Identifying nonlinear components by random fields in the US GNP growth. Implications for the shape of the business cycle*

Christian M. Dahl
Department of Economics
Purdue University
dahlc@mgmt.purdue.edu

Gloria González-Rivera
Department of Economics
University of California, Riverside
gloria.gonzalez@ucr.edu

Abstract

Within a flexible parametric regression framework (Hamilton, 2001) we provide further evidence on the existence of a nonlinear component in the quarterly growth rate of the US real GNP. We implement a battery of new tests for neglected nonlinearity based on the theory of random fields (Dahl and González-Rivera, 2003). We find that the nonlinear component is driven by the fifth lag of the growth rate. We show that our model is superior to linear and nonlinear parametric specifications because produces a business cycle that when dissected with the BBQ algorithm mimics very faithfully the characteristics of the actual US business cycle. On understanding the relevance of the fifth lag, we find that the nonparametrically estimated conditional mean supports parametric specifications that allow for three phases in the business cycle: rapid linear contractions, aggressive short-lived convex early expansions, and moderate/slow relatively long concave late expansions.

JEL classification: C12, C15, C22.
Key Words: Flexible regression; Random field; Neglected Nonlinearity; BBQ algorithm.

* We would like to thank Heather Anderson, Niels Haldrup, Jim Hamilton, Svend Hylleberg, and seminar participants at the University of Aarhus, UC Riverside, UC Davis, and the SNDE Meetings in Atlanta. We are also grateful to two anonymous referees and the Associate Editor for their constructive comments. All errors remain our responsibility. González-Rivera acknowledges the financial support of the Academic Senate of the University of California, Riverside. The scientific notation follows the standard proposed by Abadir and Magnus (2002).

1 Introduction

During the last decade, we have witnessed the development of a plethora of nonlinear models to analyze the growth rate of the US GNP. Hamilton (1989) Markov regime switching model is perhaps the most well known univariate time series specification due to its ability to classify the NBER recessions with remarkable precision. Later contributions expanded the Hamilton's model in a variety of ways, such as including three regimes, introducing a state-space model with Markov switching, introducing heteroscedastic variances, etc. Other preferred specifications were the parametric threshold models as in Potter (1995) with a variety of extensions such as threshold autoregressive, threshold moving average, self-exciting threshold with two and four regimes, etc. The performance of these models is mixed depending upon the evaluation criteria. When their performance is evaluated with statistical criteria, such as forecasting ability, we seem to have a negative view of nonlinear specifications because simple linear models tend to perform better or as well as nonlinear models (Stock and Watson, 1999). Recently, Harding and Pagan (2002) proposed to revisit the traditional approach of Burns and Mitchell (1946) grounded on the location of the turning points of a series and other graphical methods to explain the shape of the business cycle. The performance of the many statistical models is judged according to their ability to mimic the characteristics of the actual business cycle. With this yardstick, Harding and Pagan concluded that there is "little evidence that certain non-linear effects are important to the nature of business cycle". The nonlinear models they considered were the Hamilton model and a Markov switching model with duration dependence due to Durland and McCurdy (1994). A more extensive collection of nonlinear models can be found in Galvão (2002) where she evaluated fifteen models according to their ability to produce faster growth rates after a trough and slower before a peak. According to her results, the best model to reproduce such a feature of the business cycle is a three-regime Markov switching with heteroscedastic variances as in Clements and Krolzig (1998). Interestingly absent of this academic debate are nonparametric or semiparametric specifications that given their generality -there is no assumption in the functional form of the conditional mean- may seem to be a very appropriate tool to guide the choice of a parametric model. This paper contributes to this end.

Hamilton (2001) proposed a flexible parametric regression model where the conditional mean has a linear parametric component and a potential nonlinear component represented by an isotropic Gaussian random field. The model has a nonparametric flavor because no functional form is assumed but, nevertheless, the estimation is fully parametric. In this paper, we implement the flexible parametric regression model for statistical identification of a nonlinear component in the quarterly US real GNP growth rate. Previous to the estimation of the model, we test for neglected nonlinearity in the conditional mean with a battery of new powerful tests based on the theory of random fields developed by Dahl and González-Rivera (2003). We start with the choice of the linear regressors and we assume that the nonlinear component is a function of either the full set or a subset of the linear regressors. Following most of the parametric models, the preferred choices are linear and nonlinear AR(4), as in Hamilton (1989) and Durland and McCurdy (1994), and an AR(5) as in Potter (1995). Our approach is rather conservative in that we tend to choose a relative large number of lags in the linear component to guard against dynamic misspecification making the detection of nonlinearity, either in testing or in estimation, a more difficult exercise. Nevertheless, we find strong evidence for nonlinearity in the US GNP growth rate, clearly driven by one regressor, the fifth lag of the growth rate. This finding offers some support to the Potter's specification where the nonlinear component is driven by three variables: the first, second and fifth lags of the growth rate. The assessment of our model is conducted within the Harding and Pagan (2002) context. We dissect the business cycle according to measures of duration, amplitude, cumulation and excess cumulation of the contraction and expansion phases of the business cycle. Contractions are believed to be linear as opposed to expansions, which are mostly concave. This asymmetry is the challenge that any statistical specification should be able to replicate. We find that the flexible parametric regression model is able to mimic the characteristics of the actual US business cycle. It represents a clear improvement over linear models, in contrast to Harding and Pagan (2002) findings, and seems to capture just the right shape of the expansion phase as opposed to Hamilton (1989) and Durland and McCurdy (1994) models, which tend to overestimate the cumulation measure in the expansion phase. We are intrigued by the contribution of the fifth lag of the growth rate to the nonlinearity of the model. We explore this issue by examining, in the plane (y_{t-5}, y_t), the conditional mean of the growth rate estimated with the random field technology and mapping the changing signs of the autocorrelation between *growth rates* y_{t-5} and y_t into dynamic paths of the GNP *level.* We find that the expansion phase must have at least two subphases: an aggressive early expansion after the trough, and a moderate/slow late expansion before the peak implying the existence of an inflexion point that we date approximately around one-third into the duration of the expansion phase. This shape lends support to parametric models of the growth rate that allow for three regimes (Sichel, 1994), as opposed to models with just two regimes (contractions and expansions).

The organization of the paper is as follows. In section 2, we review Hamilton (2001) flexible parametric regression model where a random field models the nonlinear component of the conditional mean. In section 3, we summarize the tests for neglected nonlinearity based on the theory of random fields proposed by Hamilton (2001) and Dahl and González-Rivera (2003). In section 4, we engage in an extensive modelling exercise of the post-war growth rate of the US real GNP, which constitutes the core of this paper. We test for nonlinearity and identify the nonlinear component by estimation. Both types of statistical inference, estimation and testing, are in agreement on identifying the source of nonlinearity. We proceed with the dissection of the business cycle, exploring the shape of contractions and expansion. Finally, in section 5, we offer a brief set of conclusions.

2 Random fields and flexible parametric regression

Hamilton (2001) suggested representing the nonlinear component in a general regression model by a Gaussian homogeneous and isotropic scalar random field. A scalar random field is defined as a function $m(\omega, \boldsymbol{x}) : \Omega \times A \to \mathbb{R}$ such that $m(\omega, \boldsymbol{x})$ is a random variable for each $\boldsymbol{x} \in A$ where $A \subseteq \mathbb{R}^k$. A random field is also denoted as $m(\boldsymbol{x})$. If $m(\boldsymbol{x})$ is a system of random variables with finite dimensional Gaussian distributions, then the scalar random field is said to be Gaussian and it is completely determined by its mean function $\mu(\boldsymbol{x}) = \mathrm{E}\left[m(\boldsymbol{x})\right]$ and its covariance function with typical element $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{E}\left[(m(\boldsymbol{x}) - \mu(\boldsymbol{x}))(m(\boldsymbol{z}) - \mu(\boldsymbol{z}))\right]$ for any $\boldsymbol{x}, \boldsymbol{z} \in A$. The random field is said to be homogeneous or stationary if $\mu(\boldsymbol{x}) = \mu$ and the covariance function depends only on the difference vector $\boldsymbol{x} - \boldsymbol{z}$ and we should write $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{C}(\boldsymbol{x} - \boldsymbol{z})$. Furthermore, the random field is said to be isotropic if the covariance function depends on $d(\boldsymbol{x}, \boldsymbol{z})$, where $d(\cdot)$ is a scalar measure of distance. In this situation we write $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{C}(d(\boldsymbol{x}, \boldsymbol{z}))$.

The specification suggested by Hamilton (2001) can be represented as

$$y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \lambda m(\boldsymbol{g} \odot \boldsymbol{x}_t) + \epsilon_t, \tag{1}$$

for $y_t \in \mathbb{R}$ and $\boldsymbol{x}_t \in \mathbb{R}^k$, both stationary and ergodic processes. The conditional mean has a linear component given by $\beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1$ and a non-linear component given by $\lambda m(\boldsymbol{g} \odot \boldsymbol{x}_t)$, where $m(\boldsymbol{z})$, for any choice of $\boldsymbol{z}$, represents a realization of a Gaussian and homogenous random field with a moving average representation; $\boldsymbol{x}_t$ could be predetermined or exogenous and is independent of $m(\cdot)$, and ϵ_t is a sequence of independent and identically distributed $\mathrm{N}(0, \sigma^2)$ variates independent of both $m(\cdot)$ and $\boldsymbol{x}_t$ as well as of lagged values of $\boldsymbol{x}_t$. The scalar parameter λ represents the contribution of the nonlinear part to the conditional mean, the vector $\boldsymbol{g} \in \mathbb{R}^k_{0,+}$ drives the curvature of the conditional mean, and the symbol $\odot$ denotes element-by-element multiplication.

Let $\boldsymbol{H}_k$ be the covariance (correlation) function of the random field $m(\cdot)$ with typical element defined as $\boldsymbol{H}_k(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{E}\left[m(\boldsymbol{x})m(\boldsymbol{z})\right]$. Hamilton (2001) proved that the covariance function depends solely upon the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{z}$, rendering the random field isotropic. For any $\boldsymbol{x}$ and $\boldsymbol{z} \in \mathbb{R}^k$, the correlation between $m(\boldsymbol{x})$ and $m(\boldsymbol{z})$ is given by the ratio of the volume of the overlap of k-dimensional unit spheroids centered at $\boldsymbol{x}$ and $\boldsymbol{z}$ to the volume of a single k-dimensional unit spheroid. If the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{z}$ is greater than two, the correlation between $m(\boldsymbol{x})$ and $m(\boldsymbol{z})$ will be equal to zero. The general expression of the correlation function is

$$\begin{aligned} \boldsymbol{H}_k(h) &= \begin{cases} G_{k-1}(h,1)/G_{k-1}(0,1) & \text{if } h \leq 1 \\ 0 & \text{if } h > 1 \end{cases}, \\ G_k(h,r) &= \int_h^r (r^2 - w^2)^{k/2} dw, \end{aligned} \tag{2}$$

where $h \equiv \frac{1}{2} d_{L_2}(\boldsymbol{x}, \boldsymbol{z})$, and $d_{L_2}(\boldsymbol{x}, \boldsymbol{z}) \equiv \left[(\boldsymbol{x} - \boldsymbol{z})'(\boldsymbol{x} - \boldsymbol{z})\right]^{1/2}$ is the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{z}$.[1]

Within the specification (1), Dahl and González-Rivera (2003) provided alternative representations of the random field that permit the construction of

[1] For a formal proof, see Theorem 2.2 in Hamilton (2001)

Lagrange multiplier tests for neglected nonlinearity, which circumvent the problem of unidentified nuisance parameters under the null of linearity (as we will see in the forthcoming sections) and, at the same time, they are robust to the specification of the covariance function associated with the random field. They modified the Hamilton framework in two directions. First, the random field is specified in the L_1 norm instead of the L_2 norm, and secondly they considered random fields that may not have a simple moving average representation. The advantage of the L_1 norm, which is exploited in the testing problem, is that this distance measure is a linear function of the nuisance parameters, in contrast to the L_2 norm which is a nonlinear function. Logically, Dahl and González-Rivera proceeded in an opposite fashion to Hamilton. Whereas Hamilton first proposed a moving average representation of the random field, and secondly, he derived its corresponding covariance function, Dahl and González-Rivera first proposed a covariance function, and secondly they inquire whether there is a random field associated with it. The proposed covariance function is

$$\boldsymbol{C}_k(h^*) = \begin{cases} (1-h^*)^{2k} & if\ h^* \leq 1 \\ 0 & if\ h^* > 1 \end{cases}, \tag{3}$$

where $h^* \equiv \frac{1}{2} d_{L_1}(\boldsymbol{x}, \boldsymbol{z}) = \frac{1}{2}|\boldsymbol{x} - \boldsymbol{z}|'\boldsymbol{\imath}$. The function (3) is a permissible covariance, that is, it satisfies the positive semidefiniteness condition, which is $\boldsymbol{q}'\boldsymbol{C}_k\boldsymbol{q} \geq 0$ for all $\boldsymbol{q} \neq \boldsymbol{0}_T$. Furthermore, there is a random field associated with it according to the Khinchin's theorem (1934) and Bochner's theorem (1959). The basic argument is that the class of functions which are covariance functions of homogenous random fields coincides with the class of positive semidefinite functions. Hence, (3) being a positive semidefinite function must be the covariance function of a homogenous random field.

3 Testing for nonlinearities

We shortly describe the two most important tests for neglected nonlinearities proposed by Dahl and González-Rivera (2003). Consider the model given by equation (1). The contribution of the non-linear component to the conditional mean is driven by the parameter λ and/or by the parameter vector $\boldsymbol{g}$. It is easy to observe that a test for neglected nonlinearity will be subject to a nuisance parameter problem, where a set of parameters are identified only under the alternative hypothesis. There are two alternative approaches to specify the null hypothesis of linearity: (i) If the null hypothesis is written as $H_0 : \lambda^2 = 0$, the parameter vector $\boldsymbol{g}$ is unidentified under the null and the number of unidentified parameters increases with the dimensionality of the model. Tests with a null hypothesis $H_0 : \lambda^2 = 0$ are denoted λ-tests. (ii) If the null hypothesis is written as $H_0 : \boldsymbol{g} = \mathbf{0}_k$, the parameter λ becomes unidentified under the null. Furthermore, the contribution of the nonlinear component becomes indistinguishable from that of the constant β_0, which becomes also unidentifiable. However, in this case, the number of unidentified parameters remains the same whenever the dimensionality of the model increases. Tests with a null hypothesis $H_0 : \boldsymbol{g} = \mathbf{0}_k$ are denoted g-tests.

From model (1), we can write $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X\beta}, \lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T)$ where $\boldsymbol{y} = (y_1, y_2, ..., y_T)'$, $\boldsymbol{X}_1 = (\boldsymbol{x}_1', \boldsymbol{x}_2', ..., \boldsymbol{x}_T')'$, $\boldsymbol{X} = (\boldsymbol{\imath} : \boldsymbol{X}_1)$, $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$, $\boldsymbol{\varepsilon} = (\epsilon_1, \epsilon_2, ..., \epsilon_T)'$ and σ^2 is the variance of ϵ_t. $\boldsymbol{C}_k$ is a generic covariance function associated with the random field, which could be equal to the Hamilton spherical covariance function in (2), or to the covariance in (3). The log-likelihood function corresponding to this model is

$$\ell(\boldsymbol{\beta}, \lambda^2, \boldsymbol{g}, \sigma^2) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T| - \frac{1}{2} (\boldsymbol{y} - \boldsymbol{X\beta})' (\lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T)^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}), \quad (4)$$

which is the basis for the Lagrange multiplier tests for neglected nonlinearity. Note that the covariance function $\boldsymbol{C}_k$ depends on the parameters $\boldsymbol{g}$. The Lagrange multiplier statistic is given by $LM = \boldsymbol{s}(\tilde{\vartheta})' \boldsymbol{\mathcal{I}}^{-1}(\tilde{\vartheta}) \boldsymbol{s}(\tilde{\vartheta})$, where $\boldsymbol{s}(\tilde{\vartheta})$ denotes the score function, and $\boldsymbol{\mathcal{I}}^{-1}(\tilde{\vartheta})$ the inverse of the information matrix, both evaluated under a generic null hypothesis $H_0 : \vartheta = \tilde{\vartheta}$. We consider two different estimators of the information matrix, both consistent under the null. The first estimator, denoted $\boldsymbol{\mathcal{I}}_H$, is based on the Hessian of the log likelihood function, and the second, denoted $\boldsymbol{\mathcal{I}}_{OP}$, is based on the outer-product of the score. Under the usual regularity conditions, the Lagrange multiplier statistics will be $\chi^2(q)$-distributed where q equals the number of restrictions under the null.

3.1 λ-tests

Hamilton (2001) derived the λ-test for neglected nonlinearity based on the $\boldsymbol{\mathcal{I}}_H$ estimator of the information matrix and with $\boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) = \boldsymbol{H}_k(h)$ for $h = \frac{1}{2}d_{L_2}(\boldsymbol{x}_t, \boldsymbol{x}_s)$. We denote Hamilton's test statistic $\lambda_H^E(\boldsymbol{g})$. To deal with the identification issues, Hamilton suggested fixing $\boldsymbol{g}$ to the mean of its prior distribution and proceed to derive the Lagrange multiplier test, which then follows a standard asymptotic distribution. Dahl and González-Rivera (2003) derived the TR^2 version of the Lagrange multiplier tests based on the $\boldsymbol{\mathcal{I}}_{OP}$ estimator of the information matrix. First, they provided an analogous test to Hamilton's. Keeping $\boldsymbol{g}$ fixed, evaluating the score functions under the null $H_0 : \lambda^2 = 0$, and considering the covariance function (3), the test statistic is

$$\lambda_{OP}^E(\boldsymbol{g}) = \frac{T^2}{2}\frac{\boldsymbol{u}'\tilde{\boldsymbol{x}}(\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{x}}'\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{u}} \sim \chi^2(1),$$

where $\boldsymbol{u} = \text{vec}\,(\boldsymbol{I}_T - \frac{\varepsilon\varepsilon'}{\sigma^2})$, $\tilde{\boldsymbol{x}} = (\tilde{\boldsymbol{x}}_1 : \tilde{\boldsymbol{x}}_2)$, with $\tilde{\boldsymbol{x}}_1 = \text{vec}\,(\boldsymbol{C}_k)$ and $\tilde{\boldsymbol{x}}_2 = \text{vec}\,(\boldsymbol{I}_T)$. The statistic is easily obtained by the following procedure: 1. Estimate the model under the null and compute $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y})$ and $\widehat{\sigma}^2 = T^{-1}\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$. 2. Obtain the least squares estimate of $\boldsymbol{\nu}$ - denoted $\widehat{\boldsymbol{\nu}}$ - from the auxiliary regression $\widehat{\boldsymbol{u}} = \phi_1\tilde{\boldsymbol{x}}_1 + \phi_2\tilde{\boldsymbol{x}}_2 + \boldsymbol{\nu}$, using $\widehat{\boldsymbol{u}} = \text{vec}\,(\boldsymbol{I}_T - \frac{\widehat{\varepsilon}\widehat{\varepsilon}'}{\widehat{\sigma}^2})$. 3. Obtain the uncentered R^2 as $R^2 = 1 - \widehat{\boldsymbol{\nu}}'\widehat{\boldsymbol{\nu}}/\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}$. 4. Finally, the Lagrange multiplier statistic is given as $\lambda_{OP}^E(\boldsymbol{g}) = \frac{1}{2}T^2R^2$.

The second test proposed by Dahl and González-Rivera does not depend on the unidentified nuisance parameters $\boldsymbol{g}$ and does not depend on the exact parameterization of the covariance function. Let us start by considering the covariance function (3), i.e. $\boldsymbol{C}_k(h_{ts}^*) = (1 - h_{ts}^*)^{2k}1_{(h_{ts}^*\leq 1)}$, where $h_{ts}^* \equiv \frac{1}{2}d_{L_1}(\boldsymbol{g}\odot \boldsymbol{x}_t, \boldsymbol{g}\odot\boldsymbol{x}_s) = \frac{1}{2}\,\boldsymbol{r}_{ts}'\boldsymbol{g}$, and $\boldsymbol{r}_{ts} = \{|x_{t1}-x_{s1}|, |x_{t2}-x_{s2}|, ..., |x_{tk}-x_{sk}|\}'$. Observe the following: (i) h_{ts}^* is a linear function of the nuisance parameters $\boldsymbol{g}$; (ii) $(1-h^*)^{2k} = \sum_{j=0}^{2k}\binom{2k}{j}h^{*^j}(-1)^j$; and (iii) the indicator function $1_{(h_{ts}^*\leq 1)}$ can be approximated by a differentiable function like a logistic function, i.e. $1_{(h_{ts}^*\leq 1)} \approx (1 + \exp(-\gamma(1 - h_{ts}^*))^{-1}$ for fixed $\gamma \gg 0$. A second order Taylor expansion of the logistic function around an average value of h_{ts}^* together with the computation of the powers of h^*, that is h^{*^j}, give rise to the following auxiliary regression (step 2 of the procedure outlined in the previous paragraph)

$$\begin{aligned}
\hat{u}_{ts} &= \bar{\phi}_0 + \bar{\phi}_1 \sum_{i=1}^{k} g_i r_{ts,i} + \bar{\phi}_2 \sum_{i=1}^{k} \sum_{j \geq i}^{k} g_i g_j r_{ts,i} r_{ts,j} \\
&+ \bar{\phi}_3 \sum_{i=1}^{k} \sum_{j \geq i}^{k} \sum_{l \geq j}^{k} g_i g_j g_l r_{ts,i} r_{ts,j} r_{ts,l} + .. + \\
&+ \bar{\phi}_{2k+2} \sum_{i=1}^{k} \sum_{j \geq i}^{k} .. \sum_{m}^{k} g_i g_j .. g_m r_{ts,i} r_{ts,j} .. r_{ts,m} + \phi_2 \widetilde{x}_{2,ts} + \nu_{ts},
\end{aligned} \tag{5}$$

where $\bar{\phi}_j$ is directly proportional to ϕ_1, that is $\bar{\phi}_j = c_j \phi_1$ with c_j being the proportionality parameter. The subindex ts attached to the vectors $\hat{\boldsymbol{u}}, \widetilde{\boldsymbol{x}}_2$, and $\boldsymbol{\nu}$ means the ts^{th} entry/row in the respective vector for $t, s = 1, 2, ...T$, and g_i and $r_{ts,i}$ denote the i^{th} entry in the vectors $\boldsymbol{g}$ and $\boldsymbol{r}_{ts}$ respectively. The second order expansion of the logistic function linearizes the auxiliary regression at the expense of increasing the number of regressors. Notice that now we can proceed to estimate the auxiliary regression (5) by OLS, treating the nuisance vector $\boldsymbol{g}$ as part of the parameter space. In the computation of the test $\lambda_{OP}^{E}(\boldsymbol{g})$, $\phi_1 = 0$ implied that the null hypothesis of linearity cannot be rejected. Considering the regression (5), $\phi_1 = 0$ implies that $\bar{\phi}_j = 0,\ j = 0, 1, 2....2k + 2$. Hence, the regression (5), where $\sum_{j=1}^{2k+2} \binom{k+j-1}{k-1}$ regressors have been added, is the basis to compute a Lagrange multiplier test that is free of nuisance parameters but, in this case, the new test denoted λ_{OP}^{A} will be χ^2-distributed with $1 + \sum_{j=1}^{2k+2} \binom{k+j-1}{k-1}$ degrees of freedom.

The auxiliary regression (5) is based on the covariance function (3). Dahl and González-Rivera argued that (3) represents a very broad class of covariance functions. For a homogenous and isotropic random field, if the covariance function is differentiable, it can be approximated by a Taylor's expansion around an average value of h^*. In this case, Dahl and González-Rivera showed that the auxiliary regression needed to construct the Lagrange multiplier statistic will be the same as (5). Consequently, the λ_{OP}^{A} test, described in the previous paragraph, is also a test for nonlinearity when the covariance function is unknown but it can be approximated reasonably well by a high order Taylor's expansion. Dahl and González-Rivera provided some examples of permissible covariance functions whose characteristics can be captured by the λ_{OP}^{A} test.

3.2 g-tests

Under the null hypothesis $H_0 : \boldsymbol{g} = \boldsymbol{0}_k$, model (1) becomes $y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \lambda m(\boldsymbol{0}_k) + \epsilon_t$, where $m(\boldsymbol{0}_k) \sim \mathrm{N}(0,1)$. The model is linear on $\boldsymbol{x}_t$, but y_t will be a non-ergodic process. Consider the simplest case where $\boldsymbol{x}_t$ is deterministic. In this case, we have that $\mathrm{cov}(y_t y_{t-s}) = \lambda^2$ for any s. Ergodicity is a critical assumption for the *law of large numbers* to hold; we need consistency of the parameters of the model under the null because our tests are based on the null residuals (see step 1 in the previous section). If the process is non-ergodic, a test for nonlinearity based on the parameter vector $\boldsymbol{g}$ may not have a well defined asymptotic distribution under the null.

A simple modification of the specification of the function $m(\boldsymbol{x})$ will preserve the ergodicity of y_t under the null. If we write the modified unrestricted model as $y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \lambda\tilde{m}(\boldsymbol{g} \odot \boldsymbol{x}_t) + \epsilon_t$, where $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x}) - m(\boldsymbol{0}_k)$, notice that $\tilde{m}(\boldsymbol{0}_k) = 0$. The model under the null becomes $y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \epsilon_t$ restoring the ergodicity of y_t under the null hypothesis, provided that $\boldsymbol{x}_t$ and ϵ_t are stationary and ergodic. Let $\widetilde{\boldsymbol{C}}_k$ be the covariance function that uniquely determines the random field $\tilde{m}(\boldsymbol{x})$. The typical element in $\widetilde{\boldsymbol{C}}_k$ is defined as $\widetilde{\boldsymbol{C}}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) = \mathrm{E}\{m(\boldsymbol{x}_t) - m(\boldsymbol{0}_k)\}\{m(\boldsymbol{x}_s) - m(\boldsymbol{0}_k)\}$, hence the covariance function can be written as $\widetilde{\boldsymbol{C}}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) = \boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) + \boldsymbol{C}_k(\boldsymbol{0}_k, \boldsymbol{0}_k) - \boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{0}_k) - \boldsymbol{C}_k(\boldsymbol{0}_k, \boldsymbol{x}_s)$. Yaglom (1962, pp. 87) names $\widetilde{\boldsymbol{C}}_k$ the structure function.

To construct the g-tests, we assume that $\boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{x}_s)$ has the parametric form of (3), with $h_{ts}^* \equiv \frac{1}{2} d_{L_1}(\boldsymbol{g} \odot \boldsymbol{x}_t, \boldsymbol{g} \odot \boldsymbol{x}_s) = \frac{1}{2}\, \boldsymbol{r}_{ts}'\boldsymbol{g}$. The likelihood function is as (4) where $\boldsymbol{C}_k$ is replaced by $\tilde{\boldsymbol{C}}_k$. The derivation of the LM test proceeds as in the previous section. In the case of the g-tests, the λ parameter is unidentified under the null hypothesis and on calculating the score function we need to keep λ fixed. Evaluating the score function under the null of linearity $H_0 : \boldsymbol{g} = 0$, and keeping λ fixed, the score functions in vectorized form are

$$s(g_i)|_{\lambda^2, \boldsymbol{g}=0} = -\frac{\lambda^2}{2\sigma^2}\widetilde{\boldsymbol{x}}_i'\boldsymbol{u}, \qquad i = 1, 2, .., k, \tag{6}$$

$$s(\sigma^2)|_{\lambda^2, \boldsymbol{g}=0} = -\frac{1}{2\sigma^2}\widetilde{\boldsymbol{x}}_{k+1}'\boldsymbol{u}, \tag{7}$$

where $\tilde{\boldsymbol{x}}_i = \frac{\partial \,\mathrm{vec}\,(\widetilde{\boldsymbol{C}}_k)}{\partial g_i}|_{g=0}$, for $i = 1, 2, .., k$, $\tilde{\boldsymbol{x}}_{k+1} = \mathrm{vec}\,(\boldsymbol{I}_T)$, and $\boldsymbol{u} = \mathrm{vec}\,(\boldsymbol{I}_T - \frac{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'}{\sigma^2})$.[2]

[2] To calculate $\tilde{\boldsymbol{x}}_i$, the indicator function has been substituted for a logistic function, i.e.

With the scores (6) and (7), we compute the g-test as a TR^2 statistic, which is free of the nuisance parameter λ. We denote such a test by g_{OP}. The construction of the test statistic follows the procedure already outlined in the previous section. After having obtained $\hat{\varepsilon}$ and $\hat{\sigma}^2$ compute the uncentered R^2 from the auxiliary regression $\widehat{u}_{ts} = \sum_{i=1}^{k} \widetilde{\phi}_i \widetilde{r}_{ts,i} + \widetilde{\phi}_{k+1} \widetilde{x}_{k+1,ts} + \tilde{\nu}_{ts}$, where $\widetilde{r}_{ts,i} = -k\left(|x_{ti} - x_{si}| - |x_{ti}| - |x_{si}|\right)$, for $t, s = 1, 2, .., T$. The Lagrange multiplier statistic is then given as $g_{OP} = \frac{1}{2} T^2 R^2 \sim \chi^2(k)$.

Notice that the g-test does not depend on the unidentified nuisance parameter λ. In order to increase power under the alternative hypothesis, the auxiliary regression can be augmented with higher powers and cross products of $\widetilde{r}_{ts}$, thereby increasing the number of degrees of freedom of the asymptotic distribution of the test.

4 Modeling quarterly real US GNP growth

Our main objective is to assess the potential nonlinearities in the quarterly growth rate of the real US GNP from 1947Q1 to 2000Q4 with this new set of tools based on the theory of random fields. To this end, we implement the tests for neglected nonlinearity reviewed in the previous section. Based on the testing results, we estimate a flexible parametric random field regression model and we present the functional form of the conditional expectation of the growth rate, which reveals the asymmetric behavior of the GNP growth rate in expansions and recessions. We analyze the characteristics of the business cycle generated by the random field regression model implementing the Bry and Boschan (BB) (1971) algorithm converted to quarterly frequency (BBQ) by Harding and Pagan (2002). The BBQ algorithm permits the description of the duration and amplitude of the cycle and its phases, the asymmetric behavior of the phases, and the cumulative movements within phases.

4.1 Testing for neglected nonlinearity

The specification of the model under the null (the linear component) has a potential large impact on the family of tests for neglected nonlinearity that are based

$1_{(h^*_{ts} \leq 1)} \approx (1 + \exp(-\gamma(1 - h^*_{ts})))^{-1}$ for fixed $\gamma \gg 0$.

on the estimated residuals under the null. Including a large number of regressors in the model under the null, combined with the relative robustness of the linear model against moderate deviations from linearity (the linear model as a first order Taylor approximation to a nonlinear specification) makes the residuals behave as a white noise process. This implies that a possible neglected nonlinear component may be very hard to detect when the model under the null is not parsimoniously specified. On the other hand, imposing too many restrictions on the linear model under the null may result in too many rejections because many tests for neglected nonlinearity based on the Lagrange multiplier principle will not be able to distinguish dynamic misspecification from misspecification of the functional form.

Our approach to selecting the linear component of the model is rather conservative. We use the well known AIC and BIC criteria and we report the results in Table 1.

[Table 1 about here]

The AIC criteria selects the AR(5) model but with zero restrictions in the third and four lags providing support for the linear component in Potter's SETAR model. The BIC criteria selects a more parsimonious specification, an AR(1), however the AR(5) with some zero restrictions comes as the second choice. In any case, it is questionable whether the model selection criteria are able to select the number of lags in the linear component independently from the lags in the nonlinear component. Ideally we would like to have test statistics that are able to select the lag structure and the nature of the nonlinearity jointly. In the absence of such a procedure, when we test for neglected nonlinearity, we will maintain all the specifications considered in Table 1.

In Table 2, we report the bootstrapped p-values of the four tests for neglected nonlinearity based on the theory of random fields[3]. These are: $\lambda_H^E(\boldsymbol{g})$ is the Hamilton's Lagrange multiplier statistic based on the Hessian representation of the information matrix and on the spherical variance-covariance matrix; $\lambda_{OP}^E(\boldsymbol{g})$ is the test based on the outer product of the score and on the variance-covariance matrix (3) proposed by Dahl and González-Rivera (2003); λ_{OP}^A is the test based on a higher order Taylor approximation to the variance-covariance function; and

[3] The bootstrap procedure is described in Dahl and González-Rivera (2003).

g_{OP} is the statistic explained in section 3.2. All the four tests considered are based on the estimated residuals from the model estimated under the null hypothesis of linearity. The first column of Table 2 provides the specification of the linear component of the model and the second column the regressors considered in the nonlinear component. These are the elements of the auxiliary regressions explained in section 3.1 and 3.2. When we deal with a large number of regressors in the nonlinear component of the model, the implementation of the auxiliary regression required in the test λ_{OP}^{A} may be difficult because we may run out of degrees of freedom for moderate to small sample sizes. In this case, we reduce the auxiliary regression to second order terms or, alternatively, we could remove the terms involving cross-products and leaving the high power terms. Table 2 has three panels. In the first panel, we investigate whether Hamilton's Markov switching (MS)-AR(4) model is plausible. The MS-AR(4) model states that up to four lags of real US GNP growth should be included in the linear as well as in the nonlinear part. In the second panel, we consider a nonlinear component driven by the first five lags; and in the third panel, we consider the case suggested by Potter (1995) where only y_{t-1}, y_{t-2}, and y_{t-5} contribute to the nonlinear component.

[Table 2 about here]

In the first panel, the four tests fail to reject the null hypothesis of linearity very strongly. Their p-values are very large providing no evidence of a nonlinear component driven by the first four lags of the GNP growth rate. However, when the fifth lag is introduced (second panel), we observe a dramatic reduction in the p-values of the λ_{OP}^{A} and g_{OP} tests. On the other hand, the p-values of the $\lambda_{H}^{E}(\boldsymbol{g})$ and $\lambda_{OP}^{E}(\boldsymbol{g})$ tests either become larger or remain at the same level. This is a reflection of the nuisance parameter problem that we argued in Dahl and González-Rivera (2003) and that motivated the introduction of the λ_{OP}^{A} and g_{OP} tests. When the dimensionality of the model increases, the number of nuisance parameters grows and the $\lambda_{H}^{E}(\boldsymbol{g})$ and $\lambda_{OP}^{E}(\boldsymbol{g})$ tests, which depend on the fixed vector $\boldsymbol{g}$, tend to loose power; on the contrary, the λ_{OP}^{A} and g_{OP} tests are immune to the number of nuisance parameters and are more powerful. In the third panel, when the nonlinear component is a function of y_{t-1}, y_{t-2}, and y_{t-5}, we find that the $\lambda_{H}^{E}(\boldsymbol{g})$, λ_{OP}^{A} and g_{OP} tests reject linearity at the customary 5 or 10 % significance levels. The preferred specification is a linear component represented either by an AR(5) or an AR(2), and a nonlinear component driven by y_{t-1}, y_{t-2}, and y_{t-5}. As soon as the number of nuisance parameters is reduced, the p-values of the $\lambda_H^E(\boldsymbol{g})$ and $\lambda_{OP}^E(\boldsymbol{g})$ become smaller (compare the third panel with the first), and the $\lambda_H^E(\boldsymbol{g})$ test is able to reject the null of linearity at the 10% significance level.

In summary, the overall assessment of the four tests for neglected nonlinearity is that there is a nonlinear component in the quarterly growth rate of real US GNP most likely driven by y_{t-5}, and marginally, by y_{t-1}, and y_{t-2}. There are three potential specifications that emerge. The most parsimonious is a linear AR(2) with a nonlinear component that is a function of y_{t-1}, y_{t-2}, and y_{t-5}. The next is a model with y_{t-1}, y_{t-2}, and y_{t-5} in the linear and nonlinear components, and the last is a linear AR(5) with nonlinearities due to y_{t-1}, y_{t-2}, and y_{t-5}. We proceed conservatively choosing to estimate the latter.

4.2 Identifying the nonlinear component by estimation

We estimate a fully parameterized random field regression model as in Hamilton (2001). Letting $\lambda = \varsigma \times \sigma$, we consider the following reparameterized version of model (1)

$$\begin{aligned} y_t &= \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \sigma\{\varsigma\, m(\boldsymbol{g} \odot \boldsymbol{x}_t) + \eta_t\} \\ \eta_t &\rightarrow N(0,1) \end{aligned}$$

where $\boldsymbol{x}_t = \{y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}\}'$. The estimation and statistical significance of ς is of paramount importance to assert the existence of nonlinearities. Conditional on ς being significant, we produce parametric inference on the components that drive the nonlinear part of the model.

In Tables 3 and 4, we report the results of maximum likelihood estimation and Bayesian estimation of the above model for two different random fields. In Table 3, we consider a random field with the spherical covariance function, and, in Table 4, we deal with the covariance function (3) proposed by Dahl and González-Rivera (2003). In both tables, we supply Bayesian estimates of the parameters and their standard errors based on Hamilton's "t/normal"-mixture importance sampler (Hamilton, 2001), and on Geweke's "split-t/normal"-mixture importance sampler (Geweke, 1989). The specification of the model consists of an AR(5) in the linear component and a nonlinear component driven by y_{t-1}, y_{t-2}, and y_{t-5}.

The estimation results of Table 3 support very strongly the nonlinearity of the model. The estimate $\hat{\varsigma}$ is highly significant and the magnitude of the estimates is statistically similar across estimation techniques. The maximum likelihood estimates of g_1, g_2, and g_5 are all significant at the 5% level. The Bayesian analysis supports strongly the existence of a nonlinear component dominated by y_{t-5} and marginally by y_{t-1}. From the accuracy measures, it seems that Hamilton's importance sampler is relative better compared to Geweke's.

[Table 3 about here]

The estimation results of Table 4 are similar to those of Table 3. There is support for the nonlinearity of the model as the estimate $\hat{\varsigma}$ is statistically significant. The linear estimates are almost identical to those of Table 3, and the value of the log-likelihood function is the same for both specifications of the covariance function of the random field. Even though the statistical significance of g_1, g_2, and g_5 has been reduced, it seems that y_{t-5} is the major force driving the nonlinear component. We should note that the magnitude of g_1, g_2, and g_5 is smaller than that of the estimates in Hamilton's specification. However, we find that there is a substantial overlap between the 95% confidence intervals for these parameters when we compare both specifications.

[Table 4 about here]

4.3 Dissecting the business cycle

We assess the contribution of the estimated flexible parametric regression model to the understanding of the US business cycle within the context provided by Harding and Pagan (2002). The basic insight is that a statistical model is performing well if it is able to replicate the characteristics of the actual business cycle. The description of the cycle follows the standards established by the NBER, which consist of locating the peaks and troughs (turning points) in the level of the GNP time series, together with a set of rules to determine the duration and amplitude of phases (from peak to through, and viceversa) and full cycles (from peak to peak, and from through to through). Some of the rules are that a full cycle must have a minimum duration of 15 months (5 quarters) and that a phase must last at least 6 months (2 quarters). The algorithm to implement the NBER standards was developed by Bry and Boschan (BB) (1971) for monthly data. Harding and Pagan (2002) have adapted the BB algorithm to quarterly data, which they named BBQ.

We implement the BBQ algorithm[4] for the actual quarterly GNP from 1947Q1 to 2000Q4 and for simulated data from the flexible parametric regression models estimated in Tables 3 and 4. For comparison purposes, we have also simulated data from a linear AR(2) and a linear AR(5) model for the GNP growth rate. The results are reported in Table 5. The BBQ algorithm examines four characteristics of the business cycle: the duration of the cycle and its phases, measured in quarters; the amplitude of the cycle and its phases, measured as percentage loss (gain) of GNP from the previous peak (trough); the cumulated losses (gains) in output from peak to trough (from trough to peak) relative to the previous peak (through) in percentage terms; and the excess cumulated movements, in percentage terms, that result from approximating the actual cumulative movements with a "triangle approximation"[5].

The full business cycle for the US GNP lasts on average about 21 quarters, with short lived contractions (from peak to trough) of about 3 quarters, and longer expansions (from trough to peak) of about 17 quarters. Both random field models and the linear AR(2) are able to replicate approximately the duration of the cycle; however, the linear AR(5), while replicating the duration of contractions, produces longer expansions, extending the duration of the cycle to 27 quarters. In general, the contraction phase of the cycle is more or less faithfully replicated by the four models considered in Table 5. The random field model with the Dahl-González-Rivera (D-GR) covariance is the best to reproduce the amplitude of the contraction. All models are able to reproduce the mean cumulation and the mean excess. The latter is practically zero, corroborating that the "triangle approximation" is a good one and that the average contraction phase is linear.

[Table 5 about here]

The expansion phase of the cycle is more difficult to replicate. In the actual

[4] We thank Don Harding and Adrian Pagan for making their BBQ algorithm available from Harding's web page.

[5] For a more detailed description of these four measures, see Harding and Pagan (2002), pp. 369-370.

data, the mean excess is 1.12% and all the four models fail to replicate it. The random field model with D-GR covariance produces the largest mean excess of 0.16% with a 75% quartile of 0.40. In the remaining three counts -duration, amplitude, and cumulation- the random field models have a clear advantage over the linear models. The random field model with D-GR covariance replicates the percentual cumulation very accurately, while the linear models -in particular, the AR(5)- produce too large a cumulation. Furthermore, the dispersion of the duration, amplitude and cumulation, measured by the interquartile range, is the smaller for the random field model with D-GR covariance.

Harding and Pagan (2002) found very little evidence for nonlinear models being able to replicate the characteristics of the business cycle. They considered two popular specifications: the two-state Markov regime switching of Hamilton (1989), and an extension of this model by Durland and McCurdy (1994) where the transition probabilities depend on the duration of the expansion or contraction. In Table 6, we compare Harding and Pagan results to those of the random field model with D-GR covariance.

[Table 6 about here]

The Hamilton and the Duration Dependence models produces longer contractions as well as longer average amplitudes and cumulations than those of the actual data. In the expansion phase, the average amplitude is also larger but the most striking difference comes from the excessive average cumulation produced by the Hamilton's model.

In summary, we find that a nonlinear model as the flexible parametric regression model with a D-GR covariance function is able to replicate very faithfully the characteristics of the US GNP business cycle, though the average excess in the expansion phase remains to be explained. We address this issue examining the conditional mean function estimated with the random field model.

In the actual data, the large average excess of the expansion phase indicates that the "triangle approximation" is not very accurate and that we should expect deviation from linearity going from a trough to a peak, in particular, the expansion phase must be concave. This shape is compatible with the claim of several authors for whom the business cycle consists of three regimes, i.e. Sichel (1994), Clements and Krolzig (1998). They differentiate a high growth recovery phase after a contraction from a moderate growth phase that immediately follows. Hence, in an expansion phase we should be able to detect faster than average growth rates after a trough and slower before a peak, implying the existence of an inflexion point somewhere in the phase. When we examine the conditional mean $E(y_t|y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$ estimated with the random field model, we find an accurate description of the nonlinearity in the expansion phase that is precisely compatible with the existence of three regimes in the business cycle. We proceed in the following fashion. Let us assume that the business cycle has a linear contraction shape and a concave expansion shape. We inquire about the implications of this shape in the analysis of the functional form of the growth rate, aiming to understand the nonlinearity features that we found in the statistical testing and estimation stages. In particular, we are intrigued by the role of the fifth lag of the growth rate in driving the nonlinearity of the conditional mean of the US GNP growth rate.

In Figure 1, we draw a stylized business cycle of the GNP level (in logs) with some of the characteristics of the US business cycle.

[Figure 1 about here]

The contraction phase is linear and the average duration from peak to trough is approximately three quarters, while the expansion phase is mostly concave and its duration is about seventeen quarters. We distinguish three stages of growth in the expansion phase: aggressive growth after a trough with growth rates larger than the average and convex shape; moderate growth with rates around the average and concave shape; and slow growth before the peak where the growth rate eventually becomes zero just before the start of the contraction phase. The question we ask is the following. If the average business cycle has the characteristics described in Figure 1, which functional form should we expect when we plot the conditional expected growth rate $E(y_t|y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$ against any of the components of the information set? Note that Figure 1 describes the dynamics of the GNP *level* and we inquire about the functional form of the *growth rate*, which is the first derivative of the curve in Figure 1. According to the results obtained in the statistical testing and estimation stages, we focus mainly on the contribution of the first lag (Figure 2) and fifth lag (Figure 3) to the functional form of the conditional mean of the growth rate.

[Figure 2 about here]

Figure 2 has two panels. The left panel is a stylized version of the functional form of the conditional mean $E(y_t|y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$ as a function of y_{t-1}. The right panel is the actual functional form estimated by the random field regression model with the D-GR covariance function for all the other variables in the conditioning set fixed at their sample means. Let us focus in the left panel first. In the plane marked by coordinates (y_{t-1}, y_t), two threshold values are important: zero growth rate (at the peak or at the trough of the cycle) and the unconditional mean of the growth rate. Since contractions are linear, we should expect a constant growth rate and positive first order autocorrelation such as negative growth rates are followed by negative growth rates in the next period. Hence, we have the functional form plotted in the south-west quadrant of the left panel. In the expansion phase, if we are in the aggressive part of the phase, just after the trough, faster growth rates than the average will be followed by larger growth rates in the next period due to the convexity of the cycle, placing the functional form in the upper north-east quadrant. When we enter in the moderate growth part of the expansion phase, the cycle is already concave and growth rates around the average μ will be followed by smaller (below or around μ) growth rates. Finally, on approaching the peak of the expansion, the one-period dynamics of the growth rate is constrained to the square delimited by the points $(0, 0)$ and (μ, μ). Inspection of the right panel reveals that the estimated conditional mean seems to agree with the description of the stylized conditional mean in the left. The unconditional mean of the growth rate is 0.86% per quarter and the standard deviation is 1.02%. The aggressive growth part of the expansion seems to have growth rates of 1.5% and above, approximately one standard deviation above the mean. It is fair to say that a linear model in y_{t-1} would have basically missed only the moderate growth part of the expansion, this may be a reason why the nonlinearity in y_{t-1} is not terribly strong. However, a very different picture emerges when we analyze the contribution of y_{t-5} to the nonlinearity of the conditional mean.

[Figure 3 about here]

In Figure 3, we analyze the functional form of the conditional mean $E(y_t|y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$ as a function of y_{t-5}. The organization of this figure is the same as that of Figure 2. If the average duration of the expansion phase is about seventeen quarters, five quarters represent approximately one third of the duration. Given that the contraction phase is about three quarters, the fifth order autocorrelation between growth rates in the contraction and those in the expansion must be negative. In the plane marked by coordinates (y_{t-5}, y_t), to the left of the vertical axis that crosses the point (0,0), we picture the negative correlation between the areas T and A of Figure 1. During the expansion phase, the fifth order autocorrelation will change sign. In the aggressive part of the phase, faster growth rates than the average will be followed five periods into the future by growth rates around or below μ that we observe in the moderate part of the phase, which explains the negative autocorrelation between the areas A and M of Figure 1. Finally, the correlation between moderate growth rates and slow growth rates (areas M and S of Figure 1) will be positive since both areas are in the concave part of the phase. Inspection of the right panel confirms the stylized description of the conditional mean in the left.

In conclusion, the flexible parametric regression model lends support to a nonlinear, asymmetric business cycle with three regimes: linear contraction, aggressive early expansion, and moderate/slow late expansion, as opposed to models with just two regimes, contractions and expansions.

5 Conclusions

Nonparametric and semiparametric models for the growth rate of the US GNP are surprisingly absent from the many specifications that have being proposed during the last decade. In this paper, we have analyzed nonlinearities in the quarterly growth rate of the US GNP within the framework of Hamilton (2001) flexible parametric regression. This model is specified in a nonparametric/semiparametric fashion because no functional form for the conditional mean is assumed but, nevertheless, the estimation is fully parametric. We have implemented a battery of new tests for neglected nonlinearity based on the theory of random fields proposed by Dahl and González-Rivera (2003) and we have found that both types of statistical inference, testing and estimation, are in agreement on identifying the source of nonlinearity, which comes from the fifth lag of the growth rate. Our model produces a business cycle that when is dissected with the BBQ algorithm mimics very faithfully the characteristics of the actual US business cycle. We have shown that the flexible parametric regression model is superior to linear and nonlinear parametric specifications. On understanding why the fifth lag of the growth rate is so relevant, we have found that the nonparametrically estimated conditional mean supports parametric specifications that claim that there are three phases in the business cycle: rapid linear contractions, aggressive short-lived convex early expansions, and moderate/slow relatively long concave late expansions.

References

Abadir K. and J. Magnus (2002), "Notation in econometrics: a proposal for a standard", *Econometrics Journal* 5, 76-90.

Bochner, S. (1959), *Lectures on Fourier Integrals*, Princeton University Press, New Jersey.

Burns, A.F. and W.C. Mitchell (1946), "Measuring business cycles", NBER, New York.

Bry, G. and C. Boschan (1971), "Cyclical analysis of time series: selected procedures and computer programs", NBER, New York.

Clements, M.P. and H-M. Krolzig (1998), "A comparison of the forecasting performance of Markov-switching and threshold autoregressive models of US GNP", *Econometrics Journal* 1, C47-C75.

Dahl, C.M. and G. González-Rivera (2003), "Testing for neglected nonlinearity in regression models based on the theory of random fields", *Journal of Econometrics 114(1), 141-164.*

Durland, J.M. and T.H. McCurdy (1994), "Duration-dependent transitions in a Markov model of US GNP growth", *Journal of Business and Economic Statistics* 12, 279-288.

Galvão, A.B.C. (2002), "Can non-linear time series models generate US business cycle asymmetric shape?", *Economics Letters* 77, 187-194.

Geweke, J. (1989), "Bayesian inference in econometrics using Monte Carlos integration", *Econometrica* 57, 1317-1339.

Hamilton, J.D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica* 57, 357-384.

Hamilton, J.D. (2001), "A parametric approach to flexible nonlinear inference", *Econometrica* 69, 537-573.

Harding, D. and A. Pagan (2002), "Dissecting the cycle : A methodological investigation", *Journal of Monetary Economics* 49, 365-381.

Potter, S. (1995), "A nonlinear approach to U.S. GNP", *Journal of Applied Econometrics* 10, 109-125.

Sichel, D.E. (1994), "Inventories and the three phases of the business cycle", *Journal of Business and Economic Statistics* 12, 269-277.

Stock, J.H. and M.W. Watson (1999), "A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series" in *Cointegration,* *Causality, and Forecasting*, Oxford University Press, New York.

Yaglom, A.M. (1962), *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, Englewood Cliffs, N.J.

Table 1: Selection of the optimal number of lags under the null hypothesis of linearity. Quarterly growth rates of real US GNP, 1947Q1 - 2000Q4. The preferred specification is indicated by *.

H_0	AIC	BIC
AR(1)	-0.056	-0.024*
AR(2)	-0.053	-0.005
AR(3)	-0.056	-0.008
AR(4)	-0.061	0.019
AR(5)	-0.060	0.036
$y_{t-1}, y_{t-2}, y_{t-5}$	-0.064*	-0.016

Table 2: Bootstrapped p-values of the tests for neglected nonlinearity for the quarterly growth rates of real US GNP, 1947Q1 - 2000Q4. 1000 bootstrap resamples. Rejection of linearity at the 10 % significance level is indicated by *.

		Test Statistics			
H_0	Aux. Regres	$\lambda_H^E(\boldsymbol{g})$	$\lambda_{OP}^E(\boldsymbol{g})$	λ_{OP}^A	g_{OP}
AR(1)	$y_{t-1},..,y_{t-4}$	0.558	0.494	0.514	0.429
AR(2)	$y_{t-1},..,y_{t-4}$	0.485	0.459	0.214	0.328
AR(3)	$y_{t-1},..,y_{t-4}$	0.498	0.455	0.513	0.276
AR(4)	$y_{t-1},..,y_{t-4}$	0.787	0.528	0.503	0.474
AR(5)	$y_{t-1},..,y_{t-4}$	0.774	0.475	0.543	0.414
$y_{t-1}, y_{t-2}, y_{t-5}$	$y_{t-1},..,y_{t-4}$	0.496	0.398	0.700	0.612
AR(1)	$y_{t-1},..,y_{t-5}$	0.883	0.607	0.186	0.273
AR(2)	$y_{t-1},..,y_{t-5}$	0.986	0.544	0.086*	0.134
AR(3)	$y_{t-1},..,y_{t-5}$	0.993	0.540	0.209	0.173
AR(4)	$y_{t-1},..,y_{t-5}$	0.617	0.602	0.289	0.320
AR(5)	$y_{t-1},..,y_{t-5}$	0.508	0.576	0.143	0.103
$y_{t-1}, y_{t-2}, y_{t-5}$	$y_{t-1},..,y_{t-5}$	0.699	0.560	0.145	0.490
AR(1)	$y_{t-1}, y_{t-2}, y_{t-5}$	0.181	0.433	0.120	0.178
AR(2)	$y_{t-1}, y_{t-2}, y_{t-5}$	0.319	0.539	0.082*	0.053*
AR(3)	$y_{t-1}, y_{t-2}, y_{t-5}$	0.233	0.375	0.113	0.085*
AR(4)	$y_{t-1}, y_{t-2}, y_{t-5}$	0.121	0.433	0.171	0.209
AR(5)	$y_{t-1}, y_{t-2}, y_{t-5}$	0.056*	0.341	0.056*	0.301
$y_{t-1}, y_{t-2}, y_{t-5}$	$y_{t-1}, y_{t-2}, y_{t-5}$	0.099*	0.437	0.064*	0.623

Table 3: Maximum likelihood based estimates and Bayesian estimates of the nonlinear US-GNP model using Hamilton's Spherical Covariance Function and the mixed split-t based Importance Sampling Scheme of Geweke (1989) and mixed t based Importance Sampling Scheme of Hamilton (2001).

	Maxim. Likelihood		Bayesian Estimation			Bayesian Estimation		
			Geweke's Split-t mixture			Hamilton's t-mixture		
	Estimate	Std-err.	Estimate	Std-err.	RNE	Estimate	Std-err.	RNE
β_0	0.903	0.189	0.920	0.221	0.129	0.930	0.207	0.216
β_1	0.303	0.088	0.299	0.094	0.136	0.296	0.092	0.221
β_2	0.088	0.091	0.082	0.101	0.127	0.081	0.098	0.226
β_3	-0.081	0.071	-0.089	0.071	0.161	-0.085	0.072	0.246
β_4	-0.101	0.070	-0.091	0.071	0.156	-0.097	0.071	0.241
β_5	-0.143	0.088	-0.143	0.097	0.131	-0.147	0.096	0.220
g_1	0.568	0.242	0.607	0.360	0.358	0.566	0.253	0.461
g_2	0.652	0.301	0.838	0.637	0.329	0.727	0.469	0.565
g_5	1.058	0.316	1.115	0.597	0.241	0.997	0.418	0.438
ς	0.558	0.200	0.710	0.286	0.312	0.641	0.177	0.401
σ	0.846	0.061	0.819	0.079	0.223	0.838	0.058	0.302
logL	283.25		.			.		
$E\{w(\theta)\}$	.		0.106			0.242		
ω_1	.		177.43			60.67		
ω_{10}	.		144.01			45.23		

Notes:

RNE :	relative numerical efficiency of the importance sampling density, see Geweke (1989), pp. 1322
ω_1, ω_{10}	order statistics pertaining to the importance sampling density; the subscripts refer to the order, see Geweke (1989), pp. 1331
$w(\theta)$:	the importance sampling density as a function of the parameter vector θ.

Table 4: Maximum likelihood based estimates and Bayesian estimates of the nonlinear US-GNP model using Dahl - Gonzalez-Rivera Covariance Function and the mixed split-t based Importance Sampling Scheme of Geweke (1989) and mixed t based Importance Sampling Scheme of Hamilton (2001).

	Maxim. Likelihood		Bayesian Estimation			Bayesian Estimation		
			Geweke's Split-t mixture			Hamilton's t-mixture		
	Estimate	Std-err.	Estimate	Std-err.	RNE	Estimate	Std-err.	RNE
β_0	0.911	0.204	0.877	0.186	0.190	0.882	0.184	0.245
β_1	0.293	0.085	0.293	0.089	0.171	0.294	0.085	0.230
β_2	0.091	0.092	0.095	0.092	0.179	0.094	0.091	0.232
β_3	-0.101	0.075	-0.124	0.074	0.189	-0.106	0.071	0.245
β_4	-0.078	0.072	-0.055	0.074	0.183	-0.072	0.069	0.242
β_5	-0.140	0.089	-0.135	0.089	0.184	-0.136	0.088	0.237
g_1	0.095	0.093	0.197	0.201	0.474	0.167	0.182	0.524
g_2	0.250	0.278	0.389	0.276	0.340	0.362	0.300	0.397
g_5	0.354	0.281	0.445	0.304	0.298	0.408	0.299	0.407
ς	0.714	0.303	1.315	0.684	0.209	0.947	0.516	0.197
σ	0.783	0.102	0.643	0.151	0.206	0.734	0.114	0.196
logL	282.87		.			.		
$E\{w(\theta)\}$			0.273			0.269		
ω_1	.		56.91			84.00		
ω_{10}	.		46.53			56.47		
Notes:								
RNE :	relative numerical efficiency of the importance sampling density, see Geweke (1989) pp. 1322							
ω_1, ω_{10}	order statistics pertaining to the importance sampling density; the subscripts refer to the order, see Geweke (1989) pp. 1331							
$w(\theta)$:	the importance sampling density as a function of the parameter vector θ.							

Table 5: Business cycle characteristics. Quarterly real US GNP, 1947Q1-2000Q4. Flexible parametric regresssion models versus linear models.

	Data	Random Field	Random Field	AR(2)	AR(5)
	US GNP	Spherical cov.	D-GR covar.		
Contractions					
(peak to trough)					
Mean Duration	3.22	3.33	3.40	3.32	3.28
(in quarters)		3.00, 3.67	3.07, 3.67	2.86, 3.71	2.80, 3.67
Mean Amplitude (%)	-2.32	-1.79	-2.17	-1.87	-1.82
		-2.03, -1.57	-2.44, -1.92	-2.22, -1.54	-2.15, -1.51
Mean Cumulation (%)	-3.53	-3.13	-3.89	-3.98	-3.68
		-3.73, -2.41	-4.82, -2.98	-4.69, -2.50	-4.32, -2.36
Mean Excess(%)	0.01	0.03	0.05	0.00	-0.01
		-0.02, 0.08	-0.02, 0.10	-0.06, 0.06	-0.07, 0.05
Expansions					
(trough to peak)					
Mean Duration	17.50	19.51	16.44	20.17	23.57
(in quarters)		15.62, 21.86	12.92, 18.89	15.90, 23.43	17.87, 28.20
Mean Amplitude (%)	20.60	19.67	18.32	21.70	25.51
		15.22, 22.46	14.17, 21.21	16.47, 25.63	17.87, 30.94
Mean Cumulation (%)	259.82	318.93	251.70	373.59	537.86
		185.8, 385.0	140.1, 296.7	190.7, 466.9	250.1, 689.5
Mean Excess (%)	1.12	0.05	0.16	0.04	-0.04
		-0.23, 0.36	-0.08, 0.40	-0.31, 0.35	-0.34, 0.27
Note: The numbers below the mean are the 25% and 75% quartiles.					

Table 6: Business cycle characteristics. Quarterly real US GNP, 1947Q1-2000Q4. Flexible parametric regresssion models versus nonlinear models.

	Data	Random Field	Hamilton*	Dur Dep*
	US GNP	D-GR covar.		
Contractions				
(peak to trough)				
Mean Duration	3.22	3.40	4.4	4.8
Mean Amplitude (%)	-2.32	-2.17	-2.8	-3.3
Mean Cumulation (%)	-3.53	-3.89	-8.2	-8.5
Mean Excess(%)	0.01	0.05	0.0	0.0
Expansions				
(trough to peak)				
Mean Duration	17.50	16.44	20.0	16.9
Mean Amplitude (%)	20.60	18.32	27.3	25.0
Mean Cumulation (%)	259.82	251.70	496	293
Mean Excess (%)	1.12	0.16	-0.0	0.0
* These two columns are taken from Harding and Pagan (2002), Table 5. Their period of study is 1947Q1-1997Q1.				

Figure 1.
Stylized business cycle

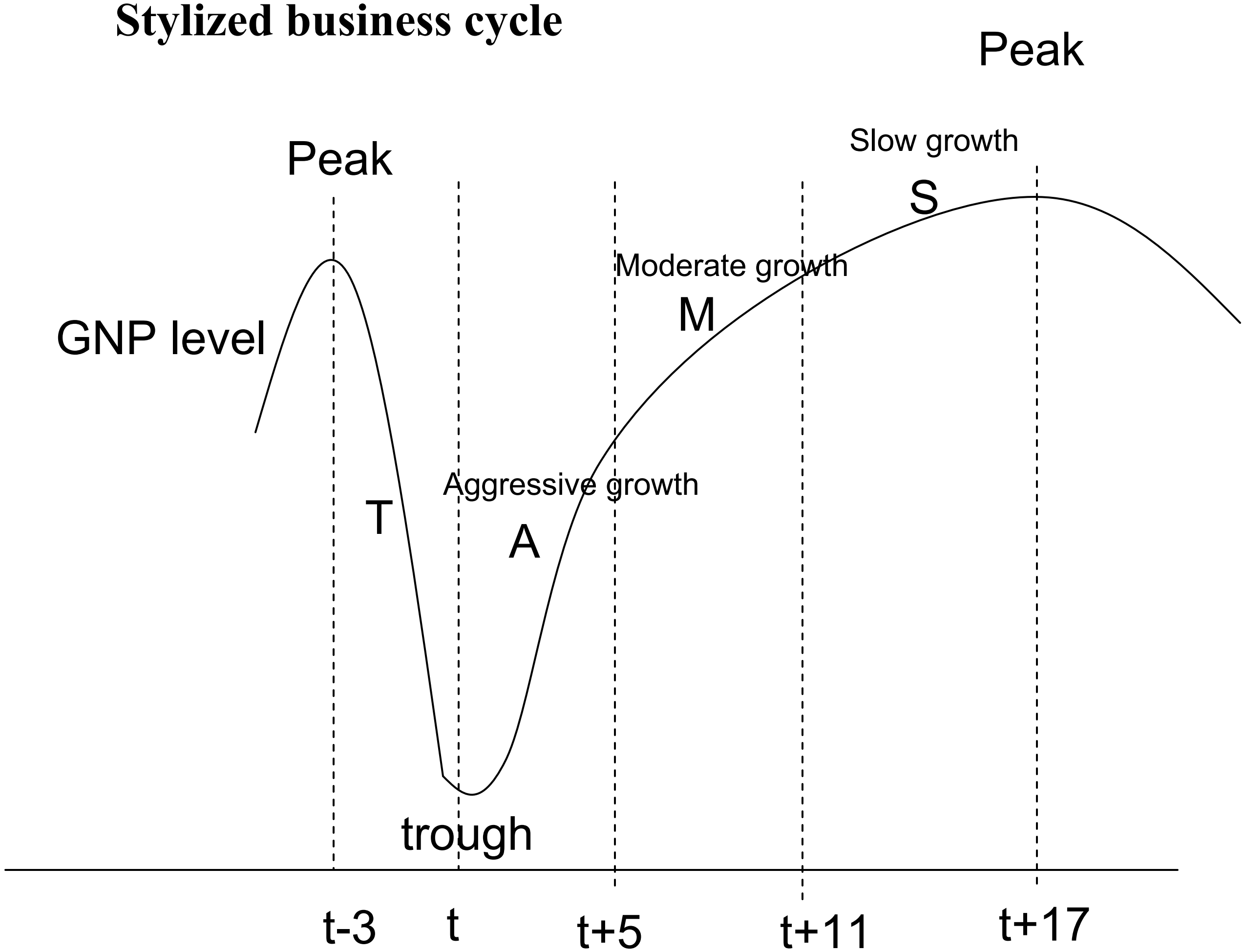
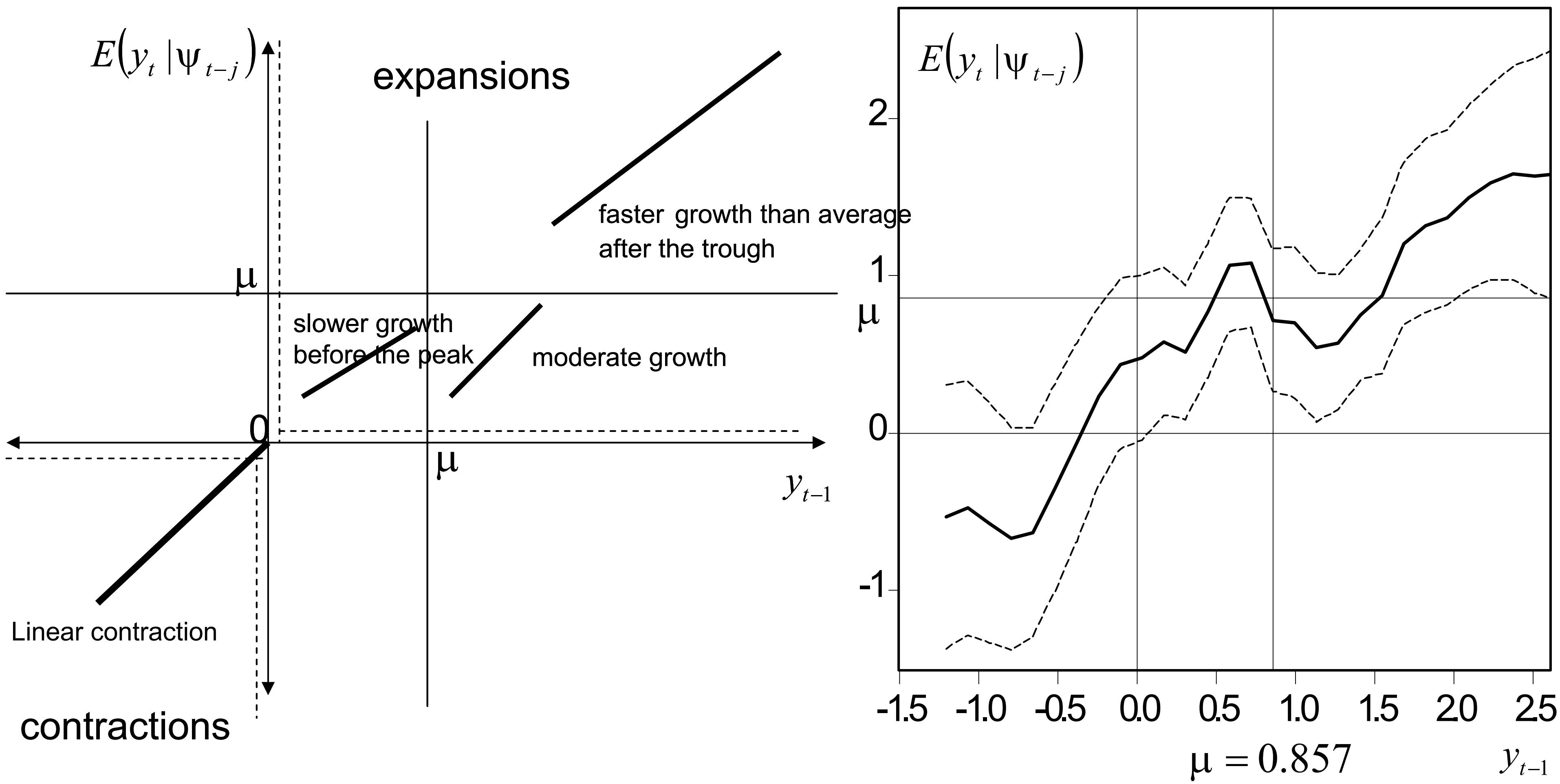


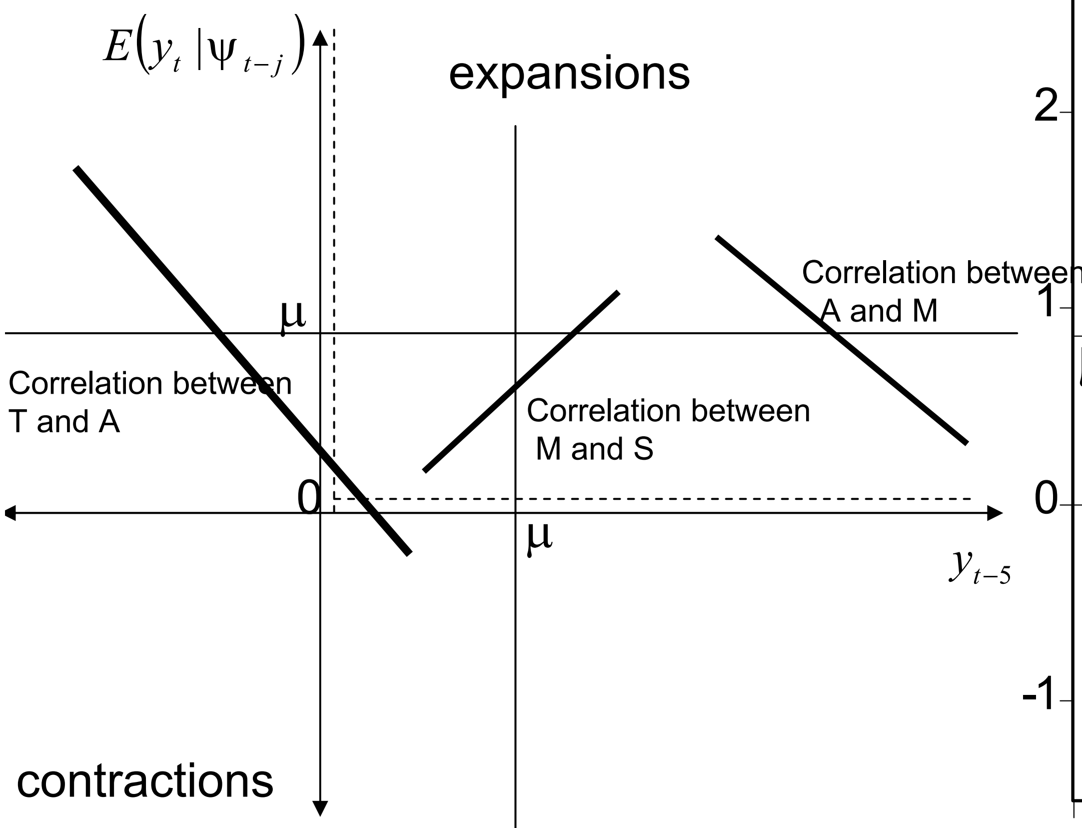
Figure 2.
Understanding nonlinearity due to the first lag of the growth rate

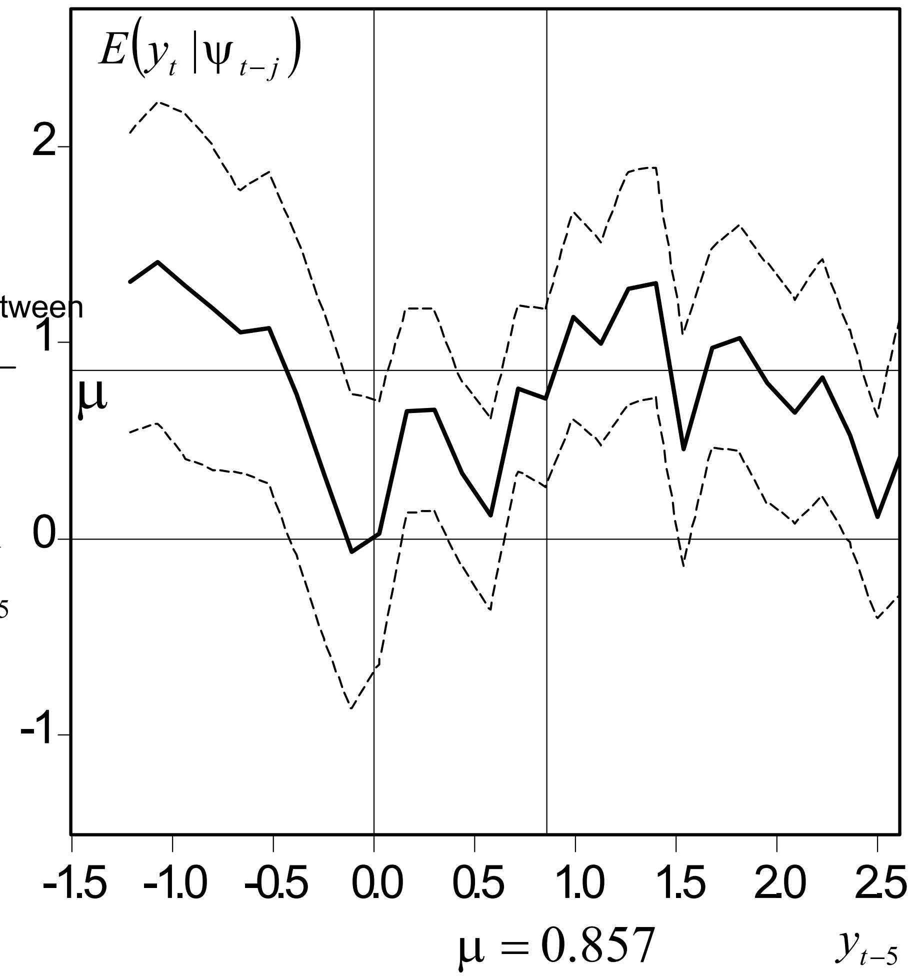


Stylized conditional mean corresponding to the stylized business cycle of Figure 1

Conditional mean and 95% confidence bands estimated with a random field regression with the Dahl & Gonzalez-Rivera covariance

Figure 3.
Understanding nonlinearity due to the fifth lag of the growth rate





Stylized conditional mean corresponding to the stylized business cycle of Figure 1

Conditional mean and 95% confidence bands estimated with a random field regression with the Dahl & Gonzalez-Rivera covariance