Federal Reserve Bank of St. Louis Working Paper 2007-047. URL https://doi.org/10.20955/wp.2007.047 |

| | |
|---|---|
| **Published In** | Review of Economics and Statistics |
| **Publisher Link** | https://doi.org/10.1162/rest_a_00215 |

# 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 to test the rationality of vector forecasts without assuming independence across variables. When only one variable is of interest, the loss function reduces to the flexible asymmetric family proposed by Elliott, Komunjer and Timmerman (2005, 2006). Following their methodology, we derive a GMM test for multivariate forecast rationality that allows the forecaster's loss to be nonseparable across variables, and takes into account forecast estimation uncertainty. We use our test to study the joint rationality of macroeconomic forecasts in the growth rate of nominal output, CPI inflation rate, and short-term interest rate. [JEL: C32, C53] * †

*Key words and phrases.* forecast rationality; multivariate loss; asymmetries

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

*Acknowledgments*. We would like to thank the Editor, Mark Watson, as well as two anonymous referees for their comments and suggestions that led to an improved version of the manuscript. We benefited from conversations with Rick Ashley, Oscar Jorda, and Lutz Kilian. Kristie Engemann, Charles Gascon, and Kate Vermann provided research assistance.

## 1. Introduction

Forecasting models typically rely on the interaction of a large number of macroeconomic variables to generate predictions. Evaluations of such forecasts, on the other hand, are largely conducted one variable at a time (see a survey by Elliott and Timmerman, 2008). Effectively, such single variable analysis imposes independence across the variables being forecast. This means, for example, that a forecaster's loss in output prediction errors is assumed to be independent of her loss in inflation prediction errors. This clearly is an undesirable feature of any output-inflation forecast evaluation procedure, especially if we believe that losses are compounded by *jointly* overpredicting output and underpredicting inflation; in this case, a policymaker might be faced with unforeseen stagflation, an outcome worse than singly missing either prediction.

The purpose of this paper is to study the properties of vector forecasts in a framework that does not impose independence across variables in the forecaster's loss. Instead, we shall assume that the forecaster's objectives can be quantified by a new family of multivariate loss functions that are finitely parameterized yet flexible enough to account for asymmetries in the forecaster's preferences as well as interactions between the variables being forecast. The proposed family of multivariate losses permits identification and estimation of the parameters of forecasters' objectives, and allows to test for rationality of vector forecasts using over-identifying restrictions. Similar to the scalar case, vector forecasts are said to be rational if they are optimal under our multivariate loss; rejections of rationality should then be interpreted as rejections of the joint hypothesis of optimality and the particular functional form of the multivariate loss.[1]

Research on forecast rationality has a long standing history. Since the seminal works of Muth (1961) and Lucas (1973), rationality in expectation formation has been the cornerstone of economic models. With the availability of survey data such as the Livingston data or the Survey of Professional Forecasters (SPF), the econometric literature on forecast evaluation and rationality testing has seen a rapid growth (for an extensive review, see Pesaran and Weale, 2006). Rationality is tested under a variety of assumptions on the forecasters' objectives. If the latter are quadratic, then testing for rationality simply amounts to testing if the forecast errors have zero mean, and are uncorrelated with any information available at the time that the forecast is made. The most popular form of these tests is the Theil–Mincer–Zarnowitz regression (see, for example, Theil, 1958; Mincer & Zanowiz, 1963; Figlewski and Wachtel, 1981; Mishkin, 1981; Zarnowitz, 1986; Keane and Runkle, 1990).

One strand of the literature has, however, argued that asymmetric losses in which positive and negative forecast errors may be weighted differently might better represent the forecasters' objectives (see, for example, Zellner, 1986; Christoffersen and Diebold, 1997; Batchelor and Peel, 1998; Elliot, Komunjer, and Timmermann, 2005; Elliot, Komunjer, and Timmerman, 2006; Patton and Timmermann, 2007a; Patton and Timmermann, 2007b). In particular, Elliot, Komunjer, and Timmermann (2006; EKT hereafter) and Capistran and Timmermann (2009) find evidence for asymmetric loss in the SPF forecasts of output and inflation.[2] Under asymmetric loss, forecast efficiency tests based on the Theil–Mincer–Zarnowitz regressions are biased. EKT (2008) quantify the extent of the bias and its impact on the size and power of standard rationality tests. They propose an alternative GMM based approach that leads to correct inference regarding forecast rationality and at the same time allows for a parsimonious parameterization of asymmetry in the forecaster's loss.

An overwhelming majority of this work focuses on one variable at a time. Indeed, despite the availability of vector forecasts in surveys, few studies have conducted tests in a multivariate framework. Existing work on vector forecast rationality testing assumes that the losses are additive separable and quadratic in individual variables—see Kirchgässner and Müller (2006), for example. Additive separability implies that the marginal loss for one variable (say, output) is independent of the others (say, inflation). In other words, no interactions between the variables are allowed under separability. Perhaps surprisingly, little work has been undertaken on vector forecast evaluation that would allow for nonseparability and asymmetries in the forecaster's loss. This is even more striking as the decision theoretic literature has long recognized the importance of complementarities in the utility functions of decision makers. The main goal of this paper is to fill this gap.

As our analysis will show, if agents have directional preferences, falsely assuming additive separability of their objectives produces two biases: First, it can alter the results of rationality tests. Second, it may be reflected in a biased evaluation of the forecaster. The latter means the econometrician may incorrectly infer a greater degree of directional preference on the part of the forecaster. For example, if the forecaster is truly trying to forecast both output and inflation, then neglecting her inflation forecast objectives when evaluating her output forecasts may result in loss function estimates that are asymmetric, even if the forecaster's loss were perfecly symmetric to start with. In this paper, we argue that incorporating nonseparable losses can, in some cases, lessen the degree of asymmetry needed to justify the rationality of multivariate forecasts.

The practical importance of this effect is significant as it may help explain large degrees of asymmetries often found in the studies of univariate forecasts. For example, EKT (2005,

2008) find that in the context of their flexible loss functions, overpredictions of output are one and a half times costlier than underpredictions, which may be deemed implausible on economic grounds. This paper shows that a joint evaluation of output forecasts together with other variables such as inflation may lead to more plausible estimates of asymmetries.

Finally, let us point out that similar to the methods developed for evaluating single variable forecasts (EKT 2005, 2008), our forecast evaluation procedure takes into account the forecast estimation uncertainty (see, for example, West, 1996; West 2006; West and McCraken, 1998; McCracken, 2000; McCracken2007; Clark and McCracken, 2001; Clark and West, 2006; Clark and West, 2007; Corradi and Swanson, 2002; Corradi and Swanson, 2006; Corradi and Swanson 2007; Hubrich and West, 2010). Hence, we explicitly recognize the fact that the observed forecasts typically depend on estimates of the forecasting model.

The remainder of the paper is organized as follows: Section 2 develops the theoretical foundation for our multivariate approach. Here, we propose a new family of multivariate loss functions and derive their properties. Where appropriate, we emphasize the differences between the separable and nonseparable losses. In Section 3 we show that the asymmetry parameters of the proposed multivariate loss are identified. Section 4 then develops the estimation and rationality testing procedures. In Section 5 we present a Monte Carlo example that illustrates the properties of the proposed methods. The same section shows that misspecifying losses as separable leads to biased loss function estimates, and exacerbates the degree of asymmetry needed to rationalize forecasts. Section 6 introduces the data used in our empirical application and presents the results. Section 7 concludes. Technical details are relegated to an Appendix. Omitted proofs and additional details can be found in an Online Appendix.

## 2. Multivariate Forecasts and Loss Function

2.1. **Setup.** Hereafter, bold letters are used to denote vectors (e.g., $\mathbf{z}_t$) and matrices (e.g., $\mathbf{B}_0$). We consider a multivariate forecasting problem in which a forecaster is interested in forecasting future values of an $n$-vector of interest $\mathbf{y}_t$ ($n \geqslant 1$). Specifically, we let $\mathbf{f}_{t+s,t}$ denote the time-$t$ forecast of $\mathbf{y}_{t+s}$, where $s$ is the prediction horizon of interest, $s \geqslant 1$. The forecast vector $\mathbf{f}_{t+s,t}$ contains all the information comprised in the forecaster's information set $\mathcal{F}_t$, which is informative for $\mathbf{y}_{t+s}$. We let $\mathcal{F}_t$ include lagged values of $\mathbf{y}_t$ in addition to other covariates used to predict $\mathbf{y}_{t+s}$.

For simplicity, we focus on the one-step-ahead predictions of $\mathbf{y}_{t+1}$, which we denote by $\mathbf{f}_{t+1,t}$, knowing that all results developed in this case can readily be generalized to any $s > 1$.[3] Using the standard notation, we let $\mathbf{e}_{t+1}$ denote the $n$-vector of time-$t+1$ forecast errors, $\mathbf{e}_{t+1} = \mathbf{y}_{t+1} - \mathbf{f}_{t+1,t}$. The distribution of $\mathbf{y}_{t+1}$ conditional on $\mathcal{F}_t$ is denoted by $F_t^0(\cdot)$, $F_t^0(\mathbf{y}) = P(\mathbf{y}_{t+1} \leqslant \mathbf{y}|\mathcal{F}_t)$ for any $\mathbf{y} \in \mathbb{R}^n$ where $\leqslant$ denotes the usual partial order on $\mathbb{R}^n$. We shall assume that:

**A1.** *For all $t = 1, 2, ...$ the conditional distribution $F_t^0(\cdot)$ is absolutely continuous with a continuous density $f_t^0(\cdot)$ such that $f_t^0(\cdot) > 0$ on $\mathbb{R}^n$ a.s.-P.*

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, given an exponent $p$, $1 \leqslant p < \infty$, EKT map an asymmetry parameter $\tau$, $-1 \leqslant \tau \leqslant 1$, into a non-negative function of a scalar error $e \in \mathbb{R}$;[4] the resulting family of losses is flexible enough to accommodate the absolute value or quadratic losses, yet allows them to become asymmetric. We now extend their definition to a vector-valued argument $\mathbf{e} \in \mathbb{R}^n$. For this, let $\|\mathbf{u}\|_p$ denote the $l_p$-norm of any $n$-vector $\mathbf{u} = (u_1, \ldots, u_n)' \in \mathbb{R}^n$, i.e., $\|\mathbf{u}\|_p = (|u_1|^p + \ldots + |u_n|^p)^{1/p}$ for $1 \leqslant p < \infty$, and $\|\mathbf{u}\|_\infty = \max_{1 \leqslant i \leqslant n}(|u_i|)$; furthermore, let $\mathcal{B}_p^n$ denote the open unit ball in $\mathbb{R}^n$, i.e., $\mathcal{B}_p^n = \{u \in \mathbb{R}^n : \|\mathbf{u}\|_p < 1\}$.

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 our $n$-variate loss function as follows:

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

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

When $p = 1$, the multivariate loss $L_1(\boldsymbol{\tau}, \mathbf{e})$ can be used to define the geometric quantiles of the $n$-vector $\mathbf{e}$, as proposed in Chaudhuri (1996), for example. In a sense, $L_1(\boldsymbol{\tau}, \cdot)$ is a multivariate extension of the univariate "check" (or "tick") loss, which is well-known in the literature on quantile estimation Koenker and Bassett (1978). When $p > 1$, the expression of the $n$-variate loss $L_p(\boldsymbol{\tau}, \cdot)$ is entirely novel and not yet seen in the literature. We start by establishing some of its useful properties.

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

A proof of Proposition 1 is in the Appendix. The shape of the $n$-variate loss $L_p(\boldsymbol{\tau}, \cdot)$ is characterized by the exponent $p$, $1 \leqslant p < \infty$. On the other hand, the $n$-vector $\boldsymbol{\tau}$ quantifies the extent of asymmetry in $L_p(\boldsymbol{\tau}, \cdot)$. When $\boldsymbol{\tau} = \mathbf{0}$, the $n$-variate loss in Equation (1) reduces to $\|\mathbf{e}\|_p^p$, which is symmetric. On the contrary, for a nonzero $\boldsymbol{\tau}$, the magnitude of $\boldsymbol{\tau}$ (given by $\|\boldsymbol{\tau}\|_q$) measures the extent of deviation of the $n$-variate loss from the 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.[5]

When the variable of interest is of dimension $n = 1$ and the forecasts are univariate, the loss in Equation (1) becomes

$$L_p(\tau, e) = [|e| + \tau e] \, |e|^{p-1} .$$

Letting $\tau = 2\alpha - 1$ so that $0 < \alpha < 1$ and $p \geqslant 1$ as in EKT, the loss reduces to

$$L_p(\tau, e) = 2[1 - \alpha + \tau \mathbb{1}(e)]|e|^p,$$

where $\mathbb{1} : \mathbb{R} \to [0, 1]$ denotes the 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, 1999).[6] In the univariate case, this flexible loss family includes: (i) squared loss function $L_2(0, e) = e^2$, (ii) absolute deviation loss function $L_1(0, e) = |e|$, as well as their asymmetric counterparts obtained when $\tau \neq 0$ (or $\alpha \neq 1/2$) which are called (iii) quad-quad loss $L_2(\tau, e)$, and (iv) lin-lin loss $L_1(\tau, e)$.

2.3. **Asymmetry and Dependence Properties.** In order to gain further insight into the features of the loss $L_p(\boldsymbol{\tau}, \cdot)$, we consider in more detail the case $n = 2$. In the bivariate case, the forecaster cares about the magnitude and the sign of her errors $e_1$ and $e_2$, committed when jointly forecasting two variables of interest $y_1$ and $y_2$. In particular, when $p = 2$ we have

$$L_2(\boldsymbol{\tau}, \mathbf{e}) = e_1^2 + e_2^2 + (\tau_1 e_1 + \tau_2 e_2) \left(e_1^2 + e_2^2\right)^{1/2} .$$

The iso-loss curves corresponding to $L_2(\boldsymbol{\tau}, \mathbf{e}) = \text{constant}$, where $\mathbf{e} = (e_1, e_2)'$ and $\boldsymbol{\tau} = (\tau_1, \tau_2)'$, are as plotted in Figure 1. Note that unless $\tau_1 = \tau_2 = 0$, the loss $L_2(\boldsymbol{\tau}, \mathbf{e})$ is nonseparable and we have

$$L_2(\boldsymbol{\tau}, \mathbf{e}) \neq L_2(\tau_1, e_1) + L_2(\tau_2, e_2)$$

for general values of the forecast errors. In other words, the bivariate loss differs from a simple sum of the individual quad-quad losses.

This property generalizes to other values of the shape parameter $p$ as well as to higher dimensions. If $p$ is strictly greater than 1, then $L_p(\boldsymbol{\tau}, \mathbf{e})$ will in general differ from the sum of coordinate-wise univariate losses $L_p(\tau_1, e_1) + \ldots + L_p(\tau_n, e_n)$. Hence, minimizing the $n$-variate loss $L_p(\boldsymbol{\tau}, \mathbf{e})$ will in general produce an optimal $n$-vector $\mathbf{e}^*$ whose coordinates $e_i^*$ do not necessarily each minimize $L_p(\tau_i, e_i)$. In other words, $L_p(\boldsymbol{\tau}, \ \mathbf{e})$ captures not only the asymmetry but also the dependence between different coordinates of $\mathbf{e}$.

In the special case in which the forecaster's loss is symmetric so $\boldsymbol{\tau} = 0$, the $n$-variate loss becomes additively separable. That is, $L_p(\mathbf{0}, \mathbf{e})$ reduces to $L_p(0, e_1) + \ldots + L_1(0, e_n)$ for any value of $p \geqslant 1$.

## 3. Multivariate Forecast Rationality Condition

We now define multivariate forecast rationality. Similar to the single variable case, an $n$-variate forecast vector is said to be rational if it minimizes the expected value of the $n$-variate loss $L_p$ in Equation (1). Since the information sets available to the forecasters change over time, the expectation of the loss is conditional on $\mathcal{F}_t$. Hence, any rational forecast necessarily satisfies a set of orthogonality conditions implied by the first-order condition of the expected loss minimization. The key idea put forth in EKT is to use the forecast rationality condition to back out the forecaster's loss function parameters. We now extend this idea to the multivariate case and establish global identifiability of the asymmetry parameter $\boldsymbol{\tau}$ given the shape parameter $p$.

Hereafter, we shall focus on the asymmetry parameter $\boldsymbol{\tau}$ alone; in other words, all our results are conditional on the shape parameter $p$, which will be held fixed in all that follows.[7]

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 rationality condition:

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

When A2 holds, we say that the multivariate forecasts $\{\mathbf{f}^*_{t+1,t}\}$ are rational under the multivariate loss $L_p$. Implicit in Assumption A2 are several important properties: (1) the forecaster is an expected loss minimizer;[8] (2) 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; and (3) the forecaster's loss is of the form $L_p(\boldsymbol{\tau}, \cdot)$ given in Equation (1) with a true value $\boldsymbol{\tau}_0$ of the asymmetry parameter $\boldsymbol{\tau}$. The shape of the loss $p$ is treated as known.

We now derive a necessary and sufficient condition for multivariate forecast rationality, which provides the basis of our identification strategy. We need the following property:

**A3.** *Given* $p$, $1 \leqslant p < \infty$ , *and for all* $t = 1, \ldots$ *we have:*

$$E\left(\|\mathbf{y}_t\|_1^{p-1}\middle| \mathcal{F}_t\right) < \infty \ \textit{a.s.-}P$$

*and*

$$\left\| \mathbf{f}^*_{t+1,t} \right\|_1^{p-1} < \infty \ a.s.\text{-}P.$$

The conditions in Assumption A3 combined with the convexity of $L_p$ (established in Proposition 1) ensure—by Lebesgue's dominated convergence theorem—that we can safely differentiate the loss $L_p$ 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 rationality.

**Proposition 2.** *Let Assumptions A1 and A3 hold. Then A2 holds if and only if for all* $t = 1, \ldots$ *we have*

$$E\left[M|\mathcal{F}_t\right] = \mathbf{0}, a.s. - P, \tag{2}$$

*where*

$$M = p\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1}) + \boldsymbol{\tau}_{\mathbf{0}} \left\| \mathbf{e}^*_{t+1} \right\|_p^{p-1} + (p-1)\tau_0' \mathbf{e}^*_{t+1} \frac{\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1})}{\left\| \mathbf{e}^*_{t+1} \right\|_p}$$

*and for any* $\mathbf{u} = (u_1, \ldots, u_n)$ *we have let*

$$\boldsymbol{\nu}_p(\mathbf{u}) \equiv (\mathrm{sgn}(u_1)|u_1|^{p-1}, \ldots, \mathrm{sgn}(u_n)|u_n|^{p-1})'.$$

A proof of Proposition 2 is in the Appendix. While the necessity of the above first-order condition is obvious, the sufficiency result relies on the convexity of the loss $L_p(\boldsymbol{\tau}_0, \cdot)$ established in Proposition 1.

3.2. **Identification of Multivariate Loss Function Parameters.** Identification of the true value $\boldsymbol{\tau}_0$ of the multivariate loss parameter $\boldsymbol{\tau}$ used by the forecaster exploits the orthogonality conditions derived in Proposition 2. Given $p$, $1 \leqslant p < \infty$, consider an $\mathcal{F}_t$-measurable $d$-vector of instruments $\mathbf{x}_t$ and denote by $\mathbf{g}_p(\cdot; \mathbf{e}^*_{t+1}, \mathbf{x}_t)$ the $nd$-vector-valued function $\mathbf{g}_p(\cdot; \mathbf{e}^*_{t+1}, \mathbf{x}_t) : \mathcal{B}^n_q \to \mathbb{R}^{nd}$ defined by

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

The key element of our identification strategy is the following: under rationality, $\left\{\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t), \mathcal{F}_t\right\}$ is a martingale difference sequence. In particular, it then holds that for every $t \geqslant 1$, $E[\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = \mathbf{0}$. If for a given $p$, $1 \leqslant p < \infty$, $\boldsymbol{\tau}_0$ is the unique value of the $n$-variate asymmetry parameter $\boldsymbol{\tau} \in \mathcal{B}^n_q$ (with $1/p + 1/q = 1$) that solves these $nd$ orthogonality conditions $E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}^*_{t+1}, \mathbf{x}_t)] = \mathbf{0}$, then $\boldsymbol{\tau}_0$ is globally identified and consistently estimable using, for example, Hansen's (1982) GMM approach.

Finding primitive conditions for global GMM identification to hold is, in general, difficult (see Komunjer, 2011). Rather than simply assuming away the identification problem, we provide primitive conditions for the asymmetry parameter $\boldsymbol{\tau}$ to be globally identified when the exponent $p$ is known. For this, we shall assume that the variables entering the moment function $\mathbf{g}_p$ in Equation (3) satisfy the following properties:

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

**A5.** *Given* $p$, $1 \leqslant p < \infty$,

*(i)* $E[\left\|\mathbf{e}^*_{t+1}\right\|_1^{p-1} \left\|\mathbf{x}_t\right\|_1] < \infty$;

*(ii)* $\mathrm{rank}\Big(E\big[\left\|\mathbf{e}^*_{t+1}\right\|_p^{p-1}(\mathbf{I}_n\otimes\mathbf{x}_t)+(p-1)\left\|\mathbf{e}^*_{t+1}\right\|_p^{-1}(\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1})\otimes\mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}\big]\Big)=n,$

where $\mathbf{I}_n$ is $n\times n$ identity matrix.

**A6.** $\mathrm{rank}\left(E[\mathbf{x}_t\mathbf{x}_t']\right)=d.$

We are now able to state our identification result, whose detailed proof can be found in the Appendix.

**Theorem 1.** *Let Assumptions A1 through A6 hold. Given $p$, $1\leqslant p<\infty$, and for any $\boldsymbol{\tau}\in\mathcal{B}_q^n$, let*

$$Q(\boldsymbol{\tau})\equiv E[\mathbf{g}_p(\boldsymbol{\tau};\mathbf{e}^*_{t+1},\mathbf{x}_t)]'\mathbf{S}^{-1}(\tilde{\boldsymbol{\tau}})E[\mathbf{g}_p(\boldsymbol{\tau};\mathbf{e}^*_{t+1},\mathbf{x}_t)],$$

*where $\tilde{\boldsymbol{\tau}}$ is any fixed value of the asymmetry parameter $\boldsymbol{\tau}$ and*

$$\mathbf{S}(\boldsymbol{\tau})\equiv E\left[\mathbf{g}_p(\boldsymbol{\tau};\mathbf{e}^*_{t+1},\mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau};\mathbf{e}^*_{t+1},\mathbf{x}_t)'\right]$$

*Then $\boldsymbol{\tau}_0$ is the unique minimum of $Q(\boldsymbol{\tau})$ on $\mathcal{B}_q^n$.*

Theorem 1 shows that the true value $\boldsymbol{\tau}_0$ of the asymmetry parameter $\boldsymbol{\tau}$ of the multivariate loss $L_p$ is *globally* identified as the unique minimum of the GMM objective function:

$$\begin{aligned}Q(\boldsymbol{\tau}) \quad &= \quad E[\mathbf{g}_p(\boldsymbol{\tau};\mathbf{e}^*_{t+1},\mathbf{x}_t)]'\\ &\qquad\times\mathbf{S}^{-1}(\tilde{\boldsymbol{\tau}})E[\mathbf{g}_p(\boldsymbol{\tau};\mathbf{e}^*_{t+1},\mathbf{x}_t)],\end{aligned}$$

where $\tilde{\boldsymbol{\tau}}$ is any fixed value of the asymmetry parameter $\boldsymbol{\tau}$. The proof of Theorem 1 exploits the linearity of the moment function $\mathbf{g}_p$ in $\boldsymbol{\tau}$ and the positive definiteness of $\mathbf{S}(\tilde{\boldsymbol{\tau}})$. It is worth pointing out that the latter holds globally on the parameter space $\mathcal{B}_q^n$ and not only at $\boldsymbol{\tau}_0$.[9]

## 4. GMM Estimation and Testing

The identification result of Theorem 1 is the starting point of our estimation and multivariate forecast rationality testing procedures, which we now discuss.

4.1. **GMM Estimation.** Having established that the asymmetry parameter $\boldsymbol{\tau}$ is globally identified on $\mathcal{B}_q^n$, we turn to the problem of estimating $\boldsymbol{\tau}_0$ by minimizing an empirical counterpart of $Q(\boldsymbol{\tau})$. It is important to note that the optimal forecast errors $\mathbf{e}_{t+1}^*$ used to define $Q(\boldsymbol{\tau})$ are unobservable in reality. Instead, for every $t$, $R \leqslant t \leqslant T = P + R - 1$, the forecast evaluator observes $\hat{\mathbf{e}}_{t+1} = \mathbf{y}_{t+1} - \hat{\mathbf{f}}_{t+1,t}$, which implicitly incorporates all of the forecast estimation uncertainty embodied in $\hat{\mathbf{f}}_{t+1,t}$. Explicit here is the assumption that the forecaster uses data from 1 to $R$ to compute her first forecast $\hat{\mathbf{f}}_{R+1,R}$ of $\mathbf{y}_{R+1}$; the estimation window is then rolled on, and data from 2 to $R+1$ is used to compute $\hat{\mathbf{f}}_{R+2,R+1}$. Hence, the evaluation exercise starts at $R+1$ and ends at $T+1 = R+P$.[10] Whether the forecaster uses one model to forecast all variables of interest or different models for different variables does not alter the validity of the proposed method, provided, however, that all the individual models contain the same conditioning variables.

Now, given $p$, $1 \leqslant p < \infty$, and given the observations $((\mathbf{x}_R', \hat{\mathbf{e}}_{R+1}')', \ldots, (\mathbf{x}_T', \hat{\mathbf{e}}_{T+1}')')'$, the GMM estimator of the $n$-variate loss asymmetry parameter $\boldsymbol{\tau}_0$, denoted by $\hat{\boldsymbol{\tau}}_P$, can be defined as a solution to the minimization problem:

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

where $\hat{\mathbf{S}}$ is a consistent estimator of

$$\mathbf{S} = E[\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)'].$$

In order to make sure that the forecast estimation uncertainty does not interfere with our rationality test, we impose a set of restrictions on how the observed forecaster's $n$-vector errors $\{\hat{\mathbf{e}}_{t+1}\}_{t=R}^{T}$ differ from their optimal counterparts $\{\mathbf{e}^*_{t+1}\}_{t=R}^{T}$.

**A7.** *For any $\varepsilon > 0$, and every $R \leqslant t \leqslant T$,*

$$\lim_{R,P\to\infty} \Pr\left(\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 > \varepsilon\right) = 0.$$

It is important to note that Assumption A7 does not presuppose any knowledge of the forecasting model on behalf of the forecast evaluator (the econometrician). When the latter has more information about the model used for forecasting, then Assumption A7 can be replaced with more primitive conditions. For instance, if the forecaster's model is some smooth data dependent function of a finite dimensional parameter $\beta$, then a primitive condition for A7 is simply that the recursive estimates $\hat{\beta}_t$ are all consistent for the pseudo-true value $\beta^*$ that minimizes the forecaster's expected loss, i.e., $\hat{\beta}_t \xrightarrow{p} \beta^*$—see West (2006), for example. Rather than putting restrictions on the underlying forecasting model, we state Assumption A7 directly in terms of the forecast errors.

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 $\{(\mathbf{e}^{*\prime}_{t+1}, \mathbf{x}'_t)'\}$ 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* $\alpha$*-mixing with mixing coefficient* $\alpha$ *of size* $-r/(r-2)$, $r > 2$, *and, given* $p$, $1 \leqslant p < \infty$, *there exist some* $\varepsilon > 0$, $\Delta_1 > 0$ *and* $\Delta_2 > 0$ *such that*

$$E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p-1)(2r+\varepsilon)}] \leqslant \Delta_1 < \infty$$

*and*

$$E[\|\mathbf{x}_t\|_1^{2r+\varepsilon}] \leqslant \Delta_2 < \infty.$$

When the forecast evaluator has additional knowledge about the forecaster's information set, then it is possible to state more primitive conditions for A8. For instance, if the forecaster's model is based on a vector of variables $\mathbf{w}_t$ that are $\mathcal{F}_t$-measurable, then, because of the rolling nature of the forecasting exercise, her forecast errors are of the form $\hat{\mathbf{e}}_{t+1} = \mathbf{e}(\mathbf{w}_t, \ldots, \mathbf{w}_{t-R+1})$. The strict stationarity and $\alpha$-mixing of the forecast errors is then implied by the strict stationarity and $\alpha$-mixing of the sequence $\{\mathbf{w}_t\}$.[11]

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

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

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

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

**Theorem 2.** *Let Assumptions A1 through A8 hold. Then, given $p$, $1 \leqslant p < \infty$, we have $\hat{\boldsymbol{\tau}}_P \xrightarrow{p} \boldsymbol{\tau}_0$ as $(R, P) \to \infty$.*

A complete proof of Theorem 2 is in the Online Appendix.

4.2. **Forecast Rationality Test Statistic.** Our test for multivariate forecast rationality defined in A2 comes in the form of a J-test. Hence, it necessitates the derivation of the asymptotic distribution of our GMM estimator $\hat{\boldsymbol{\tau}}_P$, to which we now turn. We start by strengthening our stationarity assumption A4 as follows:

**A4′.** *The process $\{(\mathbf{e}^{*\prime}_{t+1}, \mathbf{x}'_t)'\}$ is strictly stationary and $\alpha$-mixing with mixing coefficient $\alpha$ of size $-r/(r-2)$, $r > 2$, and, given $p$, $1 \leqslant p < \infty$, there exist some $\epsilon > 0$, $\Delta_3 > 0$ such that*

$$E[\left\|\mathbf{e}^*_{t+1}\right\|_1^{(p-1)(2r+\epsilon)}] \leqslant \Delta_3 < \infty.$$

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′.** *For some small $\varepsilon$ in $(0, 1/2)$ we have:*

*(i) $R^{1-2\varepsilon}/P \to \infty$ as $R \to \infty$ and $P \to \infty$, and*

*(ii) $\lim_{R,P\to\infty} \Pr\left(\sup_{R\leqslant t\leqslant T} \left\|R^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1})\right\|_1 > \delta\right) = 0$ for any $\delta > 0$.*

The above condition ensures that the forecast estimation uncertainty, embodied in $\hat{\mathbf{e}}$, does not affect the asymptotic distribution of our GMM estimator $\hat{\boldsymbol{\tau}}_P$. Note that A7′(i) imposes a condition on the relative growth of sample sizes $R$ and $P$. In particular, this assumption implies that $P/R \to 0$ as both $R$ and $P$ get large. Assumption A7$'$(ii), on the other hand, strengthens the requirement in A7 by making the latter uniform across the observation window. Finally, we need two additional new assumptions:

**A9.** *Given $p$, $1 \leqslant p < \infty$, we have:*

$$E\left(\sup_{c\in(0,1)} \left\| c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}^*_{t+1} \right\|_1^{p-2}\right) < \infty,$$

and

$$E\left(\left\|\mathbf{x}_t\right\|_1 \sup_{c\in(0,1)} \left\| c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}^*_{t+1} \right\|_1^{p-2}\right) < \infty.$$

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

We are now ready to state our asymptotic distribution result for $\hat{\boldsymbol{\tau}}_P$, whose detailed proof is in the Online Appendix.

**Theorem 3.** *Let Assumptions A1-A3, A4', A5-A6, A7', A8-A10 hold. Then, given $p$, $1 \leqslant p < \infty$, we have:*

$$\sqrt{P}(\hat{\boldsymbol{\tau}}_P - \boldsymbol{\tau}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1}), as (R,P) \to \infty,$$

*where*

$$\mathbf{S} = E[\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)'],$$

*and*

$$\mathbf{B}^* \equiv E[\left\|\mathbf{e}^*_{t+1}\right\|_p^{p-1} (\mathbf{I}_n \otimes \mathbf{x}_t) + (p-1)\left\|\mathbf{e}^*_{t+1}\right\|_p^{-1}$$
$$\times(\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1}) \otimes \mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}].$$

The asymptotic normality result of Theorem 3 is the basis for our multivariate forecast rationality test. When the dimension of the $d$-vector of instruments $\mathbf{x}_t$ used in Equation (3) is large enough ($d > 1$), then a test for overidentification provides a test of the multivariate forecast rationality condition in A2. More formally, we have the following corollary to our Theorem 3:

**Corollary 4.** *Let the assumptions of Theorem 3 hold. Then a test of rationality of the $n$-vector forecasts $\{\mathbf{f}^*_{t+1,t}\}$ under the $n$-variate loss $L_p$ can be conducted with $d > 1$ instruments $\mathbf{x}_t$ through the J-test statistic*

$$\begin{aligned}\hat{J}_P \equiv & P^{-1}\left[\sum_{t=R}^{R+P-1} \mathbf{g}_p(\hat{\boldsymbol{\tau}}_P; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)\right]' \\ & \times \hat{\mathbf{S}}^{-1}\left[\sum_{t=R}^{R+P-1} \mathbf{g}_p(\hat{\boldsymbol{\tau}}_P; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)\right] \sim \chi^2_{n(d-1)},\end{aligned}$$

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

The usual test for overidentification can be used to test the hypothesis that the vector forecasts are rational with respect to the information available in the instrument set $\mathbf{x}_t$ ($d > 1$) within the class of our multivariate loss functions $L_p$. The test provides an answer to the question whether, for a given set of instruments $\mathbf{x}_t$, there exists *some* value $\tau_0$ for which the forecasts are optimal. Note that $n$ degrees of freedom are used in the estimation of the multivariate loss parameter.

## 5. Monte Carlo Simulation Results

In this section, we examine the behavior of the GMM estimator $\hat{\boldsymbol{\tau}}_P$ and study the properties of the proposed multivariate forecast rationality test. Moreover, we illustrate the effects of wrongly assuming the forecaster's loss to be separable.

5.1. **Performance of the proposed estimator and test.** To match with the empirical application, we focus on a three-variable case, $n = 3$. We assume that the observables $\mathbf{y}_t$ are generated from a VAR(1):

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Here, $\boldsymbol{\varepsilon}_t$ is iid multivariate normal with zero mean and covariance matrix $\Sigma = 0.01 \times \mathbf{I}_3$, $\mathbf{c} = (.01, .01, .01)'$, and $\mathbf{A} = 0.5 \times \mathbf{I}_3$. We construct samples of size $T = R + P - 1$ after discarding the first 1000 periods to remove any initial values effects. The forecaster uses a rolling window of size $R$ to construct $P$ one-period-ahead forecasts $\mathbf{f}_{t,t+1}$ that minimize the expected value of the multivariate loss $L_2$ conditional on the data and a (correctly specified) VAR(1). The observed one-period-ahead forecast is then $\hat{\mathbf{f}}_{t+1,t} = \hat{\mathbf{c}} + \hat{\mathbf{A}}\mathbf{y}_t$, where

$$(\hat{\mathbf{c}}, \hat{\mathbf{A}}) \equiv \arg\min_{(\mathbf{c},\mathbf{A})} P^{-1} \sum_{t=1}^{P} L_2(\boldsymbol{\tau}_0, \mathbf{y}_{t+1} - \mathbf{c} - \mathbf{A}\mathbf{y}_t),$$

and $\boldsymbol{\tau}_0$ denotes the true value of the forecaster's asymmetry parameter. 1000 Monte Carlo simulation experiments were undertaken for different choices of $\boldsymbol{\tau}_0$, $R$, and $P$. In each simulation, using the methods described in the previous sections, we construct the GMM estimator $\hat{\boldsymbol{\tau}}_P$ for various choices of the instrument set $\mathbf{x}_t$. When overidentifying restrictions are present ($d > 1$), we also perform the J-test of multivariate forecast rationality based on the sequence of observed forecast errors

$$\{\hat{\mathbf{e}}_{t+1}\}_{t=R}^{T} = \{\mathbf{y}_{t+1} - \hat{\mathbf{c}} - \hat{\mathbf{A}}\mathbf{y}_t\}_{t=R}^{T}.$$

Panel (a) of Table 1 reports the $\boldsymbol{\tau}_0$ estimates obtained when the instrument set includes a constant only (i.e., $\mathbf{x}_t = 1$), so the model is exactly identified. The GMM estimator performs very well overall, even for values of $\boldsymbol{\tau}_0$ different from zero. Note that when $\boldsymbol{\tau}_0 \neq \mathbf{0}$, the estimator exhibits slight small sample bias towards symmetry, i.e., underestimates $\boldsymbol{\tau}_0$ in absolute value.

Panel (b) of Table 1 reports the $\boldsymbol{\tau}_0$ estimates using the instrument set $\mathbf{x}_t = (1, y_{1t})'$. The GMM estimator still performs well overall with small sample bias that seems smaller in absolute value compared to the just-identified case. Given that the dimension of the instrument is $d = 2$ there are $n(d-1) = 3$ overidentifying restrictions that allow to test whether the rationality of forecasts is compatible with some $\boldsymbol{\tau}_0$. The last column reports the empirical rejection probability of the proposed J-test. Size is well controlled for combinations of $R$ and $P$ which are compatible with our requirement in A7'(i). Not surprisingly, there are size distortions when $R/P \leqslant 1$ which confirms the importance of controlling how large the out-of-sample period is relative to the in-sample. Additional out-of-sample observations help control the size.

Additional results can be found in Panels (c) and (d) of Table 1 reported in the Online Appendix. Overall, the findings confirm that small sample bias is less of a problem when more observations are available (e.g. $P = 250$) and when the ratio $R/P$ is large. Unlike for the estimation, including extra instruments deteriorates the performance of the J-test, which becomes severely undersized as $d$ increases. Those findings are consistent with the so-called "many instrument problem" which biases GMM in the direction of least squares and thus causes size distortions in the J-test. As is well known, GMM estimator suffers from large biases when the degree of overidentification becomes large; hence, we would expect the size properties of our J-test to deteriorate whenever the ratio $n(d-1)/P$ is not close to zero.[12] This finding has important implications for the empirical analysis to follow, as it dictates the maximum number of instruments one can safely use with sample sizes around $P = 150$, which will be the case here.

5.2. **The Effect of Misspecifying Preferences.** We now examine the implications of falsely assuming separability of the forecaster's loss. For this, we construct a set of Monte Carlo experiments in which the forecaster's true loss function is our multivariate loss $L_2(\boldsymbol{\tau}, \mathbf{e})$; the forecast evaluation is, however, done under the assumption that the latter is simply the sum of univariate losses $L_2(\tau_1, e_1) + \ldots + L_2(\tau_n, e_n)$, where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)'$ and $\mathbf{e} = (e_1, \ldots, e_n)'$. Similar to previously, we consider three-dimensional $(n = 3)$ vectors $\mathbf{y}_t$ that are generated from a VAR(1): $\mathbf{y}_t = \mathbf{c} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, with $\boldsymbol{\varepsilon}_t$ that is iid multivariate normal with zero mean and covariance matrix $\boldsymbol{\Sigma}$. While $\mathbf{c} = (.01, .01, .01)'$ is kept unchanged, the remaining parameter values are different from previously. We now let both $\mathbf{A}$ and $\boldsymbol{\Sigma}$ be non-diagonal and assume that

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.2 & 0.14 \\ 0.2 & 0.5 & 0.14 \\ 0.14 & 0.14 & 0.5 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = 0.01 \times \begin{bmatrix} 1 & 0.9 & 0.17 \\ 0.9 & 1 & 0.46 \\ 0.17 & 0.46 & 1 \end{bmatrix}.$$

This choice of $\mathbf{A}$ and $\boldsymbol{\Sigma}$ results in highly correlated individual components of the vector $\mathbf{y}_t$. We further assume that $p = 2$ and the true value of the forecaster's asymmetry parameter is $\boldsymbol{\tau}_0 = (0, -0.7, -0.6)'$. Hence, the loss $L_2(\boldsymbol{\tau}_{\mathbf{0}}, \mathbf{e}_{t+1})$ is highly asymmetric in the second and third components of the forecast error; however, it is perfectly symmetric in the first component. As previously, we use a rolling window of size $R = 250$ to construct $P = 150$ one-period-ahead forecasts.

Table 2 presents the results from 1000 Monte Carlo replications of the above parameterization using five information sets. The effects of misspecifying the loss as separable when the forecaster has asymmetric preferences and the variables are correlated become evident. Under separable loss, the forecaster appears to have asymmetric preferences for the first variable (with $\hat{\tau}_1$ ranging from $-0.14$ to $-0.20$ according to the choice of instruments) even though her true preferences are symmetric with $\tau_{01} = 0$. For the two other variables, misspecification results in more asymmetric estimates for $\tau_{02} = -0.7$ and $\tau_{03} = -0.6$, with $\hat{\tau}_2$ and $\hat{\tau}_3$ ranging from $-0.91$ to $-0.97$ across information sets.

These findings have important implications on the interpretation of the univariate rationality test results. For example, consider the findings of EKT (2008) obtained when testing the rationality of SPF forecasts for GDP growth. Using their flexible univariate loss specification, EKT (2008) find the individual $\alpha$ estimates consistent with rationality to be clustered around 0.4. Translated into our setup, this would correspond to $\tau$ estimates clustered around $-0.2$. From this evidence, EKT (2008) conclude (p.141) that "asymmetry in the loss function is required to overturn rejections of the null hypothesis [that GDP growth forecasts are rational]." Our Monte Carlo experiment offers an alternative interpretation of this finding: the forecasters' losses are perfectly symmetric in their GDP forecast errors; however, those errors are not independent of the forecast errors committed in other variables.

When the true loss is nonseparable across variables, errors in one variable affect the marginal loss in the other variables. To see why this occurs, consider again the three-variable case:

$$L_2(\boldsymbol{\tau}, \mathbf{e}) = e_1^2 + e_2^2 + e_3^2 + (\tau_1 e_1 + \tau_2 e_2 + \tau_3 e_3)\left(e_1^2 + e_2^2 + e_3^2\right)^{1/2}.$$

Then,

$$\frac{\partial L_2(\boldsymbol{\tau}, \mathbf{e})}{\partial e_1} = 2e_1 + \tau_1\left(e_1^2 + e_2^2 + e_3^2\right)^{1/2} + (\tau_1 e_1 + \tau_2 e_2 + \tau_3 e_3)\frac{e_1}{\left(e_1^2 + e_2^2 + e_3^2\right)^{1/2}}.$$

Thus, even if $\tau_1 = 0$ (i.e., the forecaster's preferences are symmetric over the first variable), the marginal loss in the first variable $\partial L_2(\boldsymbol{\tau}, \mathbf{e})/\partial e_1$ depends on the remaining errors $(e_2, e_3)$. Wrongly assuming separability (i.e., that $\partial L_2(\boldsymbol{\tau}, \mathbf{e})/\partial e_1$ is a function of $e_1$ alone) then results in biased estimates of $\boldsymbol{\tau}$.

## 6. Empirical Application

We illustrate the performance of our procedure in a situation in which three macroeconomic variables are jointly forecast: growth rate in output $(y)$, CPI inflation rate $(\pi)$, and short-term interest rate $(r)$. Examples of models using these variables include Taylor's (1993) interest rate targeting rule, monetary VARs (Christiano, Eichenbaum, and Evens,k 1999), optimizing ISLM models (McCallum and Nelson, 1999), and reduced-form New Keynesian models (Clarida, Galì, 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 thus expect the forecaster's loss to be nonseparable across variables.

6.1. **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. Our sample includes forecasts from 1976:08 to 2004:12.

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. Short-term interest rate realizations are the yearly averages.

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. 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. Finally, forecasters with fewer than 80 valid observations are dropped from the sample. This leaves 57 firms with an average of 171 valid observations per firm.

The set of instruments $\mathbf{x}_t$ used in the implementation of our procedure 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.[13] The instrument sets are defined in Table 3. As a baseline for comparison, we repeat each test under the assumption of separability and the joint assumption of separability and symmetry.

6.2. **Multivariate Rationality Test Results.** Table 4 illustrates the effect of testing rationality using the nonseparable loss $L_2(\boldsymbol{\tau}, \mathbf{e})$, separable loss

$$L_2(\tau_1, e_1) + L_2(\tau_2, e_2) + L_2(\tau_3, e_3),$$

as well as separable symmetric loss

$$L_2(0, e_1) + L_2(0, e_2) + L_2(0, e_3).$$

We report the percentages of forecasters for which rationality could be rejected at the 10%, 5% and 1% levels, for each set of instruments. $p = 2$ is kept fixed in all configurations. For any instrument set, both asymmetric loss functions reject rationality for a lower percentage of forecasters than the separable symmetric baseline. The percentage of forecasters for which rationality is rejected under nonseparable loss is relatively close to that under separable loss. Rationality under separable symmetric loss is, however, overwhelmingly rejected. Interestingly, the smallest percentage of forecasters are found to be rational with respect to the unemployment rate (information set 5).

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

EKT found that the addition of asymmetric loss alone can increase the percentage of forecasters for which rationality is confirmed. However, for the separable loss functions implied by EKT, this finding often requires substantial directional asymmetry in the forecasters' loss functions. Allowing the forecaster's marginal loss to depend on all of the variables being forecast may ameliorate this problem. Recall that interpretation of the asymmetry parameters $(\tau_y, \tau_\pi, \tau_r)$ depends on their values relative to the baseline 0. Values greater (less) than

0 indicate greater losses for positive (negative) forecast errors. Table 5 provides summary statistics for the distributions of the estimated asymmetry parameters across forecasters. For the first information set, $\mathbf{x}_t = 1$, we compute the mean and median of the asymmetry parameters for all forecasters with 80 valid observations. For all other instrument sets, the mean and median asymmetry parameters are computed only for those forecasters for which rationality could not be rejected at the 5% level. Figure 2 provides graphical representations of one of these distributions.

The joint directionality in preferences is consistent across forecasters. More than half of the forecasters exhibit higher loss when jointly overpredicting output, overpredicting the short-term interest rate, and underpredicting inflation. These directional 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.

The salient result for nonseparable loss lies in the magnitude of the estimated asymmetry parameter relative to that obtained under separability. We find that the degree of directional asymmetry is reduced once separability is relaxed. To see this, we compute the norm of the preference vector for each forecaster. Because $\boldsymbol{\tau} = 0$ reflects symmetric preferences, the magnitude of $\|\hat{\boldsymbol{\tau}}\|_2$ provides a summary of the overall degree of asymmetry. Table 5 shows mean and median $\|\hat{\boldsymbol{\tau}}\|_2$ obtained under nonseparable and separable losses. For each instrument set, the overall degree of asymmetry required to make the forecaster consistent with rationality is smaller in the nonseparable case. Figure 2 plots the absolute values $|\hat{\tau}_i|$ obtained under nonseparable loss against those obtained under separable loss. With few exceptions, the estimates lie below the 45° line, indicating a decline in the estimated asymmetry once we allow the loss to be nonseparable. Assuming separability leads the econometrician to infer more directional asymmetry than may actually be warranted.

## 7. Conclusion

Recognizing the multivariate nature of most forecasting problems has important implications for the prospects of rational expectations in macroeconomic models. In a univariate setup, EKT (2005, 2008) 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 that considers the forecast series in isolation. Our findings show that imposing separability of the forecaster's loss across 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 (the short-term interest rate) when establishing their forecasts for output and inflation. The assumption of additive separability in forecast loss is akin to the assumption that forecasters believe 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.

## Appendix

*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 $L_p(\boldsymbol{\tau}, \cdot)$ is a convex function on $\mathbb{R}^n$: i.e., that

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

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

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

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 (1952) applied with $r \equiv p-1$ and $s \equiv 1$ (the latter shows that, for every $(a_1, a_2) \in \mathbb{R}_+^2$ and $0 < r < s$, we have

$$(a_1^s + a_2^s)^{1/s} \leqslant (a_1^r + a_2^r)^{1/r}),$$

and the last inequality uses $0 < \lambda < 1$. 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

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

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

*Proof of Proposition 2.* For any continuously differentiable real function $f:\mathbb{R}^n \rightarrow \mathbb{R}$, we let $\nabla_{\mathbf{u}} f(\mathbf{u})$ denote the gradient of $f(\cdot)$ with respect to $\mathbf{u}$,

$$\nabla_{\mathbf{u}} f(\mathbf{u}) \equiv \left(\partial f(\mathbf{u})/\partial u_i, \ldots, \partial f(\mathbf{u})/\partial u_n\right)'.$$

Fix $p$, $1 \leqslant p < \infty$, and $\boldsymbol{\tau}_0 \in \mathcal{B}_q^n$, where $1/p+1/q=1$. Differentiating the loss $L_p\left(\boldsymbol{\tau}_0,\cdot\right)$ in Equation (1), we have

$$\nabla_{\mathbf{e}} L_p\left(\boldsymbol{\tau}_0,\mathbf{e}\right)=p\boldsymbol{\nu}_p(\mathbf{e})+\boldsymbol{\tau}_0\|\mathbf{e}\|_p^{p-1}+(p-1)\boldsymbol{\tau}'\mathbf{e}\frac{\boldsymbol{\nu}_p(\mathbf{e})}{\|\mathbf{e}\|_p}, \tag{9}$$

for all $\mathbf{e} \in \mathbb{R}^n \backslash \mathcal{E}$ where $\mathcal{E}=\{\mathbf{e} \in \mathbb{R}^n : e_i=0 \text{ for some } i\}$ is a set of points of non-differentiability of $L_1\left(\boldsymbol{\tau}_0,\cdot\right)$ which is of measure zero. Note that in the univariate case $n=1$, the expression in Equation (9) reduces to

$$\nabla_e L_p\left(\tau_0,e\right)=p[\tau_0+sgn(e)]|e|^{p-1}$$

(see Equation (8) in EKT, p. 1121). By triangular inequality and norm equivalence,

$$\begin{aligned}\left\|\nabla_{\mathbf{e}}L_p\left(\boldsymbol{\tau}_0,\mathbf{e}\right)\right\|_1 &\leqslant p\left\|\mathbf{e}\right\|_{p-1}^{p-1}+n\left\|\mathbf{e}\right\|_{p-1}^{p-1}+\\ &\quad (p-1)n\left\|\mathbf{e}\right\|_1\left\|\mathbf{e}\right\|_{p-1}^{p-1}/\left\|\mathbf{e}\right\|_p\\ &\leqslant C_1\left\|\mathbf{e}\right\|_1^{p-1},\end{aligned}$$

with $C_1<\infty$ when $\mathbf{e}\notin\mathcal{E}$ and $\left\|\nabla_{\mathbf{e}}L_p\left(\boldsymbol{\tau}_0,\mathbf{0}\right)\right\|_1\leqslant C_2<\infty$. By Assumption A3, we have

$$E(\left\|\mathbf{y}_t\right\|_1^{p-1}|\mathcal{F}_t)<\infty\text{ a.s.-}P$$

and

$$\left\|\mathbf{f}_{t+1,1}^*\right\|_1^{p-1}<\infty\text{ a.s.-}P,$$

which together with the fact that

$$\left\|\mathbf{e}_{t+1}^*\right\|_1^{p-1}\leqslant C_3\left(\left\|\mathbf{y}_t\right\|_1^{p-1}+\left\|\mathbf{f}_{t+1,1}^*\right\|_1^{p-1}\right)\text{ a.s.-}P$$

then ensure

$$E[\nabla_{\mathbf{e}}L_p\left(\boldsymbol{\tau}_0,\mathbf{e}_{t+1}^*\right)|\mathcal{F}_t]<\infty\text{ a.s.-}P.$$

This last condition combined with the convexity of $L_p\left(\boldsymbol{\tau}_0,\cdot\right)$, which implies that $L_p\left(\boldsymbol{\tau}_0,\cdot\right)$ is locally Lipschitz, allows us to interchange the order of differentiation and expectation to get

$$\nabla_{\mathbf{e}}E\left[L_p\left(\boldsymbol{\tau}_0,\mathbf{e}_{t+1}^*\right)|\mathcal{F}_t\right]=E[\nabla_{\mathbf{e}}L_p\left(\boldsymbol{\tau}_0,\mathbf{e}_{t+1}^*\right)|\mathcal{F}_t].$$

This, combined with the gradient expression in Equation (9) and with the convexity of the loss $L_p\left(\boldsymbol{\tau}_0,\cdot\right)$, shows that the first-order condition in Equation (2) is necessary and sufficient for A2 to hold. □

*Proof of Theorem 1.* The result of Theorem 1 follows by combining Lemmas 5 and 6 below.

**Lemma 5.** *Let Assumptions A1 through A5 hold. Given* $p$*,* $1 \leqslant p < \infty$*, and for any* $\boldsymbol{\tau} \in \mathcal{B}_q^n$*, let*

$$Q(\boldsymbol{\tau}) \equiv E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)]'\mathbf{S}^{-1}E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)],$$

*with* $\mathbf{S}$ *positive definite. Then* $\boldsymbol{\tau}_0$ *is the unique minimum of* $Q(\boldsymbol{\tau})$ *on* $\mathcal{B}_q^n$*.*

Put in words, Lemma 5 gives primitive conditions for the parameter $\boldsymbol{\tau}_0$ to be globally identified by the $nd$ moment conditions $E[\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t)] = \mathbf{0}$. The key condition for the global identification of $\boldsymbol{\tau}_0$ to hold—in addition to those given by Assumptions A1-A5—is that the matrix $\mathbf{S}$ in the GMM objective function $Q(\boldsymbol{\tau})$ be positive definite. This weighting matrix is usually set to be equal to

$$\mathbf{S} \equiv E[\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t)'].$$

In order to ensure that $\mathbf{S}$ is positive definite, we need the covariance matrix of $d$-vector of instruments $\mathbf{x}_t$ to be of full rank, stated in Assumption A6.

**Lemma 6.** *Let Assumptions A1 through A6 hold. Given* $p$*,* $1 \leqslant p < \infty$*, and for any* $\boldsymbol{\tau} \in \mathcal{B}_q^n$*, let*

$$\mathbf{S}(\boldsymbol{\tau}) \equiv E\left[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)'\right].$$

*Then* $\mathbf{S}(\boldsymbol{\tau})$ *is positive definite.*

□

*Proof of Lemma 5.* First, we show that for every $\boldsymbol{\tau} \in \mathcal{B}_q^n$, $E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)]$ exists and is finite. For this, note that given $p$, $1 \leqslant p < \infty$, and for any $\boldsymbol{\tau} \in \mathcal{B}_q^n$, we can write:

$$\begin{aligned}
&\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t) \\
&= p(\boldsymbol{\nu}_p(\mathbf{e}_{t+1}^*) \otimes \mathbf{x}_t) + \\
&\left( \left\| \mathbf{e}_{t+1}^* \right\|_p^{p-1} (\mathbf{I}_n \otimes \mathbf{x}_t) + (p-1) \left\| \mathbf{e}_{t+1}^* \right\|_p^{-1} (\boldsymbol{\nu}_p(\mathbf{e}_{t+1}^*) \otimes \mathbf{x}_t) \mathbf{e}_{t+1}^{*\prime} \right) \boldsymbol{\tau} \qquad (10) \\
&= \mathbf{a}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t) + \mathbf{B}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t) \boldsymbol{\tau}, \qquad (11)
\end{aligned}$$

where we define the $nd \times 1$ vector

$$\mathbf{a}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t) \equiv p(\boldsymbol{\nu}_p(\mathbf{e}_{t+1}^*) \otimes \mathbf{x}_t) \qquad (12)$$

and the $nd \times n$ matrix

$$\begin{aligned}
\mathbf{B}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t) \equiv{}& \left\| \mathbf{e}_{t+1}^* \right\|_p^{p-1} (\mathbf{I}_n \otimes \mathbf{x}_t) + \\
& (p-1) \left\| \mathbf{e}_{t+1}^* \right\|_p^{-1} (\boldsymbol{\nu}_p(\mathbf{e}_{t+1}^*) \otimes \mathbf{x}_t) \mathbf{e}_{t+1}^{*\prime}. \qquad (13)
\end{aligned}$$

For any $m \times n$-matrix $\mathbf{A} = (a_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$, let $\|\mathbf{A}\|_\infty = \max_{1 \leqslant i,j \leqslant n}(|a_{ij}|)$. Then, it follows from (11) that for any $\boldsymbol{\tau} \in \mathcal{B}_q^n$, we have:

$$\begin{aligned}
\|\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)\|_1 &\leqslant \|\mathbf{a}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)\|_1 + \|\mathbf{B}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)\|_\infty \\
&\leqslant p \|\mathbf{x}_t\|_1 \left\| \mathbf{e}_{t+1}^* \right\|_{p-1}^{p-1} + \left\| \mathbf{e}_{t+1}^* \right\|_p^{p-1} \|\mathbf{x}_t\|_1 \\
&+ (p-1) \|\mathbf{x}_t\|_1 \left\| \mathbf{e}_{t+1}^* \right\|_{p-1}^{p-1}
\end{aligned}$$

and by Assumption A5(i) the expected value of the latter is finite; hence, $E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)]$ and $Q(\boldsymbol{\tau})$ exist and are finite on $\mathcal{B}_q^n$.

Given that $\mathbf{S}$ (and hence $\mathbf{S}^{-1}$) is positive definite, then for any $\boldsymbol{\tau} \in \mathcal{B}_q^n$ we have $Q(\boldsymbol{\tau}) \geqslant 0$ with equality if and only if $E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)] = 0$. Now, the optimality condition derived in Proposition 2 implies that $E[\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t)] = \mathbf{0}$. Hence, $\boldsymbol{\tau}_0$ is a minimum of $Q(\boldsymbol{\tau})$ on $\mathcal{B}_q^n$. We now show that this minimum is moreover unique. For this, we need to show that $E[\mathbf{g}_p( \boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)] = \mathbf{0}$ has a unique solution $\tau = \tau_0$. Since

$$E[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)] = E[\mathbf{a}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)] + E[\mathbf{B}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)]\boldsymbol{\tau}$$

a solution is unique if and only if $\operatorname{rank} E[\mathbf{B}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)] = n$; the latter holds under Assumption A5(ii), so $\boldsymbol{\tau}_0$ is the unique minimum of $Q(\boldsymbol{\tau})$. $\square$

*Proof of Lemma 6.* Given $p$, $1 \leqslant p < \infty$, and for any $\boldsymbol{\tau} \in \mathcal{B}_q^n$, let

$$\mathbf{S}(\boldsymbol{\tau}) \equiv E\left[\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau}; \mathbf{e}_{t+1}^*, \mathbf{x}_t)'\right].$$

Then

$$\begin{aligned}
\mathbf{S}(\boldsymbol{\tau}) &= E\left[\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)' \otimes \mathbf{x}_t\mathbf{x}_t'\right] \\
&= E\left[E\left(\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)'|\mathcal{F}_t\right) \otimes \mathbf{x}_t\mathbf{x}_t'\right].
\end{aligned}$$

In order to show that $\mathbf{S}(\boldsymbol{\tau})$ is positive definite, it suffices to show that with probability one:

(i) $E\left(\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)'|\mathcal{F}_t\right)$ is positive definite, and

(ii) $\mathbf{x}_t\mathbf{x}_t'$ is positive definite.

The second property holds if $E(\mathbf{x}_t\mathbf{x}_t')$ is of full rank as assumed in A6. We now show that the first property holds as well. Given that the $n$-variate loss is convex and such that $L_p(\boldsymbol{\tau}, \mathbf{e}) \geqslant 0$ with $L_p(\boldsymbol{\tau}, \mathbf{e}) = 0$ only if $\mathbf{e} = 0$ (Proposition 1), we have that given $\mathcal{F}_t$,

$$\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}_{t+1}^*)' \geq 0 \text{ a.s.-}P$$

with equality only if $\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}^*_{t+1}) = 0$ a.s.-$P$, i.e., only if $\mathbf{e}^*_{t+1} = 0$ a.s.-$P$. Since by Assumption A1 $\mathbf{y}_t$ is continuously distributed, we have that $\Pr(\mathbf{e}^*_{t+1} = 0|\mathcal{F}_t) = 0$ so

$$\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}^*_{t+1}) \nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}, \mathbf{e}^*_{t+1})' > 0 \text{ a.s.-}P.$$

Then, the property (i) holds. Hence, $\mathbf{S}(\boldsymbol{\tau})$ is positive definite. □

## References

[1] Batchelor, R. and D. A. Peel, "Rationality Testing under Asymmetric Loss," *Economic Letters* 61:1 (1998), 49-54.

[2] Bracewell, R. N., *The Fourier Transform and Its Applications* (New York: MacGraw-Hill, 1999).

[3] Capistran, C. and A. Timmermann, "Disagreement and Biases in Inflation Expectations," *Journal of Money, Credit and Banking* 41:(2-3) (2009), 365-396.

[4] Christiano, L. J., M. Eichenbaum, and C. L. Evens, "Monetary Policy Shocks: What Have We Learned and To What End?" (pp. 65-148) in J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics* (Amsterdam: Elsevier, 1999).

[5] Clarida, R., J. Galì, and M. Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics* 115:1 (2000), 147-180.

[6] Chaudhuri, P., "On a Geometric Notation of Quantiles for Multivariate Data," *Journal of the American Statistical Association* 91: 434 (1996), 862-872.

[7] Christoffersen, P. F. and F. X. Diebold, "Optimal Prediction under Asymmetric Loss," *Econometric Theory* 13:6 (1997), 808-817.

[8] Clark, T. E. and M. W. McCracken, "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105:1 (2001), 85-110.

[9] Clark, T. E. and K. D. West, "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* 135: (1-2) (2006), 155-186.

[10] —, "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics* 138:1 (2007), 291-311.

[11] Corradi, V. and N. R. Swanson, "A consistent Test for Nonlinear Out of Sample Predictive Accuracy," *Journal of Econometrics* 110:2 (2002), 352-381.

[12] —, "Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes," *International Economic Review* 48:1 (2007), 67-109.

[13] —, "Predictive Density and Conditional Confidence Interval Accuracy Tests," *Journal of Econometrics* 135:(1-2) (2006), 187-228.

[14] Elliott, G., I. Komunjer, and A. Timmermann, "Estimation and Testing of Forecast Rationality under Flexible Loss," *The Review of Economic Studies* 72:4 (2005), 1107-1125.

[15] —, "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association* 6:1 (2008), 122-157.

[16] Elliott, G. and A. Timmerman, "Economic Forecasting," *Journal of Economic Literature* 46:1 (2008), 3-56.

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

[18] Hansen, L.P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50:4 (1982), 1029-1054.

[19] Hardy, G. H., J. E. Littlewood, and George Polya, *Inequalities* (United Kingdom: Cambridge University Press, 1952).

[20] Hubrich, K. and K. D. West, "Forecast Evaluation of Small Nested Model Sets," *Journal of Applied Econometrics* 25:4 (2010), 574-594.

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

[22] Keane, M. P. and D. E. Runkle, "Testing the Rationality of Price Forecasts: New Evidence from Panel Data," *American Economic Review* 80:4 (1990), 714-735.

[23] Kirchgässner, B. and U. K. Müller, "Are Forecasters Reluctant to Revise their Predications? Some German Evidence," *Journal of Forecasting* 25:6 (2006), 401-413.

[24] Komunjer, I., "Global Identification in Nonlinear Models under Moment Restrictions," University of California, San Diego working paper (2011).

[25] Lieli, R. P. and M. B. Stinchcombe, "On the Recoverability of Forecasters' Preferences," University of Texas, Austin working paper (2009).

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

[27] McCracken, M. W., "Robust Out-of-Sample Inference," *Journal of Econometrics* 99 (2000), 195-223.

[28] —, "Asymptotics for Out of Sample Tests of Granger Causality," *Journal of Econometrics* 140 (2007), 719-752.

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

[30] Mincer, J. and V. Zarnowitz "The Evaluation of Economic Forecasts," (pp.1-46), in J. Mincer (Ed.) *Economic Forecasts and Expectations* (New York: National Bureau of Economic Research, 1969).

[31] Mishkin, Frank S., "Are Market Forecasts Rational?" *American Economic Review* 71:3 (1981), 295-306.

[32] Muth, John F., "Rational Expectations and the Theory of Price Movements," *Econometrica* 29:3 (1961), 315-335.

[33] Patton, A. J. and A. Timmermann, "Testing Forecast Optimality under Unknown Loss," *Journal of the American Statistical Association* 102 (2007a), 1172-1184.

[34] —, "Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity," *Journal of Econometrics* 140:2 (2007b), 884-918.

[35] Pesaran, M. H. and Martin Weale, "Survey Expectations" (pp. 715-776), in G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.) *Handbook of Economic Forecasting* Vol. 1 (Amsterdam: North-Holland, 2006).

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

[37] Theil, H., *Economic Forecasts and Policy* (Amsterdam: North-Holland, 1958).

[38] Taylor, J. B., "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference series on Public Policy* 39:1 (1993): 195-214.

[39] West, K. D. and M. W. McCracken, "Regression-based Tests of Predictive Ability," *International Economic Review* 39 (1998), 817-840.

[40] West, K. D., "Asymptotic Inference About Predictive Ability," *Econometrica* 64:5 (1996), 1067-1084.

[41] —, "Forecast Evaluation" (pp. 99-134), in G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Amsterdam: Elsevier, North-Holland, 2006).

[42] Zarnowtiz, V., "Rational Expectations and Macroeconomic Forecasts," *Journal of Business and Economic Statistics* 3 (1985), 293-311.

[43] Zellner, A., "Biased Predictors, Rationality and the Evaluation of Forecasts," *Economic Letters* 21:1 (1986), 45-48.

## Notes

[1]Another way to test for rationality of forecasts would be to specify an alternative forecast formation model, such as that of "adaptive" expectations, for example. In that case, rejections of rationality would be in favor of the specific alternative.

[2]Those results are obtained under a specific parametric assumption for the forecasters' loss. Quantifying the degree of asymmetry of the forecasters' losses in a nonparametric framework is still an open question. Important nonparametric identification results have been obtained in Lieli and Stinchcombe (2009) who point out that "[the] cost of this generality is that [the] identification results are more abstract and do not directly translate into a strategy for estimation and inference."

[3]Strictly speaking, such generalizations will only apply to direct forecasting methods. In particular, forecasts that are constructed iteratively would require a different treatment.

[4]The parameterization in EKT is stated in terms of $\alpha = (\tau + 1)/2$ with $0 < \alpha < 1$, which is equivalent.

[5]The two middle plots in the left panel of Figure 1 illustrate this point. The plots are obtained when $\tau = (0, -0.5)'$ and $\tau = (0.4, -0.3)$, respectively. In both cases, $\|\tau\|_2 = 0.5$; however, the directions of the two asymmetry vectors are different, thus resulting in different bivariate losses.

[6]Similarly, we use $\text{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ to denote the sign function: $\text{sgn}(u) = \mathbb{1}(u) - \mathbb{1}(-u) = 2\mathbb{1}(u) - 1$.

[7]In theory, we could estimate the shape of the loss $L_p$ together with its asymmetry parameter $\tau$. However, due to slow convergence speeds, estimation of the shape parameter is rather unreliable at sample sizes considered in our empirical application.

[8]The analogous assumption in the decision theoretic literature would be that the decision maker is an expected utility maximizer. This property, in particular, eliminates objective functions of the form $\det E[\ e_{t+1}e_{t+1}']$.

[9]That $S(\tau_0)$ is positive definite is a condition typically used to show that $\tau_0$ is locally identified.

[10]The length of the in-sample used for estimation of the forecasting model is denoted by $R$, while $P$ stands for the length of the out-of-sample forecasting period. Thus, the length of the available sample equals $T + 1 = R + P$.

[11]If instead of a rolling window scheme, we assumed that the forecasts were constructed using a recursive scheme (i.e., with an expanding estimation sample ranging from 1 to $R$ for the first forecast, then from 1 to $R + 1$ for the second forecast, and so on), then $\{\hat{\mathbf{e}}_{t+1}\}$ would not necessarily inherit the strict stationarity and $\alpha$-mixing properties of $\{w_t\}$ and A8 would fail—see West (2006), for example. Similarly, an assumption of fixed forecasting scheme would make A7 untenable.

[12]One way to correct for the "many instrument problem" would be to use estimators that are robust to the presence of many instruments (such as the GEL estimator, for example). Under many instrument asymptotics, these estimators can be shown to have convenient Gaussian limit distributions, although the form of the covariance matrix would involve an extra adjustment term relative to that obtained under the case of conventional asymptotics, which we maintain here.

[13]These data are taken from the Federal Reserve Bank of St. Louis's 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.

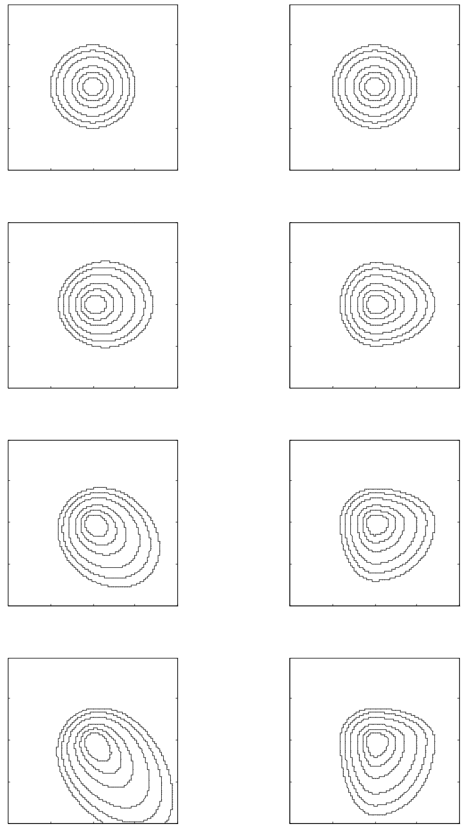

FIGURE 1. Contour plots of the bivariate loss $L_2(\tau, \cdot)$ with $\tau(\tau_1, \tau_2)'$ (left), and of the sum of univariate losses $L_2(\tau_1, \cdot)+$ $L_2(\tau_2, \cdot)$ (right), with $\tau = (0, 0)', (0, -0.5)', (0.4, -0.3)', (0.7, -0.5)'$ (top to bottom).

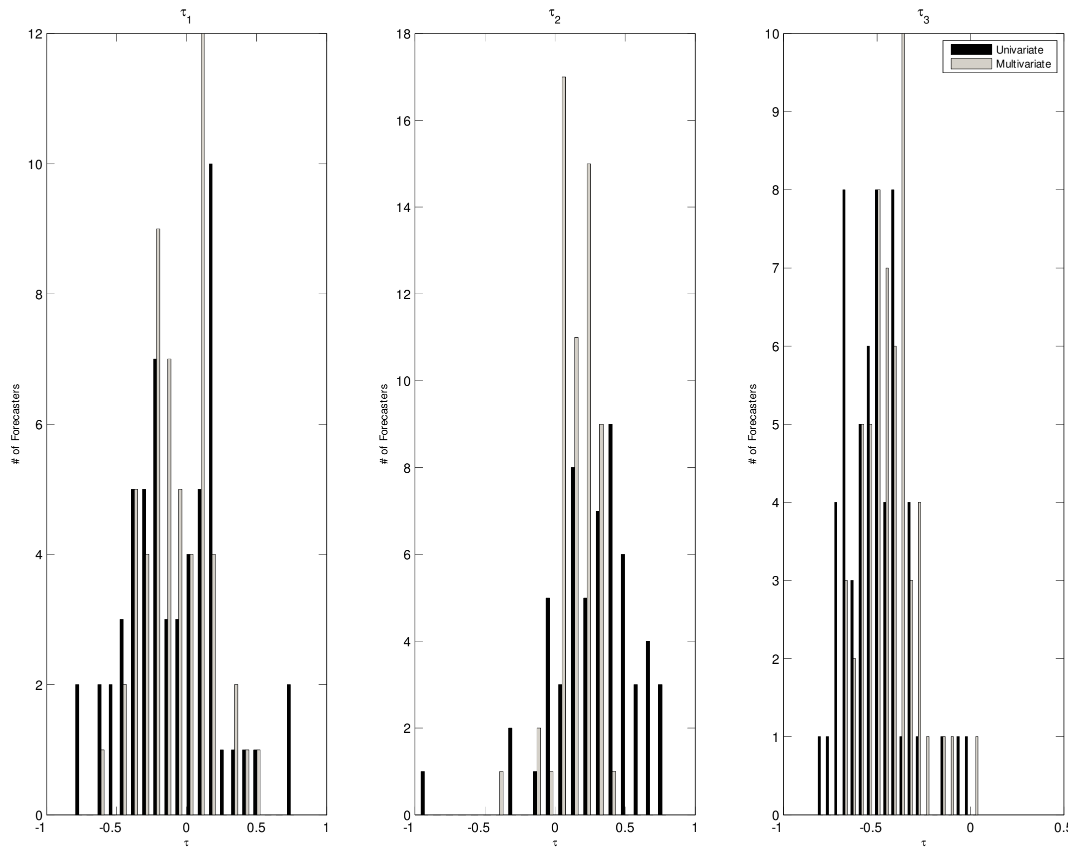


FIGURE 2. Distribution of the asymmetry parameter estimates obtained when $\mathbf{x}_t = (1, GDP_t)'$.

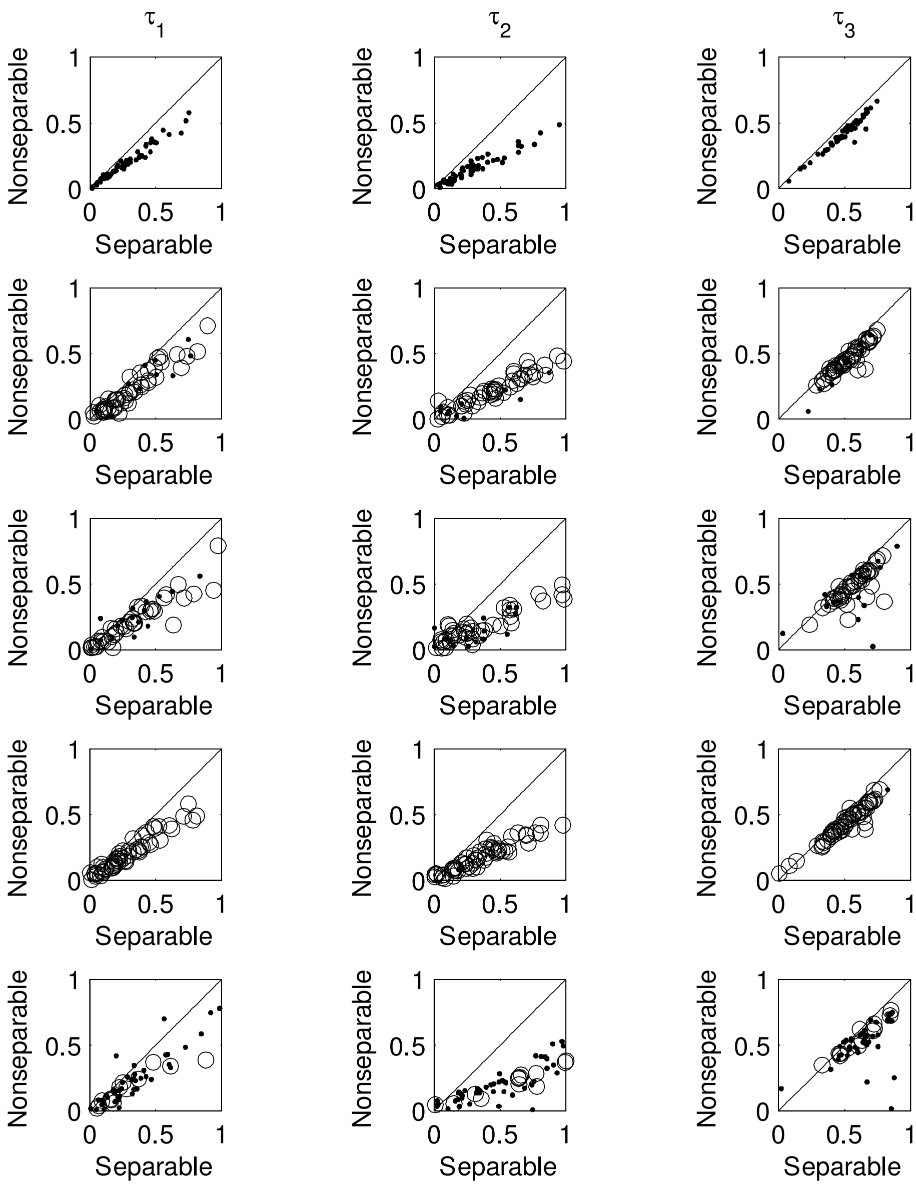


Figure 3. Absolute values of the asymmetry parameter estimates obtained under separable and nonseparable losses for each instrument set. Circles indicate values for which rationality was accepted at the 95% level using the nonseparable loss.

## Table 1: GMM Estimator and J-test of rationality (5% Nominal Size)

**(a) Instrument = (1)**

**$\tau_0 = (0,0,0)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ |
|---|---|---|---|---|---|---|
| **R=250,P=150** | -0.003 | 0.010 | 0.006 | 0.072 | 0.072 | 0.074 |
| **R=250,P=200** | -0.003 | 0.000 | 0.000 | 0.060 | 0.063 | 0.063 |
| **R=250,P=250** | 0.002 | 0.001 | -0.001 | 0.049 | 0.051 | 0.046 |
| **R=200,P=250** | -0.003 | -0.002 | -0.001 | 0.044 | 0.044 | 0.044 |
| **R=300,P=200** | 0.000 | 0.000 | 0.002 | 0.063 | 0.060 | 0.061 |

**$\tau_0 = (0,0,-0.5)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ |
|---|---|---|---|---|---|---|
| **R=250,P=150** | -0.005 | -0.007 | -0.496 | 0.070 | 0.064 | 0.060 |
| **R=250,P=200** | 0.000 | 0.000 | -0.492 | 0.054 | 0.054 | 0.048 |
| **R=250,P=250** | -0.002 | -0.002 | -0.494 | 0.046 | 0.046 | 0.039 |
| **R=200,P=250** | -0.001 | 0.002 | -0.492 | 0.039 | 0.042 | 0.035 |
| **R=300,P=200** | 0.003 | 0.001 | 0.493 | 0.056 | 0.055 | 0.048 |

**$\tau_0 = (0,0.4,-0.3)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ |
|---|---|---|---|---|---|---|
| **R=250,P=150** | 0.002 | 0.392 | -0.291 | 0.066 | 0.062 | 0.067 |
| **R=250,P=200** | 0.002 | 0.394 | -0.296 | 0.056 | 0.049 | 0.052 |
| **R=250,P=250** | 0.002 | 0.394 | -0.295 | 0.046 | 0.042 | 0.043 |
| **R=200,P=250** | -0.002 | 0.392 | -0.296 | 0.041 | 0.037 | 0.039 |
| **R=300,P=200** | -0.005 | 0.392 | -0.299 | 0.058 | 0.052 | 0.054 |

**$\tau_0 = (0,0.7,-0.5)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ |
|---|---|---|---|---|---|---|
| **R=250,P=150** | 0.001 | 0.690 | -0.495 | 0.051 | 0.034 | 0.042 |
| **R=250,P=200** | -0.001 | 0.696 | -0.493 | 0.040 | 0.027 | 0.035 |
| **R=250,P=250** | 0.000 | 0.695 | -0.494 | 0.034 | 0.024 | 0.029 |
| **R=200,P=250** | -0.001 | 0.693 | -0.496 | 0.030 | 0.021 | 0.026 |
| **R=300,P=200** | -0.002 | 0.694 | -0.499 | 0.042 | 0.028 | 0.035 |

**(b) Instrument = (1, $y_{1,t}$); 3 Overidentifying Restrictions**

**$\tau_0 = (0,0,0)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ | J-test |
|---|---|---|---|---|---|---|---|
| **R=250,P=150** | 0.001 | 0.011 | 0.004 | 0.077 | 0.075 | 0.078 | 0.04 |
| **R=250,P=200** | -0.003 | 0.000 | 0.000 | 0.063 | 0.064 | 0.065 | 0.03 |
| **R=250,P=250** | 0.002 | 0.002 | -0.001 | 0.050 | 0.053 | 0.049 | 0.01 |
| **R=200,P=250** | -0.003 | -0.003 | -0.001 | 0.046 | 0.046 | 0.045 | 0.00 |
| **R=300,P=200** | 0.000 | 0.000 | 0.002 | 0.066 | 0.063 | 0.064 | 0.03 |

**$\tau_0 = (0,0,-0.5)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ | J-test |
|---|---|---|---|---|---|---|---|
| **R=250,P=150** | -0.002 | -0.010 | -0.504 | 0.074 | 0.068 | 0.064 | 0.03 |
| **R=250,P=200** | 0.000 | 0.000 | -0.499 | 0.058 | 0.057 | 0.050 | 0.03 |
| **R=250,P=250** | -0.002 | -0.003 | -0.499 | 0.048 | 0.047 | 0.040 | 0.01 |
| **R=200,P=250** | -0.001 | 0.002 | -0.496 | 0.041 | 0.044 | 0.036 | 0.00 |
| **R=300,P=200** | 0.003 | 0.001 | 0.500 | 0.059 | 0.058 | 0.049 | 0.03 |

**$\tau_0 = (0,0.4,-0.3)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ | J-test |
|---|---|---|---|---|---|---|---|
| **R=250,P=150** | 0.002 | 0.400 | -0.297 | 0.069 | 0.064 | 0.069 | 0.02 |
| **R=250,P=200** | 0.002 | 0.399 | -0.300 | 0.058 | 0.050 | 0.054 | 0.02 |
| **R=250,P=250** | 0.002 | 0.397 | -0.297 | 0.047 | 0.043 | 0.045 | 0.01 |
| **R=200,P=250** | -0.002 | 0.396 | -0.298 | 0.043 | 0.038 | 0.040 | 0.01 |
| **R=300,P=200** | -0.005 | 0.398 | -0.304 | 0.061 | 0.054 | 0.056 | 0.04 |

**$\tau_0 = (0,0.7,-0.5)$**

| R, P | $\tau_1$ | $\tau_2$ | $\tau_3$ | SD $\tau_1$ | SD $\tau_2$ | SD $\tau_3$ | J-test |
|---|---|---|---|---|---|---|---|
| **R=250,P=150** | 0.000 | 0.695 | -0.500 | 0.054 | 0.034 | 0.042 | 0.03 |
| **R=250,P=200** | -0.001 | 0.700 | -0.496 | 0.042 | 0.028 | 0.036 | 0.02 |
| **R=250,P=250** | 0.000 | 0.698 | -0.497 | 0.035 | 0.025 | 0.030 | 0.01 |
| **R=200,P=250** | -0.001 | 0.695 | -0.497 | 0.031 | 0.022 | 0.027 | 0.00 |
| **R=300,P=200** | -0.001 | 0.700 | -0.501 | 0.044 | 0.029 | 0.036 | 0.02 |

NOTES: This table reports average GMM estimates for $\tau$ obtained across 1000 Monte Carlo simulations. Standard deviations are reported in columns marked SD. J-test represents the percentage of rejections of the null of rationality at the 5% level. $\tau_0$ indicates the true value of $\tau$. R is the size of the rolling window of data used to construct the forecasts, and P is the size of the evaluation sample.

The parameterization of the VAR(1) that generates the data is:

$$y_t = c + Ay_{t-1} + u_t,\ \text{where } u_t \sim N(0,\Sigma),\ c = 0.01*[1,1,1]',\ A = 0.5*\mathrm{I}_3,\ \text{and } \Sigma = 0.01*\mathrm{I}_3.$$

## Table 2: Monte Carlo Results

**$\tau_0$ = (0,-0.7,-0.6); R = 250; P = 150**

| Instr Set | Separable | | | Nonseparable | | | Relative MSE | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\tau_1$ | $\tau_2$ | $\tau_3$ | $\tau_1$ | $\tau_2$ | $\tau_3$ | $\tau_1$ | $\tau_2$ | $\tau_3$ |
| **1** | -0.139 | -0.943 | -0.910 | 0.003 | -0.695 | -0.596 | 15.09 | 109.79 | 132.98 |
| | (0.102) | (0.020) | (0.027) | (0.044) | (0.023) | (0.027) | | | |
| **2** | -0.159 | -0.952 | -0.917 | -0.004 | -0.699 | -0.601 | 17.14 | 108.61 | 134.23 |
| | (0.111) | (0.022) | (0.029) | (0.047) | (0.024) | (0.028) | | | |
| **3** | -0.181 | -0.959 | -0.930 | 0.010 | -0.702 | -0.606 | 18.49 | 106.49 | 128.50 |
| | (0.118) | (0.021) | (0.029) | (0.049) | (0.025) | (0.029) | | | |
| **4** | -0.203 | -0.966 | -0.939 | -0.017 | -0.706 | -0.610 | 20.06 | 105.25 | 122.48 |
| | (0.125) | (0.021) | (0.030) | (0.051) | (0.025) | (0.029) | | | |
| **5** | -0.159 | -0.952 | -0.920 | -0.038 | -0.699 | -0.600 | 16.88 | 106.24 | 131.97 |
| | (0.112) | (0.022) | (0.028) | (0.047) | (0.025) | (0.028) | | | |

NOTES: This table reports the average estimates for $\tau$ computed over 1000 Monte Carlo replications.

Relative MSE is the ratio of the mean squared error MSE = $\sigma_i^2$ + $[\text{Bias}(\tau_i)]^2$ from the separable to the nonseparable case.

The parameterization of the VAR(1) that generates the data is:

$$y_t = c + Ay_{t-1} + u_t, \text{ where } u_t \sim N(0,\Sigma),\ c = \frac{1}{100}\begin{bmatrix}1\\1\\1\end{bmatrix},\ A = \begin{bmatrix}0.5 & 0.2 & 0.14\\0.2 & 0.5 & 0.14\\0.14 & 0.14 & 0.5\end{bmatrix}, \text{ and } \Sigma = \frac{1}{100}\begin{bmatrix}1 & 0.9 & 0.17\\0.9 & 1 & 0.46\\0.17 & 0.46 & 1\end{bmatrix}.$$

The instrument sets are: (1) Constant only; (2) Constant and one lag of the first (symmetric loss) variable; (3) Constant and one lag of the second variable; (4) Constant and one lag of first two variables; and (5) Constant and one lag of all three variables. Standard deviations are reported in parentheses.

## Table 3: Instrument Sets

| Instr Set | Constant | GDP/GNP | CPI | Unemp | Short Rate |
|---|---|---|---|---|---|
| **1** | 1 | n/a | n/a | n/a | n/a |
| **2** | 1 | n/a | 1 | n/a | n/a |
| **3** | 1 | n/a | n/a | n/a | 1 |
| **4** | 1 | 1 | n/a | n/a | n/a |
| **5** | 1 | n/a | n/a | 1 | n/a |

NOTES: The table reflects the lags of variables used as instruments.
n/a indicates the variable is not included in the instrument set.
GDP/GNP, CPI, and unemployment are taken as rates. The short rate
is the 3 month T-bill rate.

## Table 4: J-tests of Rationality

| Instr Set | Rejections of Rationality | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Nonseparable Loss | | | Separable Loss | | | Separable Symmetric Loss | | |
| | **10%** | **5%** | **1%** | **10%** | **5%** | **1%** | **10%** | **5%** | **1%** |
| **2** | 0.30 | 0.18 | 0.00 | 0.30 | 0.18 | 0.04 | 1.00 | 0.99 | 0.97 |
| **3** | 0.47 | 0.28 | 0.07 | 0.56 | 0.33 | 0.07 | 1.00 | 1.00 | 1.00 |
| **4** | 0.14 | 0.02 | 0.00 | 0.14 | 0.04 | 0.00 | 1.00 | 0.99 | 0.94 |
| **5** | 0.86 | 0.77 | 0.32 | 0.91 | 0.70 | 0.21 | 1.00 | 1.00 | 0.99 |

NOTES: This table reports the percentages of forecasters for whom the null of rationality and nonseparable loss (left), rationality and separable loss (middle), and rationality and separable symmetric loss (right) could be rejected at the specified levels. The results are obtained by fixing $p=2$ . The instrument sets are as defined in Table 3.

## Table 5: Estimated Asymmetry Coefficients under Separable and Nonseparable Loss

**Nonseparable**

| Inst Set | Means | | | | Medians | | | |
|---|---|---|---|---|---|---|---|---|
| | $\tau_y$ | $\tau_\pi$ | $\tau_r$ | $\|\tau\|_2$ | $\tau_y$ | $\tau_\pi$ | $\tau_r$ | $\|\tau\|_2$ |
| **1** | -0.053 | 0.118 | -0.419 | 0.513 | -0.075 | 0.137 | -0.439 | 0.515 |
| **2** | -0.094 | 0.194 | -0.452 | 0.577 | -0.062 | 0.206 | -0.429 | 0.582 |
| **3** | -0.098 | 0.150 | -0.460 | 0.575 | -0.093 | 0.141 | -0.483 | 0.587 |
| **4** | -0.069 | 0.154 | -0.430 | 0.543 | -0.070 | 0.165 | -0.440 | 0.539 |
| **5** | 0.021 | 0.186 | -0.550 | 0.652 | 0.056 | 0.244 | -0.535 | 0.665 |

**Separable**

| Inst Set | Means | | | | Medians | | | |
|---|---|---|---|---|---|---|---|---|
| | $\tau_y$ | $\tau_\pi$ | $\tau_r$ | $\|\tau\|_2$ | $\tau_y$ | $\tau_\pi$ | $\tau_r$ | $\|\tau\|_2$ |
| **1** | -0.074 | 0.213 | -0.496 | 0.680 | -0.105 | 0.215 | -0.509 | 0.673 |
| **2** | -0.096 | 0.392 | -0.527 | 0.817 | -0.097 | 0.456 | -0.520 | 0.798 |
| **3** | -0.141 | 0.260 | -0.544 | 0.784 | -0.141 | 0.243 | -0.561 | 0.764 |
| **4** | -0.083 | 0.295 | -0.503 | 0.735 | -0.054 | 0.323 | -0.508 | 0.700 |
| **5** | 0.034 | 0.499 | -0.645 | 0.947 | 0.089 | 0.639 | -0.633 | 0.949 |

NOTES: This table reports summary statistics of GMM estimates of $\tau$ obtained across forecasters. GNP ($y$) and inflation ($\pi$) are taken as growth rates. The short rate ($r$) is the 3-month T-bill rate. The results are obtained by fixing p=2. The instrument sets are as defined in Table 3. For the first instrument set, means and medians are taken over the entire sample of valid forecasters. For all other instrument sets, they are computed only for those forecasters for whom the null of rationality could not be rejected at the 5% level.

# ONLINE APPENDIX: MULTIVARIATE FORECAST EVALUATION AND RATIONALITY TESTING

IVANA KOMUNJER AND MICHAEL T. OWYANG

This online appendix consists of three parts. Section 1 of the appendix contains additional information on the forecast data used in the empirical application (Section 6 of the paper). Section 2 of the appendix gives detailed proofs of Theorems 2 and 3 stated in the main text (Section 4 of the paper). Finally, the last part of the appendix contains additional panels of Table 1 discussed in the Monte Carlo experiment (Section 5.1 of the paper).

## 1. Information on the BCEI Forecast Data

In the empirical section of the paper, we use individual forecaster's forecasts from the Blue Chip Economic Indicators (BCEI). The data is proprietary. The survey reports monthly updates of forecasts from individual forecasters starting in 1976:08. 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.

We use forecasts for three variables: output, inflation, and a short-term interest rate. 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 (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. 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.

## 2. Supplemental Proofs

### 2.1. **Notation.** We first recall the notation.

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 scalar $u$, $u \in \mathbb{R}$, we let $\mathbb{I} : \mathbb{R} \to [0,1]$ be the indicator (or Heaviside) function, i.e., $\mathbb{I}(u) = 0$ if $u < 0$, $\mathbb{I}(u) = 1$ if $u > 0$, and $\mathbb{I}(0) = \frac{1}{2}$ (Bracewell, 2000). Similarly, we use $\mathrm{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ to denote the sign function: $\mathrm{sgn}(u) = \mathbb{I}(u) - \mathbb{I}(-u) = 2\mathbb{I}(u) - 1$, 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., $\mathbb{I}(u) = \int_a^u d\delta$, where $a$ is an arbitrary (possibly infinite) negative constant, $a \leqslant 0$.

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 $\|\mathbf{u}\|_\infty = \max_{1 \leqslant i \leqslant n}(|u_i|)$. Similarly, for any $m \times n$-matrix $\mathbf{A} = (a_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$, we let $\|\mathbf{A}\|_\infty = \max_{1 \leqslant i,j \leqslant n}(|a_{ij}|)$.

Hereafter, $\boldsymbol{\nu}_p(\mathbf{u})$, $\mathbf{V}_p(\mathbf{u})$ and $\mathbf{W}_p(\mathbf{u})$ are an $n$-vector and two $n \times n$-diagonal matrices defined as:

$$\begin{aligned}
\boldsymbol{\nu}_p(\mathbf{u}) &\equiv (\mathrm{sgn}(u_1)|u_1|^{p-1}, \ldots, \mathrm{sgn}(u_n)|u_n|^{p-1})' \\
\mathbf{V}_p(\mathbf{u}) &\equiv \mathrm{diag}(\delta(u_1)|u_1|^{p-1}, \ldots, \delta(u_n)|u_n|^{p-1}) \\
\mathbf{W}_p(\mathbf{u}) &\equiv \mathrm{diag}(|u_1|^{p-2}, \ldots, |u_n|^{p-2}),
\end{aligned}$$

respectively. Then, we have that:

$$\nabla_{\mathbf{u}} \|\mathbf{u}\|_p = \|\mathbf{u}\|_p^{1-p} \boldsymbol{\nu}_p(\mathbf{u})$$

and

$$\Delta_{\mathbf{uu}} \|\mathbf{u}\|_p = \|\mathbf{u}\|_p^{1-p} \left\{ 2\mathbf{V}_p(\mathbf{u}) + (p-1) \left[ \mathbf{W}_p(\mathbf{u}) - \|\mathbf{u}\|_p^{-p} \boldsymbol{\nu}_p(\mathbf{u}) \boldsymbol{\nu}_p'(\mathbf{u}) \right] \right\},$$

which we shall often be using in what follows.

2.2. **Proofs of Theorems 2 and 3.**

**Theorem 2.** *Let Assumptions A1 through A8 hold. Then, given $p$, $1 \leqslant p < \infty$, we have $\hat{\boldsymbol{\tau}}_P \xrightarrow{p} \boldsymbol{\tau}_0$ as $(R, P) \to \infty$.*

*Proof of Theorem 2.* The minimizer $\boldsymbol{\tau}_0$ of $Q(\boldsymbol{\tau})$ can be written as:

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

On the other hand, from Equation (4) we have $\hat{\boldsymbol{\tau}}_P \equiv -[\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_P]^{-1}\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}\hat{\mathbf{a}}_P$ with the $nd \times 1$ vector

$$\hat{\mathbf{a}}_P \equiv P^{-1}\sum_{t=R}^{T} p(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t) \tag{S-1}$$

and the $nd \times n$ matrix

$$\hat{\mathbf{B}}_P \equiv P^{-1}\sum_{t=R}^{T} \|\hat{\mathbf{e}}_{t+1}\|_p^{p-1}(\mathbf{I}_n \otimes \mathbf{x}_t) + (p-1)\|\hat{\mathbf{e}}_{t+1}\|_p^{-1}(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\hat{\mathbf{e}}'_{t+1}. \tag{S-2}$$

To show $\hat{\boldsymbol{\tau}}_P \xrightarrow{p} \boldsymbol{\tau}_0$, it is sufficient to show that *(i)* $\hat{\mathbf{a}}_P - E[\mathbf{a}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)] \xrightarrow{p} \mathbf{0}$ and *(ii)* $\hat{\mathbf{B}}_P - E[\mathbf{B}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)] \xrightarrow{p} \mathbf{0}$. Then, by using Lemma 5, 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 6, and the continuity of the inverse function (away from zero), we have that $\hat{\boldsymbol{\tau}}_P \xrightarrow{p} \boldsymbol{\tau}_0$. By the triangle inequality we have $\left\|\hat{\mathbf{a}}_P - E[\mathbf{a}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_1 \leqslant \|\hat{\mathbf{a}}_P - E[\mathbf{a}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\|_1 + \left\|E[\mathbf{a}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{a}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_1$ and $\left\|\hat{\mathbf{B}}_P - E[\mathbf{B}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_\infty \leqslant \left\|\hat{\mathbf{B}}_P - E[\mathbf{B}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\right\|_\infty + \left\|E[\mathbf{B}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{B}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_\infty$. We first show that as $P \to \infty$, $\|\hat{\mathbf{a}}_P - E[\mathbf{a}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\|_1 \xrightarrow{p} \mathbf{0}$ and $\left\|\hat{\mathbf{B}}_P - E[\mathbf{B}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\right\|_\infty \xrightarrow{p} \mathbf{0}$ by using a law of large numbers (LLN) for $\alpha$-mixing sequences [e.g., Corollary 3.48 in White (2001)]. From Theorems 3.35 and 3.49 (White, 2001) measurable functions of strictly stationary and mixing processes are strictly stationary and mixing of the same size. Hence, by A8 we have $\{p(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\}$ and $\{\|\hat{\mathbf{e}}_{t+1}\|_p^{p-1}(\mathbf{I}_n \otimes \mathbf{x}_t) + (p-1)\|\hat{\mathbf{e}}_{t+1}\|_p^{-1}(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\hat{\mathbf{e}}'_{t+1}\}$ strictly stationary and $\alpha$-mixing of size $-r/(r-2)$ with $r > 2$. Now let $\delta = \varepsilon/2 > 0$; we have

$$
\begin{aligned}
E[\|(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\|_1^{r+\delta}] &\leqslant nE[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p-1)(r+\delta)} \|\mathbf{x}_t\|_1^{r+\delta}] \\
&\leqslant n \left\{ E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p-1)(2r+2\delta)}] E[\|\mathbf{x}_t\|_1^{2r+2\delta}] \right\}^{1/2} \\
&\leqslant n \left\{ \Delta_1 \Delta_2 \right\}^{1/2} < \infty, \qquad \text{(S-3)}
\end{aligned}
$$

where the second inequality follows by Cauchy-Schwartz inequality and the third uses assumption A8. Hence, $\hat{\mathbf{a}}_P$ in Equation (S-1) satisfies the LLN and $\|\hat{\mathbf{a}}_P - E[\mathbf{a}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\|_1 \xrightarrow{p} \mathbf{0}$ as $P \to \infty$. Similarly, we have

$$
\begin{aligned}
E[\|\hat{\mathbf{e}}_{t+1}\|_p^{(p-1)(r+\delta)} \|(\mathbf{I}_n \otimes \mathbf{x}_t)\|_\infty^{r+\delta}] &\leqslant E[\|\hat{\mathbf{e}}_{t+1}\|_p^{(p-1)(r+\delta)} \|\mathbf{x}_t\|_1^{r+\delta}] \\
&\leqslant c^{(p-1)(r+\delta)} \left\{ E[\|\hat{\mathbf{e}}_{t+1}\|_1^{(p-1)(2r+2\delta)}] E[\|\mathbf{x}_t\|_1^{2r+2\delta}] \right\}^{1/2} \\
&\leqslant c \left\{ \Delta_1 \Delta_2 \right\}^{1/2} < \infty, \qquad \text{(S-4)}
\end{aligned}
$$

where the second inequality uses the norm equivalence and Cauchy-Schwartz inequality, and the third inequality uses Assumption A8. In addition,

$$
\begin{aligned}
E[\|\hat{\mathbf{e}}_{t+1}\|_p^{-(r+\delta)} \left\|(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\hat{\mathbf{e}}'_{t+1}\right\|_\infty^{r+\delta}] &\leqslant E[\|\hat{\mathbf{e}}_{t+1}\|_p^{-(r+\delta)} \|(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\|_1^{r+\delta} \|\hat{\mathbf{e}}_{t+1}\|_1^{(r+\delta)}] \\
&\leqslant (1/d)^{r+\delta} E[\|(\boldsymbol{\nu}_p(\hat{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t)\|_1^{r+\delta}] < \infty, \quad \text{(S-5)}
\end{aligned}
$$

where the second inequality uses again the norm equivalence and the third follows from Equation (S-3). Combining Equations (S-4) – (S-5) with triangular inequality and the fact that, for any $(a, b) \in \mathbb{R}$, there exists some $n_{r+\delta} > 0$ such that $|a + b|^{r+\delta} \leqslant n_{r+\delta}[|a|^{r+\delta} + |b|^{r+\delta}]$, shows that $\hat{\mathbf{B}}_P$ in Equation (S-2) satisfies the LLN and so $\left\|\hat{\mathbf{B}}_P - E[\mathbf{B}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)]\right\|_\infty \xrightarrow{p} \mathbf{0}$ as $P \to \infty$. Next we need to show that $\left\|E[\mathbf{a}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{a}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_1 \to 0$ and $\left\|E[\mathbf{B}(p, \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t)] - E[\mathbf{B}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]\right\|_\infty \to 0$ as $P \to \infty$. We have

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

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

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

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

**Theorem 3.** *Let Assumptions A1-A3, A4', A5-A6, A7', A8-A10 hold. Then, given $p$, $1\leqslant p<\infty$, we have: $\sqrt{P}(\hat{\boldsymbol{\tau}}_P-\boldsymbol{\tau}_0)\xrightarrow{d}\mathcal{N}(\mathbf{0},(\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1})$, as $R,P\to\infty$, where $\mathbf{S}=E[\mathbf{g}_p(\boldsymbol{\tau}_0;\mathbf{e}^*_{t+1},\mathbf{x}_t)\mathbf{g}_p(\boldsymbol{\tau}_0;\mathbf{e}^*_{t+1},\mathbf{x}_t)']$ and $\mathbf{B}^*\equiv E[\left\|\mathbf{e}^*_{t+1}\right\|_p^{p-1}(\mathbf{I}_n\otimes\mathbf{x}_t)+(p-1)\left\|\mathbf{e}^*_{t+1}\right\|_p^{-1}(\boldsymbol{\nu}_p(\mathbf{e}^*_{t+1})\otimes\mathbf{x}_t)\mathbf{e}^{*\prime}_{t+1}]$.*

*Proof of Theorem 3.* To simplify the notation in this proof, let it be understood that $\sum_t$ denotes $\sum_{t=R}^T$ while $\sup_t$ stands for $\sup_{R\leqslant t\leqslant T}$. In order to show that $P^{1/2}(\hat{\boldsymbol{\tau}}_P-\boldsymbol{\tau}_0)$ is asymptotically normal, note that

$$
\begin{aligned}
\sqrt{P}(\hat{\boldsymbol{\tau}}_P-\boldsymbol{\tau}_0) &= -[\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_P]^{-1}\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}[\sqrt{P}(\hat{\mathbf{a}}_P+\hat{\mathbf{B}}_P\boldsymbol{\tau}_0)] \qquad \text{(S-6)}\\
&= -[\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_P]^{-1}\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}[\sqrt{P}\hat{\mathbf{m}}^*_P+\sqrt{P}\hat{\mathbf{m}}+\sqrt{P}(\hat{\mathbf{m}}_P-\hat{\mathbf{m}}-\hat{\mathbf{m}}^*_P)],
\end{aligned}
$$

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

$$
\hat{\mathbf{m}}_P\equiv P^{-1}\sum_t\mathbf{g}_p(\boldsymbol{\tau}_0;\hat{\mathbf{e}}_{t+1},\mathbf{x}_t)=\hat{\mathbf{a}}_P+\hat{\mathbf{B}}_P\boldsymbol{\tau}_0,\ \text{ and } \hat{\mathbf{m}}^*_P\equiv P^{-1}\sum_t\mathbf{g}_p(\boldsymbol{\tau}_0;\mathbf{e}^*_{t+1},\mathbf{x}_t). \qquad \text{(S-7)}
$$

The idea then is to show that the terms $\sqrt{P}\hat{\mathbf{m}}$ and $\sqrt{P}(\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}_P^*)$ on the right-hand side of Equation (S-6) are $o_p(\mathbf{1})$. We start by showing that the first term is $o(\mathbf{1})$. Let $\mathbf{m}^* \equiv E[\mathbf{a}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)] + E[\mathbf{B}(p, \mathbf{e}_{t+1}^*, \mathbf{x}_t)]\boldsymbol{\tau}_0$. First, we show that $\nabla_{\mathbf{e}} E[\mathbf{g}_p(\boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)] = E[\nabla_{\mathbf{e}} \mathbf{g}_p(\boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)]$ 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 $\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}_0, \cdot)$ in Equation (9) we get that for any $\mathbf{e} \in \mathbb{R}^n \backslash \mathcal{E}$ ($\mathcal{E} = \{\mathbf{e} \in \mathbb{R}^n : e_i = 0 \text{ for some } i\}$),

$$\begin{aligned} \Delta_{\mathbf{ee}} L_p(\boldsymbol{\tau}_0, \mathbf{e}) &= 2p\mathbf{V}_p(\mathbf{e}) + p(p-1)\mathbf{W}_p(\mathbf{e}) \\ &\quad + (p-1)\left[2\frac{\boldsymbol{\tau}_0 \boldsymbol{\nu}_p'(\mathbf{e})}{\|\mathbf{e}\|_p} + \frac{\boldsymbol{\tau}_0'\mathbf{e}}{\|\mathbf{e}\|_p}\left((p-1)\mathbf{W}_p(\mathbf{e}) - \frac{\boldsymbol{\nu}_p(\mathbf{e})\boldsymbol{\nu}_p'(\mathbf{e})}{\|\mathbf{e}\|_p^p}\right)\right], \end{aligned} \tag{S-8}$$

where we have used the fact that for any $1 \leqslant p < \infty$, $\dfrac{\boldsymbol{\tau}_0'\mathbf{e}}{\|\mathbf{e}\|_p}\mathbf{V}_p(\mathbf{e}) = \mathbf{0}$ for all $\mathbf{e} \in \mathbb{R}^n \backslash \mathcal{E}$. Note that in the univariate case $n = 1$, the Hessian in Equation (S-8) reduces to $\Delta_{ee} L_p(\tau_0, e) = 2\{p\delta(e)|e|^{p-1} + p(p-1)[1 + \tau_0 \operatorname{sgn}(e)]|e|^{p-2}\}$ [see Equation (9) in EKT, p.1121]. Hence

$$\begin{aligned} \|\Delta_{\mathbf{ee}} L_p(\boldsymbol{\tau}_0, \bar{\mathbf{e}}_{t+1})\|_\infty &\leqslant 2p\|\mathbf{V}_p(\bar{\mathbf{e}}_{t+1})\| + p(p-1)c_3\|\bar{\mathbf{e}}_{t+1}\|_1^{p-2} \\ &\quad + (p-1)\left[2d_3\|\bar{\mathbf{e}}_{t+1}\|_1^{p-2} + (p-1)c_3\|\bar{\mathbf{e}}_{t+1}\|_1^{p-2} + c_3\|\bar{\mathbf{e}}_{t+1}\|_1^{p-2}\right] \\ &= 2p\|\mathbf{V}_p(\bar{\mathbf{e}}_{t+1})\| + 2(p-1)(pc_3 + d_3)\|\bar{\mathbf{e}}_{t+1}\|_1^{p-2}, \end{aligned} \tag{S-9}$$

where we have used the norm equivalences: $c_1\|\bar{\mathbf{e}}_{t+1}\|_1 \leqslant \|\bar{\mathbf{e}}_{t+1}\|_{p-2} \leqslant c_2\|\bar{\mathbf{e}}_{t+1}\|_1$ for some $(c_1, c_2) > 0$ and $c_3 = c_2^{p-1}$ if $p \geq 2$ and $c_3 = c_1^{2-p}$ otherwise and, similarly, $d_1\|\bar{\mathbf{e}}_{t+1}\|_1 \leqslant \|\bar{\mathbf{e}}_{t+1}\|_p \leqslant d_2\|\bar{\mathbf{e}}_{t+1}\|_1$ for some $(d_1, d_2) > 0$ and $d_3 = d_2^{p-1}$ if $p \geq 2$ and $d_3 = d_1^{2-p}$ otherwise. Under A9, we have that $E[\sup_{c\in(0,1)}\left\|c\hat{\mathbf{e}}_{t+1} + (1-c)\mathbf{e}_{t+1}^*\right\|_1^{p-2}] < \infty$. Moreover, under A10, when $p = 1$ we have $E[\|\mathbf{V}_1(\bar{\mathbf{e}}_{t+1})\|_\infty] \leqslant M$ and when $p > 1$ we have $E[\|\mathbf{V}_1(\bar{\mathbf{e}}_{t+1})\|_\infty] = 0$, so the right-hand side of Equation (S-9) 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 $\nabla_{\mathbf{e}} E[\mathbf{g}_p(\boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)] = \nabla_{\mathbf{e}} E[\nabla_{\mathbf{e}} L_p(\boldsymbol{\tau}_0, \bar{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t] = E[\Delta_{\mathbf{ee}} L_p(\boldsymbol{\tau}_0, \bar{\mathbf{e}}_{t+1}) \otimes \mathbf{x}_t] = E[\nabla_{\mathbf{e}} \mathbf{g}_p(\boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)]$.

Second, we can use a mean value expansion around $\mathbf{e}_{t+1}^*$ that yields $\mathbf{0} = \sqrt{P}\mathbf{m}^* = \sqrt{P}\hat{\mathbf{m}} - E[P^{-1}\sum_t \nabla_{\mathbf{e}} \mathbf{g}_p(\boldsymbol{\tau}_0; \bar{\mathbf{e}}_{t+1}, \mathbf{x}_t)'\sqrt{P}(\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*)]$, where for every $t$, $R \leqslant t \leqslant T$, 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

$P^{-1/2}\sum_t \nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1}) \xrightarrow{p} \mathbf{0}$ as $R\to\infty$ and $P\to\infty$. Consider $\varepsilon$ from A7$'$(i) and note that we have

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

so for any $\eta>0$ and $\delta>0$

$$\begin{aligned}
&\lim_{R,P\to\infty}\Pr\left(\left\|P^{-1/2}\textstyle\sum_t \nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1>\eta\right)\\
&\leqslant \lim_{R,P\to\infty}\Pr\left(\left\|P^{-1/2}\textstyle\sum_t \nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1>\eta,\sup_t\left\|R^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1\leqslant\delta\right)\\
&+\lim_{R,P\to\infty}\Pr\left(\sup_t\left\|R^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1>\delta\right)\\
&\leqslant \lim_{R,P\to\infty}\Pr\left(\left\|P^{-1/2}\textstyle\sum_t \nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)'(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1>\eta,\sup_t\left\|R^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1\leqslant\delta\right)\\
&\leqslant \lim_{R,P\to\infty}\Pr\left(P^{-1/2}\textstyle\sum_t\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty R^{-1/2+\varepsilon}>\dfrac{\eta}{\delta},\sup_t\left\|R^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1\leqslant\delta\right).
\end{aligned}$$

where the first inequality uses A7$'$(ii). Now, using Markov's inequality we have

$$\begin{aligned}
&\Pr\left(P^{-1/2}\textstyle\sum_t\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty R^{-1/2+\varepsilon}>\dfrac{\eta}{\delta},\sup_t\left\|R^{1/2-\varepsilon}(\hat{\mathbf{e}}_{t+1}-\mathbf{e}^*_{t+1})\right\|_1\leqslant\delta\right)\\
&\leqslant\ \frac{\delta}{\eta}E\left(P^{-1/2}\textstyle\sum_t\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty R^{-1/2+\varepsilon}\right).
\end{aligned}$$

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

$$\begin{aligned}
&E\left(\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty\right)\\
&\leqslant E\left(\sup_{c\in(0,1)}\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1},\mathbf{x}_t)\right\|_\infty\right)\\
&\leqslant 2(p-1)\left(pc_3+d_3\right)E\left(\left\|\mathbf{x}_t\right\|_1\sup_{c\in(0,1)}\left\|c\hat{\mathbf{e}}_{t+1}+(1-c)\mathbf{e}^*_{t+1}\right\|_1^{p-2}\right)<\infty.
\end{aligned}$$

Now

$$\begin{aligned}
E\left(P^{-1/2}\textstyle\sum_t\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty R^{-1/2+\varepsilon}\right)&\leqslant E\left(\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty\right)P^{-1/2}\textstyle\sum_t R^{-1/2+\varepsilon}\\
&\leqslant E\left(\left\|\nabla_{\mathbf{e}}\mathbf{g}_p(\boldsymbol{\tau}_0;\bar{\mathbf{e}}_{t+1},\mathbf{x}_t)\right\|_\infty\right)\left(\frac{P}{R^{1-2\varepsilon}}\right)^{1/2}\to 0
\end{aligned}$$

as $R \to \infty$ and $P \to \infty$, where the last limit results uses A7$'$(i). Hence $\sqrt{P}\hat{\mathbf{m}} \to \mathbf{0}$ as $R \to \infty$ and $P \to \infty$. The term $\sqrt{P}(\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}_P^*)$ on the right-hand side of Equation (S-6) is $o_p(\mathbf{1})$ provided that $\mathbf{g}$ satisfies a certain Lipshitz condition (given below) and that A7$'$ holds. Using a reasoning similar to that above, we have any $\eta > 0$ and $\delta > 0$

$$\begin{aligned}
&\lim_{R,P\to\infty} \Pr\left(P^{1/2}\left\|\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}_P^*\right\|_1 > \eta\right) \\
&\leqslant \lim_{R,P\to\infty} \Pr\left(P^{1/2}\left\|\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}_P^*\right\|_1 > \eta, \sup_t R^{1/2-\varepsilon}\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 \leqslant \delta\right).
\end{aligned}$$

Now, let $r_P(\delta) \equiv \sup\{r_{t+1}(\hat{\mathbf{e}}_{t+1}) : \left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 \leqslant \delta, R \leqslant t \leqslant T\}$, where we let

$$\begin{aligned}
&r_{t+1}(\hat{\mathbf{e}}_{t+1}) \\
&\equiv \frac{\left\|\mathbf{g}_p(\boldsymbol{\tau}_0; \hat{\mathbf{e}}_{t+1}, \mathbf{x}_t) - \mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}_{t+1}^*, \mathbf{x}_t) - [\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t](\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*)\right\|_1}{\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1},
\end{aligned} \tag{S-10}$$

where $\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right)$ is as defined in Equation (S-8). Then, by the definition of $r_{t+1}(\hat{\mathbf{e}}_{t+1})$,

$$\begin{aligned}
&P^{1/2}\left\|\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}_P^*\right\|_1 \\
&\leqslant P^{1/2}\Big\{\Big\|P^{-1}\textstyle\sum_t[\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t](\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*) - E\{[\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t](\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*)\}\Big\|_1 \\
&\qquad + P^{-1}\textstyle\sum_t r_{t+1}(\hat{\mathbf{e}}_{t+1})\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 + E\left(r_{t+1}(\hat{\mathbf{e}}_{t+1})\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1\right)\Big\} \\
&\leqslant P^{1/2}\Big\{P^{-1}\textstyle\sum_t\left\|[\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t] - E\{[\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t]\}\right\|_\infty \sup_t\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 \\
&\qquad + [r_P(\delta_R) + E(r_P(\delta_R))]\sup_t\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1\Big\}
\end{aligned}$$

and so by the Markov inequality

$$\begin{aligned}
&\Pr\left(P^{1/2}\left\|\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}_P^*\right\|_1 > \eta, \sup_t R^{1/2-\varepsilon}\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}_{t+1}^*\right\|_1 \leqslant \delta\right) \\
&\leqslant \left(\frac{P}{R^{1-2\varepsilon}}\right)^{1/2}\frac{\delta}{\eta}\Big[E\Big(P^{-1}\sum_t\left\|[\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t] - E\{[\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}_{t+1}^*\right) \otimes \mathbf{x}_t]\}\right\|_\infty\Big) \\
&\qquad + E\Big(r_P(\delta_R) + E(r_P(\delta_R))\Big)\Big].
\end{aligned}$$

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 $r_P(\delta) \to 0$ with probability 1, which by the dominated convergence theorem ensures

$E(r_P(\delta)) \to 0$. Next, we show that the sample mean of $\{\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t\}$ converges in probability to its expected value. By assumption A4′ we know that $\{\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right) \otimes \mathbf{x}_t\}$ is strictly stationary and $\alpha$-mixing with $\alpha$ 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

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

since from Equation (S-9) we know

$$
\begin{aligned}
& \left\|\Delta_{\mathbf{ee}} L_p\left(\boldsymbol{\tau}_0, \mathbf{e}^*_{t+1}\right)\right\|_\infty^{2r+\epsilon} \\
& \leqslant n_r\{[2p]^{2r+\epsilon} \left\|\mathbf{V}_p(\mathbf{e}^*_{t+1})\right\|_\infty^{2r+\epsilon} + [2(p-1)\left(pc_3 + d_3\right)]^{2r+\epsilon} \left\|\mathbf{e}^*_{t+1}\right\|_1^{(p-2)(2r+\epsilon)}\},
\end{aligned}
$$

where again $n_r$ is such that for any $(a, b) > 0$ we have $(a+b)^{2r+\epsilon} \leqslant n_r(a^{2r+\epsilon} + b^{2r+\epsilon})$; and A10 and A4′ imply that $E[\left\|\mathbf{V}_1(\mathbf{e}^*_{t+1})\right\|_\infty^{2r+\epsilon}] \leqslant M$, $E[\left\|\mathbf{V}_p(\mathbf{e}^*_{t+1})\right\|_\infty^{2r+\epsilon}] = 0$ and $E[\left\|\mathbf{e}^*_{t+1}\right\|_1^{(p-2)(2r+\epsilon)}] < \infty$. Using the weak LLN for $\alpha$-mixing sequences [e.g., Corollary 3.48 in White (2001)] then gives

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

as $P \to \infty$. Then, combining all of the above with A7′(ii) gives

$$
\lim_{R,P\to\infty} \Pr\left(\sqrt{P}\left\|\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}^*_P\right\|_1 > \eta, \sup_{R\leqslant t\leqslant T} R^{1/2-\varepsilon}\left\|\hat{\mathbf{e}}_{t+1} - \mathbf{e}^*_{t+1}\right\|_1 \leqslant \delta\right) = 0
$$

and the term $\sqrt{P}(\hat{\mathbf{m}}_P - \hat{\mathbf{m}} - \hat{\mathbf{m}}^*_P)$ on the right-hand side of Equation (S-6) is $o_p(\mathbf{1})$ as $R \to \infty$ and $P \to \infty$. Finally, we use the central limit theorem (CLT) for strictly stationary and $\alpha$-mixing sequences [e.g., Theorem 5.20 in White (2001)] to show that $\sqrt{P}\hat{\mathbf{m}}^*_P \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S})$. 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 $\{\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\}$ is strictly stationary and $\alpha$-mixing with mixing coefficient of size $-r/(r-2)$, $r > 2$. In the proof of Theorem 2 we have moreover shown that $E[\left\|\mathbf{g}_p(\boldsymbol{\tau}_0; \mathbf{e}^*_{t+1}, \mathbf{x}_t)\right\|_1^{r+\varepsilon}] < \infty$. The CLT [e.g., Theorem 5.20 in White (2001)] then ensures

$$\sqrt{P}\hat{\mathbf{m}}^*_P \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}). \tag{S-11}$$

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 $\hat{\mathbf{B}}_P \xrightarrow{p} \mathbf{B}^* \equiv E[\mathbf{B}(p, \mathbf{e}^*_{t+1}, \mathbf{x}_t)]$ as $R \to \infty$ and $P \to \infty$ ($\mathbf{B}$ was defined in Equation (12)) together with A5(ii) ensure that $(\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1}$ exists, so by using $\sqrt{P}(\hat{\boldsymbol{\tau}}_P - \boldsymbol{\tau}_0) = -[\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}\hat{\mathbf{B}}_P]^{-1}\hat{\mathbf{B}}'_P\hat{\mathbf{S}}^{-1}[\sqrt{P}\hat{\mathbf{m}}^*_P + o_p(\mathbf{1})]$, the limit result in (S-11) and the Slutsky theorem we have $\sqrt{P}(\hat{\boldsymbol{\tau}}_P - \boldsymbol{\tau}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\mathbf{B}^{*\prime}\mathbf{S}^{-1}\mathbf{B}^*)^{-1})$, which completes the proof of asymptotic normality. $\square$

## References


[1] Andrews, Donald W. K., "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinutiy," Econometrica 71:1 (1994), 43-72.

[2] Bracewell, Robert N., *The Fourier Transform and Its Applications* (New York: MacGraw-Hill, 1999).

[3] White, Halbert, Asymptotic Theory for Econometricians (San Diego: Academic Press, 2001).