Housing and the Great Depression

Mehmet Balcilar
Eastern Mediterranean University

Rangan Gupta
University of Pretoria

Stephen M. Miller
University of Nevada-Las Vegas
University of Connecticut

Housing and the Great Depression

Mehmet Balcilar
Department of Economics
Eastern Mediterranean University
Famagusta, NORTHERN CYPRUS, via Mersin 10, TURKEY

Rangan Gupta
Department of Economics
University of Pretoria
Pretoria, 0002, SOUTH AFRICA

Stephen M. Miller*
Department of Economics,
University of Nevada, Las Vegas
Las Vegas, Nevada, 89154-6005 USA

Abstract:

This paper considers the role of the real housing price in the Great Depression. More specifically, we examine structural stability of the relationship between the real housing price and real GDP per capita. We test for structural change in parameter values, using a sample of annual US data from 1890 to 1952. The paper examines the long-run and short-run dynamic relationships between the real housing price and real GDP per capita to determine if these relationships experienced structural change over the sample period. We find that temporal Granger causality exists between these two variables only for sub-samples that include the Great Depression. For the other sub-sample periods as well as for the entire sample period no relationship exists between these variables.

Keywords: Great Depression, Real House Price, Real GDP per Capita, Structural change

JEL classification: C32, E32, R31

* Corresponding author

1. Introduction

Recent events during the financial crisis and Great Recession confirm that movements in housing markets not only reflect developments in macroeconomic fundamentals, but also provide important impulses to business fluctuations (Iacoviello, 2010). In his introductory remarks at a conference on Housing and Mortgage Markets, Federal Reserve Chairman Bernanke (2008) noted that "… housing and housing finance played a central role in precipitating the current crisis.", thus emphasizing the importance of spillover effects from the housing market onto the real economy.

This possibility of such a spillover effects possesses well-grounded theoretical foundations. The permanent income hypothesis of Friedman (1957) and the life-cycle model of Ando and Modigliani (1963) imply that households allocate some of their permanent income or wealth to finance consumption. Thus, changes in housing wealth that affect permanent income or wealth alter consumption spending. Case, *et al.* (2005) provide a good recent review of wealth effects. While the original simple life-cycle model of consumption does not distinguish between different types of wealth, implicitly assuming that the marginal propensities to consume out of wealth remains the same across different wealth types, reasons exist to suggest that this implicit assumption is, in fact, invalid. Case, *et al.* (2005) offer five different possible rationalizations for different marginal propensities to consume out of different types of wealth – differing perceptions about the effects of permanent and transitory components, differing bequest motives, differing motives for wealth accumulation, differing abilities to measure wealth accumulation, and differing psychological "framing" effects.

Another possible rationalization, not mentioned by Case, *et al.* (2005), involves whether the wealth holder receives consumption services from the holding of wealth. For example, owner occupied housing and consumer durable goods provide consumption services to holders of these components of wealth. Thus, households may adjust their consumption of

nondurables and services, the usual measure of consumption for wealth-effect studies, differently to changes in the market values of owner occupied housing and consumer durables than to changes in other forms of wealth that do not possess such services.

More recent theoretical work (e.g., Bajari, *et al.* 2005; Buiter 2008) suggests that house price changes that reflect changes in fundamental value do not produce aggregate consumption changes but merely redistributed wealth between households who are long and short in housing wealth. Nevertheless, changes in house prices that constitute changes in the speculative component of house prices do exhibit a wealth effect. Finally, other authors (e.g., Aoki, *et al.* 2004; Buiter 2008; Lustig and van Niewerburg 2008) argue that changes in house prices affect the collateral value and that this can affect actual consumption for those financially constrained households who want to consume more than their financial circumstances permit.

White (2009) argues that the "forgotten" real estate boom and collapse in the 1920s shares many "familiar characteristics" with the most recent US housing price run-up and collapse. He cites several factors ("weak supervision, securitization, and a fall in lending standards" p. 4) as well as two monetary factors ("a 'Greenspan put' by the Federal Reserve" and "low interest rates" p. 4) as common elements. The housing price index constructed by Grebler, Blank, and Winnick (1956) and used as one of the components by Shiller (2005) to construct his housing price index series shows a clear peak in the nominal index in 1925, followed by a substantial decline in the index through 1934. Nicholas and Scherbina (2012) recently constructed real estate price indexes for Manhattan, New York, showing that prices peak in 1929 and then decline substantially during the 1930s. Nevertheless, the run-up and decline in housing price indexes in the 1920s and 1930s do not exhibit the magnitudes seen in the most recent housing price boom and collapse. White (2009) attributes the larger swing in

housing prices in the most recent boom and collapse to "risk-taking induced by federal deposit insurance and aggressive homeownership policies absent in the 1920s." (p. 4).

Given the role played by housing in causing the Great Recession, a pertinent question to ask is: What role, if any, did housing play during the Great Depression? Macroeconomists have long debated the proximate causes of the Great Depression. Friedman and Schwartz (1963) posited monetary policy errors or omissions as the proximate cause – the "money" view. Temin (1976), among others, argued for the collapse of aggregate demand, an autonomous fall in consumption demand, as the proximate cause – the "demand view. Others adopted elements of the "money" and "demand" points of view to develop integrated arguments explaining the Great Depression. For example, Mishkin (1978) focused on the household balance sheet, combining elements of his liquidity hypothesis (Mishkin 1976) and the life-cycle hypothesis (Ando and Modigliani 1963), as critical factors in the explanation of the Great Depression. In his model, he highlighted the role of declining demand for housing and consumer durable goods. Bernanke (1983) proposed that the financial crisis disrupted the credit allocation process, leading to higher credit allocation costs, reducing credit availability, and lowering aggregate demand. He highlighted two major factors as the root problem in the Great Depression -- the failure of financial institutions and the defaults and bankruptcies. Gordon and Wilcox (1981) stressed the roles of construction, consumption, the stock market, and the Hawley-Smoot tariff in explaining the severity of the Great Depression.

Researchers offer a variety of explanations for the causes of the Great Depression. As our main objective, and to the best of our knowledge the first such attempt, we analyze if, and how, housing, specifically real house price movements, played a role in causing the Great Depression. To achieve our goal, we consider time-varying (using a rolling window of 15 years), bootstrapped causality between the real house price (proxy for housing wealth) and real GDP per capita over the annual period from 1890 to 1952. A full-sample causality test

over this period may not help to decipher the leading role, if any, for the real house price in causing the Great Depression. That is, the relationship between real house price and real GDP per capita may prove unstable. We specifically test for such instability in both the short and long-runs. We find that we cannot rely on full-sample causality tests.

The rest of the paper is organized as follows: Section 2 outlines the econometric method. Section 3 discusses the data and presents the results on the relationship between the real housing price and real GDP per capita. Finally, Section 4 concludes.

2.   Method:

We examine whether the real house price Granger causes real GDP per capita. Our null hypothesis is Granger non-causality, which we define as follows. One variable (e.g., the real house price) does not Granger cause a second variable (e.g., real GDP per capita) when the information on the first variable (the real house price) does not improve the prediction of the second (real GDP per capita) over and above its own information. Granger non-causality tests whether the lagged values of the first variable (the real house price) jointly prove insignificant, using Wald, likelihood ratio (*LR*), and Lagrange multiplier (*LM*) statistics. These Granger-non-causality test statistics assume stationary underlying time series. If the series exhibit nonstaionarity, then standard asymptotic distribution theory does not hold. Park and Phillips (1989) and Toda and Phillips (1993, 1994), among others, illustrate the difficulties arising in the levels estimation of such nonstationary *VAR* models.

Full-Sample Analysis

Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) propose a modification to the standard Granger causality test, obtaining standard asymptotic distributions when the time series forming the *VAR(p)*, where *p* is the lag order, are *I(1)*. Employing their method, one estimates a *VAR(p+1)* in levels and the resulting modified Granger causality tests remain valid irrespective of integration-cointegration properties of the variables. That is, the

4

modification estimates a *VAR(p+1)* and performs the Granger non-causality test on the first *p* lags. Thus, one coefficient matrix, which relates to the $(p+1)^{st}$ lag, remains unrestricted under the null, giving the test a standard asymptotic distribution.

Shukur and Mantalos (1997b) evaluate the power and size properties of eight different Granger non-causality tests in standard and modified forms using Monte Carlo simulations, including the modification proposed by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996). The simulations indicate that the Wald test possesses the wrong size in small and medium size samples. Additionally, Shukur and Mantalos (1997a) demonstrate that the residual based bootstrap (*RB*) method improves the critical values and the true size of the test approaches its nominal value in models with one to ten equations. In addition, Mantalos and Shukur (1998) examine the properties of the *RB* method in *VAR* models with cointegrated time series, discovering that the *RB* critical values prove more accurate than asymptotic ones and tests based on the *RB* method also prove more robust. Further, Shukur and Mantalos (2000) explore the properties of various versions of Granger causality tests and report that *LR* tests with small sample correction exhibit relatively better power and size properties, even in small samples. They also document that all standard tests not based on the *RB* method perform poorly when no cointegration holds, especially in small samples.

Mantalos (2000) compares bootstrap, corrected-*LR*, and Wald causality tests and finds that the bootstrap test exhibits better power and size in all cases, regardless of whether the variables are cointegrated. Hacker and Hatemi-J (2006) show that the modified Wald causality test, proposed by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996), with critical values obtained from the *RB* bootstrap method exhibit much smaller size distortion compared to the tests based on the asymptotic distribution. Based on these findings, we use the *RB* based modified-*LR* statistics to examine the causality between the real house price and real GDP per capita in the US.

5

To illustrate the bootstrap modified-*LR* Granger causality test procedure, consider the following bivariate *VAR(p)* process:

$$z_t = \Phi_0 + \Phi_1 z_{t-1} + \; \ldots \; + \Phi_p z_{t-p} + \varepsilon_t, \quad t = 1, \; 2, \; \ldots, T, \tag{1}$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a white noise process with zero mean and covariance matrix Σ and p is the lag order of the process. In the empirical section, we use the Akaike Information Criterion (*AIC*) to select the lag order p. To simplify, we partition z_t into two sub-vectors, the real house price (z_{ht}) and real GDP per capita (z_{yt}) and rewrite equation (1) as follows:

$$\begin{bmatrix} z_{ht} \\ z_{yt} \end{bmatrix} = \begin{bmatrix} \phi_{h0} \\ \phi_{y0} \end{bmatrix} + \begin{bmatrix} \phi_{hh}(L) & \phi_{hy}(L) \\ \phi_{yh}(L) & \phi_{yy}(L) \end{bmatrix} \begin{bmatrix} z_{ht} \\ z_{yt} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ht} \\ \varepsilon_{yt} \end{bmatrix}, \tag{2}$$

where $\phi_{ij}(L) = \sum_{k=1}^{p+1} \phi_{ij,k} L^k$, $i, j = h, \, y$ and L is the lag operator such that $L^k z_{it} = z_{it-k}$, $i = h, \, y$.

In this setting, the null hypothesis that real GDP per capita output does not Granger cause the real house price implies that we can impose zero restrictions $\phi_{hy,i} = 0$ for $i = 1, \; 2, \; \ldots, p$. In other words, real GDP per capita does not contain predictive content, or is not causal, for the real house prices when we cannot reject the joint zero restrictions under the null hypothesis:

$$H_0^{HF}: \quad \phi_{hy,1} = \phi_{hy,2} = \cdots = \phi_{hy,p} = 0. \tag{3}$$

Analogously, the null hypothesis that the real house price does not Granger cause real GDP per capita implies that we can impose zero restrictions $\phi_{yh,i} = 0$ for $i = 1, \; 2, \; \ldots, p$. Now, the real house price does not contain predictive content, or is not causal, for real GDP per capita when we cannot reject the joint zero restrictions under the null hypothesis

$$H_0^{LI}: \quad \phi_{yh,1} = \phi_{yh,2} = \cdots = \phi_{yh,p} = 0. \tag{4}$$

We can link the Granger causality tests in equations (3) and (4) to the *HF* (housing fundamentals) and *LI* (leading indicator) hypotheses as follows. If we reject H_0^{LI} in equation

(4) but not H_0^{HF} in equation (3), then we establish evidence in favor of the *LI* hypothesis. On the other hand, if we reject the null hypothesis specified under H_0^{HF} in equation (3), but not the null hypothesis specified under H_0^{LI} in equation (4), then the findings supports the *HF* hypothesis. Of course, the (non) rejection of both null hypotheses implies (no) support for the *HF* and *LI* hypotheses.

Efron (1979) pioneered the bootstrap method, using critical or *p* values generated from the empirical distribution derived for the particular test using the sample data. In our case, we use the bootstrap approach to test for Granger non-causality. Several studies document the robustness of the bootstrap approach for testing Granger non-causality (e.g., Horowitz, 1994; Mantalos and Shukur, 1998; and Mantalos 2000). Using Monte Carlo simulations, Hacker and Hatemi-J (2006) show that the modified Wald test based on a bootstrap distribution exhibits much smaller size distortions compared to the use of asymptotic distributions. Moreover, these results hold irrespective of sample sizes, integration orders, and error-correction processes (homoscedastic or *ARCH*). In this paper, we adopt the bootstrap approach with the Toda and Yamamoto (1995) modified causality tests because of several advantages. In particular, this test applies to both cointegrated and non-cointegrated *I(1)* variables (Hacker and Hatemi-J, 2006).

Standard Granger non-causality tests assume that the *VAR* model's parameters remain constant over time, an assumption which may not hold. Granger (1996) argued that parameter non-constancy is one of the most challenging issues confronting empirical studies today. Although we can detect the presence of structural changes beforehand and we can modify our estimation to address this issue using several approaches, such as including dummy variables and sample splitting, such an approach introduces pre-test bias. In this paper, we adopt rolling bootstrap estimation to address parameter non-constancy and avoid pre-test bias. To examine

7

the effect of structural changes, we use rolling window Granger causality tests, which also use the modified bootstrap test. Structural changes may shift the parameters and the pattern of the causal relationship may change over time.

Recursive and Rolling Analysis

Rather than testing for structural breaks using the full sample, we now consider recursive and rolling estimates over the full sample period and implement the fluctuations (*FL*) test of Ploberger *et al.* (1989) and the moving-estimates (*ME*) test of Chu *et al.* (1995b). The *FL* and *ME* tests correspond to the recursive and rolling regressions. Recursive regressions start with an initial benchmark sample at the beginning of the full sample and then proceeds by expanding the sample by adding one observation at a time until reaching the end of the full sample. The rolling regression also begins with a benchmark sample at the beginning of the full sample and keeps the sample size constant as the subsample roll through the full sample adding one new observation and deleting the oldest observation each time until reaching the end of the full sample.

By structural change, we mean parameter instability in an econometric model, where by parameters estimates become worthless, statistical inference becomes invalid, and forecast accuracy becomes imprecise. Stock and Watson (1996) explore the entent and severity of this issue, using 76 major US macroeconomic time series with several thousand forecasting relationships and show that a substantial fraction of these forecasting relationships display parameter instability. In our case, the Granger causality tests applied to the full sample are invalid, if the parameters of the *VAR* model prove unstable. Moreover, the cointegration tests will also be invalid, if the parameters of the long-run equation vary over time. Therefore, we will check for the stability of the both the short-run parameters in the *VAR* model and the parameters of the long-run equation between the GDP and real house price.

A model's structure may deviate from assumed stability in numerous ways, leading to many potential alternative specifications against the null hypothesis of stability. Therefore, those tests that leave the form of instability unspecified possess desirable properties. Considering the many alternatives for practical problems, researchers require (1) a wide variety of tests to ensure that these tests exhibit power against some conceivable number of alternatives, as some tests do indeed possess little power against violations of their assumptions, such as the stationarity, no autocorrelation, no outliers, and so on and (2) tools that permit an understanding of the nature of deviations from stability so that the researcher can date the structural change along with the causes. In view of (1), we include a battery of tests that possess power against both specific alternatives and unspecified alternatives, use robust estimation methods against the known issues, such as the nonstationarity, autocorrelation, and outliers. In view of (2), we use rolling and recursive estimations and tests that permit the determination of the form of deviations from the stability and also to date the structural changes. To wit, significance tests re-combined with graphical analysis based on rolling and recursive estimates gives insights on the nature and evolution of the structural change.

To illustrate the structural change tests, let all coefficients of the $VAR(p)$ in equation (1) to vary over time and stack the coefficient matrices in the matrix $\theta_t = [\Phi_{0t}, \Phi_{0t}, \ldots, \Phi_{pt}]'$. Then, we can write the $VAR(p)$ model in the following form:

$$z_t = x_t'\theta_t + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{5}$$

where $x_t' = I_2 \otimes [1, z_{t-1}', z_{t-2}', \ldots, z_{t-p}']$, the symbol $\otimes$ denotes the Kronecker product, and ε_t is multivariate, normally distributed with variance Σ, $\varepsilon_t \sim \mathrm{MVN}(0, \Sigma)$. In addition to the stability of the $VAR(p)$ model in equation (5), we also investigate the stability of the long-run

relationship, if any, between real GDP per capita and the real house price. We can also write equation (5) for the long-run relationship as follows:

$$z_t = x'_t\theta_t + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{6}$$

where $z_t = GDP_t$ and $x_t = RHP_t$. The parameter stability tests apply equally to the parameters of the *VAR* model in equation (5) and the long-run relationship in equation (6). In the following discussion we will not make a distinction on how the tests apply to equations (5) and (6), differences will be pointed out, when they occur.

Parameter stability tests applied to equations (5) and (6) test whether the coefficients in these regressions remain constant over time. That is, we test the null hypothesis of parameter stability,

$$H_0 : \theta_t = \theta_0, \qquad \forall t = 1, 2, \ldots, T, \tag{7}$$

against the alternative that at least one of the parameters in θ_t varies over time. Several patterns of deviation from the constant parameter specification under the null in equation (7) exist, including single or multiple breaks, swift or gradual changes, and random walk parameters, and so on. Testing approaches assume that structural breaks occur in known or unknown periods. In our application, we will only consider tests that do not require prior knowledge of the dates of the structural breaks. We will also consider recursive and rolling analysis that leaves the time variation in parameters under the alternative unspecified. Hansen (2001) offers a survey of parameter stability tests and related issues.

Give the vast array of alternative to the structural stability hypothesis in equation (7), several popular branches of structural change tests have emerged in recent years. These branches include *F* tests, fluctuation tests, and maximum likelihood (*ML*) scores tests. These tests address the form of structural change in different ways, having power against different

forms of deviations from the constant parameter case. Here, we only consider tests that do not require knowing the date(s) of structural change a priori.

F tests assume a single structural change under the alternative at an unknown time. Andrews (1993) and Andrews and Ploberger (1994) propose three type of *F* tests: *Sup-F*, *Ave-F*, and *Exp-F*, either based on Wald, *LM*, or *LR* statistics. *F* tests rely on sequence of *F* statistics for a structural change at time *i*. We compute the statistics from segmented regressions (i.e., one regression estimate for each subsample determined by the break point, where the break point sequentially increases by one). That is, break points occur at $i = T_h, T_h + 1, \ldots, T - T_h$, leading to a $T - 2T_h + 1$ sequence of *F* statistics.

The test computation involves estimating two regressions -- one with no structural change and parameters $\theta_t = \theta_0$ for $t = 1, 2, \ldots, T$ and one with structural change and parameters $\theta_t = \theta_0 + \delta$ for $t = i, i+1, \ldots, T$. We construct *F* tests for $\delta = 0$ at each $i = T_h, T_h + 1, \ldots, T - T_h$. These *F* tests rejects H_0, if their supremum, average, or mean functional is too large. We can apply the tests to general classes of models fitted by ordinary least squares (*OLS*), instrumental variables, or generalized method of moments (*GMM*).[1] In our applications, we prefer *LR* based *F* tests since the *LR* tests possess advantages in our causality tests framework based on bootstrap.[2] *F* tests require trimming and we set $T_h = [hT]$ with *h*=0.15 (i.e., we trim 15 percent of the observations from the both ends of the sample).

On the other hand, the class of fluctuation tests, which use fluctuation processes computed from parameter estimates, residuals, or scores, do not assume *a priori* any specific

---

[1] As the *F* tests are easy to interpret, can determine a single break date in a fixed interval, and possess some certain weak optimality against single break alternatives, they gained popularity in the last two decades and have become the most preferred structural change tests in empirical studies.

[1] In our empirical application, we have also calculated *LM* versions of the *F* tests and results were qualitatively the same. *LM* test results are available from the authors.

form of structural break or pattern of change in the parameters, unlike the *F* test. Fluctuation tests exhibit high flexibility with regards to the alternatives. Brown *et al.* (1975) point out the fluctuation process approach " … includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey (1962). Essentially, the techniques are designed to bring out departures from constancy in a graphic way instead of parametrizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives." (pp. 149–150). Fluctuations processes do serve as explorative tools for discovering structural changes in a visual way. In addition, successful formal tests based on the fluctuation processes do exist.

Fluctuation tests first estimate the specified model in a recursive (expanding window) or rolling (fixed window) manner and then constructs a process that captures the fluctuation either in residuals (Ploberger and Kramer 1992, Chu et al. 1995a) or estimates (Ploberger et al. 1989, Chu et al. 1995b). Under the null hypothesis that parameters are constant, these fluctuation processes are governed by functional central limit theorems, converging to a functional Weiner process or Brownian bridge. Therefore, we can determine the boundaries of the limit processes with fixed probability α under the null hypothesis, allowing one to perform formal statistical tests. Under the alternative hypothesis, when true, the fluctuations in the processes generally increase. A visual inspection of the trajectory of these processes serves as an exploratory tool for determining the type of the deviation from the null hypothesis and the dating of structural breaks. We can estimate the parameters of the model by ordinary least squares or *ML* with normal error assumption.

Residual-based fluctuations tests are easy to compute and interpret. They do not give any intuition, however, about the likely cause of the rejection of parameter stability. In the estimates-based fluctuation tests, one process exists for each coefficient and we can examine each process separately. We can easily construct an overall process by aggregating over the

individual components. Recursive and rolling fluctuation tests use the following recursive and rolling estimators, respectively,

$$\hat{\theta}_j = \left(\sum_{t=1}^{j} x_t x_t' \right)^{-1} \left(\sum_{t=1}^{j} x_t z_t \right), \qquad j = k, 2, \ldots, T, \text{ and} \tag{8}$$

$$\hat{\theta}_{j,h} = \left(\sum_{t=j+1}^{j+[hT]} x_t x_t' \right)^{-1} \left(\sum_{t=j+1}^{j+[hT]} x_t z_t \right), \qquad j = 0, 2, \ldots, T - [hT] + 1, \tag{9}$$

where $0 < h < 1$ determines the window size for the rolling estimates. Now, define

$$Q_{[mT]} = \frac{1}{[mT]} \sum_{t=1}^{[mT]} x_t x_t', \quad \text{and} \quad Q_{[rT],h} = \frac{1}{[hT]} \sum_{t=[rT]+1}^{[rT]+[hT]} x_t x_t',$$

where $0 < m \leq 1$ and $0 \leq r < 1$. The full sample matrix Q_T scales matrices in constructing recursive fluctuation (*FL*) test by Ploberger et al. (1989) and also by Chu et al. (1995b) in constructing the rolling fluctuation (*ME*) test. Kuan and Chen (1994) argue that the *FL* and *ME* tests experience serious size distortions in dynamic models (i.e., in the presence of autocorrelation). Therefore, these tests more likely reject null hypotheses of parameter constancy in dynamic models. Kuan and Chen (1994) further show that when the size of these tests improve significantly when using the subsample estimates $Q_{[mT]}$ and $Q_{[rT],h}$. In our case, the residuals may exhibit significant autocorrelation. Thus, to address this issue, we use the modified *FL* and *ME* tests proposed by Kuan and Chen (1994). The modified tests are defined as follows:

$$FL = \max_{k \leq j \leq T} \frac{j}{\hat{\sigma}_T \sqrt{T}} \| Q_j^{1/2} (\hat{\theta}_j - \hat{\theta}_T) \| \text{ and} \tag{10}$$

$$ME = \max_{0 \leq j \leq T - [hT]+1} \frac{[hT]}{\hat{\sigma}_T \sqrt{T}} \| Q_{j,h}^{1/2} (\hat{\theta}_{j,h} - \hat{\theta}_T) \|, \tag{11}$$

where $\hat{\sigma}_T^2$ is the estimator of the error variance and $\| \cdot \|$ is the L_2 norm. We prefer the L_2 norm because it aggregates over the compon, which leads to better power and size properties

13

when several, or all, parameters change simultaneously. In implementing the *ME* test, we use a window parameter of *h*=0.25, implying that the ratio of the number of observations in each window to total number of observations is 0.25.[3]

Nyblom (1989) proposed an *LM* test based on the *ML* scores, denoted L_c. Hansen (1992a, 1992b) generalizes L_c the test to linear models and to models with integrated variables, respectively. We can transform the *ML* scores test into the framework of the fluctuation tests, although it possesses a quite different motivation. The fluctuation processes in the ML scores test comes from the first order conditions. Indeed, the L_c test uses the full sample parameter estimates. That is, given the parameter estimates, we evaluate the scores and form the fluctuation processes from the empirical scores. We can can estimate the parameters with *OLS*, *ML* with normal errors, or other methods such as the *GMM* and fully-modified *OLS* (*FM-OLS*) estimator (Phillips and Hansen 1990). In empirical applications, researchers usually prefer the *FM-OSL*, since it can estimate the parameters of cointegration regressions, accounting for possible endogenity of the right-hand-side variables and autocorrelations in the error term. The L_c test can also examine the stability of the parameters of *I(1)* regression models. In our application, we also use *FM-OLS* estimator due to its advantages. To examine the stability of the cointegration parameters, we emphasize the L_c tests. The L_c test is an *LM* test for parameter constancy against the alternative hypothesis that the parameters follow a random-walk process and, therefore, time-varying, since the first two moments of a random walk depend on time. The random-walk alterative makes L_c test suitable as a test for cointegration, when the alternative is that the intercept follows a random walk.

---

[3] The results that use window parameters 0.20 and 0.30 are available from authors. They produce qualitatively similar findings to those reported in the empirical section.

In sum, parameter instability can occur in many ways. This fact precludes us from covering all conceivable forms of parameter instability. We can only avoid the problem if we know the exact form of the deviation from parameter constancy. Given the difficulty test selection, we use several tests based on their optimality properties. The *Sup-F* test exhibits good power against single breaks and can usefully date structural breaks. The *Sup-F* test also performs better in detecting tail shifts in small samples. This test, however, displays low power when multiple breaks exist and in the presence of random-walk alternatives. With random-walk alternatives, the *ML* scores based test L_c possesses better power. The L_c test performs well against mid-point structural breaks, but not well against tail breaks. Moreover, the L_c test is preferred for examining the stability of long-run equilibrium regressions with *I(1)* variables. In such cases, its use with *FM-OLS* estimator also serves as a test for cointegration. The recursive estimation-based fluctuation test *FL* possesses similar properties to the *Sup-F* test, particularly for visual inspection of the pattern of the structural change. The *Sup-F* test assumes a single break and is best suited for detecting one-time structural changes. On the other, the rolling estimation based fluctuation test *ME* displays better power against multiple-breaks and random-walk alternatives. The *ME* test is particular appropriate where parameters temporarily deviate from a "normal" level. Like the *FL* test, it serves a useful explorative tool for understanding the pattern and form of the structural change.

3.   Data and Results

Data Sources

We test the relationships between the real housing price and real GDP per capita, using annual US data from 1890 to 1952. While the real house price data come from Shiller (2005), the data for real GDP at constant 2005 dollars and population to compute the real GDP per capita come from the Global Financial Database. Consistency with the theoretical models of wealth effects implies ideally that we should use data on housing wealth. The unavailability

of housing wealth for the period under consideration requires us to use the housing price index as a proxy for housing wealth, which, of course, represents a limitation on our statistical analysis. First, we test for the order of integration of the two series. Second, we perform multivariate cointegration tests. Third, we determine the full-sample Granger causality tests. Fourth, we perform various tests on parameter stability from the coefficient estimates from our rolling *VAR* regressions. Finally, we estimate rolling *VAR* regressions and perform Granger causality tests with a fixed 15-year window.

Full Sample Unit Roots, Cointegration, and Granger Temporal Causality

We first test for the presence of unit roots in the real house price and real GDP per capita series under consideration using the Phillips (1987) and Phillips and Perron (1988) test. We perform tests with both a constant and a constant and a time trend. As test statistics exhibit nonstandard distributions and critical values, we use the critical values computed by MacKinnon (1996). Table 1 reports the results of unit-root tests. We fail to reject the null hypothesis of nonstationarity for the real house price and the real GPD per capita series at 5-percent level. Further, we do reject the null of non-stationarity for the first differences of these series, implying that both series are *I(1)* processes.

We next test for a common stochastic trend, which implies a cointegrating relationship between the two series. We use Johansen's (1991) maximum likelihood method, which requires that we first identify the lag structure of the bivariate *VAR* model. We search for the optimal lag order (p) using the sequential modified likelihood ratio (*LR*) test statistic, the final prediction error (*FPE*) criteria, the Akaike Information Criteria (*AIC*) the Schwarz Information Criteria (*SIC*), and the Hannan-Quinn information criterion (*HQIC*), starting from *p*=1 to *p*=5. All lag-length selection criteria select one lag for our annual bivariate *VAR* model. Table 2 gives the results of the Johansen cointegration trace maximum eigenvalue test

statistics. We cannot reject the null hypothesis of no cointegration for the real house price and real GDP per capita series at 5-percent significance level.

Even though no cointegration exists between the real house price and real GDP per capita, these two series may still exhibit Granger temporal causality. That is, the real house price may Granger cause real GDP per capita, real GDP per capita may Granger cause the real house price, or the two series may exhibit two-way Granger causality. Table 3 reports the results of full sample Granger-causality tests. The first test is the F-test performed on the standard *VAR* model. The standard F-tests based on the *VAR*(1) estimates fail to reject the null hypothesis that real house price does not Granger cause real GDP per capita and that real GDP per capita does not Granger cause the real house price at 5-percent significance level. In sum, the results of the standard *VAR*(1) F-test indicate no predictive power running from either the real house price to real GDP per capita or from real GDP per capita to the real house price. In order to check the robustness of the F-test, we next perform bootstrap *LR* causality tests as reported in Table 3. The bootstrap *LR*-test uses the p-values obtained with 2,000 bootstrap replicates. The bootstrap *LR*-tests fails to reject the null hypotheses that the real house price does not Granger cause real GDP per capita and that real GDP per capita does not Granger cause the real house price at 5-percent significance level. Therefore, the bootstrap *LR*-test results indicate no causal links between the real house price and real GDP per capita, supporting the standard F-tests.

At the moment, we conclude based on the full sample of annual data from 1890 to 1952 that no long- or short-run relationships exist between the real house price and real GDP per capita. We now turn to examining the stability of the estimates. Structural changes may shift parameter values and the pattern of the (no) cointegration and (no) causal relationship may change over time. The results of the cointegration and Granger causality tests will show sensitivity to sample period used and order of the *VAR* model, if the parameters are

temporally instable. Therefore, studies using different sample periods and different *VAR* specifications will find conflicting results for the causal links between the real house price and real GDP per capita. The results of cointegration and Granger causality tests based on the full sample also become invalid with structural breaks because they assume parameter stability.

Full-Sample Parameter Stability

Researchers use various tests in practice to examine the temporal stability of econometric, and in our case *VAR*, models (e.g., Hansen, 1992b; Andrews, 1993; Andrews and Ploberger, 1994). Although we can apply these tests in a straightforward way for stationary models, the variables in our model are nonstationary and potentially cointegrated.[4] We need to accommodate the possibility of this integration (cointegration) property because in a cointegrated *VAR*, the variables form a vector error-correction (*VEC*) model. Thus, we need to investigate the stability of both the long-run cointegration and short-run dynamic adjustment parameters. If the long-run or cointegration parameters prove stable, then the model exhibits long-run stability. Additionally, if the short-run parameters are also stable, then the model exhibits full structural stability.

Since the estimators of cointegration parameters are superconsistent, we can split the parameter stability testing procedure into two steps. First, we test the stability of the cointegration parameters. Second, if long-run parameters prove stable, then we can test the stability of the short-run parameters. To examine the stability of cointegration parameters we use the L_c test of Nyblom (1989) and Hansen (1992a). This Nyblom-Hansen statistic tests for parameter constancy against the alternative hypothesis that the parameters follow a random

---

[4] Although the full-sample tests indicated no cointegration, we do not rule out the possibility of cointegration in our recursive and rolling analyses. That is, some sub-samples may suggest cointegration and other sub-samples may not.

walk process and, therefore, time-varying, since the first two moments of a random walk are time dependent. Next, we use the *Sup-F*, *Mean-F*, and *Exp-F* tests developed by Andrews (1993) and Andrews and Ploberger (1994) to investigate the stability of the short-run parameters. We compute these tests from the sequence of *LR* statistics that tests constant parameters against the alternative of a one-time structural change at each possible point of time in the full sample. These tests exhibit non-standard asymptotic distributions and Andrews (1993) and Andrews and Ploberger (1994) report the critical values. To avoid the use of asymptotic distributions, however, we calculate the critical values and p-values using the parametric bootstrap procedure.

We use the parameter constancy tests explained above to investigate the temporal stability of the coefficients of the *VAR* model formed by the real house price and real GDP per capita series. Table 4 reports the outcome of the tests. As just noted, these p-values come from a bootstrap approximation to the null distribution of the test statistics, constructed by means of Monte Carlo simulation using 2,000 samples generated from a *VAR* model with constant parameters. We calculate the L_c test for each equation separately using the *FM-OLS* estimator of Phillips and Hansen (1990). The *Sup-F*, *Mean-F*, and *Exp-F* tests require trimming at the ends of the sample. Following Andrews (1993), we trim 15 percent from both ends and calculate these tests for the fraction of the sample in [0.15, 0.85].

The results for L_c tests indicate that the real house price and real GPD per capita equations exhibit stable long-run parameters at the 5-percent significance level. That is, we find evidence of cointegration. In Table 4, we also report the system L_c statistics for the unrestricted *VAR*(1) model, which indicates that the *VAR* model as a whole proves unstable at the 1-percent level. This finding supports the view that the short-run parameters of the *VAR* system prove unstable.

The remaining three parameter constancy statistics also test for short-run parameter stability. The *Sup-F* statistics tests parameter constancy against a one-time sharp shift in parameters. On the other hand, the *Mean-F* and *Exp-F*, which assumes that parameters follow a martingale process, test for gradual shifting in the regime. Both the *Mean-F* and the *Exp-F* statistics test the overall constancy of the parameters (i.e., they investigate whether the underlying relationship among the variables stays stable over time). In addition, Andrews and Ploberger (1994) show that the *Mean-F* and *Exp-F* are both optimal tests. The results for the sequential *Sup-F*, *Mean-F*, and *Exp-F* tests reported in Table 4 suggest that significant evidence of parameter non-constancy exists in the real house price and real GDP per capita equations as well as the entire *VAR* system at the 1-percent level, except for the *Mean-F* test for the real house price equation at the 5-percent level.

In sum, the evidence obtained from the parameter stability tests indicate that the cointegrated *VAR* model does exhibit constant long-run parameters whereas the short-run dynamics of the model show parameter instability. The L_c, *Sup-F*, *Mean-F*, and *Exp-F* tests prove consistent in this regard.

As a set of alternative tests, we also estimated the cointegration equation between the real house price and real GDP per capita as follows:

$$RGDPC_t = \alpha + \beta \cdot RHP_t + \varepsilon_t, \tag{12}$$

where *RGDPPC* denotes real GDP per capita and *RHP* denotes the real house price. We estimate the parameters in equation (12) using the *FM-OLS* estimator. Table 5 reports the results of the various parameter stability tests. The Nyblom-Hansen L_c test cannot reject the null hypothesis of cointegration at any reasonable level. Similarly, the *Mean-F* and *Exp-F* tests cannot reject the null hypothesis of unchanging parameters in the cointegration equation. In other words, we do not find evidence of gradual shifting of the parameters of the

cointegration equation. Finally, the *Sup-F* test, however, suggests a one-time shift in the cointegration relationship.

Recursive and Rolling-Window Parameter Stability

Since the parameter constancy tests point to structural change, we estimate the *VAR* model using recursive and rolling window regression techniques. The recursive estimator starts with a benchmark sample period and then adds one observation at a time keeping all observations in prior samples so that the sample size grows by one with each iteration. The rolling-window estimator, also known as fixed-window estimator, alters the fixed length benchmark sample by moving sequentially from the beginning to the end of sample by adding one observation from the forward direction and dropping one from the end. Assume that each rolling subsample includes 15 annual observations (i.e., the window size is equal to 15). In each step for the recursive and moving window models, we determine a *VAR* model using the *LR*, *FPE*, *AIC*, *SIC*, and *HQIC* to chose the lag length and perform the Granger causality tests using *RB* bootstrap method on each subsample. This provides us with a sequence of 48 causality tests instead of just one. The recursive and rolling estimations that we adopt justified for a number of reasons. First, recursive and rolling estimations allow the relationship between the variables evolve through time. Second, the presence of structural changes introduces instability across different subsamples and recursive and rolling estimations conveniently capture this, in our case, by considering a sequence of 39 different subsamples (starting with the benchmark sample from 1890 to 1913). The rolling window uses a 15-year fixed window.

For the rolling estimations, the window size is an important choice parameter. Indeed, the window size controls the number of observations covered in each subsample and determines the number of rolling estimates, since a larger window size reduces the number of observations available for estimation. More importantly, the window size controls the precision and representativeness of the subsample estimates. A large window size increases

the precision of estimates, but may reduce the representativeness, particularly, in the presence of heterogeneity. On the contrary, a small window size will reduce heterogeneity and increase representativeness of parameters, but it may increase the standard error of estimates, which reduces accuracy. Therefore, the choice of the window size should achieve a balance of not too large or too small, thus, balancing the trade-off between accuracy and representativeness. We follow Koutris *et al.* (2008) and use a rolling window of small size (i.e., 15 annual observations) to guard against heterogeneity. Our choice of small window size may lead to imprecise estimates. Therefore, we apply the bootstrap technique to each subsample estimation to obtain more precise parameter estimates and tests.

No strict criterion exists for selecting the window size in rolling window estimation. Pesaran and Timmerman (2005) examine the window size under structural change in terms of root mean square error. They show that optimal window size depends on persistence and size of the break. Their Monte Carlo simulations shows that we can minimize the bias in autoregressive (AR) parameters with window sizes as low as 20 when frequent breaks exist. In determining the window size, we need to balance between two conflicting demands. First, the accuracy of parameter estimates depends on the degree of freedom and requires a larger window size for higher accuracy. Second, the presence of multiple regime shifts increases the probability of including some of these multiple shifts in the windowed sample. In order to reduce the risk of including multiple shifts in the subsamples, the window size should be small. Based on the simulation results in Pesaran and Timmerman (2005) we use a window size of 15 (this excludes the observations required for lags and hence is the actual number of observations in the *VAR*).[5]

---

[5] We also ran an analysis with a window size of 25. The qualitative results did not change, although some changes did occur in the quantitative findings. We report the findings for the window of 15 years.

Consider first the *VAR*(1) system. Table 4 reports the findings for the rolling window estimates in the next to last row. The *ME-L$_2$* test implies that both the real house price and real GDP per capita equations exhibit parameter instability at the 5-percent level, while the *ME-L$_2$* test for the *VAR* system also implies parameter instability at the 5-percent level. Figure 1a provides a sample-by-sample picture of the *ME-L$_2$* test statistic for the individual equation as well as the *VAR* system.[6] Based on the *ME-L$_2$* test statistic, we cannot reject the null hypothesis of stable parameters at the 5-percent level[7]. The Figure does indicate several periods of time when we can reject the null of parameter stability at the 5-percent level – 1899-1901, 1907-1908, 1925-1928, 1930, and 1934. The instability in the first two periods reflect proximately instability in the GDP equation, which occurs in 1899-1901 and 1907-08, while during the remaining periods, the instability reflects proximately instability of the RHP equation, which occurs in 1925, 1927-1928, 1930, and 1932-1934.[8]

Table 4 also reports the findings for the recursive estimates in the last row. Now, only the real GDP per capita equation shows evidence of parameter instability at the 1-percent level, whereas the real house price equation does not show evidence of parameter instability. Moreover, the *VAR* system also shows evidence of parameter instability at the 5-percent level. Figure 1b plots the *FL-L$_2$* test statistic for the individual equation as well as the *VAR* system. Based on the *FL-L$_2$* test statistic, we can reject the null hypothesis of stable parameters at the 5-percent level. Moreover, we also see on long-period of time when we reject the null hypothesis of parameter stability for the recursive subsamples – 1901-1944.

---

[6] Figure 1 only reports the significance level and mean L_2 norm test for the *VAR* system and not for the individual equations.

[7] We reject parameter stability for both individual equations and *VAR* system, when we use sup norm. This implies that we cannot reject a temporary, but somewhat persistent, deviation from the normal parameter levels, but we can reject it against a single-break alternative.

[8] The 5-percent critical values for the individual equations in the rolling and recursive specifications equal 2.2448 and 1.5444, respectively, which are not shown in the Figure.

The instability over this period always reflects instability in the GDP equation, which occurs from 1901-1944, while instability in the RHP equation only occurs from 1930-1938.

Finally, consider the long-run trend regression. Table 5 also reports the findings. The findings for the rolling and recursive window specifications paint different pictures. The *ME-L₂* test implies that long-run trend equation exhibits parameter stability for the rolling regression at the 5-percent level, but parameter instability for the recursive specification at the 1-percent level. For the rolling window regressions, Figure 1c plots the *ME-L₂* test statistic. The *ME-L₂* test statistic indicates that the parameters remain stable over the entire period. The statistics reported for each subsample over the entire period, however, suggest parameter instability at the beginning and end of the full sample – 1896-1900 and 1938-1944. For the recursive regressions, Figure 1d plots the *FL-L₂* test statistic. The *FL-L₂* test statistic indicates that the parameters do not remain stable over the entire period. In addition, the statistics reported suggest that parameter instability begins shortly after the beginning of the full sample and ends just before it ends, That is, the test suggests instability from 1895-1949.

While we find mixed evidence of parameter stability and instability across our full sample, certain patterns of stability and instability still exist. Generally, our findings support stability of the long-run parameters, but instability of the short-run parameters. When we examine, however, the stability tests for the individual rolling or recursive estimates across the full sample, we find evidence of instability in all cases for certain portions of the full sample. Any instability of parameters uncovered argues that the full-sample Granger causality tests prove unreliable. Thus, we turn now to an analysis of our rolling 15-year window estimates of Granger causality over the full-sample period from 1890 to 1952. These tests will give a better picture of the changing nature of Granger temporal causality over our sample period.

Rolling-Window Estimates

Since we want to consider how Granger temporal causality may alter as we move through the sample period 1890 to 1952, we propose to estimate the *VAR*(1) system on a rolling basis with a 15-year window.[9] In addition, we estimate the bootstrap p-value of observed *LR*-statistic rolling over the whole sample period 1898 to 1945[10] to further examine the likely temporal changes in the causality relationship. As stated above, we adopt the bootstrap approach with the Toda and Yamamoto (1995) modified causality tests because of several advantages. In particular, this test applies to both cointegrated and non-cointegrated *I(1)* variables (Hacker and Hatemi-J, 2006). We calculate the bootstrap p-values of the null hypotheses that the real house price does not Granger cause real GDP per capita and that real GDP per capita does not Granger cause the real house price using the *RB* method. More precisely, we compute the *RB* p-values of the modified *LR*-statistics that tests the absence of Granger causality from the real house price to real GDP per capita or vice-versa. We compute these from the *VAR*(1) defined in equation (2) fitted to rolling windows of 15 observations. For this reason, we only report the results with window size of 15.

We also compute the magnitude of the effect of the real house price on real GDP per capita and the effect of real GDP per capita on the real house price. We calculate the effect of the real house price on real GDP per capita as the mean of the all bootstrap estimates, that is,

$N_b^{-1}\sum_{k=1}^{p}\hat{\phi}_{hy,k}^{*}$, where N_b equals the number of bootstrap repetitions. Analogously, we

calculate the effect of real GDP per capita on the real house price as the mean of the all

---

[9] To do this, we estimate the *VAR* model in equation (1) for a time span of 15 years rolling through t = τ - 14, τ - 12, ..., τ, τ = 1905, ..., 1952. Since we estimate a *VAR*(1) system, we lose one observation at the beginning of the sample, which explains why the first 15-year sample runs from 1891 to 1905.

[10] Recall that our first 15-year sample period runs from 1890 to 1905. We report the findings for that sample at the mid-point of the 15 years from 1891 to 1905, or 1898. In other words, the point for 1898 reports the value for the 1891 to 1905 15-year window.

bootstrap estimates, that is $N_b^{-1} \sum_{k=1}^{p} \hat{\phi}_{yh,k}^*$. We calculate these results rolling through the whole sample with a fixed window size of 15 years. The estimates $\hat{\phi}_{hy,k}^*$ and $\hat{\phi}_{yh,k}^*$ are the bootstrap least squares estimates from the VAR in equation (2) estimated with the lag order of p determined by the BIC for each subsample. In our case, the number of bootstraps equals 2,000 and the number of lags equals one. We also calculate the 95-percent confidence intervals, where the lower and upper limits equal the 2.5[th] and 97.5[th] quantiles of each of $\hat{\phi}_{hy,k}^*$ and $\hat{\phi}_{yh,k}^*$, respectively.

Figure 2 plots the bootstrap p-values of the rolling test statistics, while Figure 3 plots the magnitude of the effects of each series on the other with the horizontal axes showing the mid-point observation in each of the 15-year rolling windows. For example, the value posted at year 1936 in the Figures represents the rolling window of 1929 to 1943. Figure 2 shows the bootstrap p-values of the rolling test statistics, testing the null hypotheses that the real house price does not Granger-cause real GDP per capita and vice versa. We will evaluate the non-causality tests at 5-percent significance level. Figure 3a shows the bootstrap estimates of sum of the rolling coefficients for the effect of the real house price on real GDP per capita, while Figure 3b shows the bootstrap estimates of sum of the rolling coefficients for the effect of real GDP per capita on the real house price.

Figure 2 shows that the p-values change substantially over the sample. In addition, we do not reject the null hypotheses that the real house price does not Granger-cause real GDP per capita and vice versa at the 5-percent significance level during most of the sample. We can reject the null hypothesis that the real house price does not Granger-cause real GDP per capita at the 5-percent significance level during 1925 and 1927-1934. We can also reject the null hypothesis that the real GDP per capita does not Granger-cause the real house price at

the 5-percent significance level only during 1928-1929. Figure 3a shows that the effect of the real house price on real GDP per capita proves significantly negative at the 5-percent level (two-tailed test) during 1911-1912 and significantly positive during 1925 and 1927-1934, while Figure 3b shows that the effect of the real GDP per capita on the real house price proves significantly negative at the 5-percent level (two-tailed test) during 1916-1917, 1919, 1922-1923, 1926-1930, 1932, and 1934-1936 and significantly positive during 1940-1945.

In sum, we find evidence of the real house price Granger causes real GDP per capita during the Great Depression, but not during other periods of our 1890 to 1952 sample of annual data. In other words, we continue to conclude that housing *is not* the business cycle.

4.   Conclusion

This paper considers the role of the real housing price in the Great Depression, examining structural stability between the real housing price and real GDP per capita. Using annual US data from 1890 to 1952, the paper examines the long-run and short-run dynamic relationships between the real housing price and real GDP per capita to see if these relationships change over time. More specifically, we adopt the bootstrap approach with the Toda and Yamamoto (1995) modified causality tests because of several advantages.

Overall, our tests suggest that the relationship between real GDP per capita and the real house price experienced structural shifts over the 62-observation sample from 1890 to 1952. Clear evidence suggests that the real house price Granger caused real GDP per capita only in rolling sub-samples that include a sufficient number of Great Depression years, but not other rolling sub-samples that exclude most of the Great Depression. That is, we find such evidence for 15-year subsamples that cover data from 1918 to 1941. At the same time, less evidence also exists that real GDP per capita Granger caused the real house price only during the Great Depression. Here, we find such evidence for 15-year subsample that cover data from 1921 to 1936. Furthermore, no evidence of Granger causality between real GDP

per capita and the real house price exists for any other periods in our full sample from 1890 to

1952.

References:

Ando, A., and Modigliani, F., 1963. The 'life-cycle' hypothesis of saving: Aggregate implication and tests. *American Economic Review* 53, 55-84.

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Andrews, D. W. K., and Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Aoki, K., Proudman, J., and Vlieghe, G., 2004. House prices, consumption, and monetary policy: A financial accelerator approach. *Journal of Financial Intermediation* 13, 414–435.

Bajari, P., Benkard, L., and Krainer, J., 2005. House prices and consumer welfare. *Journal of Urban Economics* 58, 474–487.

Benjamin, J. D., Chinloy, P., and Jud, G. D., 2004. Real estate versus financial wealth in consumption. *Journal of Real Estate Finance and Economics* 29, 341–354.

Bernanke, B., 1983. Nonmonetary effects of the financial crisis in the propagation of the Great Depression. *American Economic Review* 73, 257-276.

Bernanke, B. S., 2008, Housing, mortgage markets, and foreclosures. Speech, The Federal Reserve System Conference on Housing and Mortgage Markets, Washington, D.C. http://www.federalreserve.gov/newsevents/speech/bernanke20081204a.htm

Brown, R. L., Durbin, J., and Evans, J. M., 1975. Techniques for testing the tonstancy of regression relationships over time. *Journal of the Royal Statistical Society B* 37, 149–163.

Buiter, W. H., 2008. Housing wealth isn't wealth. London School of Economics and Political Science working paper.

Case, K. E., Quigley, J. M., and Shiller, R. J., 2005. Comparing wealth effects: the stock market versus the housing market. *Advances in Macroeconomics* 5, 1–34.

Chu, C. S. J., Hornik, K., and Kuan, C. M., 1995a. MOSUM Tests for parameter constancy. *Biometrika* 82, 603–617.

Chu, C. S. J., Hornik, K., and Kuan, C. M., 1995b. The moving-estimates test for parameter stability. *Econometric Theory* 11, 669–720.

Dolado, J. J., and Lütkepohl, H., 1996. Making Wald tests work for cointegrated VAR systems. *Econometrics Reviews* 15, 369-386.

Efron, B., 1979. Boot strap methods: Another look at the jackknife. *Annals of Statistics* 7, 1-26.

Fisher, I., 1933. The debt-deflation theory of Great Depressions. *Econometrica* 1, 337-357.

Friedman, M., 1957. *The Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.

Friedman, M., and Schwartz, A., 1963. *A Monetary History of the United States.* Princeton: Princeton University Press.

Grebler, L., Blank, D. M., and Winnick, L., 1956. *Capital Formation in Residential Real Estate: Trends and Prospects*. Princeton: NBER and Princeton University Press.

Gordon, R. J., and Wilcox, J. A., 1981. Monetarist interpretations of the Great Depression: An evaluation and critique. In: Brunner, K., (ed.). *The Great Depression Revisited.* Boston, MA: Martinus Nijhoff, 49-107.

Granger, C. W. J., 1996. Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics* 11, 455-473.

Hacker, R. S., and Hatemi-J, A., 2006. Tests for causality between integrated variables based on asymptotic and bootstrap distributions: theory and application. *Applied Economics* 38, 1489-1500.

Hansen, B. E. 1992a. Testing for parameter instability in linear models. *Journal of Policy Modeling* 14, 517–533.

Hansen, B. E., 1992b. Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* 10, 321-336.

Hansen, B. E., 2001. The new econometrics of structural change: dating breaks in U.S. labour productivity. *Journal of Economic Perspectives* 15, 117-128.

Hansen, B. E., and Phillips, P. C. B., 1990. Estimation and inference in models of cointegration. In Fomby, T.B., and Rhodes, G. F., (Eds.), *Advances in Econometrics* 8, Oxford: Elsevier, 225-48.

Horowitz, J. L., 1994. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics* 61, 395–411.

Iacoviello, M., 2010, Housing in DSGE models: Findings and new directions. In Bandt, O. de, Knetsch, T., Peñalosa, J., and Zollino, F., (Eds.), *Housing Markets in Europe: A Macroeconomic Perspective*, Berlin: Springer-Verlag, 3-16.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Koutris, A., Heracleous, M. S., and Spanos, A., 2008. Testing for nonstationarity using maximum entropy resampling: a misspecification testing perspective. *Econometric Reviews* 27, 363-384.

Kuan, C. M., and Chen, M. Y., 1994. Implementing the fluctuation and moving-estimates tests in dynamic econometric models. *Economics Letters* 44, 235–239.

Lustig, H., and Van Nieuwerburg, S., 2008. How much does household collateral constrain regional risk sharing? University of Chicago Working Paper.

MacKinnon, J. G., 1996. Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* 11, 601-618.

Mantalos, P., 2000. A graphical investigation of the size and power of the granger-causality tests in integrated-cointegrated VAR systems. *Studies in Non-Linear Dynamics and Econometrics* 4, 17-33.

Mantalos, P., and Shukur, G., 1998. Size and power of the error correction model cointegration test. A bootstrap approach. *Oxford Bulletin of Economics and Statistics* 60, 249-255.

Mishkin, F. S., 1976. Illiquidity, consumer durable expenditure, and monetary policy. *American Economic Review* 66, 642-654.

Mishkin, F. S., 1978. The household balance sheet and the Great Depression. *The Journal of Economic Histor*y 38, 918-937.

Nicholas, T., and Scherbina, A., 2012. Real estate prices during the roaring twenties and the Great Depression. *Real Estate Economics*, on-line at: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6229.2012.00346.x/abstract

Nyblom J., 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.

Park, J. P., and Phillips, P. C. B., 1989. Statistical inference in regression with integrated process: Part 2. *Econometric Theory* 5, 95-131.

Pesaran, M. H., and Timmermann, A., 2005. Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* 129, 183-217.

Phillips, P. C. B., 1987. Time series regression with a unit root. *Econometrica* 55, 277–302.

Phillips, P. C. B., and Hansen, B. E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economics Studies* 57, 99-125.

Phillips, P. C. B., and Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika* 75, 335-346.

Ploberger, W., and Kramer, W., 1992. The CUSUM test with OLS residuals. *Econometrica* 60, 271–285.

Ploberger, W., Krämer, W., and Kontrus, K., 1989. A new test for structural stability in the linear regression model. *Journal of Econometrics* 40, 307–318.

Shiller, R. J., 2005. *Irrational Exuberance,* 2nd Edition. Princeton University Press, Princeton: New Jersey.

Shukur, G., and Mantalos, P., 1997a. Size and power of the RESET test as applied to systems of equations: A boot strap approach. Working paper 1997:3, Department of Statistics, University of Lund, Sweden.

Shukur, G., and Mantalos, P., 1997b. Tests for Granger causality in integrated-cointegrated VAR systems. Working paper 1998:1, Department of Statistics, University of Lund, Sweden.

Shukur, G., and Mantalos, P., 2000. A simple investigation of the Granger-causality test in integrated-cointegrated VAR systems. *Journal of Applied Statistics* 27, 1021-1031.

Stock, J. H., and Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics, American Statistical Association* 14, 11-30.

Temin, P., 1976. *Did Monetary Forces Cause the Great Depression?* New York: W. W. Norton.

Toda, H. Y., and Phillips, P. C. B., 1993. Vector autoregressions and causality. *Econometrica* 61, 1367-1393.

Toda, H. Y., and Phillips, P. C. B., 1994. Vector autoregression and causality: A theoretical overview and simulation study. *Econometric Reviews* 13, 259-285.

Toda, H. Y., and Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66, 225-250.

Tukey, J. W., 1962. The future of data analysis. *The Annals of Mathematical Statistics* 33, 1–67.

White, E. N., 2009. Lessons from the great American real estate boom and bust of the 1920s. NBER Working Paper Series, Working Paper 15573. http://www.nber.org/papers/w15573

Table 1:   **Unit-Root Test Results**

Series	Level		First differences	
	Constant[a]	**Constant and Trend[b]**	**Constant[a]**	**Constant and Trend[b]**
Real House Price	-1.94	-1.73	-9.05[*]	-9.10[*]
Real GDP per Capita	-0.28	-2.05	-6.27[*]	-6.24[*]

Notes: Phillips Peron test statistics based on the Newey-West Bartlett kernel with bandwidth 3.
[a]   A constant is included in the test equation; one-sided test of the null hypothesis that a unit root exists; 1-, 5-, and 10-percent significance critical value equals -3.54, -2.91, and -2.59, respectively.
[b]   A constant and a linear trend are included in the test equation; one-sided test of the null hypothesis that a unit root exists; 1-, 5-, and 10-percent critical values equals -4.11, -3.48, and -3.17, respectively.
[*]   indicates significance at the 1-percent level.

Table 2:   **Multivariate Cointegration Test Results: Real House Price and Real GDP per Capita**

Series	Null Hypothesis	Alternative Hypothesis	Trace Test	Maximum Eigenvalue Test
Real House Price and	$r = 0$	$r > 0$	3.49	3.20
Real GDP per Capita	$r \leq 1$	$r > 1$	0.29	0.29

Notes: One-sided test of the null hypothesis that the variables are not cointegrated. The critical values for the trace and maximum eigenvalue tests come from Osterwald-Lenum (1992) and equal 5-percent critical value equals to 15.49 and 14.26, respectively, for testing $r = 0$ and 3.84 and 3.84, respectively, for testing $r \leq 1$.
[**]   indicates significance at the 5-percent level.

Table 3:   **Full-Sample Granger Causality Tests**

	H_0: Real House Price does not Granger cause Real GDP per Capita		H_0: Real GDP per Capita does not Granger cause Real House Price	
	Statistics	**p-value**	**Statistics**	**p-value**
Standard *VAR*(1) *LR*-Test	0.028	0.866	0.287	0.594
Bootstrap *LR* Test	0.027	0.887	0.268	0.711

Notes: * and *** indicate significance at the 10 and 1 percent levels, respectively.

Table 4: **Parameter Stability Tests in *VAR*(1) Model**

	Real House Price Equation		Real GDP per Capita Equation		*VAR*(1) System	
	Statistics	Bootstrap *p*-value	Statistics	Bootstrap *p*-value	Statistics	Bootstrap *p*-value
Mean-F	7.01	0.03	20.93	<0.01	16.30	<0.01
Exp-F	12.06	<0.01	24.36	<0.01	12.11	<0.01
Sup-F	31.51	<0.01	55.58	<0.01	31.32	<0.01
L_c	0.12	0.86	0.70	0.18	3.88	0.01
Rolling L_2 norm	1.23	0.32	1.16	0.34	2.39	0.22
Recursive L_2 norm	0.68	0.20	3.20	0.01	3.94	<0.01

Notes: We calculate *p*-values using 2,000 bootstrap repetitions.

Table 5: **Parameter Stability Tests in Long-Run Relationship *FM-OLS***

	Mean-F	Exp-F	Sup-F	L_c	Rolling L_2 norm	Recursive L_2 norm
RGDPPC = α + β*RHP	129.76	133.64	274.02	0.17	1.34	5.14
Bootstrap *p* value	1.00	1.00	<0.01	0.70	0.24	0.01

Notes: We calculate *p*-value using 2,000 bootstrap repetitions.

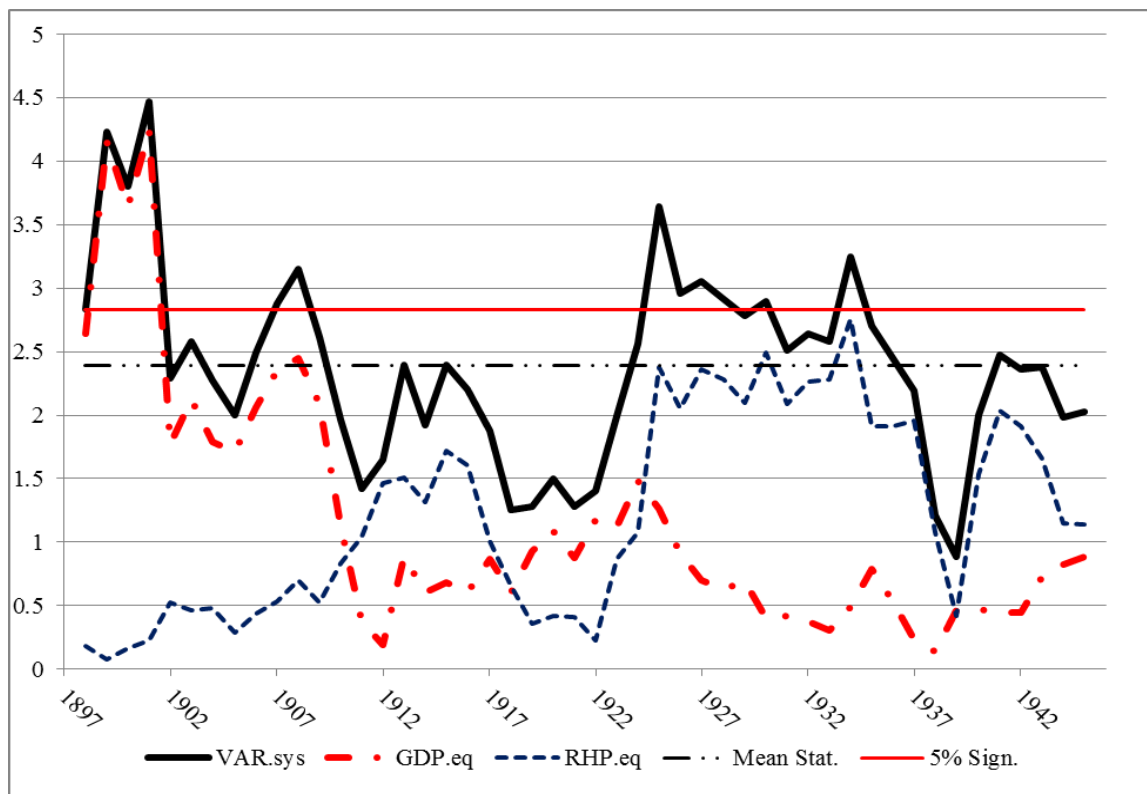**Figure 1a: Rolling VAR Stability Test with Mean L_2 Norm**

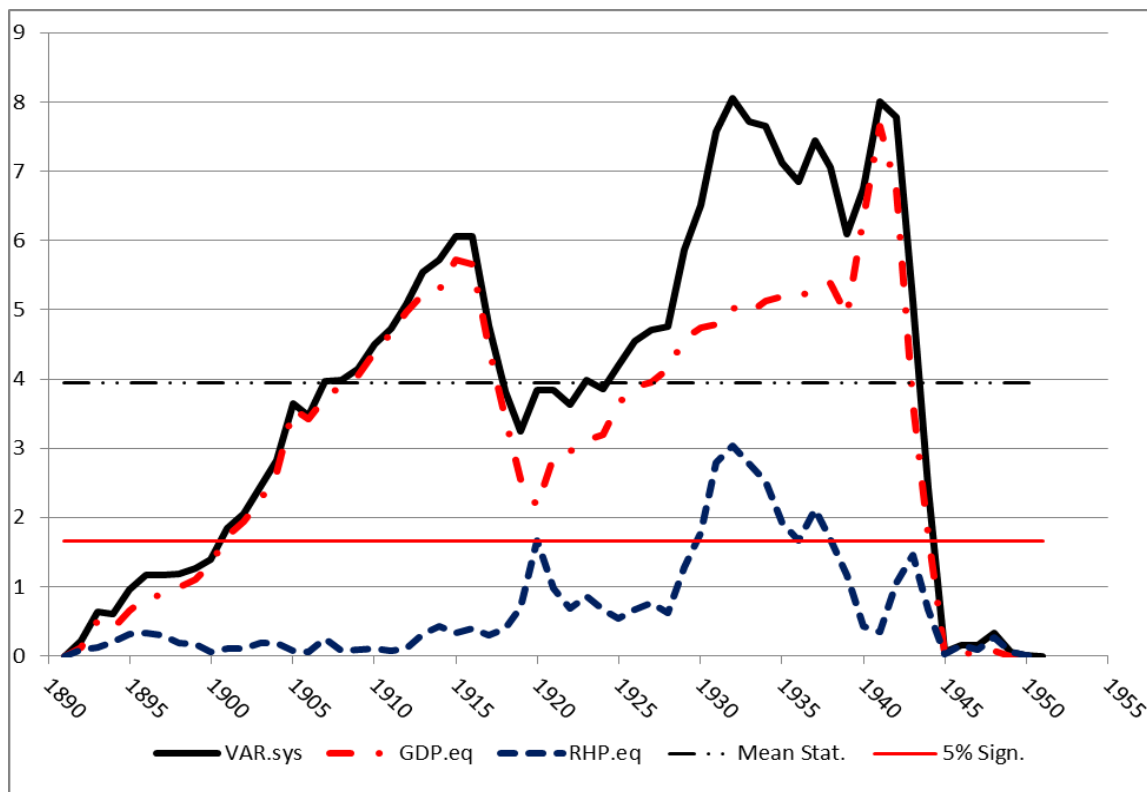**Figure 1b: Recursive VAR Stability Test with Mean L_2 Norm**

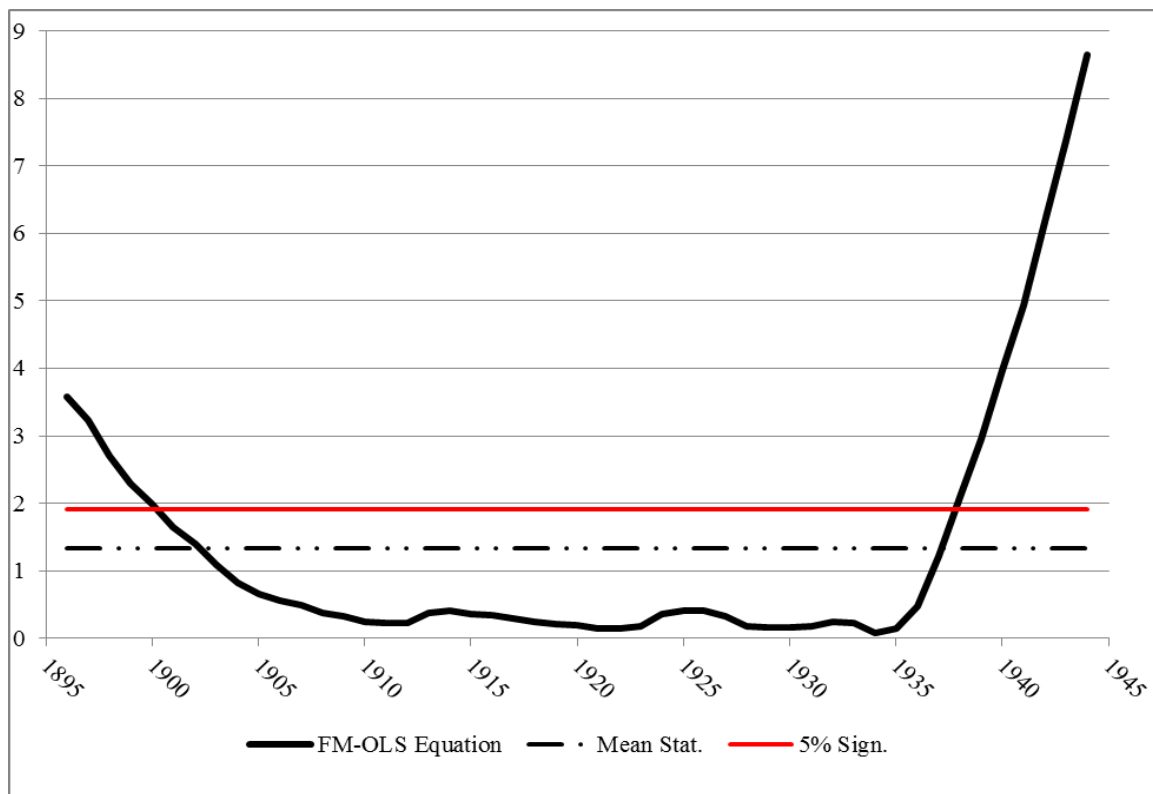**Figure 1c: Rolling Long-Run *FM-OLS* Stability Test with Mean *L₂* Norm**

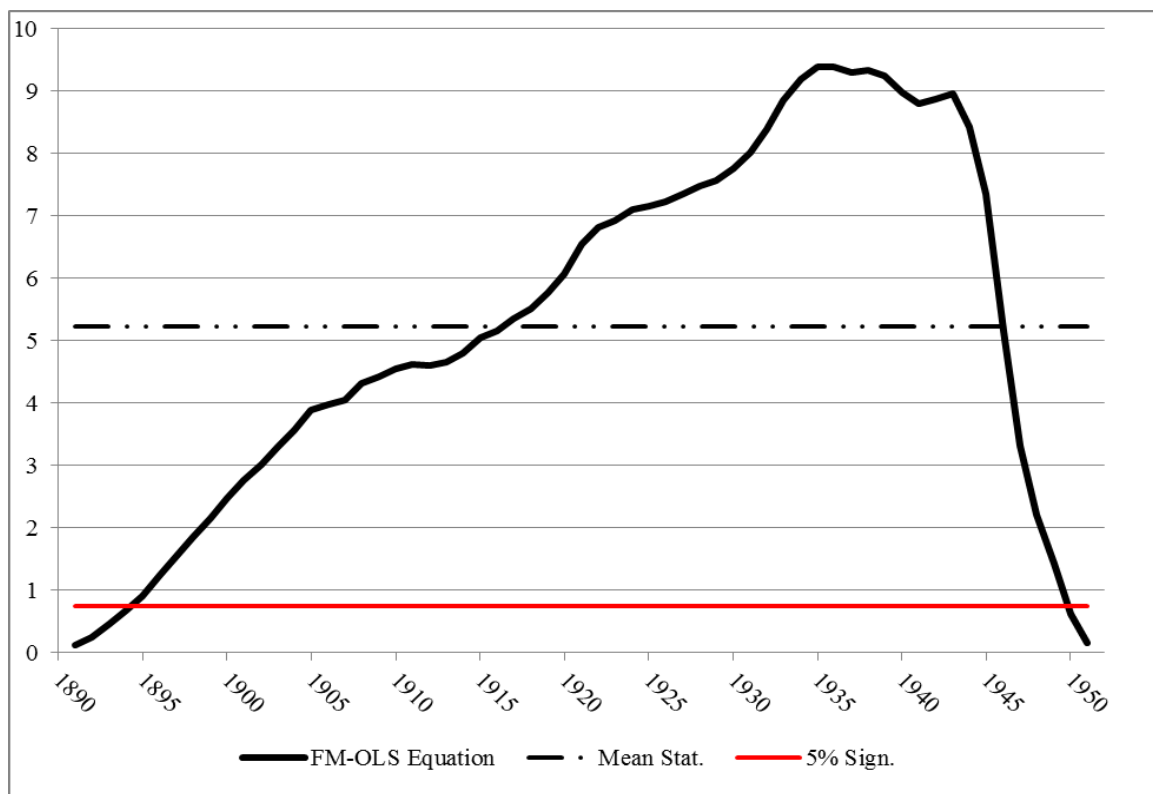**Figure 1d: Recursive Long-Run *FM-OLS* Stability Test with Mean *L₂* Norm**

Figure 2:   Granger Causality Test p-Values: Rolling Window Estimates
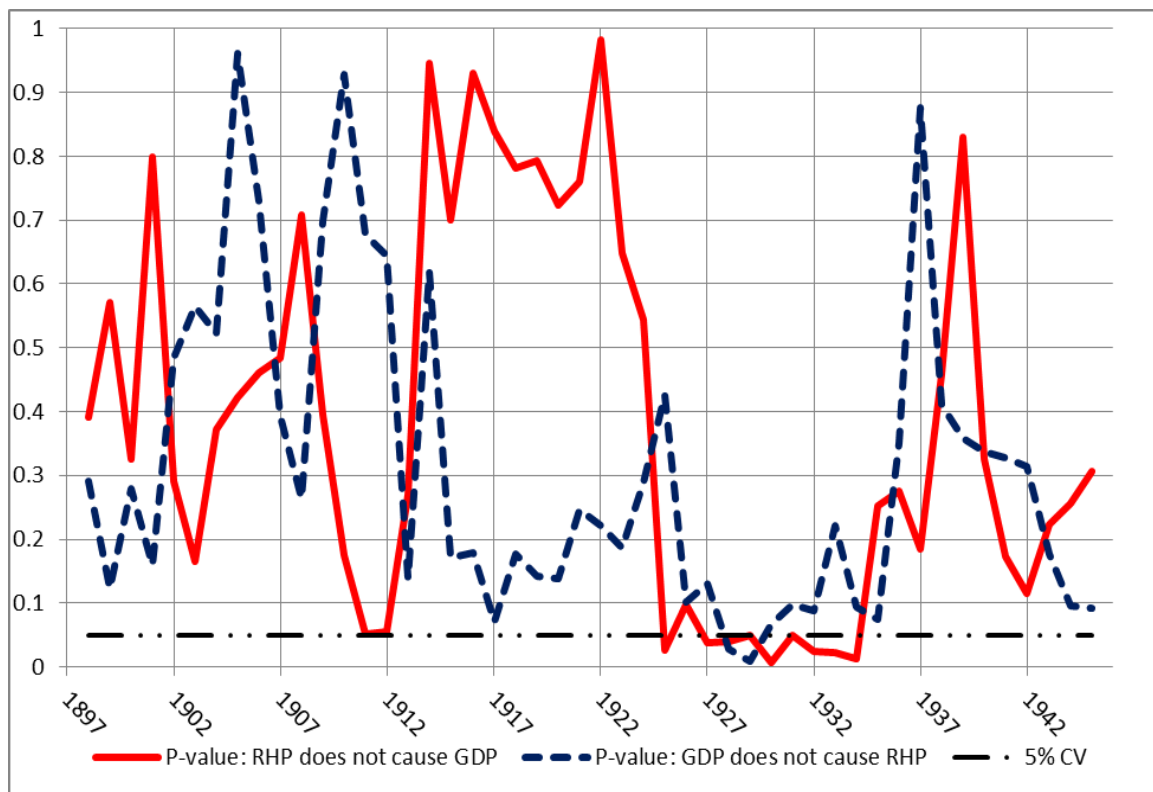
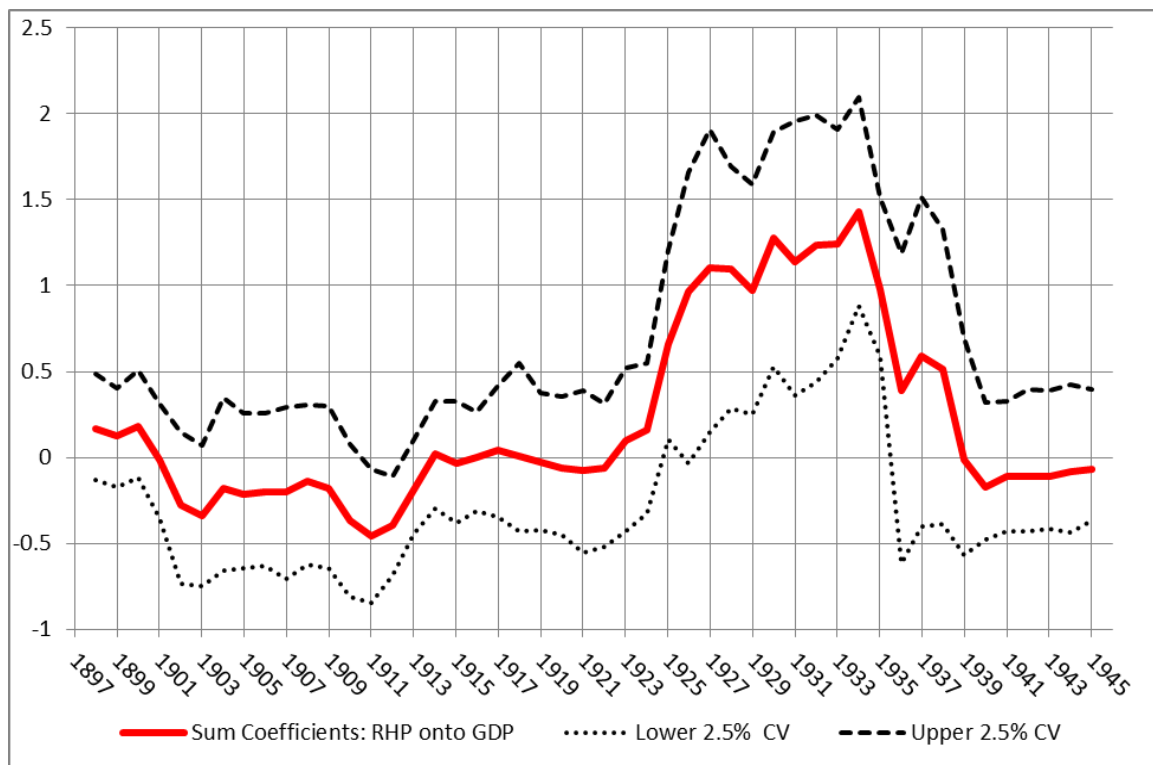**Figure 3a: Granger Causality: Sum Coefficients, RHP Causes RGDPPC**

Figure 3a: Granger Causality: Sum Coefficients, RGDPPC Causes RHP