Measuring Nonlinear Granger Causality in Mean

Abderrahim Taamouti

Universidad Carlos III de Madrid

Joint with Xiaojun Song (Carlos III de Madrid)

ESSEC Business School

March 19, 2014

Outline

1. **Motivations**

2. **Contributions**

3. **Causality Measures**

4. **Estimation**

5. **Inference**

6. **Monte Carlo Simulations: Asymptotic and Bootstrap-based Tests**

7. **Application: Nonlinear Predictability of Risk Premium**

1. Motivations

Example 1: Comparing the effectiveness of fiscal and monetary policies for economic aggregates

Policy-makers have two tools to influence the real economy:

Fiscal policy: Manipulating the level of aggregate demand in the economy through spending and revenue collection to achieve price stability, full employment, and economic growth.

Monetary policy: Manipulating the supply of money through interest rates and other tools to influence economic growth, inflation, unemployment.

How to measure and compare the effectiveness of these policies in a **model-free framework***?*

This is important since some policies might have undesirable effect on the economy:

1. Big debate on the consequences of government stimulus policies and Fed programs (in putting the money into the banking system) in terms of inducing high inflation.

2. Libertarian economists believe that the housing bubble and the subsequent financial crisis in 2008 is the consequence of cutting interest rates by the Fed and keeping them low for too long.

Thus, we need to use a policy that is more effective and has less undesirable effect (e.g. high inflation) on the economy.

The proposed measures solve the issue of comparing the effectiveness of these policies in a model-free framework.

Example 2: Comparing the predictive power of predictors of risk premium

Financial economists have used different predictors to predict risk premium: dividend-price ratio, liquidity factors, book-to-market ratio...

The econometric method used in this context is an ordinary least squares regression of stock returns onto the past of predictors, limiting the analysis to linear predictability.

How to compare the predictive power of these predictors in a **model-free framework***?*

How to quantify the **degree of nonlinear predictability** *of risk premium?*

Example 3: Measuring nonlinear causality

Consider a simple bivariate nonlinear regression model:

$$X_{t+1} = \beta \, Y_t \cdot X_t + \varepsilon_{t+1}$$

$\{Y_t\}$ and $\{X_t\}$ are mutually independent and individually i.i.d $N(0, 1)$.

Y **nonlinearly causes** the conditional mean of X. However, Y **does not linearly** cause the conditional mean of X.

How can we measure the degree of nonlinear causality-in-mean?
Existing measures do not answer this question.

Related literature

Parametric Approach:

Geweke (1982, 1984): Unconditional and Conditional measures of **linear** Granger causality in **mean** (one horizon) using bivariate and trivariate autoregressive models, respectively.

Dufour and Taamouti (2010): Measures of **linear** Granger causality in **mean** (multi-horizons) using vector autoregressive models (indirect causality).

Dufour and Zhang (2013): Measures of Granger causality in **variance** (multi-horizons) using multivariate GARCH models.

Gouriéroux, Monfort, and Renault (1987): Measures of Granger causality in **distribution** under normal assumption.

However, misspecification of parametric model may affect the structure of the causality [see Bouezmarni, Rombouts, and Taamouti (2012) and Bouezmarni and Taamouti (2011)].

Furthermore, several “causal relations” are nonlinear; see for example Gabaix, Gopikrishnan, Plerou, and Stanley (2003).

Non-Parametric Approach:

Taamouti, Bouezmarni and El Ghouch (2012): Nonparametric estimation and inference for measures of Granger causality in **distribution**.

Problem: their measures are not informative about levels of distribution where the causality exist.

2. Contributions

1. We propose model-free measures to quantify and compare nonlinear Granger causalities in mean.

2. We establish the consistency and asymptotic normality of a nonparametric estimator of causality measures.

3. The test which is based on the asymptotic normality is consistent and has nontrivial power against $\sqrt{T}$-local alternatives.

4. Monte Carlo simulations reveal that asymptotic and bootstrap-based tests behave quite well and have good finite sample size and power properties for a variety of typical data generating processes and different sample sizes.

5. We quantify and compare the degree of nonlinear predictability of equity risk premium across horizons using variance risk premium.

3. Causality measures

For X and Y two variables of interest, Wiener (1956) and Granger (1969) analysis distinguishes between three basic types of causality:

1. From Y to X

2. From X to Y, and

3. Instantaneous causality

In practice, it is possible that all three causality relations coexist, hence the importance of finding means to quantify their degree.

Remark: In the paper X and Y can be **multivariate Markov processes** of any order p, for $p \geq 1$.

Framework

Consider two processes X and Y

$$\{X_s, \quad s \quad \leq \quad t\}$$

$$\{Y_s, \quad s \quad \leq \quad t\}$$

Let $I_X(t)$ be the information contained in the history of X and $I_{XY}(t)$ the information contained in the history of X and Y.

Variance of $X_{t+1}{}'s$ forecast error, conditional on an information set B_t, is:

$$\sigma^2(X_{t+1} \mid B_t)$$

$B_t \equiv I_X(t), \, I_{XY}(t)$.

Variance characterization of non-causality

Definition: Y does not cause X iff

$$\sigma^2(X_{t+1} \mid I_X(t)) = \sigma^2(X_{t+1} \mid I_{XY}(t)) \,, \; \forall t > \omega$$

$\sigma^2(X_{t+1} \,|\, \cdot\,) = \mathsf{E}\Big\{u^2[X_{t+1} \,|\, \cdot]\Big\}$ with:

$u[X_{t+1} \,|\, \cdot] = X_{t+1} - P[X_{t+1} \,|\, \cdot]$ is the prediction error

$P[X_{t+1} \,|\, \cdot]$ is the best (nonlinear) forecast of X_{t+1} based on the information set $B_t = \cdot$

The idea

By analogy to Geweke (1982,84), we propose causality measures based on the following idea

Principle: "Y causes X" if

$$\sigma^2(X_{t+1} \mid I_X(t)) > \sigma^2(X_{t+1} \mid I_{XY}(t))$$

Measuring the difference between the two variances is the same as measuring the strength of causality from Y to X.

Measure of a feedback causality

Definition: The function

$$C(Y \to X) = \ln\left[\frac{\sigma^2(X_{t+1} \mid I_X(t))}{\sigma^2(X_{t+1} \mid I_{XY}(t))}\right]$$

defines the *mean-square causality measure* from Y to X. Similarly,

$$C(X \to Y) = \ln\left[\frac{\sigma^2(Y_{t+1} \mid I_Y(t))}{\sigma^2(Y_{t+1} \mid I_{XY}(t))}\right]$$

defines the *mean-square causality measure* from X to Y.

Important properties: **(i)** they are non-negative, and **(ii)** they cancel only when there is no causality.

Measure of an instantaneous causality

Definition: The function

$$C(X - Y) = \ln\left[\frac{\sigma^2(X_{t+1} \mid I_{XY}(t))\, \sigma^2(Y_{t+1} \mid I_{XY}(t))}{\det \Sigma[(X_{t+1},\, Y_{t+1}) \mid I_{XY}(t)]}\right]$$

defines the *mean-square instantaneous causality measure* between Y and X.

$\Sigma[(X_{t+1},\ Y_{t+1}) \mid I_{XY}(t)] = \mathsf{E}\Big\{U[Z_{t+1} \mid I_{XY}(t)]U[Z_{t+1} \mid I_{XY}(t)]'\Big\}$ for $Z_t = \Big(X_t',\, Y_t'\Big)'$.

Measure of dependence

Definition: The function

$$C(X,\,Y) = C(X \to Y) + C(Y \to X) + C(X{-}Y)$$

defines the *intensity* of the dependence between X and Y. We have:

$$C(X,Y) = \ln\left[\frac{\sigma^2(X_{t+1} \mid I_X(t))\;\sigma^2(Y_{t+1} \mid I_Y(t))}{\det \Sigma[X_{t+1},\,Y_{t+1} \mid I_{XY}(t)]}\right]$$

Causality measures for nonparametric regression models

Consider a multivariate nonparametric regression

$$Z_{t+1} = \Phi(Z_t) + u_{t+1}$$

$Z_{t+1} = \left(X_{t+1},\, Y_{t+1}\right)'$ and $u_{t+1} = \left(u^X_{t+1},\, u^Y_{t+1}\right)'$ is an error term with $E\left[u_{t+1} \mid Z_t\right] = 0$.

$\Phi(Z_t)$ is an unknown function of Z_t such that $\Phi(Z_t) = E\left[Z_{t+1} \mid Z_t\right]$.

Unrestricted and restricted nonparametric regressions

Unrestricted nonparametric regression of X_{t+1}:

$$X_{t+1} = \Phi_1 \left(Z_t \right) + u_{t+1}^X$$

Restricted nonparametric regression of X_{t+1}:

$$X_{t+1} = \bar{\Phi}_1 \left(X_t \right) + \bar{u}_{t+1}^X$$

Measure of nonlinear Granger causality-in-mean

The measure of *nonlinear* Granger causality-in-mean from Y to X is given by:

$$C(Y \to X) = \ln \left[\frac{Var\left[\left(X_{t+1} - \bar{\Phi}_1\left(X_t\right)\right)\right]}{Var\left[\left(X_{t+1} - \Phi_1\left(Z_t\right)\right)\right]} \right]$$

4. Estimation

An estimator of measure of nonlinear Granger causality-in-mean $C(Y \to X)$ is given by:

$$\hat{C}(Y \to X) := \ln\left(\frac{\hat{\hat{\sigma}}^2(X_{t+1} \mid X_t)}{\hat{\sigma}^2(X_{t+1} \mid Z_t)}\right) = \ln\left(\frac{\sum_{t=0}^{T-1}(X_{t+1} - \hat{\hat{\Phi}}_1(X_t))^2}{\sum_{t=0}^{T-1}(X_{t+1} - \hat{\Phi}_1(Z_t))^2}\right)$$

$\hat{\Phi}_1(Z_t)$ and $\hat{\hat{\Phi}}_1(X_t)$ are the Nadaraya-Watson kernel estimators of the unrestricted and restricted regression functions, respectively.

Nadaraya-Watson estimator of $\bar{\Phi}_1(\cdot)$:

$$\hat{\bar{\Phi}}_1(x) = \frac{\sum_{t=0}^{T-1} \overline{K}\left(\frac{x-X_t}{\bar{h}}\right) X_{t+1}}{\sum_{s=0}^{T-1} \overline{K}\left(\frac{x-X_s}{\bar{h}}\right)}$$

Nadaraya-Watson estimator of $\Phi_1(\cdot)$:

$$\hat{\Phi}_1(z) = \frac{\sum_{t=0}^{T-1} K\left(h^{-1}(z - Z_t)\right) X_{t+1}}{\sum_{s=0}^{T-1} K\left(h^{-1}(z - Z_s)\right)}$$

$\overline{K}\left(\frac{x-X_t}{\bar{h}}\right)$ and $K\left(h^{-1}(z - Z_t)\right)$ are univariare and multivariate kernel functions, and $\bar{h}$ and h are univariare and multivariate smoothing parameters (i.e. bandwidths).

Assumptions

(A.1.1) $\{(X_t, Y_t)' \in \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t \geq 0\}$ is a strictly stationary and ergodic process. Furthermore, $E|X_t| < \infty$ and $E|Y_t| < \infty$.

(A.1.2) The marginal density $f_X(\cdot)$ of X_t and the joint density $f_Z(\cdot)$ of $Z_t = (X_t, Y_t)'$ are bounded away from zero and bounded above.

(A.2.1) The kernel functions $K(\cdot)$ and $\overline{K}(\cdot)$ are two-product and univariate kernel functions, respectively, and they are symmetric and bounded. That is, $K(u_1, u_2) = k(u_1)k(u_2)$ and $\overline{K}(u) = k(u)$, where $k(\cdot)$ satisfies $\int k(u)\,du = 1$, $\int uk(u)\,du = 0$ and $\int u^2 k(u)\,du < \infty$.

(A.2.2) The univariate (restricted) bandwidth parameter $\bar{h}$ and bivariate (unrestricted) bandwidth parameter h satisfy $\bar{h} \longrightarrow 0$, $h \longrightarrow 0$, as $T \longrightarrow \infty$. Further, $T\bar{h} \rightarrow \infty$ and $Th^2 \rightarrow \infty$, as $T \rightarrow \infty$.

Consistency

Proposition: Under Assumptions (A.1.1)-(A.2.2)

$$\hat{C}(Y \to X) \underset{p}{\to} C(Y \to X)$$

5. Inference

We wish to test the null hypothesis

$$H_0 : C(Y \to X) = 0$$

This is equivalent to testing non causality-in-mean from Y to X.

Causality measure can be used to test for *nonlinear* Granger non causality-in-mean.

Asymptotic Normality

Theorem: Under Assumptions (A.1.1)-(A.2.2) and H_0,

$$\hat{\Gamma}_T = \sqrt{T}\hat{C}(Y \to X)/\sqrt{\Omega} \xrightarrow{d} \mathcal{N}(0,1)$$

where

$$\begin{aligned}\Omega &= (\sigma^4)^{-1}E((\bar{u}_t^X)^2 - (u_t^X)^2)^2 \\ &\quad +2(\sigma^4)^{-1}\lim_{T\to\infty}\left(\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)E[((\bar{u}_t^X)^2-(u_t^X)^2)((\bar{u}_{t-\tau}^X)^2-(u_{t-\tau}^X)^2)\right)\end{aligned}$$

For a given significance level α, we reject the null hypothesis H_0, when $\hat{\Gamma}_T > z_\alpha$, where z_α is the critical value from the standard normal distribution.

Remark: If we further impose $\bar{u}_{t+1}^X \perp u_{t+1}^X$, the asymptotic variance Ω is simplified to

$$\Omega = \bar{\kappa} + \kappa - 2$$

with $\bar{\kappa} = E(\bar{u}_{t+1}^X)^4/\sigma^4\,[X_{t+1} \mid X_t]$ and $\kappa = E(u_{t+1}^X)^4/\sigma^4[X_{t+1} \mid X_t, Y_t]$ are the kurtosis coefficients of $\bar{u}_{t+1}^X$ and u_{t+1}^X, respectively.

Consistency of the test

Proposition: Under Assumptions (A.1.1)-(A.2.2), the test is consistent for any loss functions $\sigma^2\,[X_{t+1}\,|\,X_t]$ and $\sigma^2[X_{t+1}\,|\,X_t, Y_t]$ such that:

$$\sigma^2\,[X_{t+1}\,|\,X_t] - \sigma^2[X_{t+1}\,|\,X_t, Y_t] > 0$$

Local power

We consider the following $\sqrt{T}$-local alternatives:

$$H_{1T} : C(Y \to X) = \frac{1}{\sqrt{T}}\mu$$

μ is a finite positive constant.

Proposition: Under Assumptions (A.1.1)-(A.2.2) and H_{1T}, we have

$$\sqrt{T}\hat{C}(Y \to X) \xrightarrow{d} \mathcal{N}(\mu, \Omega)$$

Ω is defined before.

Smoothed local bootstrap

(1) We draw a bootstrap sample $\{(X_t^*, Y_t^*)\}_{t=1}^T$ using the procedure developed in Paparoditis and Politis (2000);

(2) Based on the bootstrap sample $\{(X_t^*, Y_t^*)\}_{t=1}^T$, we compute the bootstrapped version of the test statistic: $\Gamma_T^* = \sqrt{T}\hat{C}^*(Y \to X)/\sqrt{\hat{\Omega}^*}$;

(3) Repeat steps (1)-(2) B times so that we get $\Gamma_{j,T}^*$, for $j = 1, \ldots, B$;

(4) We compute the bootstrapped p-value using $p^* = B^{-1}\sum_{j=1}^{B} 1(\Gamma_{j,T}^* > \hat{\Gamma}_T)$, where $\hat{\Gamma}_T$ is the test statistic based on the original sample, and for a given significance level α, we reject the null hypothesis if $p^* < \alpha$.

Validity of bootstrap

(A.3) The bootstrap bandwidth parameter h^* satisfies $h^* \to 0$ and

$Th^{*5}/(\ln T)^{\gamma} \to C$, for some $\gamma > 0$ and $0 < C < \infty$, as $T \to \infty$.

Theorem: Under Assumptions (A.1.1)-(A.2.2) and (A.3), and under the null H_0, we have

$$\Gamma_T^* = \sqrt{T}\hat{C}^*(Y \to X)/\sqrt{\Omega} \xrightarrow{d} \mathcal{N}(0,1)$$

where Ω is defined before.

6. Monte-Carlo simulations

Asymptotic-based test:

We examine the size and power properties of $\hat{\Gamma}_T$ using the DGPs in Table 1.

Univariate bandwidth $\bar{h} = T^{-1/(2+\delta)}$, for $\delta = 0.6$, 0.8.

Bivariate bandwidth $h = (h_1, h_2)'$, with $h_1 = h_2 = T^{-1/5}$.

Kernel function $k(\cdot)$ is given by the standard normal density.

Three sample sizes $T = 100$, 200, 300.

1000 simulations.

Table 1: Data-generating processes

DGPs	Variables of Interest		Direction of Causality in the DGP
	Y_t	X_t	
DGP S1	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP S2	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = \|X_{t-1}\|^{0.8} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP S3	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} \exp\{-0.5X_{t-1}^2\} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP S4	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = \sin(X_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP P1	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + 0.5Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P2	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + 0.5Y_{t-1} + 0.5\sin(-2Y_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P3	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + 0.5Y_{t-1}^2 + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P4	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1}Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P5	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = \sin(2(X_{t-1} + Y_{t-1})) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$

Table 2: Empirical size of the asymptotic-based test

Bandwidth	DGPs			
	DGP S1	DGP S2	DGP S3	DGP S4
		$T = 100$		
$\delta = 0.6$	0.050	0.055	0.026	0.054
$\delta = 0.8$	0.077	0.065	0.035	0.078
		$T = 200$		
$\delta = 0.6$	0.045	0.056	0.017	0.045
$\delta = 0.8$	0.058	0.059	0.040	0.070
		$T = 300$		
$\delta = 0.6$	0.030	0.035	0.013	0.035
$\delta = 0.8$	0.053	0.072	0.045	0.060

Table 3: Empirical power of the asymptotic-based test

Bandwidth	DGPs				
	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5
			$T = 100$		
$\delta = 0.6$	0.847	0.890	0.974	0.806	0.854
$\delta = 0.8$	0.893	0.887	0.989	0.842	0.842
			$T = 200$		
$\delta = 0.6$	0.994	0.993	1.000	0.987	1.000
$\delta = 0.8$	0.994	0.991	1.000	0.984	1.000
			$T = 300$		
$\delta = 0.6$	1.000	1.000	1.000	1.000	1.000
$\delta = 0.8$	1.000	1.000	1.000	1.000	1.000

Bootstrap-based test:

In small samples the size of the asymptotic test $\hat{\Gamma}_T$ may differ significantly from the significance level. Local bootstrap procedure of Paparoditis and Politis (2000) might help improve the size distortion of the asymptotic test.

Simulation setup is the same as in Asymptotic-based test.

Two sample sizes: $T = 50$, 75.

For each simulation we use $B = 199$ bootstrap replications.

Table 4: Empirical size of the asymptotic and bootstrap-based tests

Bandwidth	DGPs							
	DGP S1		DGP S2		DGP S3		DGP S4	
	Asy	Boot	Asy	Boot	Asy	Boot	Asy	Boot
				$T = 50$				
$\delta = 0.6$	0.062	0.046	0.071	0.048	0.031	0.064	0.049	0.048
$\delta = 0.8$	0.081	0.048	0.079	0.052	0.051	0.062	0.071	0.050
				$T = 75$				
$\delta = 0.6$	0.049	0.048	0.057	0.052	0.029	0.060	0.063	0.038
$\delta = 0.8$	0.066	0.052	0.060	0.053	0.054	0.064	0.067	0.032

Table 5: Empirical power of the asymptotic and bootstrap-based tests

Bandwidth	DGPs									
	DGP P1		DGP P2		DGP P3		DGP P4		DGP P5	
	Asy	Boot	Asy	Boot	Asy	Boot	Asy	Boot	Asy	Boot
$T = 50$										
$\delta = 0.6$	0.582	0.232	0.586	0.210	0.836	0.398	0.482	0.254	0.521	0.284
$\delta = 0.8$	0.637	0.234	0.633	0.212	0.831	0.390	0.562	0.242	0.598	0.292
$T = 75$										
$\delta = 0.6$	0.741	0.394	0.735	0.326	0.948	0.700	0.706	0.494	0.780	0.488
$\delta = 0.8$	0.789	0.406	0.774	0.400	0.949	0.624	0.720	0.504	0.811	0.494

7. Empirical application

We quantify the degree of nonlinear predictability of risk premium (expected stock excess return) across investment horizons using variance risk premium.

Variance risk premium is defined as the difference between risk-neutral and objective expectations of realized variance.

Most existing works [see Bollerslev, Tauchen and Zhou (2009)] focus on linear predictability. The econometric methodology used is an OLS regression.

Data

Monthly aggregate S&P 500 index over the period January 1996 to September 2008.

Monthly realized variance and implied variance to compute monthly variance risk premium.

Our analysis is based on the logarithmic return of the S&P 500 in excess of the 3-month T-bill rate.

Table 6: Measures of causality (predictability) from variance risk premium to risk premium

Direction of Causality	Bandwidth $T^{-1/(2+\delta)}$	Estimate of Causality Measure	Test-Statistic
		Horizon: One Month	
$VRP \rightarrow RP$			
	$\delta = 0.6$	1.298***	**6.149**
	$\delta = 0.8$	1.307***	**6.222**
		Horizon: Three Months	
$VRP \rightarrow RP$			
	$\delta = 0.6$	0.436***	**2.690**
	$\delta = 0.8$	0.468***	**2.886**
		Horizon: Six Months	
$VRP \rightarrow RP$			
	$\delta = 0.6$	0.170	0.797
	$\delta = 0.8$	0.239	1.124
		Horizon: Nine Months	
$VRP \rightarrow RP$			
	$\delta = 0.6$	0.000	0.000
	$\delta = 0.8$	0.016	0.085

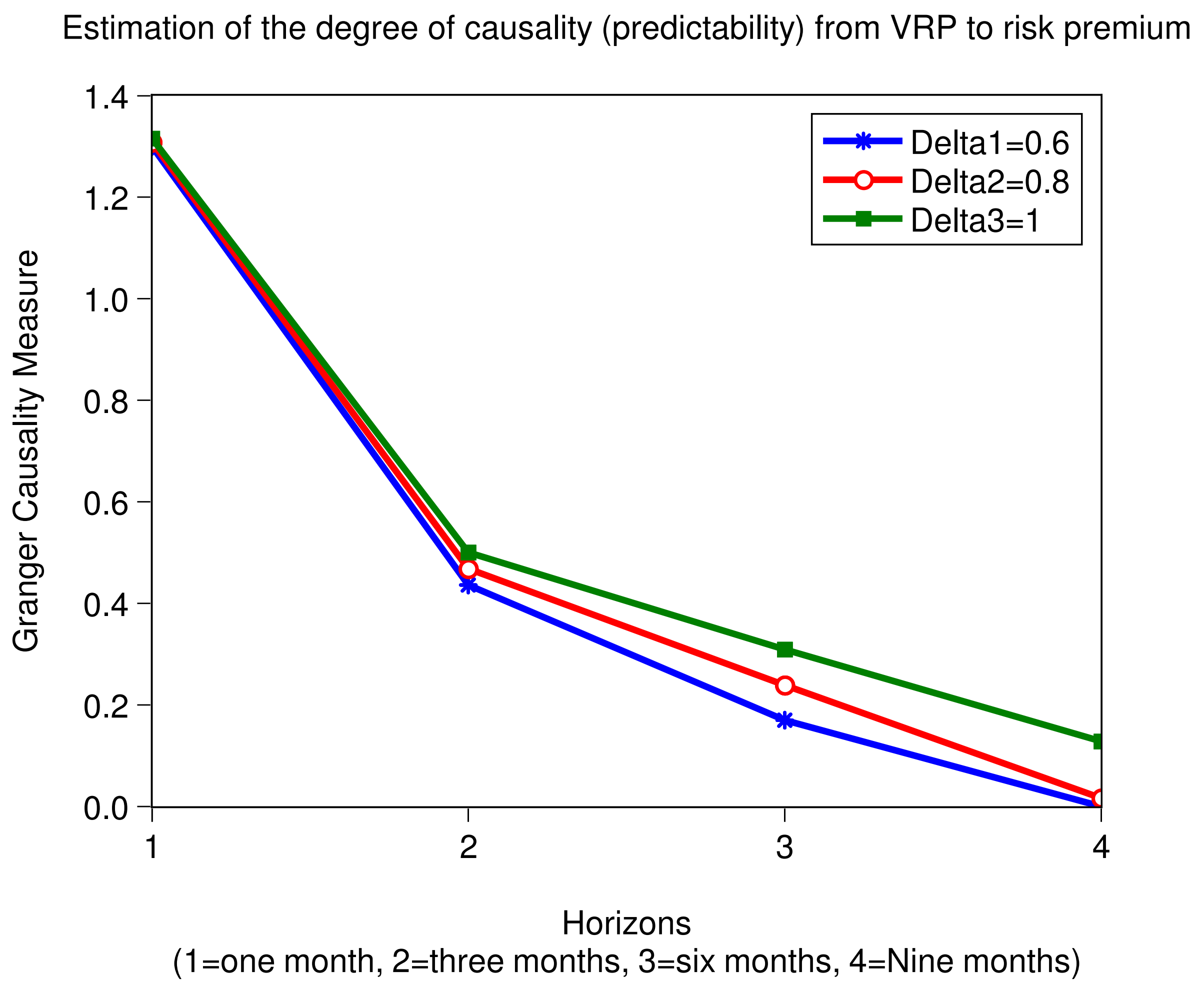

Figure 1: This figure plots the estimates of Granger causality measures from variance risk premium (VRP) to risk premium at horizons that go from one to 9 months, and using different bandwidth parameter δ (in the figure delta).

Conclusion

We propose model-free measures to quantify nonlinear causality-in-mean.

We establish the consistency and the asymptotic normality of nonparametric estimator of our measures.

The test which is based on the asymptotic normality is consistent and has nontrivial power against $\sqrt{T}$-local alternatives.

Monte Carlo simulations reveal that asymptotic and bootstrap-based tests behave quite well and have good finite sample size and power properties for a variety of typical data generating processes and different sample sizes.

Degree of nonlinear predictability of equity risk premium using variance risk premium is stronger one-month ahead and decreases with the investment horizon.

Additional Results: Comparison with Nishiyama et al. (2011):

Nishiyama, Hitomi, Kawasaki, and Jeong (2011) have proposed a nonparametric test for testing nonlinear causality-in-mean. Asymptotic distribution of their test is not normal and the critical values are obtained using simulations.

Our test $\hat{\Gamma}_T$ can also be used for testing nonlinear causality-in-mean.

We compare the size and power of our test with those of Nishiyama et al. (2011)'s test using the same simulation settings as in Nishiyama et al. (2011) [Table 6].

We compare the local power of our test with the local power of Nishiyama et al. (2011)'s test using the same $\sqrt{T}$-local-alternatives as in Nishiyama et al. (2011) [Table 7].

Table 7: Nishiyama et al. (2011)'s data-generating processes

DGPs	Variables of Interest		Direction of Causality in the DGP
	Y_t	X_t	
DGP 0	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP 1	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP 2	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1} + 0.4\sin(-2Y_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP 3	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1}^2 + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$

Table 8: Data-generating processes (DGPs): Local alternatives

DGPs	Variables of Interest		Direction of Causality in the DGP
	Y_t	X_t	
DGP 1L	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + \frac{2}{\sqrt{T}}Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP 2L	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + \frac{2}{\sqrt{T}}Y_{t-1} + \frac{4}{\sqrt{T}}\sin(-2Y_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP 3L	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + \frac{2}{\sqrt{T}}Y_{t-1}^2 + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$

Table 9: Empirical size and power of our test and Nishiyama et al. (2011)'s test

Bandwidth Parameters: $\bar{h} = T^{-1/(2.8)}$, $h = T^{-1/5}$								
Sample size	DGPs							
	DGP 0		DGP 1		DGP 2		DGP 3	
	ST	NHKJ	ST	NHKJ	ST	NHKJ	ST	NHKJ
$T = 100$	0.058	0.049	0.158	0.092	0.341	0.115	0.343	0.109
$T = 200$	0.051	0.047	0.237	0.236	0.415	0.335	0.503	0.286
$T = 300$	0.042	0.049	0.356	0.455	0.588	0.614	0.695	0.565

Table 10: Empirical local power of our test and Nishiyama et al. (2011)'s tes

Bandwidth Parameters: $\bar{h} = T^{-1/(2.8)}$, $h = T^{-1/5}$						
Sample size	DGPs					
	DGP 1L		DGP 2L		DGP 3L	
	ST	NHKJ	ST	NHKJ	ST	NHKJ
$T = 200$	0.130	0.133	0.225	0.166	0.272	0.128
$T = 300$	0.139	0.164	0.215	0.228	0.222	0.137
$T = 400$	0.138	0.151	0.239	0.213	0.289	0.138
$T = 500$	0.174	0.166	0.236	0.213	0.276	0.144
$T = 1000$	0.209	0.157	0.299	0.227	0.319	0.162