

NONLINEAR IMPULSE RESPONSE FUNCTIONS

SIMON M. POTTER *

*Domestic Research Function, Federal Reserve Bank of New York.
33 Liberty St. New York, NY 10045 U.S.A.
E-mail: simon.potter@ny.frb.org*

December 1998

ABSTRACT

The standard linear technique of impulse response function analysis is extended to the nonlinear case by defining a generalized impulse response function. Measures of persistence and asymmetry in response are constructed for a wide class of time series.

Key Words: Nonlinear, Impulse Response Function, Persistence, Generalized Impulse Response Function.

JEL: C22,C51,C52,E32.

*I would like to thank two anonymous referees, William A. Brock, Gary Koop, Hashem Pesaran, colleagues at UCLA, and seminar participants for numerous helpful suggestions and comments. Financial Support from the Graduate School of the University of Wisconsin, the Academic Senate of UCLA, the NSF under grant SES 9211726 and the Center for Computable Economics at UCLA is gratefully acknowledged. The views expressed in this paper are those of the author and do not necessarily reflect the views of the Federal Reserve bank of New York or the Federal Reserve System

1 Introduction

Macroeconomists engage in considerable debate over the time series properties of macroeconomics variables as represented in impulse response functions. The debate until recently has been exclusively based on the foundation that the economy's dynamic behavior is well approximated by (Gaussian) random impulses being propagated over time by an invariant linear structure. As discussed by Gallant, Rossi and Tauchen (1993) (GRT hereafter) there is no reason to restrict the analysis of dynamics to linear time series. They extend the notion of impulse response functions to the nonlinear time series case. This paper further develops the theory of impulse response functions for nonlinear time series.¹

GRT considered the extension of one common definition of impulse response functions for linear time series to the nonlinear case. I extend the definitions of the four types of impulse response functions in the macroeconomic literature. For linear time series the four definitions all contain the same information. In the nonlinear case they all contain different information. Call this the first reporting problem.² Instead of developing techniques for all four definitions I resolve the first reporting problem by arguing that one of the definitions has a number of advantages over the other three in the nonlinear context and call it a Generalized Impulse Response Function (GI hereafter).

If a time series is linear one can normalize the GI to produce a nonrandom function. It is not possible to construct such a nonrandom function for the GI of a nonlinear time series. Thus, there is a second reporting problem for nonlinear time series: Is the reported response a very general feature of the dynamics of the nonlinear time series? GRT took an informal approach to the second reporting problem for their definition of impulse response functions by considering visually the behavior of bundles of impulse response functions. I am able to formalize this approach by treating the GI as a random variable on the underlying probability space of the time series. This property of the GI is used to develop a number of solutions to the second reporting problem based on stochastic dominance measures of the size of random variables.

The outline of the rest of the paper is as follows. Section 2 develops some notation and some conventions. Section 3 extends the four standard definitions of linear impulse response

¹I concentrate exclusively on the univariate case. Koop, Pesaran and Potter (1996) extend the use of the impulse response functions as defined in this paper to the case of multiple time series.

²The reporting problems I examine in this paper are in addition to that usually encountered in linear impulse response function analysis. When presenting impulse response functions from estimated models there is sampling uncertainty associated with parameters defining the impulse response function. One set of possible classical solutions are described by GRT. Another set of possible Bayesian solutions are described by Koop (1995).

functions to the nonlinear case. Section 4 defines a special nonlinear time series model and illustrates the restrictions imposed by the use of Wold Representation of a time series. Section 5 develops the random variable interpretation of the GI. Section 6 provides a number of definitions of persistence. Section 7 considers various conditional versions of the GI. Section 8 contains an empirical application to persistence in U.S. GNP. Section 9 gives some brief conclusions. A short appendix contains proofs of results in the main text.

2 Notation and Conventions

All random variables are assumed to be defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Throughout the paper the convention will be maintained that lower case letters refer to a realization of a random variable (unless the argument ω of the random variable is made explicit) and upper case letters to the random variable. In order to clarify the notation with respect to the expectations operator, the double bar convention of Billingsley (1986) will be used. Thus, $E[Y_{t+n}||\mathcal{G}]$ will mean the random variable called the conditional expected value of the integrable random variable Y_{t+n} given that \mathcal{G} is a σ -field in \mathcal{F} such that

1. $E[Y_{t+n}||\mathcal{G}]$ is measurable \mathcal{G} and integrable.
2. $E[Y_{t+n}||\mathcal{G}]$ satisfies the functional equation:

$$\int_G E[Y_{t+n}||\mathcal{G}]dP = \int_G Y_{t+n}dP, \quad \forall G \in \mathcal{G}.$$

For example, if $\mathcal{G} = \{0, \Omega\}$ then clearly measurability and integrability are satisfied and $E[Y_{t+n}||\mathcal{G}] = E[Y_{t+n}]$ since this σ -field contains no information on individual realizations. Below it will be important to distinguish between realizations of the conditional expectation operator and the random variable itself. The value of $E[Y_{t+n}||\mathcal{G}]_\omega$ at $\omega \in \Omega$ is interpreted as the expected value of Y_{t+n} given that one knows for each $G \in \mathcal{G}$ whether or not it contains the point ω .

The focus will be on the sequence of increasing Borel (\mathcal{B}) σ -fields,

$$\mathcal{F}_t = \sigma(\dots, Y_{-1}, Y_0, Y_1, \dots, Y_t).$$

Hence $E[Y_{t+n}||\mathcal{F}_t]_\omega$ is the value of the conditional expectation of Y_{t+n} given information up to time t . For $n \leq 0$ the value will be the observed realization of the time series, for $n > 0$ we will denote its value by $E[Y_{t+n}|y^t(\omega)]$. Here $y^t(\omega)$ identifies the individual realization of the time series up to

time t at ω , i.e., $y^t(\omega) = (y_t, \dots, y_1, y_0, y_{-1}, \dots)$. $y^\infty(\omega)$ will represent a complete realization of the time series.

In order to simplify the exposition I will concentrate on the case of strictly stationary time series, although some of the basic definitions below will also apply to the nonstationary case with minor modification. Nonstationarity due to deterministic components or integration of stationary time series will be directly allowed for. V_t will represent independent and identically distributed random sequences, W_t will represent martingale difference sequences, U_t will represent sequences of uncorrelated random variables (white noise).

We will use the notation $X \succ_{SSD} Y$ for the statement that the random variable X second order stochastically dominates the random variable Y and $X \sim_{SSD} Y$ for the statement that X does not second order stochastically dominate Y .³ Similarly we use the notation $X \succ_{FSD} Y$ for the statement that the random variable X first order stochastically dominates the random variable Y and $X \sim_{FSD} Y$ for the statement that X does not first order stochastically dominate Y .⁴

The Wold Decomposition of any covariance stationary time series is given by:

$$Y_t = \eta(t) + \sum_{i=0}^{\infty} \psi_i U_{t-i},$$

where $\sum_{i=0}^{\infty} \psi_i^2 < \infty$, $\psi_0 = 1$ and $\eta(t)$ is a purely deterministic covariance stationary sequence. The Wold Representation is the time series minus the purely deterministic component.

3 Four Definitions of Impulse Response Functions

3.1 Linear Impulse Response Functions

Define the operator $L_n(y^t(\omega))$ as the linear predictor of Y_{t+n} given the realization of (y_t, y_{t-1}, \dots) . Using the Wold Decomposition the linear predictor is given by:

$$L_n(y^t(\omega)) = \eta(t+n) + \sum_{i=0}^{\infty} \psi_{n+i} u_{t-i}$$

³If X and Y are two random variables with the same mean and distribution functions F_X, F_Y respectively then X second order stochastically dominates Y if and only if $\int_{\lambda \geq \alpha} F_X(\lambda) - F_Y(\lambda) d\lambda \geq 0$, $\forall \alpha \in R$. Second Order Stochastic Dominance is an example of a partial ordering on the a set of distribution functions. A partial ordering on a given set D of distribution functions, F_1, F_2, F_3, \dots is a binary relation satisfying three axioms:

1. $F_1 \prec F_1$ (reflexivity).
2. If $F_1 \prec F_2$ and $F_2 \prec F_3$ then $F_1 \prec F_3$ (transitivity).
3. If $F_1 \prec F_2$ and $F_2 \prec F_1$ then $F_1 = F_2$ (anti-symmetry)

When applied directly to random variables as done here the last statement defines an equivalence class rather than equality.

⁴If X and Y are two random variables with distribution functions F_X, F_Y respectively then X first order stochastically dominates Y if and only if $F_X(\alpha) - F_Y(\alpha) \leq 0$, $\forall \alpha \in R$.

Consider the following three operations on the linear predictor:

1. The linear impulse response function ($lirf_n$), it answers the question how does my linear forecast of Y_{t+n} change if today's realization of y_t changes by δ :

$$lirf_n(\delta) = L_n(y_t + \delta, y_{t-1}, \dots) - L_n(y_t, y_{t-1}, \dots) = \psi_n \delta.$$

2. Instead of considering the perturbation, δ , one can examine the derivative of the linear predictor ($dlirf_n$). This gives the slope of the linear forecast of Y_{t+n} with respect to y_t . In the discussion of persistence of univariate time series the limiting behavior of this derivative has been of great interest.

$$dlirf_n = \frac{\partial L_n(y_t, y_{t-1}, \dots)}{\partial y_t} = \psi_n.$$

3. The linear updating function ($ludf_n$) is a crucial part of linear rational expectations models where the change in today's behavior is determined by the change in linear forecasts of the future given the arrival of information at time t :

$$ludf_n(U_t(\omega)) = L_n(y^t(\omega)) - L_{n+1}(y^{t-1}(\omega)) = \psi_n U_t(\omega).$$

It is trivial to defined scaled versions of $lirf_n$, $ludf_n$ that are equal to $lirf_n$ and remove any randomness from the linear impulse response functions. All of the three functions share a symmetry property with respect to the value of the shock and an independence from the realization of the process before time t .

4. All three are also equivalent to what one might call the **sample path operator**. Consider two sample paths of $\{U_t\}$, $u^\infty(\omega), u^\infty(\omega')$, that are same except at time t when they differ by δ .

$$Y_t(\omega) = \eta(t) + \sum_{i=1}^{\infty} \psi_i U_{t-i}(\omega) + U_t(\omega),$$

$$Y_t(\omega') = \eta(t) + \sum_{i=1}^{\infty} \psi_i U_{t-i}(\omega') + U_t(\omega'),$$

$$y_{t+n}(\omega) - y_{t+n}(\omega') = \psi_n (U_t(\omega) - U_t(\omega')) = \psi_n \delta.$$

This is a definition of an impulse response function that is used by economists examining linear models under perfect foresight where the difference in the sample path reflects an 'unanticipated shock'.

3.2 Nonlinear Impulse Response Functions

The most difficult operation to extend to the nonlinear case in a rigorous manner is the sample path one. Assume that there exists a sequence of IID random variables $\{V_t\}$ defined on the same

probability space as the time series such that $\sigma(V_t, V_{t-1}, \dots) = \mathcal{F}_t$. Define two infinite sequences of the realized values of the IID random variable, $v^\infty(\omega), v^\infty(\omega')$, that differ only in their t -th element by δ . Then a formal definition of the sample path operator would be:

$$E[Y_{t+n}|v^\infty(\omega)] - E[Y_{t+n}|v^\infty(\omega')].$$

Suppose $Y_t = f(Y_{t-1}) + V_t$, where $f(\cdot)$ is a non-affine function. Consider the value of the sample path operator at $n = 2$:

$$Y_t(\omega) = f(Y_{t-1}(\omega)) + V_t(\omega), \quad Y_t(\omega') = f(Y_{t-1}(\omega')) + V_t(\omega'),$$

$$Y_{t+2}(\omega) - Y_{t+2}(\omega') = f(f(Y_{t-1}(\omega)) + V_t(\omega)) + V_{t+1}(\omega) - f(f(Y_{t-1}(\omega')) + V_t(\omega')) + V_{t+1}(\omega'),$$

Since $f(\cdot)$ is non-affine the response depends on the value of V_{t+1} . This effect is present for all $n > 1$ and implies that the response to the perturbation δ depends on the future sequence of shocks. One possible choice would be to look at all $\omega \in \Omega$ such that the innovation sequence is zero after time t (see Beaudry and Koop (1993) for an example). This would be equivalent to analyzing the behavior of a nonlinear difference equation over a set of initial conditions defined by the distribution of the time series at time t . This approach is not followed for two important and related reasons.

- a. The behavior of a nonlinear difference equation can be very sensitive on the exact value of the parameters. For example, small changes in the parameters can alter the limiting dynamic behavior of the difference equation from a unique value to chaos. This is a very undesirable property when the models under investigation are subject to estimation error.
- b. We will need a probability measure to discuss the general dynamic properties of nonlinear stochastic models. The measure to associate with the behavior of the sample path operator using the zero innovation sequence is not as obvious since a nonstationarity is introduced by the special sample path restriction.

The natural thing to do to avoid these problems, is to average out the effect of future shocks, thus reducing the sample path operator to the nonlinear impulse response function.

We continue by replacing the operator L_n in the other three categories of impulse response functions with the conditional expectation at time t of Y_{t+n} given the sigma field \mathcal{F}_t generated by $\{Y_t, Y_{t-1}, \dots\}$ evaluated at a particular point $y^t(\omega)$.

1. A Nonlinear Impulse Response Function is defined by the difference between an expectation conditioned on the sample path $y^t(\omega')$ and an expectation conditioned on the sample path $y^t(\omega)$. Where $y^t(\omega')$ is equal to $y^t(\omega)$ except for the y_t element which is perturbed by δ :

$$nlirf_n(\delta, y^t(\omega)) = E[Y_{t+n}|y^t(\omega')] - E[Y_{t+n}|y^t(\omega)].$$

This is the definition used by GRT.

2. The ‘derivative’ of the $nlirf_n$ (conditional expectation) of Y_{t+n} with respect to y_t , assuming that it is well-defined at the point (y_t, y_{t-1}, \dots) :

$$dnlirf_n(y^t(\omega)) = \lim_{\delta \rightarrow 0} \frac{nlirf_n(\delta, y^t(\omega))}{\delta}.$$

If the derivative is not well-defined at this point then the $dnlirf_n$ will be set to infinity.

3. The Nonlinear Updating Function is defined as

$$nludf_n(y^t(\omega)) = E[Y_{t+n}|y^t(\omega)] - E[Y_{t+n}|y^{t-1}(\omega)].$$

4 Dynamics Hidden by the Wold Representation

4.1 A Canonical Nonlinear Time Series Model

Unlike the linear case there is no general causal representation of a nonlinear time series in terms of functions of IID random variables. The closest one can get is a Volterra Series Expansion. A special case of which is:

$$Y_t = \sum_{i=0}^{\infty} \psi_i V_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_{ij} V_{t-i} V_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{ijk} V_{t-i} V_{t-j} V_{t-k}.$$

with V_t IID and symmetrically distributed.

Recall that nonlinear time series do have Wold Representations. I examine the information lost by considering only dynamics present in the Wold Representation (i.e., assuming linearity) when a time series has a cubic Volterra Series Expansion.⁵ In order to highlight the loss of information I restrict the Cubic Volterra Series model so that up to scale the first term in the expansion is the Wold Representation. Under conditions given in the appendix, one can define a

⁵A cubic expansion is used because quadratic expansions are special in the sense that they share some of the properties of linear time series.

new stochastic process by:

$$U_t = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{ij} V_{t-i} V_{t-j} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_{ijk} V_{t-i} V_{t-j} V_{t-k} + V_t,$$

such that the Wold Representation of Y_t defined by a cubic Volterra series model is given by:

$$\sum_{i=0}^{\infty} h_i U_{t-i},$$

where $h_i = \psi_i / \psi_0$, $h_0 = 1$ and U_t is white noise.

4.2 Examples of Nonlinear Impulse Response Functions

In this section I give example of realizations of the impulse response functions produced by the cubic Volterra model. The calculations for the Volterra series expansion are based on the conditioning sequence v_t, v_{t-1}, \dots .⁶

1. The Nonlinear Impulse Response Function

$$\begin{aligned} nlirf_n(\delta, v^t(\omega)) &= \delta \psi_n \\ &+ \delta^2 \psi_{nn} + \delta^3 \psi_{nnn} + 2\delta \sum_{j=n+1}^{\infty} \psi_{nj} v_{t-j+n} \\ &+ 3\delta^2 \sum_{j=n+1}^{\infty} \psi_{nnj} v_{t-j+n} \\ &+ 3\delta \sum_{j=n+1}^{\infty} \sum_{k=n+1}^{\infty} \psi_{njk} v_{t-j+n} v_{t-k+n} + 3\delta \sigma^2 \sum_{j=1}^{n-1} \psi_{njj}. \end{aligned}$$

Remember that if linear methods were used on this time series the linear impulse response function would just consist of a scaling of the $\{\psi_n\}$ sequence under the restrictions stated above on the cubic Volterra series expansion. The effect of the realized history of the process and asymmetries produced by varying the size and sign of the postulated shock δ would be hidden. Also there is no sense in which the coefficients in the Wold Representation determine the average behavior of the time series to the shock. The $nlirf_n$ illustrates three possible asymmetries:

- The magnitude of the shock for the same history will produce asymmetries because of the presence of δ^2, δ^3 .

⁶It should be emphasized that in an actual forecasting exercise this sequence is unlikely to be recoverable from the observed realizations of the process $\{Y_t\}$ (see Granger and Newbold (1977)).

- The response to δ will not be the same as $-\delta$ because of the δ^2 terms.
- The same shock will have differing effects depending on the sequence of v_t, v_{t-1}, \dots

2. The Derivative of the Nonlinear Impulse Response Function

$$\begin{aligned}
dnlirf_n(v^t(\omega)) &= \psi_n \\
&+ 2v_t\psi_{nn} + 3v_t^2\psi_{nnn} + 2 \sum_{j=n+1}^{\infty} \psi_{nj}v_{t-j+n} \\
&+ 6v_t \sum_{j=n+1}^{\infty} \psi_{nnj}v_{t-j+n} \\
&+ 3 \sum_{j=n+1}^{\infty} \sum_{k=n+1}^{\infty} \psi_{nj}v_{t-j+n}v_{t-k+n} + 3\sigma^2 \sum_{j=1}^{n-1} \psi_{njj}.
\end{aligned}$$

Beaudry and Koop (1993) and Potter (1995) find evidence of asymmetric response to large negative shocks in U.S. GNP in the post-war period. Some intuition for this type of behavior can be gained from considering the $dnlirf_n$. If $v_t \ll 0$ it is possible for certain histories that the direct negative effect is out-weighed by the squared term and the terms not involving v_t .

One feature of the $dnlirf_n$ that is not clear from the Volterra series expansion, is that ergodicity of the underlying stochastic process does not necessarily imply that the $dnlirf_n$ sequence converges. Intuitively, both $nlirf_n$ and $nludf_n$ should be equal to zero with probability 1 as $n \rightarrow \infty$ if the underlying process is ergodic. Since, the individual conditional expectations are converging to the same unconditional expectation. However, for the $dnlirf_n$ measure, an interchange of limits is required for the ergodicity of the underlying process to have an effect. Apart from the high level assumption of uniform convergence of $dnlirf_n$ there does not seem to be a direct method of guaranteeing that such an interchange of limits is valid and counter-examples to convergence can be constructed.

3. The Nonlinear Updating Function

$$\begin{aligned}
nludf_n(v^t(\omega)) &= \\
&v_t\psi_n + (v_t^2 - \sigma^2)\psi_{nn} + v_t^3\psi_{nnn} + 2v_t \sum_{j=n+1}^{\infty} \psi_{nj}v_{t-j+n}
\end{aligned}$$

$$3(v_t^2 - \sigma^2) \sum_{j=n+1}^{\infty} \psi_{nnj} v_{t-j+n} + 3v_t \sum_{j=n+1}^{\infty} \sum_{k=n+1}^{\infty} \psi_{nj} v_{t-j+n} v_{t-k+n} + 3v_t \sigma^2 \sum_{j=1}^{n-1} \psi_{njj}$$

The main difference from $nlirf_n$ is that the terms involving v_t^2 are now centered by the variance of the error term (i.e. the unconditional expectation of $nludf_n$ must equal zero). If δ is small then $nlirf_n$ will be close to zero but $nludf_n$ will not necessarily be close to zero if $|v_t|$ is small. In order to illustrate the possible economic importance of this observation consider the following example.

Assume that the ex-dividend price of a stock depends on the expectation of present discount value of future dividends and that the interest rate is constant. The unanticipated change in the value of a stock is given by:

$$E[Q_t|d^t(\omega)] - E[Q_t|d^{t-1}(\omega)] = \sum_{n=1}^{\infty} \beta^n \{E[D_{t+n}|d^t(\omega)] - E[D_{t+n}|d^{t-1}(\omega)]\},$$

where Q_t is the ex-dividend price of the stock, D_t is the dividend and $0 < \beta < 1$ is the discount factor.

Suppose that dividends follow an integrated process and ΔD_t is given by the Volterra model. Then, using the linear predictor:

$$L[Q_t|u^t(\omega)] - L[Q_t|u^{t-1}(\omega)] = U_t(\omega) \sum_{n=1}^{\infty} \beta^n \sum_{i=0}^n h_n.$$

Hence in the case that $U_t(\omega)$ is small the linear prediction of change in stock price is small and zero if $U_t(\omega)$ is zero. Alternatively, using the conditional expectation and restricting $v_t, v_{t-s}, s > 0$ to be zero, implying that $U_t(\omega)$ is zero, the unanticipated movement in the nonlinear prediction of the stock price is given by:

$$E[Q_t|v^t(\omega)] - E[Q_t|v^{t-1}(\omega)] = \sigma^2 \sum_{n=1}^{\infty} \beta^n \sum_{i=1}^n \psi_{ii}.$$

5 Generalized Impulse Response Functions

The expressions, $nlirf_n$, $dnlirf_n$, $nludf_n$ all represent individual realizations of sequences of random variables produced by various operations defined on conditional expectations of a time series. Realization differ because of initial conditions and in the case of the $nlirf_n$ choice of perturbation. Instead of analyzing individual realizations of the various impulse response functions, it is possible to treat them directly as random variables defined on the probability space of the time series itself.

$nlirf_n$ is the measure suggested by GRT but for our purposes it has the major problem of requiring a distribution for the perturbation δ .⁷ For example, suppose one chose the distribution of the innovation sequence for the time series, $E[Y_t|\mathcal{F}_t] - E[Y_t|\mathcal{F}_{t-1}]$. This has the advantage of being a random variable defined on the underlying probability space of the time series itself but the disadvantage that the average of the $nlirf_n, n > 0$ across the innovation sequence is not necessarily zero.⁸ Two advantages of the $nludf_n$ are that by construction the ‘perturbation’ is the innovation to the time series and its average value over the innovation is zero.

The $dnlirf_n$ has the advantage of not requiring the choice of a perturbation but it has two major disadvantages. As noted above it has limiting behavior that is hard to classify. In particular ergodicity of the underlying time series does not necessarily imply that the $dnlirf_n$ converges to zero as n increases. Also it measures the dynamic response to an infinitesimal perturbation. In discrete time such perturbations are atypical and we are often interested in the response to large shocks.

Priestley (1988) develops a concept of ‘generalized transfer function’ for nonlinear time series and we follow his lead by calling the impulse response function obtained from the updating operation a **Generalized Impulse Response Function** and denoting it by GI . We start by considering sequences of pairs of conditional distributions generated by the σ -fields of the time series $\{Y_t\}$.

Proposition 1

If Y_t is a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$ and \mathcal{F}_t is a sequence of σ -fields in \mathcal{F} . Then there exists a sequence of functions $\mu_n(t, H, \omega)$ defined for H in $\mathcal{B}(R)$, and ω in Ω with these two properties:

1. *For each ω in Ω $\mu_n(t, H, \omega)$ is, as a function of H , a probability measure on R .*
2. *For each H in $\mathcal{B}(R)$, $\mu_n(t, H, \omega)$ is, as a function of ω , a version of $P[Y_{t+n} \in H | \mathcal{F}_t]_\omega$.*

Using this result one can define the GI as a sequence of random variables on the probability space of the time series itself:

$$GI_n(t, \omega) = \int y \mu_n(t, dy, \omega) - \int y \mu_{n+1}(t-1, dy, \omega).$$

For example, in the case that $Y_t = \phi Y_{t-1} + V_t$ we would have:

$$GI_n(t, \omega) = E[Y_{t+n} | \sigma(V_t) \bigcup \sigma(Y_{t-1})]_\omega - E[Y_{t+n} | \sigma(Y_{t-1})]_\omega = \phi^n V_t(\omega).$$

⁷GRT mainly concentrate on the behavior of $nlirf_n$ for a fixed perturbation while varying the initial condition.

⁸Refer back to the distinctions between the $nlirf_n$ and $nludf_n$ for the cubic Volterra series expansion above.

A time series $\{Y_t\}$ will have a well-defined $\{GI_n\}$ if Z_t is in $L^1(\Omega, \mathcal{F}, \mathbf{P})$, where

$$Z_t = Y_t - E[Y_t | \mathcal{F}_{-\infty}].$$

That is, the purely deterministic (in the nonlinear sense) component of Y_t is removed.⁹ For example, consider the case where $Y_t = a + bt + V_t$, $E[|V_t|] < \infty$, then $E[Y_t | \mathcal{F}_{-\infty}] = a + bt$ and $Z_t = V_t$.

6 Persistence

How can one extract information on dynamics from the GI? As in the case of all dynamic systems, dynamics are extracted by evaluating the effect of different combinations of initial conditions. The difference here is that the effects will be random except in the deterministic case of perfectly forecastable dynamics, that is $GI_n = 0$, $n = 1, 2, \dots$ for all initial conditions. Consider the canonical examples of an IID sequence (no persistent dynamics) and partial sums of IID random variables (persistent dynamics). If $Y_t = V_t$ is an IID sequence of mean zero random variables, $GI_0 = V_t$ and the rest of the sequence would be zero. In this case there are clearly no interesting dynamics. Now suppose that Y_t was a time series of sequence of partial sums of V_t , $Y_t = Y_{t-1} + V_t$. GI_n would be equal to V_t for all n . However, V_t is centered at zero so the average (across initial conditions) dynamics from the random walk are not distinguishable from those of an IID sequence. In order to avoid this conclusion we need to measure the size of the GI_n more directly.

Recall a degenerate random variable at zero is second order stochastically dominated by all random variables with mean zero. For the IID case the response at $n = 0$ second order stochastically dominates the response at all other horizons and the reverse is not true. However, for the random walk the response at any horizon second order stochastically dominates the response at any other horizon and *vice versa*. That is, the shock at time t persists indefinitely. Finally, the response generated by the random walk second order stochastically dominates the response in the IID case. That is, the response to the shock in the random walk case is much larger.¹⁰

Define the random variable:

$$P_N = \sum_{n=0}^N GI_n.$$

⁹Rosenblatt (1971) discusses the notion of purely deterministic in the nonlinear sense. $\mathcal{F}_{-\infty}$ is the σ -field produced by the intersection of all the σ -fields from $t > 0$ backwards.

¹⁰An alternative approach to the one followed here is to calculate the largest Lyapunov exponent of the time series (see for example Nychka et al., 1992). This can be used to categorize the time series as stable or unstable and give a measure of the speed of convergence or divergence. However, in the case of linear time series the Lyapunov exponent does not produce the traditional measures of persistence found in the time series literature, whereas persistence measures based on the GI will. Further, since the Lyapunov exponent is based on an infinitesimal perturbation it does not contain information on the reaction to large versus small shocks.

and \mathcal{NL} to be the space of time series satisfying

$$\lim_{N \rightarrow \infty} E\left[\sum_{n=0}^N |GI_n|\right] < \infty.$$

Unlike the linear impulse response function case, where there is a one to one mapping between covariance stationarity and square summability of the impulse response function, there is no obvious properties, other than integrability and ergodicity, that the non-deterministic components of time series in \mathcal{NL} must share. One general class of time series can be shown to be in \mathcal{NL} .

Proposition 2 *If the non-deterministic component of $\{Y_t\}$ is a geometrically ergodic time series and in $L^1(\Omega, \mathcal{F}, \mathbf{P})$ then $\{Y_t\}$ is in \mathcal{NL} .*

We need a definition of an integrated time series applicable to time series in \mathcal{NL} . The standard approach for linear time series is to check the behavior of the spectral density of the level and first difference of time series at the zero frequency. For example, if the spectral density at zero of the level of the time series is infinite but it is bounded between zero and infinity for the first difference of the time series it is integrated of order 1 (I(1)). Here a similar approach is used by considering the variability in the sum of the GI for the level and first difference of the time series.

Under absolute summability of the GI define: $P_X = \lim_{N \rightarrow \infty} P_N$. for the time series $\{X_t\}$.

Definition: Integrated Time Series

Let $\{X_t\} \in \mathcal{NL}$. If

1. $P_X \succ_{SSD} 0$ and $0 \sim_{SSD} P_X$.
2. $0 \succ_{SSD} P_{\Delta X}$ where $\Delta X_t = X_t - X_{t-1}$.

Then X_t is integrated of order zero (I(0)).

Within the class of integrated time series it is interesting to measure the persistence of shocks as the horizon goes to infinity. This will done by using second order stochastic dominance as a measure of the size of the response at the infinite horizon. Hence, unlike the case of persistence measures based on the Wold Representation only a partial ordering of the class of integrated time series is possible using the GI.

Definition: Persistence

If $\{\Delta X_t\}$ and $\{\Delta Y_t\} \in \mathcal{NL}$ and $\{X_t, Y_t\}$ are I(1). X_t is more persistent than Y_t if $P_{\Delta X} \succ_{SSD} P_{\Delta Y}$ and $P_{\Delta Y} \sim_{SSD} P_{\Delta X}$.

It is possible to obtain complete orderings for certain linear time series.

Proposition 3

Let \mathcal{L} be the set of all (purely non-deterministic) linear time series with representation $\Delta X_t = \sum_{n=0}^{\infty} a_n W_{t-n}$ where $\sum_{n \geq 0} |a_n| < \infty, E[|W_t|] < \infty$.

There is a complete ordering of persistence by \succ_{SSD}

1. For all Gaussian time series in \mathcal{L} .
2. For all linear time series generated by the equivalence class of martingale difference sequence $\{W_t\}$ in \mathcal{L} with $\sum_{n \geq 0} a_n \geq 0$ or the equivalence class of a martingale difference sequence $\{W_t\}$ with symmetric distributions in \mathcal{L}

One can achieve some relative measurements of the dynamics in nonlinear time series to linear models by comparing the ‘size’ of the random variable P with the ‘size’ the random variable GI_0 . For example, one is often interested in comparing an integrated series with a martingale.

Definition: Persistence Relative to a Martingale

For $\{Y_t\} \in \mathcal{NL}$, $X_t = X_{t-1} + Y_t$ is at least as persistent as a random walk/martingale if $P_Y \succ_{SSD} GI_0$ and less or as persistent if $GI_0 \succ_{SSD} P_Y$.

It might not be possible to rank P and GI_0 by the second order stochastic dominance criterion. In such cases and also to acquire more precise information about the size of the ‘unit root’ the following scaling procedure can be used. Define:

$$\bar{\alpha} = \inf\{\alpha > 0 : \alpha GI_0 \succ_{SSD} P_Y\}$$

$$\underline{\alpha} = \sup\{\alpha > 0 : P_Y \succ_{SSD} \alpha GI_0\}.$$

Then one would say that $\{X_t\}$ is less persistent than a martingale with innovation $\bar{\alpha}GI_0$ and more persistent than a martingale with innovation $\underline{\alpha}GI_0$. Clearly, for linear models, $\bar{\alpha} = \underline{\alpha}$.

Alternatively one can compare $\{X_t\}$ to a linear integrated processes:

$$\Delta Z_t = \sum_{n=0}^{\infty} a_n W_{t-n},$$

with martingale difference innovation sequence $\{W_t\}$ distributed as GI_0 , $\sum_{n \geq 0} |a_n| < \infty$ and $\sum_{n \geq 0} a_n > 0$.

Definition: Persistence relative to linear models

X_t is less persistent than an integrated linear process, Z_t , with innovation W_t and $\sum_{n \geq 0} a_n = \bar{\alpha}_{\Delta X}$. X_t is more persistent than an integrated linear process, Z'_t , with innovation W_t if $\sum_{n \geq 0} a'_n = \underline{\alpha}_{\Delta X}$.

7 Conditional Versions of the Generalized Impulse Response Function

In this section a less abstract approach is taken in order to discern more information on conditional aspects of the behavior of the GI . We restrict attention to time series where $\mathcal{F}_t = \sigma(V_t) \cup \mathcal{F}_{t-1}$ and V_t is an IID sequence. Allowing for a slight abuse of notation, define:

$$E[Y_{t+n} | V_t, \mathcal{F}_{t-1}] = E[Y_{t+n} | \mathcal{F}_t],$$

and adopt the convention that V_t is the ‘shock’ and \mathcal{F}_{t-1} is the ‘history’. We will also use the same notation for the GI :

$$GI_n(V_t, \mathcal{F}_{t-1}) = E[Y_{t+n} | V_t, \mathcal{F}_{t-1}] - E[Y_{t+n} | \mathcal{F}_{t-1}].$$

We signify a realization of the GI by:

$$GI_n(v_t, y^{t-1}(\omega)) = E[Y_{t+n} | v_t, y^{t-1}(\omega)] - E[Y_{t+n} | y^{t-1}(\omega)],$$

and conditional random variables by:

$$\begin{aligned} &GI_n(v_t, \mathcal{F}_{t-1}), \text{ Conditional on the Shock} \\ &GI_n(V_t, y^{t-1}(\omega)), \text{ Conditional on the History} \\ &GI_n(A, B), \text{ Conditional on } v_t \in A, y^{t-1}(\omega) \in B \end{aligned}$$

In the recent nonlinear time series literature in economics there has been much attention placed on the ‘lack of persistence of negative shocks compared to positive shocks’ (for example, Beaudry and Koop, 1993). The previous work has tended to concentrate on particular realizations of the impulse response functions to show this asymmetry, presenting the problem of selective reporting of results. Here a general method of avoiding this ‘moral hazard’ issue is given.

Consider two conditioning events for the GI . Define:

$$\begin{aligned} GI_n^+ &= GI_n(V_t > 0, \mathcal{F}_{t-1}), \\ GI_n^- &= GI_n(V_t \leq 0, \mathcal{F}_{t-1}). \end{aligned}$$

Definition: Asymmetric Response to Shocks

A time series $\{Y_t\}$ in \mathcal{NL} has larger response to positive shocks relative to negative shocks at horizon n if:

1. $GI_n^+ \succ_{FSD} GI_0^+$, $GI_0^+ \sim_{FSD} GI_n^+$ and $GI_n^- \succ_{FSD} GI_0^-$.

Or a larger response to negative shocks relative to positive shocks if:

2. $GI_0^+ \succ_{FSD} GI_n^+$, $GI_0^- \succ_{FSD} GI_n^-$ and $GI_n^- \sim_{FSD} GI_0^-$.

It is obvious how to extend the spirit of this method to other conditional dynamic features of interest.

8 Empirical Examples of Nonlinear Measures of Persistence

Much of the recent emphasis on deriving measures of persistence from linear impulse response functions has been on providing estimates with little sampling variability. Given the computational requirements of producing point estimates in the nonlinear case it is not currently feasible to assess the sampling variability. Thus, I proceed as if the distribution function of the GI generated by the estimated model's is the distribution function of the GI generated by the true model. Furthermore, I also ignore the simulation error and truncation error in constructing realizations of the persistence random variable.¹¹ The construction of the persistence random variable is truncated at 8 quarters since the response of GNP growth is approximately zero at this horizon. Despite these caveats the nonlinear measures are of independent interest as they lead one to be very wary of conclusions drawn from linear models.

The infimum and supremum above are not simple to calculate for second order stochastic dominance but they are relatively easy to calculate for the cut criterion of Karlin (see Stoyan (1983)): define $sign(f)$ to be the function returning the sign changes in the function f : if $sign(F-G) = (+, -)$ then $F \succ_{SSD} G$.¹²

The two examples I consider are Hamilton's (1989) Markov Trend model of U.S. GNP and the SETAR model for U.S. GNP of Potter (1995). Previous linear estimates of persistence have ranged from a low of zero when a unit root is rejected to the high estimates of 1.6 (see Hamilton (1989) for a representative listing). Hamilton calculated the standard deviation of the updating function for his model using the Wold Representation and obtained an estimate of 1.62. Figures

¹¹Koop, Pesaran and Potter (1996) give a detailed description of simulation methods to construct the GI.

¹²Unfortunately this is only a sufficient condition and its failure does not imply that F, G cannot be ranked by the second order stochastic dominance criterion. At the failure of the cut criterion it is possible to integrate the empirical distribution functions to check for second order stochastic dominance. Experiments indicated that this rarely affected the first decimal place. Hence given the degree of approximation involved in the exercise I suggest using the cut criterion alone.

1 and 2 contain the densities of the persistence measure for GNP derived from the SETAR model and from the Markov Trend model respectively. Each figure also contains a Gaussian density with the same variance as the persistence random variable. The estimate of $P_{\Delta GNP}$ from the SETAR model has a density that appears to be almost Gaussian except for the hump around the value of -2. The estimate of $P_{\Delta GNP}$ from the Markov Trend model produces a highly non-Gaussian distribution with the appearance of tri-modality in the density.

One interesting issue is whether the nonlinear models are more or less persistent than a martingale. This issue is examined in Figures 3 and 4 which show the distribution function of the persistence random variable. Clearly the SETAR model is more persistent than a martingale but the distinction is less clear in the case of the Markov trend model.

More precise information on persistence can be found by using the scaling procedure. For the SETAR model, this procedure produced an upper bound of $\bar{\alpha} = 2.0$ and a lower bound of $\underline{\alpha} = 1.6$ for persistence compared to a martingale with innovation GI_0 from the SETAR. For the Markov Trend model, the bands are much wider with upper bound of $\bar{\alpha} = 2.8$ and lower bounds of $\underline{\alpha} = 1.1$ compared to a martingale with innovation GI_0 from the Markov Trend. Because of the tri-modality produced by the Markov Trend model, it is not possible to order the two values of $P_{\Delta GNP}$ by second order stochastic dominance. That is, one cannot tell if the responses are more persistent in one model rather than the other.

9 Conclusion

This paper has developed a number of tools to report the dynamics of nonlinear time series models and compare these dynamics to linear time series models. An interesting issue for future research is how to provide measures of sampling uncertainty for the GI as a random variable rather than for particular realizations. In particular it would be useful to have measures of the size of the nonlinearity that take into account parameter and modeling uncertainty.

Appendix

Restrictions on Cubic Volterra Series Model

The following conditions imply that the stochastic process U_t constructed from the Cubic Volterra Series model is white noise.

$$\psi_{0i} = \psi_{0ij} = 0 \quad \forall i, j, \quad \sum_{i=1}^{\infty} \psi_{ii} = 0,$$

$$\begin{aligned}
& \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{ij}^2 E[V_{t-i}^2 V_{t-j}^2] + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_{ijk}^2 E[V_{t-i}^2 V_{t-j}^2 V_{t-k}^2] = \\
& \quad -2 \sum_{i=1}^{\infty} \psi_{iii} \psi_i \sigma_v^4 - 6 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{ijj} \psi_i \sigma_v^4, \\
& \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{ij} \psi_{i+s, j+s} E[V_{t-i}^2 V_{t-j}^2] + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_{ijk} \psi_{i+s, j+s, k+s} E[V_{t-i}^2 V_{t-j}^2 V_{t-k}^2] + \\
& \quad 3 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{i+s-1} \psi_{ijj} \sigma_v^4 + 3 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{i-1} \psi_{i+s, jj} \sigma_v^4 = 0, \quad \forall s \neq 0,
\end{aligned}$$

Proof of Proposition 1. Billingsley (1986) Theorem 33.3 applies directly for each n .

Proof of Proposition 2

I adapt a definition of geometric ergodicity from Tong (1990 Appendix 1). Without loss of generality I examine the case of a scalar first order Markov process.

Definition

$\{Y_t\}$ is geometrically ergodic if there exists a probability measure μ , a positive constant $r < 1$ and a μ -integrable non-negative measurable function h such that

$$\|P(Y_{t+n}|x) - \mu(Y_{t+n})\|_{\nu} < r^n h(x), \text{ for an arbitrary } x,$$

where $\|\cdot\|_{\nu}$ is the total variation norm. By Tweedie (1983) this exponential convergence relative to a μ -integrable non-negative measurable function h also applies to conditional moments when the unconditional moments exist.

$$\begin{aligned}
& E\left[\sum_{n=0}^N |GI_n|\right] \leq E\left[\sum_{n=0}^N |E[Y_{t+n}|Y_t] - E[Y_{t+n}|Y_t]|\right] \\
& \leq E\left[\sum_{n=0}^N |E[Y_{t+n}|Y_t] - E[Y_{t+n}]| + |E[Y_{t+n}] - E[Y_{t+n}|Y_{t-1}]|\right] \\
& \leq E\left[\sum_{n=0}^N r^n 2 \max\{h(Y_t), h(Y_{t-1})\}\right] \leq \frac{2}{1-r} E[\max\{h(Y_t), h(Y_{t-1})\}]
\end{aligned}$$

Thus, by the μ integrability of h and the dominated convergence theorem the generalized impulse response function of a geometrically ergodic Markov process is absolutely summable.

Proof of Proposition 3. For the Gaussian case $P = W_t \sum_{n \geq 0} a_n$ which is a Gaussian random variable with mean zero and variance $\sigma_w^2 (\sum_{n \geq 0} a_n)^2$. For Gaussian random variables the ranking by second order stochastic dominance is the same as by variance. Therefore, there is a complete ranking.

For the general linear model case $P = W_t \sum_{n \geq 0} a_n$, which is a scaling of the random variable W_t by $\sum_{n \geq 0} a_n$. Consider the case where $0 < a = \sum_{n \geq 0} a_n < \sum_{n \geq 0} b_n = b$. Then, we have the desired result by the cut criterion

since

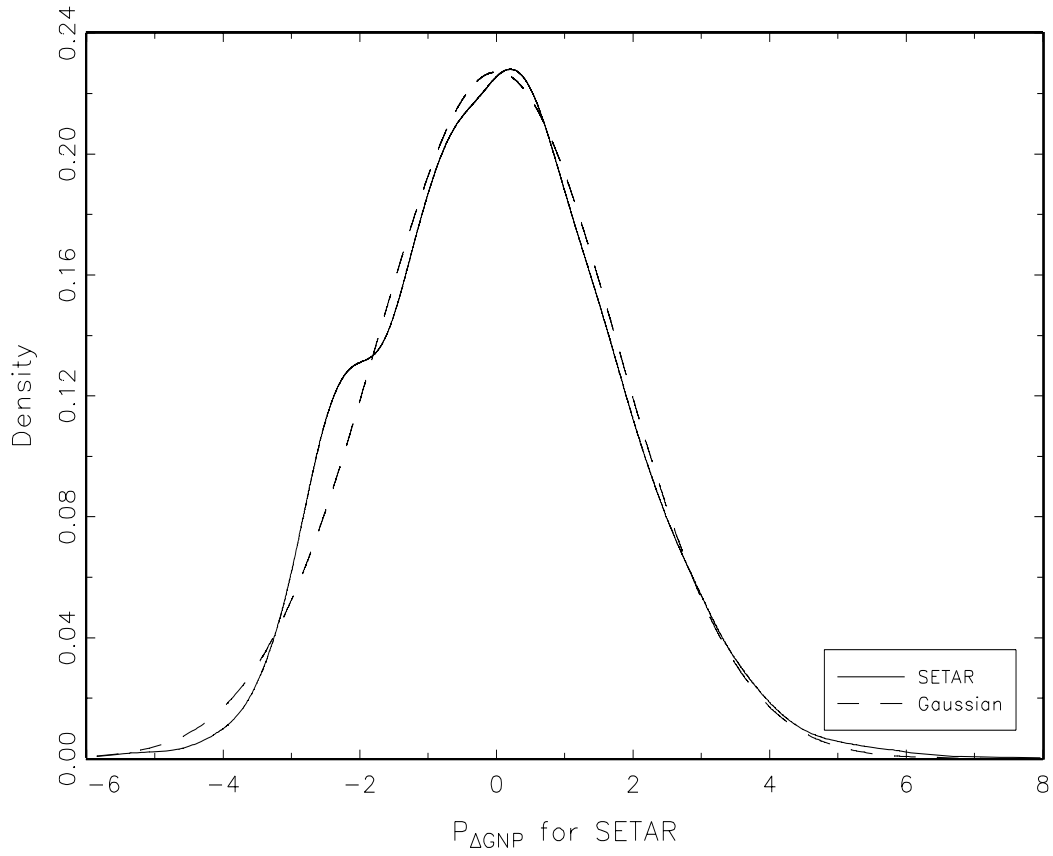
$$\text{sign}[F_W(w \leq \frac{\lambda}{b}) - F_W(w \leq \frac{\lambda}{a})] = (+, -).$$

Under the symmetry of W_t , the same argument applies to the absolute value of $\sum_{n \geq 0} a_n$ and $-W_t$.

References

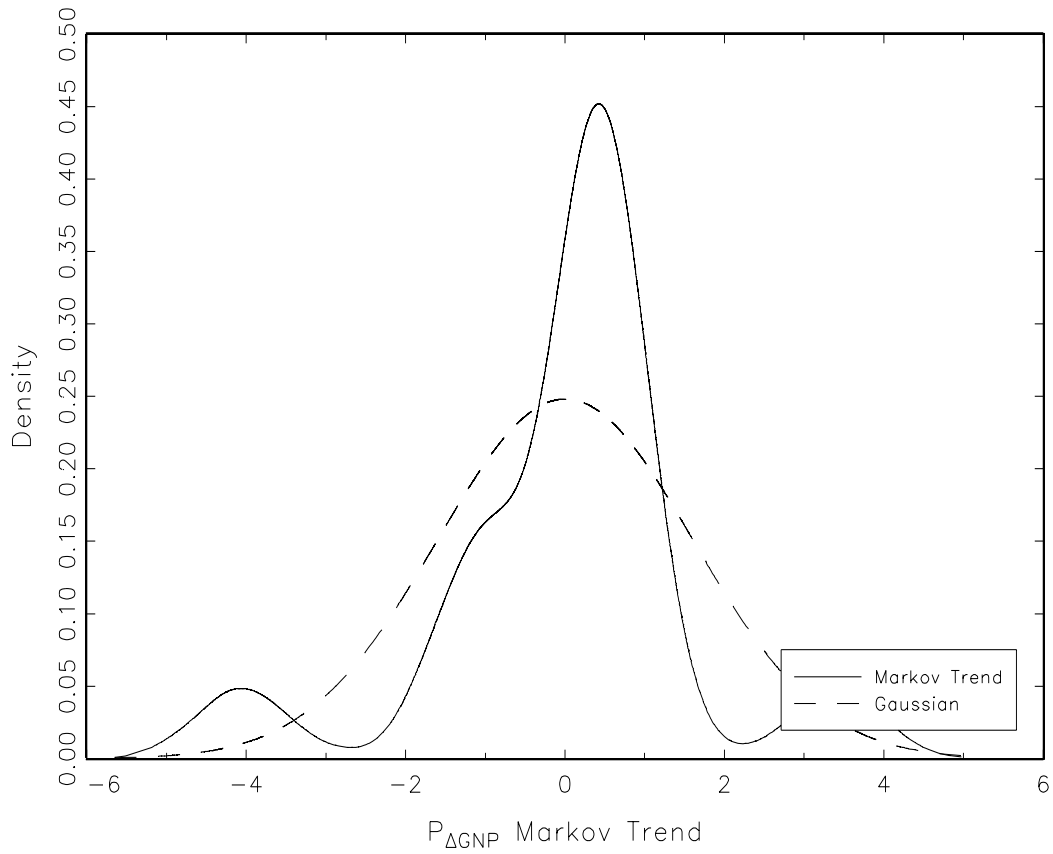
- Beaudry, P. and G. Koop (1993). "Do Recessions Permanently Affect Output?" *Journal of Monetary Economics*, 31, 149-163.
- Billingsley, P., (1986), *Probability and Measure*, (John Wiley, New York).
- Brock, W., and Potter, S., "Nonlinear Time Series and Macroeconometrics," in *Handbook of Statistics XI*, edited by Maddala, G, Rao and Vinod, (North Holland, Amsterdam)
- Granger, C.W.J., and Newbold, P., (1977), *Forecasting Economic Time Series*, (Academic Press, New York).
- Gallant, A.R., P.E. Rossi and G Tauchen (1993). "Nonlinear Dynamic Structures," *Econometrica*, 61, 871-908.
- Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary time series and the Business Cycle," *Econometrica* , 57, 357-84.
- Koop, G. (1995). "Parameter uncertainty and impulse response analysis," *Journal of Econometrics*, 72, 135-149
- Koop, G., Pesaran, H., and Potter, S., (1996): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics* 74,119-148.
- Nychka, D., Ellner, S., Gallant, R., and McCaffrey, D., (1992), "Finding Chaos in Noisy systems," *Journal of the Royal Statistical Society, Series B*, 54, 399-426.
- Potter, S., (1995), "A Nonlinear Approach to U.S. GNP," *Journal of Applied Econometrics*, 10, 109-125.
- Priestley, M.B., (1988), *Non-Linear and Non-Stationary Time Series*, (Academic Press, New York).
- Rosenblatt, M., (1971), *Markov Processes. Structure and Asymptotic Behavior*, (Springer, New York).
- Stoyan, D. (1983), *Comparison Methods for Queues and Other Stochastic Processes*, (Wiley, New York).
- Tong, H., (1990), *Non-linear Time Series: A Dynamical System Approach*, (Clarendon Press, Oxford).
- Tong, H., (1995), "A personal overview of nonlinear time series analysis from a chaos perspective," *Scandinavian Journal of Statistics* forthcoming.

Figure 1: Probability Density Function of $P_{\Delta GNP}$ For SETAR Model*



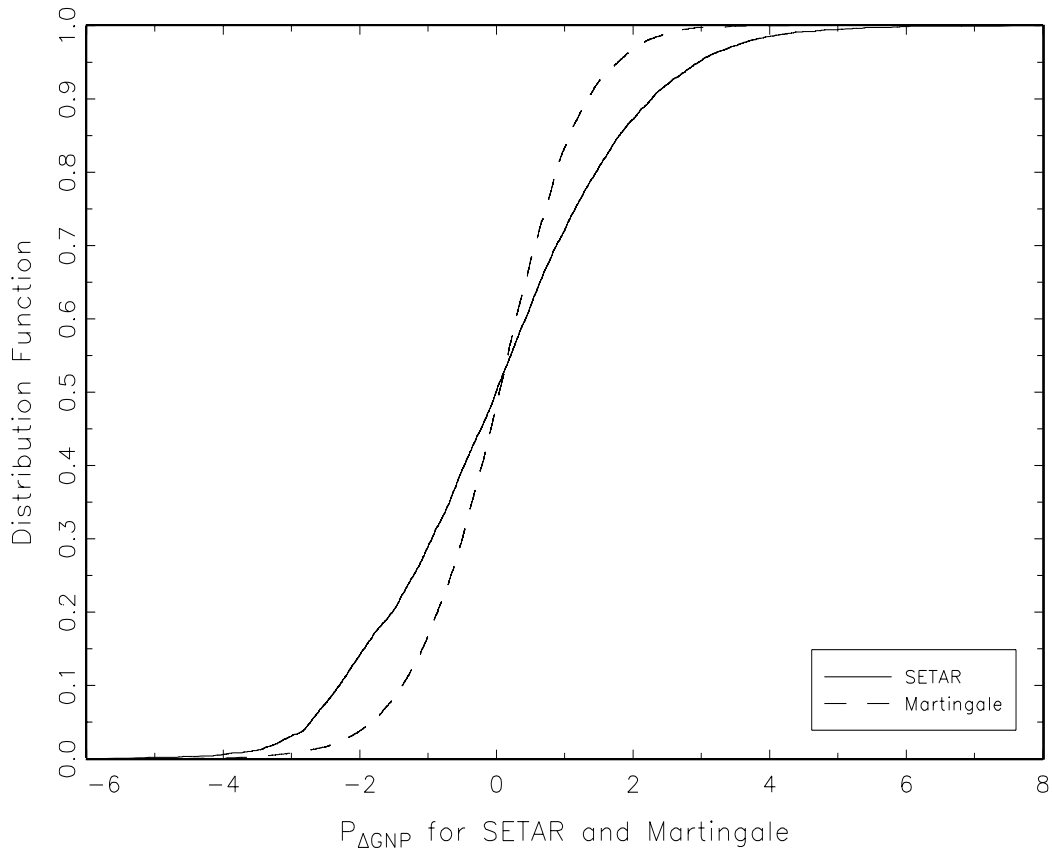
* The Persistence Random Variable ($P_{\Delta GNP}$) for the SETAR is generated from 10,000 realizations of $\sum_{n=0}^8 GI_n$ using the SETAR model of Potter (1995). The density was estimated by a normal kernel. The Gaussian density shown is for a normal random variable with the same standard deviation as the persistence random variable.

Figure 2: Probability Density Function of $P_{\Delta GNP}$ For Markov Trend Model*



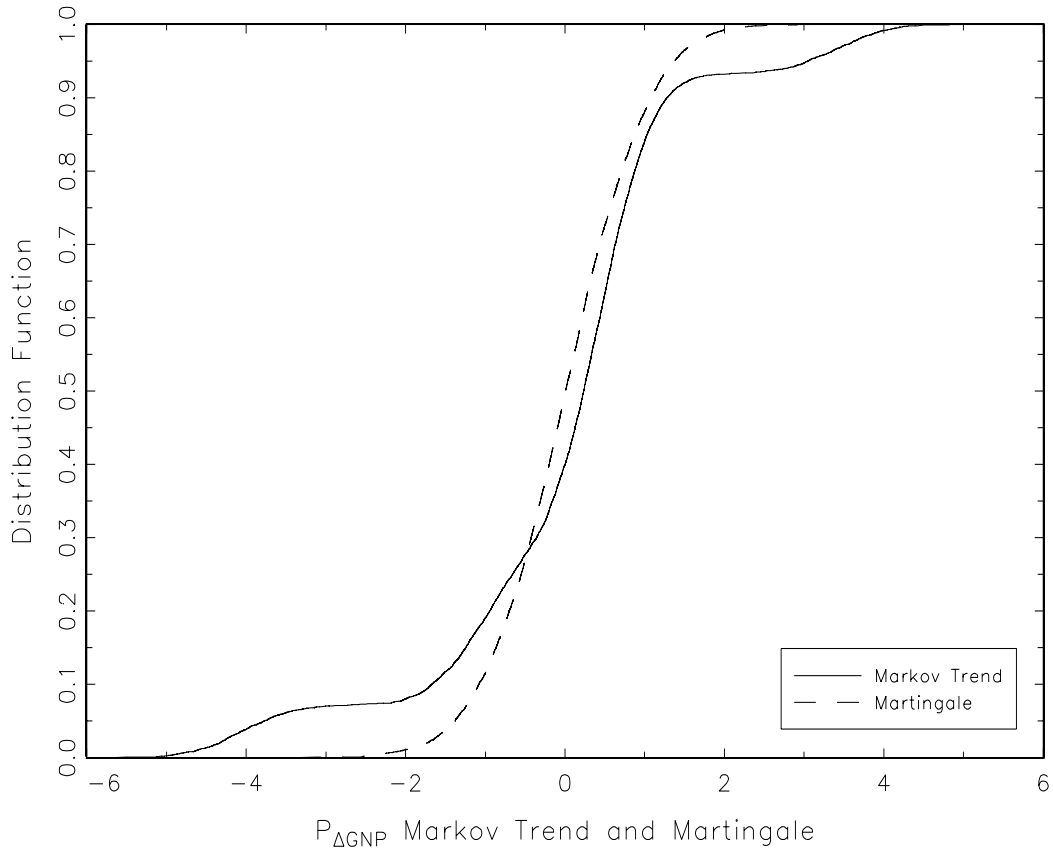
* The Persistence Random Variable ($P_{\Delta GNP}$) for the Markov Trend is generated from 10,000 realizations of $\sum_{n=0}^8 GI_n$ using the SETAR model of Hamilton (1989). The density was estimated by a normal kernel. The Gaussian density shown is for a normal random variable with the same standard deviation as the persistence random variable.

Figure 3: Distribution Function of $P_{\Delta GNP}$ for SETAR model and Martingale*



* The Persistence Random Variable ($P_{\Delta GNP}$) for the SETAR is generated from 10,000 $\sum_{n=0}^8 GI_n$ from the SETAR model of Potter (1995). The martingale it is generated from 10,000 realizations of GI_0 using the SETAR model. The SETAR \succ_{SSD} the martingale since its distribution function cuts from above.

Figure 4: Distribution Function of $P_{\Delta GNP}$ for Markov Trend model and Martingale*



* The Persistence Random Variable ($P_{\Delta GNP}$) for the Markov Trend model is generated from 10,000 realizations of $\sum_{n=0}^8 GI_n$ using the Markov Trend model of Hamilton (1989). The martingale is generated from 10,000 realizations of GI_0 for the Markov Trend model. The Markov Trend \succ_{SSD} the martingale since its distribution function cuts from above.