Strong orthogonal decompositions and non-linear impulse response functions for infinite-variance processes

Jonathan B. HILL

Abstract: The author proves that 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. His theory includes as special cases all previous Wold-type decompositions of discrete time processes, completely characterizes when non-linear heavy-tailed processes obtain a strong-orthogonal moving average representation, and easily promotes a theory of non-linear impulse response functions for infinite-variance processes. The author exemplifies his theory by developing a non-linear impulse response function for smooth transition threshold processes, and discusses how to test decomposition innovations for strong orthogonality and whether the proposed model represents the best predictor. A data set on currency exchange rates allows him to illustrate his methodology.

Décompositions orthogonales fortes et fonctions réponses à impulsion non linéaire pour des processus à variance infinie

Résumé : L'auteur démontre que pour qu'un espace de Banach réflexif lisse de processus à temps discret admette une décomposition de type Wold à innovations prévisionnelles fortement orthogonales, il faut et il suffit que l'opérateur projectif engendrant les innovations possède la propriété d'itérations. Son résultat, qui englobe comme cas particuliers toutes les décompositions de type Wold obtenues antérieurement pour des processus à temps discret, caractérise complètement les situations où les processus non linéaires à queues lourdes possèdent une représentation en moyenne mobile fortement orthogonale, en plus de conduire naturellement à une théorie des fonctions réponses à impulsion non linéaire pour les processus à variance infinie. Il se sert de cette théorie pour développer une fonction à impulsion non linéaire pour des processus à seuil de transition lisse et montre comment tester l'orthogonalité forte des innovations de la décomposition et vérifier si le modèle proposé est bel et bien le meilleur prédicteur. Un jeu de données sur des taux de change de devises lui permet d'illustrer son propos.

1. INTRODUCTION

This paper presents a complete theory of Wold-type orthogonal decompositions in smooth, reflexive Banach spaces of discrete time processes. Such Banach spaces include Hilbert spaces, the stable laws, and L_p-spaces, they contain linear and non-linear processes and include many processes with a stochastic recurrence representation (e.g., GARCH and FIGARCH processes; see Basrak, Davis & Mikosch 2001, 2002). An immediate application is the construction of non-linear impulse response functions (IRFs) 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

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

for some set of orthogonal innovations $\{Z_t\}$ and a "residual" V_n. The seminal work of Wold (1938) 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

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

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

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

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

Similarly, Cambanis, Hardin & Weron (1988) establish an asymmetric decomposition theory for $L_p(\Omega, \mathfrak{F}_t, \mu)$ processes in $\overline{\text{sp}}_t$. 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 $\overline{\text{sp}}_t - P_{t-1}\overline{\text{sp}}_t$. Specifically, the authors prove (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i), and operator linearity is expedited by the fact that $\overline{\text{sp}}_{t-1}$ is codimension one in $\overline{\text{sp}}_t$.

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 (non-linear) L_p-spaces, Cambanis, Hardin & Weron (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 $\overline{\text{sp}}_t - P_{t-1}\overline{\text{sp}}_t$ in order to promote strong orthogonality is not a trivial simplification, however. The explicit omission of non-linear best predictors and orthogonal innovations must be viewed critically in light of developments in the theory and empirical methods associated with non-linear stochastic processes. For example, moving average forms have been utilized to characterize linear dependence within processes with regularly varying tails: see Davis & Resnick (1985a,b), and Kokoszka & Taqqu (1994, 1996). The innovations in this literature are typically assumed to be independent and identically distributed , hence strongly orthogonal to far more than subspaces of $\overline{\text{sp}}_t$. Except for the special case of symmetric stable process (Cambanis, Hardin & Weron 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 à la IRFs in time series settings in which amassed evidence suggests non-linear data generating processes with heavy tails. See, e.g., Hols & de Vries (1991), Cheung (1993), Gallant, Rossi & Tauchen (1993), Phillips, McFarland & McMahon (1996), Lin (1997), Mikosch & Stărică (2000), Falk & Wang (2003), and Hill (2005b). In the economics and finance literatures, the implied "impulses" are predominantly assumed to be independent and identically distributed 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 (non-linear) forecasts, then a decomposition theory that goes substantially beyond the extant literature is required.

Toward this end, Gallant, Rossi & Tauchen (1993) and Koop, Pesaran & Potter (1996) develop non-parametric representations of impulse responses for general non-linear processes in the Hilbert space $L_2(\Omega, \mathfrak{F}_t, \mu)$. 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., $L_p(\Omega, \mathfrak{F}_t, \mu)$, $p < 2$). Gouriéroux & Jasiak (2003) develop a parametric Volterra-type expansion of independent and identically distributed 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 distributed.

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) that the property of iterated projections is necessary and sufficient for the existence of a decomposition with asymmetrically strong orthogonal innovations, $\overline{sp}(Z_{t+i}, \ldots, Z_t) \perp \mathfrak{B}_{t-1}$, for every $i \geq 1$. Using an arbitrary metric projection mapping $P_{t-1} : \mathfrak{B}_t \to \mathfrak{B}_{t-1}$, our results do not exploit operator linearity in general, and they specifically do not rely on properties of the closed linear span. Theorem 3 allows for a simple characterization of a non-linear 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 non-linear IRFs 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 non-linear IRFs 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 non-linear impulse response functions based on best L_p-metric projection and the properties of strong orthogonality and iterated projections. We construct in Section 5 an extended example demonstrating the decomposition of a non-linear smooth transition threshold model and associated non-linear 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. We therefore focus our attention on the empirical task of verifying whether the non-linear 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 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.

The rest of paper is organized as follows: Section 2 contains a preliminary metric projection

theory; Section 3 contains the main results; we develop a theory of non-linear impulse response functions in Section 4, and Section 5 contains an example and an application. The Appendix contains formal proofs.

In the sequel, we employ the following notation and definition conventions. Denote by $\mathfrak{B}_t \equiv \mathfrak{B}(\Omega, \mathfrak{F}_t, \mu, \| \cdot \|)$ 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

$$\mathfrak{B} \equiv \mathfrak{B}(\Omega, \mathfrak{F}, \mu, \| \cdot \|) = \bigcup_{t \in \mathbb{Z}} \mathfrak{B}_t, \quad \mathfrak{F} = \bigcup_{t \in \mathbb{Z}} \mathfrak{F}_t.$$

It is understood that $\|x\| < \infty$ for any $x \in \mathfrak{B}_t$, and $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$. Let $\mathfrak{L}_t \equiv L_p(\Omega, \mathfrak{F}_t, \mu)$, $p \leq 2$. We denote the signed power $\mathrm{sgn}\,(z)|z|^a$ as $z^{\langle a \rangle}$, $a \in \mathbb{R}$. 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 $S_1, \ldots, S_n$, $n > 1$, we write $S_1 + \cdots + S_n$ to denote the stochastic space $\{\sum_{i=1}^n Z_i : Z_i \in S_i\}$. For orthogonal subspaces, $S_1, \ldots, S_n$, $n > 1$, the space $S_n \oplus S_{n-1} \oplus \cdots \oplus S_1$ (synonymously $\bigoplus_{i=0}^{n-1} S_{n-i}$) denotes the space $\sum_{i=0}^{n-1} S_i$, where

$$\sum_{i=0}^{\ell-1} S_{n-i} \perp \sum_{i=\ell}^{n-1} S_{n-i}$$

for all $1 \leq \ell < n$. In general orthogonality is not symmetric. For spaces $\bigoplus_{i=0}^{n-1} S_{n-i}$, we say the subspaces S_t are *strong orthogonal*. Similarly, whenever $S_t \perp S_s$ for every $s < t$, 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 innovation Banach spaces. For background theory, see Singer (1970), Lindenstrauss & 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; see James (1947): y is *James orthogonal* to x whenever $\|y + \lambda x\| \geq \|y\|$ for every real scalar $\lambda \in \mathbb{R}$, denoted $y \perp x$. Banach space norms $\| \cdot \|$ may be supported by arbitrarily many semi-inner products $[\cdot, \cdot]$. 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; 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 $[\cdot, \cdot]$ that supports $\| \cdot \|$, such that $[U, V] = 0$: for each $u \in U$ and $v \in V$, the inner product $[\cdot, \cdot]$ satisfies $[u, v] = 0$ and $[u, u]^{1/2} = \|u\|$, $[v, v]^{1/2} = \|v\|$. Moreover, if $\mathfrak{B}$ is smooth then $[\cdot, \cdot]$ is unique.*

2.1. 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*

$$\|u - v\| \leq \|u - \tilde{v}\|$$

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 $P(u \mid V) = v$: the projection $P(u \mid V)$ is identically the "best predictor" of u. The projection $P(u \mid V)$ is continuous, bounded, and idempotent, although not in general linear; see below. For subspaces $U, V \subseteq \mathfrak{B}$, the notation $P(U \mid V)$ is understood to represent the projection space $P(U \mid V) = \{P(u \mid V) : u \in U\}$.

2.2. Iterated projections and operator linearity.

We say that the property of *iterated projections* holds in $V_1 \subseteq \mathfrak{B}$ for some projection operator $P \colon U \to V_1$ when for any subspace $V_0 \subseteq V_1 \subseteq \mathfrak{B}$, $P\{P(u \,|\, V_1) \,|\, V_0\} = P(u \,|\, V_0)$. We say that a projection operator P which maps $P \colon U \to V$ is a *linear operator* on $U \supseteq V$ if for any elements $u_1, u_2 \in U$ and any real numbers $a, b \in \mathbf{R}$, $P(au_1 + bu_2 \,|\, V) = aP(u_1 \,|\, V) + bP(u_2 \,|\, V)$, a homogenous, additive function of u_1 and u_2. If a projection operator is a linear operator then iterated projections holds; see Lemma 2, part viii.

2.3. Metric projection.

In the following, assume that u is an arbitrary element of U, and denote by $\mathrm{E}\,(u \,|\, V)$ the expectation of u conditioned on $\sigma(V)$. For a proof, see Lemma 2 of Hill (2005a), or 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* $[\,\cdot\,, \cdot\,]$ *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* $u \in U$, $\|P(u \,|\, V)\| \le k\|u\|$ *for some scalar* $0 < k < \infty$; (vi) *unbiasedness in* L_p: *if* $U, V \subseteq L_p$, $1 < p \le 2$, *and if* $\mathrm{E}\,(u \,|\, V) \in V$, *then* $P(u \,|\, V) = \mathrm{E}\,(u \,|\, V) = v$ *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* $a \in \mathbf{R}$; *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 that 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 $\mathrm{E}\,(u \,|\, V)$ may not be an element of the space V. For example, suppose $V = \overline{\mathrm{sp}}\,(v_1, \ldots, v_n)$, 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 $\mathrm{E}\,(u \,|\, V)$ need not be linear: see, for example, Hardin, Samorodnitsky & Taqqu (1991). Of course, for non-Gaussian processes in $L_2(\Omega, \mathfrak{F}_t, \mu)$, the best L_2-predictor $\mathrm{E}\,(u \,|\, V)$ need not be linear.

3. MAIN RESULTS

Denote by $P_{t,s}$ the metric projection mapping $P \colon \mathfrak{B}_t \to \mathfrak{B}_s$, $s \le t$. Construct the following spaces

$$\mathfrak{B}_{-\infty} = \bigcap_n \mathfrak{B}_n, \quad \mathfrak{B}_{+\infty} = \bigcup_n \mathfrak{B}_n.$$

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*

$$\mathfrak{B}_n = \left(\sum_{i=0}^\infty N_{n-i} \right) + \mathfrak{B}_{-\infty}, \tag{4}$$

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

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

(ii) $P_{t,t-\ell} P_{t,t-k} = P_{t,t-k}$ *for every* t, $k \le \ell$.

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

Remark 2. Cambanis, Hardin & Weron (1988) point out that operator linearity implies result (i) for processes in $L_p(\Omega, \Im_n, \mu)$ and for projection into arbitrary $L_p(\Omega, \Im_n, \mu)$-spaces. This result, however, is trivial and does not anticipate the dual relationship between orthogonality and iterated projections (i) $\leftrightarrow$ (ii) without invoking operator linearity. Assume $P_{t,s}$ is a linear operator on $\mathfrak{B}_t$, and consider any element

$$\sum_{i=k}^{\ell} \xi_{n-i} \in \sum_{i=k}^{\ell} N_{n-i}, \quad \xi_{n-i} \in N_{n-i}, \quad 0 \leq k \leq \ell.$$

Because $N_{n-i} \perp \mathfrak{B}_{n-i-1}$ by construction and by Lemma 2, part ii, we have $P_{n,n-i-1}\xi_{n-i} = 0$. By operator linearity, we conclude

$$P_{n,n-\ell-1} \sum_{i=k}^{\ell} \xi_{n-i} = \sum_{i=k}^{\ell} P_{n,n-\ell-1}\xi_{n-i} = 0,$$

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

Remark 3. 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 will be particularly useful if evidence suggests that a chosen linear or non-linear 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 $Y \in \bigoplus_{i=0}^{\infty} N_{n-i}$ obtains a unique norm-convergent series representation, we may write elements $X_n \in \mathfrak{B}_n$ in a straightforward moving-average form. See Corollary 4 of Hill (2005a).

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*

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

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*

$$\overline{\mathrm{sp}}(Z_n, Z_{n-1}, \ldots) \perp \mathfrak{B}_{-\infty}, \quad \overline{\mathrm{sp}}(Z_{t+i}, \ldots, Z_t) \perp \mathfrak{B}_{t-1}, \quad \forall t \leq n, \forall i \geq 0,$$

if and only if $P_{t,t-\ell}P_{t,t-k} = P_{t,t-\ell}$ *for every* $t, k \leq \ell$.

ii. *Moreover,* $\psi_{n,0} = 1$, *and the coefficients* $\psi_{n,i}$ *uniquely satisfy the recursive relationship for* $i = 1, 2, \ldots$,

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

Remark 4. 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}. A well-known exception holds for causal-invertible ARMA processes: see, e.g., Cline (1983). 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.

Remark 5. Theorem 5 of Hill (2005a) characterizes necessary and sufficient conditions for the innovations to be symmetrically strong orthogonal for processes in $L_p(\Omega, \mathfrak{F}_t, \mu)$. Essentially $\mathfrak{L}_{t-1} \perp N_t$, $P_{t,t-1}X_t = \mathrm{E}\left(X_t \mid \mathfrak{F}_{t-1}\right)$, and $\mathrm{E}\left[\{X_t - \mathrm{E}(X_t \mid \mathfrak{F}_{t-1})\}^{\langle p-1 \rangle} \mid \mathfrak{F}_{t-1}\right] = 0$, $\mathfrak{F}_{t-1}$-a.e., all identically imply the innovations with be symmetrically strong orthogonal. The second property identically implies $P_{t,t-1}\{X_t - P_{t,t-1}(X_t)\}^{\langle p-1 \rangle} = 0$; the latter property is simply a martingale difference property, and implies $\mathrm{E}\left[\{X_t - \mathrm{E}\left(X_t \mid \mathfrak{F}_{t-1}\right)\}^{\langle p-1 \rangle}Y_{t-1}\right] = 0$, for every $\mathfrak{F}_{t-1}$-measurable random variable Y_{t-1}. See also Cambanis, Hardin & Weron (1988).

4. NON-LINEAR IMPULSE RESPONSE FUNCTIONS IN $\mathfrak{L}_t$

In the following we develop a general theory of non-linear impulse response functions (IRFs) 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 non-linear impulse response function

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

The above definition simply generalizes the expectations based format of Koop, Pesaran & Potter (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 $I(h, v_t, \omega_{t-1}) = P(x_{t+h} \mid \omega_{t-1}, v_t) - P(x_{t+h} \mid \omega_{t-1})$ 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, Pesaran & Potter (1996). The impulse response function $I(h, V_t, \mathfrak{L}_{t-1})$ is an $\mathfrak{F}_{t-1}$-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 5. *Assume the process* $\{x_\tau : \tau \leq t\}$ *lies in* $L_p(\Omega, \mathfrak{S}_t, \mu)$ *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* $\tilde{\mathfrak{L}}_{t-k-1}$ *for any k.* *Then*

$$I(h, V_t, \mathfrak{L}_{t-1}) = \psi_{t+h,h}P(Z_{t+h,t} \mid \tilde{\mathfrak{L}}_{t-1}). \tag{5}$$

Remark 6. Strong orthogonality is required because the line of proof exploits iterated projections; see Theorem 3.

Remark 7. 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, the $x_t = \sum_{i=0}^\infty \psi_i V_{t-i}$, V_t are independent and identically distributed. It is easy to show that (5) 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 8. In L_2 the non-linear IRF $I(h, V_t, \mathfrak{L}_{t-1})$ is identically the generalized impulse response function characterized by Equation (9) in Koop, Pesaran & Potter (1996) as long as the

projection operator minimizes the mean squared error. Otherwise (5) characterizes a further generalization of the generalized IRF of Koop, Pesaran, and Potter (1996) to best L_p-projection of finite-variance processes.

Remark 9. Koop, Pesaran & Potter (1996) characterize non-parametric and bootstrap methods for estimating the conditional expectations based on draws from the empirical distributions of $\{x_{t-1}, x_{t-2}, \ldots\}$ 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 $P(Z_{n+h,n} \mid \tilde{\mathfrak{L}}_{n-1})$ 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.,

$$P\{P(x_t \mid \mathfrak{L}_{t-k}) \mid \tilde{\mathfrak{L}}_{t-k-1}\} = P(x_t \mid \tilde{\mathfrak{L}}_{t-k-1}),$$

does not diminish the generality of the result by very much. For example, if $\mathfrak{I}_t \equiv \sigma(x_\tau : \tau \leq t) = \sigma(\varepsilon_\tau : \tau \leq t)$ for some stochastic process $\{\varepsilon_t\}$, and the impulses V_t are simply ε_t, then the assumption holds because

$$\tilde{\mathfrak{L}}_t = \mathfrak{L}_t \oplus V_{t+1} = \mathfrak{L}_t \oplus \varepsilon_{t+1} = \mathfrak{L}_{t+1}.$$

This will hold for infinitely large classes of linear and non-linear processes; see Section 5 for an example.

LEMMA 6. *Let* $\mathfrak{I}_t \equiv \sigma(x_\tau : \tau \leq t) = \sigma(\varepsilon_\tau : \tau \leq t)$, *and* $V_t = \varepsilon_t$ *for all* $t \in \mathbb{Z}$. *Then*

$$P\{P(x_t \mid \mathfrak{L}_{t-k}) \mid \tilde{\mathfrak{L}}_{t-k-1}\} = P(x_t \mid \tilde{\mathfrak{L}}_{t-k-1})$$

for all $k \geq 0$. *Additionally, if* x_t *admits a strong orthogonal decomposition, then*

$$P\{P(x_t \mid \tilde{\mathfrak{L}}_{t-k}) \mid \mathfrak{L}_{t-k}\} = P(x_t \mid \mathfrak{L}_{t-k}).$$

5. THRESHOLD MODELS AND EMPIRICAL 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 non-linear IRF which requires iterated projections; see Theorems 3 and 5. In this section we focus our attention entirely on a simple threshold model for the sake of brevity.

Due to linearity and iteration properties, the predominant practice in the literature is to assume that 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 prediction errors are strong orthogonal. We then derive a sample non-linear IRF for the particular threshold model, and apply the model and specification tests to the daily returns of currency exchange rates. Proofs of each result in this section can be found in Hill (2005a).

5.1. Threshold model and orthogonal decomposition.

A growing literature suggests that the returns to many macroeconomic and financial time series have heavy tails, are serially uncorrelated, and have some form of non-linear structure. See Tong (1990), Kees & Kool (1992), Loretan & Phillips (1994), Franses & van Dijk (2000), Lundbergh, Teräsvirta & van Dijk (2003), and Lundbergh & Teräsvirta (2005), to name a few. In particular,

the daily log-returns of many currency exchange rates appear to be serially uncorrelated, to have an infinite kurtosis or infinite variance, and to display serially asymmetric extremes: see Hols & de Vries (1991), Hill (2005b), and the citations therein. Moreover, the non-linear structure of exchange rates has an intuitive regime transitional form based on "banding" policies in economic unions; see, e.g., Lundbergh & Teräsvirta (2006).

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; see Saikkonen & Luukkonen (1988), Teräsvirta (1994), Micheal, Nobay & Peel (1997), and Lundbergh & Teräsvirta (2006). Denote by x_t the log-return $\Delta \ln y_t$ of a daily exchange rate y_t. Let $x_t \in L_p(\Omega, \Im_t, P) = \mathfrak{L}_t$, $\Im_t = \sigma(x_s : s \leq t)$, and let α be the moment supremum of $\varepsilon_t : \mathrm{E}|\varepsilon_t|^p < \infty$ 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

$$x_t = \phi x_{t-1} g_{t-1}(\gamma, c) + \varepsilon_t, \quad |\phi| < 1, \quad \gamma \geq 0, \, c > 0, \tag{6}$$

where ε_t is strictly stationary,

$$\mathrm{E}(\varepsilon_t \,|\, \Im_{t-1}) = 0, \quad g_{t-1}(\gamma, c) = \exp\{-\gamma(|\varepsilon_{t-1}| - c)^2\}$$

in the ESTAR case, and in the LSTAR case

$$g_{t-1}(\gamma, c) = \left[1 + \exp\{-\gamma(|\varepsilon_{t-1}| - c)\}\right]^{-1} - \{1 + \exp(\gamma c)\}^{-1}.$$

Whether $\mathrm{E}(\varepsilon_t \,|\, \Im_{t-1}) = 0$ 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 $|\varepsilon_{t-1}|$); see van Dijk, Teräsvirta & Franses (2000).

The ESTAR form naturally articulates "inner" ($|\varepsilon_{t-1}| \approx c$) and "outer" ($|\varepsilon_{t-1}| \neq c$) regimes: when the previous period's shock $|\varepsilon_{t-1}|$ is near c, $x_t \approx \phi x_{t-1} + \varepsilon_t$, hence the return is serially persistent; when $|\varepsilon_{t-1}|$ is far from c, $x_t \approx \varepsilon_t$ such that the return is noisy. The LSTAR characterizes "lower" and "upper" regimes, respectively, when $|\varepsilon_{t-1}| \approx 0$ then $x_t \approx \varepsilon_t$ and as $|\varepsilon_{t-1}| \to \infty$, then $x_t = \phi\{1 + \exp(-\gamma c)\}^{-1}\}x_{t-1} + \varepsilon_t$. As the scale $\gamma \to \infty$ the LSTAR model converges to a Self Exciting Threshold Autoregression: $x_t = \phi x_{t-1} I(|\varepsilon_{t-1}| > c) + \varepsilon_t$. The LSTAR model naturally implies that 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 $|\varepsilon_{t-1}|$ is far from c which will predominantly occur when $|\varepsilon_{t-1}| < c$.

Although the ESTAR model has traditionally been used to capture symmetric banding policies for exchange rate levels (e.g., Micheal, Nobay & Peel 1997 and Lundbergh & Teräsvirta 2006), its ability to capture extremal asymmetries is noteworthy because daily returns likely illicit asymmetric responses from traders and policy makers; large deviations may suggest a market crisis whereas small deviations may not be noteworthy. See Engle & Ng (1993). Such volatility asymmetries have been recently modeled as smooth transition GARCH processes: see González-Rivera (1998) and McMillan & Speight (2002).

We may decompose x_t by straightforward backward substitution.

THEOREM 7. *Assume that the stochastic process $\{x_\tau : \tau \leq t\}$ lies in $L_p(\Omega, \Im_t, \mu)$, and that* (6) *holds. Then $x_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \prod_{j=1}^{i} g_{t-j}(\gamma, c)$, a.s., $\Im_t = \sigma(\varepsilon_\tau : \tau \leq t)$, and $x_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} + V_t$ where $V_t = b\phi/(1 - a\phi)$, $Z_t = \varepsilon_t$, $\psi_0 = 1$ and for $i \geq 1$*

$$Z_{t-i} = \{\varepsilon_{t-i} g_{t-i}(\gamma, c) - b\} + \{g_{t-i}(\gamma, c) - a\} \times \sum_{j=i+1}^{\infty} \phi^{j-i} \varepsilon_{t-j} \prod_{k=i+1}^{j} g_{t-k}(\gamma, c),$$

$$\psi_i = \phi^i a^{i-1}, \quad i = 1, 2, \ldots,$$

where

$$a = \mathrm{E}\left\{g_t(\gamma, c)\right\} \in [-1, 1], \quad b = \mathrm{E}\left\{\varepsilon_t g_t(\gamma, c)\right\}, \quad \mathrm{E}\left(Z_{t-i} \mid \mathfrak{S}_{t-i-1}\right) = 0.$$

If $\{Z_t\}$ is L_p-strong orthogonal for some $p \leq 2$, then $P(x_{t+h} \mid \mathfrak{S}_t) = \sum_{i=h}^{\infty} \psi_{t+h,i} Z_{t+h,t+h-i}.$

Remark 10. It is straightforward to show $\mathrm{plim}_{N \to \infty} \mathrm{E}\left(x_t \mid \mathfrak{S}_{t-N}\right) = \mathrm{E}\left(x_t\right) = b\phi/(1 - a\phi)$ 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 $a = b = 0$, $Z_{t-i} = \varepsilon_{t-i}$, and $\psi_i = \phi^i$. If $\phi = 0$ then trivially $x_t = Z_t = \varepsilon_t$.

Remark 11. If $\{Z_t\}$ is strong orthogonal, we deduce the h-step ahead forecast

$$P(x_{t+h} \mid \mathfrak{S}_t) = \sum_{i=0}^{\infty} \phi^{i+h} a^{i+h-1} \Big[\left\{\varepsilon_{t-i} g_{t-i}(\gamma, c) - b\right\} + \left\{g_{t-i}(\gamma, c) - a\right\}$$

$$\times \sum_{j=i+1}^{\infty} \phi^{j-i} \varepsilon_{t-j} \prod_{k=i+1}^{j} g_{t-k}(\gamma, c) \Big].$$

5.2. Verifying weak and strong orthogonality.

For any t write

$$Z_t = \left\{\varepsilon_t g_t(\gamma, c) - b\right\} + \left\{g_t(\gamma, c) - a\right\} \times \sum_{j=1}^{\infty} \phi^j \varepsilon_{t-j} \prod_{k=1}^{j} g_{t-k}(\gamma, c). \tag{7}$$

Exploiting $\mathrm{E}\left(\varepsilon_t \mid \mathfrak{S}_{t-1}\right) = 0$ and the definitions of a and b, we know $\mathrm{E}\left(Z_t \mid \mathfrak{S}_{t-1}\right) = 0$, such that Z_t is weakly orthogonal to $\mathfrak{L}_{t-1}$ in some sense; see Cambanis, Hardin & Weron (1988). However, if we allow $\mathrm{E}\left(\varepsilon_t \mid \mathfrak{S}_{t-1}\right) \neq 0$ for both non-linear 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* $P(Z_t \mid \mathfrak{S}_{t-1}) = 0$ (see Lemma 1, part ii) which in L_p is true *if and only if* $\mathrm{E}\left(Z_t^{\langle p-1 \rangle} \mid \mathfrak{S}_{t-1}\right) = 0$. This may be easily tested for any chosen $p > 1$. For example, Hong & White (1995) develop a nuisance parameter-free consistent non-parametric test of functional form based on the observation that if $Y_{t-1} \equiv \mathrm{E}\left(Z_t^{\langle p-1 \rangle} \mid \mathfrak{S}_{t-1}\right) \neq 0$, then $\mathrm{E}\left(Z_t^{\langle p-1 \rangle} Y_{t-1}\right) = Y_{t-1}^2 > 0$. Essentially any non-parametric estimator $\widehat{Y}_{t-1}$ 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 (7), it follows that

$$\|Z_t\|_{4(p-1)} \leq |b| + (1 + |a|)| \times |\varepsilon_t\|_{4(p-1)}(1 - \phi)^{-1} < \infty,$$

hence $\mathrm{E} |Z_t^{\langle p-1 \rangle}|^4 < \infty$ and

$$\|Z_t^{\langle p-1 \rangle} - \mathrm{E}\left[Z_t^{\langle p-1 \rangle} \mid \mathfrak{S}_{t-1}\right]\|_{4(p-1)} < 2 \times \|Z_t^{\langle p-1 \rangle}\|_{4(p-1)} < \infty,$$

such that the moment conditions of Hong & White (1995) are satisfied. Along with fairly standard regulatory assumptions the test statistic is simple to compute and is based on the sample moment $n^{-1} \sum_{t=1}^{n} \widehat{Z}_t^{\langle p-1 \rangle} \widehat{Y}_{t-1}$ for some plug-in $\widehat{Z}_t$ to be detailed below. The statistic has an asymptotic standard normal null distribution. A multitude of alternative consistent parametric and non-parametric model specification tests exist; see Bierens & Ploberger (1997) and the citations therein.

For strong orthogonality in L_p, we need

$$\mathrm{E}\left\{\left(\sum_{k=0}^{h}\pi_k Z_{t+k}\right)^{\langle p-1\rangle}\,\bigg|\,\mathfrak{F}_{t-1}\right\}=0,\quad\forall h\geq 0,\quad\forall\pi\in\mathbf{R}^h.$$

A simple method follows: randomly generate $\pi\in\mathbf{R}^h$ for various $h=1,2,\ldots$, perform the non-parametric test on the resulting $\left(\sum_{k=0}^{h}\pi_k Z_{t+k}\right)^{\langle p-1\rangle}$, 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 $\widehat{Z}_t$ will be used as an obvious plug-in for Z_t. Because ε_t is unobservable simply assume $\varepsilon_t=0\ \forall t\leq 0$ and $x_0=0$; for the ESTAR model, e.g.,

$$\begin{aligned}
x_1 &= \varepsilon_1,\\
x_2 &= \phi x_1\exp\{-\gamma(|x_1|-c)^2\}+\varepsilon_2,\\
x_3 &= \phi x_2\exp\{-\gamma\big(|x_2-\phi x_1\exp\{-\gamma x_1^2\}|-c\big)^2\}+\varepsilon_3,
\end{aligned}$$

etc. Other methods for handling the first period may be considered as well. After constructing the regressor, the threshold model (6) can then be estimated straightforwardly by M-estimation or L_p-GMM for various $1<p<2$ using standard iterative estimation techniques, generating $\hat{\varepsilon}_t$ and $\widehat{Z}_t$. The moment condition is simply $\mathrm{E}\{\varepsilon_t^{\langle p-1\rangle}\partial/\partial\theta f_t(\theta)\}=0$, where $f_t(\theta)=\phi x_{t-1}g_{t-1}(\gamma,c)$ and $\theta=(\phi,\gamma,c)^{\top}$. See Arcones (2000), de Jong & Han (2002), and Han & 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 & Han (2002).

5.3. Non-linear impulse response function.

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

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

The response to the "sole" random impulse $V_n=\varepsilon_n$ is history $\{\varepsilon_{n-i}\}_{i=1}^{\infty}$ dependent, and asymmetric with respect to the sign of V_n through $g_n(\gamma,c)$. See Koop, Pesaran & Potter (1996) for further commentary on path dependence in non-linear IRFs.

If we use estimated residuals $\{\hat{\varepsilon}_t\}$ 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 $\hat{b}=n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t g_t(\hat{\gamma},\hat{c})$, we obtain a sample IRF based on one history and one impulse $\{\hat{\varepsilon}_t\}_{t=1}^{n}$:

$$\hat{I}(h,\hat{\varepsilon}_n,\{\hat{\varepsilon}_t\}_{t=1}^{n-1})=\hat{\phi}^h\hat{a}^{h-1}\Big[\{\hat{\varepsilon}_n g_n(\hat{\gamma},\hat{c})-\hat{b}\}+\{g_n(\hat{\gamma},\hat{c})-\hat{a}\}\times\sum_{j=1}^{n-1}\hat{\phi}^j\hat{\varepsilon}_{n-j}\prod_{k=1}^{j}g_{n-k}(\hat{\gamma},\hat{c})\Big].$$

Multiple alternative strategies for handling the random history $\{\varepsilon_{n-i}\}_{i=1}^{\infty}$ and impulse V_n are available. For example, we may randomly draw a history $\{\varepsilon_{n-i}\}_{i=1}^{n-1}$ and impulse v_n from the empirical distribution of the sample path $\{\hat{\varepsilon}_t\}_{t=1}^{n}$, or simulate independent impulses if the finite distributions of $\{\varepsilon_t\}$ 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 IRFs for each horizon h. Subsequent confidence bands and kernel densities may then be computed.

5.4. 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 extending from January 1, 2000, to August 31, 2005. 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. 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, Goldie & Teugels (1987). We study the tail shape of each series by computing the tail estimator $\hat{\alpha}$ due to B. M. Hill (1975). We apply asymptotic theory and a Newey–West-type kernel estimator of the asymptotic variance of $\hat{\alpha}$ for dependent, heterogenous processes, see J. B. Hill (2005b,c). Consult J. B. Hill (2005b) 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 $\{\widehat{Z}_t\}$ are computed according to

$$\widehat{Z}_t = \{\hat{\varepsilon}_t g_t(\hat{\gamma}, \hat{c}) - \hat{b}\} + \{g_t(\hat{\gamma}, \hat{c}) - \hat{a}\} \times \sum_{j=1}^{n-1} \hat{\phi}^j \hat{\varepsilon}_{t-j} \times \prod_{k=1}^{j} g_{t-k}(\hat{\gamma}, \hat{c}),$$

and the sample IRF $\hat{I}(h, V_n, \mathfrak{L}_{n-1})$ is computed accordingly.

We use the non-parametric method of Hong & White (1995) to test the daily returns $\{x_t\}$ for evidence of $E(x_t \mid \mathfrak{S}_{t-1}) = 0$ and $P(x_t \mid \mathfrak{S}_{t-1}) = 0$ (i.e., $E(x_t^{\langle p-1 \rangle} \mid \mathfrak{S}_{t-1}) = 0$). We also test $E(\varepsilon_t \mid \mathfrak{L}_{t-1}) = 0$ and $E(Z_t \mid \mathfrak{L}_{t-1}) = 0$ for evidence the STAR model represents the best L_2-predictor. We then test the STAR residuals $\{\hat{\varepsilon}_t\}$ and the estimated decomposition innovations $\{\widehat{Z}_t\}$ for evidence of weak and strong orthogonality by using $\{\hat{\varepsilon}_t^{\langle p-1 \rangle}\}$ and $\{\widehat{Z}_t^{\langle p-1 \rangle}\}$. For tests of strong orthogonality we use the series length $h = 10$ and 20 and randomly select $\pi \in \mathbf{R}^h$. For a non-parametric estimator of the conditional mean of $\{\hat{\varepsilon}_t^{\langle p-1 \rangle}\}$ and $\{\widehat{Z}_t^{\langle p-1 \rangle}\}$, 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 (Bierens 1990)

$$Y_{t-1}^* \equiv E(Y_t \mid \mathfrak{S}_{t-1}) = \theta_0 + \sum_{i=1}^{\infty} \theta_i \exp\left(\sum_{j=1}^{\infty} \tau_j x_{t-j}\right)$$

with probability one for some sequence $\{\theta_i\}$. We use

$$\widehat{Y}_{t-1}^* = \hat{\theta}_0 + \sum_{i=1}^{100} \hat{\theta}_i \exp\left(\sum_{j=1}^{t-1} \tau_j x_{t-j}\right),$$

where the τ_j are randomly selected from $\mathbf{R}$. We repeat the Hong–White test for 100 randomly selected values of π and 100 randomly selected values of τ (10,000 repetitions) and report the average P-value.

The test of the hypothesis $E(x_t \mid \mathfrak{S}_{t-1}) = 0$ due to Hong & White (1995) requires $E(x_t^4) < \infty$ which likely fails to hold for each exchange rate return x_t (see Table 1). Similarly, the finite-variance assumptions of Bierens (1990) 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 $E(\varepsilon_t \mid \mathfrak{S}_{t-1}) = 0$ and $E(Z_t \mid \mathfrak{S}_{t-1}) = 0$. The test of $P(x_t \mid \mathfrak{S}_{t-1}) = 0$, however, requires $E(x_t^{\langle p-1 \rangle})^4 < \infty$ 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.

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	$.013$ $(.004)$	$.015$ $(.003)$	$.020$ $(.009)$
$\max(|\varepsilon_t|)$	$.031$	$.025$	$.021$
	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	$.003$ $(.001)$	$.001$ $(.003)$	$.020$ $(.000)$
$\max(|\varepsilon_t|)$	$.030$	$.027$	$.021$
	Yen	Euro	BP
$\mathrm{E}\left(x_t \mid \Im_{t-1}\right) = 0^b$	$.058$	$.096$	$.020$
$\mathrm{P}(x_t \mid \Im_{t-1}) = 0^c$	$.061$	$.072$	$.097$
$\min(x_t)$	$-.030$	$-.025$	$-.019$
$\max(x_t)$	$.025$	$.027$	$.021$
$\hat{\alpha}_m$	$2.55\pm.70^d$	3.37 ± 1.20	2.96 ± 1.03

a. Parameter estimate (heteroscedasticity robust standard error).

b. The test of Hong White (1995) for conditional mean mis-specification: values denote p-values.

c. The null is $\mathrm{E}(x_t^{\langle p-1 \rangle} \mid \Im_{t-1}) = 0$, i.e., $\mathrm{P}\left(x_t \mid \Im_{t-1}\right) = 0$.

d. Tail index estimator and 95% interval length based on a Newey-West kernel estimator with Bartlett kernel: see Hill (2005b).

Tables 1 and 2 contain tail index estimates, L_p-GMM parameter estimates of the STAR models, and Hong–White test results. We comment only on the most pertinent results with respect to the concerns of this paper. First, 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. Second, for brevity we omit all results concerning the case $p = 1.1$ because parameter estimates are uniformly insignificant. Third, significant evidence suggests the daily log return x_t has some form of unmodeled (non)linear structure: $\mathrm{E}\left(x_t \mid \Im_{t-1}\right) \neq 0$ and $\mathrm{P}\left(x_t \mid \Im_{t-1}\right) \neq 0$ when $p = 1.5$ (see Table 1).

Fourth, when the LSTAR model is used it is only the Pound for which evidence suggests both $\mathrm{E}\left(\varepsilon_t \mid \Im_{t-1}\right) \neq 0$ and $\mathrm{E}(Z_t \mid \Im_{t-1}) \neq 0$ (Tests 1 and 3 of Table 2), as well as $\mathrm{P}\left(\varepsilon_{t+k} \mid \Im_{t-1}\right) = 0$, $\mathrm{P}\left(Z_{t+k} \mid \Im_{t-1}\right) = 0$ and $\mathrm{P}\left(\sum_{k=0}^{h} \pi_k Z_{t+k} \mid \Im_{t-1}\right) = 0$ when $p = 1.5$ (Tests 2, 4–6 of Table 2). Thus, the LSTAR model reasonably articulates the best $L_{1.5}$-predictor of the Pound resulting in strong orthogonal decomposition innovations in $L_{1.5}$. Theorem 3 can be immediately used to justify iterated projections for the Pound based on an LSTAR best $L_{1.5}$-predictor, and Theorem 5 and Lemma 6 justify the existence of a non-linear IRF. 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 ($\hat{\gamma} = 67.2$), the LSTAR model suggests extremes of the Pound are persistent, non-extremes are noisy, and regime change occurs quickly, matching evidence found in Hill (2005b).

TABLE 2: Hong–White tests[a].

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

a. The P-value of the test of Hong & White (1995).

b. The null is $\mathrm{E}(\varepsilon_{t+k}^{\langle p-1\rangle} \mid \Im_{t-1}) = 0$.

c. The null is $\mathrm{E}\{(\sum_{k=0}^{h} \pi_k \varepsilon_{t+k})^{\langle p-1\rangle} \mid \Im_{t-1}\} = 0$.

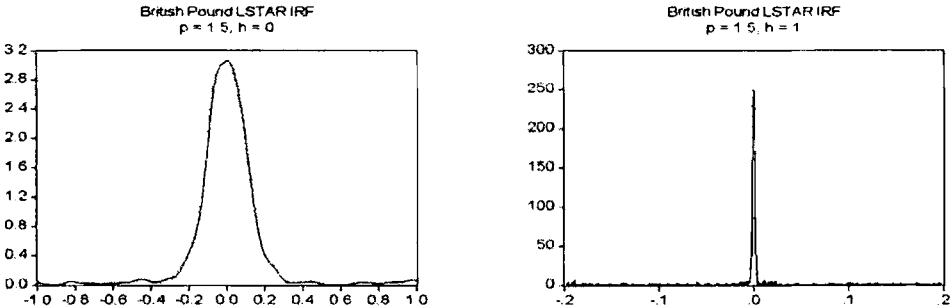d. The "h" in $\sum_{k=0}^{h} \pi_k \varepsilon_{t+k}$.

FIGURE 1: ESTAR IRF kernel densities (randomized impulses). *"h" reflects the number of steps ahead.*

Figures 1 and 2 display (Gaussian) kernel density functions of a sequence of non-linear and linear IRFs. We use the estimated residuals $\{\hat{\varepsilon}_t\}_{t=1}^{n-1}$ as one "draw" for the history, and randomly draw 500 independent and identically distributed impulses $\{v_{n,j}\}_{j=1}^{500}$ from a Pareto distribution, $\mathrm{P}\left(v_{n,j} > v\right) = v^{-\alpha}$ and $\mathrm{P}\left(v_{n,j} < -v\right) = (-v)^{-\alpha}$, $v > 0$, where $\hat{\alpha}$ is used as a plug-in for α. Sequences of IRFs are then generated for horizons $h = 0$ and 1 (when $h \geq 2$ nearly all of the probability mass of the IRFs occurs at zero). Figure 1 plots kernel densities of the IRFs for each h for the British Pound based on the LSTAR model. A prominent characteristic is the extremely heavy-tailed nature of the IRF empirical distribution; the estimated tail index of $\{\hat{I}(0, v_{n,j}, \{\hat{\varepsilon}_t\}_{t=1}^{n-1})\}_{j=1}^{500}$ is $.69 \pm .15$ due simply to the non-linear multiplicative presence of the shock history $\{\hat{\varepsilon}_t\}_{t=1}^{n-1}$. By comparison, a linear AR(1) IRF is $\phi^h \varepsilon_n$, which is Pareto distributed with index $\alpha = 2.96$. The estimated tail index of the sequence of linear IRFs is $3.24 \pm .30$ (which contains 2.96). See Figure 2.
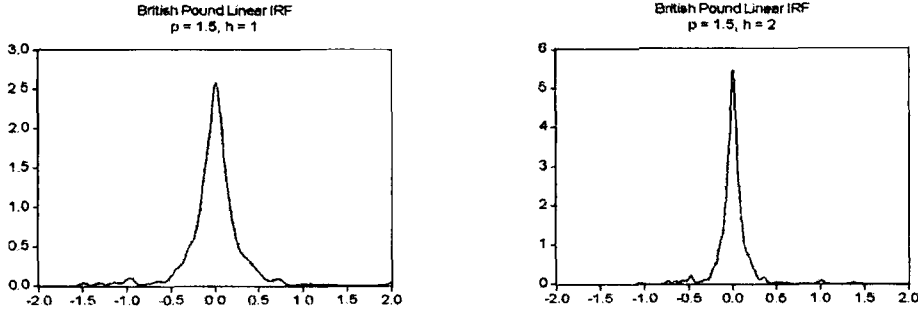
FIGURE 2: AR(1) IRF kernel densities (randomized impulses).

APPENDIX

Proof of Theorem 3. By Lemma A.1, below, for arbitrary integer $k > 0$ the finite decomposition holds,

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

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 $V_{n,k} \in \mathfrak{B}_{n-k}$, such that

$$X_n = Y_{n,k} + V_{n,k}.$$

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 $k \geq 1$

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

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 $\{V_{n,k_i}\}$; see the Bolzano–Weierstrass Theorem. In particular, define the stochastic limits as

$$\lim_{k_i \to \infty} Y_{n,k_i} = Y_n, \quad \lim_{k_i \to \infty} V_{n,k_i} = V_n.$$

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, (4) is proved.

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

Finally, the claim that every element $Y \in \bigoplus_{i=0}^{\infty} N_{n-i}$ 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 $\{N_{n-i}\}_{i=0}^{k-1}$ are weak orthogonal in that $N_{n-i} \perp B_{n-i-1}$ and $N_{n-i} \perp N_{n-j}$ for every $0 \leq i < j \leq k-1$. Moreover, the following are equivalent for any integer $k > 0$: (i) $B_n = \left(\bigoplus_{i=0}^{k-1} N_{n-i} \right) \oplus B_{n-k}$; and (ii) $P_{t,t-\ell}P_{t,t-k} = P_{t,t-\ell}$, for every $t, k \leq \ell$.*

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 $\bigoplus_{i=0}^{\infty} M_{n-i}$ 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 $X \in \bigoplus_{i=0}^{\infty} M_{n-i}$ 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 $N_t \perp \mathfrak{B}_{t-1}$, (see Lemma 2, part 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

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

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

$$
\begin{aligned}
\mathfrak{B}_n &= N_n + P_{n,n-1}\mathfrak{B}_n = N_n + \mathfrak{B}_{n-1} \\
&= N_n + N_{n-1} + \mathfrak{B}_{n-2} = \cdots \\
&= \sum_{i=0}^{k-1} N_{n-i} + \mathfrak{B}_{n-k},
\end{aligned}
$$

where for each $t \leq n$, $N_t \perp \mathfrak{B}_{t-1}$. 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 $0 \leq i < j \leq k$, it follows that $N_{n-i} \perp N_{n-j}, 0 \leq i < j \leq k$.

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

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

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

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

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

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

$$= P_{n,n-t}\left(\sum_{i=s}^{t-1}\xi_{n-i}\right) + \sum_{i=t}^{k-1}\xi_{n-i} + V_{n,k}$$

$$= \sum_{i=t}^{k-1}\xi_{n-i} + V_{n,k}.$$

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) $\Rightarrow$ (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 $1 \leq j < k$,

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

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 $0 \leq s \leq t \leq k$,

$$P_{n,n-k}\left(\sum_{i=0}^{k-1}\xi_{n-i}\right) = P_{n,n-k}P_{n,n-1}\left(\sum_{i=0}^{k-1}\xi_{n-i}\right)$$

$$= P_{n,n-k}\left(P_{n,n-1}\xi_n + \sum_{i=1}^{k-1}\xi_{n-i}\right)$$

$$= P_{n,n-k}\left(\sum_{i=1}^{k-1}\xi_{n-i}\right),$$

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

$$P_{n,n-k}\left(\sum_{i=0}^{k-1}\xi_{n-i}\right) = P_{n,n-k}\left(\sum_{i=1}^{k-1}\xi_{n-i}\right)$$

$$= P_{n,n-k}\left\{P_{\mathfrak{B}_{n-2}}\left(\sum_{i=1}^{k-1}\xi_{n-i}\right)\right\}$$

$$= P_{n,n-k}\left(P_{\mathfrak{B}_{n-2}}\xi_{n-1} + \sum_{i=2}^{k-1}\xi_{n-i}\right)$$

$$= P_{n,n-k}\left(\sum_{i=2}^{k-1}\xi_{n-i}\right) = \cdots = 0.$$

Therefore, $\sum_{i=0}^{k-1}\xi_{n-i}\perp\mathfrak{B}_{n-k}$ (see Lemma 2, part 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 $1 \leq j \leq k$, we conclude that $\sum_{i=0}^{j-1}N_{n-i}\perp\sum_{i=j}^{k-1}N_{n-i}$. It follows that $\sum_{i=0}^{k-1}N_{n-i} = \bigoplus_{i=0}^{k-1}N_{n-i}$, and $\mathfrak{B}_n = (\bigoplus_{i=0}^{k-1}N_{n-i})\oplus\mathfrak{B}_{n-k}$, which proves (ii) $\Rightarrow$ (i).

Proof of Lemma A.2. Consider an arbitrary sequence of elements $\{x_j\}, x_j \in M_j$, and recall $\bigoplus_{i=0}^{\infty}M_{n-i}$ 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 $s \leq t$,

$$\left\| \sum_{i=0}^{s} \lambda_j x_{n-j} \right\| \leq K \left\| \sum_{i=0}^{t} \lambda_j x_{n-j} \right\|. \tag{8}$$

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

$$\bigoplus_{i=0}^{s} M_{n-i} \perp \bigoplus_{i=s+1}^{t} M_{n-i}.$$

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

$$\sum_{i=0}^{s} \lambda_j x_{n-j} \perp \sum_{i=s+1}^{t} \lambda_j x_{n-j}. \tag{9}$$

By the definition of James orthogonality, it follows from (9) that

$$\left\| \sum_{i=0}^{s} \lambda_j x_{n-j} + a \sum_{i=s+1}^{t} \lambda_j x_{n-j} \right\| \geq \left\| \sum_{i=0}^{s} \lambda_j x_{n-j} \right\|$$

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

$$\left\| \sum_{i=0}^{t} \lambda_j x_{n-j} \right\| = \left\| \sum_{i=0}^{s} \lambda_j x_{n-j} + \sum_{i=s+1}^{t} \lambda_j x_{n-j} \right\| \geq \left\| \sum_{i=0}^{s} \lambda_j x_{n-j} \right\|.$$

The inequality in (8) follows with $K = 1$, which proves the result.

Proof of Theorem 5. By assumption $x_t = \sum_{i=0}^{\infty} \psi_{t,i} Z_{t,t-i}$ where the $Z_{t,t-i}$ 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 $\mathfrak{L}_t \subseteq \tilde{\mathfrak{L}}_t$, $P(x_t \mid \mathfrak{L}_{t-1}) = \sum_{i=1}^{\infty} \psi_{t,i} Z_{t,t-i}$ and

$$P(x_t \mid \tilde{\mathfrak{L}}_{t-1}) = \sum_{i=1}^{\infty} \psi_{t,i} Z_{t,t-i} + \psi_{t,0} P(Z_{t,t} \mid \tilde{\mathfrak{L}}_{t-1}),$$

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

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

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

$$\begin{aligned}
I(1, v_t, \mathfrak{L}_{t-1}) &= P(x_{t+1} \mid \tilde{\mathfrak{L}}_{t-1}) - P(x_{t+1} \mid \mathfrak{L}_{t-1}) \\
&= \sum_{i=2}^{\infty} \psi_{t+1,i} Z_{t+1,t+1-i} + P(Z_{t+1,t+1} + \psi_{t+1,1} Z_{t+1,t} \mid \tilde{\mathfrak{L}}_{t-1}) \\
&\quad - \sum_{i=2}^{\infty} \psi_{t+1,i} Z_{t+1,t-i} \\
&= P(Z_{t+1,t+1} + \psi_{t+1,1} Z_{t+1,t} \mid \tilde{\mathfrak{L}}_{t-1}).
\end{aligned}$$

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

$$
\begin{aligned}
P\big(Z_{t+1,t+1} + \psi_{t+1,1}Z_{t+1,t} \mid \tilde{\mathfrak{L}}_{t-1}\big) &= P\big\{P\big(Z_{t+1,t+1} + \psi_{t+1,1}Z_{t+1,t} \mid \mathfrak{L}_t\big) \mid \tilde{\mathfrak{L}}_{t-1}\big\} \\
&= P\big\{\psi_{t+1,1}Z_{t+1,t} + P\big(Z_{t+1,t+1} \mid \mathfrak{L}_t\big) \mid \tilde{\mathfrak{L}}_{t-1}\big\} \\
&= \psi_{t+1,1}P\big(Z_{t+1,t} \mid \tilde{\mathfrak{L}}_{t-1}\big).
\end{aligned}
$$

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

$$
\begin{aligned}
I(2, v_t, \mathfrak{L}_{t-1}) &= P\big(x_{t+2} \mid \tilde{\mathfrak{L}}_{t-1}\big) - P\big(x_{t+2} \mid \mathfrak{L}_{t-1}\big) \\
&= \sum_{i=3}^{\infty} \psi_{t+2,i}Z_{t+2,t+2-i} \\
&\quad + P\big(Z_{t+2,t+2} + \psi_{t+2,1}Z_{t+2,t+1} + \psi_{t+2,2}Z_{t+2,t} \mid \tilde{\mathfrak{L}}_{t-1}\big) \\
&\quad - \sum_{i=3}^{\infty} \psi_{t+2,i}Z_{t+2,t+2-i} \\
&= P\big(Z_{t+2,t+2} + \psi_{t+2,1}Z_{t+2,t+1} + \psi_{t+2,2}Z_{t+2,t} \mid \tilde{\mathfrak{L}}_{t-1}\big) \\
&= P\Big[P\big\{P\big(Z_{t+2,t+2} + \psi_{t+2,1}Z_{t+2,t+1} + \psi_{t+2,2}Z_{t+2,t} \mid \mathfrak{L}_{t+1}\big) \mid \mathfrak{L}_t\big\} \mid \tilde{\mathfrak{L}}_{t-1}\Big] \\
&= \psi_{t+2,2}P\big(Z_{t+2,t} \mid \tilde{\mathfrak{L}}_{t-1}\big),
\end{aligned}
$$

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

Proof of Lemma 6. Clearly $\tilde{\mathfrak{L}}_{t-k-1} = \mathfrak{L}_{t-k-1} \oplus \varepsilon_{t-k} = \mathfrak{L}_{t-k}$. Hence

$$
P\big\{P(x_t \mid \mathfrak{L}_{t-k}) \mid \tilde{\mathfrak{L}}_{t-k-1}\big\} = P\big\{P(x_t \mid \mathfrak{L}_{t-k}) \mid \mathfrak{L}_{t-k}\big\} = P(x_t \mid \mathfrak{L}_{t-k}) = P(x_t \mid \tilde{\mathfrak{L}}_{t-k-1}).
$$

If additionally x_t admits a strong orthogonal decomposition, then by Theorem 3 and the identity $\tilde{\mathfrak{L}}_{t-k-1} = \mathfrak{L}_{t-k-1} \oplus \varepsilon_{t-k} = \mathfrak{L}_{t-k}$, we deduce

$$
P\big\{P(x_t \mid \tilde{\mathfrak{L}}_{t-k}) \mid \mathfrak{L}_{t-k}\big\} = P\big\{P(x_t \mid \mathfrak{L}_{t-k+1}) \mid \mathfrak{L}_{t-k}\big\} = P(x_t \mid \mathfrak{L}_{t-k}).
$$

ACKNOWLEDGEMENTS

REFERENCES

M. A. Arcones (2000). M-Estimators converging to a stable limit. *Journal of Multivariate Analysis*, 74, 193–221.

B. Basrak, R. A. Davis & T. Mikosch (2001). *Regular Variation and GARCH Processes*. Unpublished repport, Forsikringsmatematisk Laboratorium, Københavns Universitet, Copenhagen, Denmark.

B. Basrak, R. A. Davis & T. Mikosch (2002). A characterization of multivariate regular variation. *The Annals of Applied Probability*, 12, 908–920.

H. J. Bierens (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443–1458.

H. J. Bierens & W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65, 1129–1151.

N. H. Bingham, C. M. Goldie & J. L. Teugels (1987). *Regular Variation*. Cambridge University Press.

S. Cambanis, D. Hardin & A. Weron (1988). Innovations and Wold decompositions of stable sequences. *Probability Theory and Related Fields*, 79, 1–27.

Y. W. Cheung (1993). Long memory in foreign exchange rates. *Journal of Business and Economic Statistics*, 11, 93–101.

D. B. H. Cline (1983). *Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data*. Unpublished doctoral dissertation, Department of Statistics, Colorado State University, Fort Collins, CO.

R. A. Davis & S. Resnick (1985a). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, 13, 179–195.

R. A. Davis & S. Resnick (1985b). More limit theory for the sample correlation-function of moving averages. *Stochastic Processes and Their Applications*, 20, 257–279.

R. de Jong & C. Han (2002). Properties of L_p-GMM estimators. *Econometric Theory*, 18, 491–504.

R. Engle & V. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48, 1749–1778.

B. Falk & C. Wang (2003). Testing long-run PPP with infinite-variance returns. *Journal of Applied Econometrics*, 18, 471–484.

G. D. Faulkner & J. E. Huneycutt Jr (1978). Orthogonal decompositions of isometries in a Banach Space. *Proceedings of the American Mathematical Society*, 69, 125–128.

P. H. Franses & D. van Dijk (2000). *Non-linear Time Series Models in Empirical Finance*. Cambridge University Press.

R. Gallant, P. Rossi & G. Tauchen (1993). Nonlinear dynamic structures. *Econometrica*, 61, 871–907.

J. R. Giles (1967). Classes of semi-inner-product spaces. *Transactions of the American Mathematical Society*, 129, 436–446.

J. R. Giles (2000). *Introduction to the Analysis of Normed Linear Spaces*. Cambridge University Press.

G. González-Rivera (1998). Smooth-transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

C. Gouriéroux & J. Jasiak (2003). *Nonlinear Innovations and Impulse Responses with an Application to VaR Sensitivity*. Unpublished report, Centre de recherche en économie et statistique, Institut national de la statistique et des études économiques, Paris, France.

C. Han & R. de Jong (2004). Closest moment estimation under general conditions. *Annales d'économie et de statistique*, 74, 1–13.

C. Hardin, G. Samorodnitsky & M. Taqqu (1991). Non-linear regression of stable random variables. *The Annals of Applied Probability*, 1, 582–612.

B. M. Hill (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Mathematical Statistics*, 3, 1163–1174.

J. B. Hill (2005a). Strong Orthogonal Decompositions and Projection Iterations in Banach Space. Available at http://www.fiu.edu/~hilljona/strong_orthog_banach_wp.pdf Department of Economics, Florida International University, Miami.

J. B. Hill (2005b). Gaussian Tests of "Extremal White Noise" for Dependent, Heterogeneous, Heavy Tailed Time Series with an Application. Available at http://www.fiu.edu/~hilljona/co-relation-test.pdf Department of Economics, Florida International University, Miami.

J. B. Hill (2005c). On Tail Index Estimation Using Dependent, Heterogenous Data. Unpublished report available at http://econwpa.wustl.edu:80/eps/em/papers/0505/0505005.pdf Department of Economics, Florida International University, Miami.

M. C. A. B. Hols & C. G. de Vries (1991). The limiting distribution of extremal exchange rate returns. *Journal of Applied Econometrics*, 6, 287–302.

Y. Hong & H. White (1995). Consistent specification testing via nonparametric series regression. *Econometrica*, 63, 1133–1159.

R. C. James (1947). Orthogonality and linear functionals in a normed linear space. *Transactions of the American Mathematical Society*, 61, 265–292.

G. K. Kees & C. J. M. Kool (1992). Tail estimates of East European exchange rates. *Journal of Business and Economic Statistics*, 10, 83–96.

P. S. Kokoszka & M. S. Taqqu (1994). Infinite variance stable ARMA processes. *Journal of Time Series Analysis*, 15, 203–220.

P. S. Kokoszka & M. S. Taqqu (1996). Infinite variance stable moving averages with long memory. *Journal of Econometrics*, 73, 79–99.

G. Koop, M. H. Pesaran & S. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 4, 119–147.

W. L. Lin (1997). Impulse response function for conditional volatility in GARCH models. *Journal of Business and Economic Statistics*, 15, 15–25.

J. Lindenstrauss & L. Tzafriri (1977). *Classical Banach Spaces*. Springer, New York.

M. Loretan & 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, 211–248.

S. Lundbergh, T. Teräsvirta & D. van Dijk (2003). Time-varying smooth transition autoregressive models. *Journal of Business and Economic Statistics*, 21, 104–121.

S. Lundbergh & T. Teräsvirta (2006). A time series model for an exchange rate in a target zone with applications. *Journal of Econometrics*, 131, 579–609.

D. McMillan & A. Speight (2002). Nonlinear dependence in inter-war exchange rates: Some further evidence. *Applied Economics Letters*, 9, 359–364.

R. Megginson (1998). *An Introduction to Banach Space Theory*. Springer, New York.

A. G. Miamee & M. Pourahmadi (1988). Wold decompositions, prediction and parameterization of stationary processes with infinite variance. *Probability and Related Fields*, 79, 145–164.

P. Micheal, A. R. Nobay & D. A. Peel (1997).Transaction costs and nonlinear adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy*, 105, 862–879.

T. Mikosch & C. Stărică (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process. *The Annals of Statistics*, 28, 1427–1451.

P. C. B. Phillips J. McFarland & P. McMahon (1996). Robust tests of forward exchange market efficiency with empirical evidence from the 1920s. *Journal of Applied Econometrics*, 11, 1–22.

P. Saikkonen & R. Luukkonen (1988). Lagrange multiplies tests for testing nonlinearities in time series models. *Scandinavian Journal of Statistics*, 15, 55–68.

C. Sims (1980). Macroeconomics and reality. *Econometrica*, 48, 1–48.

I. Singer (1970). *Bases in Banach Spaces*. Springer, New York.

T. Teräsvirta (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89, 208–218.

H. Tong (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press.

K. Urbanic (1964). Prediction of strictly stationary sequences. *Colloquium Mathematicum*, 12, 115–129.

K. Urbanic (1967). Some prediction problems for strictly stationary processes. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, June 21–July 18, 1965; December 27, 1965–January 7, 1966* (L. M. Le Cam & J. Neyman, eds.), University of California Press, Berkeley, Vol. II: Contributions to Probability, Part 1, pp. 235–258.

D. van Dijk, T. Teräsvirta & P. H. Franses (2000). *Smooth Transition Autoregressive Models—A Survey of Recent Developments*. Econometric Institute Research Report EI2000-23/A, Erasmus Universiteit, Rotterdam, The Netherlands.

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

Jonathan B. HILL: jonathan.hill@fiu.edu
Department of Economics
Florida International University
University Park, DM-307b, Miami, FL 33199, USA