Structural Breaks and Regime Switching Models: Theoretical Extensions and Applications

Bruce Chang-Ming Wang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2007

Program Authorized to Offer Degree:
Department of Economics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Bruce Chang-Ming Wang

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

____
Yu-chin Chen

Reading Committee:

____
Yu-chin Chen

____
Douglas Collins

____
Louis Garrison

____
Eric Zivot

Date: 4/20/2007

University of Washington

Abstract

Structural Breaks and Regime Switching Models:
Theoretical Extensions and Applications

Bruce Chang-Ming Wang

Chair of the Supervisory Committee:
Professor Yu-chin Chen
Department of Economics

In 3 essays, regime switching and structural break models are explored and used in the fields of International Economics, Health Economics, and Macroeconomics. 1) Characterizing the Real Exchange Rate in a Switching AR(1) and Unit Root Model best represents its behavior in our long horizon data for 16 countries, which raises questions regarding the common practices of utilizing single process long horizon regressions, panel analysis, and structural breaks. 2) Using the SEER-Medicare database, the burden of illness of colorectal cancer patients from 1991-2002 is shown to have a break in the average first-year cost of treatment coinciding with the FDA approval date of irinotecan, a chemotherapy agent. 3) Simulations exploring the finite sample properties of endogenous structural breakpoint tests show that the performances of the nonlinear and linear forms are identical, but bootstrapped critical values should be used in small samples. Using the NAIRU as an example, the finite sample dangers of these tests are apparent.

TABLE OF CONTENTS

Page

List of Figures ii

List of Tables iii

Chapter I: The Search for Stationarity in Real Exchange Rates: An Unobserved Component Regime Switching Approach (with Yu-chin Chen) 1

Chapter II: Trends in First-Year Costs and Survival for Colorectal Cancer Patients: Empirical Testing for Structural Breakpoints (with Louis Garrison) 46

Chapter III: Endogenous Structural Breakpoint Tests: An Investigation in Finite Sample Performance (with Jun Ma) 74

Bibliography 98

LIST OF FIGURES

Figure Number Page

1.1 Estimated Probabilities of Permanent Regime (Part I) 20
1.2 Estimated Probabilities of Permanent Regime (Part II) 21
1.3 Estimated Probabilities of Permanent Regime (US/UK only) 30
1.4 Time-Varying Transition Probabilities (PPr) 40
1.5 Time-Varying Transition Probabilities (QPr) 40
2.1 Map of SEER Registries 51
2.2 Average First-year Cost 54
2.3 First-year Survival Probability 59
2.4 Survival Curves Based on Clinical Studies 61
3.1 Flow Chart for Bootstrap Monte Carlo 81

LIST OF TABLES

Table Number Page

1.1 Program simulations: Switching AR(1) and Unit Root Model................11
1.2 Switching AR(1) and Unit Root Model..13
1.3 AR Coefficients from Switching AR(1) and Unit Root Model 14
1.4 Program simulations: Switching 2 Unit Root Model..............................17
1.5 Program simulations: Single UR Process run on 2 UR Model 18
1.6 Switching 2 Unit Roots Specification...19
1.7 Program simulations: Quasi-PPP (Switching AR(1) and UR Model).....25
1.8 Program simulations: Quasi-PPP (2 Unit Root Specification)................26
1.9 Parameters used to generate simulated data..27
1.10 Output from AR(1) with constant regression for generated data............28
1.11 Unit Root test results on generated data 28
1.12 Single processes run on Switching AR(1) and UR Model 29
1.13 OLS Regressions on Fundamentals 35
1.14 Fundamentals in Measurement Equation (US/UK) 38
1.15 Time-varying Transition Probabilities (US/UK) 41
2.1 Colorectal Cancer Drug Approval Dates 48
2.2 Observations per Stage 52
2.3 Breaks in Average First-year Cost for CRC patient 56
2.4 Use of Irinotecan by Initiation Date 57
2.5 Use of Irinotecan by Entrance into SEER Database Date 58
2.6 OLS Regressions with Time Trend..63
2.7 Breaks in Average First-year Cost with Time Trend..............................64
2.8 Breaks in Average First-year Cost (No Irinotecan Patients) 66
2.9 Breaks in Average First-year Departmental Cost for CRC patient..........68
2.10 Breaks in Average First-year Payments for CRC patient 70
3.1 Test of size for Sample Size=10 85
3.2 Test of size for Sample Size=25 85
3.3 Test of size for Sample Size=50 85
3.4 Test of power for Sample Size=10 and BE Ratio=0.25...........................87
3.5 Test of power for Sample Size=10 and BE Ratio=1................................87
3.6 Test of power for Sample Size=10 and BE Ratio=4................................87
3.7 Test of power for Sample Size=25 and BE Ratio=0.25...........................88
3.8 Test of power for Sample Size=25 and BE Ratio=1................................88
3.9 Test of power for Sample Size=25 and BE Ratio=4................................88
3.10 Test of power for Sample Size=50 and BE Ratio=0.25...........................89
3.11 Test of power for Sample Size=50 and BE Ratio=1................................89
3.12 Test of power for Sample Size=50 and BE Ratio=4................................89

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to the Department of Economics for giving me the opportunity to pursue this degree. This dissertation would never have been completed without Yu-chin Chen, Louis Garrison, Eric Zivot, Douglas Collins, and Jun Ma. I would also like to thank Charles Nelson, Chang-Jin Kim, Richard Startz, Chung-Ming Kuan, Srikanth Kadiyala, Yunmi Kim, and participants at the Western Economic Association, Asia-Pacific Economic Association, All China Economics Association, and University of Washington conferences and seminars for their useful comments. The data could not have been obtained without the help of David Papell, Ruxandra Prodan, Scott Ramsey, and the Fred Hutchinson Cancer Research Center. Lastly, I am truly thankful for the encouragement and support of my parents and dog, Happy.

DEDICATION

To my parents.

Chapter I

The Search for Stationarity in Real Exchange Rates: An Unobserved Component Regime Switching Approach

(with Yu-chin Chen)

1. Introduction

The PPP Puzzle asks whether the large and frequent short-term shocks to the Real Exchange Rate (RER) and the persistence of these shocks can be supported by theory. Empirical evidence suggests the half-life of RER shocks to be 3 to 5 years, but traditional theory predicts a quicker reversion to the process due to goods market arbitrage (Rogoff, 1996). The inconsistency between theory and evidence leads us to believe that the RER best be described by a regime switching model. There may be 2 distinct states: real shocks could prove to have permanent, lasting effects whereas monetary and financial shocks could tend to constitute volatile yet transitory disturbances. Our results indicate regime shifts in line with historical events and strong evidence of both types of states for many country pairs.

This regime switching approach allows us to overcome many of the present problems in current approaches in the literature. One of the main issues in testing for PPP is the power of the Unit Root tests in small samples.[1] Post-Bretton Woods data is often used in PPP analysis to have a sample without nominal exchange rate regime shifts, but it contains less than 40 years of data.[2] To circumvent this problem, researchers frequently choose one of two types of approaches to increase sample size: use long-horizon data series or pool the data for use in a panel framework. Some examples of long-horizon data include Lothian and Taylor (1996), Rogoff (1996), and Taylor (2000). Not only do their efforts fail to reduce the half-life of shocks down to the desired 1-2 years that would be supported by theory through price-stickiness, but the use of long-horizon data assumes a constant underlying data generating process, which is unlikely because of the numerous happenings during the span of the sample. By allowing for multiple states, our regime switching model can use the large sample advantages of long-horizon data while relaxing the constant process assumption. We show that series of RERs are characterized by shifts in the persistence of

[1] Engel (1999) and Murray and Papell (2000) explore the power of tests in small samples. Murray and Papell (2005a and 2005b) and Amara and Papell (2004) propose alternative estimation methods.

[2] Many papers explore the issue of sample selection. Grilli and Kaminsky (1989) deal with the historical background of RER through a long set of data and conclude that the RER volatility depends on its historical setting and not on the nominal regime. But, they concede that the post-Bretton Woods period exhibits very high volatility. Diebold, Husted, and Rush (1991) choose to use data from the Gold Standard because those regimes represent the greatest amount of international cooperation, which is necessary for PPP to hold. Frankel and Rose (1996) use only post-World War II data in their panel study because the data exhibits a clear shift before and after the war. Even the choice of countries has an effect on the result of the studies. Cheung and Lai (1998) claim that developed countries are less likely to exhibit stationarity than their developing counterparts.

shocks to the processes, which are not properly picked up in analyses of long horizon data using a single series.

Our model also exposes flaws in the popular method of getting around the problem of multiple regimes in a long horizon dataset by breaking up the data into separate periods. Using arbitrary sample selection methods or running structural breakpoint tests would allow the researcher to keep from mixing regimes. In Taylor (2000), he breaks up his dataset into 4 periods of history: Gold Standard, Interwar, Bretton Woods, and Float. Diebold, Husted, and Rush (1991) use the gold standard periods in their analysis to allow for the greatest amount of cooperation between countries. A problem with specifying break dates *a priori* is that if they are off by a few periods, the estimations might not fully characterize the true, underlying processes. In the regime switching framework, our model endogenously selects the dates for shifting regimes, which takes away the potential for human bias in the analysis.

Using endogenous structural breakpoint tests to find distinct regimes in long horizon data can also allow the model to select the breakdates directly, but we show the tests to be biased in the presence of highly persistent data. For example, Hedgwood and Papell (1998) reduce the half-lives of many RER series by allowing the process to shift whenever a new breakpoint is encountered. They coin their result "Quasi-PPP." We show that the regime switching approach can correctly identify Quasi-PPP, but Quasi-PPP cannot identify our regime switching specification with 2 distinct states.

Our regime switching model also has implications on the alternative way of increasing the sample size to increase the power of Unit Root tests by pooling data across countries. Frankel and Rose (1996) use 150 countries from the International Financial Statistics database to find stationarity with a half-life longer than the acceptable 1-2 years. Alba and Papell (2007) use Feasible GLS (SUR) in their analysis and conclude that one cannot characterize all countries as exhibiting stationary or nonstationary shocks.[3] The results from our estimations caution the use of this technique because there are clear differences between countries; pooling the data would assume a certain degree of homogeneity amongst the countries, which may be unrealistic in the case of RERs.

The proposed model, Switching AR(1) and Unit Root Model, characterizes the RER as a stationary process with occasional permanent shocks.[4] Unlike other regime switching approaches, our model maintains parsimony while allowing for flexibility in its characterization of the RER.[5] The stationary process governs the more common monetary shocks, and the nonstationary process

---

[3] See Wu (1996), Canzoneri, Camby, and Diba (1996), and Papell (2006) for other examples of PPP analysis using panel frameworks.

[4] The model is similar to the Innovation Regime Switching (1; 1, 0) model of Kuan, Huang, and Tsay (2005) used to model Real GDP.

[5] Other Regime Switching models in the RER context include Engel and Hamilton (1990), Engel and Kim (1999), Bergman and Hansson (2000), and Frömmel, MacDonald, and Menkhoff (2002). See Hegwood and Papell (1998), Diebold, Husted, and Rush (1991), Cheung (1993), Cheung and Lai (1993), and Papell and Prodan (2006) for other methods.

accounts for the less frequent real shocks. The empirical results strongly support our method of representing the RER as these 2 states. The model is robust to variations and encompasses the findings of previous trials using structural breakpoints and univariate Unit Root Tests. Perhaps most importantly, we show that the regimes characterized by our model are not arbitrary but are closely related to historical events, such as wars and nominal currency regime changes.

To further explain shifts between regimes, we tie fundamentals—such as GDP/Capita Differences, Commodity Price Levels and Volatilities, and Trade Openness—into our models. The fundamentals play a definitive role in explaining the RER process, but their effect depends on the country pair in question, which can be explained by different countries having different policies and dependencies on commodities. Furthermore, even though distinct states show up for each RER series, the absolute levels of the parameter estimates vary. These findings support our conclusion that panel analysis may be too restrictive.

Though our methodology does not reduce the half life of the stationary process of every country pair, there are a handful of countries that consistently show quick reversion during the stationary periods. For other countries, the shocks in the stationary periods remain highly persistent, which is consistent with the long half-life findings in the current literature and also with recent theories that suggest traditional theories of PPP reversion to be incomplete.[6]

The following section presents our models and estimation methodology. Section 3 covers the results and discussion for the Switching AR(1) and Unit Root Model and the 2 Unit Root Model; we also include robustness checks for the latter. In Section 4, we directly model fundamentals to explain the behavior of RERs as described by our model. Finally, Section 5 concludes and offers extensions to our project.

2A. Models

The Switching AR(1) and Unit Root Model allows for both stationary and nonstationary components because shocks could affect RERs in different ways depending on their inherent nature. For example, real shocks may prove to have permanent effects on RERs whereas monetary shocks are merely transitory disturbances. This suggests that the model must allow for only one type of shock

---

[6] See Benigno (2004) on how monetary policy rules may influence persistence of RER shocks. Other reasons for the persistent volatility include MacDonald and Ricci (2002), who argue that the size and competitiveness of the distribution sector of an economy impacts the price adjustment mechanisms of its tradables sector. Obstfeld and Rogoff (2000) and Imbs, Mumtaz, and Ravn (2002) show that transaction costs could impact the price levels of the individual markets as well. Imbs, Mumtaz, Ravn, and Rey (2005a) argues that an aggregation bias in the data yields larger half-lives than the individual sectors would produce. This is questioned in Chen and Engel (2005) but later reiterated in Imbs, Mumtaz, Ravn, and Rey (2005b).

each period if we wish to make this distinction. A reduction of the half-life of transitory shocks may even lessen the purchasing power parity puzzle.

In the Switching AR(1) and Unit Root Model, the series is composed of a permanent process and a stationary process, but shocks only affect one process at any given time. For our annual data, a half-life of 1 to 2 years would coincide with an AR coefficient of between 0.5 and 0.7.

Switching AR(1) and Unit Root Model:

$$y_t = x_t + z_t \tag{1}$$

$$x_t = x_{t-1} + S_t v_t$$
$$z_t = \varphi\, z_{t-1} + (1-S_t)\, e_t$$
$$v_t \sim N(0, \sigma_v^2),\ e_t \sim N(0, \sigma_e^2)$$

For the Switching AR(1) and Unit Root Model, the RER, y_t, is characterized by a Unit Root nonstationary process, x_t, and an AR(1) process with a coefficient of φ. The shocks v_t and e_t, for the Unit Root process and AR(1) process respectively, are distributed normally with mean zero and variances σ_v^2 and σ_e^2. S_t is a state parameter that takes the value of 0 or 1.

The unique feature is that the state variable, S_t, determines the allocation of the shock at time t. If $S_t = 0$ for all t in the Switching AR(1) and Unit Root Model, only the stationary shock, e_t, enters y_t. In other words, for the Switching AR(1) and Unit Root Model, the process becomes $y_t = x_0 + \varphi\, z_{t-1} + e_t$ where e_t is the transitory shock. This is merely a stationary AR process with a level shift, x_0. For the opposite case where $S_t = 1$, only the permanent process comes into effect and yields $y_t = \varphi z_0 + x_{t-1} + v_t$, where v_t is the permanent shock. Here, we have a Unit Root plus a constant, φz_0.

Unlike many other models that incorporate both permanent and transitory components for RERs, the model above suggests that the shocks are mutually exclusive. Other models, such as Bergman and Hansson (2000) are built upon the single process AR(1) model and only have switching in the intercept and coefficient, φ. The authors cannot attribute the regime shifts to historical events and, thus, fail to interpret the different states afforded by the model. We allow 2 distinct processes that can be interpreted as real and monetary shocks, which coincide with historical events. A similar model that also has switching in variances is presented by Engel and Kim (1999). There are 2 processes, 1 permanent and 1 transitory, which are always on and each have 3 possible variance states. In total, there are potentially 6 different variances. The robustness checks for our model show that any additional processes are superfluous and merely complicate the explanation of the model. So, in our model, there will either be a permanent shock *or* a transitory shock but not both in the same period. Though this may seem restrictive, it keeps the model parsimonious and allows for a clean interpretation of the empirical results. Note that it is entirely possible to express the above model in a general form of an ARMA model with state-dependent coefficients.

2B. Estimation Methodology

To estimate the Switching AR(1) and Unit Root Model, we employ the classical estimation technique for regime switching models. Using the algorithms provided in Kim and Nelson (1998), we first put the models in state-space form:

$$y_t = H_{S_t}\beta_t + A_{S_t}z_t + e_t \tag{2}$$

$$\beta_t = \tilde{\mu}_{S_t} + F_{S_t}\beta_{t-1} + G_{S_t}v_t$$

$$\begin{pmatrix} e_t \\ v_t \end{pmatrix} \sim N\left(0, \begin{matrix} R_{S_t} & 0 \\ 0 & Q^{*}_{S_t} \end{matrix}\right)$$

Equation (2) presents an N x 1 observed time-series, y_t, as a function of a J x 1 unobserved series, β_t, and a K x 1 series of weakly exogenous or lagged dependent variables, z_t. β_t is a function of the shock, v_t, which is of dimension L x 1. The dimensions for the remaining variables are as follows: H_{S_t} is N x J, A_{S_t} is N x K, F_{S_t} is J x J, and G_{S_t} is J x L. If the variable is governed by an unobserved Markov-switching state variable, it has the subscript S_t.

Then, we estimate the parameters of interest by numerically maximizing the likelihood functions constructed by their algorithms.

For the Switching AR(1) and Unit Root Model, its state-space representation is as follows:

$$y_t = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ z_t \end{pmatrix} \tag{3}$$

(Measurement Equation of the form $y_t = H\beta_t$)

$$\begin{pmatrix} x_t \\ z_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix} \begin{pmatrix} x_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} S_t & 0 \\ 0 & (1-S_t) \end{pmatrix} \begin{pmatrix} v_t \\ \varepsilon_t \end{pmatrix}$$

(Transition Equation of the form $\beta_t = F\beta_{t-1} + G_{S_t} v_t$)

Note that $R_{S_t} = 0$ and $Q^*_{S_t} = \begin{pmatrix} \sigma^2_{v_t} & 0 \\ 0 & \sigma^2_{\varepsilon_t} \end{pmatrix}$

The RER is constructed as $q_t = s_t + p_t - p^*_t$, where s_t, p_t, and p^*_t are the logarithms of the nominal exchange rate (foreign price of the US Dollar), domestic and foreign price levels, respectively. The data was obtained from Taylor (2002) and updated through 2004 (if available) using the IFS database and include the following countries in our analysis: Australia, Belgium, Canada, Denmark, Germany, Finland, France, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. We use annual data in our main estimations and quarterly data for a robustness check of the US/UK RER.

3. Results and Discussion

In the Switching AR(1) and Unit Root Model, the shocks are either permanent or transitory each period—there cannot be both types. This allows for the interpretation of the existence of real shocks *or* monetary shocks in each period. The model stems from the trend-cycle literature for economic output such as the work of Kuan, Huang, and Tsay (2005) in which they applied a similar characterization for GDP.

Table 1.1 Program simulations: Switching AR(1) and Unit Root Model

	Pr(UR|UR)	***Pr(St|St)***	φ	$\sigma_{UR}2$	$\sigma_{ST}2$
True Value	0.7	0.8	0.55	0.4	0.8
Estimated (N=150)	0.8420 (0.0816)	0.8799 (0.0605)	0.6606 (0.1026)	0.3663 (0.0492)	0.8065 (0.0827)
Estimated (N=3000)	0.6808 (0.0377)	0.7940 (0.0303)	0.5079 (0.0231)	0.3931 (0.0141)	0.8066 (0.0182)

Table 1.1 shows the validation of the Gauss program itself. Note that it depicts single instances of simulated data series in order to show the potential bias in a single, small sample series such as what we have for the actual RER data.

The probability of a Unit Root state this period given that the previous period was a Unit Root state is Pr(UR|UR). Likewise, the probability of a stationary state this period given that the previous period was stationary is Pr(St|St). σ_{UR}^2 and σ_{ST}^2 are the variances of the Unit Root process and of the stationary process respectively. The simulation indicates that the program correctly estimates the true parameters even in our small sample (N=150) example. As expected, the estimates become closer to the true parameter values as the sample size increase to 3000.

Table 1.2 shows the above program's estimations for the log RER data series for our 16 countries with the US Dollar as the base currency. Since our data frequency is annual, an estimate of 0.5 to 0.7 for the AR coefficient, $\boldsymbol{\varphi}$, would fall in the range of a 1-2 year half-life. As the results indicate, the half-lives varying tremendously based on the country pair in question. In particular, Portugal, Finland, and Belgium now have a transitory process with a half-life that falls within the acceptable range dictated by theory. France, with an AR coefficient of 0.7675, has a half-life of only 2.6 years.

Table 1.3 shows the AR coefficients from our model for all possible country pairs.[7] The large probabilities for remaining in each state demonstrate that both processes come into effect during our long horizon data series. Half-life estimates below 3 years are shown in bold. We see that a few of countries—Belgium, Denmark, Portugal, Spain, and Switzerland—frequently have low half-lives after allowing for the occasional permanent process in our regime switching model. This finding leads us to caution both the use of long horizon and panel

[7] Note that exchange rates are two-sided, so the results are the same when running US/UK and UK/US, so only one of the possible pairs are recorded in the table.

Table 1.2 Switching AR(1) and Unit Root Model. Standard Errors are in parentheses. US Dollar as base currency.

	Pr(UR|UR)	*Pr(St|St)*	φ	σ_{UR}^2	σ_{ST}^2	**LLH Value**
Australia	0.8881 (0.0562)	0.8375 (0.1028)	0.8440 (0.0681)	0.0358 (0.0079)	0.1288 (0.0209)	164.0230
Belgium	0.9751 (0.0150)	0.6997 (0.1324)	0.5277 (0.1255)	0.0806 (0.0077)	0.5832 (0.1273)	80.4131
Canada	0.9472 (0.0479)	0.9224 (0.0660)	0.8074 (0.1333)	0.0314 (0.0046)	0.0608 (0.0078)	228.9060
Denmark	0.9834 (0.0138)	0.9673 (0.0202)	0.9267 (0.0457)	0.0350 (0.0038)	0.1340 (0.0113)	125.8168
Germany	0.9398 (0.0299)	0.9075 (0.0493)	0.9580 (0.0190)	0.0270 (0.0024)	0.1280 (0.0145)	169.5601
Spain	0.9493 (0.0314)	0.8634 (0.0667)	0.9997 (0.0065)	0.0530 (0.0087)	0.1592 (0.0194)	95.4154
Finland	0.9125 (0.0353)	0.7205 (0.1141)	0.6673 (0.1390)	0.0587 (0.0072)	0.3224 (0.0508)	85.5425
France	0.8251 (0.0948)	0.8292 (0.0787)	0.7675 (0.1903)	0.0434 (0.0092)	0.1084 (0.0119)	125.6131
Italy	0.9846 (0.0122)	0.9099 (0.0470)	0.9985 (0.0010)	0.0265 (0.0025)	0.2211 (0.0223)	109.3920
Japan	0.9876 (0.0101)	0.8856 (0.0601)	0.9919 (0.0031)	0.1130 (0.0088)	0.0297 (0.0047)	104.8256
Netherlands	0.9379 (0.0308)	0.9483 (0.0320)	0.9442 (0.0257)	0.0275 (0.0026)	0.1241 (0.0117)	166.6399
Norway	0.9505 (0.0275)	0.8880 (0.0606)	0.8208 (0.0445)	0.0327 (0.0028)	0.1633 (0.0195)	163.0848
Portugal	0.9722 (0.0170)	0.6299 (0.1511)	0.5519 (0.0655)	0.0753 (0.0061)	0.3473 (0.0781)	77.8230
Sweden	0.9373 (0.0326)	0.8877 (0.0480)	0.8510 (0.0410)	0.0263 (0.0029)	0.1379 (0.0139)	154.2002
Switzerland	0.8495 (0.0597)	0.8663 (0.0622)	0.9610 (0.0142)	0.0227 (0.0031)	0.1477 (0.0150)	127.9609
United Kingdom	0.9175 (0.0417)	0.8900 (0.0655)	0.9406 (0.0351)	0.0247 (0.0026)	0.1227 (0.0135)	194.3264

Table 1.3 AR Coefficients from Switching AR(1) and Unit Root Model for All Country Pairs

	RAUS	RBEL	RCAN	RDEN	RDEU	RESP	RFIN	RFRA	RITA	RJAP	RNET	RNOR	RPRT	RSWE	RSWI	RUK	RUS
RAUS	X																
RBEL	1.00	X															
RCAN	0.81	1.00	X														
RDEN	1.00	1.00	1.00	X													
RDEU	1.00	**0.78**	1.00	0.88	X												
RESP	1.00	**0.44**	1.00	**0.65**	0.95	X											
RFIN	0.91	0.86	0.82	**0.77**	0.91	0.49	X										
RFRA	0.81	**0.53**	**0.79**	0.99	0.99	1.00	0.87	X									
RITA	0.86	**0.80**	0.99	1.00	1.00	1.00	1.00	1.00	X								
RJAP	0.99	**0.79**	0.99	0.99	0.99	0.99	0.99	0.99	0.88	X							
RNET	**0.61**	**0.69**	1.00	**0.71**	0.85	**0.75**	0.99	1.00	1.00	1.00	X						
RNOR	1.00	**0.46**	0.99	0.82	0.91	**0.44**	1.00	0.92	0.83	**0.53**	0.84	X					
RPRT	0.91	**0.32**	**0.49**	**0.64**	**0.45**	**0.23**	**0.38**	**0.50**	**0.62**	**0.61**	**0.41**	**0.43**	X				
RSWE	0.85	**0.39**	0.92	0.96	**0.88**	**0.60**	0.94	0.97	1.00	1.00	0.97	1.00	**0.46**	X			
RSWI	0.89	1.00	**0.63**	0.82	0.81	**0.79**	**0.77**	**0.66**	0.97	0.99	**0.79**	0.81	**0.46**	0.86	X		
RUK	0.98	1.00	1.00	**0.49**	1.00	1.00	0.85	1.00	0.99	0.99	0.91	1.00	0.86	1.00	1.00	X	
RUS	0.84	**0.53**	0.81	0.93	0.96	1.00	**0.67**	**0.77**	1.00	0.99	0.94	0.82	**0.55**	0.85	0.96	0.94	X

data. When working with long horizon data on a country pair that may have shifts in its process, the estimated coefficients may not reflect the true underlying process. If the RER follows a model such as ours, then the single process models become inaccurate. In a subsequent section, we present a robustness check showing our model can correctly identify a single process, but our regime switching model can be mistaken for a Unit Root process when regime shifts are not permitted.

When including countries such as Belgium in a panel framework, a researcher is then pooling together a series of RERs with dissimilar characteristics and regime shifts, which could make the estimates biased. The countries we use are in the OECD and often used in pool analysis, so it is necessary to carefully study the inclusion criteria into a panel data set. For example, Alba and Papell (2007), Murray and Papell (2004), Frankel and Rose (1996), and Wu (1996), include Belgium, Portugal, and Spain in their panel analysis. Canzoneri, Cumby, and Diba (1999) use Belgium and Spain. These countries have much quicker mean reversions when allowing for the distinct regime switching processes, so we believe using them in a panel framework would be mixing regimes and give misleading results.

On the other hand, many of the countries have results like the US/UK RER whose AR coefficient is 0.9406, which shows stationary shocks as dissipating very slowly. In fact, the highly persistent stationary process could be "mistaken" for being a Unit Root itself. For these country pairs, the Switching AR(1) and Unit Root Model can be approximated by 2 Unit Roots switching back and forth, which is when the AR coefficient, φ, is set to be 1 in our original model:

$$y_t = x_t + z_t \tag{4}$$

$$x_t = x_{t-1} + S_t v_t$$
$$z_t = z_{t-1} + (1 - S_t)\varepsilon_t$$

$$\Rightarrow y_t = x_{t-1} + z_{t-1} + S_t v_t + (1 - S_t)\varepsilon_t$$

$$= y_{t-1} + S_t v_t + (1 - S_t)\varepsilon_t$$

$$v_t \sim N(0, \sigma_v^2) \text{ and } \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

This Switching 2 Unit Root specification holds for the highly persistent country pairs because it is difficult to distinguish between highly persistent transitory shocks and "quiet" permanent shocks in small samples.[8] Given that the results from the Switching AR(1) and Unit Root Model indicate the shocks to the stationary process last a long time, the characterization imposed by the 2 Unit Root specification is a reasonable statistical approximation for those country pairs with highly persistent stationary processes.[9]

[8] See Hamilton (1994), page 444, for a detailed explanation of this equivalence.
[9] The new model can be interpreted as having "quiet" and "noisy" periods.

Table 1.4 Program simulations—Switching 2 Unit Root Model

	Pr(LO|LO)	*Pr(HI|HI)*	σ_{LO}^2	σ_{HI}^2
True Value	0.7	0.8	0.4	0.8
Estimated (N=150)	0.6366 (0.3352)	0.4309 (0.2802)	0.3165 (0.1300)	0.6860 (0.1016)
Estimated (N=3000)	0.7732 (0.0515)	0.8210 (0.0536)	0.3858 (0.0219)	0.6784 (0.0296)

Simulations show that the program can correctly identify the different variances. Table 1.4 shows the estimated parameters of a generated series for which we defined the parameter values. Pr(LO|LO) (Pr(HI|HI)) is the probability that the current state is quiet given the previous state was also quiet. Likewise, Pr(HI|HI) is the probability that the current state is noisy given the previous state was also noisy. σ_{LO}^2 and σ_{HI}^2 are the variances of the quiet state and noisy states respectively. Again, this is just a single simulation of a small sample size and of a larger sample size. The reason behind this is that we wish to show the potential variation in estimating a single, small-sample series. Monte Carlo simulations show that as the simulations increase, the average parameter estimates become increasing close to the true parameters. Nevertheless, from the above tables, we see that, for the small sample size, the true variance of the quiet process, 0.4000, is estimated to be 0.3165 with a standard error of 0.1300 by the program. As the sample size increases to 3000, the estimated parameter is 0.3858 with a standard error of 0.0219, which is not significantly different at the 95% confidence interval from the true parameter of 0.4.

Next, we check to make sure the program does not falsely break the processes into quiet and noisy states. We generate a single unit root process with normally distributed error terms to see how the program reacts. The results are in Table 1.5. The program is attempting to characterize the single series into 2 series with different variances. Regardless of the sample size, the large standard errors surrounding the probabilities of staying in their respective states indicate it does not know in which state the process resides. The variances are all around 1,

Table 1.5 Program simulations—Single Unit Root Process run on Switching 2 Unit Root Model

	Pr(LO|LO)	*Pr(HI|HI)*	σ_{LO}^2	σ_{HI}^2
Estimated (N=150)	0.6505 (4.2793)	0.6184 (1.8384)	1.1172 (0.1422)	1.1173 (0.1328)
Estimated (N=3000)	0.6514 (0.3781)	0.6171 (0.3792)	0.1011 (0.0545)	0.1011 (0.0182)

which is the true variance. So, given that the 2 identified processes are identical, it makes sense that the program does not know in which state it belongs.

Table 1.6 shows the Switching 2 Unit Root program run for our 16 countries and that this new characterization is indeed a good approximation for many country pairs. We see that the different variances are indeed present. For

Table 1.6 Switching 2 Unit Roots Model. Notes: Standard Errors are in parentheses. RERs are in terms of USD.

	Pr(LO|LO)	*Pr(HI|HI)*	σ_{LO}^2	σ_{HI}^2	**LLH Value**
Australia	0.8911 (0.0570)	0.8966 (0.0463)	0.0311 (0.0043)	0.1227 (0.0131)	164.0491
Belgium	0.7660 (0.1703)	0.9916 (0.0091)	0.0951 (0.0070)	0.8250 (0.2683)	80.0729
Canada	0.8748 (0.0665)	0.7150 (0.1178)	0.0136 (0.0033)	0.0551 (0.0049)	231.6840
Denmark	0.9811 (0.0154)	0.9401 (0.0339)	0.0360 (0.0039)	0.1360 (0.0114)	121.6769
Germany	0.9102 (0.0465)	0.9406 (0.0309)	0.0271 (0.0028)	0.1257 (0.0142)	166.8682
Spain	0.8565 (0.0681)	0.9650 (0.0231)	0.0540 (0.0073)	0.1604 (0.0790)	87.9914
Finland	0.7297 (0.1071)	0.9005 (0.0451)	0.0561 (0.0113)	0.3302 (0.0565)	81.8098
France	0.8733 (0.0541)	0.7301 (0.1141)	0.0217 (0.0042)	0.1053 (0.0093)	130.1611
Italy	0.6247 (0.2397)	0.9671 (0.0191)	0.0829 (0.0060)	2.4071 (0.6779)	80.7979
Japan	0.1053 (3.9427)	0.9914 (0.0088)	0.1000 (0.0065)	6.2768 (4.4975)	96.0614
Netherlands	0.9519 (0.0292)	0.9416 (0.0289)	0.0291 (0.0027)	0.1264 (0.0120)	164.1804
Norway	0.8946 (0.0531)	0.9550 (0.0270)	0.0349 (0.0033)	0.1564 (0.0166)	154.3472
Portugal	0.7786 (0.1624)	0.9894 (0.0146)	0.0894 (0.0089)	0.3603 (0.1000)	73.4821
Sweden	0.9039 (0.0447)	0.9500 (0.0281)	0.0291 (0.0037)	0.1409 (0.0144)	147.4977
Switzerland	0.9112 (0.0500)	0.8803 (0.0512)	0.0258 (0.0034)	0.1443 (0.0146)	125.2769
United Kingdom	0.8770 (0.0702)	0.9098 (0.0427)	0.0252 (0.0027)	0.1258 (0.0138)	191.4932

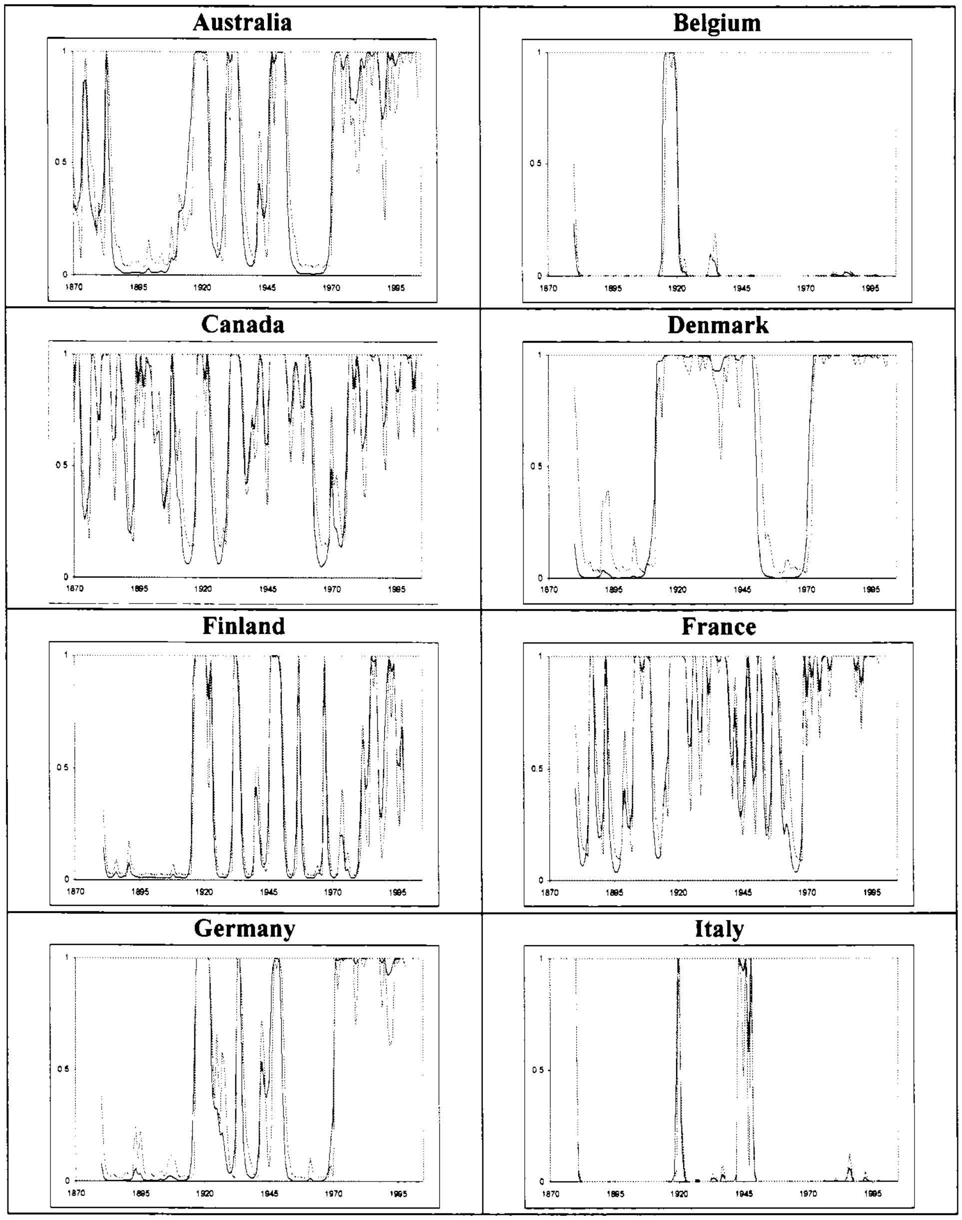

Figure 1.1 Estimated Probabilities of High Variance Regime for Switching AR(1) and Unit Root Model. The solid lines are the smoothed estimates and the dotted lines are the filtered estimates.

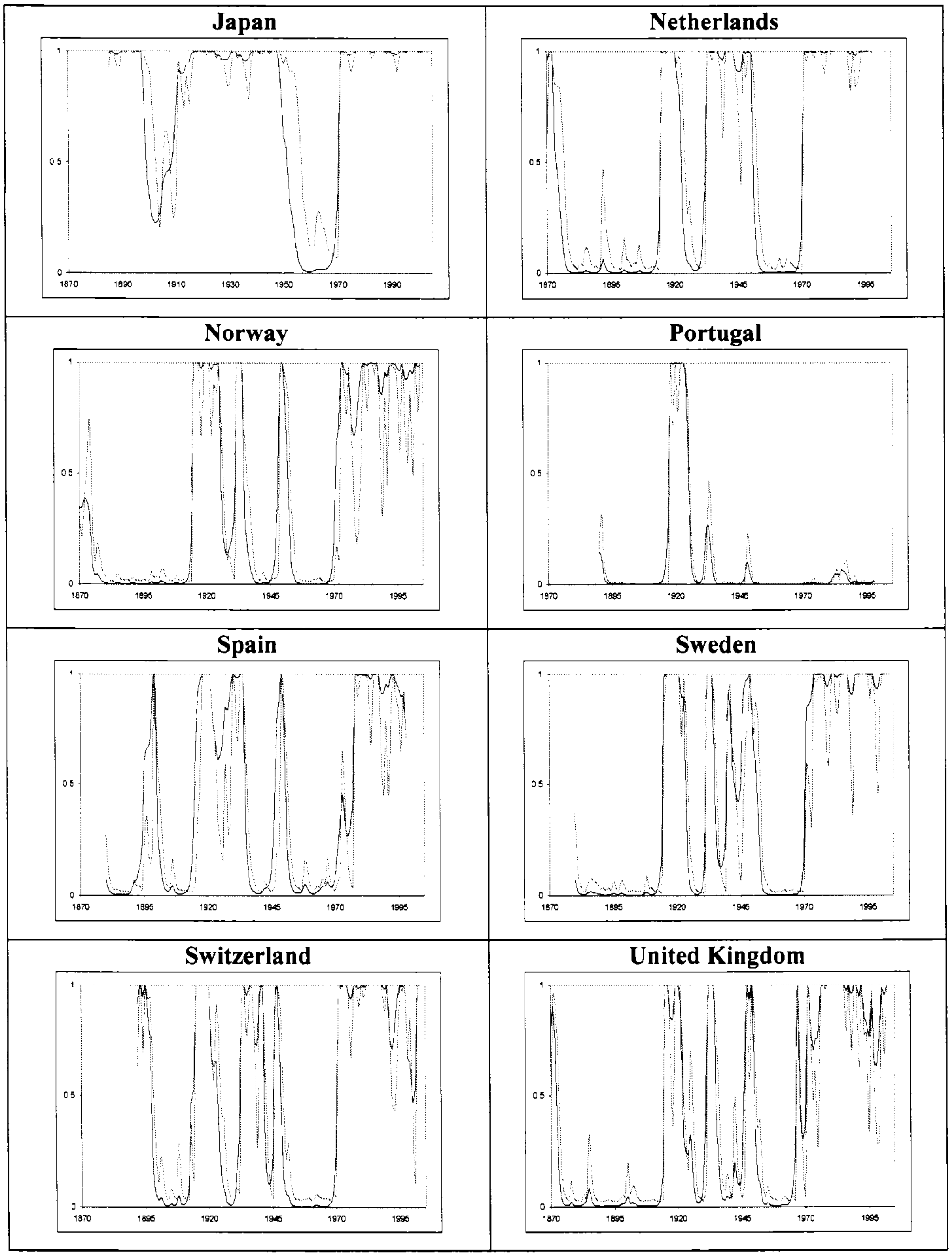

Figure 1.2 Estimated Probabilities of High Variance Regime for Switching AR(1) and Unit Root Model. The solid lines are the smoothed estimates and the dotted lines are the filtered estimates.

the US/UK RER, the quiet and noisy periods have variances of 0.0252 and 0.1258 respectively, which is identical to the Unit Root and stationary variances from the Switching AR(1) and Unit Root Model.[10]

Returning to our original Switching AR(1) and Unit Root Model, the plots in Figure 1.1 show the probability of being in a transitory state (vertical axis) during a given year (horizontal axis). For most of the countries, the plots show shocks are transitory most of the time and only becomes permanent on a few occasions. For example, Italy is shown to have transitory shocks with a couple of exceptions such as during the period around 1945 when it was involved in World War II. In the next section, we will show in more detail how the spikes indicating noisy periods coincide directly with historical events.

3. Robustness Checks

We run a series of robustness checks. For instance, we add an I(2) process to simulate a double-drift, but those results do not differ much from our current model. The addition of time trends to these models does not reduce the half-lives, which is consistent with the findings in the current literature. The model is also robust for data of other frequencies too; using Post-Bretton Woods monthly data,

[10] For the highly persistent countries, this characterization of the RER as having quiet and noisy permanent shocks is consistent with the results from the current literature showing very persistent shocks to the RER. The shocks during the quiet periods are small, permanent deviations from the mean. This suggests that there is no arbitrage possibility unless the deviations from the mean exceed a certain threshold. Transportation costs, transaction costs, and aggregation bias could explain the lack of reversion within a certain limit.

we observe the distinct regimes in our RERs. The unifying result in all of these variations implies a model exhibiting a (sometimes highly persistent) transitory process and a Unit Root process: the Switching AR(1) and Unit Root Model. The robustness checks against structural breaks and Unit Root tests are presented below. Then, we show how the regime shifts coincide with actual events by using the US/UK transitions as an example.

3A. Robustness Check: Quasi-PPP

In Hedgwood and Papell (1998), the authors find a short half-life for the shocks on the RER and call their result "Quasi-PPP." They use endogenous structural breakpoint tests on series of RERs and then run simple AR(1) regressions on the data while allowing for structural shifts for the dates indicated by the breakpoint tests. Their resulting AR coefficient is low enough to fall into the 1-2 year range for PPP to hold in the short-run. If their result holds true, the persistence exhibited by the Switching AR(1) and Unit Root Model for some country pairs is not consistent with their findings. In this exercise, we attempt to replicate their results given a data generating process from the 2 Unit Root specification. [11] Our method is as follows:

[11] If the transitory process is not persistent, then the switching AR(1) and Unit Root Model and the Quasi-PPP results are identical because the latter will always reduce the half-life down to an "acceptable" level. So, we use the extreme case of highly persistent shocks, which exist quite frequently as indicated in our earlier results (See Tables 1.2 and 1.3).

1) Generate the 2 Unit Root Model data with persistent states and different variances

2) Confirm high ϕ in a simple AR1 model:

$$q_t = \phi q_{t-1} + v_t, \; v_t \sim N(0, \sigma_v^2) \tag{5}$$

3) Run Bai and Perron (1998) programs to find multiple breakpoints

4) Estimate AR(1) model allowing for different levels for each regime:

$$q_t = \phi q_{t-1} + D_1 t_1 + \ldots + D_n t_n + v_t, \; v_t \sim N(0, \sigma_v^2), \tag{6}$$

where D_i is a dummy variable that can take the value of 0 or 1 and t_i is the level shift for the period without breaks.

In our generated series of 135 observations, which is the same number we have for US/UK RERs, we find a high coefficient (0.9983) in the simple autoregressive model. Then, the multiple breakpoint test program finds 3 breaks that would coincide with the years 1881, 1901, and 1970.[12] This is somewhat disturbing because we know our true data generating process is merely 2 Unit Roots switching back and forth and not a process with 3 breaks.[13] Nevertheless, we proceed and run the AR(1) regression allowing for level shifts in our 4 regimes.

[12] We use the tests from Bai and Perron (1998). See http://people.bu.edu/perron/code.html for code.

[13] Nunes, Kuan, and Newbold (1995) use simulations to show that unit root processes can generate "spurious breaks." Bai (1998) provides a note in which he presents a mathematical proof for this. Another issue for testing structural breaks in RERs is the low power in the Bai and Perron (1998) tests; this issue is addressed in Prodan (2005).

Using Eviews 5.1, our regression yields a φ of 0.63, which falls into the range of 1-2 years for the half-life on a shock to the system.

We have shown that the 2 Unit Root specification can be mistaken as Quasi-PPP, but can Quasi-PPP be mistaken as the 2 Unit Root specification? If the models are all equivalent in small samples, it would be impossible to identify the "true" model. We generate a series of data following the Quasi-PPP method. We assume 1) 3 breaks, 2) 4 levels that are 0.7, -0.5, 0.6, and -1 in that respective order, 3) AR(1) coefficient is 0.6, and 4) errors are *iid* (0,1). Then, we generate data for sample sizes of 150 and 1000 for both the Switching AR(1) and Unit Root Model and the 2 Unit Root specification. The break dates occur on observations 30, 60, 90 for the small sample and on 100, 500, and 750 for the large sample.

Table 1.7 Program simulations: Quasi-PPP (Switching AR(1) and Unit Root Model)

	Pr(UR|UR)	*Pr(St|St)*	φ	σ_{UR}^2	σ_{ST}^2
N = 150	0.8625 (0.1343)	0.9595 (0.0279)	0.6053 (0.1184)	1.4928 (0.2744)	0.8975 (0.0859)
N = 1000	0.0001 0.0000	0.9853 (0.0073)	0.5984 (0.0299)	1.4062 (0.3251)	0.9798 (0.0215)

As Table 1.7 indicates, the autoregressive coefficients are estimated to be close to 0.6, which is the true parameter from the data generating process. As the sample size, N, increases, the unit root process fades into the background, which is shown by the estimate of Pr(UR|UR) as 0.0001.

In Table 1.8, the program for the 2 Unit Root specification only identifies a single Unit Root process for both the small sample and the large sample. Furthermore, the variances are not very different. The results we get from using RER data for these models are different from what we just simulated above. Our RER data yield both low and high AR coefficients, 2 processes, and distinct variances. It does not appear that a Quasi-PPP data generating process can correctly replicate the empirical results brought forth by our models.[14]

Table 1.8 Program simulations: Quasi-PPP (Switching 2 Unit Root Model)

	P(UR1|UR1)	*P(UR2|UR2)*	σ_{UR1}^2	σ_{UR2}^2
N = 150	1.0000	0.0208	1.1955	1.2239
	(0.0043)	0.0000	(0.0700)	(10.1152)
N = 1000	0.5671	0.0001	1.0126	1.3354
	(0.1441)	0.0000	0.0000	(0.0805)

3B. Robustness Check: Unit Root Tests

Next, we test the robustness of the Switching AR(1) and Unit Root Model against standard Unit Root tests. Many studies using Long Horizon data, such as

[14] This phenomenon can also be interpreted as spurious structural breaks due to heteroskedasticity. On the flipside, structural breaks are not mistakenly attributed to be heteroskedastic processes.

Taylor (2002), run series of Unit Root tests to determine the stationarity of RERs. If a Unit Root is present, then the shocks are permanent and PPP does not hold. We first show that 2 commonly used Unit Root Tests (Augmented Dickey-Fuller and Dickey-Fuller GLS) cannot account for the switching processes in our model and, hence, incorrectly conclude that the process is actually a Unit Root series.[15] Then, we demonstrate the robustness of our model in that it can correctly characterize single process series.

The parameter assumptions used to generate the simulated data follow the estimates from the Belgium/US RER, which exhibits quick reversion in the transitory process. See Table 1.9 for the parameter assumptions.[16]

Table 1.9 Parameters used to generate simulated data

	Pr(UR|UR)	*Pr(St|St)*	φ	σ_{UR}^2	σ_{ST}^2
True Parameter	0.95	0.7	0.5	0.08	0.55

We double-check that a single process AR(1) with constant regression would yield persistent shocks under this data generating process. The output from a least squares regression done in Eviews 5.1 is in Table 1.10.

[15] Taylor (2002) shows the evidence for stationarity to be inconclusive in univariate settings.
[16] Sample size is 135.

Table 1.10 Output from AR(1) with constant regression for generated data

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	-0.049155	0.029627	-1.659108	0.0995
Lagged Value	0.965463	0.020188	47.82455	0.0000

The large coefficient of the lagged variable, the AR(1) parameter, is indicative of persistence and is consistent with the findings in the current literature using long horizon data. Next, we run the same series of data through 2 separate Unit Root tests: Augmented Dickey-Fuller and Dickey-Fuller GLS. The null hypothesis of the tests is that the series has a Unit Root.

Table 1.11 Unit Root test results on generated data

		ADF	DF-GLS
Test statistic		-1.7108	-0.8986
Test critical values:	1% level	-3.4797	-2.5823
	5% level	-2.8831	-1.9432
	10% level	-2.5783	-1.6151

In Table 1.11, we see that neither Unit Root test can reject the null hypothesis. So, one may incorrectly conclude that our generated series is a Unit Root process.

Next, we show that a single process series is correctly picked up by our Switching 2 Unit Root Model. We generate data from the following series:

$y_t = \varphi\, y_{t-1} + e_t$, where e_t is *iid* (0,1) shock.[17]

[17] Sample size is 135.

Table 1.12 Single processes run on Switching AR(1) and UR Model

	Pr(UR|UR)	*Pr(St|St)*	φ	σ_{UR}^2	σ_{ST}^2
True φ = 0.5	0.0013	1.0000	0.5631	0.0003	0.0720
	(0.1615)	(0.0001)	(0.0717)	(0.0570)	(0.0044)
True φ = 1	0.9681	0.1044	0.5715	0.0827	0.0094
	(0.0145)	(0.0040)	(0.0026)	(0.0057)	(0.0302)

In the first trial, the AR coefficient is set to equal 0.5 in order to produce a half-life of 1 year and in the second trial, it is merely a Unit Root process. When φ = 0.5, only the stationary process exists because the probability of entering a Unit Root state is effectively zero. Then when the series is truly a Unit Root, φ = 1, the Switching AR(1) and Unit Root Model is again correct in showing the Unit Root process to be dominant. So, if the process was indeed simply a single series without any regime shifts, then our model would characterize it as so. However, the results run on real data (see Table 1.2) clearly indicate that this is not the case and there are 2 distinct processes that govern the RER. This robustness check is evidence that the use of long horizon data without accounting for regime shifts may produce spurious results.

3C. Historical Implications

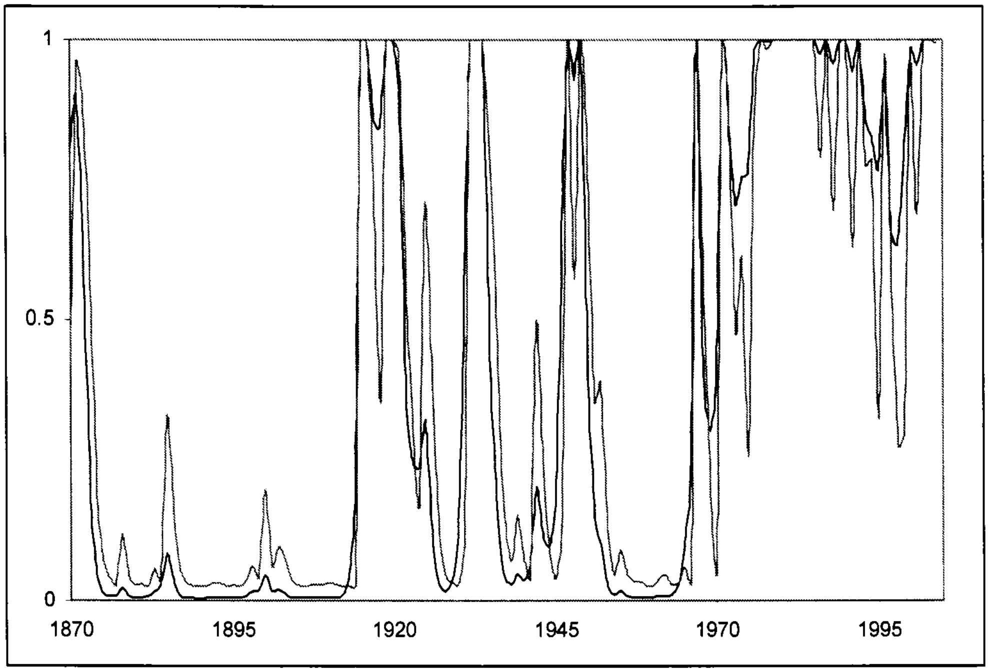

Figure 1.3 Estimated Probabilities of High Variance Regime for Switching AR(1) and Unit Root Model (US/UK). The blue/solid lines are the smoothed estimates and the pink/dotted lines are the filtered estimates.

Figure 1.2 is of the US/UK in which the vertical axis represents the probability of being in a permanent, Unit Root state at the date indicated on the horizontal axis.[18] Our goal is to explain the permanent states indicated by the model. Using the smoothed estimates, we define a period to be in the permanent state should its probability exceed 0.5. Otherwise, it is in a transitory state. The following are the resulting permanent states with potential explanations.

[18] Figure 1.1 shows plots for the other countries; again, the variations in the figures are another indication that panel analysis may be too restrictive.

1870-3: The United States government passes the "Fourth Coinage Act" in 1973 as a response to newly discovered Silver in the American West. The US leaves the bimetallism currency system where the dollar could be expressed in both Silver and Gold out of fear that the increased Silver supply would cause inflation.

1915-1921: This is a period of great instability for both the United States and the United Kingdom because of World War I. Furthermore, the world shifts towards an Anchored Dollar Standard in which the other currencies base themselves on the American Dollar.

1931-4: During the early 1930s, many governments change their currency away from being based on Gold. In 1931, the United Kingdom leaves the Gold Standard and the United States follows suit in 1933. The following year, the United States raises the price of gold from \$20/oz to \$35/oz. Another reason for the noise during this period is the Great Depression, where much of the stability is lost in the financial markets and the economy as a whole.

1946-9: After World War II, the United Nations held a conference and established the Bretton Woods institutions. The participants agreed to use Gold as the common currency standard.

1967: This spike does not have as clear of an explanation as the others. It could be a result from the American War in Vietnam or the formation of OPEC a few years earlier.

1971-present: The Bretton Woods agreement collapses after the US abandons it. Ever since, both the United States and the United Kingdom have been in a world of flexible exchange rates.

The continued Unit Root state since the end of Bretton Woods is consistent with current findings dubbed as the Exchange Rate Disconnect Puzzle. Economic fundamentals should, in theory, be closely related to exchange rates. As the previous figure indicates, the RER has experienced large, highly volatile shocks since 1971, but this volatility is not reflected in the fundamentals. There is seemingly a "disconnect" between the exchange rate and the underlying, economic variables.[19]

4. Modeling Fundamentals

In the previous section, we established the robustness of the models and showed how the different states coincide with historical events. Now, we will

[19] Flood and Rose (1995) have a thorough discussion of this disconnect, but recent work –such as Cheung and Lai (1997), which uses two efficient unit root tests to find PPP holding—challenge this finding.

apply rigorous, econometric techniques to explain the behavior of the switching. First, we apply OLS and regress the filtered probabilities on fundamentals. Then, we include the fundamentals directly within the measurement equation to see if it results in reduced residual volatility. Lastly, the model is allowed to have Time-Varying Transition Probabilities dependent on fundamentals. These exercises show that while fundamentals do influence the switching behavior of our model, the effect varies depending on the choice of bilateral exchange rates. This is more evidence that pooling the data may lead to spurious results.

4A. OLS Regressions

The filtered probability of being in a transitory state or a permanent state is the dependent variable. Even though there is little difference in the 2 series for our models, we use the filtered as opposed to the smoothed probabilities because the latter depicts an overall "trend" by utilizing the entire data sample whereas the former only uses the data up to the point in question to derive its estimate. We include both prices and volatilities of Gold, Silver, Oil, and The Economist Commodity Index. Then, we construct 3 series of dummy variables:

1) When the US is on the Gold Standard, DUM_GOLDUS takes a value of 1
2) When the UK is on the Gold Standard, DUM_GOLDUK takes a value of 1

3) If the US is involved in a war, the WAR dummy takes a value of 1.[20] Other variables in the analysis include Inflation Volatility Differences, GDP/Capita Differences, and Average Openness, which is the average Total Trade / GDP between the 2 countries.[21] The data is obtained from Global Financial Services. We log the prices and construct volatilities by taking the standard deviation of monthly data over the year in question.

Table 1.13 shows the OLS results for selected countries.[22] The variable inclusion criterion is based on Bayesian Model Averaging.[23] For the US/UK, Openness, US on the Gold Standard, Silver Volatility, and Commodity Index Volatility have significant effects on the probability of being in a low or high variance state. We note that each variable (except for Inflation Volatility Differences, which is not listed in the table) is significant under some country pair.[24] Perhaps different countries have different dependencies on fundamentals due to different industry structures, commodity endowments, monetary policies, and inflation expectations. The results point out that fundamentals do have

[20] We include DUM_GOLDUS (DUM_GOLDUK) when the US (UK) is one of the countries in the bilateral RER. For US/UK, we use DUM_GOLDUS but both dummies yield similar results because both countries were on the Gold Standard during nearly identical periods.
[21] For our selected countries, Inflation Volatility Differences are available for only US/UK, and Openness is constructed for only US/UK and Switzerland-US.
[22] Results for other country pairs are available upon request.
[23] See Raftery (1995).
[24] Inflation Volatility Differences is only available for US/UK from 1914 onwards.

Table 1.13 OLS Regressions on Fundamentals selected by BMA grouped by quick reverting (top) and persistent (bottom) country pairs.

	Gold	Gold Volatility	Silver	Silver Volatility	Oil	Oil Volatility	GDP/Capita Difference
Belgium-US	--	--	0.3035 (0.0401)	-1.3353 (0.7009)	0.2850 (0.0873)	--	--
Finland-US	--	-1.9730 (0.1337)	--	--	--	1.0260 (0.3711)	0.5469 (0.1768)
France-US	--	--	----	1.0538 (0.5011)	--	--	--
Portugal-US	--	--	0.2836 (0.0634)	--	--	--	--
Denmark-UK	0.0010 (0.0005)	--	-0.2170 (0.1059)	--	--	1.5970 (0.3504)	--
Canada-Swi.	--	--	0.2000 (0.0453)	--	0.1727 (0.0844)	--	--
Can. – Ger.	--	--	0.1241 (0.0405)	--	0.2360 (0.1335)	--	--
Canada - US	0.0009 (0.0002)	--	--	(1.2943) (0.3668)	--	--	1.4498 (0.2414)
France - UK	--	--	--	--	--	--	0.2959 (0.1119)
Ger. - UK	--	--	0.1244 (0.0552)	--	0.3575 (0.0624)	1.1101 (0.2967)	--
Swi.- US	--	2.6854 (0.9287)	--	0.3085 (0.0743)	0.3251 (0.1144)	--	--
US - UK	--	--	--	1.4389 (0.3782)	--	--	(0.3197) (0.1383)

Table 1.13 (continued) OLS Regressions on Fundamentals selected by BMA grouped by quick reverting (top) and persistent (bottom) country pairs.

	Commodity Index	Commodity Volatility	Openness	War	Gold Standard	Intercept	R^2
Belgium-US	--	--	--	--	--	-0.8039	0.4190
						(0.2063)	
Finland-US	0.1716	--	--	--	-0.2113	-0.6127	0.3220
	(0.0705)				(0.1043)	(0.2892)	
France-US	--	--	--	-0.3325	-0.2068	0.6346	0.2990
				(0.0725)	(0.1065)	(0.2968)	
Portugal-US	-0.1116	--	--	--	--	0.1961	0.4060
	(0.0759)					(0.2659)	
Denmark-UK	--	--	--	--	--	0.6333	0.3910
						(0.2076)	
Canada- Swi.	--	--	--	-0.0937	-0.2943	-0.3651	0.4590
				(0.0556)	(0.0603)	(0.2207)	
Canada – Ger.	0.1335	3.7104	--	--	--	(0.7850)	0.5158
	(0.0407)	(0.9894)				(0.1747)	
Canada - US	--	2.2652	--	(0.1495)	(0.1927)	(0.0479)	0.4193
		(0.7954)		(0.0546)	(0.0530)	(0.0985)	
France - UK	--	3.9612	--	(0.2091)	--	0.6772	0.2730
		(0.2959)		(0.0667)		(0.0625)	
Ger. - UK	--	4.2346	--	--	(0.3464)	(0.6456)	0.6151
		(0.9753)			(0.0861)	(0.1703)	
Swi. - US	(0.1173)	--	--	--	(0.4912)	(0.2448)	0.3100
	(0.0574)				(0.1226)	(0.2307)	
US - UK	--	3.5897	1.9174	--	(0.3231)	0.2558	0.6439
	(0.8532)	(0.3772)		(0.0590)	(0.0543)		

explanatory power on the behavior of shocks in our model, but which fundamentals depends on the bilateral RER being analyzed. This is another outcome that suggests pooling countries for analysis is not suitable. We do not find any systematic differences between the country pairs exhibiting quick reversion and those that do not.

4B. Parameters in the Measurement Equation

In this exercise, we incorporate the fundamentals directly in the measurement equation. We use the Switching 2 Unit Root specification because the country pair analyzed is US/UK, which has a highly persistent transitory process in the original Switching AR(1) and Unit Root model.[25] If the fundamentals affect our RER series, the model should pick up a significant coefficient in β and reduce the weight placed on the original switching UR processes.

$$\begin{aligned} y_t &= x_t + z_t + \beta * Fundamental_t \\ x_t &= x_{t-1} + S_t v_t \\ z_t &= z_{t-1} + (1-S_t)\, e_t \\ & v_t \sim N(0, \sigma_v^2)\,,\ e_t \sim N(0, \sigma_e^2) \end{aligned} \tag{7}$$

The results for the US/UK RER series are in Table 1.14.

[25] Using the Switching AR(1) and Unit Root Model yields the same results, which is further evidence that this alternative specification holds for the country pairs with slow reversion.

Table 1.14 Fundamentals in Measurement Equation (US/UK)

	Pr(LO|LO)	*Pr(HI|HI)*	σ_{LO}^2	σ_{HI}^2	β	**LLH Value**
Gold	0.9035	0.8752	0.0252	0.1146	0.1586	194.5409
	(0.0493)	(0.0801)	(0.0028)	(0.0144)	(0.1044)	
Gold Volatility	0.8890	0.8055	0.0253	0.1368	(0.4467)	193.1937
	(0.0468)	(0.0834)	(0.0033)	(0.0182)	(0.3426)	
Silver	0.9128	0.8858	0.0243	0.1241	0.0364	192.6870
	(0.0419)	(0.0632)	(0.0027)	(0.0132)	(0.0202)	
Silver Volatility	0.9106	0.8774	0.0252	0.1259	(0.0099)	191.2102
	(0.0425)	(0.0699)	(0.0027)	(0.0139)	(0.0776)	
Oil	0.9081	0.8719	0.0252	0.1263	0.0058	191.2342
	(0.0439)	(0.0750)	(0.0026)	(0.0141)	(0.0221)	
Oil Volatility	0.9098	0.8762	0.0252	0.1259	(0.0017)	191.2028
	(0.0429)	(0.0715)	(0.0027)	(0.0141)	(0.0498)	
Commodity Index	0.8994	0.8579	0.0244	0.1256	0.0669	192.6646
	(0.0494)	(0.0842)	(0.0026)	(0.0140)	(0.0383)	
Commodity Volatility	0.9097	0.8763	0.0252	0.1257	(0.0113)	191.2050
	(0.0429)	(0.0706)	(0.0027)	(0.0139)	(0.1184)	
Openness	0.8803	0.7568	0.0241	0.1379	2.6612	184.0305
	(0.0501)	(0.1088)	(0.0025)	(0.0176)	(0.4497)	
War	0.9113	0.8855	0.0231	0.1226	0.0282	195.6347
	(0.0429)	(0.0615)	(0.0025)	(0.0127)	(0.0087)	
Gold Standard	0.9122	0.8778	0.0241	0.1290	0.0432	193.7624
	(0.0406)	(0.0670)	(0.0024)	(0.0139)	(0.0177)	
GDP/Capita Difference	0.9008	0.8329	0.0257	0.1312	0.0166	189.5344
	(0.0474)	(0.0950)	(0.0033)	(0.0164)	(0.0627)	

The coefficient on the fundamental, β, is not significant for any of the series, and the other parameter estimates are very close to the original estimates.[26] The sample size is too small for the confidence intervals to be tight enough to yield significant regressors.

4C. Time Varying Transition Probabilities

In our current specification, the transition probabilities are constant for the duration of our data sample. It is quite conceivable that the behavior of the RER is linked directly to the performance of underlying, economic fundamentals. Here, we attempt to explain the switching states in terms of these fundamentals. Again, we choose to analyze the US/UK series, so we run this program on the Switching 2 Unit Root specification. If these extra parameters have a significant impact on our model, it hints that there are other processes that must be accounted for in addition to the two permanent processes. The basic methodology is to enhance the original program to allow the transition probabilities to vary based on other variables. The basic probit probabilities are:

$$\text{PPr} = \Pr(\text{St=Hi} \mid \text{St=Hi}) = 1 - \frac{1}{1+\exp(p_0 + p_1 Z_t)} \tag{8}$$

$$\text{QPr} = \Pr(\text{St=Lo} \mid \text{St=Lo}) = 1 - \frac{1}{1+\exp(q_0 + q_1 Z_t)}$$

[26] We ran the Commodity Index as a fundamental for the US-Australia RER because Australia is a commodity economy. Again, the coefficient for the fundamental is not significant. Then, we ran this exercise for the AR(1) with Unit Root Model and the other models with AR(1) processes.

p_0, p_1, q_0, and q_1 are unconstrained parameters in the optimization and Z_t is the value of the fundamental at time t. Figures 1.3 and 1.4 show the plots PPr and QPr with their 95% confidence bands for Oil.[27] Oil Prices show significant movement through time, the horizontal-axis.

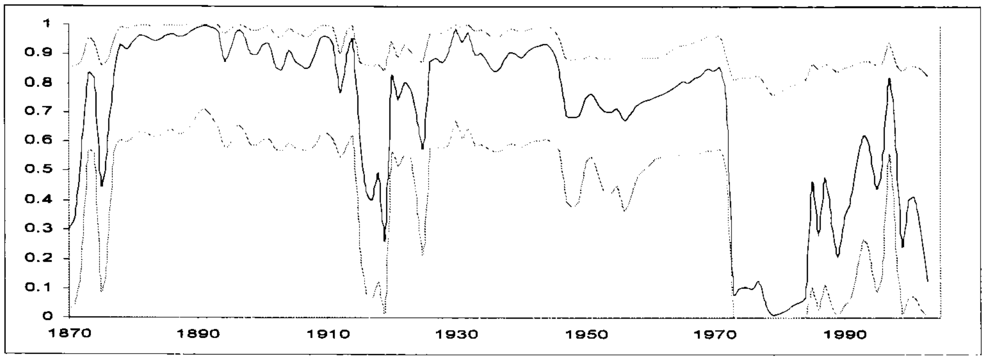

Figure 1.4 Time-Varying Transition Probabilities (PPr)

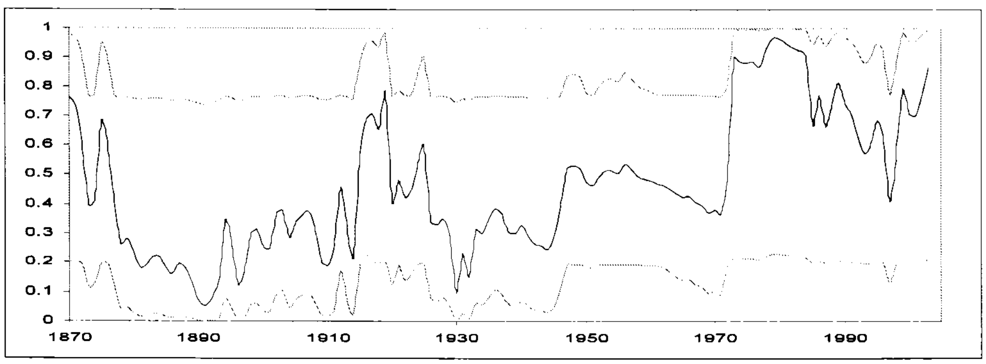

Figure 1.5 Time-Varying Transition Probabilities (QPr)

Even so, the confidence intervals are wide enough that a straight line can be drawn through the entire timeframe, so a constant probability could fit the 2 Unit Root Model. This is even more apparent for the other commodities, so the

[27] Figures for other fundamentals are available upon request.

Table 1.15 Time-varying Transition Probabilities (US/UK)

	p_0	q_0	σ_{v2}	σ_{e2}	p_1	q_1	**LLH Value**
Gold	2.4745	1.7331	0.0252	0.1293	2.4317	2.9870	195.321
	(0.6136)	(0.7046)	(0.0028)	(0.0147)	(1.1866)	(1.3996)	
Gold Vol.	0.4406	0.7762	0.0245	0.1297	(91.2717)	21.6333	196.581
	(1.0719)	(0.7685)	(0.0027)	(0.0144)	(43.2762)	(24.4934)	
Silver	2.0829	1.5488	0.0246	0.1284	1.4715	0.3384	193.722
	(0.5613)	(0.6453)	(0.0031)	(0.0145)	(0.7861)	(0.8208)	
Silver Vol.	2.5899	14.6758	0.0224	0.1139	(72.7028)	355.5128	201.585
	(0.8714)	(7.1335)	(0.0022)	(0.0100)	(36.1170)	(197.909)	
Oil	1.1290	(0.1143)	0.0245	0.1358	(3.7771)	2.4490	198.008
	(0.4707)	(0.6389)	(0.0024)	(0.0158)	(0.5661)	(1.1434)	
Oil Vol.	2.7608	2.2264	0.0249	0.1236	(9.8680)	3.0652	193.320
	(0.7493)	(0.6426)	(0.0027)	(0.0128)	(5.1328)	(0.8255)	
Comm. Ind.	2.6422	1.9619	0.0250	0.1248	3.3886	2.0477	192.188
	(0.6693)	(0.6190)	(0.0029)	(0.0133)	(3.1249)	(1.8179)	
Comm. Vol.	1.0920	1.1860	0.0215	0.1216	(58.9334)	89.7928	197.482
	(0.5184)	(0.8175)	(0.0031)	(0.0125)	(31.7486)	(51.6711)	
Openness	0.5839	0.7913	0.0248	0.1435	(58.1809)	19.5544	186.0224
	(0.9521)	(0.6960)	(0.0026)	(0.0176)	(35.8827)	(14.695)	
War	2.4331	2.0640	0.0243	0.1240	(1.8569)	(1.9195)	193.533
	(0.6458)	(0.6344)	(0.0027)	(0.0136)	(1.2003)	(1.1515)	
Gold Std.	1.8747	0.7391	0.0240	0.1266	2.2930	(2.4198)	195.831
	(0.5694)	(0.9110)	(0.0027)	(0.0145)	(1.2689)	(2.1376)	
GDP Diff	1.6572	2.1023	0.1273	0.0254	(1.3025)	2.9605	192.590
	(0.6876)	(0.5312)	(0.0145)	(0.0029)	(3.0261)	(2.5029)	

probabilities do not necessarily vary through time in our finite sample. Table 1.15 shows the numerical estimation produced by the program. We see that Oil yields significant coefficients for both probabilities. The coefficients and the probabilities move in opposite directions, so as Oil increases, PPr (probability of remaining in a Unit Root state) decreases and QPr (the probability of remaining in a stationary state) increases. In other words, as the log of Real Oil Prices increases, the probability of staying in (or moving to) a more volatile RER regime increases. Nonetheless, the smoothed and filtered plots of the states barely change after allowing for the additional parameters. In general, the confidence intervals for the probabilities are very wide, so there is not enough power in our sample size to determine whether or not the transition probabilities are indeed time-varying.

From our 3 ways of incorporating fundamentals into the regime switching model, we find it difficult to pinpoint the exact fundamentals that determine the RER process. The OLS approach shows the fundamentals to play a role in determining the state of the RER (transitory of permanent), but there is much collinearity among the regressors, and fundamentals are significant for different country pairs; perhaps, due to different industry structures, commodity endowments, monetary policies, and inflation expectations, different countries have different dependencies on fundamentals. Nevertheless, certain series—such as GDP/Capita Differences—show up frequently. This suggests the Balassa-

Samuelson effect may explain some of the variation in the RER. In general, the estimations reveal parameters to be of different values and significance for different country pairs, so the data should not be pooled.

5. Conclusion

Using over 100 years of data from 16 OECD countries, we find that the RER is best described as having both transitory and permanent shocks, a framework that overcomes many issues arising from typical long horizon, structural break, and pooled data analyses. The Switching AR(1) and Unit Root Model allows the RER to possess either stationary or Unit Root nonstationary shocks in any given period. The majority of the shocks to the RER are transitory ones though the degree of persistence varies depending on the country pair in question. Certain countries, such as Belgium, Portugal, and Spain, consistently have half-lives below 2 years during the transitory states.

This leads us to caution the common practices of both using long horizon data analysis and panel analysis as means of circumventing the testing power issues associated with the search for the existence of PPP. Our model clearly shows distinct regimes that would become mixed in simple long horizon and panel frameworks, which in turn could lead to biased estimations. One of the robustness checks for the model shows that using structural breaks to account for regime shifts is also misleading. If the stationary process in our model is highly persistent, then the endogenous structural breakpoint tests show spurious breaks.

Our specification presents regimes that are consistent with historical events. The model implies that shocks due to wars and currency standard shifts are permanent, whereas the other shocks are transitory and mean reverting an a stationary regime. After the collapse of Bretton Woods, our model consistently shows noisy periods, which supports the current literature exploring the Exchange Rate Disconnect Puzzle, where the large volatility observed in exchange rates is not reflected in economic fundamentals.

In a more rigorous analysis of the regime shifts, we find that the inclusion of fundamentals diminishes the puzzle by providing explanatory power for the changes of regimes. The fundamentals we use are commodity prices and volatilities, GDP/Capita differences, inflation differentials, war periods, openness, and gold standard periods. In addition to OLS regressions selected by BMA, we model the fundamentals as time-varying transition probabilities and also directly in the measurement equation; the significance of the variable depends on the bilateral RER used in the analysis, which could be because different countries have different policies and dependencies on fundamentals. The variation in the role of each fundamental further supports the potential problems in pooling data to increase sample size.

Though a handful of countries have reduced half-lives, the PPP Puzzle still remains for many others. Our model is in agreement with that literature that finds a long half-life for the shocks to RERs. One possible extension would be to pool highly persistent country pairs could and use the Switching 2 Unit Root framework to increase the power of the tests. Assuming proper correction for heteroskedasticity, this would allow the researcher to exploit the increased sample size while maintaining a homogenous sample. However, countries, such as Belgium, that clearly possess a transitory process must be omitted from such panels. Lastly, we could modify the model to include more states, but that would reduce the parsimony of our current framework.

Chapter II

Trends in First-Year Costs and Survival for Colorectal Cancer Patients: Empirical Testing for Structural Breakpoints

(with Louis Garrison)

1. Introduction

Colorectal cancer is the third most prevalent cancer and accounts for 10% of all cancer deaths in the United States (Jemal et al, 2005).[28] In 2005, there were over 145,000 cases of colorectal cancer and experts predict there is a 6% lifetime chance of getting the disease (Winawer et al, 2003). The monetary cost of the disease is clearly great: one report estimates the direct medical costs alone were over $5.7 billion in 1997 (Brown, 2001).

In this study, we analyze trends in both first-year costs and survival for colorectal cancer patients using time-series techniques. Table 2.1 shows the approval of new drugs during the past 45 years. Most new introductions have been in the last 10 years, which could lead to a change in average treatment costs and outcomes. Our analysis focuses on the societal impact of these new drugs on costs and survival. This information could be important not only for individuals diagnosed with colorectal cancer but also for the decisions of insurance providers and medical care providers.

[28] From 1998-2002, the incidence of colorectal cancer per 100,000 individuals was 62.1 for men, 46.0 for women, 61.7 for white men, and 72.5 for black men (Jemal et al, 2005).

We perform a time-series study of the first-year costs and first-year survival probabilities of newly diagnosed colorectal patients from 1991-2002. This is in contrast with the lifetime cost studies done by Ramsey et al. (2002) and Etzioni et al.(2001). We examine the underlying cost structure changes through time, which can provide useful information on the homogeneity assumptions, perhaps overly restrictive, sometimes made in lifetime costs analyses.

Unlike most pharmacoeconomic studies of costs and survival that are based on clinical trial data, we use "real-world" data collected in the SEER-Medicare databases. Then, we compare our empirical results to the pharmacoeconomic modeling results. Our methodology and estimates can serve as a baseline for future assessment of the real-world impact of drug regimens.

Though individual-level data is available, we use an aggregated time-series in our analysis in order to look at the overall burden of colorectal cancer to society. By not identifying the introduction of particular drug treatments ahead of time, we impose no preconceptions on the data. For example, non-pharmaceutical events, such as changes in Medicare reimbursements, may also change the average cost of illness for colorectal cancer. Using aggregated data allows us to examine the overall pattern of costs and survival before searching for explanations in their potential shifts through time.

Table 2.1 Colorectal Cancer Drug Approval Dates[29]

Drug	When Approved by FDA	General Indication
Fluorouracil (5-FU)	Apr. 25, 1962	Colon-Rectum
Leucovorin	Dec. 12, 1991	In combination with 5-FU
Irinotecan (Campto)	June 14, 1996	Metastatic CRC
Capecitabine (Xeloda)	Apr. 30, 2001 Sept. 7, 2001 June 15, 2005	Metastatic CRC (1st line) Metastatic CRC (2nd line) Adjuvant therapy-Dukes' C CRC
Oxaliplatin (Eloxatin)	Aug. 9, 2002 Jan. 9, 2004 Nov. 4, 2004	Metastatic CRC (2nd line) Metastatic CRC (1st line) Adjuvant therapy-Stage III CRC
Cetuximab (Erbitux)	Feb.12, 2004	EGFR-expressing metastatic CRC in patients refractory to irinotecan-based chemotherapy (combo) or intolerant to irinotecan
Bevacizumab (Avastin)	Feb. 24, 2004	Metastatic CRC (1st line)

For this analysis, we employ the endogenous structural breakpoint test for multiple breaks on series of costs.[30] To our knowledge, this is the first study utilizing these techniques for analyzing trends in cancer costs. There may be a break in cost and survival for a variety of reasons, such as if the impact of the new drug regimens is substantial or if there is a change in the Medicare reimbursement scheme. On the other hand, the new drugs could be priced to capture any potential savings in other costs; for example, even if the pharmacy costs are

[29] From Best (2007)
[30] See Bai and Perron (1998) and Qu and Perron (2005) for a detailed description of the tests.

greater, they could be offsetting the lower laboratory costs. The end result would be no change in the overall cost. Additionally, the changing trend in overall medical spending could be affected by medical price inflation, as reflected in the CPI Medical Deflator (base period of 1982-84), which we use to adjust nominal amounts to real amounts.

The results indicate a clear break in average first-year costs for Stage II, III, and IV patients. However, there is no break in average first-year survival for any group of patients. This would imply a break in the average cost-effectiveness of colorectal cancer treatment in the recent past.

This empirical study of costs can be compared with the pharmacoeconomic projections based on the results of clinical studies. We find significant changes in real costs in the time-series in the average cost of first-year patients by $8,433 to $13,424, corresponding in time to the introduction of irinotecan. For survival, however, we fail to find a break in the time-series that might be reflective of the 20 percentage point increase in first-year survival probability attributed to irinotecan in clinical trials.

On first glance, the break date lines up with the introduction and use of irinotecan. Then, we perform a series of robustness checks by analyzing subcomponent costs, including a time-trend, removing irinotecan patients, and an alternative measurement of cost. These exercises show that the introduction of irinotecan alone cannot fully explain the break, so there are most likely other reasons for the changing cost-effectiveness.

The following section describes the Materials and Methodology used. Section 3 presents the Results and Discussion. Section 4 has our Robustness Checks. Finally, Section 5 concludes and suggests extensions to this project.

2. Materials and Methodology

Our data combines the Surveillance, Epidemiology and End Results (SEER) and Medicare databases. SEER has 17 registries in the United States that provide individual level data on demographics, clinical diagnosis, and cause of death for cancer patients. See Figure 2.1 for a map of the SEER registries. Stage of disease is directly assessed only at baseline when each patient enters the database: thus, progression of the disease can only be inferred from other indicators. The Medicare database provides claims data for health services, which covers institutional stays, physician/supplier bills, outpatient services, home health services, hospice care, and durable medical equipment claims. Though the database records all processed claims for Medicare Part A and Part B, it does not have data on outpatient, out-of-pocket expenses.[31] The linked SEER-Medicare database allows researchers to track the medical expenses of cancer patients on an

[31] Medicare Part A covers hospital stays and Part B covers outpatient, medical expenses for services and products not included in Part A. Recently Part C (Medicare Advantage Plans) and Part D (Prescription Drug Plans) have become options to those who qualify, but these choices became available only after the end of our data series.

individual level. The database is publicly available for researchers and is updated on a continuing basis, most recently in 2006 to include Medicare data through 2002 and SEER data through 2003. For our analysis, we use Washington State patient data from 1991-2002, which provides us with 12,714 colorectal cancer patients in total.

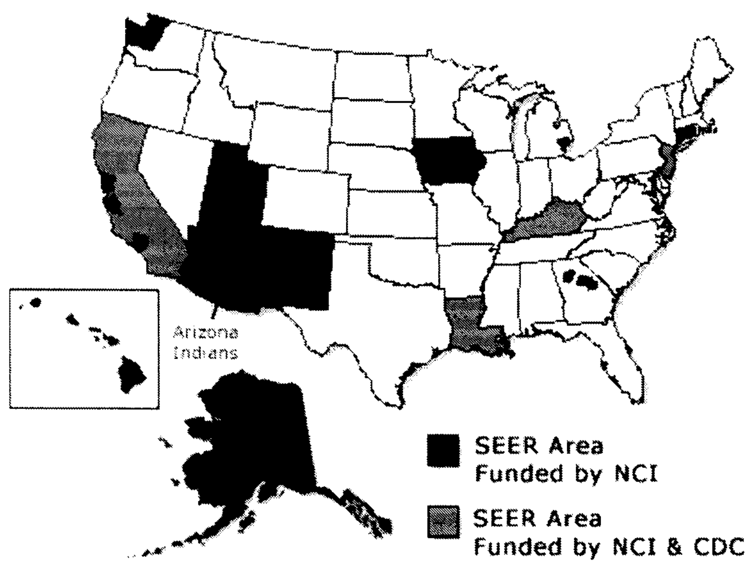

Figure 2.1 Map of SEER Registries[32]

In our final analysis, the patients must be covered by both Parts A and B for the entire first year following their initial diagnosis. The subjects must be 65 years or older unless disabled; we also exclude Medicare End Stage Renal

[32] See http://seer.cancer.gov/registries/ for more information.

Disease Program patients. Any individuals with prior diagnosis of cancer are also excluded. After processing our exclusion criteria, we are left with 10,038 individuals.

Table 2.2 Observations per Stage

Stage	1	2	3	4	0 or unknown
Observations	2,210	2,943	2,122	1,425	1,338

We combine the individual-level data to create a time-series of average first-year costs in the aggregate. Any filed claims from the Medicare files that occur within 365 days of initial diagnosis contribute to the total first year cost of the individual.[33] Then, we adjust the average first-year cost for all individuals in a given period by the CPI Medical deflator.[34]

The measure for survival is the probability of surviving the first-year after initial diagnosis. The initial date is based on the entry into the SEER database. We then calculate the percentage of patients who are alive 365 days after the initial date. Finally, we create a time-series of survival probabilities based on the entrance date in SEER.

The time-series frequency is quarterly because higher frequency data—such as monthly or daily—is not only very noisy but also too small in sample to

[33] The variable name depends on the claim file used. For Medicare Provider Analysis and Review, Carrier Claims, Outpatient Claims, Home Health Agency Claims, and Hospice Claims, the series is "Claim Total Charge Amount." For Durable Medical Equipment, it is "Submitted Charge Amount." Note that using charges may overestimate the actual payment amount. See Section 4.4 for more analysis using payment series.
[34] See Bureau of Labor Statistics for detailed information on the CPI.

aggregate for this type of time-series analysis. Lower frequency data (semi-annual and annual) would give us a small sample of data points with which to work and also assumes a constant data generating process within the longer duration. Our choice of quarterly data leaves us with 48 periods, which also allows us to use the asymptotic critical values in our analysis.[35]

To find breaks in our cost and survival series, we employ the endogenous structural breakpoint tests of Bai and Perron (1998) and Qu and Perron (2005). In particular, we use the sequential method to determine if multiple breakpoints exist. The test extends the traditional structural breakpoint tests of the Chow Test where the researcher specifies an *a priori* break date and then tests the potential subsamples for consistency using an F-test. Our tests are "endogenous" in that the break dates are not pre-specified but rather each date remaining (after 15% each end of the sample is trimmed for analytical purposes) is tested as the potential breakpoint. If there is indeed a break, the subsequent subsamples are then tested in the same manner for additional breaks. After finding breaks in our cost series, we then show the break dates coincide with the introduction and adoption of irinotecan for the treatment of colorectal cancer in Medicare patients.

[35] Chapter 3 shows that endogenous structural breakpoint tests in small samples should rely on bootstrapping techniques. A sample size of around 50 is large enough to use the asymptotic critical values.

The datasets are constructed in SAS 9.1 TS Level 1M3 for Windows XP. We utilize the code provided by Pierre Perron and run it on Gauss Light.[36]

3. Results and Discussion

Our analysis segments first-year costs, i.e., by different stages of disease. In practice and in clinical trials, staging is the determining factor in selecting treatment regimens as well as in outcomes of patients with colorectal cancer. The choice of treatment does not vary greatly with the sex and race of the individual, so we omit the results from using those specific segments.

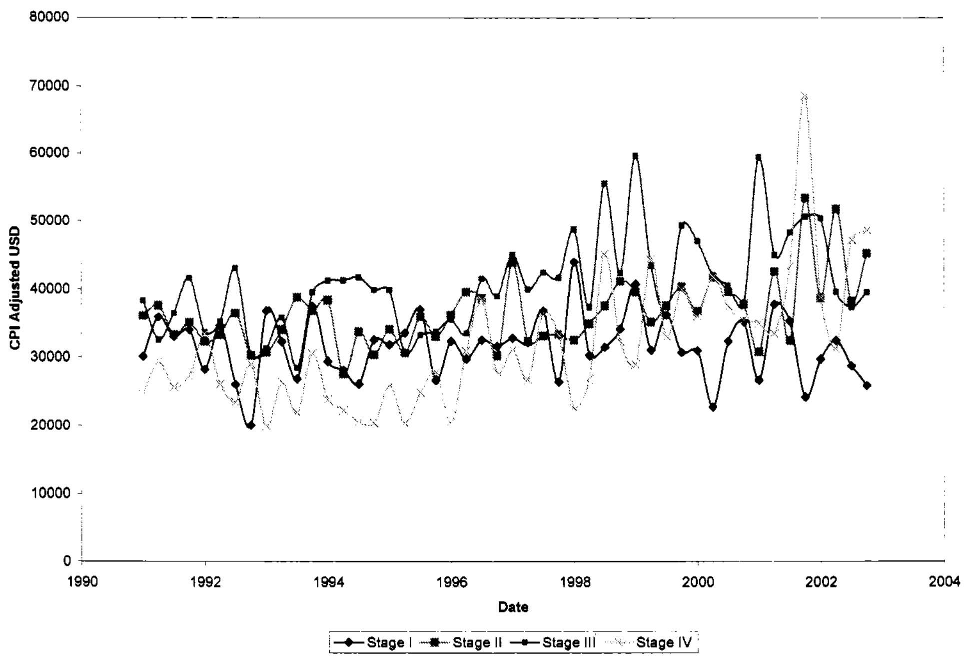

Figure 2.2 Average First-Year Cost by Stage of Disease at First Diagnosis

[36] See http://people.bu.edu/perron/code.html for code.

A simple plot of average first-year costs by stage can be seen in Figure 2.2. Since there is no clear indication of a change in costs, we perform the endogenous structural breakpoint tests. Table 2.3 shows the break date, average first-year real cost pre- and post-break, and break size for each of the stages. Aside from Stage I, each of the other stages of disease experiences a single break during our sample from 1991-2002.[37] For Stage II patients, the break occurs in the April 2001. The breaks are in January 1997 and July 1998 for Stage III and Stage IV respectively.

Also in Table 2.3, we see that both the absolute change and the percentage change in average first-year cost increases with stage of disease. The most dramatic change in costs occurs in Stage IV patients, where it jumps from \$26,623 to \$40,047 after the break date—a 50% increase in real costs. Next, we show that the break dates for Stage III and Stage IV coincide directly with the approval of irinotecan.

Irinotecan, marketed as Campto® by Yakult Honsha and Camptosar® by Pfizer, is a chemotherapy agent that inhibits Type I topoisomerase. It was given accelerated approval in June 1996 for "treatment of patients with metastatic carcinoma of the colon or rectum whose disease has recurred or progressed following 5-FU-based therapy," but regular approval was not obtained until

[37] The program allows for multiple breakpoints to be detected for each series, but it finds only a single point. This may be due to the sample size because the sub samples created after the initial break is discovered are only around 24 points each, which may be too limiting to discover subsequent breaks.

Table 2.3 Breaks in Average Real First-Year Cost (1982-4 US$) for CRC Patients by Stage (Coefficients with standard errors in parentheses and confidence intervals in brackets)

	Break Date	1st Level	2nd Level	Change	% Change
Stage I	--	--	--	--	--
Stage II	2001 Q2 [2000Q2, 2002Q1]	35407.79 (687.69)	43208.62 (1664.32)	7800.83	22%
Stage III	1997 Q1 [1995Q1, 1998Q1]	36559.35 (1150.28)	44992.01 (1150.28)	8432.65	23%
Stage IV	1998 Q3 [1996Q3, 1999Q1]	26622.99 (1230.81)	40047.36 (1588.97)	13424.37	50%

October 1998.[38] Table 2.4 shows the frequency and average cost of the irinotecan procedure based on patients' entrances into the hospital. It clearly confirms that the treatment was not available prior to 1998, so it was not used until the drug received regular approval from the FDA. The 1998 break date in Stage IV patients coincides directly with the introduction of irinotecan into oncology practice, which is a reasonable observation because it was approved for only Stage IV patients.

[38] See FDA Press Releases. http://www.fda.gov/cder/cancer/druglistframe.htm

Table 2.4 Number and Average First-Year Real Cost of Patients Initiating Irinotecan Treatment by Quarter

Date	Frequency	Avg. Cost	Date	Frequency	Avg. Cost
1991 Q1	--	--	1997 Q1	--	--
1991 Q2	--	--	1997 Q2	--	--
1991 Q3	--	--	1997 Q3	--	--
1991 Q4	--	--	1997 Q4	--	--
1992 Q1	--	--	1998 Q1	16	1031.6
1992 Q2	--	--	1998 Q2	13	1233.48
1992 Q3	--	--	1998 Q3	8	1942.5
1992 Q4	--	--	1998 Q4	12	1757.08
1993 Q1	--	--	1999 Q1	7	1442.63
1993 Q2	--	--	1999 Q2	6	1151.19
1993 Q3	--	--	1999 Q3	12	1087.03
1993 Q4	--	--	1999 Q4	9	1523.98
1994 Q1	--	--	2000 Q1	27	3127.68
1994 Q2	--	--	2000 Q2	13	1635.28
1994 Q3	--	--	2000 Q3	23	2413.15
1994 Q4	--	--	2000 Q4	17	2690.4
1995 Q1	--	--	2001 Q1	22	1650.11
1995 Q2	--	--	2001 Q2	11	1340.38
1995 Q3	--	--	2001 Q3	17	1763.99
1995 Q4	--	--	2001 Q4	26	2324.24
1996 Q1	--	--	2002 Q1	13	1798.66
1996 Q2	--	--	2002 Q2	13	2253.66
1996 Q3	--	--	2002 Q3	19	2158.48
1996 Q4	--	--	2002 Q4	21	2104.09

The question remains, though, of why the Stage III average cost series would experience the same (and perhaps earlier) break date. The estimated break date for Stage III occurs in 1997, which is slightly before the time of irinotecan.[39] One plausible explanation is that those who were first diagnosed with Stage III in 1997 progressed to Stage IV by the following year—just in time for the new drug.

[39] The confidence intervals for the Stage III and Stage IV break dates overlap, so we cannot statistically identify which break came first.

While Stage IV patients have only an average of 9 months overall survival, Stage III patients have a longer average survival duration.

Table 2.5 Use of Irinotecan by Entrance Into SEER Database Date: Frequency and Average First-Year Real Cost

Date	Frequency	Avg. Cost	Date	Frequency	Avg. Cost
1991 Q1	--	--	1997 Q1	7	2664.68
1991 Q2	1	1908	1997 Q2	3	1204.93
1991 Q3	--	--	1997 Q3	9	3508.17
1991 Q4	1	1090.56	1997 Q4	10	1394.86
1992 Q1	2	1552.4	1998 Q1	9	3258.16
1992 Q2	--	--	1998 Q2	6	2027.5
1992 Q3	1	1272	1998 Q3	10	1618.81
1992 Q4	--	--	1998 Q4	8	1981.49
1993 Q1	1	6950	1999 Q1	3	2761.51
1993 Q2	4	1348.04	1999 Q2	14	1937.84
1993 Q3	--	--	1999 Q3	7	2109.83
1993 Q4	3	1479.83	1999 Q4	8	1771.15
1994 Q1	2	740.63	2000 Q1	8	1966.16
1994 Q2	2	1233.51	2000 Q2	5	1588.71
1994 Q3	6	1582.47	2000 Q3	15	2595
1994 Q4	2	4392.5	2000 Q4	5	1851.31
1995 Q1	5	1403	2001 Q1	7	1455.24
1995 Q2	2	3680.02	2001 Q2	8	1574.47
1995 Q3	1	1481	2001 Q3	5	1558.7
1995 Q4	5	1623.38	2001 Q4	14	1776.99
1996 Q1	2	1395.12	2002 Q1	7	1654.95
1996 Q2	5	3368.09	2002 Q2	3	1814.92
1996 Q3	12	1126.38	2002 Q3	13	2531.14
1996 Q4	6	1864.52	2002 Q4	10	1983.77

Table 2.5 shows the use of irinotecan by entrance into the SEER database date. It indicates that patients who entered SEER in earlier years were also given the drug. Since we know that the drug was not available until 1998 and was approved only for Stage IV patients, this indicates irinotecan was indeed also

given to patients whose cancer progressed in addition to those who were only first diagnosed by SEER as Stage IV post-1998.[40]

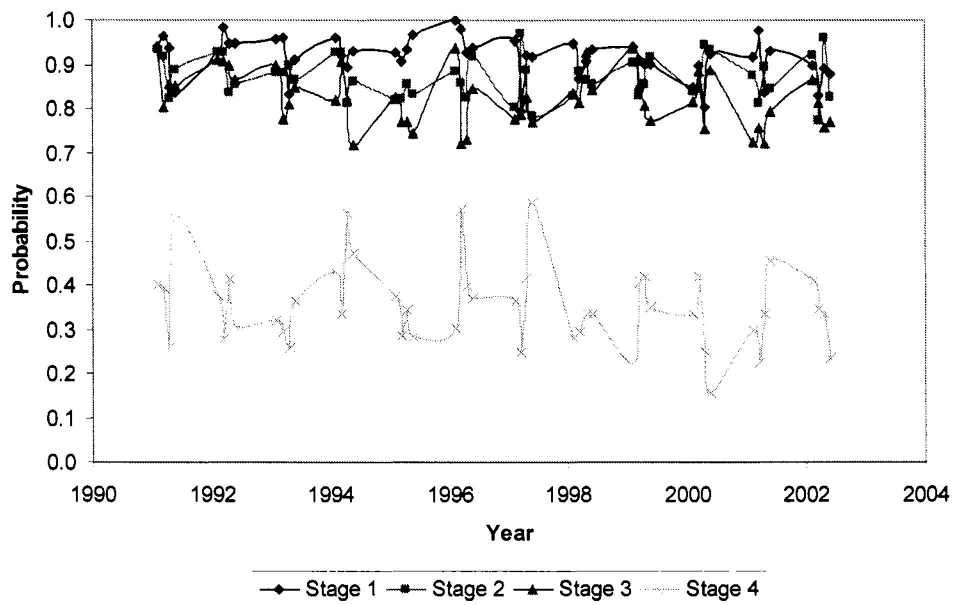

Figure 2.3 First-Year Survival Probability

Although the break dates correspond with the approval of irinotecan, the break size appears to be significantly larger than could be explained by projected incremental cost of this drug alone. From Table 2.3, we see that the breaks in Stage III and Stage IV are characterized with an increase in average first-year real costs of \$8,433 and \$13,424 respectively. In nominal terms, these amounts would be about 2.4 times greater—so roughly \$20,000-\$30,000 in current dollars. Based

[40] Average progression time from Stage III to Stage IV cannot be obtained by this dataset because SEER only records the initial stage of the patient.

on clinical trials, Levy-Piedbois et al. (2000) calculate that the increased survival benefits attributed to irinotecan are worth an additional $9,344 to $10,137 for each life year; Iveson (1999) shows the gain of irinotecan over 5-FU to be between £7,685 and £11,947 per life year. Although the incremental cost of irinotecan could be $20,000-30,000 in the first year, this could not fully explain the break in the aggregate times series, where fewer than 10% of patients are receiving irinotecan.

Next, we examine the trend in first-year survival probability segmented by stage of disease. The plot of survival probabilities in Figure 2.3 shows no clear pattern, so we proceed to perform endogenous structural breakpoint tests for multiple breakpoints. However, the tests do not find any breaks in any of our survival series, so the first-year survival outcomes for colorectal cancer patients have not changed during our period of analysis.

If the uptake of irinotecan significantly shifted the first-year costs, then we would expect to see a similar shift in first-year survival probabilities as indicated in economic studies based on clinical trials. Figure 2.4 shows roughly a 20 percentage point gain in 12 month survival probability in irinotecan over other care in this key trial, but our empirical tests do not suggest that this is reflected in the aggregated series.

With an increase in cost after 1998 and no comparable increase in survival probability over the same period, the data indicates a potential shift in the cost-effectiveness of colorectal cancer care. If our preliminary analysis is correct in attributing the break in cost to the use of irinotecan, these aggregate results might suggest that the drug is not cost-effective because we cannot detect an increase in survival. However, given the small proportion of irinotecan patients, more analyses would be needed to examine and identify the underlying causes of the shift in first-year cost of illness based on our aggregate analysis.

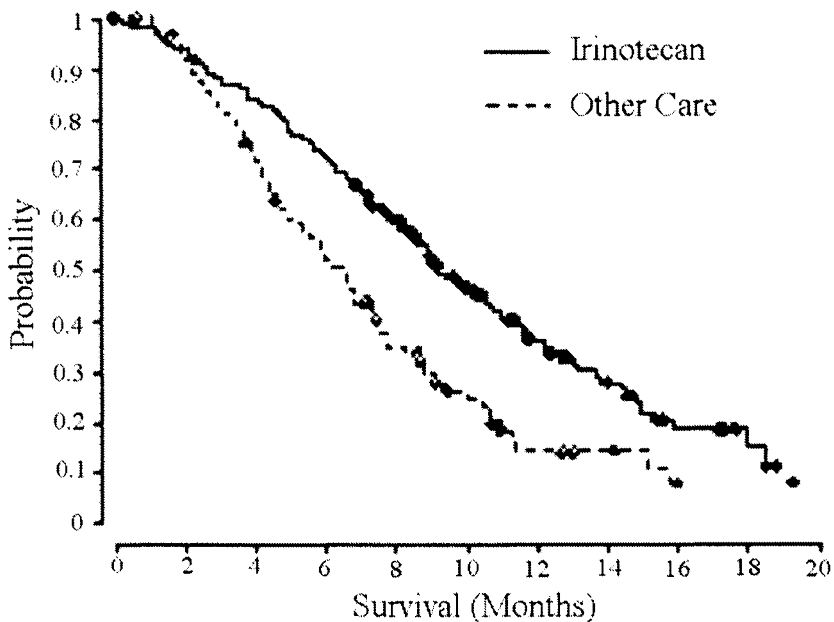

Figure 2.4 Survival Curves Based on Clinical Studies[41]

[41] From Cunningham et al. (1998)

4.1. Robustness Check: Time trend

Can a time trend help explain variations in average first-year real costs in our quarterly time-series from 1991-2002? The plots of first-year cost in Figure 2.2 do not provide a definitive conclusion. Such a trend could be significantly present if new treatments and improved care drive a steady increase in colorectal cancer spending. However, by adjusting by the CPI Medical Deflator, it is clear that there have been significant increases in total current spending, but it is not clear how much a time trend in spending is due to increase in real services received. We explore this factor using a trend analysis of real costs.

First, we estimate a simple OLS regression on a time trend: $cost_t = mean + \beta \cdot time_t + \varepsilon_t$ where $\varepsilon_t \sim iid\,(0,1)$. If the coefficient, β, is significant, it indicates that first-year costs have increased over time. Then, we can include the time trend in our original analysis of structural breaks.

The OLS results in Table 2.6 show a significant time trend for Stages II, III, and IV. As the severity—stage of disease—increases, there is an increase in spending. For our sample of Stage IV patients, each subsequent quarter results in an increase of $452.92 in average first-year real costs.. The R-squared also increases with stage, so we can interpret this as the time trend playing a larger role in explaining variations in cost for later-stage patients.

Table 2.6 OLS Regressions with Time Trend (standard errors in parentheses)

	Stage I	Stage II	Stage III	Stage IV
Mean	31604.82	32132.53	34496.44	21013.42
	(1324.34)	(1278.62)	(1705.53)	(1978.15)
Trend	-1.18	187.78	267.20	452.92
	(48.55)	(46.87)	(62.52)	(72.51)
R-squared	0.000013	0.258673	0.28422	0.458902

First-year costs for Stage I patients do not have a significant time trend, so the CPI Medical Deflator accounts for the trend in spending for the least sick patients.

Next, we perform the endogenous structural breakpoint tests on the average first-year costs for each stage while allowing for a time-trend.[42] Table 2.7 records breaks in Stages II, III, and IV. However, the direction and size of the breaks are unclear. For example, the average first-year cost after the break is less than it was before the break for Stage II and Stage IV patients; but for Stage III patients, the break results in an increase in cost of $8089.76. Furthermore, the break dates no longer correspond with the introduction with irinotecan. Instead, the confidence intervals overlap around 1995, where no colorectal cancer drugs were first introduced. The time trend is still significant for Stage II and Stage IV patients, but it is no longer significant for Stage III patients.

[42] Running structural breakpoint tests on time trends yield no breaks for all stages.

These inconsistencies serve as motivation for deeper analysis of the basis of the breaks in average first-year cost. This robustness check no longer points to the introduction irinotecan as the possible culprit behind the changing cost of illness. But we do learn two things from adding the time trend. First, colorectal cancer spending outpaces general medical price inflation (as accounted for by the CPI Medical Deflator) for all but the healthiest, Stage I, patients. Secondly, there is no break in the time trend, so the increase in spending is steady over time.

Table 2.7 Breaks in Average First-year Cost with Time Trend (Coefficients with standard errors in parentheses and confidence intervals in brackets)

	Break Date	**1st Level**	**2nd Level**	**Change**	**Trend**
Stage I	--	--	--	--	--
Stage II	1994 Q1	32609.42	27254.04	-5355.38	320.04
	[1992Q3,1995Q4]	(1280.67)	(2277.99)		(69.70)
Stage III	1996 Q4	36380.77	44470.53	8089.76	14.29
	[1995Q1,1997Q4]	(1885.91)	(4488.76)		(118.78)
Stage IV	1995 Q2	24574.4	14731.62	-9842.78	623.88
	[1995Q1,1997Q4]	(2301.82)	(2692.59)		(87.52)

4.2. Robustness Check: Excluding Irinotecan Patients

To analyze the impact of the uptake of irinotecan on average first-year costs, we perform the structural breakpoint tests on the sample of colorectal cancer patients after removing those who received irinotecan treatments. If there is no break, it would provide evidence that the original breaks in Table 2.3 are indeed associated with the introduction of the new drug. However, it is also conceivable that a break still remains because the irinotecan patients are relatively small in number, so their impact cannot be large enough to be the driving force behind such breaks.

First, we exclude all patients who were given irinotecan in their lifetimes.[43] Then, we perform the endogenous structural breakpoint tests on each stage. The results are in Table 2.8.

In our new sample without irinotecan patients, the tests show breaks in Stage I, III, and IV patients. The confidence intervals around the break dates in Stages III and IV contain the original break dates from the original sample as indicated in Table 2.3. Furthermore, the absolute size of the change in average first-year cost pre- and post-break date are similar to the original analysis. So, if irinotecan patients are no longer in this sample and the breaks are still present, then there must be other reasons for the break.

[43] The drug code for irinotecan takes a value of J9206 in the HCPCS series of the NCH Medicare claims file.

Table 2.8 Breaks in Average First-year Cost (No Irinotecan Patients) (Coefficients with standard errors in parentheses and confidence intervals in brackets)

	Break Date	1st Level	2nd Level	Change	% Change
Stage I	1999 Q3	25678.84	20988.20	-4690.64	-18%
	[1997Q4, 2001Q3]	(721.59)	(1183.99)		
Stage II	--	--	--	--	--
Stage III	1996 Q3	26655.82	33709.77	7053.95	26%
	[1994Q2, 1997Q3]	(1105.29)	(1060.13)		
Stage IV	1998 Q2	18593.49	31443.66	12850.17	69%
	[1997Q1, 1999Q1]	(1212.24)	(1564.99)		

From this analysis, even more questions arise. For example, why did the break in Stage II disappear after removing irinotecan patients? The break in Stage I is also puzzling because the early stage patients are not treated with irinotecan. So, why would there be a decrease of $4,690.64 in average first-year cost among newly diagnosed Stage I patients in the 1999 Q3? However, for the Stage IV patients the structural breakpoint changes by only one quarter from the when the irinotecan patients were included, and the size of the break is similar. Clearly, something else is happening post 1998 besides the introduction of irinotecan. More analysis must be carried out in order to answer these questions. One possible avenue is changes in reimbursement or service use for specific types or subcomponents of cost.

4.3. Robustness Check: Subcomponents of Costs

In order to analyze the break, we break the cost into different sub-components, such as laboratory costs, departmental costs, and accommodation costs. The structural breakpoint tests show that departmental cost is the only sub-component with structural breaks across all stages. Table 2.9 shows the break dates, pre- and post-break levels, absolute change, and percentage change of the average first-year departmental costs. This series comes from the MEDPAR hospitalization file which records Medicare Part A claims. Specifically, this series refers to "the total charge amounted (rounded to whole dollars) for all ancillary departments (other than routine room and board, CCU, and ICU) related to a beneficiary's stay."[44]

The breakpoint in departmental cost occurs from April 2000 to January 2001 depending on the cancer stage of the patient; nevertheless, the dates themselves are not significantly different because the confidence intervals overlap. Stage II has the largest change in costs from $7,991 to $12,037, which is an increase of 51%. This $4,046 increase explains much of the $7,801 increase in total costs of Stage II patients that also occur in 2001 (see Table 2.1).

[44] See the MEDPAR Expanded Modified Record from CMS Repository. http://healthservices.cancer.gov/seermedicare/

Table 2.9 Breaks in Average First-year Departmental Cost for CRC patient (Coefficients with standard errors in parentheses and confidence intervals in brackets)

	Break Date	1st Level	2nd Level	Change	% Change
Stage I	2000 Q3	6789.98	8939.50	2149.52	32%
	[1999Q4, 2002Q4]	(199.51)	(388.92)		
Stage II	2001 Q1	7991.33	12037.38	4046.06	51%
	[2000Q2, 2001Q2]	(257.30)	(575.34)		
Stage III	2001 Q1	8501.65	11308.03	2806.38	33%
	[2000Q3, 2001Q4]	(231.30)	(517.21)		
Stage IV	2000 Q2	7462.98	11106.94	3643.96	49%
	[1999Q2, 2000Q4]	(250.47)	(459.38)		

We believe that the Stage II break in Total Costs is driven by this break in Departmental Costs. The reason it occurs only in Total Cost for Stage II is because its impact is relatively smaller for other stages. Departmental Costs only increase by $2,149 for Stage I. For Stage III and Stage IV, the sub-cost increases by $2,806 and $3,643 respectively, which is small relative to the break size in the aggregate series occurring in 1998. More research is needed to uncover the underlying reason behind the break in this subcomponent series, which is used by many researchers to calculate "actual" cost (Asper, 2005).

4.4. Robustness Check: Payment Amount Series

Using total charges as a measure of cost may overestimate actual medical payments, so we also perform the endogenous structural breakpoint tests on a cost series based on payment amounts.[45] Though total charges may represent the value of the procedure from the provider's prospective, payment amount may be a better proxy for direct medical costs incurred by society. However, the payment amounts are predetermined for certain procedures (e.g., by diagnosis-related group), so it removes much of the variation for individual patients. Therefore often the payment amount (an expected average) is even larger than the total charged amount due to the payment rules.

The payment and charge series are historically highly correlated, so there should be similar breaks in the first-year payments of colorectal patients. If the Medicare payment rules are slow to adopt new treatments, then there might be a lag in the break date or even no break if the changes are smooth.

Table 2.10 shows the results from the tests. For Stage IV patients, there is a break in the first-year payments in 2000 Q1, though the confidence interval of the breaks encompasses the original break date, 1998 Q3, of the total charges series of costs. Similarly, the break size is 48% larger than the pre-break level, which is close to the increase of 50% using the original series (See Table 2.3).

[45] For Carrier Claims, Outpatient Claims, Home Health Agency Claims, and Hospice Claims, the series is "Payment Amount." For Durable Medical Equipment, it is "Line Payment Amount." For Medicare Provider Analysis and Review, it is "Reimbursement Amount."

Table 2.10 Breaks in Average First-year Payments for CRC patient (Coefficients with standard errors in parentheses and confidence intervals in brackets)

	Break Date	1st Level	2nd Level	Change	% Change
Stage I	1992 Q3 [1991Q1, 1993Q1]	12724.16 (303.72)	10972.54 (326.76)	-1751.62	-14%
Stage II	--	--	--	--	--
Stage III	--	--	--	--	--
Stage IV	2000 Q1 [1997Q3, 2000Q4]	16566.96 (649.21)	24597.36 (2103.46)	8030.40	48%

Unlike the results based on total charges, there is a break in Stage I patients but none for Stage II and Stage III patients. Though these breaks are smaller in magnitude in both samples, this discrepancy requires additional analysis. We confirm the existence of a break in the payment series of Stage IV colorectal patients in this analysis. The results do not tell us the reason behind the break, but they also support the finding of a potential unfavorable shift in the cost-effectiveness of colorectal cancer care in the past decade.

5. Conclusion

We study the change in important components of the burden of illness of colorectal patients in the recent past to explore the impact of the introduction of new drug regimens versus other possible factors. In particular, we examine the average first-year costs recorded by Medicare Part A and Part B and first-year survival probability for newly diagnosed colorectal patients in Washington State. We use the SEER-Medicare linked database and analyze over 10,000 patients from 1991-2002. It is difficult to discern any noticeable pattern by simply analyzing the time-series plots of cost and survival, so we employ endogenous structural breakpoint tests to econometrically test for breaks.

First, our tests show indicate a break in costs in 1998, which coincides with the FDA regular approval of irinotecan. We track the uptake of irinotecan through time and confirm its potential impact on first-year costs of newly diagnosed colorectal cancer patients. However, we find that it is not likely to have been the major source of the change. Furthermore, there is no corresponding break in survival, so, if anything, the average cost-effectiveness of treatments for colorectal patients appears to have suffered a decline from 1998-2002.

Next, the cost changes coincide in time with the introduction of irinotecan. The pharmacoeconomic estimates of incremental costs based on clinical trials suggest that the use of irinotecan would definitely increase average costs. We estimate the incremental aggregate first-year real cost of treatment to be between $8,363 and $13,424 depending on the initial stage of diagnosis. Converted to current dollars, these differences would not be explained by the use of irinotecan in fewer than 10%.[46] Also, the aggregate survival estimates do not show that

[46] In 1998 dollars, the cost would be between $20,238 and $32,486.

irinotecan had a significant, measurable impact, despite clinical trials showing improvements on the order of 20 percentage points due to irinotecan.

We perform a series of robustness checks to further explore the source of the breaks. First, adding a time-trend proves to be significant, but it does not remove the break from the cost series. The resulting breaks, though, no longer match up with the uptake of irinotecan. By removing the irinotecan patients, our second robustness check clearly shows a break not associated with the new treatment drug. An analysis of a subcomponent of costs—departmental cost—shows an underlying break, which might illuminate the reason behind the changing costs and cost-effectiveness of care. Lastly, we use payment amount as a measurement of cost to confirm the existence of a break in Stage IV patients.

The next step would be to study potential changes due to Medicare changes in reimbursement schemes or other procedural shifts. Also, the choice of price deflators and control samples may have an impact on the break date. Another extension of this study would be to examine the effects of new drug regimens on an individual-level to study the direct effects of specific drug regimens.

Due to the limitations of data, there are a few obvious shortcomings of our research. For example, our analysis only goes through 2002, so we cannot account for the effects of capecitabine, oxaliplatin, cetuximab, bevacizumab, or any other drug introduced within the last 5 years.[47] Nevertheless, we believe our characterization of average cost and survival in first-year patients is accurate through the end of our sample. A second notable drawback due to our data availability is that we are limited to Washington State. There is the possibility that regional differences could affect the uptake of irinotecan and the costs associated with the treatment. A wider geographic area must be used before we can generalize our result to encompass the entire United States.

[47] Though capecitabine was approved in 2001 for 1st line and 2nd line treatment in metastatic colorectal patients, the "trimming" criteria of endogenous structural breakpoint tests does not allow us to include it in our testable sample.

Chapter III

Endogenous Structural Breakpoint Tests: An Investigation in Finite Sample Performance

(with Jun Ma)

1. Introduction

In recent decades, much research in the fields of econometrics and statistics has been done on the subject of structural parameter breakpoints. Though exhaustive in its theoretical treatment of the issue, the validity of the proposed procedures in practice involving finite samples is still up for question. The obvious reason that the interest exists for this topic is that economic theory often formulates parameters and characteristics within their models. For example in macroeconomics, one estimates the NAIRU or the marginal propensity to consume, and in microeconomics, the returns to schooling, which could have drastic implications in their respective models should those parameters shift overtime. The major shortcoming of the available theory for structural breakpoint tests is that they are based on asymptotic conditions, which clearly are not viable in practice. Moreover, there is no reason to assume that a finite sample of, say, 50 observations should behave in the same manner as its asymptotic counterpart.

In this paper, we explore the finite-sample properties of the asymptotic structural breakpoint tests, namely, the supremum tests proposed in Andrews

(1993). The theory also suggests that the tests hold asymptotically for linear and nonlinear models, but again their finite sample behavior has not been touched. Certain parameters, such as the NAIRU, are naturally estimated in nonlinear form but are estimated indirectly via linear regressions for tractability (see Ball and Mankiw (2002)).

Though it seems plausible that processing a small-sample of data in structural breakpoint tests could yield differing results depending on whether or not the equations are in linear or nonlinear form, we show that there is no noticeable difference for regressions that can be transformed between the forms. This result is consistent with the asymptotic theory. There are major differences in the performance of differing forms of the supremum tests in finite samples, which are all equivalent asymptotically in theory. Lastly, we explore the merits and limitations of bootstrapping critical values versus merely using the asymptotic critical values that can be calculated in closed form.

The literature on structural breakpoint tests has been numerous with extensive survey papers already appearing decades ago: Krishnaiah and Miao (1988) and Zacks (1983). Starting from Hansen (1990) and Andrews (1993), the theory behind these procedures has been focused on the asymptotics, such as in Andrews (2002), Bai and Perron (1998), Chu et al. (1996), Diebold and Inoue (2000), Gagliardini et al. (2003), and Perron and Qu (2004). Though they provide a solid foundation on the subject, these above papers mention little regarding the practicality of their proposed methods in finite samples.

Aside from Andrews (1993), Andrews (2002), and Hansen (1990), the literature deals with the linear properties of the tests since the nonlinear versions are identical. Recently, extending the original theory to allow for multiple breakpoints has also been tackled in Andrews (2002), Bai and Perron (1998), and Perron and Qu (2004).

When applying these methods in practice, the majority of authors merely uses the asymptotic properties and assumes that they hold in the finite samples. Diebold and Chen (1996) propose bootstrapping the critical values for the supremum tests and show that they provide greater size than the asymptotic values do. Nonetheless, this practice remains scarce, though Basci et al. (2000) bootstrap critical values in their research that applies the structural breakpoint tests in the Turkish financial markets.

The last major niche of this breakpoint literature has to do with the characteristics of the data series themselves. Zivot and Andrews (1992) show that these asymptotic tests do not hold if there are unit roots (and that the unit root tests do not hold in the presence of structural breaks). Diebold and Inoue (2000) explore the issue of long-memory in data series. Since there has been no clear answer to this caveat in using these tests, we will use stationary data in our simulations.

Section 2 details the procedure and methodology behind our simulations of these breakpoint tests in finite samples. We present and analyze the results in Section 3, which is further divided into 6 subsections for each of our findings. Since we find that bootstrapping critical values have the best performance in finite samples, Section 4 applies this methodology to the NAIRU, and we compare the results with those using the traditional asymptotic critical values. Lastly, a conclusion follows in Section 5 where we propose an extensive set of extensions that follow from our results.

2. Procedure

We specify our model under the null of no breaks as follows:

Nonlinear – $y_t = g(x, \alpha, \beta_N) = \alpha(x_t + \beta_N) + u_t$

Linear – $y_t = f(x, \alpha, \beta_L) = \alpha x_t + \beta_L + u_t$

Here, $\beta_L = \alpha\beta_N$

Notice that the equations for the linear and nonlinear forms represent the same relationship between the regressors, regressands, and parameters. This simulates the problem often faced in practice where the econometrician estimates β_N indirectly by using the linear equation by simply calculating β_L/α. Our model allows us explore the variation of the test results for equations that can be written in linear and nonlinear representations. Note that x_t*'s* and y_t*'s* are observed data values and the α and the β*'s* are parameters to be estimated. u_t*'s* are shocks to the system. The data generating process that we use for the model is as follows:

$\{x_t\}$ are *i.i.d.* random numbers uniformly distributed between [0,100].

$\{u_t\}$ are *i.i.d.* random numbers normally distributed with zero mean and unit variance.

As expected, the model changes under the alternative to include a break. This is necessary to explore the power characteristics of the finite sample tests.[48] In evaluating the power, we impose a break on β_L by introducing a dummy variable D, which is 0 before our breakdate and 1 after the break. Thus, the model under the alternative changes to the following:

$$\text{Nonlinear} - \; y_t = G(x, \alpha, \beta_N, \Delta\beta_N) = \alpha(x_t + \beta_N + \Delta\beta_N \cdot D) + u_t$$

$$\text{Linear} - y_t = F(x, \alpha, \beta_L, \Delta\beta_L) = \alpha x_t + \beta_L + \Delta\beta_L \cdot D + u_t$$

Here, $\Delta\beta_N$ and $\Delta\beta_L$ are the size of the breaks in the parameters for the nonlinear and linear models respectively.

In formulating the test statistics, we adhere to the common practice of using the supremum Wald-type (SupF), Lagrange Multiplier-type (SupLM), and Likelihood Ratio-type tests (SupLR).[49] Though theory also puts forth exponential and average versions of these test statistics, the supremum ones are most

[48] Diebold and Chen (1996) has a detailed analysis of the size of the structural breakpoint tests in finite samples and give preliminary results which state that power should be an easy property to extend from their findings.

[49] We use 5% significance in our analyses.

commonly used in practice, so they are of the most interest to us.[50] Research has also shown that the data must be truncated to allow for accuracy and feasibility of testing for these breaks. Andrews (2002) explores shrinking the truncation percentage, but we use the common practicing of trimming 30% of the data, so our truncation is set to be 0.15 and the range to be tested for the structural break is [t1, t2], where, t1 is [0.15*T], t2 is [0.85*T], and T is the entire sample.

$$\text{supF} = \sup_{[t1,t2]} T \cdot \frac{RSSR - USSR}{USSR} \text{ in both linear and nonlinear cases}$$

Here, RSSR and USSR are the restricted sum of squared residuals and unrestricted sum of squared residuals respectively for both linear and nonlinear cases. (See Diebold and Chen (1996) or Engle (1984))

$$\text{supLM} = \sup_{[t1,t2]} T \cdot \frac{RSSR - USSR}{RSSR} \text{ for linear case;}$$

$$= \sup_{[t1,t2]} T \cdot \frac{\hat{u}' x^* (x^{*\prime} x^*)^{-1} x^{*\prime} \hat{u}}{RSSR} \text{ for nonlinear case;}$$

here, $\hat{u}$ is the estimated residual in the restricted model and x^* is the gradient vector of $G(x, \alpha, \beta_N, \Delta\beta_N)$ evaluated at the restricted estimators.[51] (See Greene (2003))

$$\text{supLR} = \sup_{[t1,t2]} T \cdot \log(RSSR/USSR) \text{ for both cases.}$$

---

[50] It follows that the exact same simulations can be implemented for these various types of tests.

[51] The LM-style test for nonlinear has been advocated by many because one needs only to estimate one equation thereby saving much computing time.

The bootstrapping procedure we implement involves an extra step since we are interested in the supremum values of each loop and not the individual values. We bootstrap 100 times for each simulation.[52] For both the linear and nonlinear cases, we simulate the data series 100 times to start; then, after finding that even at such small simulations, the linear and nonlinear results are identical, we simulate the linear case 1000 times (the nonlinear case is not done at this higher simulation because it is highly computer intensive).

A detailed diagram of the flow of the bootstrap technique is included (Figure 1), but the general concept can be outlined as follows. We start with a generated data series each time that is simulated under the null (to test for size) or under the alternative (to test for power). Then, we estimate the nonlinear and linear models under both the unrestricted and restricted forms to get the necessary information to compute the three supremum test statistics for structural breakpoints. Using the estimated coefficients and residuals, the semi-parametric

[52] See Hall (1986) for a theoretical explanation of the ability of the bootstrap to produce satisfactory results with few replications of the bootstrap sampling process.

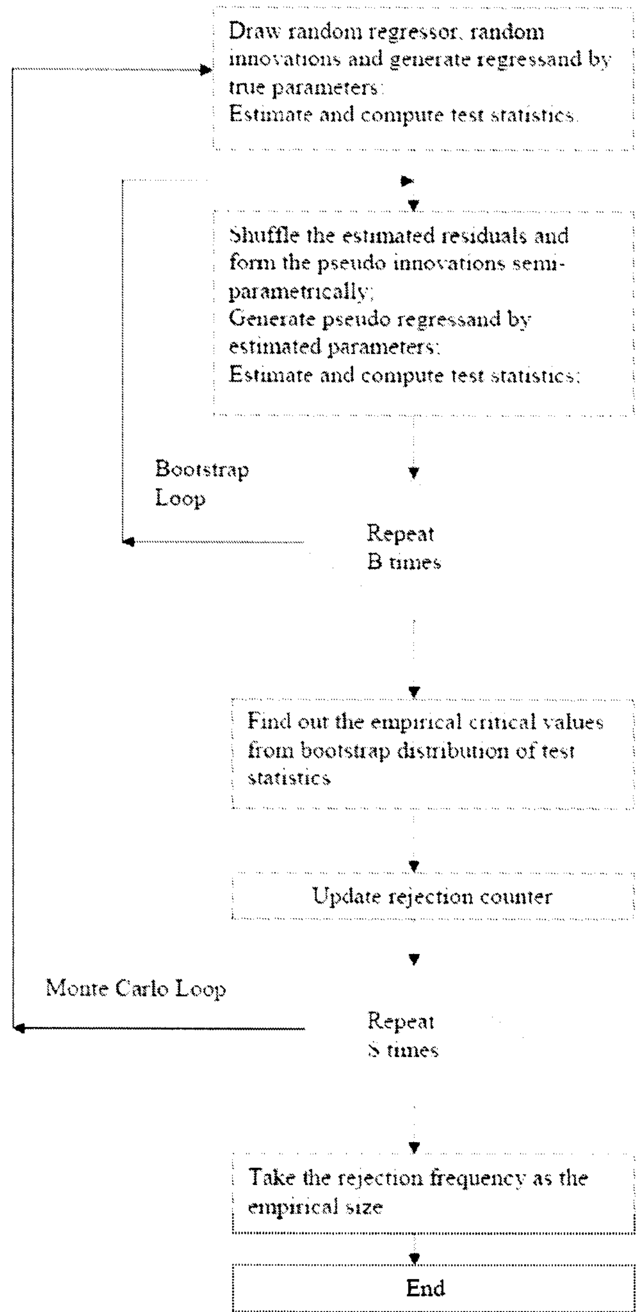

Figure 3.1 Flow Chart for Bootstrap Monte Carlo

bootstrap method is implemented to bootstrap the critical values for all three test statistics.

With the model, data generating process, and bootstrap all specified, we proceed to test both the asymptotic and the bootstrapped size of each test statistic. This serves to double-check our results with Diebold and Chen (1996) that exalt the bootstrap's size in the linear context and to see whether or not it holds true in the nonlinear framework. Next, we test both the asymptotic and bootstrapped power of each test statistic. Lastly, we change the data generating process to see how robust the methods are when the break size varies. We change the break size to error variance (BE) ratio from 0.25, 1, to 4 to test the effect on the power for the asymptotic and bootstrapped test statistics.

Our programs are written in Mathwork's Matlab 6.5 and incorporate the maximization and optimization algorithms of the software necessary for nonlinear estimation. They are then run on a cluster of DEC and IBM UNIX workstations at the University of Washington, which use AIX 5.2 as their operating system.

3. Results and Analysis

The major findings from the simulations outlined in the previous section can be divided into 6 subsections.

i. Equivalence of forms

A question that often arises is whether or not the structural breakpoint tests are robust when an equation can be written identically in linear and nonlinear forms. Which form should one use? In our model, where the parameters could be estimated directly in nonlinear form or indirectly in linear form, we find that the actual form makes no noticeable difference in the results. The asymptotic tests of power and size for each of the three supremum tests yield the identical values in linear and nonlinear form. This result is a relief since it can easily be shown that the algebraic forms of the critical values should be the same in small samples. From our model, the estimation of the coefficients in the linear model are:

$$\hat{\alpha} = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}$$

$$\hat{\beta}_L = \bar{y} - \hat{\alpha}\bar{x} = \bar{y} - \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}\bar{x}$$

To get the estimation of the nonlinear model, we first minimize the following:

$$\text{Min} \sum (y_t - \alpha x_t - \alpha\beta_N)^2$$

Take the first order condition and simplify to get:

$$\hat{\alpha} = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}$$

$$\hat{\alpha}\hat{\beta}_N = \bar{y} - \hat{\alpha}\bar{x} = \bar{y} - \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}\bar{x} = \hat{\beta}_L$$

This indicates that $\hat{u}_L = \hat{u}_N$, so $USSR_L = USSR_N$ and $RSSR_L = RSSR_N$. On a practical level, the ability to find robust results from linear forms saves immense computing power. For example, merely bootstrapping a sample of 50 critical values for a nonlinear model took nearly 20 hours, while bootstrapping its linear form took no more than 5 minutes. We propose that this equivalence always holds for any model that can be written in both linear and nonlinear forms. Therefore, all of the subsequent tables and results in this paper will, thus, show that the results of the linear and nonlinear models are the same.[53]

ii. Performance of different tests

Asymptotically, the SupF, SupLR, SupLM tests yield the exact same results, thus having the same performance. Our simulations for small samples give us a clear sense of hierarchy when ranking the tests for size and power using the asymptotic critical values. Tables 3.1-3.3 provide the results of our simulations for size when varying the sample size at 10, 25, and 50. Since we use the 5% critical value for 1 break from Andrews (1992), we expect a rejection rate given no breaks to be around 5%. In all the various sample sizes, the simulations show that the rejection rate was highest for the SupF tests and lowest for the

[53] When bootstrapping 1000 times, there are slight variations in the results. See Section III.3 for further information

Table 3.1 Test of size for Sample Size=10 (based on 100 simulations)

	Sample Size = 10		
Linear		Asymptotic	Bootstrap
	supF	0.24	0.08
	supLM	0	0.08
	supLR	0.11	0.08
Nonlinear		Asymptotic	Bootstrap
	supF	0.24	0.08
	supLM	0	0.08
	supLR	0.11	0.08

Table 3.2 Test of size for Sample Size=25 (based on 100 simulations)

	Sample Size = 25		
Linear		Asymptotic	Bootstrap
	supF	0.09	0.04
	supLM	0.01	0.04
	supLR	0.03	0.04
Nonlinear		Asymptotic	Bootstrap
	supF	0.09	0.04
	supLM	0.01	0.04
	supLR	0.03	0.04

Table 3.3 Test of size for Sample Size=50 (based on 100 simulations)

	Sample Size = 50		
Linear		Asymptotic	Bootstrap
	supF	0.07	0.07
	supLM	0.04	0.07
	supLR	0.05	0.07
Nonlinear		Asymptotic	Bootstrap
	supF	0.07	0.07
	supLM	0.04	0.07
	supLR	0.05	0.07

SupLM tests with the SupLR tests lying somewhere in between. In Table 3.1 where the sample size is 10, we see that SupF rejects the null of no breaks nearly 25% of the time and SupLM never rejects the null. As the sample sizes increase, the rejection rates for all three tests become closer together. In Table 3.3, where the sample size is 50, the difference in the rejection rates of the SupF and the SupLM is only 3%. Continuing with this pattern, we expect the variation in the rejection rates between the different supremum tests to disappear altogether as theorized in asymptotic theory.

When we test for power, our model has a break and we simulate the rejection rate of the model under this alternative form. Ideally, the tests should pick up the break, so the rejection rate should be high. In addition to the original model where the BE ratio is 1, we vary the BE ratio to include 0.25 and 4 in our simulations. Nonetheless, in the 9 possible simulations that arise—since we have 3 sample sizes and 3 BE ratios—the rejection rates can all be ranked as follows: SupF, SupLR, and SupLM. Note that this order is the same as for the simulations for size. Again, the variation of the rejection rates for the tests diminishes as the sample sizes increase as indicated in Tables 3.4-3.12.

Table 3.4 Test of power for Sample Size=10 and BE Ratio=0.25

	Sample Size = 10		
	Break size/error ratio = 0.25		
Linear		Asymptotic	Bootstrap
	supF	0.25	0.06
	supLM	0	0.06
	supLR	0.14	0.06
Nonlinear		Asymptotic	Bootstrap
	supF	0.25	0.06
	supLM	0	0.06
	supLR	0.14	0.06

Table 3.5 Test of power for Sample Size=10 and BE Ratio=1

	Sample Size = 10		
	Break size/error ratio = 1		
Linear		Asymptotic	Bootstrap
	supF	0.42	0.14
	supLM	0	0.14
	supLR	0.21	0.14
Nonlinear		Asymptotic	Bootstrap
	supF	0.42	0.14
	supLM	0	0.14
	supLR	0.21	0.14

Table 3.6 Test of power for Sample Size=10 and BE Ratio=4

	Sample Size = 10		
	Break size/error ratio = 4		
Linear		Asymptotic	Bootstrap
	supF	1	0.94
	supLM	0.33	0.94
	supLR	0.99	0.94
Nonlinear		Asymptotic	Bootstrap
	supF	1	0.94
	supLM	0.33	0.94
	supLR	0.99	0.94

Table 3.7 Test of power for Sample Size=25 and BE Ratio=0.25

	Sample Size = 25		
	Break size/error ratio = 0.25		
Linear		Asymptotic	Bootstrap
	supF	0.11	0.09
	supLM	0.03	0.09
	supLR	0.08	0.09
Nonlinear		Asymptotic	Bootstrap
	supF	0.11	0.09
	supLM	0.03	0.09
	supLR	0.08	0.09

Table 3.8 Test of power for Sample Size=25 and BE Ratio=1

	Sample Size = 25		
	Break size/error ratio = 1		
Linear		Asymptotic	Bootstrap
	supF	0.59	0.51
	supLM	0.37	0.51
	supLR	0.50	0.51
Nonlinear		Asymptotic	Bootstrap
	supF	0.59	0.51
	supLM	0.37	0.51
	supLR	0.50	0.51

Table 3.9 Test of power for Sample Size=25 and BE Ratio=4

	Sample Size = 25		
	Break size/error ratio = 4		
Linear		Asymptotic	Bootstrap
	supF	1	1
	supLM	1	1
	supLR	1	1
Nonlinear		Asymptotic	Bootstrap
	supF	1	1
	supLM	1	1
	supLR	1	1

Table 3.10 Test of power for Sample Size=50 and BE Ratio=0.25

	Sample Size = 50		
	Break size/error ratio = 0.25		
Linear		Asymptotic	Bootstrap
	supF	0.12	0.12
	supLM	0.08	0.12
	supLR	0.12	0.12
Nonlinear		Asymptotic	Bootstrap
	supF	0.12	0.12
	supLM	0.08	0.12
	supLR	0.12	0.12

Table 3.11 Test of power for Sample Size=50 and BE Ratio=1

	Sample Size = 50		
	Break size/error ratio = 1		
Linear		Asymptotic	Bootstrap
	supF	0.80	0.78
	supLM	0.72	0.78
	supLR	0.77	0.78
Nonlinear		Asymptotic	Bootstrap
	supF	0.80	0.78
	supLM	0.72	0.78
	supLR	0.77	0.78

Table 3.12 Test of power for Sample Size=50 and BE Ratio=4

	Sample Size = 50		
	Break size/error ratio = 4		
Linear		Asymptotic	Bootstrap
	supF	1	1
	supLM	1	1
	supLR	1	1
Nonlinear		Asymptotic	Bootstrap
	supF	1	1
	supLM	1	1
	supLR	1	1

iii. Equivalence in the bootstrap

In each of the simulations for size and power, we calculate the rejection ratio using the asymptotic critical values and using the bootstrapped critical values. While there is a clear ranking for the rates of rejection using the asymptotic values (see Section 3, Part ii), the rejection variation between the bootstrapped Wald-type, LR-type, and LM-type supremum critical values is completely eliminated.[54] This suggests that when using the bootstrap to generate critical values in small samples, there is no need to use all of the various supremum tests. The SupLM test, which involves only estimating 1 equation, is sufficient in providing accurate results for the bootstrapped versions.

Though the variation is eliminated, it does not mean that the performance of the bootstrapped critical values do not leave more to be desired. In the tests for power, for example, the rejection rate is not always very high, so there may be under rejection. But, in scenarios such as those depicted in Table 3.6—where there is a small sample size and a large break size to error variance ratio—the bootstrapped critical values take the guesswork out of which test to use since the asymptotic tests results vary greatly from depending on the form of the supremum test used.

[54] We increased the bootstrap number to 1000 to see if we could detect any variations. There is a small variation in the bootstrap rejection values for some of the tables, but they remain nearly identical. See Tables 3.13-3.24.

iv. Bootstrapping for Size

When the sample size is very small, say 10, the simulations for size based on the asymptotic critical values are all over the place. SupF rejects 24% of the time and SupLM never rejects (See Table 3.1). Neither result is satisfactory when it comes to giving the econometrician confidence of his results. As mentioned in Section 3.iii, the bootstrapped versions of the simulations are all the same and, in the case of the sample size being 10, has decent size with 8% rejection. This indicates that when one is faced with a very small sample of data but wishes to find a structural break, bootstrapping the critical values is most reliable. As we increase the sample size to 25 in Table 3.2, the variance of the rejection rate diminishes for the asymptotic simulations, but the 4% rejection rate exhibited by the bootstrapped simulation shows that it has better performance. When the sample size increases to 50, the size of the test improves for the asymptotic version and the advantages the bootstrap had in the smaller samples decreases.

v. Relative Power of Bootstrap Unclear

The advantage of the power exhibited by the bootstrap is not as evident as its advantage in size over its asymptotic counterparts. As shown in the tables, the rejection rate using the bootstrapped critical values fall within the range of the rates given by the simulations using the asymptotic values. In a small sample size of 10, the power of the bootstrap is below that of SupF and SupLR but above that of SupLM (see Tables 3.1-3.3). As the sample size increases to 25 and 50, the power of the bootstrap is still lower than that of SupF and above that of SupLM, but it is nearly identical to that of SupLR, which falls in between the other two supremum tests as indicated in Section 3, Part ii.

While the advantage in size that the bootstrapped critical values possess in the tests for structural change has already been shown in Diebold and Chen (1996), the power result is unexpected. Unfortunately, we cannot say that the power of the bootstrap is without doubt greater than using the asymptotic critical values. In the cases of small samples, the SupF criterion has the best power and poor size due to its tendency to over reject.

vi. Tradeoff in Break Size to Error Variance Ratio and Sample Size

When testing for the power of the different types of critical values, we varied the BE ratio as well as the sample sizes. As expected, the power of the tests increase as the sample size increases to a point where the rejection rate is near 100% as shown in Tables 3.9 and 3.12. Nevertheless, the deciding factor in whether or not the tests have good power is not just the sample size but, rather, the BE ratio. A small ratio would indicate that breaks could be mistaken for shocks to the system, which are of a greater magnitude. As intuition would suggest, even if we impose a small break in a large sample, the power of the test should be small. Tables 3.4, 3.7, and 3.10 show that the power of the test given a

BE ratio of 0.25 improves only minimally as the sample size increases from 10 to 25 to 50.

If the BE ratio grows to 1, it means that the size of the break is only as big as the variance of the error term. In this case, though, the gains in power as sample size increases are much more substantial. The bootstrapped version for sample size 10 has a power of only 14%, but when the sample size is increased to 50, the power increases to 78%. We arrive at the same results using the asymptotic power results because, as suggested in Section 3.v, the bootstrap power is representative of the power of all 3 supremum tests on average.

Once the BE ratio becomes large, such as 4, the power of the bootstrapped critical values in the small sample size is still 94% (Table 3.6). With the large ratio, the rejection rate of the null when there is a break is 100% for both asymptotic and bootstrapped critical values when the sample size is increased to 25 (Table 3.9) and 50 (Table 3.12). These results indicate that if one expects a very small BE ratio, then it is necessary to have a larger amount of data in order for the test to have good power. However, if the break is large in itself, then a very small sample size already exhibits good power.

4. Application: NAIRU

There have been many papers devoted to estimating the NAIRU in the United States, most of which assume that this measure of the "natural rate" of unemployment has not changed post World War II. When the endogenous structural breakpoint tests of Andrews (1993) are programmed by Bai and Perron (1998), they use the asymptotic critical values in their analysis. We can estimate the equation in Section II with β_N as the NAIRU and α as the coefficient if x_t is the unemployment rate in year t and y_t is the change in inflation calculated from the CPI. Using a sample size of 50, we proceed to run the asymptotic program in Gauss by Bai and Perron (1998) and see that there is no structural break.

Our research suggests that we must be careful in using asymptotic theory when working with finite samples for the structural breakpoint tests. So, we use the same data and bootstrap our critical values in the same manner as described in Section II. Keeping in mind that the 5% asymptotic critical value is 8.85, the bootstrapped values for each of our 3 critical values are as follows:

SupF: 8.8883
SupLM: 7.6113
SupLR: 8.2168

The results are consistent with the asymptotic values, so we still cannot reject the null that there is a no break. This is not surprising since we have 50 years of data with which to work; our analysis in Section III shows that once we reach a sample size of 50, the asymptotic critical values have decent size and power since the bootstrapped critical values begin to resemble them.

However, if we use a smaller sample size, the benefits of having bootstrapped critical values are evident. Since our program allows for only 1 break at a time, it is conceivable that during such a long duration (50 years), there could have been multiple breaks, which makes our estimation biased. We can revise the question to ask whether or not there is a shift in the NAIRU in the 1990s. Then, we will use only data after the post-Volcker monetary policy influence, which could have been a shock in itself, to cut our data sample down to include just the past 20 years (1983-2003) where we expect no more than 1 break. As before, the 5% asymptotic critical value is 8.85, but, now, the bootstrapped critical values are as follows:

SupF: 12.7962
SupLM: 7.9482
SupLR: 9.9899

The respective supremum test statistics for each test are not large enough to reject the hypothesis of no break:

SupF: 11.517
SupLM: 7.4379
SupLR: 9.1821

If the researcher chooses simply to use the asymptotic critical values, we see that the test statistics for SupF and SupLR would reject the null (suggesting a break in 1998). But, in a sample size of 20, the bootstrapped critical values have more power and better size, so one must apply great caution drawing conclusions from the asymptotic critical values. In this case, the conclusions of the bootstrapped critical values differ from those using the asymptotic ones.

5. Conclusion

Our findings indicate that, even in finite samples, there is no difference in the performance of structural breakpoint tests when the model is written in linear or in nonlinear form. This is supported by the mathematical equivalency of the test statistics that hold even in small samples. When using asymptotic critical values derived from Brownian Bridges (see Andrews (1993)), one must use caution because in small samples, SupF always rejects more often than SupLR, which rejects more often than SupLM. The differences in rejection rates are eliminated if the critical values are bootstrapped. We show the size advantage that the bootstrapped values have over the asymptotic values but fail to see any advantage in power for either method. Lastly, when dealing with finite samples, there is a clear tradeoff in the test performance between the BE ratio and the sample size.

Though we explore many issues regarding structural breakpoint tests in finite samples, many possible extensions still exist. First and foremost, the analysis can easily be extended to allow for multiple breaks. The bootstrap technique necessary to do so has been proposed in a few papers, such as in Banerjee et al. (2002). Moreover, we can adopt another test for power, which Andrews and Ploberger (1994) dubs as the "optimal power test." Finally, there is the issue of the behavior of these test statistics in the presence of unit roots and other non-stationary processes.

The simulations of small sample behavior indicate that one needs only to bootstrap a single type of supremum test statistic in its linear form. If we use the SupLM test, then we only need to estimate the equation once for each bootstrap; this clearly is more advantages than bootstrapping all of the test statistics in their nonlinear forms! We caution, though, that bootstrapping critical values can only do so much in small samples. If the BE ratio is small, then its power is still weak. In that case, we recommend a sample size of at least 50 observations. On the other hand, if there is reason to suspect that the break in question has a large break-error ratio, then a sample size as small as 10 will offer sufficient power when bootstrapped critical values are used.

BIBLIOGRAPHY

Alba, J., Papell, D., 2007. Purchasing Power Parity and Country Characteristics: Evidence from Panel Data Tests, Journal of Development Economics 83(1), 240-251.

Amara, J., Papell, D., 2004. Testing for Purchasing Power Parity Using Stationary Covariates, working paper, University of Houston.

Andrews, D., 1993. Tests for Parameter Instability and Structural Change With Unknown Change Point, Econometrica 61, 821-856.

Andrews, D., 2002. End-of-Sample Instability Tests, manuscript, Cowles Foundation for Research in Economics, Yale University.

Andrews, D., and Ploberger, W., 1994. Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative, Econometrica 62, 1383-1414.

Asper, F., 2005. Using Medicare Hospital Cost Reports Cost-to-Charge Ratios in Research, ResDAC, TN-008.

Bai, J., 1998. A Note on Spurious Break, Econometric Theory 14, 663-669.

Bai, J., and Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes, Econometrica 66, 47-78.

Ball, L., and Mankiw., N.G., 2002. The NAIRU in Theory and Practice, manuscript, Department of Economics, Johns Hopkins University.

Basci, S., Basci, E., and Zaman., A., 2000. A Method for Detecting Structural Breaks and an Application to the Turkish Stock Market, manuscript, Department of Economics, Bilkent University.

Banerjee, A., Lazarova, S., and Urga., G., 2002. Bootstrapping Sequential Tests for Multiple Structural Breaks, manuscript, European University Institute (Florence, Italy).

Banerjee, A., Lumsdaine, R., and Stock, J., 1992. Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence, Journal of Business & Economic Statistics 10 (3), 271-287.

Benigno, G., 2004. Real exchange rate persistence and monetary policy rules, Journal of Monetary Economics 51, 473-503.

Bergman, U.M., Hansson, J.,2000. Real Exchange Rates and Switching Regimes, working paper, Lund University.

Brown, ML., Riley, GF., et al, 1999. Estimating Health Care Costs Related to Cancer Treatment from SEER-Medicare Data, Medical Care 40 (8), IV104-IV117.

Brown, ML., Riley, GF., et al, 1999. Obtaining Long-Term Disease Specific Costs of Care: Application to Medicare Enrollees Diagnosed with Colorectal Cancer, Medical Care 37 (12), 1249-1259.

Brown, ML., Lipscomb, J., and Snyder, C., 2001. The Burden of Illness of Cancer: Economic Cost and Quality of Life, Annual Review of Public Health 22, 91-113.

Canzoneri, M, Cumby, R., Diba, B., 1999. Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence For a Panel of OECD Countries, Journal of International Economics 47, 245-266.

Chen, S.S., Engel, C., 2005. Does 'Aggregation Bias' Explain the PPP Puzzle? Pacific Economic Review 10 (1), 49-72.

Cheung, Y.W., 1993. Long Memory in Foreign-Exchange Rates, Journal of Business & Economic Statistics 11, 93-101.

Cheung, Y.W., Lai, K.S., 1993. A Fractional Cointegration Analysis of Purchasing Power Parity, Journal of Business & Economic Statistics 11, 103-112.

Cheung, Y.W., Lai, K.S., 1997, Parity Reversion in Real Exchange Rates During the Post-Bretton Woods Period, working paper, University of California, Santa Cruz.

Cheung, Y.W., Lai, K.S., 2000a. On The Purchasing Power Parity Puzzle, Journal of International Economics 52, 321-330.

Cheung, Y.W., Lai, K.S., 2000b. On Cross-Country Differences in the Persistence of Real Exchange Rates, Journal of International Economics 50 (2), 375-397.

Chow, G., 1960. Tests of Equality between sets of coefficients in two linear regressions, Econometrica 28(7), 597-605.

Chu, C.S., Stinchcombe, M., and White, H., 1996. Monitoring Structual Change, Econometrica 64, 1045-1065.

Clarida, R., Gali, J., 1994. Sources of Real Exchange Rate Fluctuations: How Important are Nominal Shocks? NBER Working Paper 4658.

Cunningham, D., Humblet, Y., Siena, S., et al., 2004. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer, New England Journal of Medicine 351(4):337-345.

Delea, T., et al., 2002. The Incidence and Cost of Hospitalization for 5-FU Toxicity among Medicare Beneficiaries with Metastatic Colorectal Cancer, Value In Health 5(1), 35-43.

Diebold, F., and Chen., C., 1996. Testing Structural Stability with Endogenous Break Point: A Size Comparison of Analytic and Bootstrap Procedures, manuscript, Department of Economics, University of Pennsylvania.

Diebold, F., Husted, S., Rush, M., 1991. Real Exchange Rates under the Gold Standard, The Journal of Political Economy 99 (6), 1252-1271.

Diebold, F., and Inoue., A., 2000. Long Memory and Regime Switching, manuscript, Department of Economics, University of Pennsylvania.

Duggan, M. and Evans, W., 2005. Estimating the Impact of Medical Innovation: The Case of HIV Antiretroviral Treatments, NBER Working Paper 11109.

Engel, C., 1994. Can the Markov Switching Model Forecast Exchange Rates, Journal of International Economics 36, 151-165.

Engel, C., 2000. Long-Run PPP May Not Hold After All, Journal of International Economics 51(2), 243-273.

Engel, C., Hamilton, J., 1990. Long Swings in the Dollar: Are They in the Data and Do Markets Know It? The American Economic Review 80 (4), 689-713.

Engel, C., Kim, C.J., 1999. The Long-Run US/UK Real Exchange Rate, Journal of Money, Credit, and Banking 31 (3), 335-356.

Engel, C., West, K., 2004. Exchange Rates and Fundamentals, Journal of Political Economy 113 (3), 485-517.

Engle, R., 1984. Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, *Handbook of Econometrics*, 11.

Engle, R., Smith, A.,1999. Stochastic Permanent Breaks, The Review of Economics and Statistics 81 (4), 553-574.

Etzioni, R., Ramsey, S., et al, 2001. The Impact of Including Future Medical Care Costs When Estimating the Costs Attributable to a Disease: A Colorectal Cancer Case Study, Health Economics 10, 245-256.

Flood, R., Rose, A., 1995. Fixing Exchange Rates: A Virtual Quest for Fundamentals, Journal of Monetary Economics 36, 3-37.

Frankel, J., Rose, A.,1996. A Panel Project On Purchasing Power Parity: Mean Reversion Within and Between Countries, Journal of International Economics 40, 209-224.

Friedman, M., Schwartz, J.,1963. A Monetary History of the United States. Princeton, New Jersey: Princeton University Press.

Frommel, M., MacDonald, R., Menkhoff, L., 2005. Markov Switching Regimes in a Monetary Exchange Rate Model, Economic Modelling 22 (3), 485-502.

Gagliardini, P., Trojani, F., and Urga, G., 2003. Robust GMM Tests for Structural Breaks, manuscript, Institute of Finance, University of Lugano.

Greene, W. 2003. *Econometric Analysis*, Fifth Edition, New Jersey: Pearson Hall.

Grilli,V., Kaminsky, G., 1991. Nominal Exchange Rate Regimes and the Real Exchange Rate: Evidence from the United States and Britain, 1885-1986. Journal of Monetary Economics (27), 191-212.

Hall, P., 1986. On the Number of Bootstrap Simulations Required to Construct a Confidence Interval, The Annals of Statistics 14, 1453-1462.

Hall, P., and Joel Horowitz 1996. Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators, Econometrica 64, 891-916.

Hamilton, J., 1994. Time Series Analysis. Princeton, New Jersey: Princeton University Press.

Hansen, B., 1990. Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models, manuscript, Department of Economics, University of Rochester.

Hegwood, N., Papell, D., 1998. Quasi Purchasing Parity Puzzle, International Journal of Finance and Economics 3, 279-289.

Hershman, D., Jacobson, JS., et al, 2004. Effectiveness of platinum-based chemotherapy among elderly patients with advanced ovarian cancer, Gynecologic Oncology 94, 540-549.

Hillner, B., Schrag, D., et al., 2005. Cost-effectiveness projections of oxaliplatin and infusional fluorouracil versus irinotecan and bolus fluorouracil in first-line therapy for metastatic colorectal carcinoma, Cancer 104(9), 1871-84.

Imbs, J., Mumtaz, H., Ravn, M., Rey, H., 2003. Nonlinearities and Real Exchange Rate Dynamics, Journal of the European Economic Association 1, 639-649.

Imbs, J., Mumtaz, H., Ravn, M., Rey, H., 2005a. PPP Strikes Back: Aggregation and the Real Exchange Rate, The Quarterly Journal of Economics 120 (1), 1-43.

Imbs, J., Mumtaz, H., Ravn, M., Rey, H., 2005b. 'Aggregation Bias' DOES Explain the PPP Puzzle, NBER Working Paper 11607.

Iveson, T., et al., 1999. Irinotecan in second-line treatment of metastatic colorectal cancer: improved survival and cost-effect compared with infusional 5-FU, European Journal of Cancer 35(13), 1796-1804.

Iwashyna, T., and Lamont, E., 2002. Effectiveness of Adjuvant Flourouracil in Clinical Practice: A Population-Based Cohort Study of Elderly Patients with Stage III Colon Cancer, Journal of Clinical Oncology 20 (19), 3992-3998.

Jemal, A., et al., 2005. Cancer Statistics, A Cancer Journal for Clinicians 55, 10-30.

Krishnaiah, P. and Miao, B.Q., 1988. Review about Estimation of Change Points, *Handbook of Statistics*, Vol. 7, ed. by Krishnaiah and Rao. New York: Elsevier.

Kuan, C.M., Huang, Y.L., Tsay, R., 2005. An Unobserved-Component Model with Switching Permanent and Transitory Innovations, Journal of Business & Economic Statistics 23 (4), 443-454.

Levy-Piedbois, C., et al., 2000. Cost-effectiveness of second-line treatment with irinotecan or infusional 5-flourouracil in metastatic colorectal cancer, Annals of Oncology 11(2), 157-161.

Lothian, J., Taylor, M., 1996. Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries, The Journal of Political Economy 104 (3), 488-509.

Lopez, C., Murray, C., Papell, D., 2004. State of the Art Unit root Tests and Purchasing Power Parity, working paper, University of Cincinnati.

Lumsdaine, R., Papell, D., 1997. Multiple Trend Breaks and the Unit-Root Hypothesis, The Review of Economics and Statistics 79 (2), 212-218.

MacDonald, R., Ricci, L., 2001. PPP and the Balassa Samuelson Effect: The Role of the Distribution Sector, IMF Working Papers.

Meyerhardt, J., and Mayer, R., 2005. Systemic Therapy for Colorectal Cancer, The New England Journal of Medicine, 352.

Murray, C., Papell, D., 2002. The Purchasing Power Parity Persistence Paradigm, Journal of International Economics 84, 1-19.

Murray, C., Papell, D., 2005a. The Purchasing Power Parity is Worse Than You Think, Empirical Economics 30 (3), 783-790.

Murray, C., Papell, D., 2005b. Do Panels Help Solve the Purchasing Power Parity Puzzle? Journal of Business & Economic Statistics 23 (4), 410-415.

Nunes, L., Kuan, C.M., Newbold, P., 1995. Spurious Break, Econometric Theory 11, 736-749.

Obstfeld, M., Rogoff, K., 2000. The Six Major Puzzles in International Macroeconomics: Is There a Common Cause? NBER Working Paper 7777.

Papell, D., Prodan, R., 2006. Additional Evidence of Long Run Purchasing Power Parity with Restricted Structural Change, Journal of Money, Credit, and Banking 38, 1329-1349.

Perron, P., and Qu, Z., 2004. Estimating Restricted Structural Change Models, manuscript, Department of Economics, Boston University.

Prodan, R., 2005. Potential Pitfalls in Determining Multiple Structural Changes with an Application to Purchasing Power Parity, manuscript, University of Alabama.

Qu, Z., and Perron, P., 2005. Estimating and testing structural changes in multivariate regressions, Boston University Working Paper Series.

Raftery, A., 1995. Bayesian Model Selection in Social Research, Sociological Methodology 25, 111-163.

Ramsey, S., Howlader, N., et al., 2004. Chemotherapy Use, Outcomes, and Costs for Older Persons with Advanced Non-Small-Cell Lung Cancer: Evidence from Surveillance, Epidemiology and End Results-Medicare, Journal of Clinical Oncology 22 (24), 4971-4978.

Ramsey, S., Berry, K., et al., 2002. Lifetime Cancer-Attributable Cost of Care for Long Term Survivors of Colorectal Cancer, The American Journal of Gastroenterology 97 (2), 440-445.

Rogoff, K., 1996. The Purchasing Parity Puzzle, Journal of Economic Literature 34 (2), 647-668.

Taylor, A., 2002. A Century of Purchasing-Power Parity, The Review of Economics and Statistics 84 (1), 139-150.

Venook, A., 2005. Critical Evaluation of Current Treatments in Metastatic Colorectal Cancer, The Oncologist 10, 250-261.

Wang, B., and Ma, J., 2005. Endogenous Structural Breakpoint Tests: An Investigation in Finite Sample Performance, University of Washington, Manuscript.

Winawer, S., Fletcher, R., et al., 2003. Colorectal cancer screening and surveillance: Clinical guidelines and rationale-update based on new evidence, Gastroenterology 124 (2), 544-560.

Wu, Y, 1996. Are Real Exchange Rates Nonstationary? Evidence from a Panel-Data Test, Journal of Money, Credit, and Banking 28, 54-63.

Zachs, S., 1983, Survey of Classical and Bayesian Approaches to the Change-point Problem: Fixed Sample and Sequential Procedures of Testing and Estimation, *Recent Advances in Statistics*, ed. by Rivzi, Rustagi, and Siegmund. New York: Academic Press.

Zivot, E., and Andrews, D., 1992. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, Journal of Business & Economic Statistics 10, 251-270.

VITA

Bruce Wang was born and raised in Seattle, Washington, where his father is Professor Emeritus of Comparative Literature at the University of Washington. He earned a Bachelor of Arts degree in Economics from Columbia University and a Master of Arts degree in Economics from the University of Washington. In 2007, he earned a Doctor of Philosophy at the University of Washington in Economics. His primary fields are Applied Econometrics and Health Economics. Currently, he is a strategist at Goldman Sachs in New York. He is an accomplished violinist and is fluent in English, Chinese, and French.