Professorship for statistics and empirical research
Institute of statistics and econometrics
Faculty of Business, Economics and Social Sciences
Kiel University

Predictive regressions and IVX-methodology

Seminar paper statistics
Submission Date: September 24, 2021

Author: Peer Lasse Hinrichsen
Subject: Statistics
Semester: 4
Module: Seminar statistic
Stu-Email-Adress: stu201157@mail.uni-kiel.de
Matriculation number: 1104322

Assessor: Prof.Matei Demetrescu

1 Introduction

In economic timeseries-data often AR-processes with a near-unit root instead of a unit root need to be assumed. It is undoubtedly intuitive that in the real world effects of shocks do not last into the infinite future but rather decay over a very long time span. However when using variables with such persistent behavior in the framework of estimation and inference, many problems arise. This is due to the standard non-stationary methodology relying on the assumption of a unit root. Leading in turn to nonstandard asymptotic distributions for inference conducted on these persistent variables. A solution proposed to this problem is the concept of IVX-estimation. An instrument for the persistent regressor is constructed from the regressor itself without relying on external information. The article *Regime-Specific Predictability in Predictive Regressions* by *Gonzales, Pitarakis (2012)* developed some new limit theory for threshold models with persistent regressors. This new theory was then directed at predicting excess stock returns in a nonlinear way using dividend yields as a naturally persistent regressor.

While the theorems derived are proofed rigorously, this seminar paper aims to clarify the context with respect to the literature of near integrated process. To do so the proofs are broken down and in conjunction with the theory for persistent variables, explained. Of special interest are the origins of the stochastic integrals such as the Ornstein-Uhlenbeck processes which play a central role in the limiting theory proposed. To do so the limiting results for sample moments of persistent regressors as in *Phillips (1987,1988)* will be discussed.

The plan for this seminar paper is as follows. Section 2 briefly revises the general model setup and the leading assumption used in the original paper by *Gonzales, Pitarakis (2012)*. Section 3 then goes through the details of the proof for proposition 1 and subsequently proposition 2. Section 4 introduces the concepts of IVX estimation and afterwards uses those to revise the proof for proposition 3. Section 5 contains some additional simulations complementary to those in the original paper. Besides different covariance structures, simulations for varying parameter are presented and compared to the results obtained by *Gonzales, Pitarakis (2012)*. Section 6 concludes. Some tables referenced in section 5 are delegated to the appendix.

2 General Setup

2.1 The Model

The general setup will be the following threshold specification.

$$y_{t+1} = \begin{cases} \alpha_1 + \beta_1 x_t + u_{t+1} & q_t \leq \gamma \\ \alpha_2 + \beta_2 x_t + u_{t+1} & q_t > \gamma \end{cases}$$

with:

$$\begin{aligned} x_t &= \rho_T x_{t-1} + \nu_t, \\ \rho_T &= 1 - \frac{c}{T} && c > 0 \\ q_t &= \mu_q + u_{q_t} \end{aligned}$$

u_t, u_{qt} and ν_t are stationary random disturbances. The threshold variable q_t will be governing the regime switches. The threshold parameter γ is assumed to be unknown with $\gamma \in \Gamma = [\gamma_1, \gamma_2]$ where γ_1, γ_2 are selected such that not all but at least some values of the threshold variable q_t are below γ. Further defining

$$\begin{aligned} I_{1t} &\equiv I(q_t \leq \gamma) \\ I_{2t} &\equiv I(q_t > \gamma) \end{aligned}$$

and making use of the probability integral transform, we obtain

$$I(q_t \leq \gamma) = I(F(q_t) \leq F(\gamma)) = I(U_t \leq \lambda)$$

where $F(.)$ is the marginal distribution of q_t and U_t is a standard uniformly distributed random variable.

Now using these indicator functions the model can be expressed in matrix notation, which will be useful for the discussion on limiting behavior. Therefore let :

- y denote the vector stacking y_{t+1}
- X_i denote the matrix stacking $(I_{it}\ x_t I_{it})$
- $Z = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$, $\theta = \begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix}$
- $\theta_i = (\alpha_i, \beta_i)'$, $i = 1, 2$

such that the model can be expressed as :

$$y = Z\theta + u$$

2.2 Leading Assumptions

The innovations of the persistent variable will follow a general linear process.

$$\nu_t = \Psi(L)e_t \quad \text{with} \quad \Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$$

Where the lag polynomial $\Psi(L)$ satisfies $\Psi(1) \neq 0$ and its coefficients are j-summable $\sum_{j=0}^{\infty} j|\psi_j| < \infty$.

The shocks to the dependent variable y_t namely u_t will be of martingale difference type and thus its expectation with respect to its own past is zero. This way we can impose relatively weak assumptions on the processes while still being able to justify a series of CLT.

Letting $\tilde{\omega}_t = (u_t, e_t)'$ and $F_t^{\tilde{\omega}q} = \{\tilde{\omega}_s, u_{qs}|s \leq t\}$ the filtration generated by $(\tilde{\omega}_t, u_{qt})$. We can state our Assumptions more formally.

$$E[\tilde{\omega}_t|F_{t-1}^{\tilde{\omega}q}] = 0$$
$$E[\tilde{\omega}_t\tilde{\omega}_t'|F_{t-1}^{\tilde{\omega}q}] = \tilde{\Sigma} > 0$$
$$sup_t E\tilde{\omega}_{it}^4 < \infty$$

We further assume that the threshold variable $q_t = u_q + u_{qt}$ has a continuous and strictly increasing distribution $F(.)$ and is such that u_{q_t} is a strictly stationary, ergodic and strong mixing sequence with mixing numbers α_m satisfying $\sum_{m=1}^{\infty} \alpha^{\frac{1}{m}-\frac{1}{r}} < \infty$ for some $r > 2$. I.e The process for the threshold variable exhibits asymptotic independence.

3 Limiting Distributions

Gonzales, Pitarakis (2012) discussed two hypotheses. First the null of linearity $H_0^A : \alpha_1 = \alpha 2$, $\beta_1 = \beta_2$. And second the hypothesis testing linearity and predictability jointly H_0^B: $\alpha_1 = \alpha 2$, $\beta_1 = \beta_2 = 0$. It is important to note that the threshold value γ is unidentified. Therefore the focus will be on the sup-Wald formulation from here onward. Also note that for the sup-Wald formulation ten percent trimming is assumed at both ends of the samples to ensure that enough data is in each regime.

3.1 Limiting Distributions H_0^A

The Wald statistic for testing H_0^A is given by

$$W_T^A(\lambda) = \frac{\hat{\theta}' R_A' (R_A (Z'Z)^{-1} R_A')^{-1} R_A \hat{\theta}}{\hat{\sigma}_u^2}$$

Where $X = X1 + X2$ and $\hat{\theta} = (Z'Z)^{-1} Z'y$ is of OLS type. The restriction matrix for H_0^A is $R_A = [I - I]$ where I is as 2×2 identity matrix. The residual variance is given by

$$\hat{\sigma}_u^2 = \frac{y'y - \sum_{i=1}^2 y'X_i \left(X_i'X_i\right)^{-1} X_i'y}{T}$$

Using this notation the Wald statistic can be reformulated conveniently:

$$W_T^A(\lambda) = \left[u'X_1 - u'X\left(X'X\right)^{-1} X_1'X_1\right] \times \left[X_1'X_1 - X_1'X_1\left(X'X\right)^{-1} X_1'X_1\right]^{-1} \times \left[X_1'u - \left(X_1'X_1\right)\left(X'X\right)^{-1} X'u\right] / \hat{\sigma}_u^2$$

3.1.1 Proof of Proposition 1

Now scaling the individual elements of this reformulated Wald statistic with $D_T = diag(\sqrt{T}, T)$, convergence results for moments of nearly integrated processes can be applied. These are readily available in the literature.
Caner,Hansen (2001); Hansen (1996); Phillips (1987,1988,1998); Phillips, (2007); White (2001).

Starting with

$$D_T^{-1} X_1'X_1 D_T^{-1} = \begin{pmatrix} \frac{\sum I_{1t}}{T} & \frac{\sum x_t I_{1t}}{T^{\frac{3}{2}}} \\ \frac{\sum x_t I_{1t}}{T^{\frac{3}{2}}} & \frac{\sum x_t^2 I_{1t}}{T^2} \end{pmatrix}$$

it can be seen that $\frac{\sum I_{1t}}{T} \xrightarrow{p} \lambda$. This follows from a LLN which is proposed by White (2001). Recall the previously defined indicator function.

$$I_{1t} = I\left(q_t \leq \gamma\right) = I\left(F\left(q_t\right) \leq F\left(\gamma\right)\right) \equiv I\left(U_t \leq \lambda\right)$$

using the LLN it is evident that

$$\frac{1}{T} \sum I\left(q_t \leq \gamma\right) \xrightarrow{p} E\left[I\left(q_t \leq \gamma\right)\right] = F\left(\gamma\right) = \lambda$$

In order to comprehend the limiting result for the second entry, a discussion about the convergence results for sample-moments of persistent variables is required. For this matter concentrate on the notation for the AR-coefficient of *Phillips (1987)* which slightly differs but still models AR-processes with near unit roots.

$$x_t = ax_{t-1} + \nu_t$$
$$a = exp(\frac{c}{T}) \qquad , (-\infty < c < \infty)$$

Note that the limiting results will carry over to the model used in *Gonzales, Pitarakis (2012)*, since the behavior of the AR coefficient is, albeit it being formulated slightly different, similar. To obtain the result first construct an auxiliary process from the sequence of partial sums of the error term.

$$S_T(r) = T^{-\frac{1}{2}}\sigma^{-1}S_{[Tr]} = T^{-\frac{1}{2}}\sigma^{-1}S_{j-1}$$

with

$$\frac{(j-1)}{T} \leq r < \frac{j}{T} \qquad , (j = 1, \ldots, T)$$

and

$$\sigma^2 = E\left(u_1^2\right) + 2\sum E\left(u_1 u_k\right) \qquad , k = 2, \ldots, \infty.$$

Here $S_t = \nu_1 + \ldots + \nu_t$ denotes the partial sum where ν_t are the innovations for the persistent variable. One can observe that since T and r uniquely define j that $S_T(r)$ is a random walk and thus a functional CLT can be applied. It is then recognizable that $S_T(r) \to W(r)$ where $W(r)$ is a standard Brownian motion. Proceeding with the MA-representation of the possibly persistent process x_t

$$x_t = \sum_{j=1}^{t} e^{(t-j)\frac{c}{T}}\nu_j + e^{\frac{tc}{T}}x_0$$

reformulate using $S_T(r)$, to obtain

$$\begin{aligned}
T^{-\frac{1}{2}}x_{[Tr]} &= \sigma \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \int_{\frac{(j-1)}{T}}^{\frac{j}{T}} dS_T(s) + O_p\left(T^{-\frac{1}{2}}\right) \\
&= \sigma \sum_{j=1}^{[Tr]} \int_{\frac{(j-1)}{T}}^{\frac{j}{T}} e^{(r-s)c}\, dS_T(s) + O_p\left(T^{-\frac{1}{2}}\right) \\
&= \sigma \int_0^r e^{(r-s)c}\, dS_T(s) + O_p\left(T^{-\frac{1}{2}}\right)
\end{aligned}$$

Using integration by parts onto the first term.

$$\sigma \left\{ X_T(r) + c \int_0^r e^{(r-s)c} S_T(s)\, ds \right\} + O_p\left(T^{\frac{1}{2}}\right)$$

It then follows by applying the functional CLT and the CMT that for $T \to \infty$ the following convergence result holds:

$$T^{-\frac{1}{2}} x_{[Tr]} \to \sigma W(r) + c \int_0^r e^{(r-s)c} W(s)\ ds = \sigma K_c(r)$$

$K_c(r) = \int_0^r e^{(r-s)c}\, dW(s)$ is a so called Ornstein-Uhlenbeck process which can be recognized as the solution to the stochastic differential equation $dK_c(r) = cK_c(r)dr + dW(r)$ and describes a Gaussian process with mean zero and variance $\frac{1}{2}(\frac{e^{2rc}-1}{c})$ where c is the mean-reversion parameter.

Now reformulate the sample moment $\frac{\sum x_t^2 I_{1t}}{T^2}$:
Letting $X_{T,t} = \frac{x_t}{\sqrt{T}}$ and $X_T(r) = \frac{x_{[Tr]}}{\sqrt{T}}$

$$\frac{\sum x_t^2 I_{1t}}{T^2} = \frac{1}{T}\sum X_{T,t}^2 I_{1t} = \lambda \frac{1}{T}\sum X_{T,t}^2 + \frac{1}{T}\sum X_{T,t}^2 (I_{1t} - \lambda)$$

By letting $a = \frac{p-2}{2p}$ and requiring $E\,|e_t|^p < \infty, p \geq 4$ the strong Approximation result derived by *Phillips (1988)* and discussed above holds, such that $sup_{r\in[0,1]}\,|X_T(r) - K_c(r)| = o_p(T^{-a})$. It then follows by the proof of *lemma1 f) A.3) Gonzales, Pitarakis (2012)* that

$$|\int_0^1 X_T(r)^2\, dr - \int_0^1 K_c(r)^2 dr| = o_p(T^{-a})$$

Applying this to the reformulated sample moment one obtains

$$\begin{aligned}&\frac{1}{T}\sum X_{T,t}^2 I_{1t} - \lambda \int_0^1 K_c^2(r)\, dr \\ &\quad = \frac{1}{T}\sum X_{T,t}^2 (I_{1t} - \lambda) + o_p\left(T^{-a}\right) \qquad ,\forall\lambda\end{aligned}$$

Since $sup_{r\in[0,1]}|X_T(r)|$ is bounded in probability and $\frac{\sum I_{1t}}{T} \xrightarrow{p} \lambda$ one can see that

$$\begin{aligned}&sup_\lambda |\frac{1}{T}\sum X_{T,t}^2 I_{1t} - \lambda \int_0^1 K_c^2(r)\, dr| = o_p(1) \\ &\Rightarrow \frac{1}{T}\sum X_{T,t}^2 = \int_0^1 K_c^2(r)\, dr + o_p\left(T^{-a}\right)\end{aligned}$$

The asymptotics for $\frac{\sum x_t I_{1t}}{T^{\frac{3}{2}}}$ are entirely similar. To summarize the convergence results, by letting $\overline{K_c}(r) = (1, K_c(r))$, write

$$D_T^{-1} X_1' X_1 D_T^{-1} \to \begin{pmatrix} \lambda & \lambda \int_0^1 K_c(r)\, dr \\ \lambda \int_0^1 K_c(r)\, dr & \lambda \int_0^1 K_c^2(r)\, dr \end{pmatrix}$$
$$\equiv \lambda \int_0^1 \overline{K_c}(r)\, \overline{K_c}(r)' \tag{1}$$

And analogously

$$D_T^{-1} X' X D_T^{-1} \to \int_0^1 \overline{K_c}(r)\, \overline{K_c}(r)' \tag{2}$$

Piecing these two results together using the CMT, the middle term of the reformulated Wald statistic becomes

$$\left[D_T^{-1} X_1' X_1 D_T^{-1} - D_T^{-1} X_1' X_1 \left(X' X \right)^{-1} X_1' X_1 D_T^{-1} \right]^{-1}$$
$$\Rightarrow \frac{1}{\lambda (1 - \lambda)} \left(\int_0^1 \overline{K_c}(r)\, \overline{K_c}(r)' \right)^{-1} \tag{3}$$

What's left is to discuss the behavior of

$$D_T^{-1} X_1' u = \begin{pmatrix} \frac{\sum I_{1t} u_{t+1}}{\sqrt{T}} \\ \frac{\sum x_t I_{1t} u_{t+1}}{T} \end{pmatrix}$$

and

$$D_T^{-1} X' u = \begin{pmatrix} \frac{\sum u_{t+1}}{\sqrt{T}} \\ \frac{\sum x_t u_{t+1}}{T} \end{pmatrix}$$

Notice that the first entry matches, under the assumptions mentioned before, exactly the term for the two-parameter generalized functional CLT proposed by *Caner,Hansen (2001)*. Such that the immediate convergence result follows.

$$\frac{\sum I_{1t} u_{t+1}}{\sqrt{T}} \to B_u(r, \lambda)$$

Where $B_u(r, \lambda)$ is a two parameter Brownian motion. For the second entry consider theorem 2 by *Caner,Hansen (2001)*. By combining the convergence result from above together with the fact that $T^{-\frac{1}{2}}x_{[Tr]} \to K_c(r)$ we obtain that under the given assumptions

$$\frac{1}{T}\sum x_t I_t u_{t+1} \to \int_0^1 K_c(r)\, dB_u(r,\lambda)$$

Note that for this stochastic integral integration is with respect to the first argument of $B_u(r,\lambda)$. The asymptotics for $D_T^{-1}X'u$ are then again entirely similar, such that in summary

$$D_T^{-1}X_1'u \ \to \begin{pmatrix} B_u(r,\lambda) \\ \int_0^1 K_c(r)\, dB_u(r,\lambda) \end{pmatrix} \tag{4}$$

$$D_T^{-1}X'u \ \to \begin{pmatrix} B_u(r,1) \\ \int_0^1 K_c(r)\, dB_u(r,1) \end{pmatrix} \tag{5}$$

Using again $\overline{K_c}(r) = (1, K_c(r))$ and also the so called Kiefer-process $G_u(r,\lambda) = B_u(r,\lambda) - \lambda B_u(r,1)$ the components can be pieced together using the CMT, to obtain

$$D_T^{-1}X_1'u - \lambda D_T^{-1}X'u \to \int_0^1 \overline{K_c}(r)\ dG_u(r,\lambda) \tag{6}$$

Combining (1) -(6) the limiting distribution of the Wald statistic follows.

proposition 1

$$\begin{aligned} sup_\lambda W_T^A(\lambda) \ \to & sup_\lambda \frac{1}{\lambda(1-\lambda)\sigma_u^2}\left[\int_0^1 \overline{K_c}(r)\ dG_u(r,\lambda)\right]' \\ & \times \left[\int_0^1 \overline{K_c}(r)\ \overline{K_c}(r)'\right]^{-1}\left[\int_0^1 \overline{K_c}(r)\ dG_u(r,\lambda)\right] \end{aligned}$$

While this distribution does not look very promising, it can be simplified substantially.

3.1.2 Simplification to Brownian bridge-type

In fact independence between $K_c(r)$ and $G_u(r,\lambda)$ can be verified. For this matter consider the joint expectation.

$$\begin{aligned} & E\left[G_u(r_1,\lambda_1)K_c(r_2)\right] \\ =& E\left[(B_u(r_1,\lambda_1) - \lambda_1 B_u(r_1,1))\left(B_\nu(r_2) + c\int_0^{r_2} e^{(r_2-s)c}B_\nu(s)\,ds\right)\right] \end{aligned}$$

multiplying out the Brackets yields

$$
\begin{aligned}
=&E\left[B_u\left(r_1,\lambda_1\right)B_\nu\left(r_2\right)\right]-\lambda_1 E\left[B_u\left(r_{1,}1\right)B_\nu\left(r_2\right)\right]\\
&+c\int_0^{r_2}e^{(r_2-s)c}E\left[B_u\left(r_1,\lambda_1\right)B_\nu\left(s\right)\right]ds\\
&-\lambda_1 c\int_0^{r_2}e^{(r_2-s)c}E\left[B_u\left(r_{1,}1\right)B_\nu\left(s\right)\right]ds
\end{aligned}
$$

Due to Brownian motions being zero mean Gaussian, it follows that

$$E\left[B_u\left(r_1,\lambda_1\right)B_\nu\left(r_2\right)\right]=\lambda_1 E\left[B_u\left(r_{1,}1\right)B_\nu\left(r_2\right)\right]$$

and thus the joint expectation above becomes 0. This proves Independence since the elements of the joint expectation also have zero expectation.

Now since the Ornstein-Uhlenbeck process $K_c\left(r\right)$ is Gaussian and independent of $G_u\left(r,\lambda\right)$ it can be used together with the covariance function of $G_u\left(r,\lambda\right)$.

$$E\left[G_u\left(r_1,\lambda_1\right)G_u\left(r_2,\lambda_2\right)\right]=\sigma_u^2\left(r_1\wedge r_2\right)\left(\left(\lambda_1\wedge\lambda_2\right)-\lambda_1\lambda_2\right)$$

to obtain the result for the conditional distribution.

$$\int K_c\left(r\right)\,dG_u\left(r,\lambda\right)\equiv N\left(0,\sigma_u^2\lambda\left(1-\lambda\right)\int K_c^2\left(r\right)\right)$$

However normalizing this expression by $\sigma_u^2\int K_c^2(r)$ the distribution no longer depends on a realization on $K_c^2(r)$ and is thus the unconditional distribution. It turns out that this distribution is a quadratic form of a Brownian Bridge process, which is again zero mean Gaussian and with covariance $\lambda_1\wedge\lambda_2-\lambda_1\wedge\lambda_2$, where $\wedge$ denotes the minimum operator. This distribution can be directly applied to the result in *proposition 1*, since

$$E\left[\overline{K_c}\left(r_2\right)G_u\left(r_1,\lambda\right)\right]'=E\left[G_u\left(r_1,\lambda_1\right)K_c\left(r_2\right)G_u\left(r_1,\lambda_1\right)\right]'=\left[0\ 0\right]'$$

The limiting distribution of the Wald statistic is then given by

$$sup_\lambda W_T^A\left(\lambda\right)\ \rightarrow sup_\lambda\frac{BB\left(\lambda\right)'BB\left(\lambda\right)}{\lambda\left(1-\lambda\right)}$$

where $BB(\lambda)$ denotes a standard Brownian bridge. The task to develop a nuisance parameter free limiting distribution for testing linearity in a threshold specification with possibly persistent regressors is therefore accomplished.

3.2 Limiting Distributions H_0^B

Starting from the initial setup, the model can be written as

$$y = \alpha_1 I_1 + \beta_1 x_1 + \alpha_2 I_2 + \beta_2 x_2 + u$$

By letting x_i stack $x_t I_{1t}$ and I_i stacking $I_i t$ it is clear that the two model specifications are equivalent. Also recall that X_i stacks $(I_{it}\ x_t I_{it})$ and thus $X_i = (I_i\ x_i)\ ,\ i = 1, 2.$

Re-parameterizing $\alpha = \alpha_1, \beta = \beta_2$ and $\eta = (\gamma, \delta)'$ with $\gamma = \alpha_2 - \alpha_1,\ \delta = \beta_2 - \beta_1$ leads to the model specification

$$y = \alpha + \beta x + X_2 \eta + u$$

such that the null hypotheses can be stated as

$$\begin{aligned} H_0^A &: \eta = 0 \\ H_0^B &: \eta = 0, \beta = 0 \end{aligned}$$

This convenient reformulation allows to segregate the parts of the Wald statistic, which asymptotics are standard from those which are not.

3.2.1 Proof of Proposition 2

The proof from here on, is straight forward and can be looked up in the appendix of the original paper *Gonzales, Pitarakis (2012)*. The result is that through this reformulation, the Wald statistic for testing H_0^B can be expressed as a linear combination of $W_T^A(\lambda)$ and an auxiliary Wald statistic $W_T(\beta = 0)$, such that for large samples

$$\begin{aligned} W_T^B(\lambda) &\approx W_T(\beta = 0) + W_T^A(\lambda) \\ sup_\lambda W_T^B(\lambda) &\approx W_T(\beta = 0) + sup_\lambda W_T^A(\lambda) \end{aligned}$$

The limiting behavior of $W_T(\beta = 0)$ can be established in a similar fashion as for $W_T^A(\lambda)$. Reformulate the Wald statistic and scale, such that the convergence results for sample moments of persistent regressors can be applied.

proposition 2

$$\begin{aligned} sup_\lambda W_T^B(\lambda) \rightarrow & \frac{\left[\int K_c^*(r)\, dB_u(r,1)\right]^2}{\sigma_u^2 \int K_c^*(r)^2} + sup_\lambda \frac{1}{\lambda(1-\lambda)\sigma_u^2} \\ & \times \left[\int \overline{K}_c^*(r), dG_u(r,\lambda)\right]' \left[\int \overline{K}_c^*(r)\, \overline{K}_c^*(r)'\right]^{-1} \\ & \times \left[\int \overline{K}_c^*(r), dG_u(r,\lambda)\right]' \end{aligned}$$

with $\overline{K}_c^*(r) = (1, K_c^*(r))'$ *and* $K_c^*(r) = K_c(r) - \int_0^1 K_c(r)\,dr$.

Note that the second part of this expression is as before of Brownian bridge type. The first component however, being nonstandard, can not be simplified further since $K_c^*(r)$ depends on the random disturbances in the dividend yields (ν_t) and $B_u(r,1)$ on the disturbances in the stock returns(u_t), which have possibly non zero correlation.

This nonstandard result is not surprising since, as shown by *Elliot(1998)*, standard inference about β cannot be mixed normal if x_t is a local to unit root process. This is due to the approximation of a unit root when using cointegration methods, leading to large size distortions of tests conducted. Hypothesis tests on β will tend to over reject on average when testing for the true value of β.

4 IVX

Elliot(1998) suggests to use an instrument for the near unit root process. This instrument needs to be exogenous to avoid the size distortion and reestablishing robustness of χ^2 inference with respect to the degree of persistence. *Phillips, Magdalinos (2009)* then proposed the IVX approach were the possibly mild stationarity of the persistent regressor is used to create an instrument for itself without using exogenous information. This instrumentalization removes long run endogeneities irrespective of x_t being integrated, near integrated or mildly integrated.

Formally the Instrument is constructed using an artificial persistence coefficient given by

$$R_t = 1 - \frac{c_z}{T^\delta} \quad , c_z > 0, \delta < 1$$

Which is then used to construct the instrument from the differences of the persistent variable.

$$\tilde{z}_t = R_T \tilde{z}_{t-1} + \Delta x_t$$

or in case of initialization at zero

$$\tilde{z}_t = \sum_{j=1}^{t} R_T^{t-j} \Delta x_j$$

As shown by *Phillips, Magdalinos (2009)* this instrument does indeed satisfy both, asymptotic relevance and asymptotic orthogonality restrictions. It was also shown that δ governs the trade off between asymptotic bias and long run endogeneities. Thus the restriction $\delta \in (2/3, 1)$ is required in order for the IV estimator based on this constructed instrument, to be asymptotically mixed-normally distributed. Comparing with *Theorem 3.4 PM09* it can be seen that in case of an mildly integrated process the long run endogeneities can be removed. Note that due to the MDS assumption for u_t the autocovariance $\lambda_{u\nu}$ is zero, such that no bias correction term is needed for the purposes of this paper.

Even though *PM09* have only covered the case of no fitted intercept, the follow up literature by *Kostakis et al. (2010)* has shown that the theorems are also valid when including an intercept. It is also important to highlight that the results obtained here do come at the cost of a decreased rate of convergence. This however is very much acceptable, especially at times of rapidly increasing data availability and computational power.

4.1 Proof proposition 3

By denoting demeaned quantities with an asterisk, the following IV estimator is obtained.

$$\tilde{\beta}^{ivx} = \frac{\sum y_t^* \tilde{z}_{t-1}^*}{\sum x_{t-1}^* \tilde{z}_{t-1}^*}$$

such that the modified Wald statistic can be written as

$$W_T^{ivx}(\beta = 0) = \frac{(\tilde{\beta}^{ivx})^2 (\sum x_{t-1}^* \tilde{z}_{t-1}^*)^2}{\tilde{\sigma}_u^2 \sum (\tilde{z}_{t-1}^*)^2}$$

Where $\tilde{\sigma}_u^2$ is given by $\sum (y_t^* - \tilde{\beta}^{ivx} x_{t-1}^*)^2 / T$, which is due to consistency of the IV-Estimator asymptotically equivalent to $\hat{\sigma}_{lin}^2$. Under H_0^B these residual variances are also equivalent to $\hat{\sigma}_u^2$.

As shown by *Park, Phillips (1988)* mixed normality of the estimator implies that inference using Wald tests with linear restrictions will be asymptotically χ^2 distributed. Thus by plugging in the IVX based Wald statistic in the decomposition from before

$$W_T^{B,ivx}(\lambda) = W_T^{ivx}(\beta = 0) + W_T^A(\lambda)$$

the nuisance parameter free limiting distribution of the Wald statistic testing linearity and no predictability jointly, is obtained as

proposition 3

$$sup_\lambda W_T^{B,ivx}(\lambda) \rightarrow W(1)^2 + sup_\lambda \frac{BB(\lambda)'BB(\lambda)}{\lambda(1-\lambda)}$$

Where $W(1)$ is a univariate standard normally distributed random variable. In summary *Gonzales, Pitarakis (2012)* have demonstrated a useful application of the IVX methodology. It is this segregation of the limiting distribution where one part follows standard and one non-standard asymptotics, that reveals the usefulness of IVX estimation. Using the IVX approach standard asymptotics can be recovered at relatively low costs.

5 Finite sample results

Complementary to the simulations conducted by *Gonzales, Pitarakis (2012)* some extensions are presented in this section. The following experiments were done using $N = 5000$ replications. For the power calculations a confidence level of 2.5% is assumed. The following covariance structures will be considered.

$$DGP_3 = \{\sigma_{ue}, \sigma_{uu_q}, \sigma_{eu_q}\} = \{-0.8, 0.5, 0.6\}$$
$$DGP_4 = \{\sigma_{ue}, \sigma_{uu_q}, \sigma_{eu_q}\} = \{-0.5, 0.8, 0.8\}$$
$$DGP_5 = \{\sigma_{ue}, \sigma_{uu_q}, \sigma_{eu_q}\} = \{-0.1, 0.3, 0.4\}$$

5.1 Testing $H_0^A : \alpha_1 = \alpha_2, \beta_1 = \beta_2$

To begin with, consider simulations of the critical values for the SupWald^A statistic, with several different covariance structures. For this simulations the parameters are exactly what they are in the original paper. The first configuration allows in general for stronger correlation between all three variables. After that a configuration was tested which allowed for moderate correlation between u and e and a strong correlation between the errors of the dependent/independent variable and the error of the threshold variable. The last configuration allows for moderate correlation with the threshold variable, while limiting the correlation between e and u to a rather small amount.

For the critical values of the SupWald^A only very small changes can be observed[1]. One thing that is worth noting is that even though the general effect of the changes in covariances is marginal, the most relevant change

[1]See Appendix A1

did occur when letting the correlation between u and e be more pronounced. The values obtained by changing the covariances of x and y with q, were nearly indistinguishable from those reported in *Gonzales, Pitarakis (2012)*. This indicates that the simulated CVs are robust to different covariance structures, as long as the degree of correlation with the threshold errors is held constant.

Now turning to the size properties of SupWald^A. For the three configurations above we see that the findings from before are again verified[2]. Only marginal decreases in size can be observed for lower quantiles, most prominently in the configuration of DGP_3 where a stronger correlation all around was allowed for. [3] In **Table 1** additional size estimates for the $sup_\lambda W_T^A$ statistic are presented. This simulation was performed using the covariance structure referred to as DGP_1.

$$DGP_1 : \{\sigma_{ue}, \sigma_{uu_q}, \sigma_{eu_q}\} = \{-0.5, 0.3, 0.4\}$$

Further the parameter were set as $\{\alpha, \beta, \phi\} = \{0.01, 0.1, 0.50\}$ while ρ takes on higher values then those reported in *Gonzales, Pitarakis (2012)* $(0.7, 0.8, 0.9)$. In contrast to the simulations in the original paper, the size estimates presented here are slightly smaller. However the estimates seem to remain robust to changes in the non centrality parameter c. Even though a slight decrease in size can be observed when c increases.

Table 2 contains power estimates for SupWald^A. In this simulation larger values for the AR coefficient ρ have been considered. What immediately stands out is that for larger values of ρ the power of the test increases strongly. In fact when beta is allowed to shift, for $c = 1$ and $\rho = 0.8$ we get a power of unity even at sample sizes as small as $T = 200$. The decreasing power for larger c can still be observed. When looking at DGP_3 where $\beta_1 = \beta_2$, notice that the power did not change substantially when compared to the results in the original paper. It is observed that the further ν moves towards a non stationary AR-process the higher the power of our tests. When setting $\rho = 1$ for $T = 200$ both the power for DGP_1 and DGP_2 are at unity. This might be useful in cases when the underlying economic theory allows for such high values of ρ.

[2]See Appendix A2

[3]One could also check what happens when flipping the sign of the covariances, however the change in economic interpretation has to be considered

Table 1: Size properties SupWald^A

	$T = 200$			$T = 400$		
$\rho = 0.7$	2.5%	5%	10%	2.5%	5%	10%
$c = 5$	1.86	3.92	8.04	1.94	4.10	8.78
$c = 10$	1.82	3.60	8.06	1.74	3.68	8.58
$\rho = 0.8$	2.5%	5%	10%	2.5%	5%	10%
$c = 5$	2.10	3.98	8.72	1.92	4.06	8.58
$c = 10$	1.80	3.52	7.90	1.72	3.90	7.94
$\rho = 0.9$	2.5%	5%	10%	2.5%	5%	10%
$c = 5$	2.34	4.72	9.40	1.96	4.00	8.82
$c = 10$	1.90	4.28	8.74	1.64	3.54	8.00

Table 2: Power properties SupWald^A

	$c = 1$			$c = 10$		
$\rho = 0.7$	DGP_1^A	DGP_2^A	DGP_3^A	DGP_1^A	DGP_2^A	DGP_3^A
$T = 200$	0.97	0.97	0.13	0.67	0.70	0.13
$T = 400$	1.00	1.00	0.36	1.00	1.00	0.33
$\rho = 0.8$	DGP_1^A	DGP_2^A	DGP_3^A	DGP_1^A	DGP_2^A	DGP_3^A
$T = 200$	1.00	1.00	0.13	0.92	0.92	0.13
$T = 400$	1.00	1.00	0.35	1.00	1.00	0.34

5.2 Testing $H_0^B : \alpha_1 = \alpha_2, \beta_1 = \beta_2 = 0$

Considering the size estimates for the SupWald^{B,ivx} statistics again using the covariance structures from before it becomes clear that the IVX based test statistic retains its relatively good size properties, even when allowing for substantial endogeneity in the data generating process[4]. Differences to the simulation in the original paper are again marginal. Now the behavior at the boundaries of suitable values for $\delta \in (2/3, 1)$ will be displayed in **Table 3**. The remaining parameter and the covariances are set according to the simulation in the original paper. It can be seen that even when approaching the limits of the range of suitable δ the size properties are still good. In fact it seems that in this simulation the size slightly decreases comparing with the simulations where δ takes on values in between $(0.66, 0.99)$. The robustness with respect to c can be confirmed too.

To asses the power properties of SupWald^{B,ivx} for different values of delta, consider the results in **Table 4**. Besides δ all parameters and the covariance structure remain unchanged. Looking at the results it can be observed that the power of the test slowly degrades when approaching the boundaries of suitable δ. It is also evident that for the two DGP with $\beta 1 \neq \beta 2$, higher values for c lead to less power when looking at small sample sizes. However both of this does not matter for large samples. [5] Already for $T = 400$ the power approaches unity with DGP_1 and DGP_2. Looking at the figures displayed by *Gonzales, Pitarakis (2012)* it can be suspected that even for DGP_3, power will reach unity when sample sizes are large enough.

[4]See Appendix A3

[5]Simulations for even larger samples and or different values of δ could be performed when equipped with better computational ressources.

Table 3: Size properties of $\text{SupWald}^{B,ivx}$

	$T = 200$			$T = 400$			$T = 1000$		
$\delta = 0.66$	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
$c = 1$	2.36	4.80	9.50	1.60	3.94	8.66	1.72	3.92	8.44
$c = 10$	1.82	3.90	9.16	1.90	3.96	8.86	1.68	4.02	8.58
$\delta = 0.99$	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
$c = 1$	2.18	4.88	10.36	2.18	5.04	10.96	1.82	3.62	9.00
$c = 10$	2.38	4.70	10.10	1.82	4.44	8.98	1.94	4.02	8.54

Table 4: Power properties $\text{SupWald}^{B,ivx}$

	$c = 5$			$c = 10$		
$\delta = 0.66$	DGP_1^B	DGP_2^B	DGP_3^B	DGP_1^B	DGP_2^B	DGP_3^B
$T = 200$	0.63	0.78	0.09	0.43	0.67	0.09
$T = 400$	0.99	1.00	0.23	0.99	1.00	0.20
$\delta = 0.99$	DGP_1^B	DGP_2^B	DGP_3^B	DGP_1^B	DGP_2^B	DGP_3^B
$T = 200$	0.83	0.90	0.09	0.61	0.78	0.09
$T = 400$	1.00	1.00	0.23	1.00	1.00	0.21

6 Conclusion

This Seminar paper aimed at shining some light onto the results derived by *Gonzales, Pitarakis (2012)* and to further examine the limiting behavior of the corresponding SupWald statisitcs. To do so, first the general concept of the proofs was recited. A second step consisted of evaluating the literature on near integrated process backing these proofs. By diving deeper into the works done prior to the analysis of *Gonzales, Pitarakis (2012)* the context of their work was clarified substantially. Looking at the general results for moments of persistent regressors as derived by *Phillips (1987,1988)* enables to understand the convergence results, which consist of multiple stochastic integrals such as Ornstein-Uhlenbeck processes and other derivatives[6] of Brownian motions. At this point the usefulness of segregating the SupWald^B statistic in order to use the IVX methodology to recover standard asymptotics, shall be highlighted.

In general the simulation results did not differ too much from those of *Gonzales, Pitarakis (2012)*. However when examining the power properties of SupWald^A it was observed that when ρ was allowed to increase the power of the test increases and reached unity even for small sample sizes. When looking at the properties for SupWald^{b,ivx} it becomes evident that letting δ being close to the boundaries of its suitable values leads to small decrease in size and less power for small sample sizes.

Gonzales, Pitarakis (2012) mentioned that further extensions of their work could consist of adding more regimes. This could be an excellent addition. When working with economic data one can often observe that while recessions often hit almost immediate and can be well approximated using thresholds, the recovery often is less rapid and occurs steadily over time. One could even use smooth transition models to explore this and model the reality more thoroughly.

[6] In the sense of these processes being derived using Brownian motions, not being literal derivatives.

References

Caner, M., & Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, *69*(6), 1555–1596.

Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica*, *66*(1), 149–158.

Gonzalo, J., & Pitarakis, J.-Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics*, *30*(2), 229–241.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, 413–430.

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, *28*(5), 1506–1553.

Park, J. Y., & Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*(3), 468–497.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, *74*(3), 535–547.

Phillips, P. C. (1988). Regression theory for near-integrated time series. *Econometrica: Journal of the Econometric Society*, 1021–1043.

Phillips, P. C. (1998). New tools for understanding spurious regressions. *Econometrica*, 1299–1325.

Phillips, P. C., & Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, *136*(1), 115–130.

Phillips, P. C., Magdalinos, T., et al. (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, *7*.

White, H. (2014). *Asymptotic theory for econometricians.* Academic press.

7 Appendix

A1

A1.1

Table 1: Critical values of SupWald^A using DGP_3

$T = 200$	$c = 1$	$c = 5$	$c = 10$	$c = 20$
2.5%	2.235126	2.199667	2.279775	2.194146
5%	2.585451	2.579798	2.571831	2.561704
10%	3.072148	3.058625	3.069692	3.056029
90%	10.286226	10.242859	10.273899	10.193191
95%	12.133219	11.985353	12.312315	12.194345
97,5%	14.091052	13.808342	13.680701	13.892900
$T = 400$	$c = 1$	$c = 5$	$c = 10$	$c = 20$
2.5%	2.247267	2.209213	2.253528	2.236187
5%	2.579902	2.544711	2.597636	2.624149
10%	3.079296	3.034250	3.034367	3.074554
90%	10.077608	10.170321	10.168687	10.297120
95%	11.596533	11.843554	12.041287	12.005275
97,5%	13.603813	13.466795	13.651400	13.392880

A1.2

Table 2: Critical values of SupWald^A using DGP_4

$T = 200$	$c = 1$	$c = 5$	$c = 10$	$c = 20$
2.5%	2.181005	2.176156	2.167967	2.139738
5%	2.536220	2.549974	2.580814	2.536516
10%	2.989260	3.061161	3.081808	2.983224
90%	10.050641	10.393884	10.480988	10.312545
95%	11.771356	12.035687	12.520682	12.294659
97,5%	13.688936	13.956737	14.221089	14.196536
$T = 400$	$c = 1$	$c = 5$	$c = 10$	$c = 20$
2.5%	2.246823	2.178327	2.238088	2.312498
5%	2.608373	2.591387	2.610439	2.626736
10%	3.106593	3.128877	3.073905	3.090717
90%	10.300818	10.128346	10.219072	10.224954
95%	12.200278	11.898967	11.924512	12.247952
97,5%	14.141966	13.605808	13.681100	14.042373

A1.3
Critical values of SupWald^A using DGP_5

$T = 200$	$c = 1$	$c = 5$	$c = 10$	$c = 20$
2.5%	2.184147	2.137154	2.135950	2.161834
5%	2.559134	2.567031	2.494475	2.535744
10%	3.012641	3.013868	2.967961	3.072727
90%	10.231071	10.119924	10.436441	10.178015
95%	12.094198	11.887370	12.182816	12.126929
97,5%	13.827477	13.821674	14.190522	13.987551
$T = 400$	$c = 1$	$c = 5$	$c = 10$	$c = 20$
2.5%	2.253581	2.173322	2.193451	2.149608
5%	2.601410	2.548339	2.562767	2.514793
10%	3.080575	2.994641	2.973710	3.083704
90%	10.399873	10.096074	10.049979	10.292524
95%	12.190549	12.096529	11.781594	12.131211
97,5%	14.287968	13.766618	13.723472	14.209371

A2

A2.1

Size properties SupWald^A for DGP_3

	$T=200$			$T=400$		
	2.5%	5%	10%	2.5%	5%	10%
$c=1$	1.84	4.40	9.72	1.92	4.60	9.06
$c=5$	1.60	3.56	8.36	1.90	4.12	8.76
$c=10$	2.20	4.26	9.22	1.82	4.14	9.28
$c=20$	1.64	3.92	8.78	1.70	4.12	8.64

A2.2

Size properties SupWald^A for DGP_4

	$T=200$			$T=400$		
	2.5%	5%	10%	2.5%	5%	10%
$c=1$	1.82	3.96	8.06	2.34	4.92	9.68
$c=5$	2.12	4.22	8.50	1.92	3.74	8.50
$c=10$	2.06	4.02	8.96	2.12	4.54	8.78
$c=20$	1.82	3.80	8.10	1.96	4.32	8.52

A2.3

Size properties SupWald^{A} for DGP_5

	$T = 200$			$T = 400$		
	2.5%	5%	10%	2.5%	5%	10%
$c = 1$	1.96	4.12	8.44	2.16	4.34	9.22
$c = 5$	2.32	4.26	9.00	1.82	3.58	7.96
$c = 10$	2.06	3.80	8.34	2.08	4.24	8.58
$c = 20$	2.10	4.00	8.28	2.26	4.32	8.78

A3

A3.1

Size properties SupWald^{B,ivx} for DGP_3

	$T = 200$			$T = 400$		
$\delta = 0.7$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.52	5.00	9.92	2.12	4.24	9.42
c= 10	2.38	4.62	9.82	1.90	3.84	8.56
$\delta = 0.8$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.64	5.46	10.62	2.54	5.32	9.92
c= 10	2.32	4.72	9.70	2.20	4.84	9.54
$\delta = 0.9$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.40	4.98	10.46	2.76	5.34	10.60
c= 10	1.94	4.14	8.94	2.22	4.62	9.46

A3.2

Size properties SupWald^{B,ivx} for DGP_4

	$T = 200$			$T = 400$		
$\delta = 0.7$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.38	4.54	9.16	2.18	4.30	10.06
c= 10	2.34	4.50	9.26	1.98	4.32	8.92
$\delta = 0.8$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.26	4.76	9.88	1.96	4.20	9.04
c= 10	1.90	4.12	8.90	2.34	4.26	9.38
$\delta = 0.9$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.22	4.98	10.18	2.44	4.84	10.08
c= 10	1.98	4.50	9.16	1.72	4.14	9.16

A3.3

Size properties SupWald^{B,ivx} for DGP_5

	$T = 200$			$T = 400$		
$\delta = 0.7$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.08	4.06	8.38	1.72	4.42	8.96
c= 10	1.88	3.76	8.30	1.98	4.14	8.86
$\delta = 0.8$	2.5%	5%	10%	2.5%	5%	10%
c= 5	2.16	4.46	9.40	1.90	4.00	8.88
c= 10	2.04	4.14	8.98	1.92	3.66	8.24
$\delta = 0.9$	2.5%	5%	10%	2.5%	5%	10%
c= 5	1.88	4.34	9.64	2.20	4.50	9.26
c= 10	2.04	4.30	8.90	1.90	4.10	8.88