Strong Orthogonal Decompositions and Non-Linear Impulse Response Functions for Infinite Variance Processes*

[Original version : November 2003]

Jonathan B. Hill†
Dept. of Economics
Florida International University

July 27, 2006

Abstract

In this paper we prove Wold-type decompositions with strong-orthogonal prediction innovations exist in smooth, reflexive Banach spaces of discrete time processes *if and only if* the projection operator generating the innovations satisfies the property of iterations. Our theory includes as special cases all previous Wold-type decompositions of discrete time processes; completely characterizes when nonlinear heavy-tailed processes obtain a strong-orthogonal moving average representation; and easily promotes a theory of nonlinear impulse response functions for infinite variance processes. We exemplify our theory by developing a nonlinear impulse response function for smooth transition threshold processes, we discuss how to test decomposition innovations for strong orthogonality and whether the proposed model represents the best predictor, and we apply the methodology to currency exchange rates.

1. Introduction We develop a complete theory of Wold-type orthogonal decompositions in smooth, reflexive Banach spaces of discrete time processes.

* *AMS* 2000 *subject classifications*. Primary 60G25; secondary 47N30.
Key words and phrases. Orthogonal decompositions; Banach spaces; projection iterations; infinite variance; moving average; nonlinear impulse response function; smooth transition autoregression; Lp-metric projection; Lp-GMM.
I would like to thank an anonymous referee, an Associate Editor and the Editor for comments that led to substantial improvements. All errors, of course, are mine alone.
This paper is dedicated to my father, Walter Howard Hill, 1922-1996.
† Dept. of Economics, Florida International University; *hilljona@fiu.edu*; www.fiu.edu/~hilljona.

Such spaces include the Hilbert spaces, and contain linear and nonlinear heavy-tailed processes[1]. An immediate application is the construction of nonlinear impulse response functions (IRF's) for possibly non-stationary and/or long memory heavy-tailed processes based on moving average representations.

In particular, we provide necessary and sufficient conditions for the existence of orthogonal decompositions in the time-domain with asymmetric strong orthogonal innovations. For some Banach space stochastic process $\{X_t\} = \{X_t : -\infty < t < \infty$, we consider the decomposition

$$(1) \quad X_n = \sum_{i=0}^{\infty} \psi_{n,i} Z_{n-i} + V_n$$

for some set of orthogonal innovations $\{Z_t\}$ and a "residual" V_n. Wold's (1938) seminal work provides a foundation for characterizing stationary finite variance processes with covariance orthogonal innovations. In the classic setting the innovations necessarily satisfy the strong orthogonality condition

$$(2) \quad \overline{sp}(Z_{t+i}, ..., Z_t) \perp \overline{sp}(X_{t-1}, ..., X_{t-j}) \ \ \forall i \geq 0, \forall j \geq 1,$$

where $\overline{sp}$ denotes the closed linear span. In general Banach spaces, however, the "covariance" may not exist, conditional expectations and the best predictor may not equate, metric projection operators need not be linear, and innovations may not satisfy (2) although they will satisfy (3), below. A related decomposition theory with strong orthogonal innovations for processes with an unbounded variance, or for any process based on metric projection other than minimizing the mean-squared-error, is relatively limited and the most promising contributions to the literature focus entirely on closed linear spans.

Let $\overline{sp}_t = \overline{sp}(X_s : s \leq t)$ denote the closed linear span of $\{X_t\}$, and denote by P_t a metric projection operator (e.g. $P_{t-1} : \overline{sp}_t \rightarrow \overline{sp}_{t-1}$). Urbanic (1964, 1967) considered decompositions of strictly stationary infinite variance processes which admit independent metric projection innovations. Faulkner and Huneycutt (1978) consider decompositions with innovations $\{Z_t\}$ that only satisfy a *weak* asymmetric orthogonality condition:

$$(3) \quad \overline{sp}(Z_t) \perp \overline{sp}(X_{t-1}, ...X_{t-j}) \ \ \forall j \geq 1.$$

Miamee and Pourahmadi (1988) develop a weak-orthogonal decomposition theory for p-stationary processes based on innovations in $\overline{sp}_t - P_{t-1}\overline{sp}_t$. The theory, however, fundamentally exploits the codimension one property of closed linear spans: $\overline{sp}_t = \overline{sp}(X_t, \overline{sp}_{t-1})$.

Similarly, Cambanis *et al* (1988) establish an asymmetric decomposition theory for $L_p(\ , \mathfrak{F}_t, \mu)$ processes in $\overline{sp}_t$. The authors prove (i) projection operator linearity, (ii) iterated projections and (iii) the existence of strong orthogonal innovations are equivalent when the innovations are restricted to the space $\overline{sp}_t$

[1] For example, all stable random variables with characteristic exponent $\alpha \in (0, 2]$ belong to an L_p-space, $p < \alpha$, and all L_p-spaces are Banach spaces. Smooth, reflexive Banach spaces are complete normed spaces endowed with a semi-inner product which induces the norm.

$- P_{t-1}\overline{sp}_t$. Specifically, the authors prove $(i) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$, and operator linearity is expedited by the fact that $\overline{sp}_{t-1}$ is codimension one in $\overline{sp}_t$.

In the literature, therefore, either independence is assumed; only weak orthogonality is proven; or explicit properties of closed linear spans are exploited to promote strong orthogonality. For projections into arbitrary (nonlinear) L_p-spaces, Cambanis *et al* (1988) point out that operator linearity sufficiently renders strong orthogonal innovations (2). This result, however, is trivial: see Theorem 3, below. Moreover, no result exists (that we know of) characterizing strong orthogonal decompositions for finite variance processes based on best L_p-metric projection, $p < 2$.

The construction and use of orthogonal innovation spaces $\overline{sp}_t - P_{t-1}\overline{sp}_t$ in order to promote strong orthogonality is not a trivial simplification, however. The explicit omission of nonlinear best predictors and orthogonal innovations must be viewed critically in light of developments in the theory and empirical methods associated with nonlinear stochastic processes. For example, moving average forms have been utilized to characterize linear dependence within processes with regularly varying tails: see Davis and Resnick (1985a,b), and Kokoszka and Taqqu (1994, 1996). The innovations in this literature are typically assumed to be *iid*, hence strongly orthogonal to far more than subspaces of $\overline{sp}_t$. Except for the special case of symmetric stable process (Cambanis *et al*, 1988), nowhere in this literature are necessary and sufficient conditions for the existence of such moving averages derived.

Moreover, we see the use of moving averages a la IRF's in time series settings in which amassed evidence suggests nonlinear data generating processes with heavy tails[2]. In the economics and finance literatures the implied "impulses" are predominantly assumed to be *iid* finite variance innovations computed from inherently linear vector autoregression [VAR] representations (e.g. Sims, 1980). A linear structure ensures symmetry with respect to how positive and negative shocks persist over time, and renders shocks independent of the history of the process. If we wish to track heavy-tailed shocks with asymmetric impacts on the level process, based on best (nonlinear) forecasts, then a decomposition theory that goes substantially beyond the extant literature is required.

Toward this end, Gallant *et al* (1993) and Koop *et al* (1996) develop nonparametric representations of impulse responses for general nonlinear processes in the Hilbert space $L_2(\ , \mathfrak{F}_t, \mu)$. The impulses are assumed to be independent, and the responses are simply defined as differences between conditional expectations. As stated above, the conditional expectations may not be the best predictor in a general Banach space (e.g. $L_p(\ , \mathfrak{F}_t, \mu)$, $p < 2$). Gourieroux and Jasiak (2003) develop a parametric volterra-type expansion of *iid* Gaussian innovations for strongly stationary, square integrable processes that do not display long memory properties. In this case the level process has a finite variance and limited memory, and the innovations are assumed to be symmetrically dis-

[2] Separate evidence suggests that many foreign exchange rates and asset returns are heavy-tailed, and the best one-step ahead forecasts are nonlinear. See, e.g., Hol and de Vries (1991), Cheung (1993), Gallant *et al* (1993), Phillips *et al* (1996), Lin (1997), Mikosch and Stărică (2000), Falk and Wang (2003), and Hill (2005a).

tributed.

In this paper, we extend orthogonal decomposition theory to its arguable limit. For any smooth, reflexive Banach space $\mathfrak{B}_t$ we prove in Theorem 3 (the main result) the property of iterated projections is necessary and sufficient for the existence of a decomposition with asymmetrically strong orthogonal innovations, $\overline{sp}(Z_{t+i}, ..., Z_t) \perp \mathfrak{B}_{t-1}$, $\forall i \geq 1$. Using an arbitrary metric projection mapping $P_{t-1} : \mathfrak{B}_t \rightarrow \mathfrak{B}_{t-1}$, our results do not exploit operator linearity in general, and specifically do not rely on properties of the closed linear span. Theorem 3 allows for a simple characterization of a nonlinear IRF based on best L_p-metric projection. Our results include as special cases Wold decompositions of Hilbert space processes; of L_p-space processes; of processes in Banach spaces which do not admit a linear metric projection operator; of long memory, or non-stationary, or non-square integrable processes; of processes with asymmetrically distributed innovations/impulses; and does not restrict projection mappings to closed linear spans. Moreover, our primitive result linking iterated projections to strong orthogonality holds for any appropriate L_p-metric projection operator even if the process belongs to L_2. For example, our theory fully characterizes when the best L_1-predictor of a finite variance process generates strong orthogonal errors.

If the operator P_t does not iterate then a strong orthogonal moving average does not exist. We lose moving average-based nonlinear IRF's with adequately noisy impulses, and theories of linear dependence for moving averages with independent innovations do not apply. Conversely, if a strong orthogonal moving average form does not exist then the projection operator does not iterate, $P_s P_t \neq P_s$ for some or all $s < t$. In this case we lose an array of prediction-based results which rely on iterated projections, including iterative multi-step ahead forecasts and nonlinear IRF's based on L_p-metric projection.

We make the theory concrete by constructing in Corollary 4 a parametric decomposition of the form (1) with solutions for $\{\psi_{n,i}\}$. In Section 4 we then develop a theory of nonlinear impulse response functions based on best L_p-metric projection and the properties of strong orthogonality and iterated projections. We construct an extended example in Section 5 demonstrating the decomposition of a nonlinear smooth transition threshold model and associated nonlinear impulse response function. Although Theorem 3 characterizes the dual relationship between prized prediction characteristics, it says nothing about when they will hold or how to verify that they hold. This is compounded by the inherent difficulty associated with computing the best L_p-predictor[3]. We therefore focus our attention on the empirical task of verifying whether the nonlinear model actually presents the conditional expectation and/or the best L_p-predictor, and whether the proposed decomposition innovations are strong orthogonal. We apply the methods to daily returns of the Yen, Euro and British Pound exchange

[3] See, for example, Cline (1983) and Cline and Brockwell (1985) for best prediction of ARMA processes with infinite variance. Although operator theory represents an important branch of mathematics and statistics, theoretical and applied work aimed at functionally characterizing and verifying best non-linear predictors in general Banach spaces is apparently non-existent.

rates against the U.S. Dollar. We find significant evidence that the threshold model adequately characterizes the best L_p-predictor for some $p < 2$ for some exchange rates. In particular evidence suggests the threshold model generates strong orthogonal innovations for the Pound based on best $L_{1.5}$-metric projection.

The rest of paper is organized as follows: Section 2 contains preliminary metric projection theory; Section 3 contains the main results; we develop a theory of nonlinear impulse response functions in Section 4; and Section 5 contains an example and an application. The appendix contains formal proofs, and tables and figures are placed at the end of the paper.

In the sequel, we employ the following notation and definition conventions. Denote by $\mathfrak{B}_t \equiv \mathfrak{B}(\ ,\mathfrak{F}_t, \mu, \|\cdot\|)$ a closed, smooth, reflexive Banach measure space of nondeterministic stochastic processes $\{X_\tau : \tau \leq t\}$ endowed with the norm $\|\cdot\|$, measure μ, and σ-field $\mathfrak{F}_t = \sigma(X_\tau : \tau \leq t)$. Denote $\mathfrak{B} \equiv \mathfrak{B}(\ ,\mathfrak{F}, \mu, \|\cdot\|)$ $= \cup_{t\in\mathbb{Z}}\mathfrak{B}_t$, $\mathfrak{F} = \cup_{t\in\mathbb{Z}}\mathfrak{F}_t$. It is understood that $\|x\| < \infty$ for any $x \in \mathfrak{B}_t$, and $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$. Let $\mathfrak{L}_t \equiv L_p(\ ,\mathfrak{F}_t, \mu)$, $p \leq 2$. We denote the signed power $sgn(z)|z|^a$ as $z^{<a>}$, $a \in \mathbb{R}$. Denote by $\perp$ any orthogonality condition in $\mathfrak{B}_t$, and let $\mathfrak{B}_t^\perp$ denote the orthogonal complement of $\mathfrak{B}_t$.

For closed linear subspaces of $\mathfrak{B}_t$, say $S_1, ..., S_n$, $n > 1$, we write $S_1 + ... + S_n$ (synonymously $\sum_{i=1}^n S_i$) to denote the stochastic space $\{\sum_{i=1}^n Z_i : Z_i \in S_i\}$. For orthogonal subspaces, $S_1, ..., S_n$, $n > 1$, the space $S_n \oplus S_{n-1} \oplus ... \oplus S_1$ (synonymously $\sum_{i=0}^{n-1} \oplus S_{n-i}$) denotes the space $\sum_{i=0}^{n-1} S_i$ where

(4) $$\sum_{i=0}^{l-1} S_{n-i} \perp \sum_{i=l}^{n-1} S_{n-i}$$

for all $1 \leq l < n$. In general orthogonality is not symmetric. For spaces $\sum_{i=0}^{n-1} \oplus S_{n-i}$, we say the subspaces S_t are *strong orthogonal*. Similarly, whenever $S_t \perp S_s$ $\forall s < t$, we say the subspaces S_t are *weak orthogonal*. Clearly, strong orthogonality implies weak orthogonality.

2. Projection Operators and Orthogonality in Banach Space

The subsequent decomposition theory is based on orthogonal innovations Banach spaces. For background theory, see Singer (1970), Lindenstrauss and Tzafriri (1977), Giles (1967, 2000) and Megginson (1998). For arbitrary random variables $(x, y) \in \mathfrak{B}$, we work with the property of James Orthogonality: y is *James orthogonal* to x whenever

(5) $$||y + \lambda x|| \geq ||y||$$

for every real scalar $\lambda \in \mathbb{R}$, denoted $y \perp x$: see James (1947). Banach space norms $\|\cdot\|$ may be supported by arbitrarily many semi-inner products $[,]$. However, for smooth spaces $\mathfrak{B}$ and $(x, y) \in \mathfrak{B}$, if y is orthogonal to x there exists one inner-product that supports $[y, x] = 0$ (see, e.g., Giles, 1967, and Singer, 1970).

Lemma 1 *Let $\mathfrak{B}$ be a Banach space with norm $\|\cdot\|$. For any subspaces $U, V \subseteq \mathfrak{B}$ such that $U \perp V$, there exists a semi-inner product $[,]$ that supports $\|\cdot\|$ such* *that* $[U,V] = 0$: *for each* $u \in U$ *and* $v \in V$, *the inner-product* $[,]$ *satisfies* $[u,v] = 0$ *and* $[u,u]^{1/2} = \|u\|$, $[v,v]^{1/2} = \|v\|$. *Moreover, if* $\mathfrak{B}$ *is smooth then* $[,]$ *is unique.*

Metric Projection Operators Consider arbitrary subspaces $U,V \subseteq \mathfrak{B}$, $\sigma(V) \subset \sigma(U)$, where $\sigma(V)$ denotes the sigma algebra induced by the elements of V. For some element $u \in U$, we say $v \in V$ is the "best predictor" of u with respect to V *if and only if*

(6) $\|u - v\| \leq \|u - \tilde{v}\|$

for every element $\tilde{v} \in V$. Because the space $\mathfrak{B}$ is reflexive, the predictor v exists and is unique. We define, then, the metric projection operator that maps $P : U \to V$ as $P(u|V) = v$: the projection $P(u|V)$ is identically the "best predictor" of u . The projection $P(u|V)$ is continuous, bounded, and idempotent, although not in general linear: see below. For subspaces $U,V \subseteq \mathfrak{B}$, the notation $P(U|V)$ is understood to represent the projection space: $P(U|V) = \{P(u|V) : u \in U\}$.

Iterated Projections We say that the property of *iterated projections* holds in $V_1 \subseteq \mathfrak{B}$ for some projection operator $P : U \to V_1$ when for any subspace $V_0 \subseteq V_1 \subseteq \mathfrak{B}$, $P(P(u|V_1)|V_0) = P(u|V_0)$. The property exists in any real Hilbert space, however iterated projections need not hold in an arbitrary Banach space: see Theorem 3.

Operator Linearity We say that a projection operator P which maps $P : U \to V$ is a *linear operator* on $U \supseteq V$ if for any elements $u_1, u_2 \in U$ and any real numbers $a,b \in \mathbb{R}$, $P(au_1 + bu_2|V) = aP(u_1|V) + bP(u_2|V)$, an homogenous, additive function of u_1 and u_2. Metric projection operators are always linear in real Hilbert spaces because inner products are linear in both arguments. Projection operators in $\mathfrak{B}$, however, need not be linear, although Lemma 2.*iv* depicts a quasi-linearity property. If a projection operator is a linear operator, however, the property of iterated projections holds: see Lemma 2.*viii*.

Metric Projection In the following, assume u is an arbitrary element of U, and denote by $E[u|V]$ the expectation of u conditioned on $\sigma(V)$. Consult Singer (1970), Giles (2000) and Megginson (1998).

Lemma 2 (i) *Orthogonality: the element* $v \in V$ *satisfies* $P(u|V) = v$ *if and only if* $(u - v)\perp V$ *if and only if* $[u - v, \tilde{v}] = 0$ *for a unique* $[,]$ *and every* $\tilde{v} \in V$; (ii) $Pu = 0$ *if and only if* $u \perp V$ *if and only if* $[u, \tilde{v}] = 0$ *for every* $\tilde{v} \in V$; (iii) *for any* $v \in V$, $P(v|V) = v$; (iv) *quasi-linearity: for any* $z \in V$ *and any* $u \in U$, $P(u + z|V) = P(u|V) + z$; (v) *norm-boundedness: for any element* $u \in U$, $||P(u|V)|| \leq k||u||$ *for some scalar* $0 < k < \infty$; (vi) *unbiasedness in* L_p : *if* $U, V \subseteq L_p$, $1 < p \leq 2$, *and if* $E[u|V] \in V$, *then* $P(u|V) = E[u|V] = v$ *with probability one if and only if* $(u - v)\perp V$ *and* $V \perp (u - v)$; (vii) *scalar-homogeneity:* $P(au|V) = aP(u|V)$ *for every real scalar* $a \in \mathbb{R}$; *and* $(viii)$ *if* P *is a linear operator on* $U \supseteq V$ *then* P *satisfies the property of iterated projections.*

Remark 1: Property (ii) implies if $u \perp V$ then $P(u|V_0) = 0 \; \forall V_0 \subseteq V$. Property (iii) identically implies idempotence: $PP = P$. The conditions for property (vi) are non-trivial: the conditional expectations $E[u|V]$ may not be an element of the space V. For example, suppose $V = \overline{sp}(v_1, ..., v_n)$, the closed linear span of stable random variables $(v_i)_{i=1}^{n}$ with tail index $\alpha < 2$, and suppose u is a stable random variable with tail index α. Then $P(u|V) \in V$ by construction, yet $E[u|V]$ need not by linear: see, e.g., Hardin *et al* (1991). Of course, for non-Gaussian processes in $L_2(\ ,\mathfrak{F}_t, \mu)$ the best L_2-predictor, $E[u|V]$, need not be linear.

3. Main Result: Strong Orthogonal Decomposition in $\mathfrak{B}_n$ The main orthogonal decomposition result follows. Denote by $P_{t,s}$ the metric projection mapping $P : \mathfrak{B}_t \to \mathfrak{B}_s$, $s \leq t$. Construct the following spaces

$$(7) \quad \mathfrak{B}_{-\infty} = \bigcap\nolimits_n \mathfrak{B}_n \qquad \mathfrak{B}_{+\infty} = \bigcup\nolimits_n \mathfrak{B}_n.$$

We assume the spaces $\mathfrak{B}_t$ contain only non-deterministic processes such that $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$ and $\mathfrak{B}_{t-1} \subset \mathfrak{B}_t$. Consequently, prediction generates non-trivial errors: $||X_t - P_{t,s}X_t|| > 0 \; \forall X_t \in \mathfrak{B}_t, \forall s < t$.

Theorem 3 *For any space $\mathfrak{B}_n$, there exists a sequence of subspaces $\{N_{n-i}\}_{i=0}^{\infty}$, $N_t \subseteq \mathfrak{B}_t$, such that*

$$(8) \quad \mathfrak{B}_n = \left(\sum\nolimits_{i=0}^{\infty} N_{n-i}\right) + \mathfrak{B}_{-\infty},$$

where $N_t \perp \mathfrak{B}_{t-1}$ and $N_t \perp N_s$ for every $s < t \leq n$. Moreover, the following are equivalent:

i. $\mathfrak{B}_n = (\sum_{i=0}^{\infty} \oplus N_{n-i}) \oplus \mathfrak{B}_{-\infty}$, *where* $\sum_{i=0}^{k-1} \oplus N_{n-i} \perp \mathfrak{B}_{n-k}, \forall k \geq 1$.

ii. $P_{t,t-l}P_{t,t-k} = P_{t,t-k}$, *for every* t, $k \leq l$.

Furthermore, provided (i) holds, every element $Y \in \sum_{i=0}^{\infty} \oplus N_{n-i}$ obtains a unique norm-convergent expansion, $Y = \sum_{i=0}^{\infty} \xi_{n-i}$, for some $\xi_t \in N_t$.

Remark 1: Cambanis *et al* (1988) point out that operator linearity implies result (i) for processes in $L_p(\ ,\Im_n, \mu)$ and for projection into arbitrary $L_p(\ ,\Im_n, \mu)$-spaces. This result, however, is trivial and does not anticipate the dual relationship between orthogonality and iterated projections, $(i) \iff (ii)$, without invoking operator linearity. Assume $P_{t,s}$ is a linear operator on $\mathfrak{B}_t$, and consider any element $\sum_{i=k}^{l} \xi_{n-i} \in \sum_{i=k}^{l} N_{n-i}$, $\xi_{n-i} \in N_{n-i}$, $0 \leq k \leq l$. Because $N_{n-i} \perp \mathfrak{B}_{n-i-1}$ by construction and by Lemma 2.ii we have $P_{n,n-i-1}\xi_{n-i} = 0$. By operator linearity we conclude

$$(9) \quad P_{n,n-l-1} \sum\nolimits_{i=k}^{l} \xi_{n-i} = \sum\nolimits_{i=k}^{l} P_{n,n-l-1}\xi_{n-i} = 0$$

hence $\sum_{i=k}^{l} \xi_{n-i} \perp \mathfrak{B}_{n-l-1}$, cf. Lemma 2.$ii$. Because the element $\sum_{i=k}^{l} \xi_{n-i} \in \sum_{i=k}^{l} N_{n-i}$ is arbitrary, we deduce $\sum_{i=k}^{l} N_{n-i} \perp \mathfrak{B}_{n-l-1}$ for any $0 \leq k \leq l$. This identically implies strong orthogonality of the innovations spaces N_t.

Remark 2: Theorem 3 characterizes the existence of a decomposition for any process in a smooth reflexive Banach space based on any appropriate metric-projection operator. This includes processes with a finite variance where the metric projection operator is not restricted to the minimum mean-squared-error operator. This will be particularly useful if evidence suggests a chosen linear or nonlinear model of a finite variance process does not represent the best L_2-predictor but does characterize the best L_p-predictor for some $p < 2$: see Section 4.

Because any element $Y \in \sum_{i=0}^{\infty} \oplus N_{n-i}$ obtains a unique norm-convergent series representation, we may write elements $X_n \in \mathfrak{B}_n$ in a straightforward moving-average form.

Corollary 4 *Consider any Banach space $\mathfrak{B}_n$ such that a Wold decomposition exists.*
i. For every $X_n \in \mathfrak{B}_n$ there exists a sequence of orthogonal subspaces $\{N_{n-i}\}_{i=0}^{\infty}$, $N_t \subseteq \mathfrak{B}_t$, a sequence of stochastic elements $\{Z_t\}$, $Z_t \in N_t$, an element $V_n \in \mathfrak{B}_{-\infty}$ and real-numbers $\{\psi_{n,i}\}_{i=0}^{\infty}$,, such that

$$(10)\quad X_n = \sum_{i=0}^{\infty} \psi_{n,i} Z_{n-i} + V_n$$

where the series $\sum_{i=0}^{\infty} \psi_{n,i} Z_{n-i}$ is norm-convergent, and the innovations Z_t are strong orthogonal in the sense that $Z_t \perp \mathfrak{B}_{t-1}$ and

$$(11)\qquad \overline{sp}(Z_n, Z_{n-1}...) \perp \mathfrak{B}_{-\infty}$$

$$(12)\qquad \overline{sp}(Z_{t+i}, ..., Z_t) \perp \mathfrak{B}_{t-1}, \quad \forall t \leq n,\ \forall i \geq 0,$$

if and only if $P_{t,t-l} P_{t,t-k} = P_{t,t-l}$, for every t, $k \leq l$.
ii. Moreover, $\psi_{n,0} = 1$, and the coefficients $\psi_{n,i}$ uniquely satisfy the recursive relationship for $i = 1, 2, ...$,

$$(13)\quad \psi_{n,i} = \frac{[Z_{n-i}, X_n]}{[Z_{n-i}, Z_{n-i}]} - \sum_{j=0}^{i-1} \psi_{n,j} \frac{[Z_{n-i-j}, Z_n]}{[Z_{n-i}, Z_{n-i}]}.$$

Remark: Although $Z_n \equiv X_n - P_{n,n-1} X_n$, by definition, for an arbitrary process $\{X_n\}$ we cannot in general say $Z_{n-i} = X_{n-i} - P_{n-i,n-i-1} X_{n-i}$, the best one-step ahead prediction error of X_{n-i}[4]. If iterated projections hold (equivalently, if the innovations are strong orthogonal), then $Z_{n-i} \equiv P_{n,n-i} X_n - P_{n,n-i-1} X_n = P_{n,n-i} X_n - P_{n,n-i-1} P_{n,n-i} X_n$, the innovation based on a one-step ahead projection of the i-step ahead forecast.

Metric projection errors in Banach space need not be serially independent. Indeed, they need not be symmetrically orthogonal. In general they need only satisfy a weak orthogonality condition $\overline{sp}(Z_t) \perp \mathfrak{B}_{t-1}$. Nonetheless, we can establish necessary and sufficient conditions symmetric orthogonality to hold for processes in $L_p(\ \ , \mathfrak{F}_t, \mu)$, $1 < p \leq 2$. We write $\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} S_{n-i}$ to denote the *symmetrically strong orthogonal* space $\sum_{i=0}^{n-1} S_i$ where

$$(14)\quad \sum_{i=0}^{l-1} S_{n-i} \perp \sum_{i=l}^{n-1} S_{n-i} \ \text{ and } \ \sum_{i=l}^{n-1} S_{n-i} \perp \sum_{i=0}^{l-1} S_{n-i},$$

[4] An exception is for causal-invertible ARMA processes: see, e.g., Cline (1983).

for all $1 \leq l < n$.

Denote by $P_{t,s}$ the mapping $P : \mathfrak{L}_t \rightarrow \mathfrak{L}_s$, $s \leq t$. Provided a strong orthogonal decomposition exists at all, the innovations will be symmetrically strong orthogonal *if and only if* the decomposition sub-spaces are symmetrically orthogonal, which occurs *if and only if* the best predictor $P_{t,t-1}X_t$ is the conditional expectations $E[X_t|\mathfrak{F}_{t-1}]$. Of course, if the best predictor coincides with the conditional expectations, then operator linearity and iterated projections trivially follow from properties of expectations, and therefore a strong orthogonal decomposition exists.

Theorem 5 *Denote by N_t the projection error space, $\mathfrak{L}_t - P_{t,t-1}\mathfrak{L}_t$. The following are equivalent:*

i. $\mathfrak{L}_n = \left(\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} N_{n-i}\right) \overleftrightarrow{\oplus} \mathfrak{L}_{-\infty}$ *for arbitrary* n;

ii. $\mathfrak{L}_{t-1} \perp N_t$, $\forall t \leq n$;

iii. $P_{t,t-1}X_t = E\left(X_t|\mathfrak{F}_{t-1}\right)$, $\forall t \leq n$;

iv. $E\left[\left(X_t - E\left(X_t|\mathfrak{F}_{t-1}\right)\right)^{<p-1>}|\mathfrak{F}_{t-1}\right] = 0$, $\mathfrak{F}_{t-1}$*-a.e.*

Remark 1: If the decomposition innovations are symmetrically orthogonal, then for $\{Z_t\}$ in (10), $[Z_s, Z_t] = 0, \forall s \neq t$, and $\psi_{n,i} = [Z_{n-i}, X_n]\,/\,[Z_{n-i}, Z_{n-i}]$, $i = 1, 2, \ldots$

Remark 2: Condition (iv) is simply a martingale difference property for $L_p(\ \ , \mathfrak{F}_t, \mu)$ processes, and implies $E[(X_t - E(X_t|\mathfrak{F}_{t-1}))^{<p-1>}Y_{t-1}] = 0$, for every $\mathfrak{F}_{t-1}$-measurable random variable Y_{t-1}. By (iii) we may identically write $P_{t,t-1}(X_t - P_{t,t-1}(X_t))^{<p-1>} = 0$.

Remark 3: The above conditions constitute necessary conditions for innovations spaces to be independent.

4. Nonlinear Impulse Response Functions in $\mathfrak{L}_t$ In the following we develop a general theory of nonlinear Impulse Response Functions [IRF] based on strong orthogonal decomposition innovations. Let V_t be an $\mathfrak{L}_t$-valued random variable, define the sequence of spaces $\{\tilde{\mathfrak{L}}_t\} = \{\mathfrak{L}_t \oplus V_{t+1}\}$, and let $\mathfrak{L}_{-\infty} = \{0\}$ for simplicity. Define the h-step ahead nonlinear impulse response function

$$(15)\quad I(h, V_t, \mathfrak{L}_{t-1}) = P(x_{t+h}|\tilde{\mathfrak{L}}_{t-1}) - P(x_{t+h}|\mathfrak{L}_{t-1}).$$

The above definition simply generalizes the expectations based format of Koop *et al* (1996): the response at horizon h is the best h-step ahead prediction response to a random shock V_t at time t, conditioned on all past histories $\mathfrak{L}_{t-1}$. We could write $I(h, v_t, \omega_{t-1}) = P(x_{t+h}|\omega_{t-1}, v_t) - P(x_{t+h}|\omega_{t-1})$ to make explicit a particular history $\omega_{t-1} \equiv \{x_{t-1}, x_{t-2}, \ldots\}$ and particular shock v_t in the manner of Koop *et al* (1996). The impulse response function $I(h, V_t, \mathfrak{L}_{t-1})$ is an $\Im_{t-1}$-measurable random variable, and $I(h, v_t, \omega_{t-1})$ is simply a realization. We may compute $I(h, V_t, \mathfrak{L}_{t-1})$ for a large number of draws $\{v_t, \omega_{t-1}\}$ from the joint distribution of V_t and $\{X_{t-1}, X_{t-2}, \ldots\}$. An empirical distribution function and confidence bands of the responses $I(h, v_t, \omega_{t-1})$ can then be estimated: see Section 5.

Theorem 6 *Assume the process $\{x_\tau : \tau \leq t\}$ lies in $L_p(\ , \Im_t, \mu)$ and obtains a strong orthogonal decomposition $x_t = \sum_{i=0}^{\infty} \psi_{t,i} Z_{t,t-i}$ with respect to the subspaces $\{\mathfrak{L}_\tau : \tau \leq t-1\}$. Assume the metric projection operator $P : \mathfrak{L}_t \to \mathfrak{L}_{t-1}$ iterates from $\mathfrak{L}_{t-k}$ to $\mathfrak{L}_{t-k-1}$ for any k. Then*

(16) $I(h, V_t, \mathfrak{L}_{t-1}) = \psi_{t+h,h} P(Z_{t+h,t} | \tilde{\mathfrak{L}}_{t-1})$

Remark 1: Strong orthogonality is required, cf. Theorem 3, because the line of proof exploits iterated projections.

Remark 2: An h-step ahead "impulse response" is simply a scaled predicted strong orthogonal innovation, where the prediction exploits information contained in the random impulse V_t. In a standard linear setting, $x_t = \sum_{i=0}^{\infty} \psi_i V_{t-i}$, $V_t \sim iid$, (16) reduces to a classic representation: for any particular history ω_{t-1} and impulse v_t, $I(h, v_t, \omega_{t-1}) = \psi_h v_t$.

Remark 3: In L_2 the nonlinear IRF $I(h, V_t, \mathfrak{L}_{t-1})$ is identically the generalized impulse response function characterized in Koop *et al* (1996: eq. (9)) as long as the projection operator minimizes the mean-squared-error. Otherwise (16) characterizes a further generalization of Koop *et al*'s (1996) generalized IRF to best L_p-projection of finite variance processes.

Remark 4: Koop *et al* (1996) characterize non-parametric and bootstrap methods for estimating the conditional expectations based on draws from the empirical distributions of $\{x_{t-1}, x_{t-2}, ...\}$ and V_t. It is beyond the scope of the present paper to consider such comparable bootstrap methods for approximating a best L_p-predictor. In the sequel we estimate $\psi_{n+h,h}$ and $P(Z_{n+h,n} | \tilde{\mathfrak{L}}_{n-1})$ directly using in-sample information and either imputed or simulated impulses, under assumed stationarity (e.g. $\psi_{n+h,h} = \psi_{n,h} = \psi_h$ for all n).

Requiring the operator to iterate from $\mathfrak{L}_{t-k}$ to $\tilde{\mathfrak{L}}_{t-k-1}$ (i.e. $P(P(x_t | \mathfrak{L}_{t-k}) | \tilde{\mathfrak{L}}_{t-k-1})$ $= P(x_t | \tilde{\mathfrak{L}}_{t-k-1})$) does not diminish the generality of the result by very much. For example, if $\Im_t \equiv \sigma(x_\tau : \tau \leq t) = \sigma(\epsilon_\tau : \tau \leq t)$ for some stochastic process $\{\epsilon_t\}$, and the impulses V_t are simply ϵ_t then the assumption holds because $\tilde{\mathfrak{L}}_t$ $= \mathfrak{L}_t \oplus V_{t+1} = \mathfrak{L}_t \oplus \epsilon_{t+1} = \mathfrak{L}_{t+1}$.This will hold for infinitely large classes of linear and nonlinear processes: see Section 5 for an example.

Lemma 7 *Let $\Im_t \equiv \sigma(x_\tau : \tau \leq t) = \sigma(\epsilon_\tau : \tau \leq t)$, and $V_t = \epsilon_t$ for all $t \in \mathbb{Z}$. Then $P(P(x_t | \mathfrak{L}_{t-k}) | \tilde{\mathfrak{L}}_{t-k-1}) = P(x_t | \tilde{\mathfrak{L}}_{t-k-1})$ for all $k \geq 0$. Additionally, if x_t admits a strong orthogonal decomposition, then $P(P(x_t | \tilde{\mathfrak{L}}_{t-k}) | \mathfrak{L}_{t-k}) = P(x_t | \mathfrak{L}_{t-k})$.*

5. ARCH-M and Threshold Models, and an Application In practice the analyst will need to verify whether a particular decomposition actually generates strong orthogonal innovations, and indeed whether the predictor used to generate the innovations actually represents the best predictor. The verification of such properties is required as a necessary foundation for generating an exact nonlinear IRF which requires iterated projections, cf. Theorems 3 and 6. In this section we focus our attention entirely on a simple threshold model for the sake of brevity.

Due to linearity and iterations properties the predominant practice in the literature is to assume a particular model represents the conditional mean, which may not be the best L_p-predictor for some, or finitely many, $p > 0$. As a nod toward convention and practical simplicity, we explore a conditional expectations-based decomposition and discuss model specification tests to verify whether the conjectured model represents the best L_p-predictor for any $p \in (1, 2]$ and whether the resulting predictor errors are strong orthogonal. We then derive a sample nonlinear IRF for the particular threshold model, and apply the model and specification tests to the daily returns of currency exchange rates[5].

5.1 Linear "ARCH-in-Mean" in $\mathfrak{L}_t$

Let

(17) $x_t = \sum_{i=0}^{\infty} \theta_i u_{t-i} + \sum_{i=0}^{\infty} \beta_i |u_{t-i}|$,

where $u_t \in \mathfrak{L}_t = L_p(\ \ , \mathfrak{F}_t, \mu)$ are mean-zero independent shocks, $p > 1$, $\theta_0 = 1$ and $\beta_0 = 0$. The data generating process can be thought as a nonlinear "ARCH-in-Mean" with respect to absolute conditional volatility[6]: $x_t = \sum_{i=0}^{\infty} \theta_i u_{t-i} + h_t$, where, say, $h_t = P(|x_t| \mid \overline{sp}(|u_{t-1}|, |u_{t-2}|, ...)) = \sum_{i=0}^{\infty} \beta_i |u_{t-i}|$.

Because the $u_t^{\prime}s$ are serially independent, any function of u_t is independent of $\mathfrak{L}_{t-1}$, and the property

(18) $i.\ P_{t-j,t-j-1}(\theta_j u_{t-j} + \beta_j |u_{t-j}|) = a_j$

holds for some $a_j \in \mathfrak{L}_{-\infty}$ for all $j = 1, 2, ...$ For example, if $\beta_j = 0$ then $a_j = 0$.

Now consider the following property:

(19) $ii.\ P_{t,t-m}(u_t + Y_{t-1}) = P_{t,t-m} Y_{t-1},\ \forall Y_{t-1} \in \mathfrak{L}_{t-1},\ u_t \bot Y_{t-1},\ \forall m \geq 1.$

Property (ii) is sufficient for the existence of a strong orthogonal decomposition of x_t.

Theorem 8 *Assume (ii) holds with respect to $\{\tilde{\mathfrak{L}}_t\}$ and $\{\mathfrak{L}_t\}$. Then x_t obtain a strong orthogonal decomposition $x_t = \sum_{i=0}^{\infty} \psi_{t,i} Z_{t,t-i} + V_t$, $V_t \in \mathfrak{L}_{-\infty}$, $Z_{t,t-i} = u_{t-i} + (1/\theta_i)(\beta_i |u_{t-i}| - a_i)$ and $\psi_{t,i} = \theta_i$ if $\theta_i \neq 0$; $Z_{t-i} = |u_{t-i}| - a_i/\beta_i$ and $\psi_{t,i} = \beta_i$ if $\theta_i = 0$ and $\beta_i \neq 0$; $Z_{t,t-i} = u_{t-i}$ and $\psi_{t,i} = \theta_i$ if $\beta_i = 0$. Moreover, the nonlinear IRF is exactly*

$$
\begin{aligned}
(20)\quad I(h, V_t, \mathfrak{L}_{t-1}) &= \psi_h P(u_t + (1/\theta_h)(\beta_h |u_t| - a_h) | \tilde{\mathfrak{L}}_{t-1}) \text{ if } \theta_h \neq 0; \\
&= P(\beta_h |u_t| - a_h | \tilde{\mathfrak{L}}_{t-1}) \text{ if } \theta_h = 0,
\end{aligned}
$$

where $a_h = 0$ if $\beta_h = 0$. If $V_t = u_t$, then simply $I(h, V_t, \mathfrak{L}_{t-1}) = \theta_h V_t + \beta_h |V_t| - a_h$.

[5] The present section is in no way offered as an exhaustive study of the tail, memory or nonlinear properties of daily exchange rates. The reader may consult the vast extant literature for background theory and evidence.

[6] The "mean" traditionally refers to the conditional mean in which the ARCH term occurs. In the present general environment, however, "mean" is simply an incorrect euphemism for "best predictor".

Remarks: Property (ii) is similar in spirit to quasi-linearity, cf. Lemma 2: linearity holds if $u_t \perp Y_{t-1}$. Of course, if $m = 1$ then the equality follows directly from quasi-linearity. Notice that $u_t's$ are strong orthogonal (they are independent), but the decomposition innovations $Z_{t,t-i}$ need not be. Moreover, the $u_t's$ may be heterogenous: we do not assume they are *iid*.

5.2 Threshold Models and Orthogonal Decomposition

A growing literature suggests the returns to many macroeconomic and financial time series have heavy tails, are serially uncorrelated, and have some form of nonlinear structure. See Tong (1990), Kees and Kool (1992), Loretan and Phillips (1994), Franses and van Dijk (2000), Lundbergh *et al* (2003), and Lundbergh and Teräsvirta (2005), to name a few. In particular, the daily log returns of many currency exchange rates appear to be serially uncorrelated, have an infinite kurtosis or infinite variance, and display serially asymmetric extremes: see Hols and de Vries (1991) and Hill (2005b) and the citations therein. Moreover, the nonlinear structure of exchange rates has an intuitive regime transitional form based on "banding" policies in economic unions: see, e.g., Lundbergh and Teräsvirta (2005).

Together, the characteristics of noisy returns, persistent extremes, and currency policies suggest daily exchange rate returns may be governed by Smooth Transition Autoregression [STAR] data generating process, cf. Saikkonen and Luukkonen (1988), Teräsvirta (1994), Micheal *et al* (1997), and Lundbergh and Teräsvirta (2005). Denote by x_t the log return $\Delta \ln y_t$ of a daily exchange rate y_t. Let $x_t \in L_p(\ , \Im_t, P) = \mathfrak{L}_t$, $\Im_t = \sigma(x_s : s \leq t)$, and let α be the moment supremum of ϵ_t: $E|\epsilon_t|^p < \infty$ for all $p < \alpha$. Assume $\alpha > 1$ and consider any $1 < p < \min\{\alpha, \alpha/4+1\}$. Simple STAR models which capture the above stylized traits include the exponential and logistic STAR

$$(21) \quad x_t = \phi x_{t-1} g_{t-1}(\gamma, c) + \epsilon_t, \quad |\phi| < 1, \ \gamma \geq 0, \ c > 0,$$

where ϵ_t is strictly stationary, $E[\epsilon_t|\Im_{t-1}] = 0$, $g_{t-1}(\gamma, c) = \exp\{-\gamma(|\epsilon_{t-1}| - c)^2\}$ in the ESTAR case, and in the LSTAR case $g_{t-1}(\gamma, c) = [1 + \exp\{-\gamma(|\epsilon_{t-1}| - c)\}]^{-1} - [1 + \exp\{\gamma c\}]^{-1}$. Whether $E[\epsilon_t|\Im_{t-1}] = 0$ is supported in practice will be considered below. There are many available variations on this theme, and numerous alternative choices for the threshold variable (here we use the previous period's shock $|\epsilon_{t-1}|$): see van Dijk *et al* (2000).

The ESTAR form naturally articulates "inner" ($|\epsilon_{t-1}| \approx c$) and "outer" ($|\epsilon_{t-1}| \neq c$) regimes: when the previous period's shock $|\epsilon_{t-1}|$ is near c, $x_t \approx \phi x_{t-1} + \epsilon_t$ hence the return is serially persistent; when $|\epsilon_{t-1}|$ is far from c, $x_t \approx \epsilon_t$ such that the return is noisy. The LSTAR characterizes "lower" and "upper" regimes: respectively when $|\epsilon_{t-1}| \approx 0$ then $x_t \approx \epsilon_t$ and as $|\epsilon_{t-1}| \to \infty$ then $x_t = \phi[1 + \exp\{-\gamma c\}]^{-1}\}x_{t-1} + \epsilon_t$[7]. As the scale $\gamma \to \infty$ the LSTAR model

[7] Although the ESTAR functional form has been used to capture symmetric banding policies for exchange rate levels (e.g. Micheal *et al*, 1997; and Lundbergh and Teräsvirta, 2005), daily returns may illicit asymmetric responses from traders and policy makers: large deviations may suggest a market crisis whereas small deviations may not be noteworthy. See Engle and Ng (1993). Such volatility asymmetries have been recently modeled as smooth transition GARCH processes: see González-Rivera (1998) and McMillan and Speight (2002).

converges to a Self Exciting Threshold Autoregression: $x_t = \phi x_{t-1} I(|\epsilon_{t-1}| > c) + \epsilon_t$. The LSTAR model naturally implies extremes are persistent and non-extremes are noisy. The ESTAR model can also capture this asymmetry if c is extremely large: the returns will be noisy if $|\epsilon_{t-1}|$ is far from c which will predominantly occur when $|\epsilon_{t-1}| < c$.

We may decompose x_t by straightforward backward substitution.

Theorem 9 *Assume the stochastic process $\{x_\tau : \tau \leq t\}$ lies in $L_p(\ ,\Im_t,\mu)$, and assume (21) holds. Then $x_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \prod_{j=1}^{i} g_{t-j}(\gamma, c)$, a.s., $\Im_t = \sigma(\epsilon_\tau : \tau \leq t)$, and $x_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} + V_t$ where $V_t = b\phi/(1 - a\phi)$, $Z_t = \epsilon_t$, $\psi_0 = 1$ and for $i \geq 1$*

$$
\begin{aligned}
(22) \qquad Z_{t-i} &= [\epsilon_{t-i} g_{t-i}(\gamma, c) - b] + [g_{t-i}(\gamma, c) - a] \\
&\quad \times \sum_{j=i+1}^{\infty} \phi^{j-i} \epsilon_{t-j} \prod_{k=i+1}^{j} g_{t-k}(\gamma, c) \\
\psi_i &= \phi^i a^{i-1},\ i = 1, 2, ...
\end{aligned}
$$

where $a = E[g_t(\gamma, c)] \in [-1, 1]$, $b = E[\epsilon_t g_t(\gamma, c)]$, and $E[Z_{t-i}|\Im_{t-i-1}] = 0$. If $\{Z_t\}$ is L_p-strong orthogonal for some $p \leq 2$ then $P(x_{t+h}|\Im_t) = \sum_{i=h}^{\infty} \psi_{t+h,i} Z_{t+h,t+h-i}$.

Remark 1: It is straightforward to show

$$
(23) \quad \text{plim}_{N \to \infty} E[x_t|\Im_{t-N}] = E[x_t] = b\phi/(1 - a\phi),
$$

where the limit holds almost surely. If ϵ_t is symmetrically distributed then $b = 0$. If $\gamma = 0$ such that x_t is a simple AR(1) then $a = b = 0$, $Z_{t-i} = \epsilon_{t-i}$, and $\psi_i = \phi^i$. If $\phi = 0$ then trivially $x_t = Z_t = \epsilon_t$.

Remark 2: If $\{Z_t\}$ is strong orthogonal we deduce the h-step ahead forecast

$$
\begin{aligned}
(24) \quad P(x_{t+h}|\Im_t) &= \sum_{i=0}^{\infty} \phi^{i+h} a^{i+h-1} \{[\epsilon_{t-i} g_{t-i}(\gamma, c) - b] + [g_{t-i}(\gamma, c) - a] \\
&\quad \times \sum_{j=i+1}^{\infty} \phi^{j-i} \epsilon_{t-j} \prod_{k=i+1}^{j} g_{t-k}(\gamma, c)\}.
\end{aligned}
$$

5.2.1 Verifying Weak and Strong Orthogonality

For any t write

$$
(25) \quad Z_t = [\epsilon_t g_t(\gamma, c) - b] + [g_t(\gamma, c) - a] \times \sum_{j=1}^{\infty} \phi^j \epsilon_{t-j} \prod_{k=1}^{j} g_{t-k}(\gamma, c).
$$

Exploiting $E[\epsilon_t|\Im_{t-1}] = 0$ and the definitions of a and b, we know $E[Z_t|\Im_{t-1}] = 0$ such that Z_t is weakly orthogonal to $\mathfrak{L}_{t-1}$ in some sense: see Cambanis *et al* (1988). However, if we allow $E[\epsilon_t|\Im_{t-1}] \neq 0$ for both nonlinear forms $g_t(\gamma, c)$ such that the STAR model does not represent the best L_2-predictor, it may nonetheless represent the best L_p-predictor for some $p < 2$ and the innovations $\{Z_t\}$ may be weak and/or strong orthogonal in the sense of Section 3.

The decomposition innovations Z_t are weakly orthogonal to $\mathfrak{L}_{t-1}$ *if and only if* $P(Z_t|\Im_{t-1}) = 0$ (cf. Lemma 1.*ii*) which in L_p is true *if and only if*

$E[Z_t^{<p-1>}|\Im_{t-1}] = 0$. This may be easily tested for any chosen $p > 1$. For example Hong and White (1995) develop a nuisance parameter-free consistent nonparametric test of functional form based on the observation that if $Y_{t-1} \equiv E[Z_t^{<p-1>}|\Im_{t-1}] \neq 0$ then $E[Z_t^{<p-1>} Y_{t-1}] = Y_{t-1}^2 > 0$. Essentially any nonparametric estimator $\hat{Y}_{t-1}$ may be substituted for Y_{t-1}, including use of Fourier series, a flexible Fourier form, regression splines, etc. See Section 5.4. From $p < \alpha/4 + 1$, Minkowski's inequality, stationarity, the fact that $|g_t(\gamma, c)| \leq 1$ with probability one, and (25), it follows that $||Z_t||_{4(p-1)} \leq |b| + (1 + |a|)| \times |\epsilon_t||_{4(p-1)}(1-\phi)^{-1} < \infty$, hence $E|Z_t^{<p-1>}|^4 < \infty$ and $||Z_t^{<p-1>} - E[Z_t^{<p-1>}|\Im_{t-1}]||_{4(p-1)} < 2 \times ||Z_t^{<p-1>}||_{4(p-1)} < \infty$, such that the moment conditions of Hong and White (1996) are satisfied. Along with fairly standard regulatory assumptions the test statistic is simple to compute and is based on the sample moment $n^{-1}\sum_{t=1}^{n} \hat{Z}_t^{<p-1>} \hat{Y}_{t-1}$ for some plug in $\hat{Z}_t$ to be detailed below[8]. The statistic has an asymptotic standard normal null distribution.

For strong orthogonality in L_p we need $E[(\sum_{k=0}^{h} \pi_k Z_{t+k})^{<p-1>}|\Im_{t-1}] = 0$ for all $h \geq 0$ and every $\pi \in \mathbb{R}^h$. A simple method follows: randomly generate $\pi \in \mathbb{R}^h$ for various $h = 1, 2, ...$, perform the non-parametric test on the resulting $(\sum_{k=0}^{h} \pi_k Z_{t+k})^{<p-1>}$, repeat by generating a large number of sequences $\{\pi_k\}_{k=0}^{h}$ and subsequent test statistics, and average the resulting p-values.

In practice an estimated $\hat{Z}_t$ will be used as an obvious plug-in for Z_t. Because ϵ_t is unobservable simply assume $\epsilon_t = 0 \ \forall t \leq 0$ and $x_0 = 0$: for the ESTAR model, for example, $x_1 = \epsilon_1$, $x_2 = \phi x_1 \exp\{-\gamma(|x_1| - c)^2\} + \epsilon_2$, $x_3 = \phi x_2 \exp\{-\gamma(|x_2 - \phi x_1 \exp\{-\gamma x_1^2\}| - c)^2\} + \epsilon_3$, etc. Other methods for handling the first period may be considered as well. After constructing the regressor, the threshold model (15) can then be estimated straightforwardly by M-estimation or L_p-GMM for various $1 < p < 2$ using standard iterative estimation techniques, generating $\hat{\epsilon}_t$ and $\hat{Z}_t$. The moment condition is simply $E[\epsilon_t^{<p-1>} \partial/\partial\theta f_t(\theta)] = 0$ where $f_t(\theta) = \phi x_{t-1} g_{t-1}(\gamma, c)$ and $\theta = (\phi, \gamma, c)'$. See Arcones (2000), de Jong and Han (2002), and Han and de Jong (2004). For brevity, we assume all conditions which ensure the L_p-GMM estimator is consistent and asymptotically normally distributed hold: see de Jong and Han (2002).

5.2.2 Nonlinear Impulse Response Function

Assume $\{Z_t\}$ forms a sequence of strong orthogonal innovations and set $v_n = \epsilon_n$. From Theorem 5, Lemma 6 and Theorem 7 the h-step ahead nonlinear impulse response function $I(h, V_n, \mathfrak{L}_{n-1})$ is $\psi_{n+h,h} P(Z_{n+h,n}|\tilde{\mathfrak{L}}_{n-1})$, hence

$$(26) \qquad \begin{aligned} I(h, V_n, \mathfrak{L}_{n-1}) &= \psi_{n+h,h} P(Z_{n+h,n}|\tilde{\mathfrak{L}}_{n-1}) \\ &= \phi^h a^{h-1}\{[V_n g_n(\gamma, c) - b] + [g_n(\gamma, c) - a] \\ &\qquad \times \sum_{j=1}^{\infty} \phi^j \epsilon_{n-j} \prod_{k=1}^{j} g_{n-k}(\gamma, c)\}. \end{aligned}$$

The response to the "sole" random impulse $V_n = \epsilon_n$ is history $\{\epsilon_{n-i}\}_{i=1}^{\infty}$ dependent, and asymmetric with respect to the sign of V_n through $g_n(\gamma, c)$. See Koop

[8] There exists a multitude of consistent parametric and non-parametric model specification tests: see Bierens and Ploberger (1997) and the citations therein.

et al (1996) for further commentary on path dependence in nonlinear IRF's.

If we use estimated residuals $\{\hat{\epsilon}_t\}$ generated from estimates $\hat{\phi}$, $\hat{\gamma}$ and $\hat{c}$, and sample estimators $\hat{a} = n^{-1}\sum_{t=1}^{n} g_t(\hat{\gamma},\hat{c})$ and $\hat{b} = n^{-1}\sum_{t=1}^{n}\hat{\epsilon}_t g_t(\hat{\gamma},\hat{c})$, we obtain a sample IRF based on one history and one impulse $\{\hat{\epsilon}_t\}_{t=1}^{n}$: $\hat{I}(h,\hat{\epsilon}_n,\{\hat{\epsilon}_t\}_{t=1}^{n-1}) = \hat{\phi}^h\hat{a}^{h-1}\{[\hat{\epsilon}_n g_n(\hat{\gamma},\hat{c}) - \hat{b}] + [g_n(\hat{\gamma},\hat{c}) - \hat{a}] \times \sum_{j=1}^{n-1}\hat{\phi}^j\hat{\epsilon}_{n-j}\prod_{k=1}^{j} g_{n-k}(\hat{\gamma},\hat{c})\}$. Multiple alternative strategies for handling the random history $\{\epsilon_{n-i}\}_{i=1}^{\infty}$ and impulse V_n are available. For example, we may randomly draw a history $\{\epsilon_{n-i}\}_{i=1}^{n-1}$ and impulse v_n from the empirical distribution of the sample path $\{\hat{\epsilon}_t\}_{t=1}^{n}$, or simulate independent impulses if the finite distributions of $\{\epsilon_t\}$ are known (e.g. Pareto, stable, t, normal). We may repeat either method J-times generating sequences of histories and/or impulses and sequences of IRF's for each horizon h. Subsequent confidence bands and kernel densities may then be computed.

5.2.3 Empirical Study

Finally, we perform a limited empirical application of the STAR model to currency exchange rates. We study log returns $x_t = \Delta \ln y_t$ to the Yen/Dollar, Euro/Dollar and British-Pound/Dollar daily spot exchange rates $\{y_t\}$ for the period 1/1/00 - 8/31/05[9]. We assume the extreme tails are regularly varying with shape $P(x_t < \varepsilon) = \varepsilon^{-\alpha}L_1(\varepsilon)$ and $P(x_t > \varepsilon) = \varepsilon^{-\alpha}L_2(\varepsilon)$ where $L_i(\varepsilon)$ are slowly varying and $\alpha > 0$ denotes the tail index. Consult Bingham *et al* (1987). We study the tail shape of each series by computing B. Hill's (1975) tail estimator $\hat{\alpha}$. We apply asymptotic theory and a Newey-West-type kernel estimator of the asymptotic variance of $\hat{\alpha}$ for dependent, heterogenous processes, cf. Hill (2005a,b). Consult Hill (2005a) for a method for determining the sample tail fractile for computing the Hill estimator.

The STAR models are estimated by identity matrix weighted L_p-GMM for $p = 1.1$ and 1.5, 500 estimated decomposition innovations $\hat{Z}_t's$ are computed according to $\hat{Z}_t = [\hat{\epsilon}_t g_t(\hat{\gamma},\hat{c}) - \hat{b}] + [g_t(\hat{\gamma},\hat{c}) - \hat{a}] \times \sum_{j=1}^{n-1}\hat{\phi}^j\hat{\epsilon}_{t-j} \times \prod_{k=1}^{j} g_{t-k}(\hat{\gamma},\hat{c})$, and the sample IRF $\hat{I}(h, V_n, \mathfrak{L}_{n-1})$ is computed accordingly. We use Hong and White's (1995) non-parametric method to test the daily returns $\{x_t\}$ for evidence of $E[x_t|\Im_{t-1}] = 0$ and $P(x_t|\Im_{t-1}) = 0$ (i.e. $E[x_t^{<p-1>}|\Im_{t-1}] = 0$). We also test $E[\epsilon_t|\mathfrak{L}_{t-1}] = 0$ and $E[Z_t|\mathfrak{L}_{t-1}] = 0$ for evidence the STAR model represents the best L_2-predictor. We then test the STAR residuals $\{\hat{\epsilon}_t\}$ and the estimated decomposition innovations $\{\hat{Z}_t\}$ for evidence of weak and strong orthogonality by using $\{\hat{\epsilon}_t^{<p-1>}\}$ and $\{\hat{Z}_t^{<p-1>}\}$. For tests of strong orthogonality we use the series length $h = 10$ and 20 and randomly select $\pi \in \mathbb{R}^h$. For a non-parametric estimator of the conditional mean of $\{\hat{\epsilon}_t^{<p-1>}\}$ and $\{\hat{Z}_t^{<p-1>}\}$ we exploit Corollary 1 of Bierens (1990) which states the conditional mean of any Y_t, $E[Y_t^2] < \infty$, satisfies the Fourier series expansion $Y_{t-1}^* \equiv E[Y_t|\Im_{t-1}] = \theta_0 + \sum_{i=1}^{\infty}\theta_i \exp\{\Sigma_{j=1}^{\infty}\tau_j x_{t-j}\}$ with probability one for some sequence $\{\theta_i\}$. We use $\hat{Y}_{t-1}^* = \hat{\theta}_0 + \sum_{i=1}^{100}\hat{\theta}_i \exp\{\Sigma_{j=1}^{t-1}\tau_j x_{t-j}\}$, where the $\tau_j's$ and randomly selected

[9] The data were obtained from the New York Federal Reserve Bank statistical releases. Observations with missing values are removed (e.g. weekends, holidays), leaving a sample of 1424 daily returns. We filter each series through a standard daily dummy regression in order to control for day effects.

from $\mathbb{R}$. We repeat the Hong-White test for 100 randomly selected $\pi' s$ and 100 randomly selected $\tau' s$ (10,000 repetitions), and report the average p-value[10].

Tables 1 and 2 contain tail index estimates, L_p-GMM parameter estimates of the STAR models, and Hong-White test results. For brevity we omit all results concerning the case $p = 1.1$ because parameter estimates are uniformly insignificant for each variable in both exponential and logistic models.

Only the Yen provides unambiguous evidence for heavy-tails: for the Yen, Euro and Pound, the tail estimates and 95% interval widths are respectively $2.55 \pm .70$, 3.37 ± 1.20 and 2.96 ± 1.03. Because the theory of this paper holds for finite variance processes and L_p-metric projection, the question of whether the variance is infinite or not is not ultimately important. The question, then, is whether model (21) adequately represents the best L_p-predictor.

We find for each series significant evidence that the daily log return x_t has some form of unmodeled (non)linear structure: $E[x_t|\Im_{t-1}] \neq 0$ and $P(x_t|\Im_{t-1}) \neq 0$ when $p = 1.5$ (see Table 1). In either STAR model the L_p-GMM estimates uniformly suggest a large and significant scale γ for each exchange rate when $p = 1.5$, which implies regime change is quite fast. For the ESTAR model, the British Pound, in particular, is essentially in either one extreme regime state or the other. The LSTAR scale estimates are large enough that the implied switching dynamic is roughly identical to a SETAR: each exchange rate return is either noisy of persistent. The LSTAR estimated threshold is .020 and the largest absolute shock is .021. This provides concrete support of Hill's (2005a) finding that extremes in the daily return to the Pound are highly persistent and may be governed by an extremal switching process. This also points out the inappropriateness of the standard practice in the STAR literature of restricting estimates of c to lie between the lower 15% and upper 85% sample percentiles of the threshold variable (e.g. Teräsvirta, 1994; van Dijk *et al*, 2000).

A test that the ESTAR model represents the conditional expectations of the daily return (i.e. $E[\epsilon_t|\Im_{t-1}] = 0$) is rejected for the Euro at the 10%-level (see Test 1 of Table 2). Similarly, evidence suggests the LSTAR model only characterizes the conditional mean of the Yen.

Moreover, for the Yen and Pound the LSTAR model reasonably represents the best one-step ahead $L_{1.5}$-predictor based on tests of the hypothesis $P(\epsilon_{t+k}|\Im_{t-1}) = 0$ (see Test 2 in Table 2). The ESTAR model apparently only characterizes the best $L_{1.5}$-predictor of the Yen. Both ESTAR and LSTAR models generate weak orthogonal decomposition innovations as evidenced by Test 4 (i.e. $P(Z_{t+k}|\Im_{t-1}) = 0$). Evidence that the decomposition innovations

[10] Hong and White's (1995) test of $E[x_t|\Im_{t-1}] = 0$ requires $E[x_t^4] < \infty$ which likely fails to hold for each exchange rate return x_t (see Table 1). Similarly Bierens' (1990) finite variance assumptions may not hold for x_t, in particular for the Yen. Thus some caution should be taken when interpreting a test of this hypothesis. The same is true for tests of $E[\epsilon_t|\Im_{t-1}] = 0$ and $E[Z_t|\Im_{t-1}] = 0$. The test of $P(x_t|\Im_{t-1}) = 0$, however, requires $E[x_t^{<p-1>}]^4 < \infty$ which we assume holds for small enough $p < 2$. The smallest estimated tail index is $2.55 \pm .70$ (for the Yen), and $p < 2.55/4 + 1 = 1.6375$ is satisfied when $p = 1.1$ or 1.5. The lower bound of the 95% interval is 1.85, and $1.85/4 + 1 = 1.4875$, hence estimation and test results for the Yen when $p = 1.5$ should be interpreted with some caution. The preceding discourse identically applies to tests of weak and strong orthogonality of ϵ_t and Z_t.

are symmetrically orthogonal, however, is rather weak: we only find evidence supporting $E[Z_t|\Im_{t-1}] = 0$ for the Pound in the ESTAR case, and the Yen and Euro in the LSTAR case (see Test 3). It is the LSTAR model, in particular, that renders reasonably strong orthogonal decomposition innovations for each return series (see Test 6). The ESTAR model leads to reasonably strong orthogonal innovations for the Euro and Pound.

In summary, both STAR models may be the best L_2- and $L_{1.5}$-predictor of the daily return to the Yen. The decomposition innovations of the Yen are weak orthogonal but apparently not strong orthogonal in $L_{1.5}$ at the 10% level. The ESTAR model may represent the conditional mean of the Pound but not the best $L_{1.5}$-predictor, and the LSTAR may represent the best $L_{1.5}$-predictor but not the conditional expectation. The decomposition innovations from either smooth transition model suggest weak and strong orthogonality. For the Euro, evidence that the STAR models characterize the best predictor in either space is extremely weak. However, evidence suggests the decomposition innovations are weak and strong orthogonal. Such conflicting evidence may simply be due to model mis-specification evidenced by Tests 1 and 2, or due to an unbounded variance or kurtosis suggesting the Hong-White test of $E[\epsilon_t|\Im_{t-1}] = 0$ or $E[Z_t|\Im_{t-1}] = 0$ has a degenerate or non-standard limiting null distribution.

Coupled with evidence that the LSTAR model may be the best $L_{1.5}$- *and* L_2-predictor of the Yen, and weak evidence suggesting $E[Z_t|\Im_{t-1}] = 0$, implies the LSTAR model may be used to construct symmetrically *weak* orthogonal decompositions for the Yen.

Finally, when the LSTAR model is used notice that it is only the Pound for which evidence suggests both $E[\epsilon_t|\Im_{t-1}] \neq 0$ and $E[Z_t|\Im_{t-1}] \neq 0$, as well as $P(\epsilon_{t+k}|\Im_{t-1}) = 0$, $P(Z_{t+k}|\Im_{t-1}) = 0$ and $P(\sum_{k=0}^{h} \pi_k Z_{t+k}|\Im_{t-1}) = 0$. Thus, the LSTAR model reasonably articulates the best $L_{1.5}$-predictor of the Pound (but not best L_2-predictor), generating a strong orthogonal decomposition. Theorem 3 can be immediately used to justify iterated $L_{1.5}$-projections for the Pound based on an LSTAR best predictor, and Theorem 6 and Lemma 7 justify the existence of a nonlinear IRF based on best $L_{1.5}$-prediction. Finally, for the LSTAR model of the Pound notice the estimated threshold $\hat{c}$ is .02 and the largest shock in absolute value is .021: coupled with the large estimated scale, the LSTAR model suggests extremes of the Pound are persistent, non-extremes are noisy, and regime change occurs quickly, matching evidence found in Hill (2005a).

Figures 1 and 2 display nonlinear IRF's generated from ESTAR and LSTAR models based on two different methods for handling the random history $\{\epsilon_t\}_{t=1}^{n-1} \in \mathfrak{L}_{n-1}$ and random impulse $V_n \in \mathfrak{L}_n$. First, we generate four sequences of $\{\epsilon_t\}_{t=1}^{n}$ and put $v_n = \epsilon_n$: we set $\epsilon_t = .01$, $\forall t \neq n-i$, and fix $\epsilon_{n-i} = \hat{c}$, $i = 0$, 1, 2, or 3. Thus, we study impulse responses when x_{n-i+1} is assumed to lie in the "inner" (persistent) regime. Figure 1 plots IRF's for horizons $h = 0...15$ for each exchange rate based on the ESTAR model. The asymmetric impact impulses and histories have is clearly revealed. Nonetheless, persistence is visibly shallow: fixing the past impulse $v_{n-2} = \hat{c}$ or $v_{n-3} = \hat{c}$ results is nearly identical response paths.

Second, we use the estimated residuals $\{\hat{\epsilon}_t\}_{t=1}^{n-1}$ as one "draw" for the history, and randomly draw 500 iid impulses $\{v_{n,j}\}_{j=1}^{500}$ from a Pareto distribution, $P(v_{n,j} > v) = v^{-\alpha}$ and $P(v_{n,j} <$ -$v) = (-v)^{-\alpha}$, $v > 0$, where $\hat{\alpha}$ is used as a plug-in for α. Sequences of IRF's are then generated for horizons $h = 0$ and 1 (when $h \geq$ 2 nearly all of the probability mass of the IRF's occurs at zero). Figure 2 plots (Gaussian) kernel densities of the IRF's for each h for the British Pound based on the LSTAR model and an AR(1) model. A prominent characteristic is the extremely heavy-tailed nature of the nonlinear IRF empirical distribution: the estimated tail index of $\{\hat{I}(0, v_{n,j}, \{\hat{\epsilon}_t\}_{t=1}^{n-1})\}_{j=1}^{500}$ is .69 $\pm$ .15 due simply to the nonlinear multiplicative presence of the shock history $\{\hat{\epsilon}_t\}_{t=1}^{n-1}$. By comparison, the linear AR(1) IRF is $\phi^h \epsilon_n$ which is Pareto distributed with index $\alpha = 2.96$. The estimated tail index of the sequence of IRF's is 3.24 $\pm$ .30 (which contains 2.96).

Appendix

Proof of Lemma 2. *i.* Let $v \in V$ be the best predictor of $u \in U$, $P(u|V) = v$, hence for every vector element $\tilde{v} \in V$

$$(27) \quad \|u - v\| \leq \|u - \tilde{v}\|$$

by definition. Notice that for any real scalar λ the affine combination $v - \lambda\tilde{v}$ is also an element of V by the linearity of $\mathfrak{B}$. Thus, the element $v = P(u|V)$ satisfies

$$(28) \quad \|u - v\| \leq \|u - v + \lambda\tilde{v}\|$$

for every $\lambda \in \mathbb{R}$ and every $\tilde{v} \in V$. The above inequality denotes the condition of James orthogonality, cf. (5). This proves $P(u|V) = v$ *if and only if* $(u - v)\perp V$. The result $(u - v)\perp V$ *if and only if* $[u - v, \tilde{v}] = 0$ for every $\tilde{v} \in V$ and a unique inner-product $[,]$ follows directly from Lemma 1.

ii. From (*i*) we know $P(u|V) = v$ *if and only if* $(u - v)\perp V$ *if and only if* $[u - v, \tilde{v}] = 0$ for every $\tilde{v} \in V$. Simply consider the case when $v = 0$: $P(u|V) = 0$ *if and only if* $(u - 0)\perp V$ *if and only if* $[u, \tilde{v}] = 0$ for every $\tilde{v} \in V$.

iii. By definition the metric projection error $u - P(u|V)$ satisfies $||u - P(u|V)|| \leq \|u - \tilde{v}\|$ for every $\tilde{v} \in V$. Since $v \in V$, we deduce the best predictor of v, $P(v|V)$, satisfies

$$(29) \quad \|v - P(v|V)\| \leq \|v - v\| = 0,$$

which is true *if and only if* $v = P(v|V)$ with probability one.

iv. For any element $z \in V$ and any element $u \in U$ we need to show $P(u + z|V) = P(u|V) + z$. The projection $P(u + z|V)$ is the best predictor of $u + z$, by construction, hence

$$(30) \quad \|u + z - P(u + z|V)\| \leq \|u + z - \tilde{v}\|$$

for every $\tilde{v} \in V$. Because $z + P(u|V) \in V$, the above inequality holds for $\tilde{v} = z + P(u|V)$, hence

$$(31) \quad \|u + z - P(u + z|V)\| \leq \|u - P(u|V)\|.$$

However, the projection $P(u|V)$ is the best predictor of u in V, hence

$$(32) \quad \|u - P(u|V)\| \leq \|u + \tilde{v}\|$$

for every vector $\tilde{v} \in V$, and therefore for the element $\tilde{v} = z - P(u + z|V) \in V$, giving

$$(33) \quad \|u - P(u|V)\| \leq \|u + z - P(u + z|V)\|.$$

Combining (31) and (33), we obtain

$$(34) \quad \|u - P(u|V)\| \leq \|u + z - P(u + z|V)\| \leq \|u - P(u|V)\|,$$

hence equality holds

$$(35)\quad \|u+z-P(u\ +\ z|V)\| = \|u-P(u|V)\| = \|u+z-P(u|V)-z\|\,.$$

Using (31) and (35), we deduce

$$(36)\quad \|u+z-P(u\ +\ z|V)\| = \|u+z-P(u|V)-z\| \le \|u+z-\tilde{v}\|\,,$$

which is true *if and only if* the element $P(u|V)+z \in V$ is identically the best predictor of $u+z$, which proves the result.

v. Define $v = P(u|V)$. First, notice that the projection error $u-v$ satisfies

$$(37)\quad \begin{aligned}\|u-v\|^2 &= [u-v, u-v]\\ &= [u-v,u]-[u-v,v] = [u-v,u]\end{aligned}$$

due to $v \in V$ and orthogonality $(u-v)\perp V$, by result (i). By Hölder's inequality, we deduce

$$(38)\quad \|u-v\|^2 = [u-v,u] \le \|u-v\|\,\|u\|\,,$$

and because $\|u-v\| > 0$ by assumption, we conclude $||u-v|| \le \|u\| < \infty$.

Similarly, by Minkowski's inequality and inequality (38), we deduce

$$(39)\quad \begin{aligned}||v|| &= ||u-v+u\|\\ &\le ||u-v||+||u\| \le \|u\|+||u\| = 2||u||,\end{aligned}$$

which proves the result with $k=2$.

vi. Let $U,V \subseteq L_p$, $1<p\le 2$, and assume $E[u|V] \in V$. For the following, we employ the result that James orthogonality equivalently implies L_p-orthogonality, $x\perp y$ *if and only if* $E[x^{<p-1>}y]=0$: see, e.g., Samorodnitsky and Taqqu (1994). From result (i) we know v is the best predictor of u, $P(u|V)$, *if and only if* $(u-v)\perp V$. First, assume the best predictor v satisfies $V\perp(u-v)$. Then for any $\tilde{v} \in V$, by the definition of orthogonality $V\perp(u-v)$ we deduce

$$(40)\quad \begin{aligned}0 &= E\left[\tilde{v}^{<p-1>}\,(u-v)\right]\\ &= E\left[\tilde{v}^{<p-1>}\,(u-P(u|V))\right]\\ &= E\left(E\left[\tilde{v}^{<p-1>}\,(u-P(u|V))\,|V\right]\right)\\ &= E\left(\tilde{v}^{<p-1>}E\left[(u-P(u|V))\,|V\right]\right)\\ &= E\left(\tilde{v}^{<p-1>}\,(E[u|V]-P(u|V))\right).\end{aligned}$$

Because equality (40) holds for every $\tilde{v} \in V$, and because $E[u|V] \in V$ and $P(u|V) \in V$, it follows that $E[u|V]-P(u|V) \in V$ by the convexity of L_p. Hence, equality (40) holds for $\tilde{v} = E[u|V]-P(u|V) \in V$,

$$(41)\quad \begin{aligned}0 &= E\Big((E[u|V]-P(u|V))^{<p-1>}\,(E[u|V]-P(u|V))\Big)\\ &= E\,|E[u|V]-P(u|V)|^p\,,\end{aligned}$$

which holds *if and only if* $E[u|V] - P(u|V) = 0$ with probability one. Because equality holds for each i, we conclude $E[u|V] = P(u|V)$ with probability one.

Conversely, assume $E[u|V] = P(u|V) = v$ (in which case $E[u|V] = P(u|V) \in V$ automatically holds). Then for any vector $\tilde{v} \in V$,

$$
\begin{aligned}
(42)\quad E\left[\tilde{v}^{<p-1>}(u-v)\right] &= E\left[\tilde{v}^{<p-1>}(u-E[u|V])\right] \\
&= E\left[E\left(\tilde{v}^{<p-1>}(u-E[u|V])\right)|V\right] \\
&= E\left[\tilde{v}^{<p-1>}(E[u|V]-E[u|V])\right] \\
&= 0,
\end{aligned}
$$

hence $V \perp (u-v)$.

vii. By result (i), if v is the best predictor of u then for any real scalar $a \in \mathbb{R}$ we know $P(au|V) = aP(u|V) = av$ *if and only if* $(au - av) \perp V$. By homogeneity of the semi-inner product, for any $\tilde{v} \in V$

$$
\begin{aligned}
(43)\quad [au-av,\tilde{v}] &= [a(u-v),\tilde{v}] \\
&= a[u-v,\tilde{v}] = 0
\end{aligned}
$$

therefore $au - av \perp V$. By result (i) and the uniqueness of metric projections in $\mathfrak{B}$, we conclude av is the best predictor of au, hence $P(au|V) = aP(u|V)$.

viii. Consider subspaces $V_0 \subseteq V_1 \subseteq \mathfrak{B}$, and assume $P: U \to V_1$ is a linear operator on V_1. By (i) we know the error $u - P(u|V_1)$ is orthogonal to V_1, therefore $u - P(u|V_1)$ is orthogonal to any subspace of V_1, say V_0. By (ii) we therefore obtain $P(u - P(u|V_1)|V_0) = 0$. Because we assume P is a linear operator on V_1, we deduce

$$
\begin{aligned}
0 &= P(u-P(u|V_1)|V_0) \\
(44)\qquad &= P(u|V_0) - P(P(u|V_1)|V_0) \\
P(u|V_0) &= P(P(u|V_1)|V_0),
\end{aligned}
$$

hence iterated projections holds. ■

Proof of Theorem 3. Consider the claim $\mathfrak{B}_n = \sum_{i=0}^{\infty} N_{n-i} + \mathfrak{B}_{-\infty}$. By Lemma A.1, below, for arbitrary integer $k > 0$ the finite decomposition holds,

$$
(45)\quad \mathfrak{B}_n = \left(\sum_{i=0}^{k-1} N_{n-i}\right) + \mathfrak{B}_{n-k}, \quad N_t \equiv \mathfrak{B}_t - P_{t,t-1}\mathfrak{B}_t.
$$

Consider an arbitrary element $X_n \in \mathfrak{B}_n$, and define the sequences $Y_{n,k} \in \sum_{i=0}^{k-1} N_{n-i}$ and $V_{n,k} \in \mathfrak{B}_{n-k}$ and such that

$$
(46)\quad X_n = Y_{n,k} + V_{n,k}.
$$

By orthogonality $V_{n,k} \perp \sum_{i=0}^{k-1} N_{n-i}$, norm-boundedness $\|X_n\| < \infty$, and the triangular inequality, for any $k \geq 1$

$$
\begin{aligned}
(47)\quad \|V_{n,k}\| &\leq \|V_{n,k} + Y_{n,k}\| = \|X_n\| < \infty \\
\|Y_{n,k}\| &= \|X_n - V_{n,k}\| \leq 2\|X_n\| < \infty.
\end{aligned}
$$

Because the sequences $Y_{n,k}$ and $V_{n,k}$, are norm bounded in a reflexive Banach space, $\mathfrak{B}_n$, they have simultaneously weakly convergent subsequences, say $\{Y_{n,k_i}\}$ and $\{V_{n,k_i}\}$, cf. the Bolzano-Weierstrass Theorem. In particular, define the stochastic limits as

(48) $\lim_{k_i\to\infty} Y_{n,k_i} = Y_n \qquad \lim_{k_i\to\infty} V_{n,k_i} = V_n.$

Now, because the equality $X_n = Y_{n,k} + V_{n,k}$ holds for any integer $k > 0$ and the sequences $Y_{n,k_i} \in \sum_{i=0}^{k_i-1} N_{n-i}$ and $V_{n,k_i} \in \mathfrak{B}_{n-k_i}$ converge, we deduce by continuity for arbitrary $X_n \in \mathfrak{B}_n$, $X_n = Y_n + V_n$, where clearly $Y_n \in \sum_{i=0}^{\infty} N_{n-i}$ and $V_n \in \mathfrak{B}_{-\infty}$. Because $X_n \in \mathfrak{B}_n$ is arbitrary, (8) is proved.

The proof of $(i) \Longleftrightarrow (ii)$ follows in a manner identical to the line of proof of Lemma A.1, below.

Finally, the claim that every element $Y \in \sum_{i=0}^{\infty} \oplus N_{n-i}$ obtains a unique norm-convergent expansion, $Y = \sum_{i=0}^{\infty} \xi_{n-i}$, $\xi_t \in N_t$, follows from a direct application of Lemma A.2, below. ■

Lemma A.1 *For any Banach space B_n, there exists a sequence of subspaces $\{N_{n-i}\}_{i=0}^{k-1}$, $N_t \subseteq B_t$, such that $B_n = \sum_{i=0}^{k-1} N_{n-i} + B_{n-k}$ where the spaces N_t are weak orthogonal in the sense that $N_{n-i} \perp B_{n-i-1}$ and $N_{n-i} \perp N_{n-j}$ for every $0 \le i < j \le k-1$. Moreover, the following are equivalent for any integer $k > 0$: i. $B_n = \left(\sum_{i=0}^{k-1} \oplus N_{n-i}\right) \oplus B_{n-k}$; and ii. $P_{t,t-l}P_{t,t-k} = P_{t,t-l}$, for every t, $k \le l$.*

Lemma A.2 *Consider a sequence of orthogonal subspaces $\{M_{n-i}\}_{i=0}^{\infty}$, $M_{t-1} \subseteq M_t \subseteq B_t$, such that $\sum_{i=0}^{\infty} \oplus M_{n-i}$ exists, and consider a sequence of elements $\{x_j\}$, $x_j \in M_j$. The space $\{x_t : x_t \in M_t\}$ forms a Schauder basis for its closed linear span. Consequently, every element $X \in \sum_{i=0}^{\infty} \oplus M_{n-i}$ obtains a unique norm convergent expansion $X = \sum_{i=0}^{\infty} a_i x_{n-i}$, $x_t \in M_t$, for some sequence of real constants $\{a_t\}$.*

Proof of Lemma A.1. Define the sequence $\{N_t\}$, $N_t \equiv \mathfrak{B}_t - P_{t,t-1}\mathfrak{B}_t$, where $N_t \perp \mathfrak{B}_{t-1}$, cf. Lemma 2.$i$, and $P_{t,t-1}\mathfrak{B}_t = \mathfrak{B}_{t-1}$ due to $\mathfrak{B}_{t-1} \subset \mathfrak{B}_t$. Because Banach spaces are linear and $P_{t,t-1}\mathfrak{B}_t = \mathfrak{B}_{t-1} \subset \mathfrak{B}_t$, we deduce $N_t \subseteq \mathfrak{B}_t$. We obtain the tautological expression

(49) $\mathfrak{B}_n = N_n + P_{n,n-1}\mathfrak{B}_n.$

Recursively decomposing $\mathfrak{B}_{n-1}$, etc., it follows that for arbitrary $k \ge 1$

$$
\begin{aligned}
(50)\quad \mathfrak{B}_n &= N_n + P_{n,n-1}\mathfrak{B}_n = N_n + \mathfrak{B}_{n-1} \\
&= N_n + N_{n-1} + \mathfrak{B}_{n-2} = \ldots \\
&= \sum\nolimits_{i=0}^{k-1} N_{n-i} + \mathfrak{B}_{n-k},
\end{aligned}
$$

where for each $t \le n$, $N_t \perp \mathfrak{B}_{t-1}$. Observe that given the orthogonality property $N_{n-i} \perp \mathfrak{B}_{n-i-1}$ and $N_{n-j} \subseteq \mathfrak{B}_{n-j} \subseteq \mathfrak{B}_{n-i}$ for every $0 \le i < j \le k$, it follows that $N_{n-i} \perp N_{n-j}$, $0 \le i < j \le k$.

Assume (i) holds. Then for any $k \geq 1$, $\mathfrak{B}_n = \left(\sum_{i=0}^{k-1} \oplus N_{n-i}\right) \oplus \mathfrak{B}_{n-k}$, hence we may write for any $X_n \in \mathfrak{B}_n$,

$$(51)\quad X_n = \sum_{i=0}^{k-1} \xi_{n-i} + V_{n,k},$$

where $\xi_i \in N_i$ and $V_{n,k} \in \mathfrak{B}_{n-k}$. Thus, by quasi-linearity, cf. Lemma 2.iv, we deduce for any $0 \leq t \leq k$,

$$\begin{aligned}(52)\quad P_{n,n-t}X_n &= P_{n,n-t}\left(\sum_{i=0}^{k-1} \xi_{n-i} + V_{n,k}\right) \\ &= P_{n,n-t}\left(\sum_{i=0}^{t-1} \xi_{n-i}\right) + \sum_{i=t}^{k-1} \xi_{n-i} + V_{n,k} \\ &= \sum_{i=t}^{k-1} \xi_{n-i} + V_{n,k},\end{aligned}$$

where $P_{n,n-t}\left(\sum_{i=0}^{t-1} \xi_{n-i}\right) = 0$, cf. Lemma 2.$ii$, due to $\sum_{i=0}^{t-1} \xi_{n-i} \in \sum_{i=0}^{t-1} \oplus N_{n-i}$ and $\sum_{i=0}^{t-1} \oplus N_{n-i} \perp \mathfrak{B}_{n-t}$ by assumption. Similarly, for any $1 \leq s \leq t \leq k$,

$$\begin{aligned}(53)\quad P_{n,n-t}P_{n,n-s}X_n &= P_{n,n-t}P_{n,n-s}\left(\sum_{i=0}^{k-1} \xi_{n-i} + V_{n,k}\right) \\ &= P_{n,n-t}\left(P_{n,n-s}\sum_{i=0}^{s-1} \xi_{n-i} + \sum_{i=s}^{k-1} \xi_{n-i} + V_{n,k}\right) \\ &= P_{n,n-t}\left(\sum_{i=s}^{k-1} \xi_{n-i} + V_{n,k}\right) \\ &= P_{n,n-t}\left(\sum_{i=s}^{t-1} \xi_{n-i}\right) + \sum_{i=t}^{k-1} \xi_{n-i} + V_{n,k} \\ &= \sum_{i=t}^{k-1} \xi_{n-i} + V_{n,k}.\end{aligned}$$

This proves $P_{n,n-t}X_n = P_{n,n-t}P_{n,n-s}X_n$ for arbitrary X_n in $\mathfrak{B}_n$ and any s, t such that $0 \leq s \leq t \leq k$. Because $X_n \in \mathfrak{B}_n$ is arbitrary, we deduce the operators satisfy $P_{n,n-t}P_{n,n-s} = P_{n,n-t}$ for any $0 \leq s \leq t \leq k$, hence $(i) \Longrightarrow (ii)$.

Next, assume (ii) holds. It suffices to prove for any $k > 0$ the subspaces $\{N_{n-i}\}_{i=0}^{k-1}$ and $\mathfrak{B}_{n-k}$ are strong orthogonal such that for every $1 \leq j < k$

$$(54)\quad \sum_{i=0}^{j-1} N_{n-i} \perp \sum_{i=j}^{k-1} N_{n-i}, \quad \sum_{i=0}^{k-1} N_{n-i} \perp \mathfrak{B}_{n-k}.$$

Consider any element $\sum_{i=0}^{k-1} \xi_{n-i} \in \sum_{i=0}^{k-1} N_{n-i}$. By iterated projections $P_{n,n-t}P_{n,n-s}$ $= P_{n,n-t}$ for arbitrary $0 \leq s \leq t \leq k$,

$$\begin{aligned}(55)\quad P_{n,n-k}\left(\sum_{i=0}^{k-1} \xi_{n-i}\right) &= P_{n,n-k}P_{n,n-1}\left(\sum_{i=0}^{k-1} \xi_{n-i}\right) \\ &= P_{n,n-k}\left[P_{n,n-1}\xi_n + \sum_{i=1}^{k-1} \xi_{n-i}\right] \\ &= P_{n,n-k}\left(\sum_{i=1}^{k-1} \xi_{n-i}\right),\end{aligned}$$

where $P_{n,n-1}\xi_n = 0$, cf. Lemma 2.*ii*, due to $\xi_n \in N_n$ and $N_n \perp \mathfrak{B}_{n-1}$ by construction. Proceeding with subsequent $P_{n,n-h}$, $h = 1...k$, we obtain

$$(56)\quad \begin{aligned} P_{n,n-k}\left(\sum_{i=0}^{k-1}\xi_{n-i}\right) &= P_{n,n-k}\left(\sum_{i=1}^{k-1}\xi_{n-i}\right) \\ &= P_{n,n-k}\left[P_{\mathfrak{B}\,n-2}\left(\sum_{i=1}^{k-1}\xi_{n-i}\right)\right] \\ &= P_{n,n-k}\left[P_{\mathfrak{B}\,n-2}\xi_{n-1} + \sum_{i=2}^{k-1}\xi_{n-i}\right] \\ &= P_{n,n-k}\left(\sum_{i=2}^{k-1}\xi_{n-i}\right) = ... = 0. \end{aligned}$$

Therefore, $\sum_{i=0}^{k-1}\xi_{n-i}\perp\mathfrak{B}_{n-k}$, cf. Lemma 2.*ii*, for any integer $k > 0$ and any elements $\xi_t \in N_t$. Because the elements $\xi_t \in N_t$ are arbitrary, we deduce for every $k > 0$, $\sum_{i=0}^{k-1} N_{n-i}\perp\mathfrak{B}_{n-k}$.

Finally, because $\sum_{i=j}^{k-1} N_{n-i}$ is a subspace of $\mathfrak{B}_{n-j}$ for any $1 \le j \le k$, we conclude $\sum_{i=0}^{j-1} N_{n-i}\perp\sum_{i=j}^{k-1} N_{n-i}$. It follows that $\sum_{i=0}^{k-1} N_{n-i} = \sum_{i=0}^{k-1}\oplus N_{n-i}$, and $\mathfrak{B}_n = \left(\sum_{i=0}^{k-1}\oplus N_{n-i}\right)\oplus\mathfrak{B}_{n-k}$, which proves $(ii) \Longrightarrow (i)$. ■

Proof of Lemma A.2. Consider an arbitrary sequence of elements $\{x_j\}$, $x_j \in M_j$, and recall $\sum_{i=0}^{\infty}\oplus M_{n-i}$ exists. A necessary and sufficient condition for a sequence of Banach space elements $\{x_j\}$ to form a Schauder basis (hereafter referred to as a *basis*) is the existence of some scalar constant $0 < K < \infty$ such that for all scalar real-valued sequences $\{\lambda_j\}$ and integers $s \le t$,

$$(57)\quad \left\|\sum_{i=0}^{s}\lambda_j x_{n-j}\right\| \le K\left\|\sum_{i=0}^{t}\lambda_j x_{n-j}\right\|.$$

See, e.g., Proposition 4.1.24 of Megginson (1998); see also Singer (1970). In our case, by the existence of the space $\sum_{i=0}^{\infty}\oplus M_{n-i}$, the subspaces M_j are strong orthogonal by construction, for any $s < t$ it follows that

$$(58)\quad \sum_{i=0}^{s}\oplus M_{n-i}\perp\sum_{i=s+1}^{t}\oplus M_{n-i}.$$

Synonymously, for all scalar real-valued sequences $\{\lambda_j\}$ and components $x_j \in M_j$

$$(59)\quad \sum_{i=0}^{s}\lambda_j x_{n-j}\perp\sum_{i=s+1}^{t}\lambda_j x_{n-j}.$$

By the definition of James orthogonality, it follows from (59) that

$$(60)\quad \left\|\sum_{i=0}^{s}\lambda_j x_{n-j} + a\sum_{i=s+1}^{t}\lambda_j x_{n-j}\right\| \ge \left\|\sum_{i=0}^{s}\lambda_j x_{n-j}\right\|$$

for all real scalars a. For $a = 1$, we conclude that $\forall s < t$,

$$(61)\quad \begin{aligned} \left\|\sum_{i=0}^{t}\lambda_j x_{n-j}\right\| &= \left\|\sum_{i=0}^{s}\lambda_j x_{n-j} + \sum_{i=s+1}^{t}\lambda_j x_{n-j}\right\| \\ &\ge \left\|\sum_{i=0}^{s}\lambda_j x_{n-j}\right\|. \end{aligned}$$

The equality in (57) follows with $K = 1$, which proves the result. ■

Proof of Corollary 4. *i.* The result is an immediate consequence of the definition of weak-orthogonality, and Theorem 3. In particular, for real-valued $\psi_{n,i}$, simply define the components $\psi_{n,i}Z_{n-i} \equiv \xi_{n-i}$, $Z_{n-i} \neq 0$, where $\xi_t \in N_t = \mathfrak{B}_t - P_{t,t-1}\mathfrak{B}_t$ are defined in Theorem 3. The strong-orthogonality result follows from Theorem 3.

ii. Recall the decomposition

$$(62) \quad X_n = \sum\nolimits_{i=0}^{\infty} \psi_{n,i} Z_{n-i} + V_n,$$

and recall the orthogonality properties

$$(63) \quad Z_t \perp \mathfrak{B}_{t-1}$$

$$(64) \quad Z_t \perp Z_s, \ \forall s < t.$$

Consider the inner-product between Z_n and X_n. Using decomposition (61) and linearity in the second argument of the semi-inner product [,], we obtain

$$\begin{aligned}(65) \quad [Z_n, X_n] &= \left[Z_n, \sum\nolimits_{i=0}^{\infty} \psi_{n,i} Z_{n-i} + V_n\right] \\ &= \sum\nolimits_{i=0}^{\infty} \left[Z_n, \psi_{n,i} Z_{n-i}\right] + [Z_n, V_n] \\ &= \psi_{n,i} \sum\nolimits_{i=0}^{\infty} [Z_n, Z_{n-i}] + [Z_n, V_n] \\ &= \psi_{n,0} [Z_n, Z_n],\end{aligned}$$

due to $[Z_n, Z_{n-i}] = 0 \ \forall i > 0$, cf. (64) and Lemma 2.*i*; and $[Z_n, V_n] = 0$ due to $V_n \in \mathfrak{B}_{-\infty} \subseteq \mathfrak{B}_{n-1}$ and (63). Because $Z_{n-i} \in \mathfrak{B}_{n-i} - P_{n-i,n-i-1}\mathfrak{B}_{n-i}$, and the spaces $\mathfrak{B}_t$ contain only non-deterministic components, we can always choose $Z_{n-i} \in N_{n-i}$ such that $\psi_{n,i}Z_{n-i} = 0$ *if and only if* $\psi_{n,i} = 0$, and deduce $[Z_n, Z_n] = \|Z_n\|^2 > 0$. Hence, we conclude the first coefficient is uniquely determined as

$$(66) \quad \psi_{n,0} = \frac{[Z_n, X_n]}{[Z_n, Z_n]} = \frac{[Z_n, X_n - P_{n,n-1}X_n]}{[Z_n, Z_n]} = \frac{[Z_n, Z_n]}{[Z_n, Z_n]} = 1,$$

where $[Z_n, X_n] = [Z_n, X_n - P_{n,n-1}X_n]$ follows from linearity in the second argument, and orthogonality: $Z_n \perp \mathfrak{B}_{n-1}$ and $P_{n,n-1}X_n \in \mathfrak{B}_{n-1}$.

Similarly, the inner-product between Z_{n-1} and X_n reduces to

$$\begin{aligned}(67) \quad [Z_{n-1}, X_n] &= \psi_{n,i} \sum\nolimits_{i=0}^{\infty} [Z_{n-1}, Z_{n-i}] + [Z_n, V_n] \\ &= \psi_{n,0} [Z_{n-1}, Z_n] + \psi_{n,1} [Z_{n-1}, Z_{n-1}],\end{aligned}$$

Observe that we cannot in general deduce

$$(68) \quad [Z_{n-1}, Z_n] = 0,$$

because the orthogonality condition $\perp$ is in general asymmetric. Using (67) the second coefficient is uniquely determined as

$$(69) \quad \psi_{n,1} = \frac{[Z_{n-1}, X_n] - [Z_{n-1}, Z_n]}{[Z_{n-1}, Z_n]}$$

Proceeding recursively, we deduce for $i = 1, 2, \ldots$

$$(70)\quad \psi_{n,0} = 1, \quad \psi_{n,i} = \frac{[Z_{n-i}, X_n]}{[Z_{n-i}, Z_n]} - \sum_{j=0}^{i-1} \psi_{n,j} \frac{[Z_{n-j}, Z_n]}{[Z_{n-i}, Z_n]}.$$

■

Proof of Theorem 5. We will prove in order $(ii) \iff (iii)$, $(i) \iff (ii)$, and $(iii) \iff (iv)$. First consider $(ii) \iff (iii)$, and let (ii) hold such that $\mathfrak{L}_{t-1} \perp N_t$. Because $N_t = \mathfrak{L}_t - P_{t,t-1}\mathfrak{L}_t$, by Lemma 2.$i$ it follows that for any $X_t \in \mathfrak{L}_t$ and any $Y_{t-1} \in \mathfrak{L}_{t-1}$

$$(71)\quad EY_{t-1}^{<p-1>}(X_t - P_{t,t-1}X_t) = 0.$$

Denoting by E_{t-1} expectations with respect to the σ-field $\mathfrak{F}_{t-1}$, it follows that

$$(72)\quad E_{t-1}Y_{t-1}^{<p-1>}(E[X_t|\mathfrak{F}_{t-1}] - P_{t,t-1}X_t) = 0,$$

for any $Y_{t-1} \in \mathfrak{L}_{t-1}$. Because $E(X_t|\mathfrak{F}_{t-1}) - P_{t,t-1}X_t \in \mathfrak{L}_{t-1}$, it must be the case that

$$(73)\quad E\left|E[X_t|\mathfrak{F}_{t-1}] - P_{t,t-1}X_t\right|^p = 0,$$

or $E(X_t|\mathfrak{F}_{t-1}) = P_{t,t-1}X_t$. Thus $(ii) \Longrightarrow (iii)$. Next, let (iii) hold, and observe that $\mathfrak{L}_{t-1} \perp N_t$ follows immediately because for any $X_t \in \mathfrak{L}_t$ and any $Y_{t-1} \in \mathfrak{L}_{t-1}$,

$$\begin{aligned}(74)\quad & EY_{t-1}^{<p-1>}(X_t - P_{t,t-1}X_t) \\ = \; & EY_{t-1}^{<p-1>}(X_t - E(X_t|\mathfrak{F}_{t-1})) \\ = \; & E_{t-1}Y_{t-1}^{<p-1>}(E[X_t|\mathfrak{F}_{t-1}] - E[X_t|\mathfrak{F}_{t-1}]) = 0.\end{aligned}$$

Hence, $(iii) \Longrightarrow (ii)$, and therefore $(ii) \iff (iii)$.

Now consider $(i) \iff (ii)$. We will prove $(i) \Longrightarrow (ii)$ and $(ii) \Longrightarrow (iii) \Longrightarrow (i)$. First, let (i) hold, and observe that symmetric orthogonality $\left(\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} N_{n-i}\right) \overleftrightarrow{\oplus} \mathfrak{B}_{-\infty}$ implies for any $t \leq n$,

$$\begin{aligned}(75)\quad & N_t \perp \left(\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} N_{t-1-i}\right) \overleftrightarrow{\oplus} \mathfrak{L}_{-\infty} \\ & \left(\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} N_{t-1-i}\right) \overleftrightarrow{\oplus} \mathfrak{L}_{-\infty} \perp N_t.\end{aligned}$$

But, by definition of a symmetrically orthogonal decomposition, $\mathfrak{L}_{t-1}$ is exactly decomposed as

$$(76)\quad \mathfrak{L}_{t-1} = \left(\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} N_{t-1-i}\right) \overleftrightarrow{\oplus} \mathfrak{L}_{-\infty}.$$

Thus (75) and (76) imply

$$(77)\quad \left(\sum_{i=0}^{\infty} \overleftrightarrow{\oplus} N_{t-1-i}\right) \overleftrightarrow{\oplus} \mathfrak{L}_{-\infty} \perp N_t \Longrightarrow \mathfrak{L}_{t-1} \perp N_t,$$

giving $(i) \Longrightarrow (ii)$.

Nest, let (ii) hold such that $\mathfrak{L}_{t-1} \perp N_t$. Then $(ii) \Longrightarrow (iii)$ follows from above. Consequently, the best predictor coincides with the conditional expectations, $E(X_t|\mathfrak{F}_{t-1}) = P_{t,t-1}X_t$. By the property of iterated expectations, in this case iterated projections must hold, and therefore a strong orthogonal decomposition exists, cf. Theorem 3,

$$(78) \quad \mathfrak{L}_n = \left(\sum\nolimits_{i=0}^{\infty} \oplus N_{n-i}\right) \oplus \mathfrak{L}_{-\infty},$$

Now, observe that the construction $N_t = \mathfrak{L}_t - P_{t,t-1}\mathfrak{L}_t$ implies $N_t \perp \mathfrak{L}_{t-1}$, cf. Lemma 2.$i$, and the strong orthogonal decomposition implies

$$(79) \quad \sum\nolimits_{i=0}^{l-1} N_{n-i} \perp \sum\nolimits_{i=l}^{n-1} N_{n-i} \text{ and } \sum\nolimits_{i=0}^{\infty} N_{n-i} \perp \mathfrak{L}_{-\infty},$$

for all $1 \leq l < n$. Therefore, under the maintained assumptions, in order to prove the decomposition is symmetrically orthogonal, it suffices to prove for any $1 \leq l < n$,

$$(80) \quad \sum\nolimits_{i=l}^{n-1} N_{n-i} \perp \sum\nolimits_{i=0}^{l-1} N_{n-i} \text{ and } \mathfrak{L}_{-\infty} \perp \sum\nolimits_{i=0}^{\infty} N_{n-i}.$$

Because $\mathfrak{L}_{-\infty} \subseteq \mathfrak{L}_{t-1}$, the condition $\mathfrak{L}_{t-1} \perp N_t$ implies $\mathfrak{L}_{-\infty} \perp N_t$, $\forall t \leq n$. Hence, by linearity in the second argument of the $L_p(\ \ ,\mathfrak{F}_t,\mu)$ semi-inner product, for any $T \geq 0$,

$$(81) \quad \mathfrak{L}_{-\infty} \perp \sum\nolimits_{i=0}^{T} N_{n-i}.$$

A simple continuity argument suffices for $\mathfrak{L}_{-\infty} \perp \sum_{i=0}^{\infty} N_{n-i}$.

Moreover, for all $1 \leq l < n$ and $i = 0...l-1$, the properties $\mathfrak{L}_{n-l} \subseteq \mathfrak{L}_{n-i-1}$ and $\mathfrak{L}_{n-i-1} \perp N_{n-i}$ imply $\mathfrak{L}_{n-l} \perp N_{n-i}$. So, because $\sum_{i=l}^{n-1} N_{n-i} \subseteq \mathfrak{L}_{n-l}$, we deduce for each $i = 0...l-1$,

$$(82) \quad \sum\nolimits_{i=l}^{n-1} N_{n-i} \perp N_{n-i}.$$

Again, by linearity in the second argument of the $L_p(\ \ ,\mathfrak{F}_t,\mu)$ semi-inner product, we conclude

$$(83) \quad \sum\nolimits_{i=l}^{n-1} N_{n-i} \perp \sum\nolimits_{i=0}^{l-1} N_{n-i}.$$

Thus, $(ii) \Longrightarrow (i)$.

Finally, consider the claim $(iii) \Longleftrightarrow (iv)$. The conditional expectations coincides with the best predictor *if and only if*

$$(84) \quad E\left(X_t - E[X_t|\mathfrak{F}_{t-1}]\right)^{<p-1>} Y_{t-1} = 0,$$

for all such elements $Y_{t-1} \in \mathfrak{L}_{t-1}$, cf. Lemma 2.$i$. This is synonymous to

$$(85) \quad \begin{aligned} E\left(X_t - E[X_t|\mathfrak{F}_{t-1}]\right)^{<p-1>} Y_{t-1} &= 0 \\ E_{t-1}E\left[\left(X_t - E[X_t|\mathfrak{F}_{t-1}]\right)^{<p-1>} Y_{t-1}|\mathfrak{F}_{t-1}\right] &= 0 \\ E_{t-1}Y_{t-1}E\left[\left(X_t - E[X_t|\mathfrak{F}_{t-1}]\right)^{<p-1>} |\mathfrak{F}_{t-1}\right] &= 0 \end{aligned}$$

for all elements $Y_{t-1} \in \mathfrak{L}_{t-1}$. It follows that the zero identity holds for the element

$$(86)\quad Y_{t-1} = \left(E\left[(X_t - E[X_t|\mathfrak{F}_{t-1}])^{<p-1>}|\mathfrak{F}_{t-1}\right]\right)^{<p-1>} \in \mathfrak{L}_{t-1}.$$

Together, (85) and (86) imply

$$(87)\quad E_{t-1}\left|E\left[(X_t - E[X_t|\mathfrak{F}_{t-1}])^{<p-1>}|\mathfrak{F}_{t-1}\right]\right|^p = 0,$$

or $E[(X_t - E[X_t|\mathfrak{F}_{t-1}])^{<p-1>}|\mathfrak{F}_{t-1}] = 0$, $\mathfrak{F}_{t-1}$-$a.e.$ ■

Proof of Theorem 6. By assumption $x_t = \sum_{i=0}^{\infty}\psi_{t,i}Z_{t,t-i}$ where the $Z_{t,t-i}^{\prime}s$ are strong orthogonal to $\mathfrak{L}_{t-i-1}$, and the projection operator iterates from $\mathfrak{L}_{t-k}$ to $\mathfrak{L}_{t-k-1} \oplus v_{t-k}$ for any t and $k \geq 0$. Project x_t separately onto $\mathfrak{L}_{t-1}$ and $\tilde{\mathfrak{L}}_{t-1}$: using quasi-linearity and homogeneity, and observing $\mathfrak{L}_t \subseteq \tilde{\mathfrak{L}}_t$, $P\left(x_t|\mathfrak{L}_{t-1}\right) = \sum_{i=1}^{\infty}\psi_{t,i}Z_{t,t-i}$ and

$$(88)\quad P(x_t|\tilde{\mathfrak{L}}_{t-1}) = \sum\nolimits_{i=1}^{\infty}\psi_{t,i}Z_{t,t-i} + \psi_{t,0}P(Z_{t,t}|\tilde{\mathfrak{L}}_{t-1})$$

hence the 0-step ahead nonlinear impulse response $I(0, v_t, \mathfrak{L}_{t-1})$ at time t is simply

$$(89)\quad I(0, v_t, \mathfrak{L}_{t-1}) = \psi_{t,0}P(Z_{t,t}|\tilde{\mathfrak{L}}_{t-1}).$$

For the 1-step ahead impulse response $I(1, v_t, \mathfrak{L}_{t-1})$, by strong orthogonality and quasi-linearity

$$\begin{aligned}(90)\qquad & I(1, v_t, \mathfrak{L}_{t-1}) \\ &= P(x_{t+1}|\tilde{\mathfrak{L}}_{t-1}) - P\left(x_{t+1}|\mathfrak{L}_{t-1}\right) \\ &= \sum\nolimits_{i=2}^{\infty}\psi_{t+1,i}Z_{t+1,t+1-i} + P\left(Z_{t+1,t+1} + \psi_{t+1,1}Z_{t+1,t}|\tilde{\mathfrak{L}}_{t-1}\right) \\ &- \sum\nolimits_{i=2}^{\infty}\psi_{t+1,i}Z_{t+1,t-i} \\ &= P\left(Z_{t+1,t+1} + \psi_{t+1,1}Z_{t+1,t}|\tilde{\mathfrak{L}}_{t-1}\right).\end{aligned}$$

By iterated projections, quasi-linearity, homogeneity, and orthogonality

$$\begin{aligned}(91)\qquad & P\left(Z_{t+1,t+1} + \psi_{t+1,1}Z_{t+1,t}|\tilde{\mathfrak{L}}_{t-1}\right) \\ &= P\left(P\left(Z_{t+1,t+1} + \psi_{t+1,1}Z_{t+1,t}|\mathfrak{L}_t\right)|\tilde{\mathfrak{L}}_{t-1}\right) \\ &= P\left(\psi_{t+1,1}Z_{t+1,t} + P\left(Z_{t+1,t+1}|\mathfrak{L}_t\right)|\tilde{\mathfrak{L}}_{t-1}\right) \\ &= \psi_{t+1,1}P(Z_{t+1,t}|\tilde{\mathfrak{L}}_{t-1}).\end{aligned}$$

Similarly, using iterated projections and $\tilde{\mathfrak{L}}_{t-1} \subseteq \mathfrak{L}_t \subseteq \mathfrak{L}_{t+1}$, the 2-step ahead impulse response function is

$$
\begin{aligned}
(92)\ & I(2, v_t, \mathfrak{L}_{t-1}) \\
&= P\left(x_{t+2}|\tilde{\mathfrak{L}}_{t-1}\right) - P\left(x_{t+2}|\mathfrak{L}_{t-1}\right) \\
&= \sum_{i=3}^{\infty} \psi_{t+2,i} Z_{t+2,t+2-i} \\
&+ P\left(Z_{t+2,t+2} + \psi_{t+2,1} Z_{t+2,t+1} + \psi_{t+2,2} Z_{t+2,t}|\tilde{\mathfrak{L}}_{t-1}\right) \\
&- \sum_{i=3}^{\infty} \psi_{t+2,i} Z_{t+2,t+2-i} \\
&= P\left(Z_{t+2,t+2} + \psi_{t+2,1} Z_{t+2,t+1} + \psi_{t+2,2} Z_{t+2,t}|\tilde{\mathfrak{L}}_{t-1}\right) \\
&= P\left(P\left(P\left(Z_{t+2,t+2} + \psi_{t+2,1} Z_{t+2,t+1} + \psi_{t+2,2} Z_{t+2,t}|\mathfrak{L}_{t+1}\right)|\mathfrak{L}_t\right)|\tilde{\mathfrak{L}}_{t-1}\right) \\
&= \psi_{t+2,2} P\left(Z_{t+2,t}|\tilde{\mathfrak{L}}_{t-1}\right),
\end{aligned}
$$

and so on. Therefore $I(h, v_t, \mathfrak{L}_{t-1}) = \psi_{t+h,h} P(Z_{t+h,t}|\tilde{\mathfrak{L}}_{t-1})$. ■

Proof of Lemma 7. Clearly $\tilde{\mathfrak{L}}_{t-k-1} = \mathfrak{L}_{t-k-1} \oplus \epsilon_{t-k} = \mathfrak{L}_{t-k}$. Hence $P(P(x_t|\mathfrak{L}_{t-k})|\tilde{\mathfrak{L}}_{t-k-1}) = P(P(x_t|\mathfrak{L}_{t-k})|\mathfrak{L}_{t-k}) = P(x_t|\mathfrak{L}_{t-k}) = P(x_t|\tilde{\mathfrak{L}}_{t-k-1})$. If additionally x_t admits a strong orthogonal decomposition, then by Theorem 3 and the identity $\tilde{\mathfrak{L}}_{t-k-1} = \mathfrak{L}_{t-k-1} \oplus \epsilon_{t-k} = \mathfrak{L}_{t-k}$ we deduce $P(P(x_t|\tilde{\mathfrak{L}}_{t-k})|\mathfrak{L}_{t-k}) = P(P(x_t|\mathfrak{L}_{t-k+1})|\mathfrak{L}_{t-k}) = P(x_t|\mathfrak{L}_{t-k})$. ■

Proof of Theorem 8. Project $P_{t-1}x_t$ and define $\xi_t = x_t - P_{t-1}x_t$ (recall $\beta_0 = 0$)

$$
\begin{aligned}
(93)\qquad & P_{t-1}x_t = \sum_{i=1}^{\infty} \theta_i u_{t-i} + \sum_{i=1}^{\infty} \beta_i |u_{t-i}| \\
& \xi_t = \theta_0 u_t + \beta_0 |u_t| = \theta_0 u_t.
\end{aligned}
$$

Similarly, by (i)

$$
\begin{aligned}
(94)\qquad P_{t-2}P_{t-1}x_t &= \sum_{i=2}^{\infty} \theta_i u_{t-i} + \sum_{i=2}^{\infty} \beta_i |u_{t-i}| + P_{t-2}\left(\theta_1 u_{t-1} + \beta_1 |u_{t-1}|\right) \\
&= \sum_{i=2}^{\infty} \theta_i u_{t-i} + \sum_{i=2}^{\infty} \beta_i |u_{t-i}| + a_1 \\
\xi_{t-1} &= \theta_1 u_{t-1} + \beta_1 |u_{t-1}| - a_1,
\end{aligned}
$$

etc. In general

$$
(95)\quad \xi_{t-j} = \theta_j u_{t-j} + \beta_j |u_{t-j}| - a_j,
$$

giving the decomposition

$$
\begin{aligned}
(96)\quad x_t &= \sum_{i=0}^{\infty} \xi_{t-i} + V_t = \sum_{i=0}^{\infty} \left(\theta_i u_{t-i} + \beta_i |u_{t-i}| - a_i\right) + V_t \\
&= \sum_{i=0}^{\infty} \psi_i Z_{t-i} + V_t,
\end{aligned}
$$

where $V_t = \lim_{N\to\infty} P_{t-N} \cdots P_{t-1}x_t \in \mathfrak{L}_{-\infty}$, and

$$
\begin{aligned}
(97)\quad Z_{t-i} &= u_{t-i} + (1/\theta_i)\left(\beta_i |u_{t-i}| - a_i\right) \text{ and } \psi_i = \theta_i \text{ if } \theta_i \neq 0; \\
Z_{t-i} &= |u_{t-i}| - a_i/\beta_i \text{ and } \psi_i = \beta_i \text{ if } \theta_i = 0 \text{ and } \beta_i \neq 0; \\
Z_{t-i} &= u_{t-i} \text{ and } \psi_i = \theta_i \text{ if } \beta_i = 0.
\end{aligned}
$$

For any $k \geq 1$, by quasi-linearity and (i)

$$(98)\quad P_{t-1}\left(\sum_{j=0}^{k-1} \xi_{t-j}\right) = \sum_{j=1}^{k-1} \theta_j u_{t-j} + \sum_{j=1}^{k-1} (\beta_j |u_{t-j}| - a_j)$$

$$P_{t-2}\left(\sum_{j=0}^{k-1} \xi_{t-j}\right) = \sum_{j=2}^{k-1} \theta_j u_{t-j} + \sum_{j=2}^{k-1} (\beta_j |u_{t-j}| - a_j) + P_{t-2}\left(u_t + \theta_1 u_{t-1} + \beta_1 |u_{t-1}|\right) - a_1.$$

By serial independence and $\theta_1 u_{t-1} + \beta_1 |u_{t-1}| \in \mathfrak{L}_{t-1}$, $u_t \perp (\theta_1 u_{t-1} + \beta_1 |u_{t-1}|)$. If (ii) holds, then

$$(99)\quad P_{t-2}\left(u_t + \theta_1 u_{t-1} + \beta_1 |u_{t-1}|\right) = P_{t-2}\left(\theta_1 u_{t-1} + \beta_1 |u_{t-1}|\right) = a_1$$

$$P_{t-2}\left(\sum_{j=0}^{k-1} \xi_{t-j}\right) = \sum_{j=2}^{k-1} \theta_j u_{t-j} + \sum_{j=2}^{k-1} (\beta_j |u_{t-j}| - a_j),$$

and so on. Repeating, we obtain $P_{t-k}(\sum_{j=0}^{k-1} \xi_{t-j}) = 0$ for any $k \geq 1$ if (ii) holds.

Using Theorem 6, we deduce

$$(100)\quad I(k, v_t, \mathfrak{B}_{t-1}) = \psi_{t+h,h} P\left(Z_{t+h,,t} | \tilde{\mathfrak{L}}_{t-1}\right)$$

$$= \psi_{t+h,h} P\left(u_t + (1/\theta_h)(\beta_h |u_t| - a_h) | \tilde{\mathfrak{L}}_{t-1}\right) \text{ if } \theta_h \neq 0$$

$$= P(\beta_h |u_t| - a_h | \tilde{\mathfrak{L}}_{t-1}) \text{ if } \theta_h = 0,$$

where $a_h = 0$ if $\beta_h = 0$, hence if $v_t = u_t$

$$(101)\quad I(k, v_t, \mathfrak{B}_{t-1}) = \psi_{t+h,,h} u_t + (1/\theta_h)(\beta_h |u_t| - a_k) \text{ if } \theta_h \neq 0$$

$$= \beta_h |u_t| - a_h \text{ if } \theta_h = 0.$$

■

Proof of Theorem 9. From backward substitution

$$(102)\quad x_t = \phi x_{t-1} g_{t-1}(\gamma, c) + \epsilon_t$$

$$= \phi^2 x_{t-2} g_{t-2}(\gamma, c) g_{t-1}(\gamma, c) + \phi \epsilon_{t-1} g_{t-1}(\gamma, c) + \epsilon_t$$

$$= \phi^N x_{t-N} \prod_{i=1}^{N} g_{t-j}(\gamma, c) + \sum_{i=0}^{N-1} \phi^i \epsilon_{t-i} \prod_{j=1}^{i} g_{t-j}(\gamma, c).$$

Because $|\phi| < 1$ and $\limsup_N \prod_{i=1}^N g_{t-j}(\gamma, c) \in [1, -1]$, $a.s.$, the series representation $x_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \prod_{j=1}^{i} g_{t-j}(\gamma, c)$ $a.s.$ follows.

Using $E[\epsilon_t | \Im_{t-1}] = 0$ and the $\Im_{t-1}$-measurability of x_{t-1},

(103)

$$
\begin{aligned}
E\left[x_t|\Im_{t-1}\right] &= \sum_{i=1}^{\infty}\phi^i\epsilon_{t-i}\prod_{j=1}^{i}g_{t-j}(\gamma,c)\\
E\left[x_t|\Im_{t-2}\right] &= \sum_{i=2}^{\infty}\phi^i\epsilon_{t-i}\prod_{j=2}^{i}g_{t-j}(\gamma,c)E\left[g_{t-1}(\gamma,c)|\Im_{t-2}\right]+\phi E\left[\epsilon_{t-1}g_{t-1}(\gamma,c)|\Im_{t-2}\right]\\
&= \sum_{i=2}^{\infty}\phi^i\epsilon_{t-i}\prod_{j=2}^{i}g_{t-j}(\gamma,c)a+\phi b\\
E\left[x_t|\Im_{t-3}\right] &= \sum_{i=3}^{\infty}\phi^i\epsilon_{t-i}\prod_{j=3}^{i}g_{t-j}(\gamma,c)a^2+\phi^2 ab+\phi b\\
E\left[x_t|\Im_{t-N}\right] &= \phi^N a^{N-1}\sum_{i=N}^{\infty}\phi^{i-N}\epsilon_{t-i}\prod_{j=N}^{i}g_{t-j}(\gamma,c)+b\phi\sum_{i=0}^{N-2}\phi^i a^i
\end{aligned}
$$

By iterated expectations

$$
\begin{aligned}
(104)\; x_t &= \phi x_{t-1}g_{t-1}(\gamma,c)+\epsilon_t=E\left[x_t|\Im_{t-1}\right]+\epsilon_t\\
&= E\left[x_t|\Im_{t-2}\right]+\left(E\left[x_t|\Im_{t-1}\right]-E\left[x_t|\Im_{t-2}\right]\right)+\epsilon_t\\
&= \ldots=\sum_{i=0}^{\infty}\psi_i Z_{t-i}+V_t
\end{aligned}
$$

where $V_t=\lim_{N\to\infty}E\left[x_t|\Im_{t-N}\right]=\lim_{N\to\infty}(\sum_{i=N}^{\infty}\phi^i\epsilon_{t-i}\prod_{j=N}^{i}g_{t-j}(\gamma,c)a^{N-1}$ $+\,b\phi\sum_{i=0}^{N-2}\phi^i a^i)=b\phi/(1-\phi a)$, $a.s.$, $\psi_0=1$, $Z_t=\epsilon_t$, and for all $l\geq 1$

(105)

$$
\begin{aligned}
\psi_l Z_{t-l} &= E\left[x_t|\Im_{t-l}\right]-E\left[x_t|\Im_{t-l-1}\right]\\
&= \phi^l a^{l-1}\sum_{i=l}^{\infty}\phi^{i-l}\epsilon_{t-i}\prod_{j=l}^{i}g_{t-j}(\gamma,c)+b\phi\sum_{i=0}^{l-2}\phi^i a^i\\
&\quad -\phi^{l+1}a^l\sum_{i=l+1}^{\infty}\phi^{i-l-1}\epsilon_{t-i}\prod_{j=l+1}^{i}g_{t-j}(\gamma,c)-b\phi\sum_{i=0}^{l-1}\phi^i a^i\\
&= \phi^l a^{l-1}\left(\left[\epsilon_{t-l}g_{t-l}(\gamma,c)-b\right]+\left[g_{t-l}(\gamma,c)-a\right]\sum_{i=l+1}^{\infty}\phi^{i-l}\epsilon_{t-i}\prod_{j=l+1}^{i}g_{t-j}(\gamma,c)\right)
\end{aligned}
$$

hence $\psi_l=\phi^l a^{l-1}$,

(106)

$$
Z_{t-l}=\left[\epsilon_{t-l}g_{t-l}(\gamma,c)-b\right]+\left[g_{t-l}(\gamma,c)-a\right]\sum_{i=l+1}^{\infty}\phi^{i-l}\epsilon_{t-i}\prod_{j=l+1}^{i}g_{t-j}(\gamma,c),
$$

and

(107)

$$
\begin{aligned}
E\left[Z_{t-l}|\Im_{t-l-1}\right] &= E\left[\epsilon_{t-l}g_{t-l}(\gamma,c)-b\right]\\
&\quad +\sum_{i=l+1}^{\infty}\phi^{i-l}\epsilon_{t-i}\prod_{j=l+1}^{i}g_{t-j}(\gamma,c)E\left[g_{t-l}(\gamma,c)-a\right]\\
&= 0.
\end{aligned}
$$

■

References

[1] Arcones, M.A.. (2000). M-Estimators Converging to a Stable Limit, Journal of Muiltivariate Analysis 74, 193-221.

[2] Bierens, H. J. and W. Ploberger (1997). Asymptotic Theory of Integrated Conditional Moment Tests, Econometrica 65, 1129-1151.

[3] Bingham, N.H., C.M. Goldie and J. L. Teugels (1987) Regular Variation. Cambridge Univ. Press, Great Britain.

[4] Cambanis, S.C.D. Hardin and A. Weron (1988). Innovations and Wold Decompositions of Stable Sequences, Probability Theory and Related Fields 79, 1-27.

[5] Cheung, Y.W. (1993). Long Memory in Foreign Exchange Rates, Journal of Business and Economic Statistics 11, 93-101.

[6] Cline, D.B.H. (1983). Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data, Unpublished Ph.D. Dissertation, Department of Statistics, Colorado State University.

[7] Cline, D.B.H. and P. Brockwell (1985). Linear Prediction of ARMA Processes with Infinite Variance, Stochastic Processes and Their Applications 19, 281-296.

[8] Davis, R. and S. Resnick (1985a). Limit Theory for Moving Averages of Random-Variables with Regularly Varying Tail Probabilities, Annals of Probability 13, 179-195.

[9] Davis, R. and S. Resnick (1985b). More Limit Theory for the Sample Correlation-Function of Moving Averages, Stochastic Processes and Their Applications 20, 257-279.

[10] de Jong, R., and C. Han (2002). Properties of Lp-GMM estimators, Econometric Theory 18, 491-504.

[11] van Dijk, D., T. Teräsvirta and P.H. Franses 2000. Smooth Transition Autoregressive Models-A Survey of Recent Developments, Econometric Institute Research Report EI2000-23/A, Erasmus University.

[12] Engle, R., and V. Ng (1993). Measuring and Testing the Impact of News on Volatility, Journal of Finance 48, 1749-78.

[13] Falk, B. and C. Wang (2003). Testing Long-Run PPP with Infinite-Variance Returns, Journal of Applied Econometrics 18, 471-484.

[14] Faulkner, G.D. and J.E. Huneycutt Jr. (1978). Orthogonal Decompositions of Isometries in a Banach Space, Proc. Am. Math. Soc. 69, 125-128.

[15] Feller, W. (1971). An Introduction to Probability Theory and its Applications, 2 ed., vol. 2, (Wiley: New York).

[16] Franses, P.H. and D. van Dijk (2000). Non-linear Time Series Models in Empirical Finance (Cambridge University Press).

[17] Gallant, R., P. Rossi, and G. Tauchen (1993). Nonlinear Dynamic Structures, Econometrica 61, 871-907.

[18] Giles, J.R. (1967). Classes of Semi-Inner-Product Spaces, Trans. Amer. Math. Soc. 129, 436-446.

[19] Giles, J.R. (2000). Introduction to the Analysis of Normed Linear Spaces (University Press: Cambridge).

[20] González-Rivera, G. (1998). Smooth-Transition GARCH Models, Nonlinear Dynamics and Econometrcs 3, No. 2, Article 1.

[21] Gourieroux, C., and J. Jasiak (2003). Nonlinear Innovations and Impulse Responses with an Application to VaR Sensitivity, mimeo, Centre de Recherche en Économie et Statistique, Institut National de la Statistique et des Etudes Economique.

[22] Han, C., and R. de Jong (2004). Closest Moment Estimation under General Conditions, Annales d'Economie et de Statistique 74, 1-13.

[23] Hardin, C., G. Samorodnitsky and M. Taqqu (1991). Non-linear Regression of Stable Random Variables, The Annals of Applied Probability 1, 582-612.

[24] Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, Annals of Mathematical Statistics 3, 1163-1174.

[25] Hill, J.B. (2005a). Gaussian Tests of "Extremal White Noise" for Dependent, Heterogeneous, Heavy Tailed Time Series with an Application. Dept of Economics, Florida International University; available at http://econwpa.wustl.edu:80/eps/em/ papers/0411/0411014.pdf.

[26] Hill, J.B. (2005b). On Tail Index Estimation Using Dependent, Heterogenous Data. Dept of Economics, Florida International University; available at http://econwpa.wustl.edu:80/eps/em/ papers/0505/0505005.pdf.

[27] Hols, M.C.A.B., and C.G. de Vries (1991) TheLimiting Distribution of Extremal Exchange Rate Returns, Journal of Applied Econometrics 6, 287-302.

[28] James, R.C. (1947). Orthogonality and Linear Functionals in a Normed Linear Space, Transactions of the American Mathematical Society 61, 265-292.

[29] Kees, G.K., and C.J.M. Kool (1992). Tail Estimates of East European Exchange Rates, Journal of Business & Economic Statistics 10, 83-96 .

[30] Kokoszka, P.S. and M.S. Taqqu (1994). Infinite Variance Stable ARMA Processes, Journal of Time Series Analysis 15, 203-220.

[31] Kokoszka, P.S. and M.S. Taqqu (1996). Infinite Variance Stable Moving Averages with Long Memory, Journal of Econometrics 73, 79-99.

[32] Koop, G., M.H. Pesaran, and S. Potter (1996). Impulse Response Analysis in Nonlinear Multivariate Models, Journal of Econometrics 4, 119-147.

[33] Lin, W. L. (1997). Impulse Response Function for Conditional Volatility in GARCH Models, Journal of Business and Economic Statistics 15, 15-25.

[34] Lindenstrauss, J. and L. Tzafriri (1977). Classical Banach Spaces (Springer, New York).

[35] Loretan, M. and P.C.B. Phillips (1994). Testing the Covariance Stationarity of Heavy-Tailed Economic Time Series: An Overview of the Theory with Applications to Financial Data Sets, Journal of Empirical Finance 1, 211-48.

[36] Lundbergh, S, T. Teräsvirta, and D. van Dijk D, Time-Varying Smooth Transition Autoregressive Models (2003). Journal of Business & Economic Statistics 21, 104-121.

[37] Lundbergh, S., and T. Teräsvirta (2005). A Time Series Model for an Exchange Rate in a Target Zone with Applications, Journal of Econometrics: in press.

[38] McMillan, D., and A. Speight (2002). Nonlinear Dependence in Inter-War Exchange Rates: Some Further Evidence, Applied Economics Letters 9, 359-364.

[39] Megginson, R. (1998). An Introduction to Banach Space Theory (Springer-Verlag: New York).

[40] Miamee, A.G. and M. Pourahmadi (1988). Wold Decompositions, Prediction and Parameterization of Stationary Processes with Infinite Variance, Probability and Related Fields, 79, 145 - 164.

[41] Micheal, P., A.R. Nobay, and D.A. Peel (1997).Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation, Journal of Political Economy 105, 862-879.

[42] Mikosch, T., and C. Stărică (2000). Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1,1) Process, Annals of Statistics 28, 1427-1451.

[43] Phillips, P.C.B., J. McFarland, and P. McMahon (1996). Robust Tests of Forward Exchange Market Efficiency with Empirical Evidence from the 1920s, Journal of Applied Econometrics 11, 1-22.

[44] Saikkonen, P. and R. Luukkonen (1988). Lagrange Multiplies Tests for Testing Nonlinearities in Time Series Models, Scandinavian Journal of Statistics 15, 55-68.

[45] Samorodnitsky, G. and M.S. Taqqu (1994). Stable Non-Gaussian Random Processes (Chapman and Hall, New York).

[46] Sims, C. (1980). Macroeconomics and Reality, Econometrica 48, 1-48.

[47] Singer, I. (1970). Bases in Banach Spaces (Springer: Berlin-Heidelberg-New York).

[48] Teräsvirta, T. (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, Journal of the American Statistical Association 89, 208-218.

[49] Tong, H. (1990). Nonlinear Time Series: A Dynamical System Approach (Oxford, U.K.: Oxford University Press).

[50] Urbanic, K. (1964). Prediction of Strictly Stationary Sequences, Colloquium Mathematicum 12, 115-129.

[51] Urbanic, K. (1967). Some Prediction Problems for Strictly Stationary Processes. In: Proc. Fifth Berkeley Symp. Math. Stat. Probab., Vol. 2, Part I, 235-258 (University of California Press: Berkeley).

[52] Wold, H.O.A. (1938). A Study in the Analysis of Stationary Time Series (Uppsala, Almqvist and Wiksell).

Table 1

L_p-GMM Estimates (p = 1.5) and Tail Indices			
ESTAR			
Parameter	Yen	Euro	BP
ϕ	-.089 (.027)[a]	.296 (.073)	.099 (.026)
γ	23.3 (13.7)	43.0 (11.9)	15.5 (7.57)
c	.013 (.004)	.015 (.003)	.020 (.009)
$\max\{\lvert\epsilon_t\rvert\}$	.031	.025	.021
LSTAR			
Parameter	Yen	Euro	BP
ϕ	-.093 (.053)	-.043(.814)	.130 (.107)
γ	40.7 (26.7)	68.3 (159)	67.2 (29.1)
c	.003 (.001)	.001 (.003)	.020 (.000)
$\max\{\lvert\epsilon_t\rvert\}$	.030	.027	.021
	Yen	Euro	BP
$E[x_t \lvert \Im_{t-1}]=0^b$	.058	.096	.020
$P(x_t \lvert \Im_{t-1})= 0^c$	.061	.072	.097
$\min\{x_t\}$	-.030	-.025	-.019
$\max\{x_t\}$	.025	.027	.021
$\hat{\alpha}_m$	2.55±.70[d]	3.37±1.20	2.96±1.03

Notes: a. Parameter estimate (heteroscedasticity robust standard error).
b. Hong and White's (1995) test of conditional mean mis-specification: values denote p-values.
c. The null is $E[x_t^{<p-1>} \lvert \Im_{t-1}]=0$ (i.e. $P(x_t \lvert \Im_{t-1})=0$).
d. Tail index estimator and 95% interval length based on a Newey-West kernel estimator with Bartlett kernel: see Hill (2005b).

Table 2

Hong-White Tests[a]								
			ESTAR			LSTAR		
Test	Null Hypothesis	h^d	Yen	Euro	BP	Yen	Euro	BP
1.	$E[\epsilon_t \lvert \Im_{t-1}] = 0$	-	.264	.096	.384	.258	.078	.100
2.	$P(\epsilon_{t+k} \lvert \Im_{t-1}) = 0^b$	-	.941	.091	.540	.582	.088	.126
3.	$E[Z_t \lvert \Im_{t-1}] = 0$	-	.099	.000	.249	.188	.197	.109
4.	$P(Z_{t+k} \lvert \Im_{t-1}) = 0$	-	.161	.813	.159	.401	.465	.217
5.	$P(\sum_{k=0}^{h} \pi_k \epsilon_{t+k} \lvert \Im_{t-1}) = 0^c$	10	.191	.127	.799	.075	.199	.159
		20	.233	.108	.100	.076	.702	.386
6.	$P(\sum_{k=0}^{h} \pi_k Z_{t+k} \lvert \Im_{t-1}) = 0$	10	.091	.831	.292	.937	.351	.179
		20	.114	.339	.119	.332	.332	.376

Notes: a. The p-value of Hong and White's (1995) test. b. The null is $E[\epsilon_{t+k}^{<p-1>} \lvert \Im_{t-1}] = 0$.
c. The null is $E[(\sum_{k=0}^{h} \pi_k \epsilon_{t+k})^{<p-1>} \lvert \Im_{t-1}] = 0$. d. The "h" in $\sum_{k=0}^{h} \pi_k \epsilon_{t+k}$.

Figure 1
IRF's for ESTAR (p = 1.5), Impulses: $\mathbf{v}_{n-i} = \mathbf{c}$

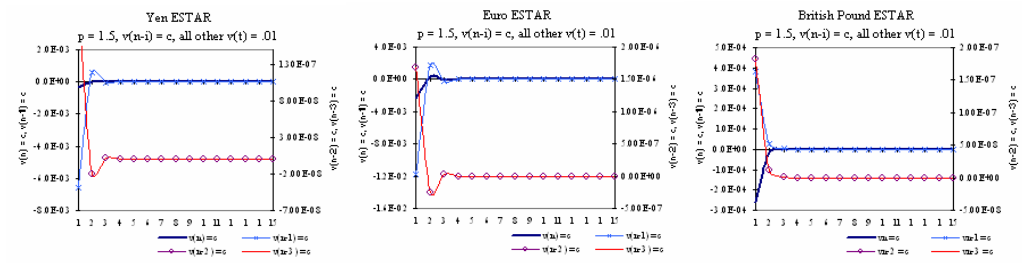

Figure 2
IRF Densities (randomized impulses)