Title
Essays in macroeconomics

Permalink
https://escholarship.org/uc/item/76v0631s

Authors
Daula, Thomas Anthony
Daula, Thomas Anthony

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in Macroeconomics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Thomas Daula

Committee in charge:

Professor James D. Hamilton, Chair
Professor Julie Cullen
Professor Jun Liu
Professor Alan Timmermann
Professor Valerie Ramey

2012

The dissertation of Thomas Daula is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Chair

University of California, San Diego

2012

DEDICATION

To my parents, who provided unwavering support for my odyssey through higher education. The journey was long and, at times, uncertain but I persevered. I could not have done it without them.

EPIGRAPH

For society to be at once humane and to give opportunity for great human achievements, it is necessary that a small minority of people who do not have materialistic objectives have the greatest degree of freedom.

—Milton Friedman

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

VITA

2001	B. S. in Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA
2003	M. S. in Statistics, University of California, Los Angeles, Los Angeles, CA
2004-2005	Fixed Income Trading Analyst/Associate, Morgan Stanley, New York, NY
2005-2012	Graduate Teaching Assistant, University of California, San Diego, La Jolla, CA
2012	Ph. D. in Economics, University of California, San Diego, La Jolla, CA

ABSTRACT OF THE DISSERTATION

Essays in Macroeconomics

by

Thomas Daula

Doctor of Philosophy in Economics

University of California, San Diego, 2012

Professor James D. Hamilton, Chair

Motivated by the frictions found in the theoretical literature on the credit channel of monetary policy, chapter one investigates whether credit default swaps' (CDS) superior measurement of credit risk can be used to forecast real economic activity. In a simple model, we find that CDS improve forecasts when compared to standard measures of credit risk and portfolios of corporate bonds compiled by Gilchrist et al. (2010), particularly in the middle of the credit distribution. However the CDS data is limited by only having a short time series. Two additional models are then estimated to overcome this hurdle, a dynamic factor model and Bayesian model averaging (BMA). In these more robust settings, CDS lose their explanatory power. This suggests that modern dimension reduction techniques can successfully extract a parsimonious forecasting model in a data rich environment.

The recent global financial crisis suggests the post-1984 Great Moderation has

come to an abrupt end. How we obtained nearly 25 years of stability and why it ended are ongoing puzzles. Chapter two departs from traditional monetary policy explanations and considers two empirical regularities in US employment: i) the decline in the procyclicality of labor productivity with respect to output and labor input and ii) the increase in the volatility of labor input relative to output. We first consider whether these stylized facts are robust to statistical methodology. We find that the widely reported decline in the procyclicality of labor productivity with respect to output is fragile. Using a new international data set on total hours constructed by Ohaninan and Raffo (2011) we then consider whether these moments are stylized facts of the global Great Moderation. We document significant international heterogeneity. We then investigate whether the role of labor market frictions in the US as found in Galí and van Rens (2010) can explain the international results. We conclude that their stylized model does not appear to account for the differences with the US experience and suggest a direction for future research.

Chapter three examines state-level dynamics of revenues and expenditures. In contrast to previous literature that consider budget deficits in a panel setting, we consider the income elasticity of revenues and expenditures in the cross-section. We examine three budget levels: 1) total budget, 2) excluding liquor stores, utilities, and insurance trusts (LUSI) revenue and expenditures, and 3) excluding capital expenditures. We find that every 1% increase in revenue elasticity is associated with a 0.57% increase in expenditure elasticity for the non-LUSI budget, and this is robust to alternative specifications.

1 Do Credit Default Swaps Improve Forecasts of Real Economic Activity?

1.1 Introduction

There is a large literature using asset prices to forecast output and recessions. The intuition is two-fold. First the basic asset pricing equation is the expectation of discounted future dividends, cash flows, returns, etc. Second, the efficient market hypothesis in its various forms says, roughly, that current prices reflect all of the current information available. The forward looking aspect coupled with some form of the efficient market hypothesis suggests current prices should be good leading indicators of future economic activity. Stock and Watson (2003) provide an exhaustive survey of the recent literature. They trace the interest in asset prices to the instability of forecasts in the 1970s and 1980s using monetary aggregates. They find that certain asset prices have seen success during various periods but none has succeeded consistently. They group financial variables into two broad categories, monetary and credit. Monetary variables include the short rate and term structure, and provide information on monetary policy. Credit variables include various corporate-Treasury bond spreads which back out a measure of the default risk and provide information on the future profit opportunities of firms.

Early empirical evidence suggests credit variables provide information on the real economy whereas monetary variables only provide information prior to 1984. Thus credit variables have had a more stable forecasting performance. Two explanations for this difference are the pivotal effect of the Volcker tightening on the estimation and the low inflation years following 1984 which deemphasized the role of monetary policy shocks. However recent empirical results for the decline of monetary variables are mixed, Black et al. (2000)

1

find support for this hypothesis whereas Thoma and Gray (1998) find no predictive power from any financial variables.

In addition to their empirical support, credit variables also have independent theoretical support. Modern monetary transmission models allow for financial market imperfections. These are predominantly, but need not be, introduced in credit markets. Therefore this class of models have become known as the *credit view* of monetary transmission. Boivin, Kiley, and Mishkin (2010) provide a history of monetary transmission models. There are three main credit channels, however the most widely used is the financial accelerator framework of Bernanke and Gertler (1989) and Bernanke, Gertler, and Gilchrist (1999). In this model, decreases in net worth exacerbate information asymmetries in debt financing thereby deepening downturns. Similarly, as net worth rises excess credit is extended which, in turn, amplifies expansions. In their DSGE model, the amplification mechansim is tuned by the external finance premium which, in turn, is proportional to the credit risk of the borrower. Thus if the financial accelerator is an accurate representation of the economy, credit risk should be intimately tied to the business cycle.

Complicating the empirical inquiry is the fact that the corporate-Treasury spread is an imperfect measure of credit risk, and the results depend on the choice of credit quality and maturity. In fact credit derivatives have overtaken corporate bonds to become the preferred method of gaining credit exposure due to frictions in the bond market, in particular liquidity concerns and the high cost of executing a short position. Also, since CDS are relatively new contracts, there may be concerns about whether their prices are fundamental, i.e. not driven by non-credit forces, and robust to market stress. Preliminary evidence suggests this is not a concern. The CDS market continued to function during the general market meltdowns in 2007 and 2008 following the implosion of the US housing market.[1] They also provided a market signal of the credit worthiness of major banks and borrowers that was otherwise absent in both Libor and corporate bonds, as documented in the Wall Street Journal's coverage of the crisis.[2] Further properties of CDS will be detailed in the next section.

Thus their main shortcoming, empirically, is that they are new contracts having only been liquid since 2003. Adding to the difficulty, macroeconomic data is generally

---

[1] Documented in the November 8, 2008 issue of *The Economist*, "The Great Untangling".

[2] *The Wall Street Journal,* "Bankers cast doubt on key rate amid crisis," 16 April 2008.

published on a quarterly basis exacerbating the problems of the short time series. A final complication is the fact that they have existed only through a single business cycle. In order to assess their forecast ability despite these shortcomings, we will implement two dimension reduction methods, dynamic factor models and Bayesian model averaging. This allows for a richer set of candidate models to be compared.

The paper is structured as follows. Section 1.2 examines the structure and deficiencies of the bond market followed by how CDS can overcome those problems. Section 1.3 examines a very simple forecasting model to motivate the inquiry. Section 1.4 compares CDS to the bond portfolios in Gilchrest *et al.* (2009). After documenting encouraging results we move to two robustness checks. Section 3.6 implements the dynamic factor model found in Hatzius *et al.* (2010) and Bayesian model averaging found in Faust *et al.* (2011). Section 1.6 concludes.

1.2   Market Structure and Credit Default Swaps

The CDS market has seen tremendous growth, rising from $5.1tn in 2004 to $33.4tn gross notional at its peak in 2008.[3] This growth can be explained by the standardization of the market in 1999 and the revisions in 2001 and 2003 prompted by intervening credit events. This proved both that the contracts could withstand actual defaults and also corrected errors in the contracts those events elicited. Once investors were comfortable with the contracts, market participants entered in droves to capitalize on the improvements over the traditional credit risk instrument: corporate bonds. The corporate bond market is characterized by two main deficiencies, low liquidity and difficulty in establishing shorts.

Corporate bond illiquidity can be traced to the asset-liability matching of the investor base, mainly pension funds and insurance companies. They require investment grade assets (generally via statute) and demand high yields in order to cover both the principal and flow of their liabilities. Consequently most investment grade bonds are locked up and do not trade on the secondary market resulting in relatively little market liquidity.[4] They

---

[3]The gross market value rose from $112bn to $1.9tn. In June 2010, the respective values were $18.3tn and $993bn. Data from the Bank of International Settlements: http://www.bis.org/statistics/derstats.htm

[4]Alexander *et al.* (1998) cites anecdotal evidence for initial liquidity in corporate bonds that diminishes over time as these buy-and-hold investors lock-up the securities. Schultz (2001) and Blanco et al. (2005) discuss this as well.

are also a large enough part of the market that even the largest issuances can be held by 200 or fewer institutions. Evidence for the lack of liquidity in the corporate bond market has most recently been examined by Bao, Pan, Wang (2009) who find economically significant levels of illiquidity. In contrast to earlier work, they use transaction level data which allows for a robust model-free measure of liquidity. However, the reliance on transaction level data means they only examine relatively liquid bonds. Since most bonds are traded infrequently this means they obtain a conservative estimate of the full market's illiquidity, i.e. the full market is even more illiquid than their results suggest.

Forecasting models use the excess spread of corporate bonds over a matched maturity Treasury bond. However Duca (1999) points out that this spread actually includes three risk measures; prepayment risk[5], liquidity risk, and credit risk. Typically an investment grade (Baa or better) spread is used which is relatively remote from the default threshold, thus it is expected to mainly reflect the first two risks. Gertler and Lown (1999) agree with this intuition and propose using the high yield-Treasury spread to increase the contribution of default risk to the spread with a noticeable improvement in predictive power. Longstaff, Mithal, and Neis (2005) also found that a significant fraction of the bond spread was due to non-default factors. They provide a number of compelling reasons to prefer CDS over corporate bonds as a measure of credit risk. They then jointly model corporate bond and CDS spreads, assuming CDS are a better measure of credit risk, in order to back out the non-default component. They find that the non-default component of the spread ranges monotonically from 49% for AAA/AA rated bonds to 17% for BB (junk) bonds, confirming Gertler and Lown's intuition. Further they regress this component on a variety of liquidity and tax factors. Their results indicate the non-default component is strongly related to liquidity factors and find little evidence for the tax hypothesis.

In contrast to the corporate bond market, the CDS market is much more liquid. The primary reason is because CDS are contracts rather than securities. This means they can be created whenever there is a willing counterparty rather than relying on new supply which depends on the funding requirements of corporations. In addition, the pressure brought by buy-and-hold investors mentioned in Alexander *et al.* (1998) is alleviated by the ability to create new contracts. Couple this with the standardized nature of the contracts and it

---

[5]This is due to embedded call options allowing the company to purchase the bonds prior to maturity. However this can be corrected for by using Option-Adjusted Spreads (OAS).

is often cheaper simply to enter a new offsetting contract rather than unwind an existing position. Thus the liquidity of an investor's portfolio is less relevant than the liquidity of the market that can be used to replicate its cash flows.

The second institutional problem is the expense in shorting corporate bonds. This is intimately related to liquidity since the small secondary market in conjunction with an aversion to securities lending by pension and insurance companies makes shorting corporate bonds expensive. For CDS this is a non-issue. Since they are contracts, they involve a transfer of credit risk rather than its creation. This means that one counterparty is always short in every CDS transaction. An additional difficulty of shorting in the bond market is the increased the likelihood of being caught in a short squeeze due to closing out the short (buying the security) in an illiquid market.[6] The contractual nature of CDS, on the other hand, implies that shorting is safe from "squeezes" due to the ability to always create new securities. In addition to the implications of transient departures from fair value caused by short squeezes for forecasting, there are two other important reasons to desire an efficient means of shorting; information dissemination and hedging.

Information dissemination is clearly desirable from a forecasting perspective as it reduces the bias and/or increases the signal-to-noise ratio of the security prices we are including in our regressions. The former is directly related to shorting by allowing investors' negative credit views to be reflected in the security's price. For the latter, consider the extreme case of 100% buy-and-hold investors, credit risk is only subject to market forces at the date of issuance with no signal in between issuance dates.[7] While an active two-way market is all that is necessary to transmit information in the interim, in the more realistic case of a small secondary market efficient shorting makes it easier for a two-way market to thrive.[8]

Finally there is no cost effective way for naturally long market participants to hedge

---

[6]Longstaff *et al.* (2005) note that shorting costs are relatively minor (~5 bps) for liquid bonds but it is precisely the fact that the market generally is *not* liquid which makes the costs relevant. In particular they note that firms can sometimes briefly trade special by 50-75 bps but they are generally distressed firms. Blanco, Brennan, Marsh (2005) note that the corporate bond repo market is illiquid leading to sometimes excessive shorting costs.

[7]Similarly, the lack of liquidity leads to so-called Matrix Pricing, where prices for securities without an active market are inferred from similar credits. For empirical work, it is not clear whether prices are true prices or matrix prices unless transaction data is available.

[8]Two recent papers who find shorting in the stock market increases efficiency are Saffi and Sigurdsson (2010) and Boehmer and Wu (2010). I am unaware of a systematic analysis of short selling in the corporate bond market.

their credit risk in the corporate bond market. This is especially true of banks which have large corporate loan portfolios. As loans are bilateral contracts, the due diligence performed by banks does not have a simple outlet to the wider market. Bank's hedging activities via CDS provide that outlet. Additionally credit has been identified as a separate asset class with returns that are relatively uncorrelated with other assets such as equities and commodities. The inability to short efficiently limits the effectiveness of trading strategies that seek to isolate particular risks, e.g. macro, fundamental, credit, etc. Thus the ability to hedge aids in the price discovery process by bringing these arbitrage players into the market. The high transaction cost, large probability of being squeezed, and small secondary market all combine to discourage investors from exerting market discipline on corporate bond prices. In fact, Blanco, Brennan, and Marsh (2005) find that short-run deviations between bond spreads and CDS do exist and can be accounted for by a lead for CDS prices over bonds in the price discovery process.

1.3 Usefulness of CDS to Forecasting the Real Economy

Having established that CDS are a good candidate for forecasting macroeconomic phenomena, it now becomes an empirical question. To begin, we establish a base case that examines the macroeconomic forecasting power of CDS in isolation before turning to more complicated models.

1.3.1 Data

The CDS data are daily five year spreads from DataStream covering the period 1/2004-12/2009. I construct two separate indexes, one for investment grade and one for high yield. In order to control for liquidity I include all entities that have ever been constituents of the traded MarkIt Partners CDX.IG and CDX.HY indexes.[9] This yields 167 investment grade credits and 139 high yield credits. The portfolio spreads are then converted to a monthly frequency by averaging over all credits within each portfolio over the month. This gives two time series, IG and HY, sampled monthly from 1/2004-12/2009.

---

[9]I use the exclusive nor of the two sets, e.g. only those entities that are in one or the other, but not both. This prevents double counting of individual credits. There are also 10 year spreads, however that market was still in its infancy and the integrity of the spreads are questionable.

The measures of macroeconomic activity are industrial production and employment. These were chosen for two reasons. First, these are the principal series considered by the NBER for dating recessions. The committee states,

> A recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in production, employment, real income, and other indicators...The committee believes that domestic production and employment are the primary conceptual measures of economic activity.[10]

Second, they are available at a monthly frequency, unlike GDP. Considering the relatively short time series for CDS, this is important to obtain any statistical power. The data is obtained from the Federal Reserve's statisical database, FRED,[11] and covers the period 3/1989-12/2009. This range is firmly in the modern, post-Volcker period and it is reasonable to assume a stable forecasting relationship. It also covers both of the recessions prior to the introduction of CDS, 7/1990-3/1991 and 3/2001-11/2001.

1.3.2 Model

Let Y_t denote a measure of economic activity in month t, define

$$\nabla^h Y_{t+h} = \frac{1200}{h} \ln\left(\frac{Y_{t+h}}{Y_t}\right)$$

where h denotes the forecast horizon. Consider the following forecasting equations:

$$\nabla^h \text{EMP}_{t+h} = \beta_0 + I_{\{t>t_{CDS}\}}\beta_1 + \sum_{i=0}^{3} \beta_{2i}\nabla\text{EMP}_{t-i} + \sum_{i=0}^{3} \beta_{3i}\nabla\text{IP}_{t-i} + \eta_1' CDS_{1t} + \varepsilon_{1,t+h}$$

$$\nabla^h \text{IP}_{t+h} = \gamma_0 + I_{\{t>t_{CDS}\}}\gamma_1 + \sum_{i=0}^{3} \gamma_{2i}\nabla\text{EMP}_{t-i} + \sum_{i=0}^{3} \gamma_{3i}\nabla\text{IP}_{t-i} + \theta_1' CDS_{1t} + \varepsilon_{2,t+h}$$

Where $I_{\{t>t_{CDS}\}}$ is an indicator for the introduction of CDS. This autoregressive framework accounts for the serial dependence in Y_t and follows Stock and Watson (2003). Since the forecasts are overlapping, we follow the literature and use a Newey-West estimator with lag parameter equal to $h+1$ in order to correct for autocorrelation and heteroskedacity. The hypothesis of interest is whether *CDS* has any predictive power above and beyond that found in the lags of EMP_t and IP_t: $\eta_1 = 0$ and $\theta_1 = 0$.

---

[10]NBER dating committee December 2008 press release: http://www.nber.org/cycles/dec2008.html
[11]http://research.stlouisfed.org/fred2/

1.3.3 Results

The model is estimated with and without the CDS at a 3 month horizon and the results are given in Table 1.1. The null hypothesis is soundly rejected with both p-values being zero out to 3 decimal points. Thus there is additional information in the CDS that is not present in the economic activity variables themselves. The adjusted R^2 also shows that the fit improves more for IP than EMP, although EMP is fit better overall. Figure 1.6 graphs the fitted and realize values. It appears that for both IP and EMP, the improvement from including CDS is in the fit to the large downturn in 2008-2009. This is true not only in the depth of the downturn but also in the timing. Having established that CDS improve the fit to economic activity versus a distributed lag model, we now compare them to a wider array of financial indicators.

1.4   Comparison with other credit variables

There are a number of credit indexes available, as well as the underlying corporate bonds themselves. The standard credit indexes used in the literature are 1) *paper-bill spread*: the difference between the yield on one-month nonfinancial AA-rated commercial paper and the constant maturity one-month Treasury bill[12]; 2) *Aaa corporate bond spread*: the difference between the yield on an index of seasoned long-term Aaa-rated corporate bonds and the yield on the constant maturity 10-year Treasury note; 3) *Baa corporate bond spread:* the difference between the yield on an index of seasoned long-term Baa-rated corporate bonds and the yield on the constant maturity 10-year Treasury note; and 4) *High yield bond spread*: the difference between the yield on an index of high yield corporate bonds and the yield on the constant maturity 10-year Treasury note. [13]

The corporate bond market is notoriously opaque, although the introduction of TRACE in 2000 has alleviated that criticism somewhat. Consquently we use the corporate

---

[12]The one-month Treasury Bill is only available starting in 7/2001. To go back to 2/1990 I splice it with the one-month implied Treasury yields obtained from Gurkaynak, Sack, Wright (2006).

[13]Commercial paper rates are taken from the "Commercial Paper Rates and Outstanding" Federal Reserve statistical release. The source of the Treasury rates and the Aaa- and Baa-rated corporate bonds is "Selected Interest Rates" (H.15) Federal Reserve statistical release. The high yield spread is Bank of America/Merrill Lynch's High Yield Master II index obtained from Datastream. The spread is not adjusted for embedded options and is not directly comparable to that used by GYZ. This point is taken up later when discussing the results.

bond portfolios compiled by Gilchrist *et al.* (2010) (henceforth GYZ) using proprietary data. They have a large panel of corporate bonds which are sorted by credit risk and maturity. Here we briefly review their construction, for full details please see their paper. The initial panel data are month-end secondary market prices of corporate bonds for a sample of 899 publicly traded firms. The credit quality spans the full set, from junk "D" to high quality "Aaa" and the median observation is Bbb (investment grade). After controling for outliers and liquidity, they arrive at 5,045 individual securities. These securities are then sorted according to credit risk. Of particular note, they acknowledge the contamination of corporate spreads by factors other than credit risk, as mentioned in Section 1.2.[14] For this reason they use a measure of credit risk, independent of corporate bonds, marketed by Moody's/KMV (MKMV) corporation.

The default risk is assessed using expected default frequencies (EDFs) provided by MKMV. The EDF is derived using a Merton model[15] of the firm coupled with the empirical distribution of defaults. In the first step, Merton derives a distance-to-default measure:

$$[\text{Distance to Default}] = \frac{[\text{Mkt. Value of Assets}] - [\text{Default Point}]}{[\text{Mkt. Value of Assets}] \times [\text{Asset Volatility}]}$$

However, the market value of the assets and the associated volatility are not directly observable. Merton overcame this deficiency by noting the market value of the firm's equity can be viewed as a call option on the firm's assets with a strike price equal to the current book value of the firm's debt. MKVM uses this to back out the distance to default measure and then transforms it (nonparametrically) into an EDF using the empirical distribution of actual defaults. Two advantages of this approach are that it uses the equity market rather than the bonds themselves and the equity prices provide more timely and nuanced information than credit ratings, both of which can also be said about the CDS market.

Finally the portfolios are sorted according to maturity: 1) less than 3 years; 2) 3 to 7 years; 3) 7 to 15 years; and 4) more than 15 years. This results in 20 (5 credit by 4 maturity) portfolios. Since I only have 5 year CDS, I ignore the maturity dimension and group the portfolios by credit quality in the following regressions. The time period examined is from

---

[14]"A *portion* of observed credit spreads reflects compensation demanded by investors for bearing the risk that a firm...will default on its payment obligations." p. 474. (emphasis added)

[15]Originally published in Merton (1973, 1974). A modern treatment can be found in most option pricing texts, for instance Hull (2002).

2/1990 to 12/2008.[16]

GYZ examine a similar question to our own and a brief summary of their results is warranted. They look at both the portfolios' ability to explain in-sample variation and forecast pseudo out-of-sample. GYZ find that their portfolios contain information not found in the standard indexes, however this information is largely complementary rather than supplementary. Their in-sample results mirror our base case with a modest improvement in the fit to employment and substantial improvement in industrial production. Their pseudo out-of-sample results also support the presence of additional information in corporate bond data particularly at the 12 month forecast horizon. Also the increase in forecasting performance is concentrated in the middle of quality distribution and longer maturity bonds. Extending their paper allows for a direct comparison between CDS and corporate bonds (for reasons outlined in Section 1.2) as well as the standard credit indexes.

1.4.1 Model

Let Y_t denote a measure of economic activity in month t, define

$$\nabla^h Y_{t+h} = \frac{1200}{h} \ln\left(\frac{Y_{t+h}}{Y_t}\right)$$

where h denotes the forecast horizon. GYZ estimate the following bivariate direct h-step ahead forecasting relation:

$$\nabla^h \text{EMP}_{t+h} = \beta_0 + \sum_{i=0}^{11} \beta_{1i} \nabla \text{EMP}_{t-i} + \sum_{i=0}^{11} \beta_{2i} \nabla \text{IP}_{t-i} + \eta_1' Z_{1t} + \eta_2' Z_{2t,j} + \varepsilon_{1,t+h}$$

$$\nabla^h \text{IP}_{t+h} = \gamma_0 + \sum_{i=0}^{11} \gamma_{1i} \nabla \text{EMP}_{t-i} + \sum_{i=0}^{11} \gamma_{2i} \nabla \text{IP}_{t-i} + \theta_1' Z_{1t} + \theta_2' Z_{2t,j} + \varepsilon_{2,t+h}$$

where $\underset{(3\times1)}{Z_{1t}}$ denotes the standard credit spreads (paper-bill, Aaa, Baa) and $\underset{(4\times1)}{Z_{2t,j}}$ denotes the four maturity differentiated EDF portfolios in the jth credit risk quantile constructed using Moody's/KMV's method.

We are interested in the forecasting ability of CDS, however generalizing the above model to include CDS raises two difficulties. First, the CDS data only exists after 1/2004.

---

[16]Gilchrest *et al.* (2009) look at 2/1990 to 9/2008, they were kind enough to send me their EDF portfolios updated to 12/2008.

In order to allow maximum flexibility in the model, I allow all of the coefficients and the constant to be different in the period before the introduction of CDS and afterward. Second, the eleven lags of the dependent variable puts tremendous strains on the data due to the short time series for CDS and the addition of the structural change parameters. Consequently I only include three lags of the dependent variable. These changes then give us the forecasting equations:

$$
\begin{aligned}
\nabla^h \text{EMP}_{t+h} &= \beta_0 + I_{\{t>t_{CDS}\}}\beta_1 + \sum_{i=0}^{3}\beta_{2i}\nabla\text{EMP}_{t-i} + \sum_{i=0}^{3}\beta_{3i}\nabla\text{IP}_{t-i} + \eta'_{11}Z_{1t} \quad (1.1)\\
&\quad + \eta'_{12}I_{\{t>t_{CDS}\}}Z_{1t} + \eta'_{21}Z_{2t,j} + \eta'_{22}I_{\{t>t_{CDS}\}}Z_{2t,j} + \eta'_{3}Z_{3t} + \varepsilon_{1,t+h}\\
\nabla^h \text{IP}_{t+h} &= \gamma_0 + I_{\{t>t_{CDS}\}}\gamma_1 + \sum_{i=0}^{3}\gamma_{2i}\nabla\text{EMP}_{t-i} + \sum_{i=0}^{3}\gamma_{3i}\nabla\text{IP}_{t-i} + \theta'_{12}Z_{1t}\\
&\quad + \theta'_{12}I_{\{t>t_{CDS}\}}Z_{1t} + \theta'_{21}Z_{2t,j} + \theta'_{22}I_{\{t>t_{CDS}\}}Z_{2t,j} + \theta'_{3}Z_{3t} + \varepsilon_{2,t+h}
\end{aligned}
$$

where Z_{3t} $\underset{(2\times1)}{}$ denotes CDS spreads (IG and HY), and $I_{\{t>t_{CDS}\}}$ is an indicator for the time period that CDS data begins.

The hypothesis of interest is whether the coefficients on Z_{3t}(CDS) remain significant. Including the other two variables, $\{Z_{1t}, Z_{2t,j}\}$, allows us to test nested hypotheses of whether CDS improve the fit over standard indexes used in the literature or the EDF portfolios. Since we only include one EDF quantile at a time, this gives ten different specifications; five quantiles of the EDF portfolios for each dependent variable.

1.4.2   Results

1.4.2.1   In-sample results

Tables 1.2 and 1.3 show the results for the in-sample GYZ forecast given in equation 1.1.[17] The left column gives the regressors included in addition to the autoregressive lags. For each specification the Wald statistic for the null hypothesis that the coefficients are equal to zero is given for the standard credit indexes W_1, the EDF portfolio W_2, and the CDS indexes W_3.

We find that the CDS indexes remain significant even in the presence of the standard indexes for both IP and EMP. This suggests they contain some additional information

---

[17]These mirror Table 3 found in Gilchrest *et al.* (2009).

beyond that found in broad quality constrained credit indexes. To judge the goodness-of-fit, compare the adjusted R^2 of the model just with the standard indexes to that with just CDS. The adjusted R^2 attempts to account for the possibility of overfitting by penalizing additional terms. The increases in adjusted R^2 from a model with just standard indexes to one that also includes CDS means the improvement in fit is more than the penalty due to the additional terms. Comparing CDS in isolation, we see that CDS do a much poorer job fitting IP than EMP compared to the standard indexes. However the overall fit is better for EMP than IP (0.821 vs. 0.503).

Turning to the corporate bond portfolios, CDS generally remain significant for the middle of the credit quality distribution and appear to have more information for EMP. They do not displace any of the existing covariates and are thus complementary measures of real activity.[18] This is somewhat intuitive, the bond portfolios and credit indexes are contaminated measures of default risk. However a prime contaminant is interest rates[19], and the large literature on forecasting real activity using the term structure suggests this is not white noise. CDS complement these measures by refining the default measurement, as indicated by Longstaff, *et al.* (2005). This suggests a natural extension to considering a pure interest rate measure (e.g. Treasuries) in conjunction with a pure default measure, CDS, in order to decompose the variation in real activity. This extension is left to future research.

1.4.2.2 Pseudo out-of-sample results

An additional way to view the predictive content of credit spreads is to examine how the forecasting relation would hold out-of-sample. Since we do not have any observations beyond our estimation sample, we are not able to do a true out-of-sample experiment. However the pseudo out-of-sample exercise can give some indication of the forecast stability.

The choice set for splitting our sample is limited by the short CDS time series.

---

[18]These results aren't directly comparable to GYZ due to the different HY index. The time series was independently verified to be much different, and I believe this is due to the embedded options for which I cannot control. However the qualitative results were largely unchanged when estimated without the HY index. Quantitatively, all of the CDS results were stronger, were significant in a few more cases, and actually displaced the standard indexes in one case.

[19]See Duca (1999) for this and other problems with corporate bonds not pursued here.

We need a long enough estimation sample for the model to have any chance at fitting the data but also need a long enough test sample in order for the forecast error statistic to be meaningful. In addition, we do not want to include the financial crisis in the estimation sample in order to minimize any bias for the time period we are restricted to. In light of these considerations, the first regression is estimated on the sample from February 1990 to January 2005. Those estimates are used to construct forecasts for the next three months and the forecast error is recorded. The estimation sample is then augmented by an additional month of data and the exercise is repeated until we reach the end of our sample. Tables 1.4 and 1.5 record the root mean squared forecast error (RMSFE). In addition, the forecast accuracy is compared across models by taking the ratio of the MSFE with respect to the model with just the standard credit indexes in Table 1.4 and with respect to the model with both standard credit indexes and the matched EDF portfolio in Table 1.5.

The CDS do not improve the out-of-sample forecast for industrial production. However for employment there appears to be some gains. Focusing first on Table 1.4, we see that CDS significantly improve the out-of-sample performance relative to the standard credit indexes for employment. In addition, EDF portfolios and CDS indexes together do not seem to improve the out-of-sample performance except for the lowest rated portfolios. Turning to Table 1.5 the CDS indexes are seen to significantly improve the performance for the middle of the credit distribution. This is in line with the in-sample results which also showed a better fit for the middle of the credit distribution.

A possible explanation for why the lowest rated EDF portfolios perform better than CDS is the peculiarities of the high-yield bond market. Depending on whether these were issued as HY or were issued at a higher rating (so-called "fallen angels") they may have various credit enhancements and embedded options. These additional non-linearities may be providing non-credit forms of information not present in the CDS.[20] Unfortunately, without access to the underlying bonds this can not be tested.

1.4.3   Caveats

The GYZ portfolios provide a direct comparison to the CDS. In this exercise we saw forecasting improvements for employment but not for industrial production, however

---

[20]In an extension to GYZ, Faust, *et al.* (2011) devote much of their paper to accounting for embedded options in the EDF portfolios.

two problems remain. First, the short time period necessitated rather restricted forms of the regressions in order to maintain precision. Second, without access to the underlying bonds it is unclear what is driving the difference in forecasting performance. An ideal dataset would match each CDS to its floating rate corporate bond and a floating rate risk-free security.

The next section implements two modern estimation techniques to work around these issues. They are inspired mainly to deal with the first issue - short time series. They accomplish this by, loosely speaking, reducing the dimension of the covariates. The reduction is either by finding a common source of variation or by model averaging over many univariate regressions. This permits a wider range of covariates to be entertained in a single model. Additionally, they also allow us to bring in a larger variety of covariates. This expands the scope beyond simply comparing credit variables and addresses the value of CDS in a larger context of forecasting exercises.

1.5   Robustness Checks

Absent a matched portfolio of bonds and CDS, the GYZ results provide the closest intuitive comparison between the forecasting performance of CDS and other widely used measures of credit risk. However the short time span asks much of the data. This section provides two modern approaches to dealing with the short time period, dynamic factor models and bayesian model averaging (BMA).

1.5.1   Dynamic Factor Model

We implement the dynamic factor model introduced in Hatzius, *et. al.* (2010) (henceforth HHMSW). HHMSW are interested in summarizing the relationship between the financial markets and economic activity in a robust, reduced form framework. This is particularly relevant given the unprecedented central bank policy actions of the recent crisis and corresponding dearth of structural models. Their model is marketed as a useful guide for motivating analysis of monetary policy with imperfect financial markets, otherwise known as the "credit view" of monetary transmission summarized in Boivin, *et al.* (2010). They improve upon the workhorse principal components framework to summarize

the information in 45 separate financial time series into a financial conditions index (FCI) and then relate the resulting factors to economic activity.

First they survey the existing FCIs and then construct a new index based upon the identified deficiencies. FCIs fall into two broad categories, a weighted-sum approach and principal-components approach. The weighted-sum approach assigns weights to the components series based upon their relative predictive power for real GDP. The weights are generated from some underlying model ranging from structural macroeconomic models to reduced form VARs. The principal-components approach, on the other hand, extracts a common factor from a set of variables.[21] This common factor captures the greatest common variation in the set of variables. In contrast to the weighted-sum approach, principal components is primarily a statistical representation rather than an economic one. Based upon their analysis they introduce three innovations to the construction of FCIs.

Their first innovation is to expand the components beyond the traditional price and credit variables (e.g. interest rates, stock market return, corporate bond/Treasury spread) to also include quantity (e.g. commercial paper and ABS issuance, bank credit) and survey variables (e.g. Michigan survey). Second, they implement a new estimation technique for principal components with unbalanced panels that allows them to coherently examine a longer time series. This allows for series to be added later in time as financial innovation provides new securities that isolate specific risks. The final innovation is to control for past changes in real measures (in their case, GDP growth and inflation) in order to isolate the components' predictive power for future activity.

Their method is interesting for two reasons. First, it is built specifically to allow for financial innovation. Our principal motivation is that the introduction of CDS provided a better measure of credit risk. Their model automatically incorporates that new information. Second, this is a broad selection of financial variables. One of the conclusions in Stock and Watson (2003) is the instability of univariate forecasts. By optimally summarizing the information in a disparate set of financial variables, this should overcome the problem of covariates only having power in certain subperiods. The question then is whether CDS can provide information that is either not already attainable in the market or provide it more efficiently.

---

[21]More broadly it extracts a set of factors as we will see later. See Stock and Watson (2006) for an overview.

1.5.1.1 Data

Two separate exercises are performed, one on the data provided by HHMSW and a second on a subsample. First, HHMSW focus on GDP and thus construct quarterly factors from 45 separate financial time series. These are broadly defined as interest rates (15), prices (5), quantities (15), surveys (7), and 2nd moments (3). 29 of the variables have never been included in an FCI. Following the discussion in Section 1.3.1 our dependent variables are non-farm payrolls (EMP) and industrial production (IP). The predictors are the HHMSW factors introduced above and two CDS indexes. EMP, IP, and the HHMSW factor are quarterly data that span 5/1970 to 12/2009 and are available on Mark Watson's webpage.[22] The CDS indexes used previously are converted to a quarterly frequency and span 2/2004 - 12/2009.

There are two reasons to construct our own factors. First, monthly data is more desireable given our short CDS time series. Second, the HHMSW model is intended to incorporate new financial innovations. Thus it is of interest whether CDS improve upon the constructed factors. The factors are constructed as follows. Let X_{it} be the ith financial variable at time t, and Y_t be the vector of real activity variables. First regress X_{it} on lags of Y_t:

$$X_{it} = A_i(L)Y_t + v_{it}$$

thus v_{it} is uncorrelated with current and lagged Y. This isolates the innovations in the financial market that are uncorrelated with the current state of the real economy. Further suppose v_{it} can be decomposed as

$$v_{it} = \lambda_i' F_t + u_{it}$$

This is the standard PCA assumption. However recall that we have an unbalanced panel, in which case the standard eigenvector decomposition can not be applied. This is instead estimated via iterated least-squares.[23] F_t are the factors of interest.

---

[22] The non-proprietary data used for constructing their factor is also available on their website: http://www.princeton.edu/~mwatson/wp.html

[23] See the Hatzius *et al.* (2010) for more details, as well as the statistical properties that allow these factors to be used in later regressions without a loss in efficiency.

For the second exercise our dependent variables are non-farm payrolls (EMP) and industrial production (IP). The predictors are the HHMSW factors constructed on the publicly available subset of financial time series with the inclusion of two CDS indexes. EMP, IP, and the HHMSW factor are monthly data that span 5/1970 to 12/2009.

1.5.1.2 Model

For our first specification, we estimate a direct h-step ahead forecast.

$$
\begin{aligned}
\nabla^h \text{EMP}_{t+h} &= \beta_0 + I_{\{t>t_{CDS}\}}\beta_1 + \sum_{i=0}^{3} \beta_{2i} \nabla \text{EMP}_{t-i} + \eta'_{11}Z_{1t} + \eta'_{12}I_{\{t>t_{CDS}\}}Z_{1t} \quad (1.2) \\
&\quad + \eta'_2 Z_{3t} + \varepsilon_{1,t+h} \\
\nabla^h \text{IP}_{t+h} &= \gamma_0 + I_{\{t>t_{CDS}\}}\gamma_1 + \sum_{i=0}^{3} \gamma_{2i} \nabla \text{EMP}_{t-i} + \theta'_{11}Z_{1t} + \theta'_{12}I_{\{t>t_{CDS}\}}Z_{1t} \\
&\quad + \theta'_2 Z_{3t} + \varepsilon_{2,t+h}
\end{aligned}
$$

Where $\underset{(1\times1)}{Z_{1,t}}$ is the factor derived by HHMSW, $\underset{(2\times1)}{Z_{3t}}$ denotes CDS spreads (IG and HY), and $I_{\{t>t_{CDS}\}}$ is an indicator for the time period that CDS data begins. This is done on quarterly data with the published HHMSW factors. Here we are interested in whether the CDS contain relevant information not found in the constructed factors.

For the second exercise we also estimate a direct h-step ahead forecast, but we use the HHMSW method directly.

$$
\begin{aligned}
\nabla^h \text{EMP}_{t+h} &= \beta_0 + \sum_{i=0}^{3} \beta_{2i} \nabla \text{EMP}_{t-i} + \sum_{i=0}^{3} \eta'_{1i}Z_{1,t-i} + \varepsilon_{1,t+h} \quad (1.3) \\
\nabla^h \text{IP}_{t+h} &= \gamma_0 + \sum_{i=0}^{11} \gamma_{2i} \nabla \text{IP}_{t-i} + \sum_{i=0}^{3} \theta'_{1i}Z_{1,t-i} + \varepsilon_{2,t+h}
\end{aligned}
$$

Here $\underset{(1\times1)}{Z_{1,t}}$ is the factor estimated on the public time series and either does or does not include CDS as a candidate in its construction. If CDS contain pertinent information then it should shift the estimated factors.

1.5.1.3 Results

Table 1.7 contains the results of the first exercise. In contrast to the standard method, CDS marginally improve the in-sample fit of IP but are insignificant in the EMP regression. That there is any power is significant given the small sample (24 quarterly observations) used to estimate the contribution from CDS. It is possible the power may be unique to the time period considered, especially the financial crisis of 2008-2009, however the wide variety of sources used in the construction of the HHMSW factor suggests that possibility is less than one might imagine. The CDS had to improve the information set of not only the other credit factors already included but also the quantity and survey variables.

Since there appears to be some additional information in the CDS beyond the constructed factors, the next step is to incorporate the CDS into the factors themselves and ascertain whether there is any improvement. Thus we now estimate equation 1.3 where the factors are constructed on a subset of the series and either include CDS or do not. Tables 1.8 and 1.9 show the in-sample fit of industrial production and employment, respectively. We see that the factors provide nearly identical fits and when both are included, there are significant collinearity problems. This is unsurprising as the two factors themselves are nearly identical as can be seen in figure 1.6 which plots the single factor with and without CDS in the covariate set. Regressing the contemporaneous CDS factor on the contemporaneous factor without the CDS yields an R^2 of 1.

Although the CDS do not seem to be altering the path, the factor appears to be picking up the business cycle. The recessions in the mid-1970s and early 1980s are clearly evident, as is the 1987 crash and 2008 crisis. The only difficulty appears to be the 1990 recession and subsequent boom. Table 1.11 has the RMSFE for the pseudo out-of-sample exercise, following the procedure outlined in the previous section. Commensurate with the single factor figure, we see that the introduction of the CDS does not alter the forecasting performance.

This suggests that factor models are able to extract the "business cycle" signal from the existing time series and there is no need to include CDS. One possibility is to view the standard credit indexes as contaminated estimates of credit risk, then the factor model is able to extract that information in conjunction with the interest rate variables. Even if CDS are less noisy, the model weights the standard indexes more due to the longer time series.

This allows the factor structure to disentangle the business cycle effects on the interest rate and credit markets. Another possibility is that the iterated least-squares methodology itself hinders the incorporation of new information. For the longer time series it must minimize a larger number of errors. Thus even if a new series provides additional information, it is hard for the least-squares algorithm to shift the entire path of the factors. A heuristic check of the latter hypotheses is to start the estimation much closer to the CDS, such that they now are a larger fraction of the time periods.

Figure 1.6 plots the dynamic factors for estimation starting in 1970, 1980, 1990, and 2000. Note that there isn't significant change until we begin 1990, but by then all the series except CDS and CP issuance are already included, see Table 1.6. Also note that the most recent factor diverges around 2005 and is overly optimistic, although this is unsurprising since the only recession in the data is in 2001. Table 1.10 provides the pairwise correlations. This confirms the divergence of the factors estimated from 2000, but also shows that the extra volatility in the 1990 estimation mask a fairly close similarity to the previous time periods.

Perhaps the change in the factors is due to the greater importance of CDS, but I suspect the removal of the 1981 recession also eliminates the so-called "Volcker" effect alluded to in the introduction. Namely the recessions in the 1970s and early 1980s were more intimately related to monetary policy and removing them from our estimation allows greater freedom to fit the modern business cycle regime. Increasing the number of factors can possibly disentangle the credit and monetary policy effects. Figure 1.6 plots the factors from a three factor model. The third factor appears to pick up the recessions in the 1970s and early 1980s, but it shows improving conditions in 2008. The 1980s recessions were associated with the Volcker tightening whereas the 2008 recession was accompanied by unprecedented loosening of monetary policy. Additionally, note that the first two factors pick up the 1987 crash, whereas the third factor shows little change. This evidence suggests that the third factor is indeed a monetary policy factor. While this appears to be a success of the dynamic factor model, the interpretation of the other two factors remains ambiguous and this separation does not appear in the two factor model (available upon request).

We have shown that dynamic factor models appear to be able to capture the salient features of the business cycle. The HHMSW factor structure provides several key extensions of the basic principal components analysis but there remain some potential concerns.

The variation with the starting date highlights the delicate balance between relevance and precision. Namely, we require enough data on rare recession events but want to make use of the more recent and, therefore, more informative data. Additionally, if we want to retain interpretability the choice of the number of factors remains important.

1.5.2 Bayesian Model Averaging

An alternative method to overcome the short time series is to appeal to Bayesian methods. Even without an informative prior, Bayesian methods intuitively incorporate uncertainty about parameter values with that contained in the data. Bayesian model averaging (BMA) is one way to incorporate the "prior" information contained in multiple independent regressions into a single final forecast. The weights used to construct our final forecast summarize the information in that regression and thus, as a byproduct, we are also able to compare the information content of competing models.

We follow the BMA procedure in Fernandez, et al. (2001).[24] Following the notation in Faust et al. (2011), let M_i denote model i which is parameterized by θ_i, and let D be the observed data. The researcher has prior $P(M_i)$ and then updates beliefs to form the posterior

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum P(D|M_j)P(M_j)} \tag{1.4}$$

where

$$P(D|M_i) = \int P(D|\theta_i, M_i)P(\theta_i|M_i)d\theta_i$$

We consider regression forecast models given by

$$y_{t+h} = \beta_i Z_{it} + \gamma' X_t + \varepsilon_{t+h} \tag{1.5}$$

where h is the forecast horizon, Z_{it} are the model-specific covariates, and X_t is a $(p \times 1)$ vector of covariates common to every model. Assume Z_{it} is orthogonal to X_t and $\varepsilon_{t+h} \sim N(0, \sigma^2)$. In our case X_t includes the standard credit indexes and Z_{it} is either CDS or the Gilchrist EDF portfolios, with one caveat. The CDS data is available starting in 2004 so

---

[24]Faust, *et al.* (2011) also implement a version of the Fernandez, *et al.*(2011) BMA procedure however their objective is much broader than ours. We are concerned specifically with the forecasting ability of CDS versus other credit instruments, whereas they look at many more macro and financial variables to determine if credit (in any form) is useful. They find significant positive results with updated Gilchrist, *et al.* (2010) portfolios.

Z_{it} also includes $I_{\{t>t_{CDS}\}}X_t$ for those models, similar to our specifications in the previous sections.

Now we need to specify the priors. Following the literature we specifiy an un-informative prior for $p(\gamma, \sigma)$ that is proportional to $1/\sigma$, and the g-prior specification of Zellner (1986) for β_i conditional on σ. The g-prior is given by $N(0, \phi\sigma^2(X_i'X_i)^{-1})$ where the shrinkage hyperparameter $\phi > 0$ measures the strength of the prior.[25] For the model given here, the marginal likelihood of the i-th model reduces to[26]

$$P(D|M_i) \propto \left[\frac{\phi}{1+\phi}\right]^{-\frac{(p+p_i+1)}{2}} \times \left[\frac{1}{1+\phi}SSR_i + \frac{\phi}{1+\phi}SSE_i\right]^{-\frac{T}{2}} \tag{1.6}$$

Where SSR_i is the regression sum of squares, $(\hat{y}-\bar{y})'(\hat{y}-\bar{y})$, and SSE_i is the sum of squared errors, $(y-\hat{y})'(y-\hat{y})$.

Let $\hat{\beta}_i$ and $\hat{\gamma}$ be the OLS estimates from equation (1.5), then $\tilde{\beta}_i = \left(\frac{\phi}{1+\phi}\right)\hat{\beta}_i$ is the posterior mean of β_i and the Bayesian h-period ahead forecast from model M_i at time T is given by

$$\tilde{y}^i_{T+h|T} = \tilde{\beta}_i X_{it} + \hat{\gamma}'Z_t$$

The final BMA forecast is then given by the individual forecasts weighted by the posterior probabilities of each model

$$\tilde{y}_{T+h|T} = \sum_{i=1}^{n} P(M_i|D)\tilde{y}^i_{T+h|T}$$

1.5.2.1 Results

The model specific covariates, Z_{it}, in equation 1.5 are the standard credit indexes, GYZ quantile portfolios, and CDS quantile portfolios.[27] A benefit of the univariate mod-els is we can look at a larger number of covariates, in particular disaggregating the credit risk in the CDS data into quantiles. In the basic specification, the short time series lim-ited the number of covariates that were feasible. Table 1.12 has the posterior probabilities from equation 1.4 grouped by type, either CDS or GYZ, for each model specification. In other words, $\sum_i P(M_i|D)$ where the sum is over all models classified as CDS or GYZ, re-spectively. We see that the BMA procedure loads on the CDS models over the standard

---

[25]This correpsonds to g_0 in Fernandez et al. (2001).

[26]This differs slightly from Faust, et al. (2011) due to the additional elements in Z_{it} for the CDS specifications.

[27]Similar results held when it was just split between IG and HY.

indexes but the EDF portfolios dominate if all three sets are included. Table 1.13 has the pseudo out-of-sample RMSFE at the 3 month horizon. We see that the CDS do no better than the standard indexes in forecasting out-of-sample. However it is not due to the CDS being uninformative, since the posterior loads almost entirely on the CDS. Instead the CDS appear to provide complementary information that is equally as effective in forecasting out-of-sample. The EDF portfolios dominate the out-of-sample exercise for industrial production and provide a minor improvement in employment. This is in contrast to the basic specification where CDS provided significant improvements in out-of-sample forecasts for employment and there is little improvement in industrial production.

GYZ were interested in decomposing the information content between asset maturity and credit risk. There are significant differences in the performance among those dimensions, and the BMA procedure is able to combine them in order to achieve results on par with the best in the simpler case. BMA however is robust to data mining and is easily capable of introducing new time series. This suggests the earlier results were specific to the time period and short time series. Both the dynamic factor model and BMA procedures show that given a rich enough data set, an applied researcher can extract the relevant information in a straightforward, atheoretic way.

1.6 Conclusions

Asset prices have the potential to improve forecasts of the real economy. In recent history, credit variables have performed the best out of candidate assets. However, institutional problems in the corporate bond market and difficulties in extracting a clean measure of credit risk plagued inquiry. Here we have shown that CDS are theoretically superior to corporate bonds in terms of providing a measure of credit risk. We then showed that CDS improves the performance of a simple forecasting model, particularly for employment. However this result is from a limited time series over a particularly benign credit period.

Two dimension reduction methods are then employed to efficiently combine the information from a multitude of time series over the short time period. Both dynamic factor models and Bayesian model averaging improve the forecasting performance relative to a benchmark autoregression. In addition to finding a parsimonius model, they also perform

well and nearly identically in a pseudo out-of-sample exercise. We conclude that given a sufficiently rich data environment, modern dimension reduction techniques can successfully extract relevant information. In addition, while the underlying theory may appear daunting they are relatively easy to implement and may highlight fragility in statistical results from simpler models. Finally, the focus here was on various candidates for measures of credit risk. In a richer data environment with intelligent groupings, this can be extended to look at what types of information are relevant for the forecasting problem of interest, e.g. Faust *et al.* (2011), and motivate focus for structural inquiry.
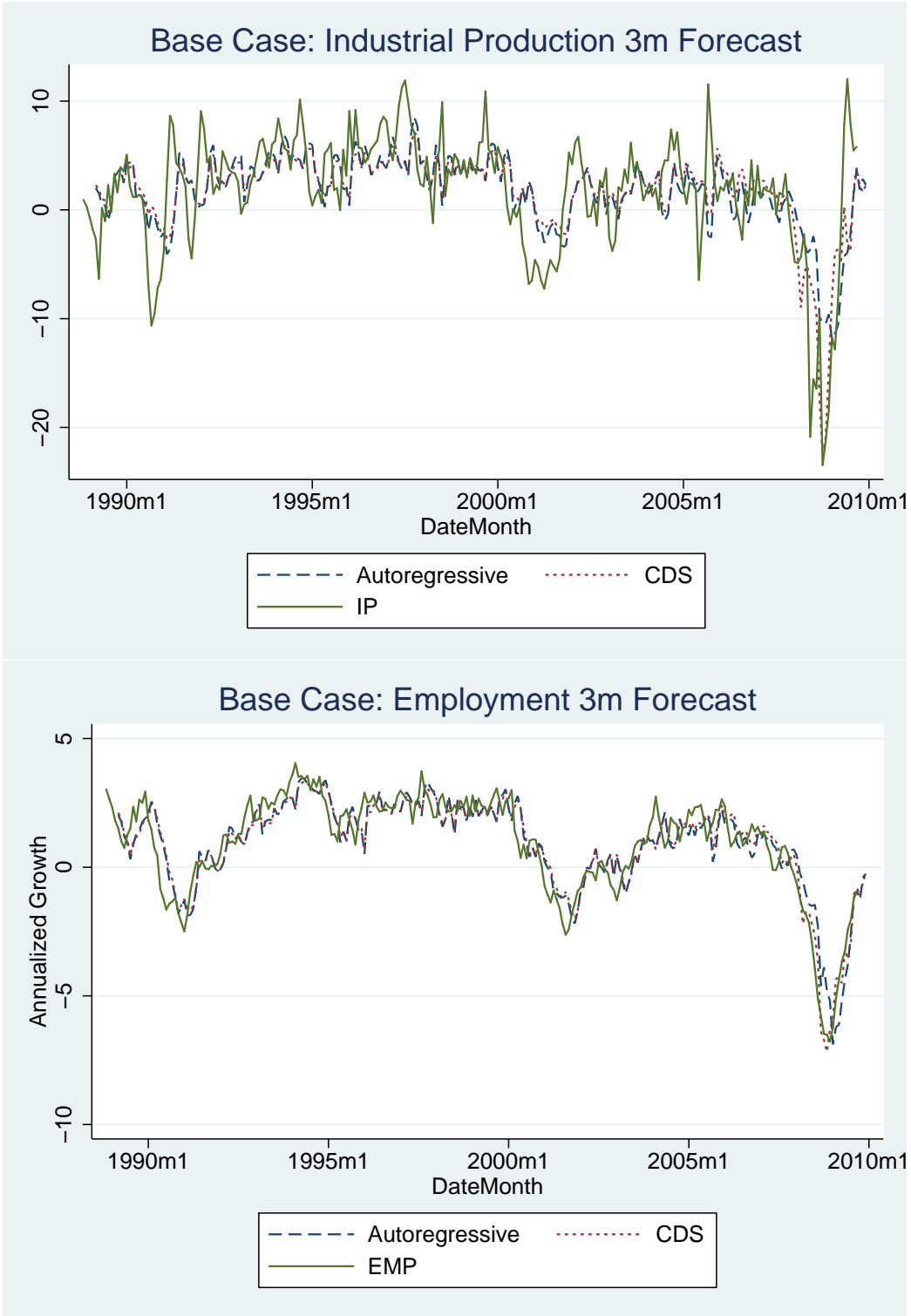
Figure 1.1: Forecasts of Industrial Production and Emloyment using CDS
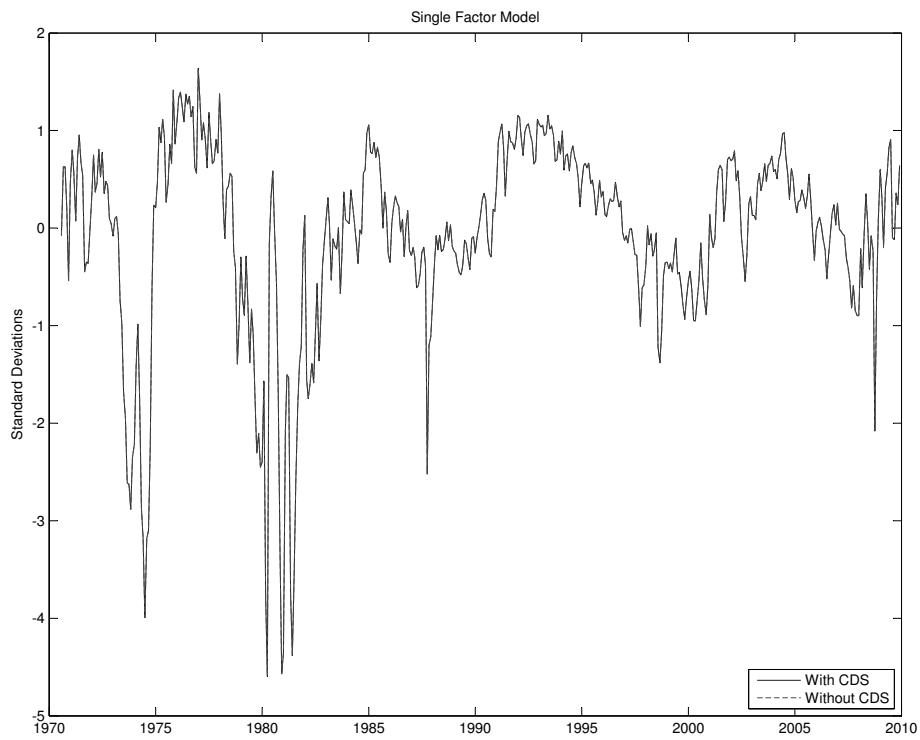
Figure 1.2: Dynamic factor model: 1 factor
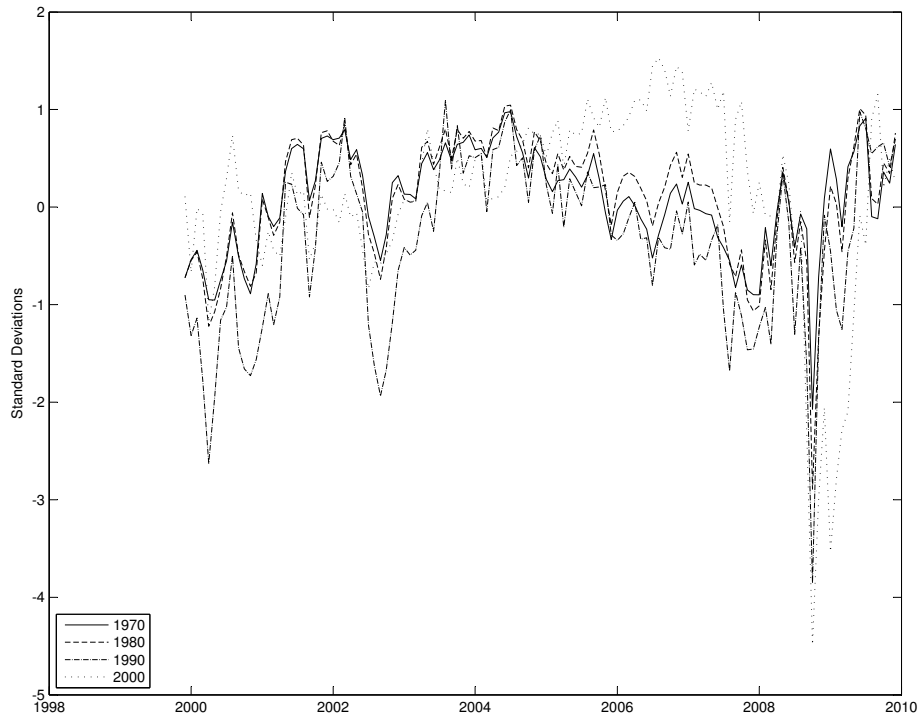
Figure 1.3: Variation of dynamic factors for different estimation periods

Table 1.1: Base Case: 3m Forecast : 3/1989 - 12/2009

	Employment			Industrial Production	
	Pr $W_1 > 0$	Adj. R^2		Pr $W_1 > 0$	Adj. R^2
Autoregressive	-	0.80		-	0.34
CDS	0.000	0.84		0.000	0.44

Note: Pr $W_1 > 0$ denotes the robust Wald test that the coefficients on CDS are jointly

nonzero.
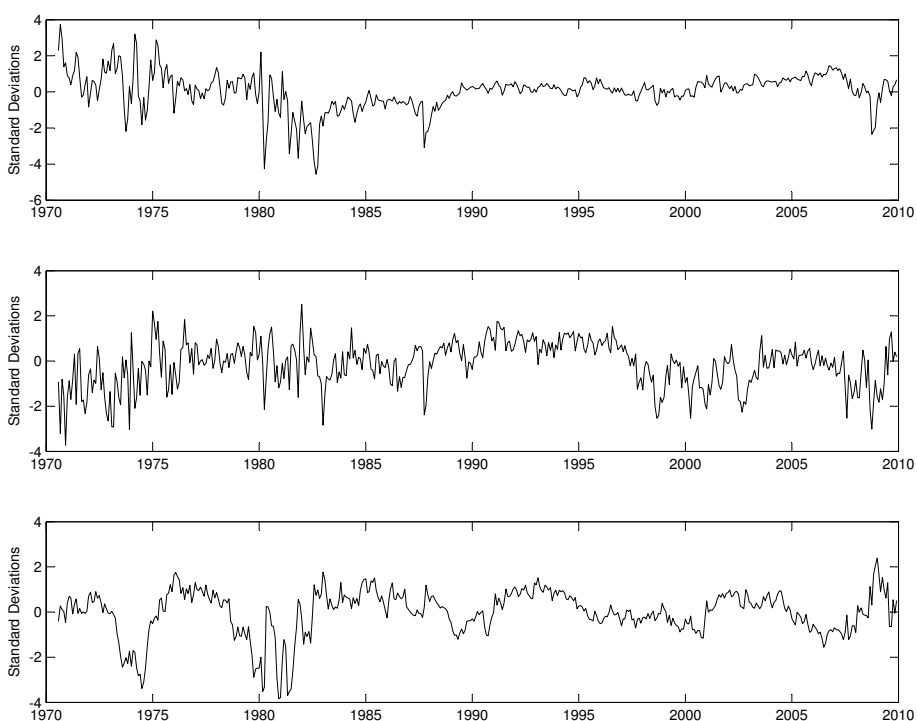
Figure 1.4: Dynamic Factors: 3 factor model

Table 1.2: In-sample predictive content of credit spreads: 3 Month Forecast of Industrial Production

	Industrial Production			
	Pr > W_1	Pr > W_2	Pr > W_3	Adj. R^2
Standard	0.000	-	-	0.627
CDS	-	-	0.000	0.503
Standard & CDS	0.000	-	0.008	0.652
Std & EDF-Q1 & CDS	0.002	0.000	0.275	0.694
Std & EDF-Q2 & CDS	0.001	0.004	0.070	0.659
Std & EDF-Q3 & CDS	0.000	0.072	0.004	0.666
Std & EDF-Q4 & CDS	0.000	0.058	0.139	0.675
Std & EDF-Q5 & CDS	0.002	0.188	0.276	0.659

Note: Sample period: Monthly Februrary 1990 to December 2008 except for CDS (begins in Januarly 2004). Pr > W_i for $i \in \{1,2,3\}$ denotes the Wald test for the null hypothesis that the coefficients on the 1) standard credit indexes or 2) EDF based credit spreads in a particular quantile or 3) CDS spreads are jointly equal to zero.

Table 1.3: In-sample predictive content of credit spreads: 3 Month Forecast of Employment

	Employment			
	Pr > W_1	Pr > W_2	Pr > W_3	Adj. R^2
Standard	0.000	-	-	0.852
CDS	-	-	0.000	0.821
Standard & CDS	0.000	-	0.000	0.870
Std & EDF-Q1 & CDS	0.000	0.002	0.109	0.877
Std & EDF-Q2 & CDS	0.000	0.039	0.077	0.873
Std & EDF-Q3 & CDS	0.000	0.041	0.000	0.875
Std & EDF-Q4 & CDS	0.000	0.000	0.004	0.886
Std & EDF-Q5 & CDS	0.000	0.000	0.242	0.883

Note: Sample period: Monthly Februrary 1990 to December 2008 except for CDS (begins in Januarly 2004). Pr > W_i for $i \in \{1,2,3\}$ denotes the Wald test for the null hypothesis that the coefficients on the 1) standard credit indexes or 2) EDF based credit spreads in a particular quantile or 3) CDS spreads are jointly equal to zero.

Table 1.4: Out-of-sample predictive content of credit spreads relative to standard indexes:
3 Month Horizon

	Employment			Industrial Production	
	RMSFE	Ratio		RMSFE	Ratio
Standard	0.0174	-		0.0735	-
Standard & CDS	0.0146	0.704		0.0811	1.217
Std & EDF-Q1 & CDS	0.0241	1.918		0.0885	1.450
Std & EDF-Q2 & CDS	0.0187	1.155		0.0970	1.742
Std & EDF-Q3 & CDS	0.0188	1.167		0.0855	1.353
Std & EDF-Q4 & CDS	0.0166	0.910		0.0801	1.188
Std & EDF-Q5 & CDS	0.0162	0.867		0.1157	2.478

Note: Sample period: Monthly Februrary 1990 to December 2008 except for CDS (begins in Januarly 2004). "Ratio" denotes the ratio of the RMSFE of each model relative to the model that just contains the standard credit indexes.

Table 1.5: Out-of-sample predictive content of credit spreads relative to standard indexes and EDF: 3 Month Forecast

	Employment			Industrial Production	
	RMSFE	Ratio		RMSFE	Ratio
Std & EDF-Q1	0.0227	-		0.0618	-
Std & EDF-Q1 & CDS	0.0241	1.127		0.0885	2.051
Std & EDF-Q2	0.0232	-		0.0869	-
Std & EDF-Q2 & CDS	0.0187	0.650		0.0970	1.246
Std & EDF-Q3	0.0227	-		0.0772	-
Std & EDF-Q3 & CDS	0.0188	0.686		0.0855	1.227
Std & EDF-Q4	0.0195	-		0.0778	-
Std & EDF-Q4 & CDS	0.0166	0.725		0.0801	1.060
Std & EDF-Q5	0.0153	-		0.1059	-
Std & EDF-Q5 & CDS	0.0162	1.121		0.1157	1.194

Note: Sample period: Monthly Februrary 1990 to December 2008 except for CDS (begins in Januarly 2004). "Ratio" denotes the ratio of the RMSFE of each model relative to the model that just contains the standard credit indexes and same EDF portfolio.

Table 1.6: Data used in dynamic factor models

	Introduction
Rate Variables:	
10y Treasury	5/1970
3m T-bill / Fed Funds spread	5/1970
2y Treas / 3m T-bill spread	6/1976
10y Treas / 3m T-bill spread	5/1970
TED spread	9/1981
Baa / 10y Treasury spread	5/1970
Auto financing / 2y Treasury	6/1976
30yr Mortgage / 10y Treasury spread	4/1971
Exchange Rate	1/1973
Price and Quantity Variables:	
Bank Credit	5/1970
Commercial Paper Oustanding	5/1970
Commercial Paper Issuance	1/2003
Money Stock	1/1974
Price of crude oil	5/1970
Correlation and Volatilities:	
Correlation betwee equities and treasuries	6/1976
Idiosyncratic bank stock volatility	1/1973
VIX	1/1986

NOTE: See Hatzius *et al.* (2010) for complete description and data sources. All data available from Mark Watson's website.

Table 1.7: In-sample predictive content of credit spreads: 3 Month horizon

	Industrial Production			Employment	
	$\Pr > W_1$	$\Pr > W_2$		$\Pr > W_1$	$\Pr > W_2$
HHMSW	0.000	-		0.000	-
CDS	-	0.003		-	0.001
HHMSW & CDS	0.000	0.235		0.000	0.052

Note: Sample period: Quarterly April 1970 to December 2009 except for CDS (begins in Januarly 2004). $\Pr > W_i$ for $i \in \{1, 2\}$ denotes the Wald test for the null hypothesis that the coefficients on the 1) HHMSW factors or 2) CDS spreads are jointly equal to zero.

Table 1.8: Industrial Production : In-sample predictive content of constructed factors at 3 Month horizon

	(1)	(2)	(3)
L0.IP	0.204	0.204	0.197
	(5.92)**	(5.93)**	(5.77)**
L1.IP	0.155	0.155	0.149
	(4.47)**	(4.48)**	(4.33)**
L2.IP	0.083	0.083	0.083
	(2.40)*	(2.40)*	(2.45)*
L3.IP	0.006	0.006	0.005
	(0.19)	(0.18)	(0.154)
Factor with CDS	0.019	-	0.199
	(3.17)**	-	(0.14)
L1.cds	0.01	-	0.624
	(1.17)	-	(0.30)
L2.cds	0.002	-	0.171
	(0.28)	-	(0.08)
L3.cds	-0.005	-	2.59
	(0.83)	-	(1.74)
Factor without CDS	-	0.019	-0.18
	-	(3.17)**	(0.12)
L1.no-cds		0.01	-0.614
		(1.16)	(0.30)
L2.no-cds		0.002	-0.168
		(0.28)	(0.08)
L3.no-cds		-0.005	-2.596
		(0.84)	(1.74)
Constant	0.015	0.015	0.017
	(5.51)**	(5.50)**	(6.16)**
Observations	469	469	469
R-Squared	0.37	0.37	0.4

Absolute value of t-statistics in parantheses.

* significant at 5%

** significant at 1%

Table 1.9: Employment : In-sample predictive content of constructed factors at 3 Month horizon

	(1)	(2)	(3)
L0.IP	0.087	0.087	0.086
	(8.46)**	(8.46)**	(8.35)**
L1.IP	0.069	0.069	0.069
	(6.64)**	(6.64)**	(6.62)**
L2.IP	0.056	0.056	0.055
	(5.42)**	(5.41)**	(5.36)**
L3.IP	0.036	0.036	0.038
	(3.92)**	(3.92)**	(4.04)**
Factor with CDS	0.005	-	0.416
	(2.70)**	-	(0.94)
L1.cds	0.002	-	0.295
	(0.72)	-	(0.47)
L2.cds	0.002	-	0.244
	(0.61)	-	(0.39)
L3.cds	-0.003	-	-0.291
	(1.59)	-	(0.64)
Factor without CDS	-	0.005	-0.411
	-	(2.69)**	(0.92)
L1.no-cds	-	0.002	-0.294
	-	(0.72)	(0.47)
L2.no-cds	-	0.002	-0.243
	-	(0.61)	(0.39)
L3.no-cds	-	-0.003	0.289
	-	(1.59)	(0.64)
Constant	0.015	0.015	0.017
	(13.02)**	(13.01)**	(13.25)**
Observations	469	469	469
R-Squared	0.53	0.53	0.53

Absolute value of t-statistics in parantheses.

* significant at 5%

** significant at 1%

Table 1.10: Pairwise factor correlation across estimation periods

	1970	1980	1990	2000
1970	1.000			
1980	0.956	1.000		
1990	0.870	0.913	1.000	
2000	0.208	0.457	0.468	1.000

Table 1.11: Dynamic factor mode: pseudo out-of-sample 3m forecast

	Employment		Industrial Production
	RMSFE		RMSFE
HHMSW	0.0190		0.0753
CDS	0.0190		0.0752

Table 1.12: BMA posterior attribution

	Employment			Industrial Production	
	CDS	EDF		CDS	EDF
Std & CDS	0.985	-		0.997	-
Std & EDF	-	0.999		-	1.000
Std & EDF & CDS	0.040	0.959		0	1.000

NOTE: Entries are $\sum_i P(M_i|D)$, where the sum is taken over all models classified as CDS models or EDF, respectively.

Table 1.13: BMA pseudo out-of-sample predictive content: 3 month horizon

	Employment	Industrial Production
	RMSFE	RMSFE
Standard	0.0192	0.0753
Std & CDS	0.0192	0.0753
Std & EDF	0.0191	0.0629
Std & EDF & CDS	0.0191	0.0629

Note: Sample period: Monthly Februrary 1990 to December 2008 except for CDS which begins in January 2004.

2 International Employment and the Business Cycle: New Stylized Facts with an Application to the Great Moderation

2.1 Introduction

Since documenting the substantial reduction in the volatility of output in the early 1980s, empirical research has uncovered a widespread reduction in the volatility of macroeconomic time series roughly coinciding with that in output. Stock and Watson (2002) and Sensier and van Dijk (2004) examined the break in volatility in more than 100 US macroeconomic time series. The vast majority coincided with the decline in output volatility. The mechanisms underlying this decline remain controversial and falls broadly into three categories; i) good luck, ii) good policy, or iii) structural change. The "good luck" hypothesis posits the recent period of stability is due to a string of smaller than normal exogenous shocks. This represents the alternative hypothesis to the other methods considered, which we discuss in turn. The "good policy" hypothesis suggests the improvements in monetary policy have "tamed" the business cycle.[1] Finally, the "structural change" hypothesis posits a fundamental change in institutions (e.g. labor markets) or household preferences has altered the business cycle.

The regime shift to Volcker's intense focus on inflation in the early 1980s coupled with increased transparency at the Federal Reserve and rapid innovation in macroeconomic theory all point to increasingly effective monetary policy.[2] Note that the decline in inflation

---

[1] See Galí and Gambetti (2009) for a brief survey of these mechanims.

[2] In particular the departure from large-scale macroeconomic models following the Lucas' (1976) critique and Sims' (1980) cogent case for VAR analysis. The empirical success of Volcker-Greenspan policy

volatility is not restricted to the US. Cecchetti *et al.* (2006) examine 24 countries and find that in 11 countries both inflation and output volatility fell. In an additional 9 countries inflation volatility fell substantially while output volatility rose modestly or was unchanged. Importantly, none of the countries saw an increase in both inflation and output volatility.

DSGE models provide primarily support for good policy results. Theoretically, Clarida et al. (2000) show that sufficiently passive monetary policy is unable to overcome individuals' expectations. Consequently expectations become self-fulfilling resulting in multiple-equilibria. In their calibrated model, Volcker shifted the US from an indeterminate to a determinate equilibrium thereby reducing the volatility. Lubik and Schorfheide (2004) subsequently provide an estimator that is valid for both determinate and indeterminate regimes. They apply their estimator to a DSGE model and find that the pre-Volcker regime was in an indeterminate equilibria.

VAR models mostly point to good luck. For those VAR models finding a role for monetary policy, the effect is not large and depends upon the greater role of demand shocks impacting the economy.[3] Internationally, Canova *et al.* (2007) consider the US, UK, and Euro Zone in a structural Time-Varying Parameter (TVP)-VAR. They find international co-movement in inflation and nearly independent output which is inconsistent with a good policy story. Furthermore they find that the interaction of supply and monetary shocks drive output volatility in the US, whereas it is demand and monetary shocks in the Euro Zone, and solely supply shocks in the UK.[4]

One possible reconciliation between the DSGE and VAR results is that VARs fail to account for multiple equilibria. Using the Lubik and Schorfheide (2004) estimator, Benati and Surico (2009) find that the difference between DSGE and VAR results are due to the failure of VAR models to account for multiple equilibria pre-Volcker. They show that VAR results with indeterminacy are observationally equivalent to those without, however only the former implies a role for monetary policy.

Despite their theoretical advantages, DSGEs have estimation problems as well. Canova [2006b]

[2006a] shows that the Benati and Surico (2009) critique is specific to their methodology

---

culminated in the now-standard Taylor (1993) rule for monetary policy.

[3]See, e.g., Boivin and Giannoni (2006).

[4]Further VAR results include Primiceri (2005), Sims and Zha (2006), and Gambetti, Pappa, Canova (2008).

and provide a model with determinacy that is able to reproduce the dynamics with indeterminacy. Their more general conclusion is that greater care needs to be taken in order to match the identifying restrictions in the structural VAR to the underlying DSGE model. Furthermore, Canova and Sala (2009) show that there are substantial identification problems with DSGEs themselves.

Beyond the technical difficulties in estimation, there are two further problems for the good policy hypothesis. First, it is difficult to point to examples of monetary regime change outside of the USA under Volcker and the UK's exit from the European Exchange Rate Mechanism (ERM) in 1992. Not only is the UK the only other example, but Benati (2008) finds that monetary policy cannot even explain the UK's volatility dynamics. Second, the recent financial crisis is marked by its lack of monetary policy change. Whereas Volcker's chairmanship marked a clear change in monetary policy and the beginning of the Great Moderation, Bernanke was appointed to extend the successful Volcker-Greenspan regime. With plausible continuity in policy there must be a non-policy explanation for the Great Moderation's abrupt end.

Several structural (non-monetary) policy changes have been proposed; inventory management, financial frictions, and labor frictions. Improved inventory management was originally proposed in McConnell and Peres-Quiros (2000) and Kahn and McConnell (2002). This view has fallen out of favor based on both theoretical and empirical grounds.[5] Financial frictions have not undergone the same sort of scrutiny in this context. One example is Justiniano and Primiceri (2008) who propose a large-scale DSGE model with time-varying volatilities that accounts for the pre-Volcker indeterminate equilibrium. They find that the majority of the decline in output volatility is due to a shock which they interpret as representing financial frictions. However they stress this is, at best, a reduced form interpretation. Explicitly including financial frictions will alter the model beyond just the investment relation. While this may explain the onset of the Great Moderation, it does not explain its end. If anything, financial services continued to be deregulated and financial innovation expanded the universe of credit instruments during this time.[6] The Great Moderation was punctuated by several financial crises prior to 2008.[7] Perhaps the reduction of

---

[5]For theoretical arguments see Maccini and Pagan (2003), Kahn and Thomas (2007). For empirical results see Ahmed et al (2004), Ramey and Vine (2006).

[6]For a survey of the scope and benefits of financial innovation, see Litan (2010)

[7]Most notably LTCM in 1998 and the S&L Crisis in the mid-1980s.

explicit financial frictions was offset by a rise in systemic risk that is not captured by the reduced form of the model. We leave that for future research.[8]

The remaining hypothesized structural change is labor market frictions. Labor represents two thirds of national income.[9] Consequently, a small change in labor frictions could cause large changes in aggregate fluctuations.

In addition to the decline in hours volatility, there are three additional stylized facts in the literature. The first is the large decline in the correlation between labor productivity and output in the US that coincides with the Great Moderation. Kydland and Prescott (1982) used the high correlation between labor productivity and output to support the real business cycle (RBC) model. In the RBC model positive aggregate technology shocks increase the marginal productivity of labor leading to an increase in employment during booms, and vice versa during recessions. However, the rapid decline in the correlation to near zero in the mid-1980s undermined their supporting stylized fact. Subsequent research has tried to revive the role of technology shocks in driving the business cycle by introducing frictions with mixed results.[10] Second, Stiroh (2009) documents a stark decline in the correlation between labor productivity and hours in the US also coinciding with the Great Moderation. Furthermore, this result holds using both aggregate and disaggregate manufacturing data. A final stylized fact was uncovered by Galí and Gambetti (2009). They consider a model with labor productivity and output in a time-varying parameter (TVP-) VAR with stochastic volatility. They find that the volatility of hours and output both decline in the US, however hours declined less than output. Interestingly they also calculate the correlation between labor productivity and output and find that it remains significantly pro-cyclical. However they do not discuss this anomaly.

We provide three contributions. First, the stylized facts given above are obtained using different data sets and different statistical methodologies. There are several methods to deal with the nonstationarity induced by the structural change of the Great Moderation.

---

[8]Stock and Watson (2005), in a VAR, study the international business cycle for the G7 and find that international shocks have declined. This may be due to the reduction in international trade barriers and globalization of finance. Additionally the countries seem to have split into two cyclically coherent groups: English and non-English speaking countries.

[9]This is true internationally, see Gollin (2002).

[10]For early arguments for and against the RBC model see the Summer 1989 issue of The Journal of Economic Perspectives. Specifically, Mankiw (1989) and Plosser (1989). A brief review of the merger between the RBC and the search and matching literature can be found in Ramey (2011).

We find, in the US, most of the labor market stylized facts are robust to statistical method. The exception is the correlation between labor productivity and output, the linchpin of the RBC model.

Second, we extend the set of stylized facts to thirteen OECD countries using a new data set constructed by Ohanian and Raffo (2011). Previously, international labor data was only available on a capitation employment rather than hourly basis. Since modern macroeconomic models find support for adjustment of both the intensive and extensive margins, we would expect this to be an important innovation. Furthermore the extensive literature in the US utilizes hourly data. This data provide the first comprehensive means of comparing labor and output internationally. We document significant international heterogeneity and provide a new set of stylized facts.[11]

Finally, we consider the model of Galí and van Rens (2010) as a possible explanation for the observed heterogeneity. Galí and van Rens (2010) provide a theoretical model that is able to match employment and output statistics solely using labor market frictions. Specifically, they consider three empirical regularities in US employment: i) the decline in the procyclicality of labor productivity with respect to output and labor input [Stiroh (2009), Uhlig (2010)], ii) the increase in the volatility of labor input relatve to output [Galí and Gambetti (2009)], and iii) the rise in the absolute and relative volatility of the real wage [Galí and van Rens (2010)]. Their model allows for endogenous effort and labor adjustment costs. Calibrated to US data, moving from a completely rigid to a completely flexible labor market generates all of the stylized facts, including a reduction in output volatility. However, they caution that the calibrated model only delivers a modest reduction in output volatility.

In addition, their results are illustrative rather than quantitative. The US labor market was never completely rigid and is not today completely flexible. Hence whether the US experienced an economically meaningful shift in labor market frictions remains an empirical question. The stylized facts we uncover thus predict a change in labor market frictions within their model. We compare those predictions to a set of labor market frictions using

---

[11]Ohanian and Raffo (2011) look at similar moments, however they use only a single method and a single breakpoint. They estimate the HP filter over two subsamples and calculate the difference. This will be discussed in more detail in section 2.2.2. Their use of a single breakpoint corresponding to the US change in 1984 is inappropriate given the asynchronous onset of the Great Moderation. See, e.g. Blanchard and Simon (2001) and Stock and Watson (2005).

data collected by the Amsterdam Institute for Advanced labor Studies (AIAS). We find that these measures do not match the predictions and tentatively conclude that a reduction in labor market frictions is not the main driver of these moments internationally. However this does not refute the possibility that it could still apply to the US. We then suggest an alternative explanation of the data.

The paper proceeds as follows. Section 2.2 introduces the statistical methods and applies them to the US data. Section 2.4 then considers the international evidence and highlights those countries consistent with the US experience. Section 2.5 presents the labor market frictions and possible explanations for our international results. Section 2.6 concludes.

2.2   US Stylized Facts

To begin we review the stylized facts found in the literature for the US. These stylized facts are statements on the non-stationarity of the respective series. To be precise, the series are typically assumed to be stationary in two subperiods with a one-time structural break. However, that is not the only possible form of non-stationarity. Consequently, we consider whether the stylized facts are robust to alternative specifications of non-stationarity. It is critical that the stylized facts we use to build models are robust to statistical method. To quote den Haan (2000): "Macroeconomic models are judged on their ability to reproduce key correlations in the data. Using these kind of empirical results to judge theories presupposes that there is a set of correlations upon which everyone can agree." In particular, we allow for continuous change in the second moments.

In addition to the decline in output volatility we consider two stylized facts found in the literature. These are that i) labor productivity has become less procyclical with respect to output and labor input, and ii) the volatility of labor input relative to output has increased. Let y_t denote output, ℓ_t labor input, and $x_t = y_t - \ell_t$ be labor productivity, all in logs. Then the stylized facts can be expressed as:

$$\sigma_y \downarrow$$
$$corr(x_t, y_t) \downarrow$$
$$corr(x_t, \ell_t) \downarrow$$
$$\frac{\sigma_\ell}{\sigma_y} \uparrow$$

Our techniques fall into two categories: i) first estimate the structural break date, and then compare the two sub-periods, and ii) estimate a continuous, time varying measure of the covariance matrix. In the first case, each subperiod is stationary and we compare a set of statistics restricted to each subperiod. In other words, the only dynamic is the change across the regimes.

While this two stage estimation is relatively simple an important caveat deserves mention. Cogley and Sargent (2005) and Benati (2007) provide Monte Carlo evidence for the lack of power in state-of-the art structural break tests when the true model is a random walk. Unfortunately, in our application labor productivity is typically assumed to follow a random walk, e.g. identification of technology shocks from long-run restrictions on labor productivity [Galí (1999)]. In fact, Benati (2007) concludes that "when time-variation in equilibrium productivity growth does take place, it takes place most likely gradually...so that the best way of analysing it is via time-varying parameters models, rather than via break tests."

Consequently, we allow continuous variation and use the entire time series to estimate the dynamics. This allows a much more detailed view of the dynamics, rather than a simple sign comparison. An additional benefit noted by Benati (2007), is that allowing varying VAR coefficients is a good approximation even if there is but a single break whereas first estimating a break and comparing subperiods is valid only if trend-breaks is the correct DGP. However, this requires estimating a much larger, more complex model and the time-varying parameters are therefore less precisely estimated.

In the rest of this section we will present our five statistical models and apply them to the US. Our goal is to ascertain whether each method delivers all of the US stylized facts found in the literature. This documents a baseline to compare international results to in section 2.4. Before summarizing the models we first present our data.

Table 2.1: Hours per Worker: Sample Periods

Australia	1970-2010		Italy	1960-2010
Austria	1965-2010		Japan	1960-2010
Canada	1960-2010		Norway	1960-2010
Finland	1960-2010		Sweden	1975-2010
France	1960-2010		UK	1971-2010
Germany	1960-2010		U.S.	1960-2010
Ireland	1960-2010			

2.2.1 Data

The data are GDP and hours for thirteen countries roughly over the period 1960q1-2010q4, with the analysis limited by the availability of hours data. Table 2.1 provides the time periods for which hours per capita are available for each country. GDP data are real chain-weighted indices. Hours are establishment data. Both are constructed to match National Income and Product Account (NIPA) conventions. In addition, both series are standardized by the working age population 15-64 years and converted to logs. Labor productivity is given as the log difference between GDP and hours.

The data are from a new publicly available dataset constructed in Ohanian and Raffo (2011).[12] They assemble annual data from the OECD, national statistical agencies, and Groningen Growth and Development Centre (GGDC). They then backcast these official series using quarterly International Labor Orginization (ILO) and, rarely, OECD Main Economic Indicators (MEI) data on hours. They use the method from Deaton (1971)[13] to ensure the quarterly series matches the more accurate annual time series. The exact details of its construction can be found in Ohanian and Raffo (2011), but we note the estimation used to construct hours does not use GDP.

Note that Brügemann, Hagedorn, and Manovskii (2010) and Hagedorn and Manovskii (2011) find that the choice of labor input has significant effects on the labor market statistics in the US. Specifically, there are significant differences between the Current Population (household) Survey and the more popular Current Employment (establishment)

---

[12]The data was kindly provided by the authors.

[13]This is used, for instance, to construct the Industrial Production series.

Survey. Ohanian and Raffo (2011) use the latter definition. We admit that the results may vary with alternative definitions of employment, however we make several observations. First, prior to this data set, European labor and productivity analyses have predominantly relied upon per-worker concepts whereas US research has focused per-hour measures. Insofar as we believe stylized facts in the labor market correspond to adjustments along both the intensive and extensive margins, the most appropriate measure is hours data. Second, for the international comparisons to be sensitive to the choice of series the discrepancy between the series would have to vary systematically across countries. Third, this is the first internationally consistent data set for hours. While the results may be sensitive to the choice of household versus establishment survey, neither was previously available. Lastly, the sensitivity has only recently been examined in the US. Using the much more popular establishment data allows us to compare our results to a much larger US literature.

2.2.2   Filters and Rolling Window

Our first method considers two detrending methods that are used to extract the business cycle: the Hodrick-Prescott (HP) and Baxter-King (BK) filters[14]. The former's trend can be formulated as a ridge regression on time with a smoothing parameter (λ)[15], while the latter directly considers the frequency domain via a band-pass filter. For the BK filter this involves setting the frequency band directly, $[\phi_{LO}, \phi_{HI}]$ and the number of lead/lags to use in the approximation, k.[16] In either case, the goal is to recover the business cycle which is typically defined as frequencies of 6-32 quarters. Following Baxter and King (1999) and Ravn and Uhlig (2002) the tuning parameters are set to the optimum values for quarterly data under standard assumptions; $\lambda = 1600$ in the HP filter and $\{\phi_{LO}, \phi_{HI}, k\} = \{6, 32, 12\}$ in the BK filter. Although filters are used extensively in the literature, they differ in several aspects. First, they differ in their end point properties. As our sample ends just after the 2008 financial crisis, these end point problems can be particular severe. Second, they differ in how they react to aggregation. To be precise, let y_t denote log output, ℓ_t denote labor log input, and $x_t = y_t - \ell_t$ denote log labor productivity. Then we have that

---

[14]Details can be found in Hodrick and Prescott (1997) and Baxter and King (1999).

[15]See Ley (2006).

[16]The BK filter is an approximation to an ideal filter. An ideal filter, among other things, requires an infinite series. k determines the length of the approximating series. For more details, see Baxter and King (1999).

$HP(x_t) = HP(y_t - \ell_t) = HP(y_t) - HP(\ell_t)$, however the same does not hold for the BK filter. Consequently we may obtain significantly different results when we look at the correlations with labor productivity.[17]

In practice, researchers account for the structural break in the early 1980s by splitting the sample, filtering on each subperiod, and calculating the relevant statistic. We then take the difference between the two subperiods as the evidence for the structural change.[18] The break date is the onset of the Great Moderation as calculated in, e.g. McConnell and Perez-Quiros (2000) or Stock and Watson (2005). We estimate an American break date of 1983q4 which compares favorably with 1983q2 in Stock and Watson (2005) and 1984q1 in McConnell and Perez-Quiros (2000). We also calculate the break date for all countries in our sample, details of which can be found in section 2.4.1.

Table 2.2 contains the standard deviations and correlations for HP and BK filtered data. The results are consistent with the stylized facts. The standard deviation of GDP and hours declines, as does the correlation between productivity and GDP and productivity and hours. The results are similar between the two methods.

The difference between the two subperiods provides a single summary statistic. However, the dynamics of these moments are also of interest, especially mean reversion and break speed. Six-year rolling windows can give us a sense of these dynamics. The results appear in figures 2.1 and 2.2. Volatility of GDP and hours both decline dramatically in the mid-1980s and remain subdued. Even in the recent recession the GDP volatility does not ascend to the heights reached in the 1960s. The volatility of hours, however, returns to levels not seen since prior to the hyperinflation in the 1970s.

The correlations also decline in the mid-1980s, however the procyclicality between productivity and GDP steadily increases after the initial decline. This suggests a more nuanced story than the typical narrative based on simple two period correlations. We see here that the decline in correlation was temporary and actually increased over most of the second half of the sample.

---

[17]Brügemann, et al. (2010) and Hagedorn and Manovskii (2011) consider the impact of the definition of labor input and filter choice on the stylized facts found in the US. They find that the choice of filter is unimportant in the US, however the definition of labor input significantly changes the results. See the description of the data in section 2.2.1.

[18]See, e.g., Galí and van Rens (2010).

Table 2.2: Second Moments of Filtered Data

	HP			
	Pre-84	Post-84	Difference	p-value
Std error GDP	7.60	4.38	-3.22	0.00
Std error Hours	6.46	5.23	-1.23	0.02
Std error Productivity	3.68	2.84	-0.84	0.05
Corr(prod, GDP)	0.53	-0.01	-0.54	0.00
Corr(prod, Hours)	0.05	-0.55	-0.60	0.00
	BK			
	Pre-84	Post-84	Difference	p-value
Std error GDP	7.29	3.74	-3.55	0.00
Std error Hours	6.37	4.43	-1.94	0.00
Std error Productivity	3.26	2.27	-0.99	0.01
Corr(prod, GDP)	0.49	-0.03	-0.52	0.00
Corr(prod, Hours)	0.05	-0.54	-0.59	0.00

2.2.3 Den Haan VAR forecasting errors

In section 2.2.2 we examined the volatility and correlations without imposing any structure on the data beyond a single structural break. Here we impose the minimal multivariate structure of an (non-structural) VAR. However, it is still a time-invariant VAR and so we retain the single break assumption. In sections 2.2.4 and 2.2.5 we relax the single break assumption by estimating two continuous volatility techniques: GARCH and TVP-VAR.

Using VAR forecasting errors to consider the comovement of multiple time series was introduced by den Haan (2000). Den Haan (2000) was motivated by the disagreement over empirical results stemming from the use of a single unconditional correlation. This single correlation can be very sensitive to the methods used to calculate it. Instead he proposed the use of calculating the correlation of VAR forecasting errors calculated over a set of horizons. This provides information on the dynamics of the correlation structure that is lost when considering a single summary statistic. This method can also be used to
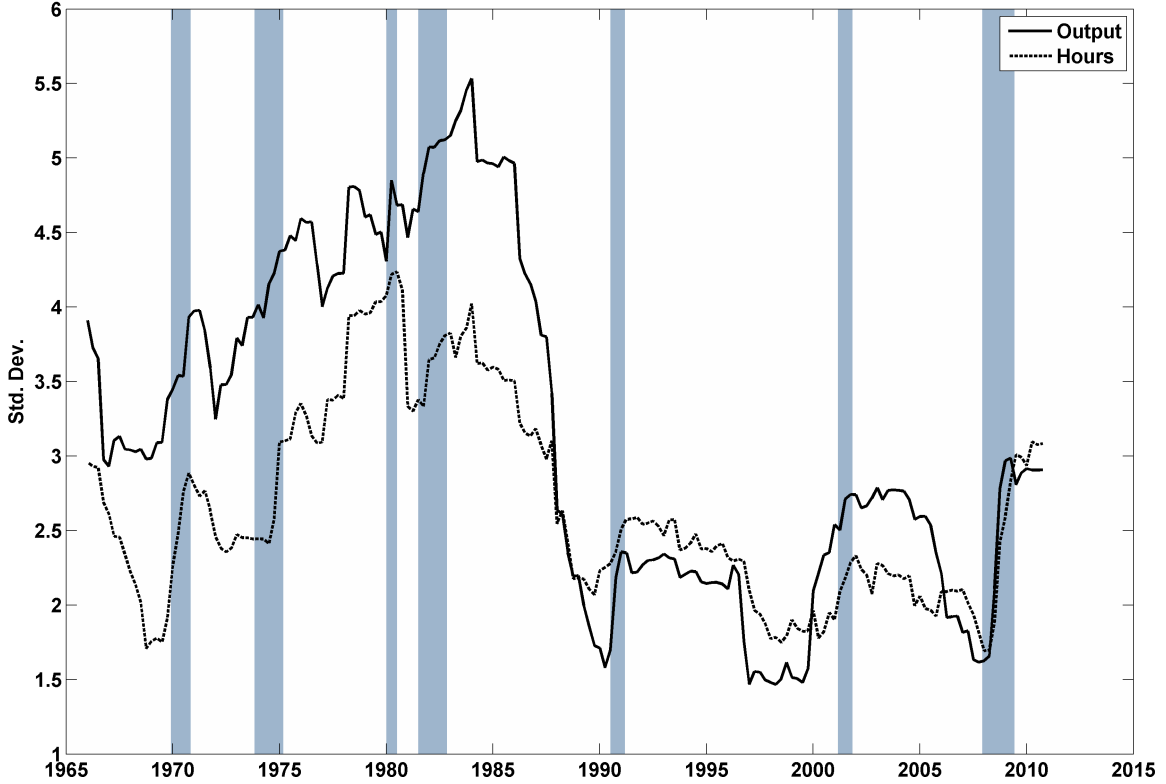
Figure 2.1: Annualized US Quarterly Output and Hours Growth: 6-year rolling window

investigate volatilities, although that possibility is not considered in den Haan (2000).[19] We follow the intuition found in den Haan (2000) below and present volatility estimates at the end.

To fix ideas, assume we are interested in the comovement of y_t and x_t. Let Z_t be an N-vector of endogenous regressors, which includes at least y_t and x_t, and consider the following VAR:

$$Z_t = \mu_t + Bt + Ct^2 + \sum_{j=1}^{L} A_j Z_{t-j} + \varepsilon_t \qquad (2.1)$$

where A_j is an $N \times N$ matrix of regression coefficients, μ, B, and C are N-vectors of constants, and ε_t is an N-vector of innovations, and the total number of lags is equal to L. Denote the K-period forecast errors as y_{t+K}^{ue} and x_{t+K}^{ue}. Then we are interested in the covariance between these two errors, COV(K).

The components of Z_t can be any combination of stationary and arbitrarily inte-

---

[19]We thank James Hamilton for pointing out this possibility.
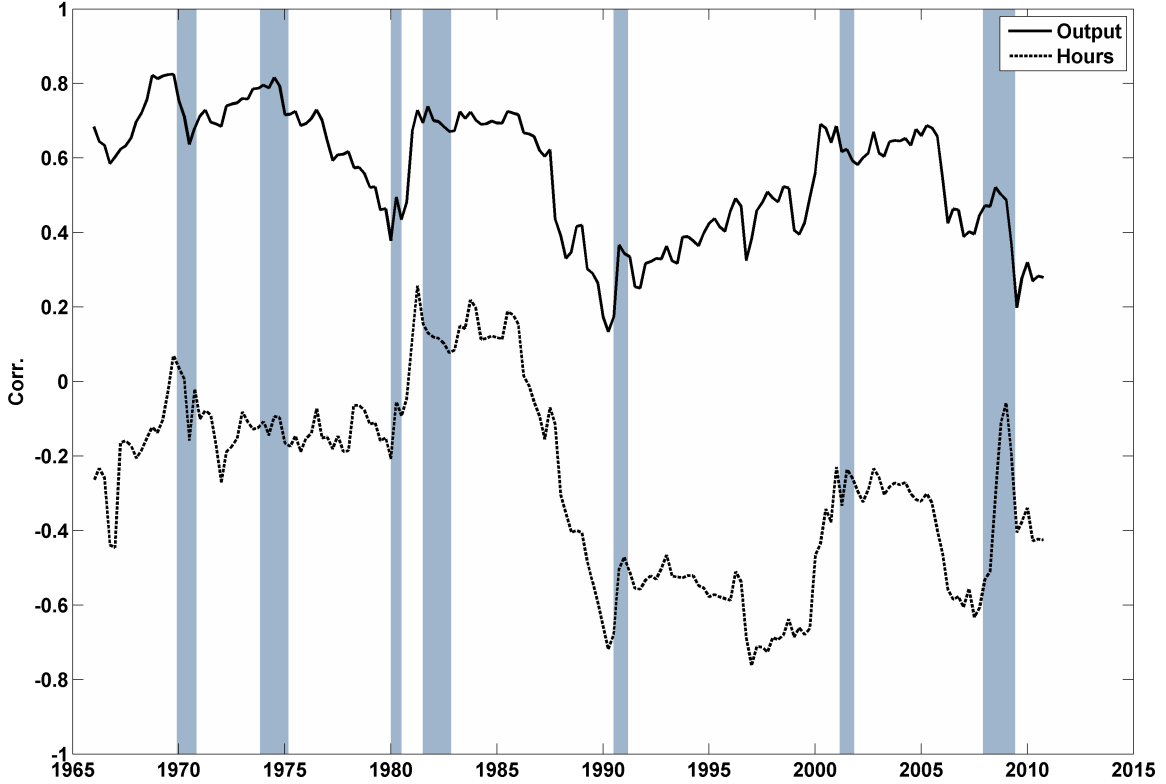
Figure 2.2: US Quarterly Productivity Correlation: 6-year rolling window

grated time series. Den Haan shows that COV(K) will be consistently estimated, for fixed K, even if Z_t is not stationary. This is true so long as (2.1) is well-specified. In particular, if it contains sufficient lags to ensure ε_t is not integrated.

In addition the forecast error covariances can be considered consistent estimates of the covariances implied by the true impulse-response functions. To see this, rewrite the K-period forecast error as the sum of forecast updates:

$$y^{ue}_{t+K,t} = (y_{t+K} - E_{t+K-1}y_{t+K}) + (E_{t+K-1}y_{t+K} - E_{t+K-2}y_{t+K}) + ... + (E_{t+1}y_{t+K} - E_t y_{t+K})$$

Denote the covariance between the kth terms as:

$$COV^{\triangle}(k) = COV\left[(E_{t+K-k+1}y_{t+K} - E_{t+K-k}y_{t+K}), (E_{t+K-k+1}x_{t+K} - E_{t+K-k}x_{t+K})\right]$$

then, since the forecast errors are serially uncorrelated, there is a simple relationship be-

tween $\text{COV}^{\triangle}(k)$ and $\text{COV}(K)$.

$$COV(K) = \sum_{k=1}^{K} COV^{\triangle}(k)$$

den Haan then shows that $\text{COV}^{\triangle}(k)$ is equal to the sum-product across all the fundamental shocks of the impulse-responses after k periods of the underlying series of interest. In other words, assume there are M fundamental shocks and let $y_k^{imp,m}$ denote the response after k periods to a one standard deviation change to the mth fundamental shock. Then we have

$$COV^{\triangle}(k) = \sum_{m=1}^{M} y_k^{imp,m} x_k^{imp,m} \qquad (2.2)$$

For $M = 1$, $\text{COV}^{\triangle}(k)$ is exactly equal to the product of the impulse responses. For $M > 1$, first note that the average absolute value is a good estimate of the standard deviation. Then equation (2.2) implies $\text{COV}^{\triangle}(k)$ measures the comovement after k periods where each model's fundamental shocks are set equal to its mean absolute value. Therefore $\text{COV}(K)$ measures the cumulative impact of these average impulse-responses.

Why is this important? A popular strategy to estimate DSGE models is to match the impulse-responses from an identified structural VAR and the analogous objects in the DSGE.[20] Those impulse-responses then characterize our object, $COV^{\triangle}(k)$, however they are subject to strict identifying restrictions. Instead, we obtain a consistent estimate of $COV^{\triangle}(k)$ directly under minimal assumptions. $COV^{\triangle}(k)$ then imposes restrictions on the impulse-response dynamics even though the impulse-responses themselves are unobserved and, indeed, unidentified.

This is especially important here given the criticisms by Fernald (2007) and Francis and Ramey (2009) of the long-run restrictions for identification introduced in Galí (1999).[21] More generally this addresses the criticism of VAR analyses brought by Benati and Surico (2009). They criticize the lack of connection between structural VARs and the underlying theoretical models. Here we bypass those identifying assumptions by consistently estimating a moment of the impulse-responses, although it yields weaker conclusions.

Figures 2.3 and 2.4 show the correlation results for the US. The figures depict the correlations between productivity and GDP, and productivity and hours, respectively. The

---

[20]See, e.g. Christiano, Eichenbaum, and Evans (2005)

[21]These restrictions are used in section 2.2.5 in the time-varying parameter VAR of Galí and Gambetti (2009).
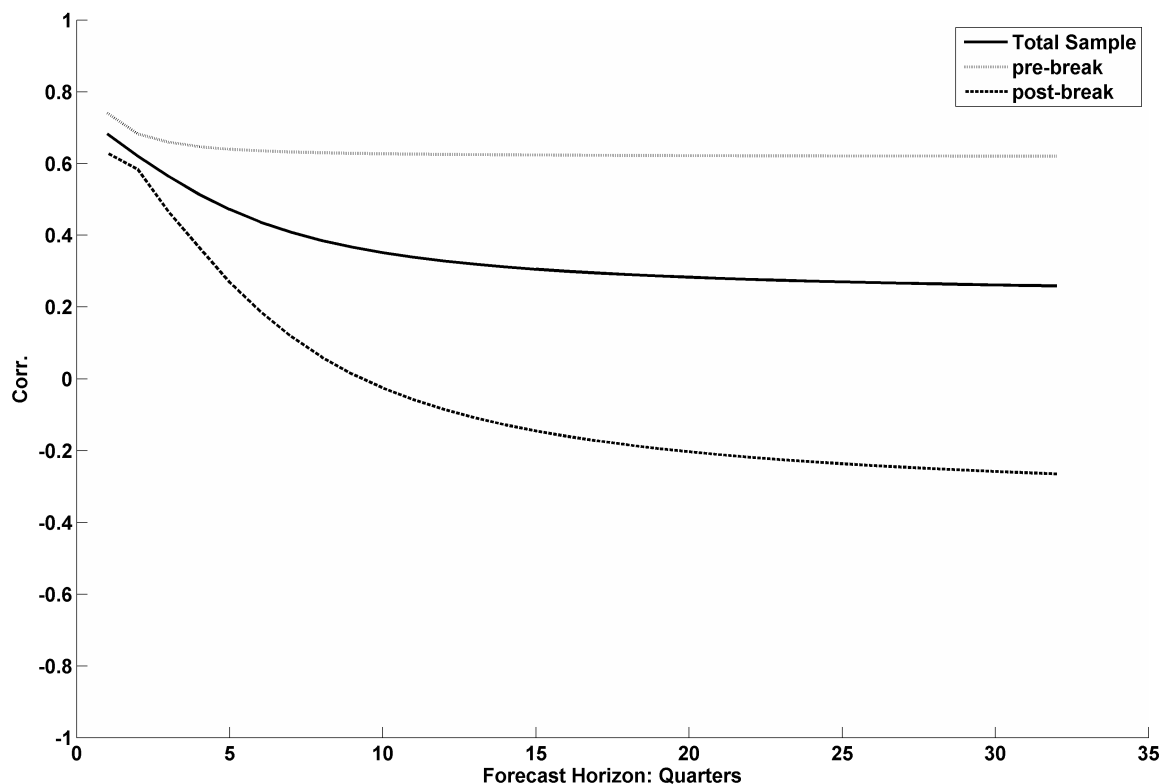
Figure 2.3: den Haan VAR forecast correlation: US hourly productivity and output

VAR lag length is determined by BIC for each country. The correlations are for horizons from 1 to 32 quarters, with 6 to 32 quarters considered the business cycle. Finally three separate estimates are depicted for each correlation, i) the full sample, ii) prior to the volatility break in 1984 and iii) after 1984. For both GDP and hours we see the correlations decline uniformly in the post-1984 period. Further for GDP the correlations are generally less than zero at business cycle horizons, and are all below zero for hours. This matches results found using separately filtered data for GDP and the volatility accounting procedure for hours in Stiroh (2009).

Although the explanation given above is given in terms of covariances, the identity $Cov(y,y) = Var(y)$ shows that the analysis goes through in terms of variances. Figure 2.5 depicts the results. We see that output volatility declines after the break date for all horizons. For hours, volatility rises for horizons greater than or equal to ten quarters but declines for horizons less than ten quarters. Thus, hours volatility generally rises for business cycle frequencies. This is in contrast to the filtered and rolling window results, which
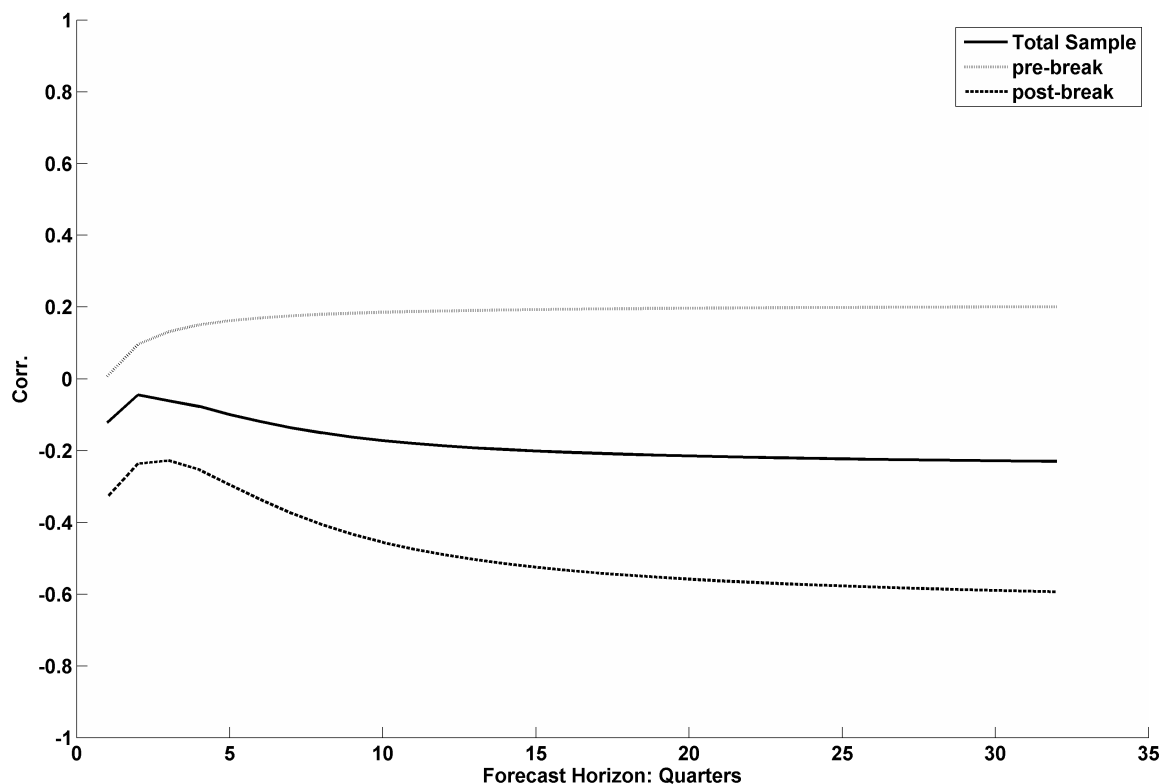
Figure 2.4: den Haan VAR forecast correlation: US hourly productivity and hours

show a decline in hours volatility at business cycle frequencies. Here we are conditioning on past productivity as well as past hours, rather than just past hours. This suggests that the business cycle response is more nuanced than the individual filtering results implies. In this framework, the Great Moderation is associated with a rotation in hours volatility conditional on the business cycle; a decline in short-term volatility but an increase in long-term volatility.

2.2.4 Multivariate GARCH

Next we consider explicitly modeling the time-varying volatilties and correlations. This relaxes the single break point assumption made in the previous sections while retaining an explicit multivariate structure. We gain power and efficiency by eliminating the problematic first-stage structural break tests[22] and using the entire sample period. There are two approaches, GARCH and stochastic volatility models. Here we consider GARCH

---

[22]See the discussion in section 2.2. Also Cogley and Sargent (2005) and Benati (2007).
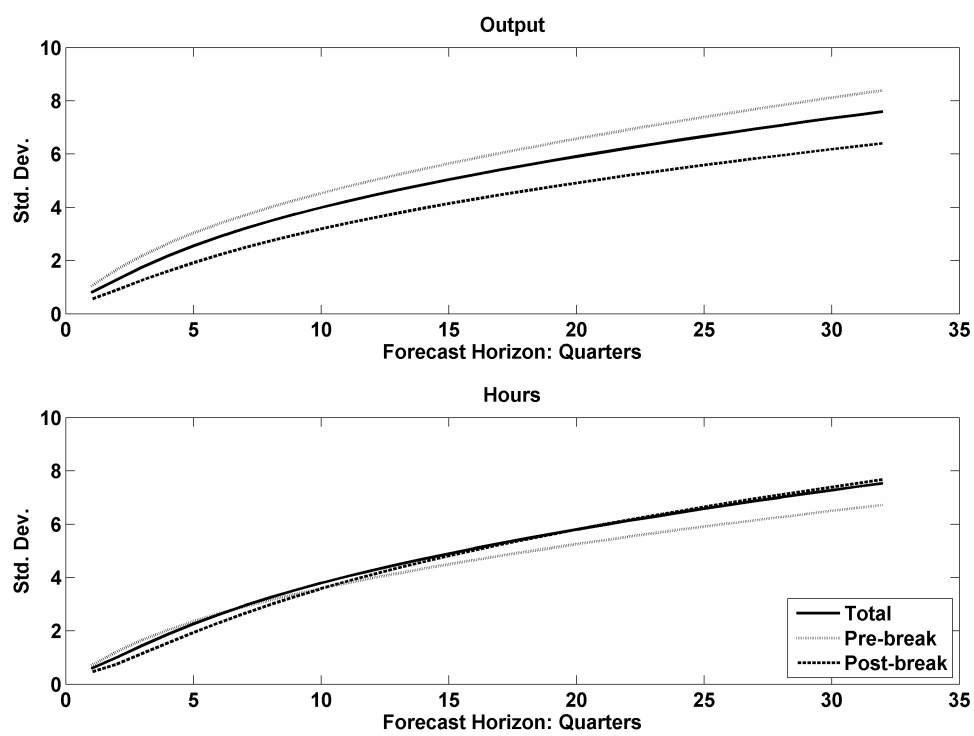
Figure 2.5: den Haan VAR forecast volatility: US output and hours

models and in the following section present a stochastic volatility model.

The unconstrained multivariate GARCH model is too complicated to bring to data. There exist a number of simplifying parameterizations that involve restricting the off-diagonal elements of the covariance matrix or the autoregressive structure. For instance, Bollerslev *et al.* (1988) use the vec operator to eliminate cross-terms from the ARCH specification. However this had difficulty ensuring positive definiteness of the covariance matrix. The BEKK model of Engle and Kroner (1995) provides sufficient conditions on the VEC model parameters to ensure positive definiteness. Reducing the dimension can be accomplished by assuming a factor structure, leading to the F-GARCH specification of Engle *et al.* (1992), but this is just a special case of the BEKK model. For a recent survey of the many methods and their complications, see Bauwens *et al.* (2006).

Here we implement the Engle (2002) dynamic conditional correlation (DCC-) GARCH model.[23] This has two appealing features. First, it is a two step procedure whereby we first estimate separate univariate GARCH models and then, taking those parameters as given, estimate the correlation structure. This estimation is simpler than fully multivariate GARCH and avoids those models' difficulties with convergence and strong assumptions required to ensure positive definiteness. Second, this fits neatly into our investigative paradigm. We are interested in both the time-series of the volatilities for each series as well as their comovement. This procedure returns the volatilities as an intermediate output in the first step. Anticipating our conclusion, it also provides a simple extensible framework to consider a larger space of structural mechanisms for the Great Moderation.

Following the notation in Engle (2002) the model is given by:

$$
\begin{aligned}
v_{t|t-1} &\sim N(0, D_t R_t D_t) \\
D_t^2 &= \mathrm{diag}\{\omega_i\} + \mathrm{diag}\{\kappa_i\} \circ v_{t-1} v_{t-1}' + \mathrm{diag}\{\lambda_i\} \circ D_{t-1}^2 \\
\varepsilon_t &= D_t^{-1} v_t \\
Q_t &= S \circ (\iota \iota' - A - B) + A \circ \varepsilon_{t-1} \varepsilon_{t-1}' + B \circ Q_{t-1} \\
R_t &= \mathrm{diag}\{Q_t\}^{-1/2} Q_t \mathrm{diag}\{Q_t\}^{-1/2}
\end{aligned}
$$

where $\circ$ is the Hadamard product, v_t is a zero-mean residual, D_t is the diagonal volatility

---

[23]An alternative model with similar flexibility is the Varying Conditional Correlation (VCC-) GARCH model of Tse and Tsui (2002).

matrix, R_t is the correlation matrix, and S is the unconditional covariance of the epsilons. The normality assumption ensures to a likelihood function which can be maximized directly. However, Engle (2002) provides a simpler two-step estimator that is consistent but inefficient. This proceeds by decomposing the log-likelihood into two parts, one that governs the volatility and another for correlation. Let θ denote the volatility parameters and ϕ denote the correlation parameters. Then we have

$$
\begin{aligned}
L(\theta, \phi) &= L_V(\theta) + L_C(\theta, \phi) \\
L_V(\theta) &= -\frac{1}{2}\sum_t (n\log(2\pi) + \log|D_t|^2 + v_t' D_t^{-2} v_t \\
L_C(\theta, \phi) &= -\frac{1}{2}\sum_t (\log|R_t| + \varepsilon_t' R_t^{-1}\varepsilon_t - \varepsilon_t'\varepsilon_t)
\end{aligned}
$$

It turns out that $L_V(\theta)$ is the log-likelihood of the sum of univariate GARCH likelihoods, which is optimized by maximizing each term separately. This can be done with standard software routines.[24] When we have a consistent estimator for these GARCH models (as we do here), denoted by $\hat{\theta}$, then we can substitute that into $L_C(\hat{\theta}, \phi)$ to obtain a consistent estimator for ϕ.

Our conditional mean model is a VAR in GDP and hours growth with two lags. Since GDP and hours are measured in logs, productivity is a linear combination of GDP and hours. Thus, the correlations with productivity are calculated as linear combinations of the variances and covariances of GDP and hours. See appendix A. The univariate errors from the first stage estimation are modeled as GARCH(1,1) processes.[25] Figure 2.6 shows the results for GDP and hours. Reassuringly, the figure corresponds closely with the rolling window figures from section 2.2.2, albeit much less smooth. The stark decline in volatility in the mid-1980s is readily apparent, as is the subsequent extended period of calm. There are isolated spikes in GDP volatility but it largely remains below the levels seen in the 1970s. Hours, as with the rolling windows, returns to the subdued levels before the 1970s. Figure 2.7 depicts the volatility of hours relative to output. The relative volatility of hours is higher, on average, in the post-84 period although with frequent sharp drops coinciding

---

[24]We use Kevin Shepherd's Oxford MFE toolbox for Matlab. This is a significant rewrite of his earlier UCSD_GARCH toolbox. http://www.kevinsheppard.com/wiki/MFE_Toolbox

[25]The qualitative results were unchanged for other GARCH(p,q) processes and excluding the exogenous regressors. The adequecy of GARCH(1,1) specifications, absent leverage effects, can be found in, e.g., Hansen and Lunde (2005).

roughly with recessions. Looking at the previous two figures we see that these declines are generated by transient output volatility spikes.
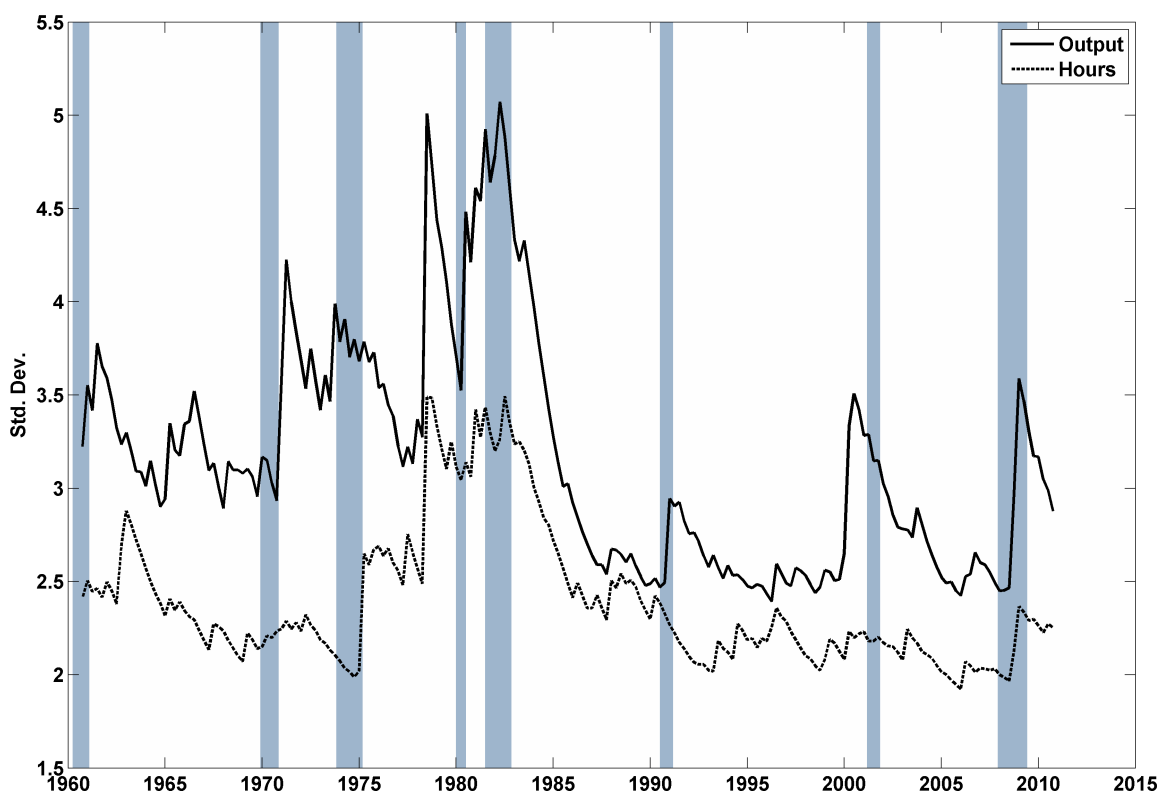
Figure 2.6: Annualized US quarterly growth volatility: GARCH(1,1)

The volatility results are all obtained using standard GARCH methods. We now turn to the conditional correlation estimates. Figure 2.8 depicts the correlation between labor productivity and both output and hours. First note the correlation between productivity and hours declines in the mid-1980s and then recovers during the 2000s. This mimics the rolling window results although the decline is not as stark. Interestingly, the GARCH estimates of the correlation between labor productivity and output *rise* over this time period. This is the first example of a discrepancy between the estimation methods, and corresponds with perhaps the most important moment we consider. Recall the decline in the procyclicality of labor productivity is given as *a priori* evidence against the RBC model. This suggests that empirical fact is sensitive to estimation method. Further evidence for the sensitivity of correlation estimates is found in the next section where we find similar results for the TVP-VAR model.
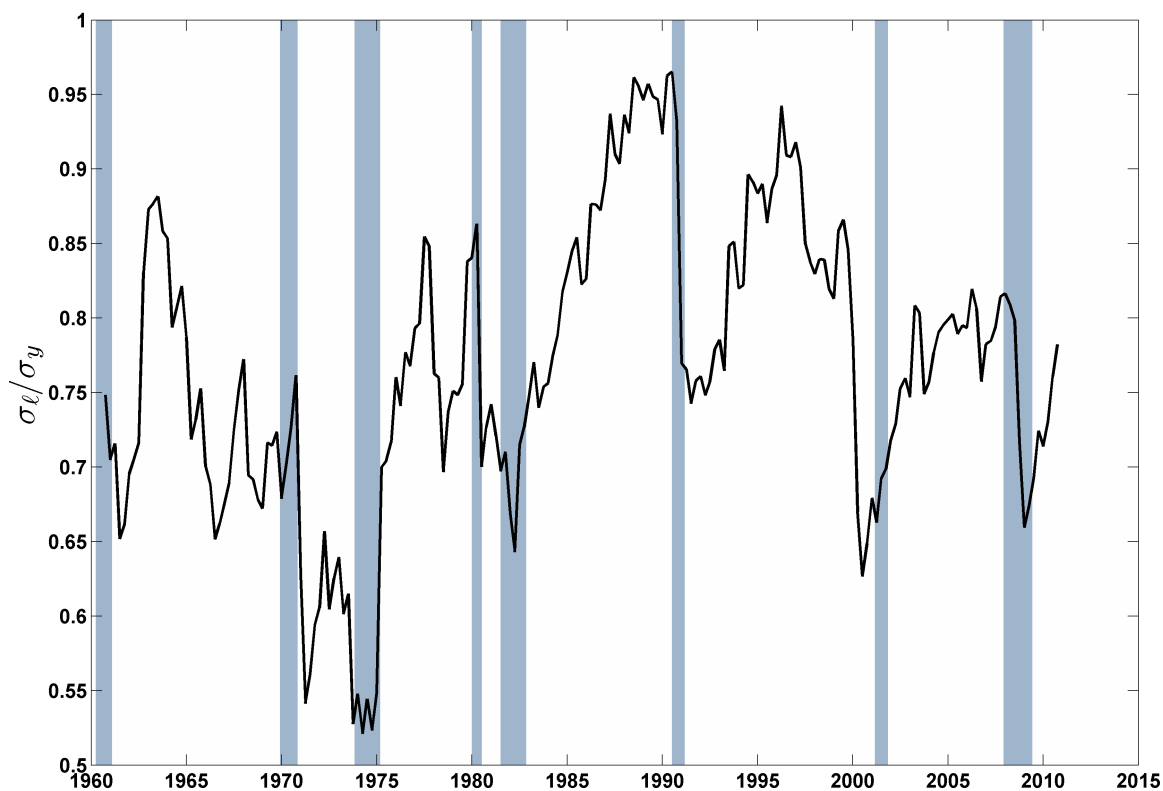
Figure 2.7: Ratio of US labor input volatility to output volatility: GARCH(1,1)

2.2.5 Time-Varying Parameter VAR

Our final model considers stochastic volatility, an alternative to GARCH for modeling time-varying volatility. It also adds a final innovation, time-varying conditional means via the time-varying coefficients. Both features are necessary in a TVP-VAR due to the model's flexibility. If one feature is missing, then the other will compensate in order to match the time variation found in the data, thereby biasing the estimates.[26] Given the model's complexity, we separate the technical discussion and empirical results into the following two subsections.

2.2.5.1 Model

We use the time-varying parameter VAR model found in Galí and Gambetti (2009), which incorporates elements of Primiceri (2005) and Cogley and Sargent (2005). The VAR

---

[26]See the discussion in Cogley and Sargent (2005) and Stock (2002).
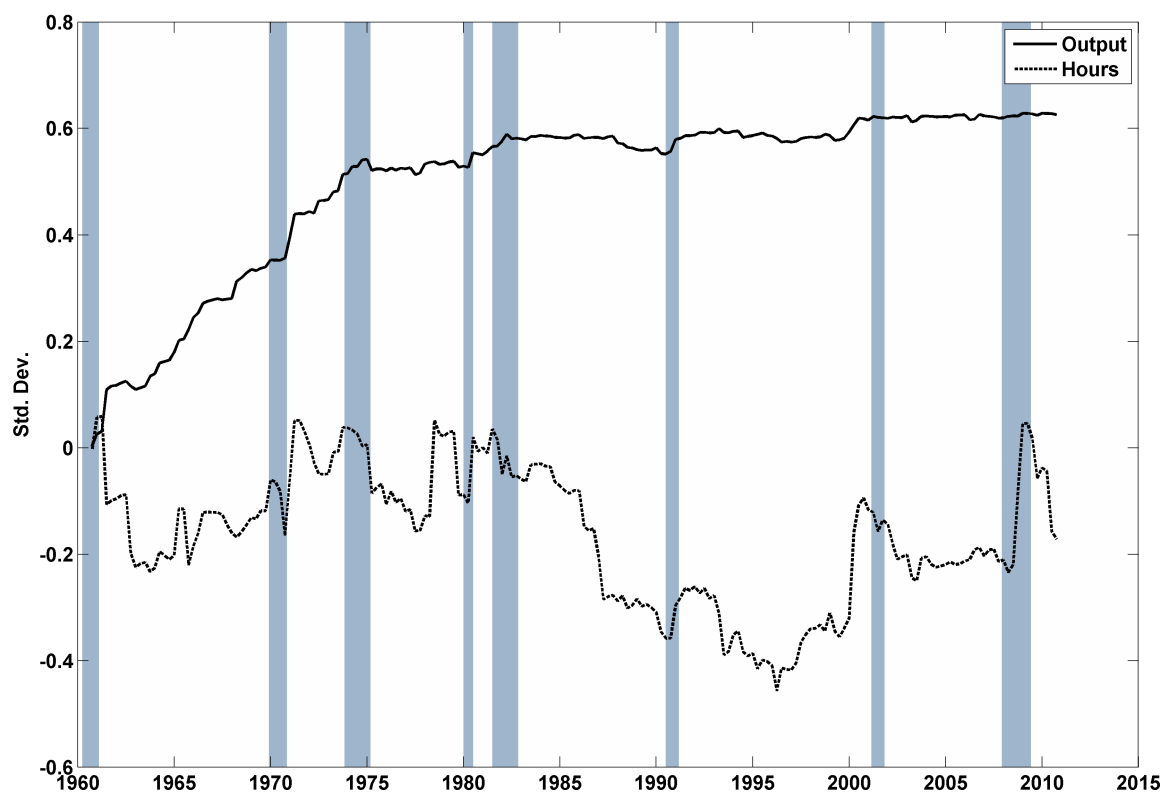
Figure 2.8: Correlation with US labor productivity: DCC-GARCH

model is given by

$$z_t = c_t + B_{1,t}z_{t-1} + \ldots + B_{k,t}z_{t-k} + u_t \qquad t = 1, \ldots, T \qquad (2.3)$$

where z_t is an $n \times 1$ vector of observed endogenous variabels; c_t is an $n \times 1$ vector of time varying intercepts; $B_{j,t}$, $j = 1, \ldots, k$, are $n \times n$ matrices of time varying coefficients; u_t are heteroscedastic unobservable shocks with variance covariance matrix Ω_t. We assume that the roots of the VAR polynomial lie outside the unit circle for all t.

Further, consider the triangular decomposition of Ω_t

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t'$$

where A_t is the lower triangular matrix

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{n1,t} & \cdots & \alpha_{n(n-1),t} & 1 \end{bmatrix}$$

and Σ_t is the diagonal matrix

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t} \end{bmatrix}$$

Thus we model the time-varying volatility and covariances separately.

Stacking in a vector θ_t all the R.H.S. coefficients, (2.3) can be rewritten as

$$\begin{aligned} z_t &= G_t' \theta_t + A_t^{-1} \Sigma_t \varepsilon_t & (2.4) \\ G_t' &= I_n \otimes [1, z_{t-1}', \ldots, z_{t-k}'] \end{aligned}$$

where the symbol $\otimes$ denotes the Kronecker product, and $V(\varepsilon_t) = I_n$.

Let α_t be the $[n \times (n-1)]/2$ vector of non-zero and non-one elements of the matrix A_t stacked by row and σ_t be the vector of the diagonal elements of the matrix Σ_t. The dynamics of the time varying parameters is then given by

$$\theta_t = \theta_{t-1} + v_t \tag{2.5}$$
$$\alpha_t = \alpha_{t-1} + \zeta_t$$
$$\log \sigma_t = \log \sigma_{t-1} + \eta_t$$

where all the innovations are assumed to be jointly normally distributed. The variance covariance matrix, V, is assumed to have the following block diagonal form

$$V = \mathrm{Var}\left(\begin{bmatrix} \varepsilon_t \\ v_t \\ \zeta_t \\ \eta_t \end{bmatrix}\right) = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}$$

where I_n is an n-dimensional identity matrix; Q, S, and W are positive definite matrices. Additionally, we will assume that W is diagonal and S is block diagonal with the blocks corresponding to the parameters from separate equations.

The model is cast as a set of state-space models following Primiceri (2005). In general, each state-space model is a simple transformation of the measurement and state equations given by equations 2.4 and 2.5, respectively. Each model is then estimated using a Bayesian state-space Gibbs sampling algorithm developed in Carter and Kohn (1994) . Details of the prior specifications and estimation algorithm are found in the appendix. Briefly, the Gibbs sampler allows us to sample from the joint distribution by iteratively sampling from a set of conditional distributions. The gain comes from converting an intractable high dimensional joint distribution into a series of much simpler conditional distributions. After a sufficient number of initial iterations (the burn-in period), the draws will be from the true joint distribution. Our estimates are the corresponding sample moments from a series of draws from this joint distribution.

2.2.5.2 US Results Discussion

The VAR is in the growth of productivity and hours, in logs, with two lags. This matches the specifciation found in Galí and Gambetti (2009). The volatility of and correlation with GDP is calculated using the linearity of productivity induced by the log specification. See appendix A. Figures 2.9 and 2.10 depict the US results for volatility and

correlation, respectively. These results compare favorably with Galí and Gambetti (2009), although we have a longer time series and consider the whole economy rather than the non-farm business sector.[27] The volatility of GDP and hours decline precipitously in the early 1980s while labor productivity declines steadily over the sample. Also note that the volatility of hours declines more and remains below the pre-1970s level, unlike our previous results. However we lose much of the 1960s in order to center the prior and thus do not have a long pre-peak sample to compare against.
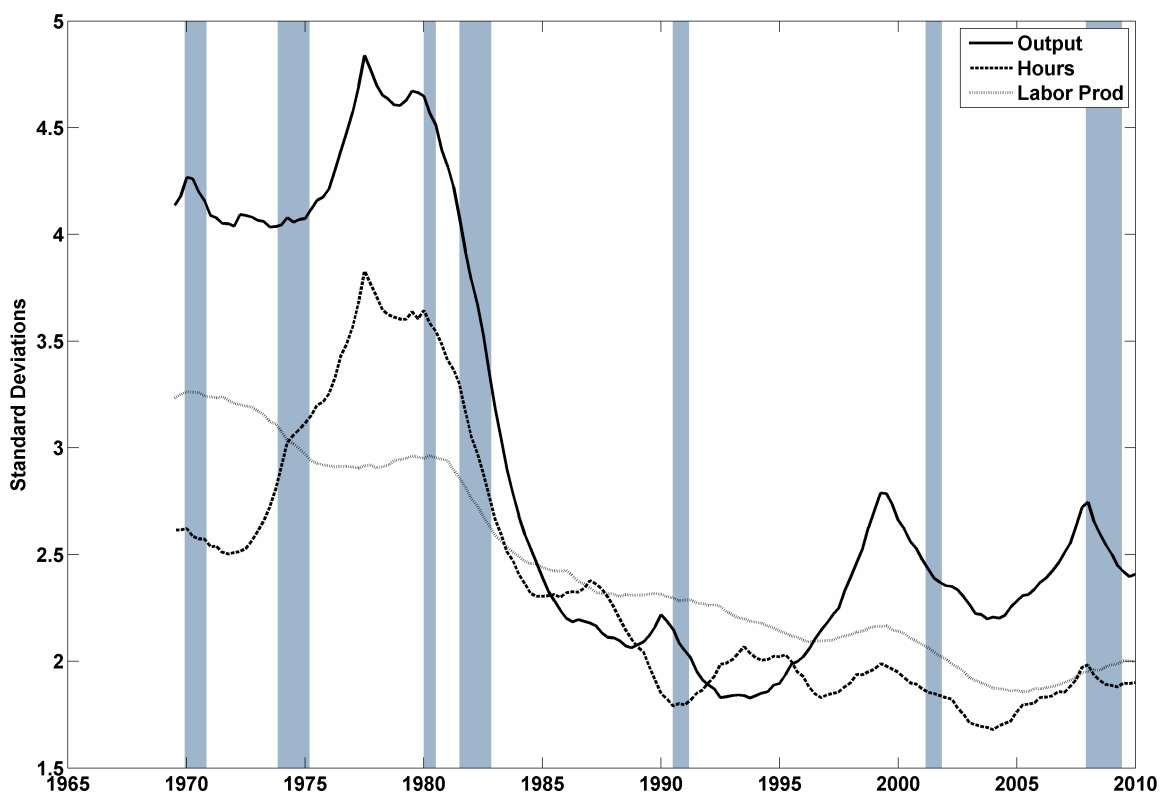
Figure 2.9: US conditional volatility: TVP-VAR

The decline in correlation between productivity and hours coincides with output in the early 1980s as we have found previously. Notice, however, the correlation between productivity and output declines until the late 1980s and then subsequently increases. It also does not decline as much as with the previous techniques. It clearly remains procyclical. This matches the results found in Galí and Gambetti (2009). However in a subsequent

---

[27]These correspond to figures 1a and 2 from Galí and Gambetti (2009). They use proprietary data from the Haver USECON database (see their footnote 10).

paper, Galí and van Rens (2010) ignore the TVP-VAR results and use the univariate filter results to justify their theoretical model instead.
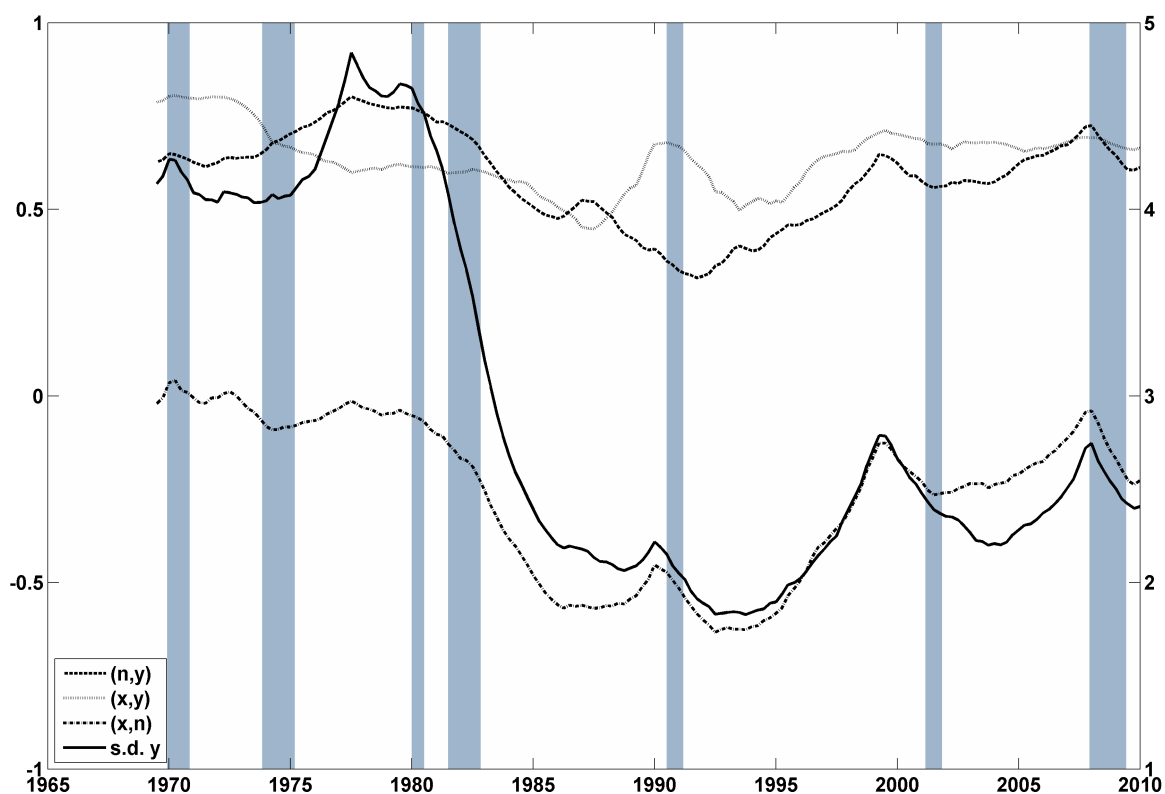
Figure 2.10: US unconditional correlation: TVP-VAR

2.2.6 Summary of US Results

This section sought to confirm the US stylized facts found in the literature using our data set and ascertain their sensitivity to statistical method. We considered both continuous volatility measures and multivariate conditional means. We found that all the stylized facts are robust to how we model the non-stationarity except for the correlation between labor productivity and output. The two continuous measures, DCC-GARCH and TVP-VAR, provided estimates that do not support labor productivity becoming acyclic. This contradicts the received wisdom that the cyclicality of productivity has declined and diminished the usefulness of RBC models. However, it is unclear whether the moments matched in calibration exercises are the objects of interest. Alternative statistical specifications may

provide a better fit to the procyclicality of labor productivity while leaving the other (already well fit) moments unchanged.[28] Section 2.4 compares international data to the US experience. Given these results, the correspondence with the US on the correlation of labor productivity and output must be treated with some skepticism.

2.3   Labor Market Frictions Model

The model of Galí and van Rens (2010) can generate all of the macro-moment stylized facts via a decline in labor market frictions. This section sketches the intuition behind their result.

They consider two extremes of fully flexible and completely rigid labor markets. They model this with two labor market frictions; endogenous effort choice and convex labor adjustment costs, i.e. hiring costs. Endogenous effort choice provides an intensive margin that is not subject to the adjustment cost. This provides a labor margin that is able to adjust to shocks in the completely rigid environment. Their two driving shocks are to technology and preferences. In addition, labor adjustment costs generate wage rigidity. This is because existing matches generate a surplus, equal to the adjustment cost, that is split between workers and the firm. Effort is assumed to have a higher marginal disutility and stronger diminishing returns than employment. With no adjustment costs, the intensive margin is never adjusted since it is dominated by employment. With infinite adjustment costs, only effort is adjusted.

The signs of the correlations depend upon the parameters governing the intertemporal elasticity of consumption, disutility of effort, diminishing returns to total labor, and diminishing returns to effort, as well as the relative size of technology and preference shocks. The primary difference between the regimes is the excess sensitivity to the underlying shocks along the intensive margin. We show in appendix C the inequalities necessary to generate the observed changes in the US. In particular, if technology shocks are sufficiently more volatile than preference shocks then we obtain all of the labor market moments.

The decline in output volatility is generated by a change in the flexibility of wages.

---

[28]See Morris (2011) for a discussion of this and related issues in the context of calibration. In addition, higher order approximations and other moment restrictions have recently been shown to have first order effects in DSGE models. He also provides an alternative estimation methodology that overcomes these issues.

More flexible wages are able to counteract technology shocks and reduce output volatility. More flexible wages follow from the smaller surplus, and smaller bargaining set, generated by smaller adjustment costs. While they are able to generate the correct *sign*, their calibrated wage rigidity is too small to generate the observed *magnitude* of decline in output volatility. However, in the US the change in labor market dynamics coincides with the change in output dynamics, as shown in section 2.2.

We do not estimate their DSGE model directly. Instead, we consider the moments it can theoretically generate and then we will compare them to the observed pattern of labor market frictions. Their model is highly stylized and a close correspondence between the structural labor market parameters and data is difficult to achieve. We consider the model under a best-case scenario; it can already generate all of the desired moments and perhaps can generate the US output moments with better data. Therefore, when we compare the international data to the US, we will include the well-documented international change in output volatility among our classifying variables. We stress this is a stronger statement than advocated by Galí and van Rens (2010) but we believe it illustrates how closely the international Great Moderation experience matches the US, and whether their mechanism is capable of delivering the disparate results.

2.4 International Results

In section 2.2 we showed that the stylized facts in the US are robust to statistical method with the exception of the correlation between labor productivity and output. Here we extend the analysis to thirteen additional countries. We observe significant heterogeneity across countries. As an organizational framework, we therefore classify the countries according to whether their moments match the US. By organizing the countries according to whether they match the US, in section 2.5 we will be able to examine whether labor market frictions declined in those countries.

To be precise, we classify the countries into three categories: 1) those that agree with the US experience, 2) those that are the exact opposite, and 3) those with mixed results. The first two categories have clear, but opposite, predictions of labor market frictions according to Galí and van Rens (2010); decreasing in the first case and increasing in the latter. For countries with mixed results, there are several possibilities. First, the underlying

structural parameters can be significantly different than those found in the US. This suggests consumption preferences, labor preferences, or the returns to labor vary significantly across countries. Second, the shocks themselves may vary across countries. Specifically, the economies have the same response to the shocks however technology shocks are much less volatile than preference shocks. Lastly, the shocks and structural parameters may be the same however a complex time series of labor frictions may generate complex movements in the moments. Our classification thus extracts the cleanest predictions regarding labor market frictions with a minimum of assumptions. This is important since measures of labor market frictions are controversial and relatively coarse.

We emphasize that this is not definitive evidence for or against Galí and van Rens (2010) but rather an indication of the relative importance of their mechanism if we assume countries are similar to the US. Their strong conclusions are supported by implicitly extrapolating the US experience to the rest of the world. Otherwise, their argument rests on the idiosyncratic experience in the US, i.e. data-mining.

2.4.1   Output Volatility Break Dates

Recall that two methods, univariate filters and den Haan, require a break date. We estimate the volatility break following Sensier and van Dijk (2004). Let $W(\tau)$ denote the Heteroskedasticity and Autocorrelation Consistent (HAC) Wald test of the null hypothesis $H_0 : \delta_1 = \delta_2$ in the regression

$$|y_t - \hat{\mu}| = \delta_1\{1 - I(t > \tau)\} + \delta_2 I(t > \tau) + \varepsilon_t, \qquad t = 1, ..., T$$

where $\hat{\mu}$ is the sample mean, T is the number of time periods, τ is the specified break date, and $I(\cdot)$ is the indicator function. If we treat τ as unknown then we can test for the presence of a break using a variety of statistics: sup-Wald [Andrews (1993)], AveW or ExpW [Andrews and Ploberger (1994)]. Point estimates of the break date are given by the τ that minimizes the sum of squared errors in the regression (equivalently, the τ in the SupW statistic).

$$SupW = \sup_{\tau_1 < \tau < \tau_2} W(\tau)$$

The middle 70% of the sample is used to estimate the change point. This means that $\tau_1 = [0.15T]$ and $\tau_2 = [0.85T]$, where $[\cdot]$ denotes the integer part. Approximate asymptotic

p-values are obtained using the method of Hansen (1997).

Table 2.3 gives the estimated break dates for hours and GDP as well as the Stock and Watson (2005) results for comparison. Relative to Stock and Watson (2005), we obtain very similar results for output. The hours break date differs from that of output by more than a decade for 5 of the 13 countries, and does not always lead GDP. We also calculate a second break date conditional on the first. This is done primarily for Japan in order to center all the calculations in the Great Moderation narrative time period. However note the second break in GDP is much closer to that of hours for France. The earlier GDP break date is due to the idiosyncratic May 1968 strike that brought France to a standstill and will be readily apparent in the sections to follow. Interestingly Finland's break date is in the early 1980s rather than the early 1990s when its trade collapsed with the Soviet Union and it mismanaged financial market deregulation following a domestic credit crisis [Nickell (1997)]. This confluence of events over the period 1990-1993 saw the unemployment rate more than triple from 3.4% to 17.7%. Unlike France, this period appears as a modest increase in output volatility although we will see an increase in hours volatility in the mid 1980s that presages the crisis to come.

2.4.2   Output Volatility

We begin our classification by splitting the countries according to whether output volatility declined. International output volatility has been well documented. For the G7, Stock and Watson (2005) estimate an instantaneous volatility measure. Cecchetti et al. (2006) consider 21 countries using the HP filter and split-sample. Given the problems documented with filters in section 2.2.2 and to economize on space, we omit those results in the following sections. In general, they confirm the continuous results or are insignificant, and are available in the online appendix. Here we instead extend the evidence to continuous measures and the den Haan VAR forecasts.

Figures 2.11 and 2.13 depict output volatility measured using a six-year rolling window, GARCH, and TVP-VAR for decreasing and increasing output volatility countries, respectively. The GARCH results are plotted versus the right axis due to isolated spikes in

Table 2.3: Break in Volatility for Hours and GDP

	Estimated Break Date				Stock and Watson (2005)
	Hours	GDP (1st)	GDP (2nd)		GDP
Australia	1980q1	1983q3	2005q3		-
Austria	1992q2	1988q2	1972q4		-
Canada	1983q2	1987q1	1966q2		1991q2
Finland	1983q3	1981q1	1992q3		-
France	1979q2	1969q1	1979q1		1968q1
Germany	1970q1	1993q1	1984q2		1993q1
Ireland	1997q3	1996q2	1986q1		-
Italy	1968q2	1980q1	2006q2		1980q1
Japan	1976q1	2003q2	1990q3		$p > 5\%$
Norway	1982q4	1977q4	1998q2		-
Sweden	1990q3	1992q2	2007q4		-
UK	1991q1	1980q4	2006q1		1980q1
US	1984q3	1983q4	1999q4		1983q2

Notes: GDP (1st) and GDP (2nd) correspond to the primary break date and the secondary break date conditional on the first, respectively. Stock and Watson (2005) only estimate break dates for G7 countries.

Figure 2.11: Decreasing Output Volatility: International Results

Note: The rolling window and TVP-VAR results are on the left axis, and the GARCH results are on the right axis. The rolling window is based upon a six-year window. The standard deviation is in annualized percent.

a few of the countries. Figures 2.12 and 2.14 depict the den Haan output volatility results for decreasing and increasing output volatility countries, respectively.

Figure 2.11 shows those countries in which output volatility declines. First note that the US results are remarkably clean. All three methods show a sudden decrease in volatility in the early- to mid-1980s. However that pattern is not repeated elsewhere. In general, the rolling window and TVP-VAR results are similar. The GARCH results, on the other hand, are either erratic around the trend (Australia, Finland, UK) or trendless with rapidly mean-reverting spikes in volatility (Austria, France, Italy). The change in volatility is generally not sharp but tends to trend down. Figure 2.12 shows the den Haan results. They also show
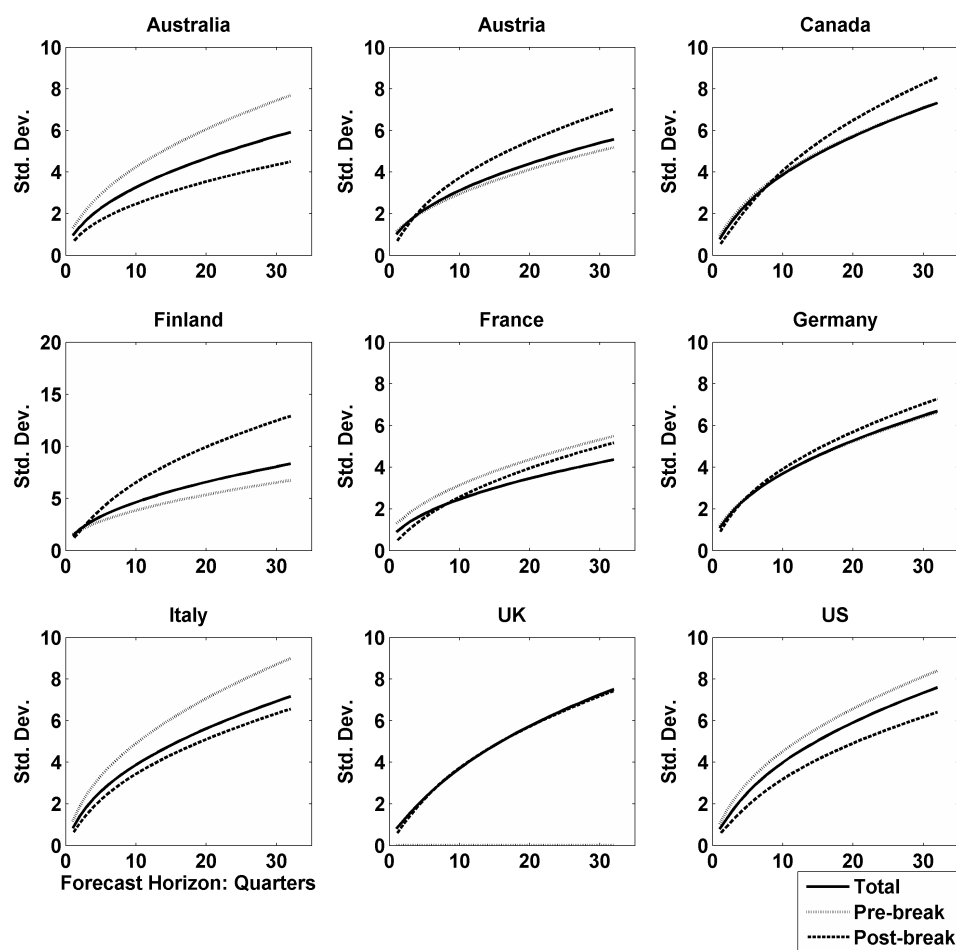
Figure 2.12: Decreasing Output Volatility: den Haan Results

Note: The standard deviation is in annualized percent.

output volatility declining for half the countries (Australia, France, Italy, and US) but show output volatility increasing for the other half (Austria, Canada, Finland, and Germany), and no change for the last country (UK). Since there generally is not a sharp break in the continuous volatility measures, the mixed results for the den Haan measure may be from imposing such a break *a priori*.

Figure 2.13 displays the three countries whose GDP volatility increased; Ireland, Norway, and Sweden. Ireland and Norway exhibit sharp increases in volatility that are near mirror-images to the US, although separated temporally. Ireland increases in the late 1990s well outside the traditional Great Moderation period. Norway, on the other hand, increases in the late 1970s. Sweden's volatility increases more gradually. Note that Sweden's volatility nearly triples over the 1990s decade for all three measures. However the figure is distorted by the tripling again during the recent financial crisis. This brings into sharp relief the impact of the financial crisis. For the countries that already saw an increase in volatility, only Sweden exhibited a further increase during the recent crisis. Figure 2.14 shows the results for the den Haan measure. Norway and Sweden also exhibit an increase in volatility, and Ireland is virtually unchanged.

The three methods had conflicting results for the final country, Japan, as seen in figure 2.15. According to the rolling window output volatility initially declines and remains stable in the early 1980s, rises and stabilizes at pre-moderation levels for most of the 1990s, and then subsequently declines again. The GARCH results are inconclusive and the TVP-VAR has a roughly downward trend. Figure 2.16 shows the den Haan results. This suggests Japan's output volatility decreased after the break but figure 2.15 suggests this depends critically on when we choose that break.

For the six countries not in the G7, we find that the instantaneous volatility follows a clear trend. This confirms the robustness in Cecchetti et al.'s (2006) international output volatility results by showing that the standard practice of splitting the sample does not obscure any complex dynamics.[29] For the G7, our results are similar to those found in Stock and Watson (2005). They estimate an autoregression with stochastic volatility using a non-Gaussian smoother. This is similar to our TVP-VAR setup. There are three noteworthy

---

[29]They examine the relationship between detrended inflation volatility and output volatility. This suggests that their methodology adequately captures the output volatility dynamics, but similar issues may pertain to the inflation dynamics.
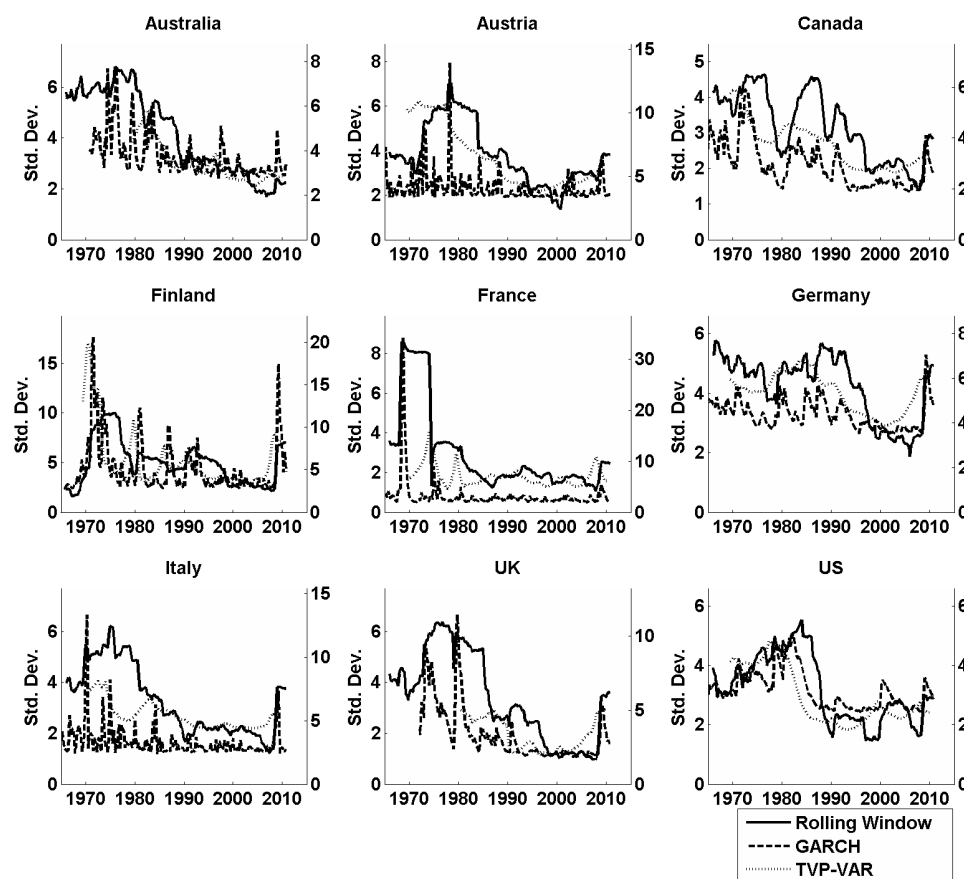
Figure 2.13: Increasing Output Volatility: International Results

Note: The rolling window and TVP-VAR results are on the left axis, and the GARCH results are on the right axis. The rolling window is based upon a six-year window. The standard deviation is in annualized percent.
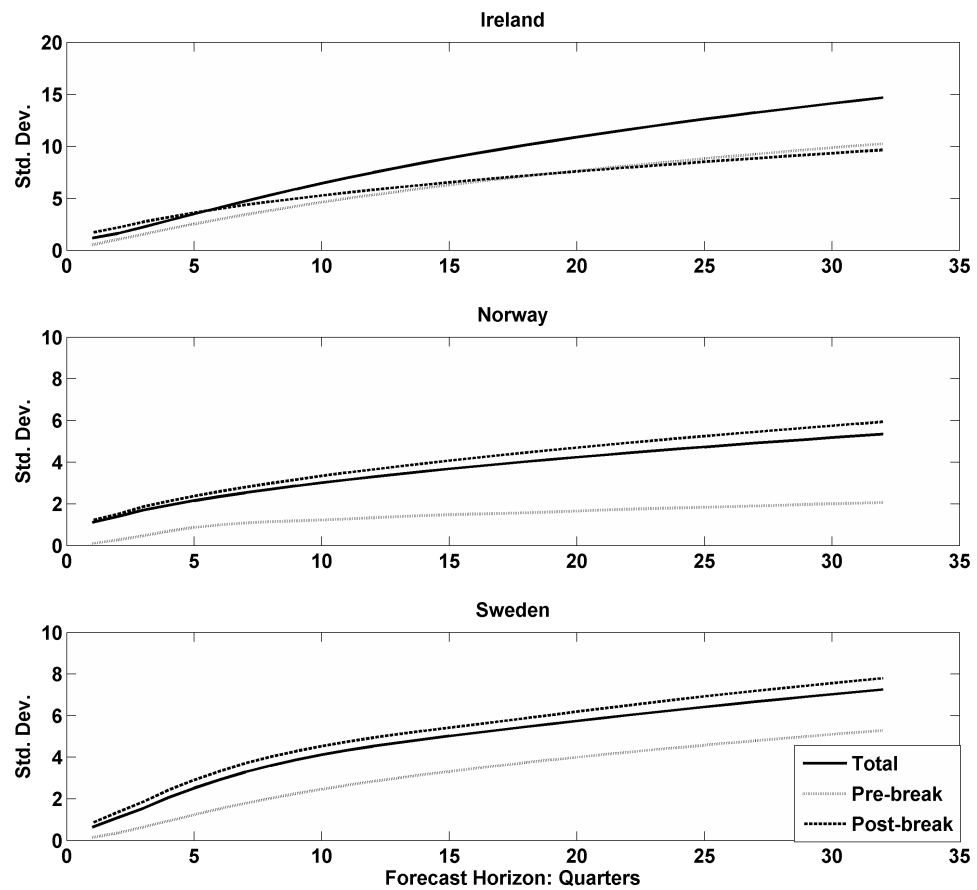
Figure 2.14: Increasing Output Volatility: den Haan

Note: The standard deviation is in annualized percent.
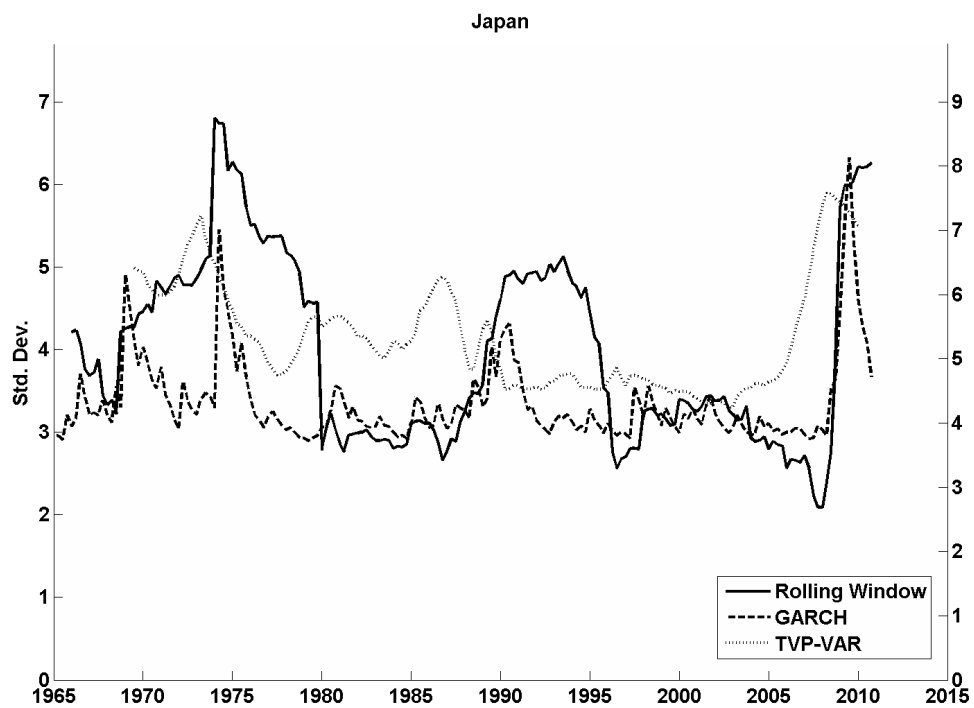
Figure 2.15: Neither Increasing nor Decreasing Output Volatility: International Results
Note: The rolling window and TVP-VAR results are on the left axis, and the GARCH results are on the right axis. The rolling window is based upon a six-year window. The standard deviation is in annualized percent.
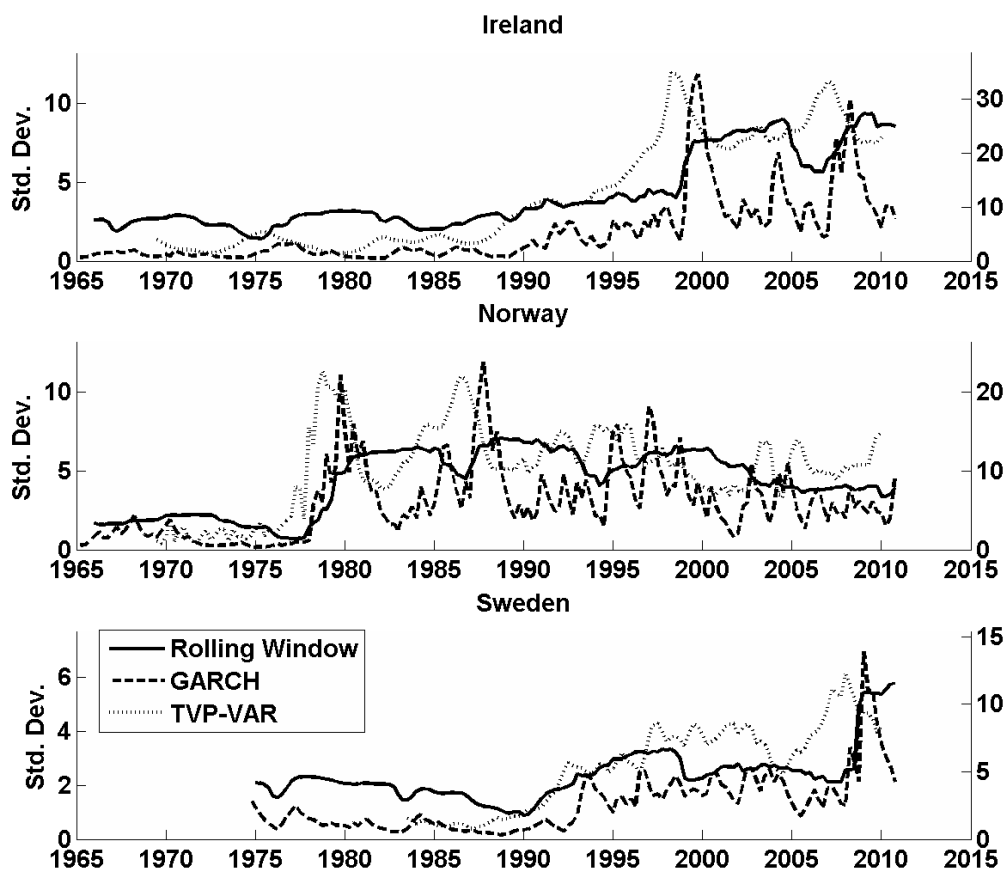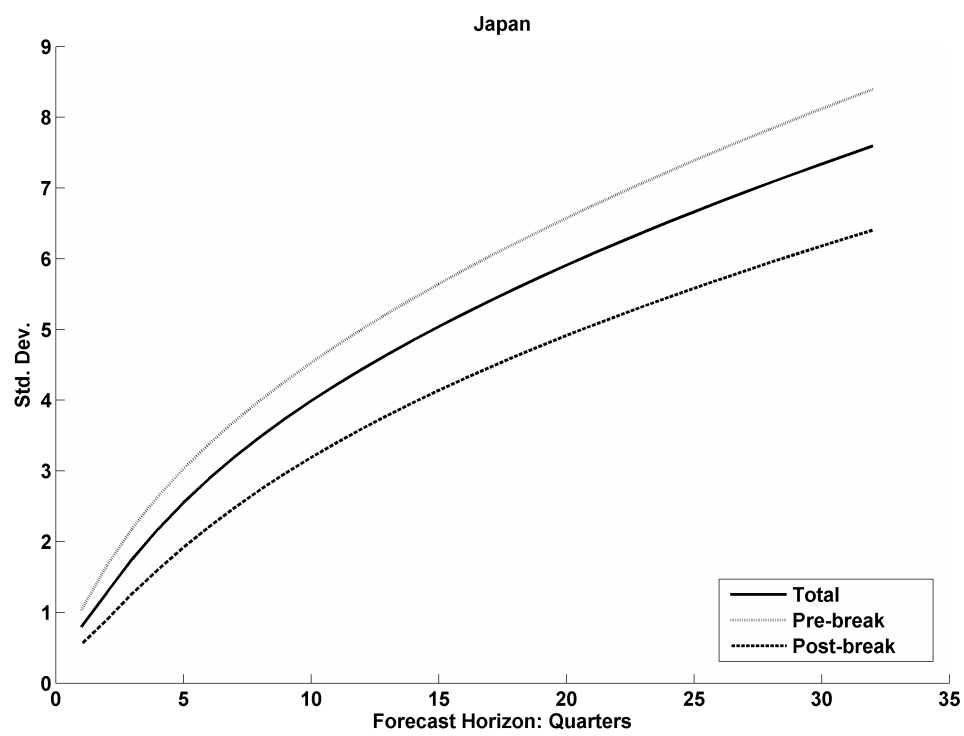
Figure 2.16: Neither Increasing nor Decreasing Output Volatility: den Haan

Note: The standard deviation is in annualized percent.

differences. First, they find the volatility in France to be relatively constant. France's results are dominated by a series of national strikes in the late 1960s and early 1980s. In particular the May 1968 strike which is the largest strike on record. Rather than remove the outlying observation, which represents real labor frictions present in the economy, we consider how extreme results may influence the methods under consideration. Second, they find German volatility declines over the entire sample. The series they use has been discontinued but the difference is likely how East and West German data are combined prior to 1993. Lastly, they find Japan's volatility declines until the mid-1980s and then rises monotonically over the later period.[30]

2.4.3 Labor Market Stylized Facts

2.4.3.1 Hours Volatility and Relative Volatility

Galí and Gambetti (2009) present empirical evidence for the US that hours volatility decreases but it decreases less than output volatility. Thus the volatility of hours relative to the volatility of output increases.

In order to match the US, the countries where output volatility declined must also have the relative volatility of hours increase. Figure 2.17 shows the results for the countries for which we had found a decline in output volatility. We see that, consistent with Galí and Gambetti's findings for the US, labor volatility increased relative to output volatility for all but two of these eight countries, the two exceptions being Finland and France. In Finland the ratio increases only for the rolling window. The other two methods do not show a clear trend. For France the ratio initially increases but begins decreasing around the final major strikes in the early 1980s, i.e. when output volatility begins settling down. The patterns in these two countries therefore are not consistent with the US.

Figure 2.18 shows the den Haan results for the countries for which we had found a decline in output volatility. The relative volatility increases for half the countries (Australia, Austria, France, UK, and US), and decreases for half the countries (Canada, Finland, Germany, and Italy). As with output volatility, we have mixed results, however it is not the same partition. One criticism of the continuous measures is that whether the volatility is increasing or decreasing is subjective. However, even with the clean interpretation from den

---

[30]These are also similar to G7 results found in Blanchard and Simon (2001).
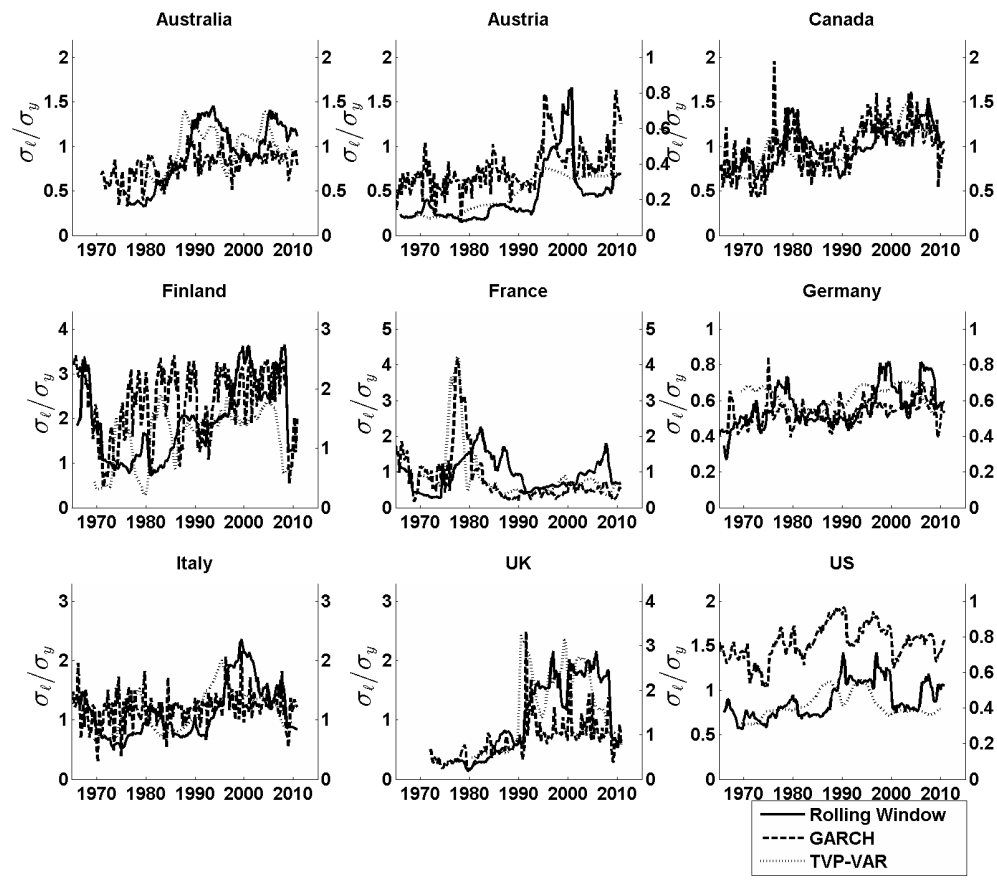
Figure 2.17: Volatility of Hours Relative to Output: Predicted Increase
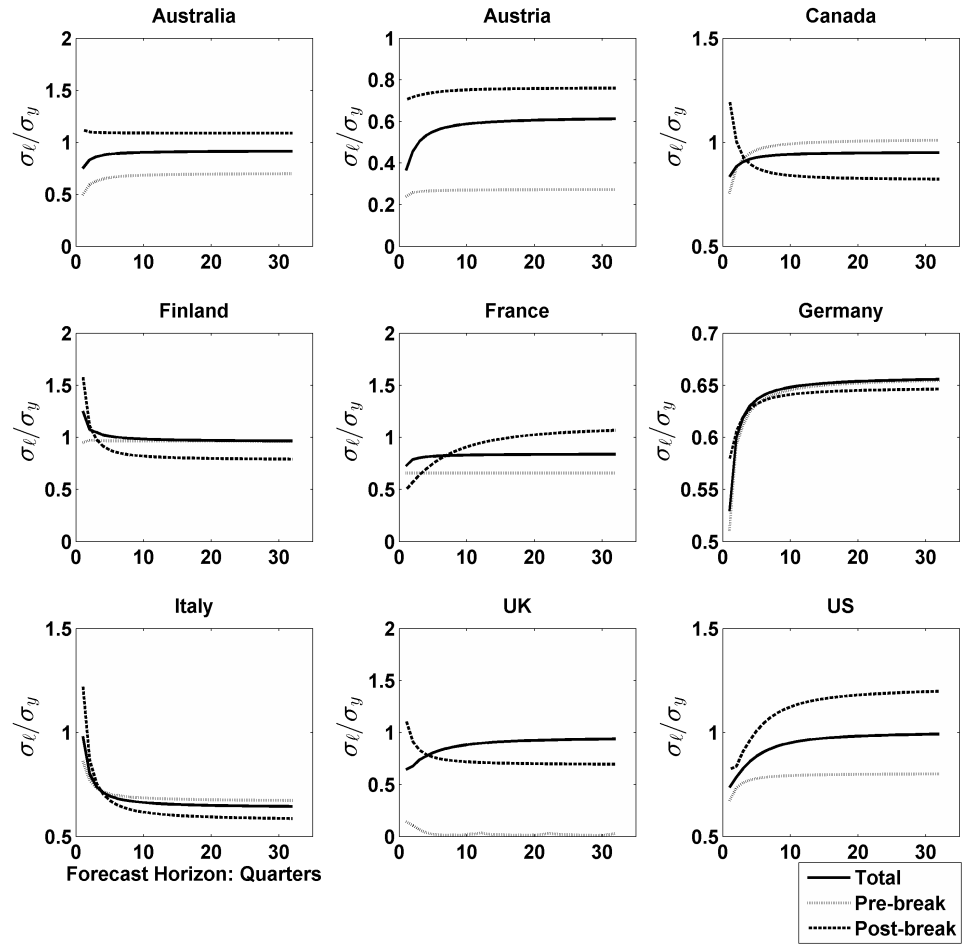
Figure 2.18: Volatility of Hours Relative to Output: Predicted Increase
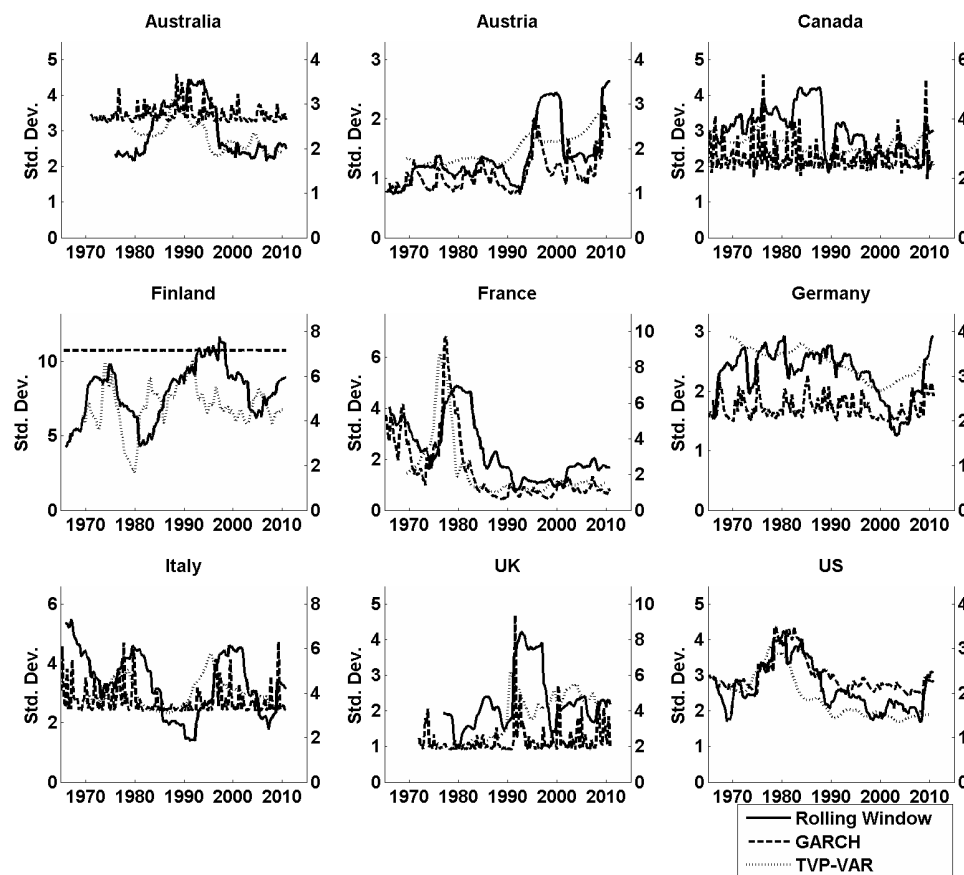
Figure 2.19: Hours Volatility: Predicted Decrease

Haan, we still do not get a consistent partitioning of the countries. And den Haan remains reliant on the choice of break date.

For the remaining countries, the increase in the ratio can also be due to an increase in the hours volatility. Figure 2.19 examines this possibility. We see that hours volatility increases for Austria, Finland, Italy, and the UK. Thus the only countries that remain consistent with the US are Australia, Canada, and Germany. Figure 2.20 shows the den Haan measure. As seen in section 2.2.3, the US displays a rotation in hours volatility; decreasing at high frequencies (< 6 quarters) but increasing at business cycle frequencies (6-32 quarters). That being said, hours volatility decreases for Australia, Canada, and Italy and increases for the remainder.

We now turn to the countries where output volatility increased. Here, if increased output volatility resulted from more labor market frictions, we would expect the relative
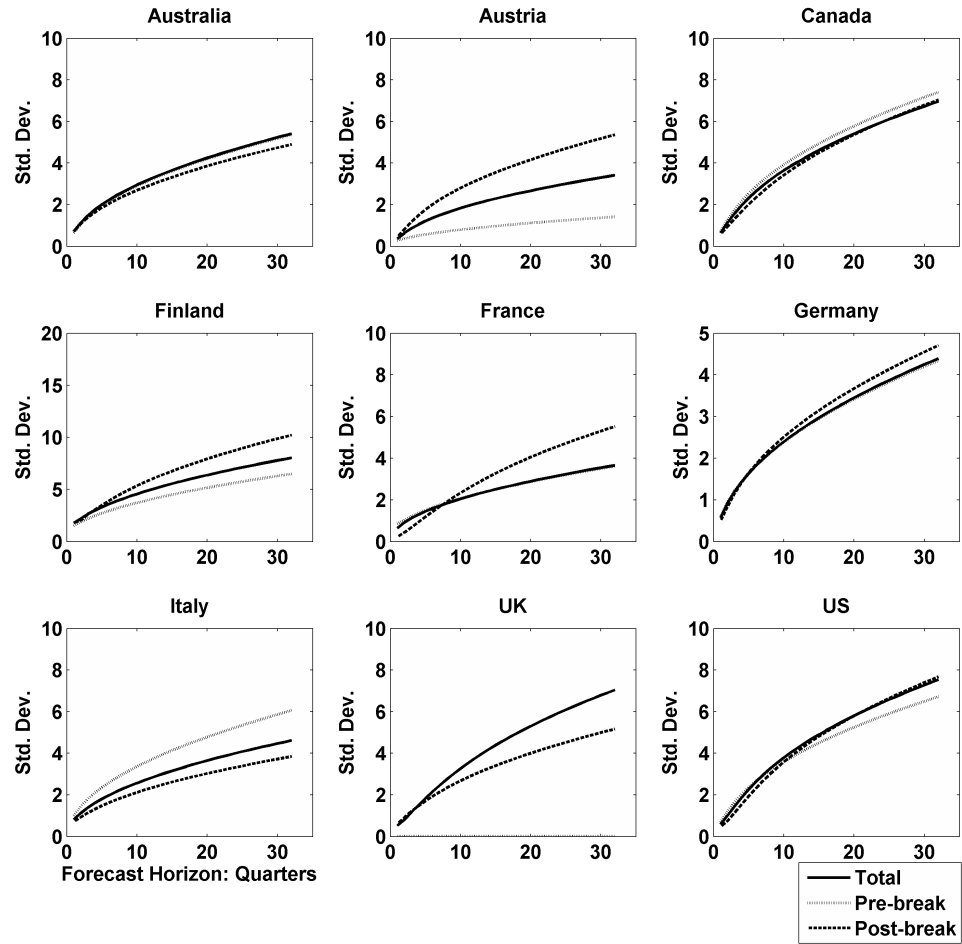
Figure 2.20: Hours Volatility: Predicted Decrease
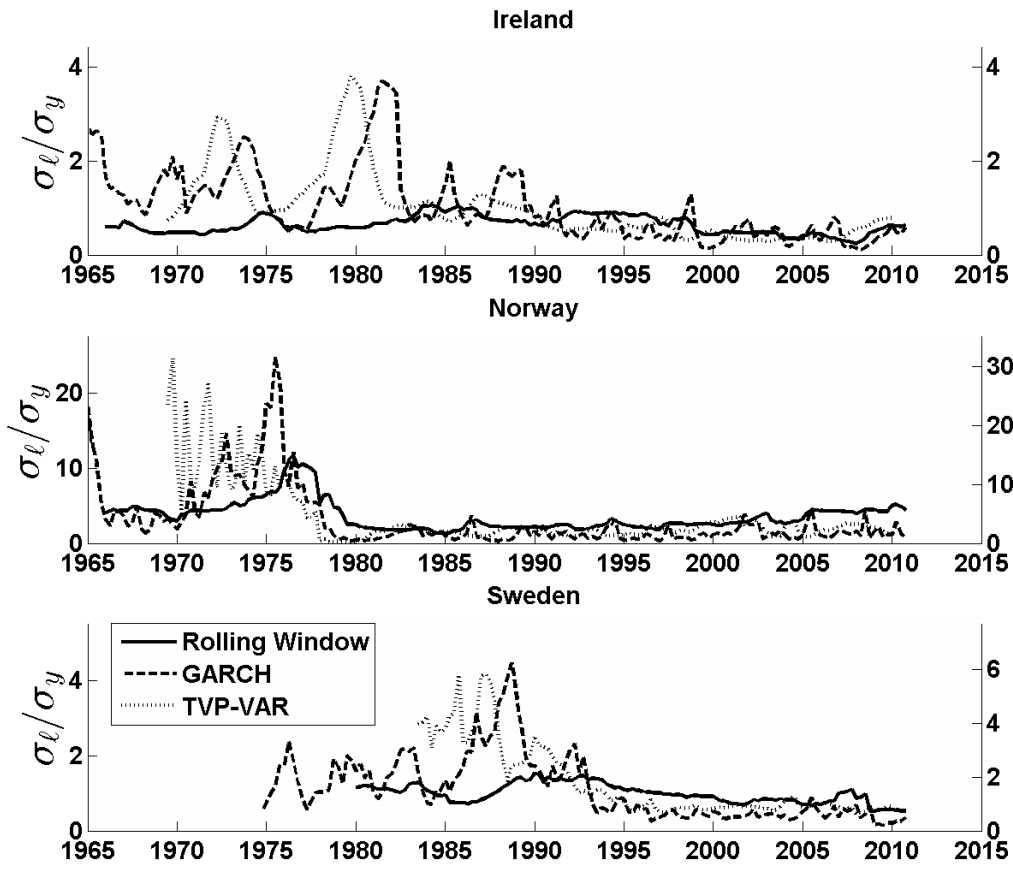
Figure 2.21: Volatility of Hours Relative to Output: Predicted Decrease
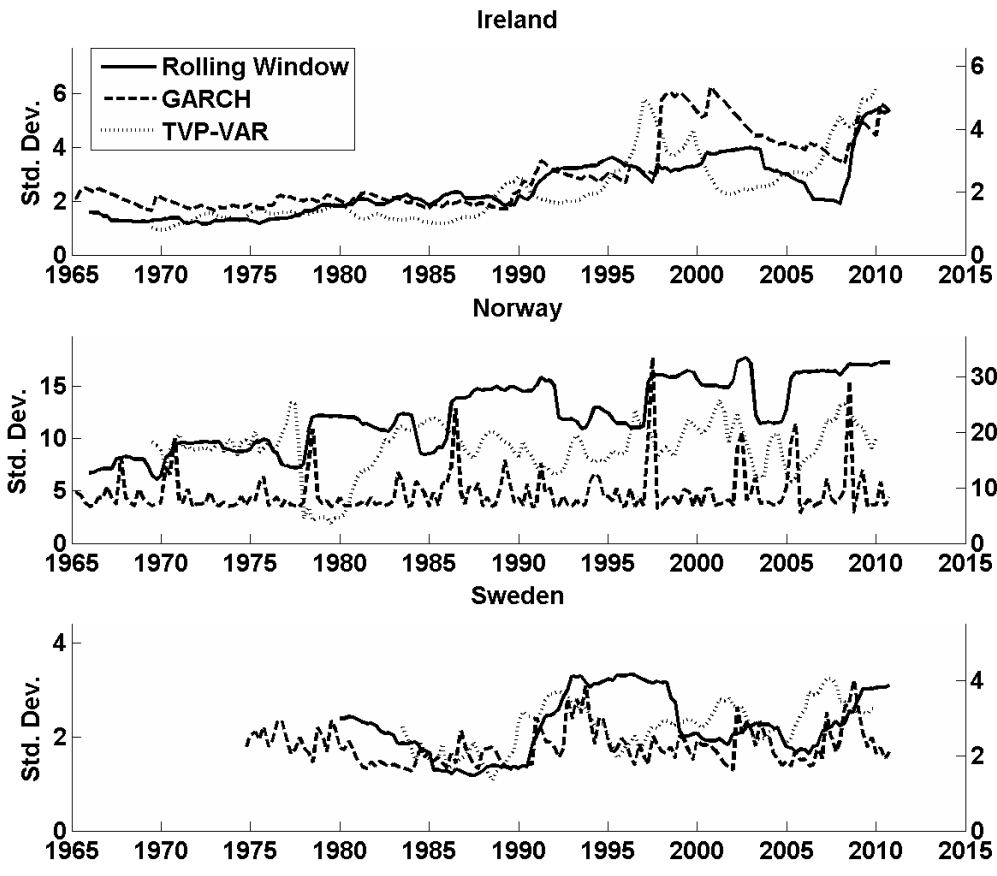
Figure 2.22: Hours Volatility: Predicted Increase
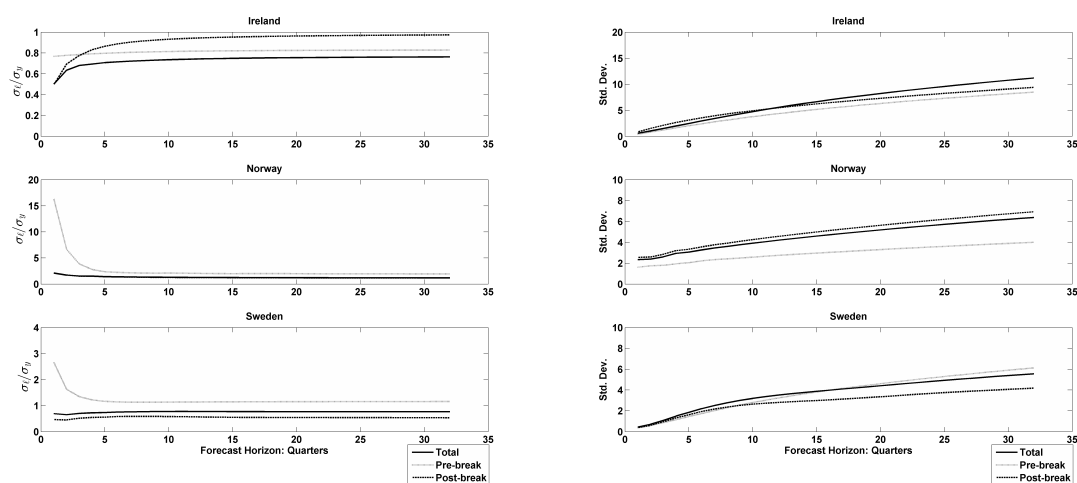
(a) Volatility of Hours Relative to Output: Predicted Decrease

(b) Hours Volatility: Predicted Increase

Figure 2.23: Den Haan Hours Volatility and Relative Volatility

volatility to decrease and hours volatility to increase. Figure 2.21 shows that the relative volatility decreases for all these countries. Figure 2.22 confirms that this is not due to hours volatility decreasing. Figure 2.23 shows the den Haan results. Norway is consistent with the continuous measures, but Ireland and Sweden have mixed results. Relative volatility increases for Ireland and hours volatility declines for Sweden.

2.4.3.2   Cyclicality of Labor Productivity

In the US, the correlation between labor productivity and both output and hours declines. We begin with the correlation between productivity and hours. Figure 2.24 plots the results for the countries that can still be consistent with the US. Figure 2.24.A plots the three instantaneous volatility measure and Figure 2.24.B plots the den Haan correlations. The US has consistent and clear results across all four methods. The instantaneous correlations in Australia and Canada decline although it is marginal. The den Haan correlations, however, are clear. Australia is the only country other than the US that exhibits a decline.

Figure 2.25 plots the productivity-hours correlation for the countries that are opposite the US. Here we expect the correlation to increase. There is substantial variation in the instantaneous correlation results however the correlation marginally increases. The in-

(a) Instantaneous Correlations   (b) Den Haan Correlations
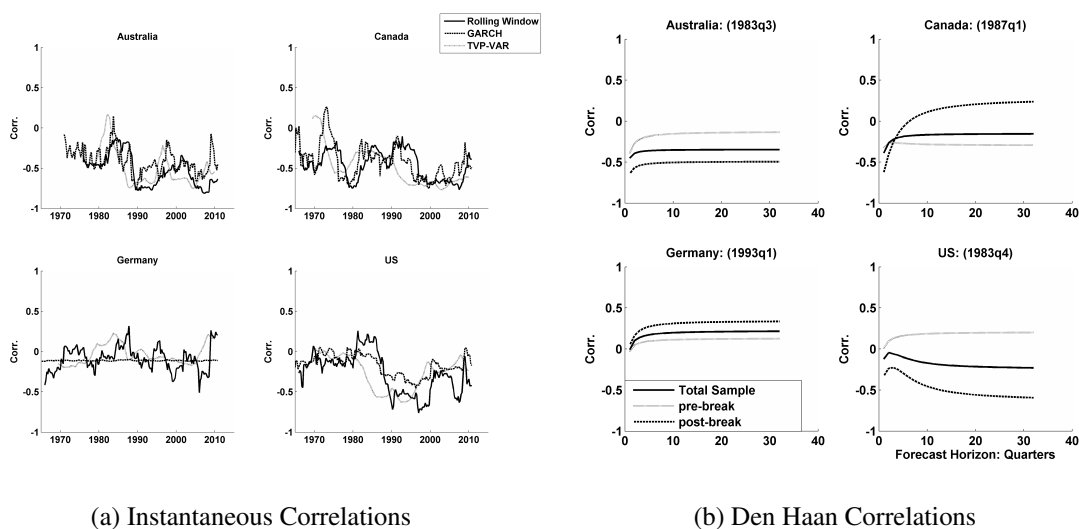
Figure 2.24: Correlation between productivity and hours: Predicted Decline

Note: The GDP break dates calculated in section 2.4.1 are in parantheses.

(a) Instantaneous Correlations   (b) Den Haan Correlations
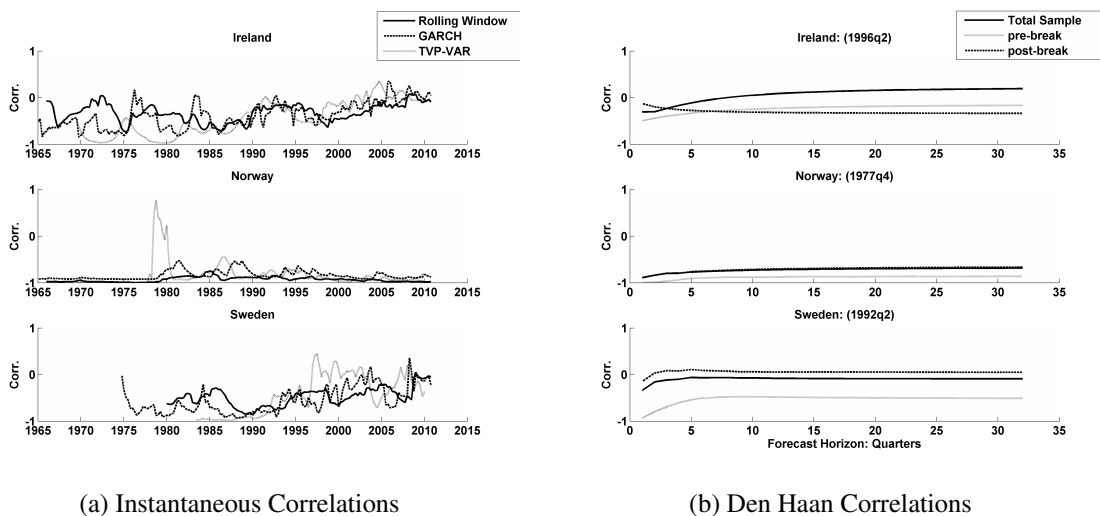
Figure 2.25: Correlation between productivity and hours: Predicted Increase

Note: The GDP break dates calculated in section 2.4.1 are in parantheses.

<div align="center">(a) Instantaneous Correlations   (b) Den Haan Correlations</div>
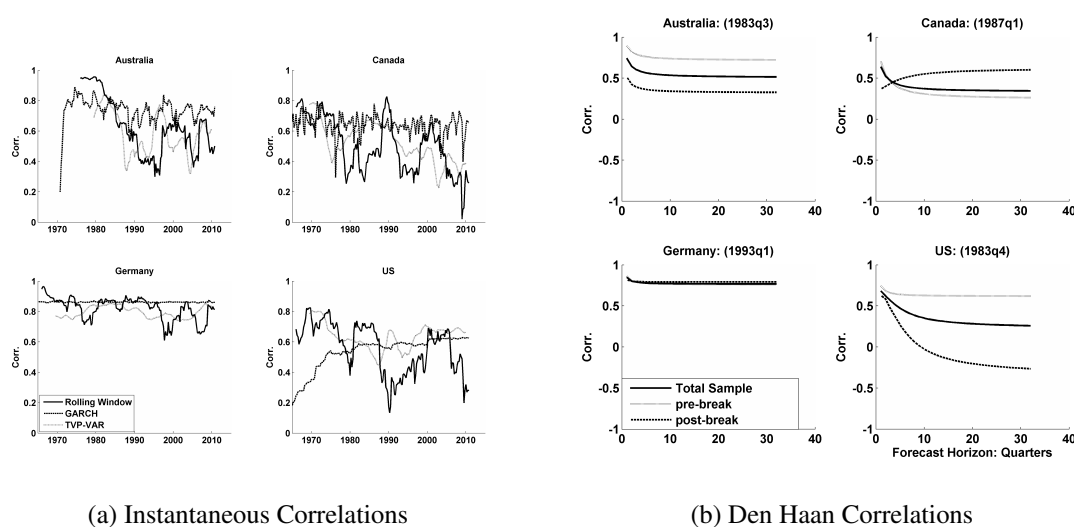
Figure 2.26: Correlation between productivity and output: Predicted Decline

Note: The GDP break dates calculated in section 2.4.1 are in parantheses.

crease is particularly pronounced for the TVP-VAR estimation in Norway in the late 1970s, however the increase is transient and is relatively mild for the rest of the time period. The other two methods also increase from essentially perfectly negative correlation. The den Haan correlations are more dramatic. Norway and Sweden clearly increase. Ireland is essentially unchanged at business cycle frequencies but exhibits a substantial increase at higher frequencies. Given the marginal instantaneous correlation results, we exclude Ireland based on the den Haan correlations.

Figure 2.26 depicts the correlation between productivity and output for the remaining countries that can still be consistent with the US. Recall that for the US the decline in the procyclicality of labor productivity manifests itself only in the split sample methods (filtering and den Haan) and the univariate conditional mean models (filtering and rolling window) and not the VAR and GARCH methods. Hence these results are already suspect. We see that only Canada declines under the instantaneous measures. However the den Haan correlations match what we found with hours.

Figure 2.27 depicts the results for the countries predicted to increase. Looking at the instantaneous correlations, apart from Sweden, no clear pattern emerges. For Sweden there is a rapid decline in the early 1990s that reverses over the rest of the decade and subsequently plateaus at a higher level. However in the den Haan correlations, Norway and

(a) Instantaneous correlations   (b) Den Haan Correlations
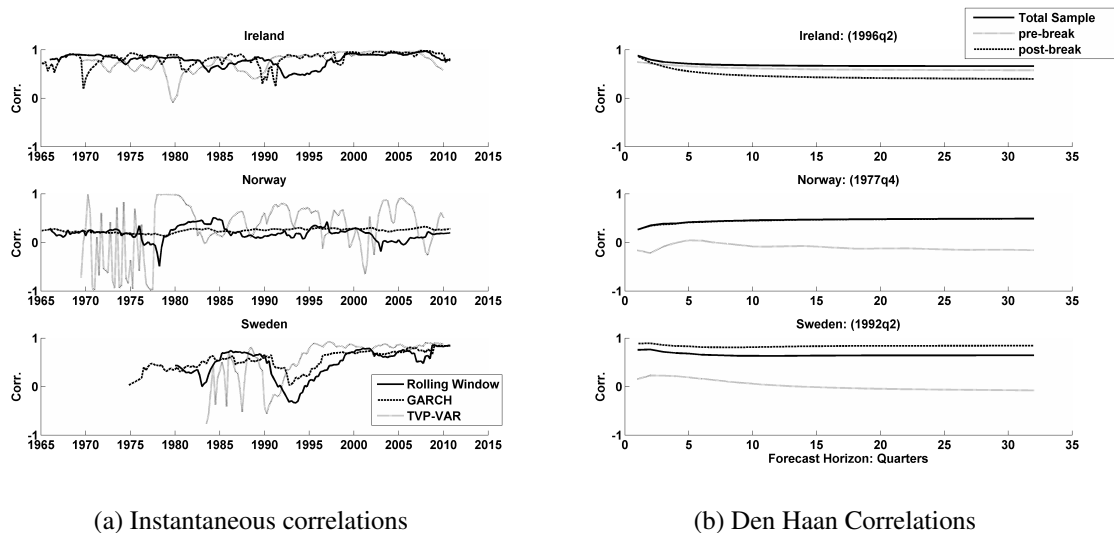
Figure 2.27: Correlation between productivity and output: Predicted Increase

Note: The GDP break dates calculated in section 2.4.1 are in parantheses.

Sweden increase substantially while Ireland declines modestly. The den Haan correlation thus match the expected signs whereas the instantaneous correlations provide much weaker evidence.

2.4.4   Conclusions

The results vary considerably across countries. In particular, only two countries, Australia and Canada, exhibit the same set of patterns as the US. However, neither country displays moments as clean as the US. Furthermore, we have two countries that appear to have the exact opposite experience as the US; Norway and Sweden. The largest disagreement is the increase in the hours volatility underlying the increase in the relative volatility. This heterogeneity, in turn, suggests labor market frictions can still play a role but there is a more complicated relationship with GDP. In the next section we explore changes in labor market friction over this time period.

2.5   The Role of Labor Market Institutions

We explore the relationship between labor market institutions (LMI) and macroeconomic volatility and correlation dynamics in a panel setting. Our empirical setup follows Rumler and Scharler (2011). They consider how LMI's affect the volatility of the output gap and inflation using a panel of 20 OECD countries. They focus on three labor market frictions: wage bargaining centralization, union density, and employment protection legislation obtained from Nickell (2001). The output gap is represented as the difference from the HP filtered trend. Volatility of the output gap is calculated over non-overlapping 5-year periods using data over 1970-1995. This results in 6 data points for each country.

We extend this in several ways. First, we update and expand the number of LMI measures to 10. We use updated data from Nickell (2006) as well as data from the OECD and Amsterdam Institute for Advanced labour Studies (AIAS)[31] over the period 1970-2003. Second, we control for non-stationarity using a time-varying parameter (TVP-) VAR. Unlike the HP filter, this explicity takes into account the multivariate structure of output and the labor market. Third, the TVP-VAR provides an annual conditional volatility measure. This expands the time series dimension from 6 to 33 points.

An alternative approach is taken by Gnocchi and Pappa (2011). They consider 13 LMI's drawn from Nickell (2006), OECD, and AIAS. Due to the lack of time variation in most of the series, they ignore the time dimension and collapse the LMI dimension using Principal Component Analysis (PCA) on the time-averages. They then consider the cross-sectional relationship between univariate filtered (HP, BK, 4D) second moments (volatility and correlation) of GDP, employment, and productivity, and LMI factors. This is inapplicable given our interest in the dynamics of the second moments. Moreover, the cross-country heterogeneity in hours dynamics that we identified in section 2.4 implies time-averaging is misleading. It is not enough to know that Italy had on average higher labor market frictions and higher hours volatility than the US. It is crucial to know if changes in LMI's coincide with Italy's U-shaped pattern. If not, then the higher volatility is due to an unexplained factor in the residual correlated with the LMI's.

---

[31]Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts (ICTWSS): http://www.uva-aias.net/208. This expands the coverage, depth, and time period of the OECD *Employment Outlook* (1994, 1997) special chapters on collective bargaining.

2.5.1 Labor Market Institution Data

The heterogeneity observed in section 2.4 was discussed in relation to the Galí and van Rens (2010) model. Specifically, in their model a decline in labor market frictions explains the US labor market and productivity dynamics. The labor frictions are modeled as wage bargaining power and hiring and firing costs. Wage bargaining is proxied by various union measures encompassing both their extent and their power and unemployment benefits. Hiring and firing costs are proxied by employment protection legislation.

Our full specification search and details of the LMI data are documented in Appendix D. Our LMI data comes from two sources: AIAS and Nickell (2006). The labor market frictions are: i) *BRR*, OECD wage Benefit Replacement Rate (%) averaged over first 5 years of unemployment for 3 family situations and 2 money levels (Nickell); ii) *NRW*, Net Replacement Wage (%) due to Allard (2005b), incorporates tax treatment, duration, and conditions necessary to collect (Nickell) iii) *UD*, union density, percentage of workforce that is unionized (AIAS); iv) *UC*, union coverage, percentage of contracts that are negotiated by unions (Nickell).

Greater clarity on the relationship between union density and union coverage will aid in later interpretation. Union density is the proportion of all wage earners that are members of a union. Union coverage is the proportion of all wage earners *that are eligible to join a union* that are members of a union. For this reason union coverage is actually *adjusted* union coverage.[32] There are three ways union coverage can differ from union density. First, employees can be prohibited from wage bargaining. All else equal, this will reduce the denominator in *UC* and lead to *UC* being greater than *UD*. Second, union-negotiated wage contracts can be imposed on non-union employers. All else equal, this will raise the numerator in *UC* and lead to *UC* being greater than *UD*. Lastly, union members may work for non-union firms. All else equal, this will raise the numerator of *UD* and lead to *UC* being smaller than *UD*. This last effect is only significant in Japan, therefore we expect *UC* to exceed *UD*. However, small negative differences between *UC* and *UD* do sometimes exist. This is due to different data being used to calculate the two measures and the assumptions necessary to calculate *UC*. For more information, see the OECD Employment Outlook (1997) and (2004). The difference between *UC* and *UD* approximates the *de facto*

---

[32]See Ochel (2001).

net effect of legislation and 3rd party agreements that extend union contracts to non-union members and restrictions on union organizing.[33]

Our specification is dominated by traditional wage bargaining factors. However, unions may have *de facto* influence over hiring and firing costs, even though the *de jure* government employment protection legislation is not found to be significant. A similar problem accounts for the absence of international minimum wage statistics in LMI analyses. Many European countries do not have minimum wage laws, however unions impose a *de facto,* albeit opaque, minimum wage.

2.5.2   Empirical Investigation

Following Rumler and Scharler (2011), we regress the volatility or correlation on our LMI's, control variables (X_{it}), firm fixed effects (μ_i), and time fixed effects (λ_t).

$$
\begin{aligned}
\sigma(y_{it}) &= \alpha_1 + \beta_1' LMI_{it} + \mu_{1i} + \lambda_{1t} + \varepsilon_{1,it} \\
\sigma(n_{it}) &= \alpha_2 + \beta_2' LMI_{it} + \mu_{2i} + \lambda_{2t} + \varepsilon_{2,it} \\
\rho(x_{it}, y_{it}) &= \alpha_3 + \beta_3' LMI_{it} + \mu_{3i} + \lambda_{3t} + \varepsilon_{3,it} \\
\rho(x_{it}, n_{it}) &= \alpha_4 + \beta_4' LMI_{it} + \mu_{4i} + \lambda_{4t} + \varepsilon_{4,it}
\end{aligned}
$$

The results of Galí and van Rens (2010) predict that $\beta_1, \beta_2 > 0$, i.e. more friction results in more volatile output and hours. Similarly, we expect $\beta_3, \beta_4 > 0$, i.e. more friction exacerbates the labor hoarding incentive and results in more procyclical productivity.

We calculate each regression on two samples, all of the countries and restricted to the countries consistent with the Galí and van Rens (2010) identified in section 2.4. The latter are Australia, Canada, Germany, Norway, Sweden, and US. We lose Ireland due to a lack of data on union coverage. The restricted sample provides cleaner dynamics with maximum cross-country separation for our limited LMI data to fit. Whether the gain in power is overcome by the loss in sample size is an empirical question. However, these countries are not selected at random so care should be taken in interpreting the restricted results.

---

[33] A coarser direct measure of the *de jure* extension of union wage contracts is *ext* from AIAS that we exclude.

Table 2.4: Macroeconomic Dynamics and Labor Market Frictions

Panel regression of Time-Varying Parameter-VAR volatilities and correlations on labor market institutions for 12 countries from 1969-2000. *Full* denotes the 12 countries with data. Ireland is excluded due to lack of *UC* data. *Restricted* denotes the 6 countries consistent with Galí and van Rens (2010) identified in section 2.4. They are Australia, Canada, Germany, Norway, Sweden, and US.

	$\sigma(Y)$		$\sigma(L)$		$\rho(X,Y)$		$\rho(X,L)$	
	Full	Restricted	Full	Restricted	Full	Restricted	Full	Restricted
BRR	0.0203	**0.0223**	**0.0357**	0.0255	0.0072	**0.0174**	-0.0052	0.0062
	(0.0284)	**(0.0078)**	**(0.0096)**	(0.0116)	(0.0076)	**(0.0034)**	(0.0063)	(0.0062)
NRW	-0.0450	0.0216	0.0304	0.0529	**-0.0392**	**-0.0462**	-0.0122	**-0.0535**
	(0.0281)	(0.0188)	(0.0268)	(0.0289)	**(0.0112)**	**(0.0103)**	(0.0140)	**(0.0141)**
UD	-0.0343	**0.0479**	0.0295	0.0435	0.0004	0.0229	-0.0032	0.0198
	(0.0525)	**(0.0114)**	(0.0201)	(0.0396)	(0.0105)	(0.0115)	(0.0089)	(0.0138)
UC	0.0377	0.0599	-0.0478	0.0005	**0.0288**	0.0224	**0.0350**	0.0301
	(0.0362)	(0.0242)	(0.0244)	(0.0421)	**(0.0099)**	(0.0113)	**(0.0101)**	(0.0141)
F p-value	0.411	0.0000	0.008	0.0001	0.012	0.001	0.043	0.023

NOTES: White standard errors clustered over countries are reported in parantheses. Statistically significant coefficients at the 5% level are displayed in bold face. Regressions include both country and time fixed-effects. The *F*-statistic p-value is for the null that all the labor friction coefficients are equal to zero.

The results are shown in Table 2.4 with standard errors in parentheses. The dependent variable is given in the column heading with both the full and restricted results. All variables are measured in percent, therefore coefficient interpretations are given as percent added rather than percent growth. Since Gali and van Rens link reduced labor market frictions to moderation and since we did witness both growth moderation and labor market deregulation, we would expect to find the predicted relationship. Similarly, the heterogeneous labor volatility dynamics illustrated in section 2.4 would require more complicated LMI dynamics to find a relationship.

However, we find precisely the opposite. The first column shows that none of the LMI's are individually or jointly statistically significant in the full sample for GDP volatility. The third column shows that gross replacement rates are statistically significant in explaining hours volatility. Furthermore, the relationship with hours volatility is the cor-

rect sign and economically significant. The observed 15% average increase in benefits is associated with an increase of about 0.5% in annual hours volatility.

It is possible the Galí and van Rens (2010) (GvR) LMI mechanism is does not hold for or is not strong in all of the countries. In their calibration to US data, GvR find that their model is able to match the sign but not the magnitude of the decline in GDP volatility. Therefore, we expect the relationship between LMI's and GDP volatility to be weak. Since the model was designed to match the decline in employment volatility, that dimension provides a fairer test of the model. Unfortunately, in the full sample, reduced labor market frictions are associated with greater employment volatility.

Restricting the sample to countries consistent with GvR, once again we arrive at counterintuitive results. Looking at the second column, we find that unadjusted unemployment benefits and union density are statistically significant and have the anticipated sign for GDP volatility. However, nothing is individually significant for hours volatility which GvR was designed to explain. Labor frictions jointly explain labor volatility as seen in the significant F-statistic. Thus, if we take the sign of the coefficients as given, broadly speaking an increase in labor frictions is associated with greater labor volatility.

Turning to the correlation between labor productivity and GDP we find that union coverage, but not union density, is statistically significant and the correct sign.

In addition, net replacement rates are statistically significant for the correlation with GDP, but it has the wrong sign. When we restrict ourselves to the GvR sample, union coverage loses it's significance however now both gross and net benefits are significant. In addition, they enter with opposite signs. This says that more generous unemployment benefits exacerbates the labor hoarding incentive by driving up the reservation wage, however increasing access to or lowering distortions in unemployment benefits lowers this incentive.[34] Although the net benefits coefficient is considered to have the wrong sign, there is an alternative general equilibrium interpretation. Easier access to unemployment benefits, holding the level constant, reduces the drag on the economy from the decrease in aggregate demand when unemployment rises. Consequently, observed equilibrium output grows more for a technology driven unit decrease in employment. Alternatively, unemployment benefits ended prior to the adjustment speed to technology driven unemployment shocks and greater duration provided more efficient aggregate demand support. In effect,

---

[34]The p-value on the test of the equality of the two coefficients is 0.0014.

the OECD gross benefit variable measured the *de jure* outcome whereas the net benefit variable comes closer to the *de facto* outcome by incorporating duration, tax distortions, and barriers to acquiring unemployment benefits.[35]

Finally, we find that union coverage is statistically significant and has the correct sign for the correlation between productivity and hours when using the whole sample. When we restrict the estimation to the GvR sample union coverage is no longer significant but net benefits become significant with the "wrong" sign. Again this appears to be capturing a friction in the provision of unemployment benefits.

Our panel results suggest LMI's may have a greater quantitative impact on the great moderation in GDP than the GvR model suggests. However, LMI's have surprisingly little to say about the volatility in hours. One possibility is that the GvR model studied employment rather than hours. This suggests that households are able to optimize on their intensive margin to such an extent as to overcome the calibrated dynamics found in GvR. Union coverage, or the extent that employers are bound by union negotiated contracts, is associated with higher correlation between labor productivity and both GDP and hours. This *de facto* union influence is more important, in this sample, than the proportion of employees that are union members. Lastly, we uncover a negative relationship between Allard's net replacement wages and the correlation of labor productivity and both output and, to a lesser extent, hours. We suggested a possible general equilibrium interpretation that warrants further study of the barriers to receiving unemployment benefits, the adjustment speed to technology driven unemployment, and the role unemployment benefits play in setting the reservation wage.

2.6 Conclusion

Three possible explanations for the Great Moderation have been proposed: good luck, good policy, or structural change. Here we consider the possibe role of labor market frictions. The stylized facts in the labor market are statements of the form of non-stationarity of labor productivity correlations with output and hours and the volatility of hours. We show that for the US these stylized facts are robust to various methods of modeling non-stationarity with the exception of the correlation between labor productivity and

---

[35]Estimation excluding *NRW* also finds *BRR* to be insignificant.

output. We find that this correlation remains positive except for univariate filtering and den Haan correlations. This is consistent with the existing literature using TVP-VARs however the disagreement of this moment between the models has not been stressed. In fact, theoretical models have relied solely upon filter evidence to the best of our knowledge. However, we note that the den Haan correlations have a more theoretically appealing connection to DSGE models and confirm the widely held notion that labor productivity has become less procyclical with respect to output.

Using a new internationally consistent data set on total hours, we then extend the stylized facts to thirteen additional countries. Since existing theories rely disproportionately on US data, we use the stylized facts in the US as a base case. We find significant international heterogeneity. Only two countries, Australia and Canada, have moments similar to the US. Galí and van Rens (2010) present a model that can explain all of the stylized facts using only a reduction in labor market frictions. An additional two countries, Norway and Sweden, are consistent with Galí and van Rens (2010) although they predict an increase in labor market frictions. Using data from the Amsterdam Institute for Advanced labor Studies (AIAS) on unionization and government intervention, we find that labor market frictions do not explain even these five countries. This suggests this mechanism for labor market frictions is mainly a US phenomena.

With a few notable exceptions, the literature has focused almost exclusively on the US in examining the Great Moderation. The newly available hours data from Ohanian and Raffo (2011) and heterogeneity in the labor market moments argues for international comparisons of medium scale DSGEs.

Chapter 2 represents coauthored work with Benjamin Kay. The dissertation author and Thomas Daula are co-first-authors

3 The dynamics of revenue and expenses: US states 1977-2006

3.1 Introduction

The recent financial crisis and recession has strained all levels of government financing. The record US federal deficits and ongoing European sovereign debt crisis have brought critical attention to the fiscal multiplier, in particular because of the zero lower bound, and sustainability of the welfare state. However, US states have not received the same attention. States face record drops in revenue[1] and, unlike the Federal government, nearly all states have enacted Balanced Budget Requirements (BBRs).[2] This prevents deficit spending during recessions when government revenues are expected to be low. Thus, state fiscal policy is more pro-cyclical which, in turn, amplifies business cycles. However, not all states have fared equally poorly in this recession or, for that matter, in past recessions.

The focus of this paper is the role a progressive revenue structure plays in state level budget dynamics. Progressivity is measured using the income elasticity of revenue. We classify states into two categories, *progressive* and *regressive*, depending on whether income elasticity is greater or less than one, respectively. Progressive revenues vary more than 1-for-1 with state income. This means that if we hold income dynamics constant across states, more progressive revenues are more volatile. Furthermore, if BBRs are enforced, this volatility in revenues is transmitted to expenditures. To study budget dynamics, we

---

[1]Dadayan and Boyd (2009)

[2]According to different interpretations of the statutes, either Vermont or North Dakota has no BBRs. Vermont according to the Advisory Commission on Intergovernmental Relations (1987) and North Dakota according to Hou and Smith (2006). See Mahdavi and Westerlund (2011) for a discussion.

employ impulse-response functions (IRF) from a standard VAR model that includes state income, revenue, and expenditures. The cross-sectional relationship is then investigated in a second-stage regression of expenditure elasticity on revenue elasticity. To account for estimation error, weighted least-squares is used with the IRF standard errors as weights. We find that more elastic revenues are associated with more elastic expenditures.

The two papers closest to ours are Bruce, Fox, Tuttle (2006) and Fatas and Mihov (2006) which investigate revenues and expenditures, respectively. Bruce, et al. (2006) consider the short- and long-term elasticities of personal income and sales taxes across US states in a cointegration framework. Unlike the general tax elasticity literature, they examine the tax bases on a state-by-state basis. This allows them to explain the cross-sectional variation in elasticities using policy variables. Fatas and Mihov (2006) decompose government expenditure into responsiveness and discretion. Responsiveness is measured as the income elasticity. The innovation is measuring discretion as the residual from a forecasting equation. They then consider the effect of BBRs on responsiveness and discretion. These two fiscal policies have offsetting effects on output volatility. Responsiveness helps reduce output volatility whereas discretion increases output volatility. Empirically, they find that the latter effect dominates: BBRs reduce output volatility via the larger reduction in discretion.

These two papers do not consider revenue and expenditure jointly. A related paper that examines the budget's relationship with the business cycle is Sorensen et al. (2001). They examine the cyclicality of state surpluses with respect to state GDP using a panel covering 48 states over the period 1978-1994. They find that surpluses are procyclical, driven by strongly procyclical revenues and only weakly procyclical expenditures. Their short time series precludes formal cointegration tests and they do not pursue how revenue and expenditure cyclicality are related cross-sectionally.

We do not use BBR's directly but, instead, as motivation for the link between revenues and expenditures. Previous research has primarily focused on the success of BBRs insofar as how well they limit deficit spending. States' BBRs vary in strength due to when they require balance in the budget cycle, whether they allow budget deficits to be carried forward, and the number of various technical provisions that are not subject to political machinations.[3] This variation is used to examine the speed of adjustment [Poterba (1994)],

---

[3]Advisory Commission on Intergovernmental Relations' (1987), Hou and Smith (2006)

cost shifting to accounts not subject to BBRs [Bohn and Inman (1996)], and the sustain-ability of budget deficits [Quintos (1995), Mahdavi and Westerlund (2011)]. This literature has generally found that BBRs limit deficits.

However, prohibiting deficit spending neutralizes the option to use counter-cyclical state fiscal policy to smooth business cycles. Levinson (1998), and Krol and Svorny (2007) examine whether BBR stringency is related to output volatility. These papers confirm that BBRs limit deficits, but they do so at the expense of flexibility in responding to output. In other words, states respond swiftly to budget gaps but this leads to procyclical fiscal policy that exacerbates business cycles. However, Fatás and Mihov (2006) find that states' discretionary spending destabilizes the economy, and that this effect dominates the auto-matic adjustments to macroeconomic conditions. Empirically, BBRs reduce macroeco-nomic volatility by limiting states' discretion.

3.2   Data

The dataset is annual and covers the period 1977-2009.[4] Disaggregated data are only available for the period 1977-2006. State finances and population are provided by *Census of Governments* (CoG), U.S. Bureau of the Census. I convert all monetary values to per capita real (2005) dollars using the GDP deflator from the Federal Reserve's FRED database.[5] We exclude Alaska and Hawaii due to unique fiscal structures. Wyoming is also generally excluded, see Bohn and Inman (1996), Mahdavi and Westerlund (2011). We retain Wyoming, however the qualitative results do not change with its exclusion.

The CoG annual data is for each state's fiscal year which typically ends June 30. The four exceptions are Alabama and Michigan (September 30), New York (March 31), and Texas (August 31). However, the personal income provided by CoG is for calendar year. Following Reed, et al. (2011), we construct fiscal year personal income using quarterly Bureau of Economic Analysis data.

We use the combined state-local government data for three reasons. First, state and local governments are inextricably linked via intergovernmental transfers. Second,

---

[4]Following Mahdavi and Westerlund (2011), we interpolate missing local values in 2001 and 2003

[5]Mahdavi and Westerlund (2011) adjust for inflation using the BEA's "state and local government con-sumption and investment" price index.

states evade BBRs by shifting costs to local governments. Third, local governments also operate under BBRs so this cost-shifting is cosmetic. If states shifted costs to the federal government then this would actually evade BBRs by utilizing the federal government's deficit spending authority.[6]

3.3 Elasticity and Progressivity

The tax system is represented, in reduced form, as

$$R_t = AY_t^{\eta}$$

where R_t is tax revenue, Y_t is income, and we have suppressed the state index. Progressivity is measured by η. If $\eta = 1$ then we have a flat tax of A. If $\eta > 1$, then tax revenue rises faster than income, i.e. it is progressive. Lastly, if $\eta < 1$ then tax revenue rises slower than income, i.e. it is regressive. To make this a linear regression we take logs

$$r_t = \alpha + \eta y_t + \varepsilon_t \tag{3.1}$$

where lower-case denotes logs and $\alpha = ln(A)$. This log specification reinforces the definition of η as the income elasticity of tax revenue.

The elasticity of revenues is related to the extensive literature on tax progressivity. A popular local measure of tax progressivity is the income elasticity of tax revenue, $\frac{dLog(TAX_t)}{dLog(Y_t)}$, however there are numerous measures of progressivity, ranging from local to global to uniform measures.[7] The relationship between tax progressivity and budget dynamics has not been studied, to the best of my knowledge. However, there is an extensive literature on the relationship between tax progressivity and economic growth, particularly across US states. Rather than elasticity, the economic growth literature has predominantly used a measure introduced in Koester and Kormendi (K&K) (1989) and subsequently extended by Becsi (1996). Becsi's extension turns out to be the inverse of the elasticity, however this is never mentioned. The primary reason is that the focus of this literature is on how the distortionary effects of *marginal tax rates* (MTR) influence economic growth. The connection to progressivity is made once one controls for the *average tax rates* (ATR).

---

[6]See Primo (2008) and Mahdavi and Westerlund (2011) for further discussion.
[7]See Seidl (2009) for a survey.

3.4   Empirical Methodology

If balanced budget requirements (BRRs) induce a linkage between revenues and expenditures then higher revenue elasticities (more progressive revenues) should be associated with higher expenditure elasticities. Consequently, we estimate income elasticities for revenues and expenditures for each state i in a three variable VAR.

$$x_{i,t} = \alpha_i + \sum_{j=1}^{2} \Phi_{i,j} x_{i,t-j} + \varepsilon_{it} \tag{3.2}$$

where

$$x_{i,t} = \begin{pmatrix} y_{i,t} \\ r_{i,t} \\ e_{i,t} \end{pmatrix}$$

$y_{i,t}$ is personal income in state i in year t; $r_{i,t}$ is government revenue; and $e_{i,t}$ is government expenditure, all in logs. Two lags are chosen based upon the Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC).[8]

We consider impulse-response functions (IRFs) for shocks to personal income. IRFs trace out the dynamic response of an endogenous variable to a unit shock to one of the endogenous variables. If we let $\Omega_i = E(\varepsilon_{it} \varepsilon_{it}')$, then a shock to an element of ε_{it} is correlated with the other elements, except for the special case where Ω is diagonal. Orthogonalized innovations are constructed using the triangular decomposition of the variance-covariance matrix.

This construction is recursive and thus depends on the ordering of the variables in the VAR. We order personal income first so that, in the triangular decomposition of the variance-covariance matrix, personal income can influence revenue and expenditure contemporaneously but not vice versa. Although this can be viewed as imposing identifying restrictions on the structural relationship, in this context we use it to isolate shocks to personal income in a forecasting exercise.[9]

---

[8]For the aggregate budget, BIC chooses one lag for 41/48 states, AIC chooses 3 or 4 lags for 28/48 states.
[9]See appendix E and Hamilton (1994) Chapter 11.4 and 11.6.

We classify states into two categories, *progressive* and *regressive,* based on whether the contemporaneous income elasticity is greater or less than 1, respectively. We average the IRFs for each group and compare the mean dynamics. We also regress the income elasticities of expenditures on those of revenues to summarize the association between revenues' response to state income and expenditures' response to state income in the cross-section. We account for different precision across states in the first stage estimation by using weighted least-squares (WLS).[10]

We consider three levels of the budget.

1. Total revenue vs. total expenditure: This includes all revenue sources except bond issuance. These are taxes, charges, misc, LUSI (liquor stores, utilities, and insurance trusts), and inter-governmental (IG). This includes all expenditures.

2. General revenues vs. total general expenditure: This excludes LUSI but retains revenue and expenditure IG transfers. LUSI represents, on average, 16.1% of revenue and 13.7% of spending.

3. Total revenue vs. non-capital expenditure: This includes all revenue sources and excludes capital expenditures. Capital expenditures represent, on average, 12.9% of state spending.

LUSI is the most complicated state revenue statistics Census collects.[11] Public pensions are generally agreed to be underfunded due to rosy actuarial projections of returns, although the degree of underfunding and remedies are disputed.[12] Shoag (2010) uses a new dataset on state pension returns to identify government spending shocks from unexpected pension return windfalls. He finds pension return windfalls lessen the need for direct contributions to the pension fund and those funds can be diverted to other general expenditure priorities. Excluding LUSI may increase the power of BBRs by ameliorating the difficulties in attributing public pension returns and the ease state legislatures have in cost-shifting to these opaque accounts.[13] Similarly, capital expenditures are less salient to voters and BBRs

---

[10]Lane (2003)Lane [73] and Afonso *et al.* (2010) apply the same methodology to international data.

[11]See Ch. 11 of Government Finance and Employment Classification Manual: http://www.census.gov/govs/www/class_ch11.html.

[12]For competing viewpoints see, e.g. Biggs (2010) and Baker (2011).

[13]There are other general accounting gimmicks states can employ, such as shifting "on-budget" items to "off-budget", however Poterba (1996) suggests these effects are small.
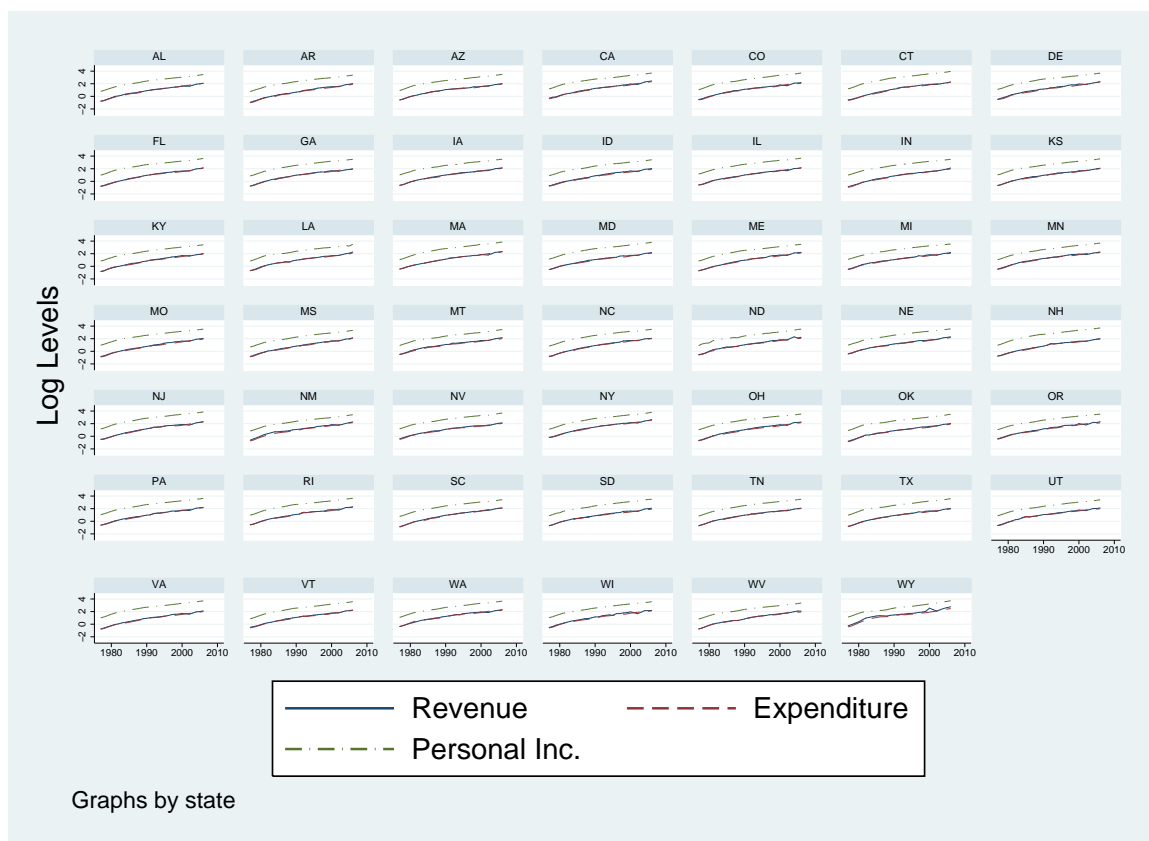
Figure 3.1: Time trend for total revenue, total expenditure, and personal income

generally apply to current expenditures.

3.4.1 Cointegration

With respect to tax revenue, cointegration has been applied inconsistently. Reed et al. (2011) is the first paper in the tax progressivity literature to apply cointegration. The income elasticity literature, however, has generally employed error-correction models (ECM) to deal with the cointegration between tax revenues and income. Although, even recent papers are mixed in whether they address nonstationarity.[14] The paramount objective is the distinction between short- and long-term consequences of tax structures. [15] In the budget

---

[14]See Reed, et al. (2011) footnote 11.

[15]For example, Wolswijk (2009) considers asymmetric responses to three levels of taxes (personal and corporate income tax, VAT) in the Netherlands. Fricke and Süssmuth (2011) examine a panel of Latin American countries and also consider an alternative to cointegration, an earlier deterministic time trend framework due to White (1983).

cyclicality literature, Sorensen et al. (2001) assumes cointegration but first-difference the variables. As discussed below, this is inconsistent with cointegration since it does not impose the long-run restriction. However, they check the robustness of their results using a regression in levels, which is the approach we follow.

Furthermore, the limited sample size is often used to justify not testing for cointegration. Rather, they justify cointegration theoretically.[16] For example, Wolswijk (2009) states (citations omitted and emphasis added)

> In view of the limited size of our sample and possible non-linear adjustment that are known to reduce the power of the test, we did not opt for the Johansen cointegration test. Instead, there is a *strong theoretical presumption of cointegration* by the fact that the equations, while including behavioural elements, mostly are of an arithmetic nature as there is only limited possibility to avoid taxation if the taxable event increasing the tax base occurs.

Figure 3.1 displays the time-series for total revenue, total expenditure, and personal income for each state. Revenue and expenditure are nearly indistinguishable in the figure. All of the series for each state are trending up over the sample period. This means the elasticities obtained from equation 3.2 are either spurious regressions due to independent trends or the series are *cointegrated.*

Looking at the series individually, Augmented Dickey-Fuller (ADF) tests suggest that revenue and expenditure are I(1), personal income is I(1) for over half the states, but cointegration does not hold.[17] This leaves us with three options.[18] First, differencing variables to obtain stationarity is common. However, a regression in differences is not consistent with cointegration, since it ignores the long-run equilibrium information found in the levels.[19] Second, we can carefully consider revenue, expenditure, and personal income for each state, determine if cointegration holds, and estimate an appropriate model. Unfortunately, the various unit-root tests have low power and are often contradictory. In addition, the model is misspecified if we impose cointegration when it does not hold (Type II error) or difference the variables when it does hold (Type I error). Third, we can estimate

---

[16]A theoretically justified cointegration vector can improve the power in testing for cointegration via bypassing the first stage estimation. See, e.g. Hamilton (1994) Ch. 19.2, p.582-586.

[17]We implement the residual cointegration test. This tests if the residuals are I(0) after finding the individual series to be I(1). This requires adjusted critical values found in Phillips and Ouliaris (1990) or Hamilton (1994) Table B.8.

[18]This discussion follows Hamilton (1994) Ch. 20.4 p. 651-653.

[19]See Engle and Granger (1987).

(a) Impulse-Response

(b) Cumulative Impulse-Response
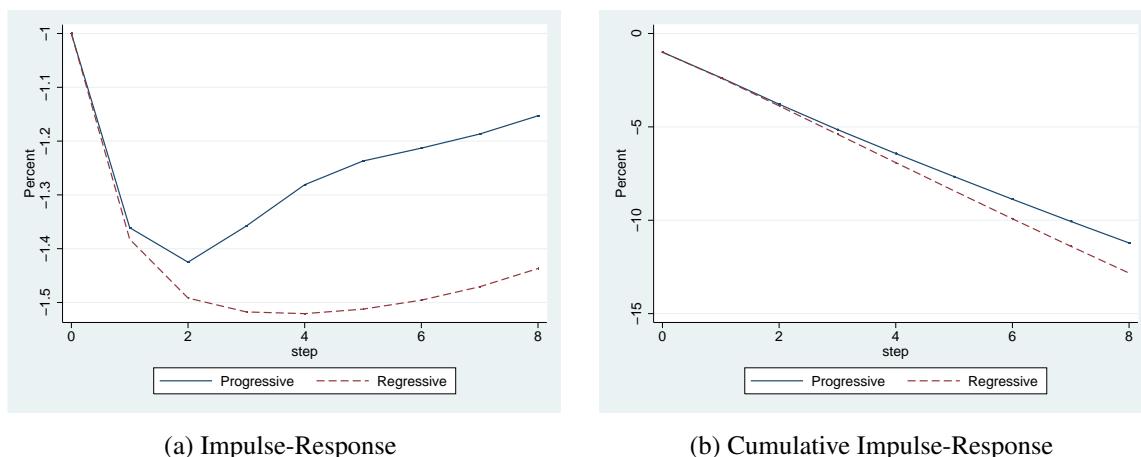
Figure 3.2: Total budget: IRF and CIRF of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income.

a VAR in levels. A VAR in levels is consistent with cointegration, if it exists, even though it ignores the non-stationarity. However, it is inefficient since it does not impose the long-run equilibrium restrictions. Also, even if the true DGP is a VAR in differences, certain functions of the parameters still have the same distribution in levels. Thus, we can still perform valid hypothesis tests.

Our specification in equation 3.2 is consistent with cointegration, albeit inefficient if true, while not introducing model misspecification via incorrect constraints.

3.5 Results

3.5.1 Total Revenues and Expenditures

We split the sample according to whether the impact, or contemporaneous, elasticity is greater or less than unity, i.e. whether revenues have a greater or lesser than 1:1 response to personal income shocks. We label impact elasticities of revenue greater than one *progressive*, and those less than one *regressive*. We find that 22/48 (~46%) of the states have progressive revenues according to this definition. We next average the impulse-response functions, subject to a -1% shock to personal income, over this partition. Impact and 2-year elasticities for each state can be found in appendix F.

(a) Impulse-Response   (b) Cumulative Impulse-Response
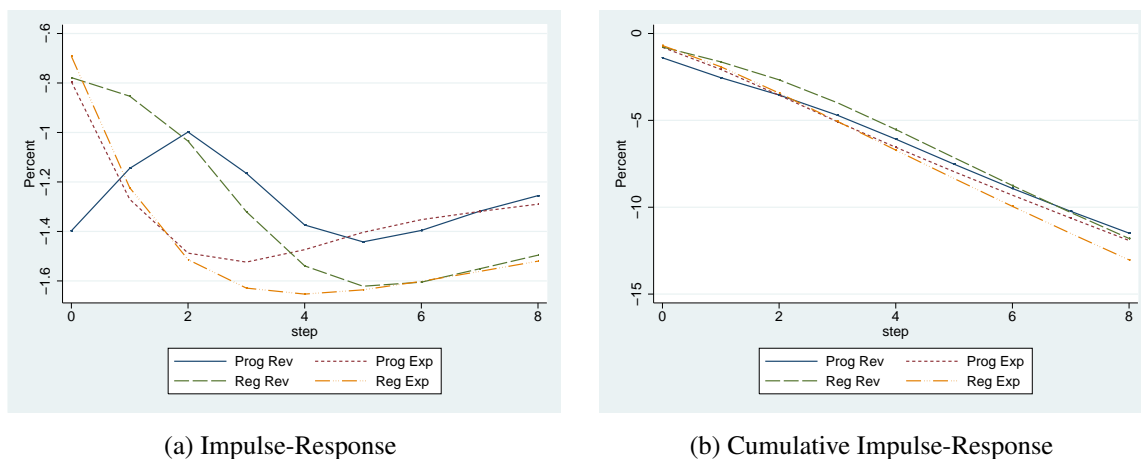
Figure 3.3: Total budget: IRF and CIRF of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

First, we check whether personal income behaved similarly in the two groups. Figure 3.2 shows the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income. In figure 3.2.A we see that the two categories have a similar response in the first year, however progressive states recover more quickly than regressive states. In figure 3.2.B we see that the regressive states have an excess persistence of about (12.8% - 11.2%) = 1.6% in the point estimates.

Figure 3.3 displays the IRF and CIRF for revenues and expenditures, for both progressive and regressive revenues. In figure 3.3.A we see that progressive revenues respond much more strongly than expenditures. For a 1% shock to personal income, the initial budget gap rises by about 0.6% for progressive states compared to 0.1% for regressive states. In order to close that gap, both revenues and expenditures adjust in progressive states; relative to the initial losses revenues rise and expenditures fall over the first two years. In contrast, regressive states continue to lose revenue and the adjustment occurs solely in spending. Interestingly, the spending response in the first two years is the same between progressive and regressive states. Thus, there appears to be limited scope for spending cuts and progressive states require revenue increases to close their larger initial budget deficit. In progressive states, the response of revenues and expenditures begin to track each other after about four years. This is earlier than in regressive states which take about five years.

Table 3.1: WLS regression of expenditure elasticity on revenue elasticity

	η_0^E
η_0^R	0.025
	(0.098)
constant	-0.692*
	(0.110)
R^2	0.001

Notes: η_0^R and η_0^E from equation 3.2.

* p<0.05, ** p<0.01, *** p<0.001

In figure 3.3.B, we see the cumulative response. The persistence is large and matches that found in personal income. All three measures decline by about 12% over 8 years in response to an initial 1% drop in personal income. Progressive states appear to be more successful in closing their budget gap. After two years the cumulative loss in revenues is less than that in expenditures and the two track each other closely thereafter. In regressive states, sharp cuts in spending relative to a minor budget gap closes the budget gap in less than a year. However, the persistent deep decline in spending relative to revenue results in a large surplus after 8 years according to the point estimates.

These results are the average dynamics for the two groups. In the cross-section we are interested in whether the more progressive revenue states also have more progressive expenditures. Table 3.1 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. The weights are the asymptotic standard errors of the IRF estimates. This says that a 1% increase in revenue elasticity is associated with a 0.025% increase in expenditure elasticity, although the result is not statistically significant. This is consistent with figure 3.3.A which shows a similar pattern in expenditure and very different patterns in revenues between the two groups.

3.5.2 General Revenues and Expenditures: Exclude LUSI

The LUSI (liquor stores, utilities, and insurance trusts) budget is opaque. Therefore, the BBRs may have more force, *de jure* or *de facto*, excluding LUSI. 32/48 (~67%) of states have progressive revenues once LUSI are excluded.

(a) Impulse-Response

(b) Cumulative Impulse-Response

Figure 3.4: Excluding LUSI: IRF and CIRF of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income.

Figure 3.4 shows the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income. In figure 3.4.A we see that, in contrast to the total budget, regressive states now recover more quickly from a personal income shock than progressive states. This is due to two factors. First, our partition is different than the total budget. The states now determined to be progressive with respect to the non-LUSI budget do not have the same income dynamics profile. Second, the non-LUSI budget contains different information than the total budget. To be precise, this is less information however we gain precision by eliminating the opaque LUSI budget, as discussed in section 3.4. Consequently, in figure 3.4.B we see that the excess persistence of personal income in progressive states is about (11.8% - 9.4%) = 2.4% in the point estimates after 8 years.

Figure 3.5 displays the IRF and CIRF for revenues and expenditures, for both progressive and regressive revenues. The results are a remarkable contrast to the total budget. In figure 3.5.A we see that the dynamics are virtually identical, however the progressive states IRFs are parallel shifted downward. Both groups rely predominately on spending cuts and reach equilibrium at around five years. In figure 3.5.B, we see the cumulative response. Both groups close the budget gap in about two years. In addition, the budget gap remains small for both groups. In the total budget, regressive states obtained a persistent budget surplus. The cumulative paths track those for income but the gap between progressive and regressive states is amplified. The final gap in revenues between progressive and

(a) Impulse-Response

(b) Cumulative Impulse-Response

Figure 3.5: Excluding LUSI: IRF and CIRF of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

Table 3.2: WLS regression of expenditure elasticity on revenue elasticity: Excluding LUSI

	η_0^E
η_0^R	0.569***
	(0.108)
constant	-0.193*
	(0.092)
R^2	0.374

Notes: η_0^R and η_0^E from equation 3.2. LUSI revenues and expenditures are excluded.
* p<0.05, ** p<0.01, *** p<0.001

regressive states is (12.6% - 8.5%) = 4.1% in the point estimates.

Table 3.2 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. This says that a 1% increase in revenue elasticity is associated with a 0.57% increase in expenditure elasticity and it is statistically significant. Following Shoag (2010) the stronger result may be due to the ability of state legislatures to use LUSI budget items to evade BBRs.

(a) Impulse-Response   (b) Cumulative Impulse-Response
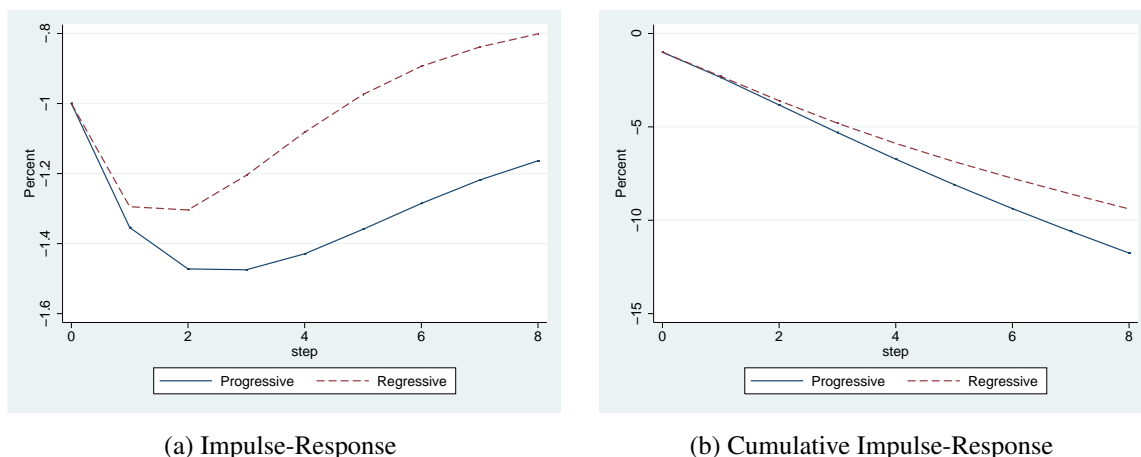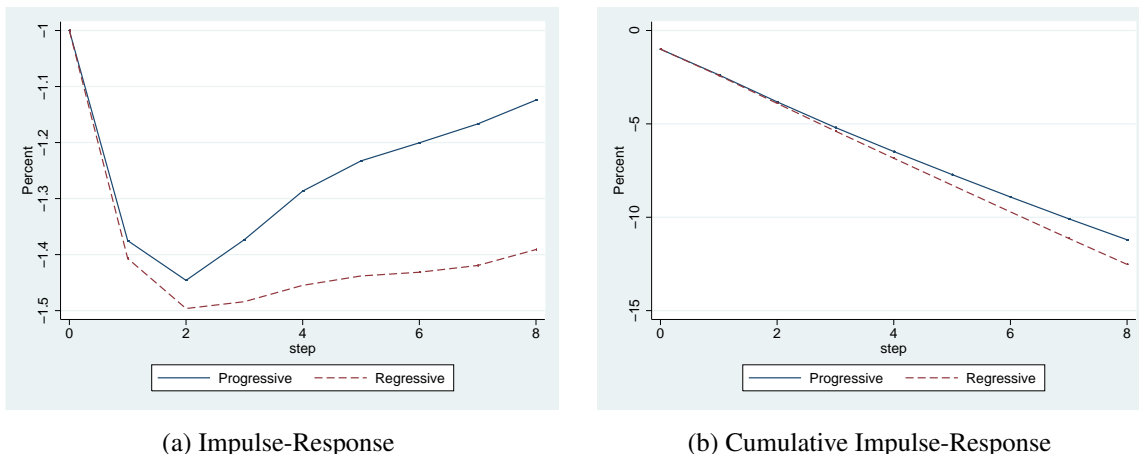
Figure 3.6: Excluding capital expense: IRF and CIRF of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income.

3.5.3   Total Revenues and Non-Capital Expenditures

Capital expenditures may be needed for public welfare, however they are not as salient as current expenditures on, e.g., education. Therefore, the BBRs may have more force, *de jure* or *de facto*, excluding capital expenditures. 35/48 (~73%) of states have progressive revenues once capital expenditures are excluded. This is close to the total budget since we are also using total revenues.

Figure 3.6 shows the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income. Since we have nearly the same partition as for the total budget, the responses are nearly identical. The excess persistence of regressive income is approximately (12.5%-11.2%)=1.3% in the point estimates.

Figure 3.7 displays the IRF and CIRF for revenues and expenditures, for both progressive and regressive revenues. The results are similar to those found in the total budget. In figure 3.7.A we see that progressive states adjust both revenues and expenditures to close a much larger budget gap than regressive states. Regressive states rely predominately on spending cuts. Spending dynamics are also similar for the first two years between progressive and regressive states, as we saw in the total budget.

In figure 3.7.B, we see the cumulative response. As in the total budget, we see that the the persistence is large and matches that found in personal income. Progressive states close their budget gap in two years and it remains small thereafter. In regressive states,

(a) Impulse-Response
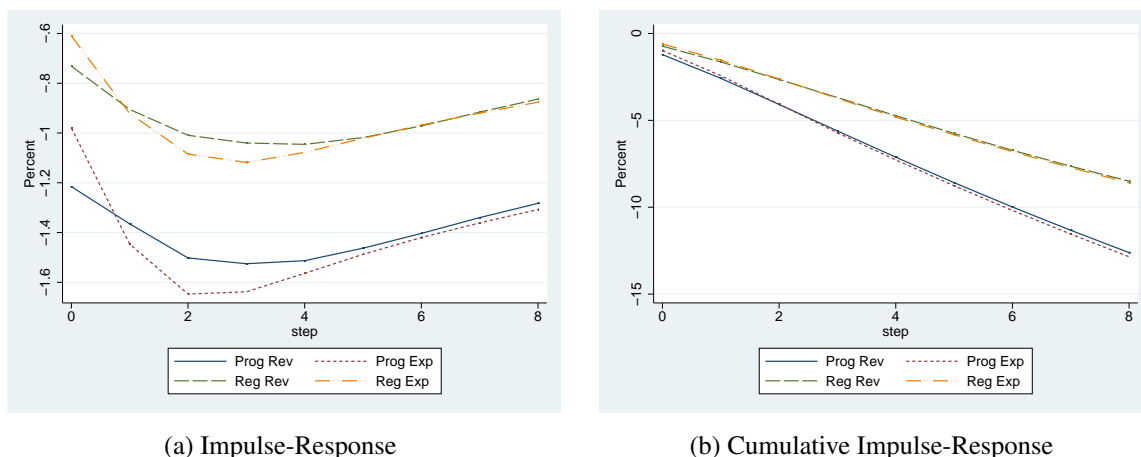
(b) Cumulative Impulse-Response
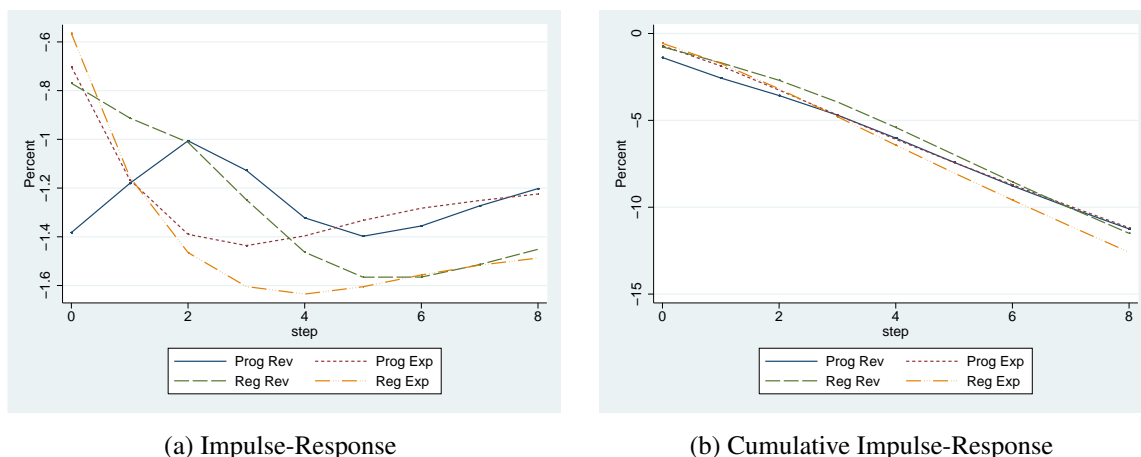
Figure 3.7: Excluding LUSI: IRF and CIRF of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

Table 3.3: WLS regression of expenditure elasticity on revenue elasticity

	η_0^E
η_0^R	0.145
	(0.106)
constant	-0.447***
	(0.117)
R^2	0.039

Notes: η_0^R and η_0^E from equation 3.2.

* p<0.05, ** p<0.01, *** p<0.001

sharp cuts in spending relative to a minor budget gap closes the budget gap in less than a year. However, the persistent deep decline in spending relative to revenue results in a large surplus after 8 years according to the point estimates.

Table 3.3 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. This says that a 1% increase in revenue elasticity is associated with a 0.15% increase in expenditure elasticity. However, the result is not statistically significant.

Taken as a whole, this suggests that capital expenditures are not used to shift budget items. Capital may be more salient to voters or more readily observable compared to LUSI.

3.6 Robustness Checks

In section 3.5 we documented stylized facts for three levels of state budgets. Due to the limited and conflicting evidence on the statistical form of state budget dynamics, we employ an estimation strategy that is robust to mispecification, specifically cointegration, as discussed in section 3.4. Here we consider two robustness checks. First, we check the stability of the time-invariant VAR coefficients by splitting the sample into two subperiods. Second, we include a time trend in the level VAR.

3.6.1 Impulse-Response Stability

We implicitly assume that the classification into progressive and regressive revenues, and their relationship with expenditures, is constant over the entire time-period, 1977-2006. We consider whether the relationship between revenues and expenditures is the same in two equal sub-periods in two ways. First, we employ the same methodology in section 3.4, however we restrict the time period to (1977-1991) and (1992-2006). This is the *in-sample* method. We are not attempting to identify the endogenous response of the taxing authority to state budget dynamics. However, it is known that states lowered taxes broadly in response to the excess, and ultimately unsustainable, capital gains revenue in the 1990s expansion. Our measure of progressivity may be sensitive to the mix of taxes and unduly influenced by those states with large exposures to capital gains.

Second, we classify the states using the first time period and estimate the dynamics in the second time period. This is the *pseudo out-of-sample* method. Here we examine whether states with observable higher volatility in the first period continue to have higher volatility in the second period, and vice versa.

We estimate the same VARs in the two methods. The difference is which states we choose to average the impulse-responses over. Due to space considerations and our previous results, we only examine the dynamics for the total budget and excluding LUSI items.

3.6.1.1   Total Revenues and Expenditures

For the total budget, we find that 35/48 (~73%) of the states have progressive revenues in 1977-1991 and this falls to 6/48 (12.5%) in 1992-2006. State budgets clearly behave much differently in the second half of the sample. During this period, states lowered personal income taxes and raised sales taxes, however sales tax revenues does not keep up with personal income growth. Therefore, this may reflect lagging state incomes during the expansion in the late 1990s.

We average the impulse-response functions, subject to a -1% shock to personal income, over these partitions for each time period.

First, we check whether personal income behaved similarly in the two groups. Figures 3.8 and 3.9 show the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income for each time period. In figure 3.8.A we see regressive states begin to recover a year earlier than progressive states and the two series appear to be converging in the first half of the sample. In figure 3.8.B, for the second half of the sample, we see that the shock immediately begins to die out for regressive states but it appears to be permanent for the progressive states, the point estimates exceeding unity at all horizons. In figure 3.9 we see the cumulative gap nearly double as we move to the second half of the sample.

Figures 3.10 and 3.11 display the IRF and CIRF for revenues and expenditures both sub-periods. In figure 3.10.A, for 1977-1991, we see regressive states have essentially no initial budget gap. Furthermore, there is a permanent -0.5% response in revenues but exploding expenditures. Progressive states have an initial budget gap and correct using both revenues and expenditures which converge at about 4 years. In figure 3.10.B, for 1992-2006, regressive states start with a surplus and have high oscillations in the budget over the next 8 years. Progressive states have a large initial budget gap and adjust using solely revenues.

Comparing figures 3.11.A and 3.11.B we see that the progressive states have roughly the same expenditure response however the larger initial revenue response in the second time period is never closed in the point estimates. The gap closes in about a year in 1977-1991 but converges for the first three years in 1992-2006 before trending with a minor budget gap. The point estimates for regressive states are more erratic, despite being an

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006
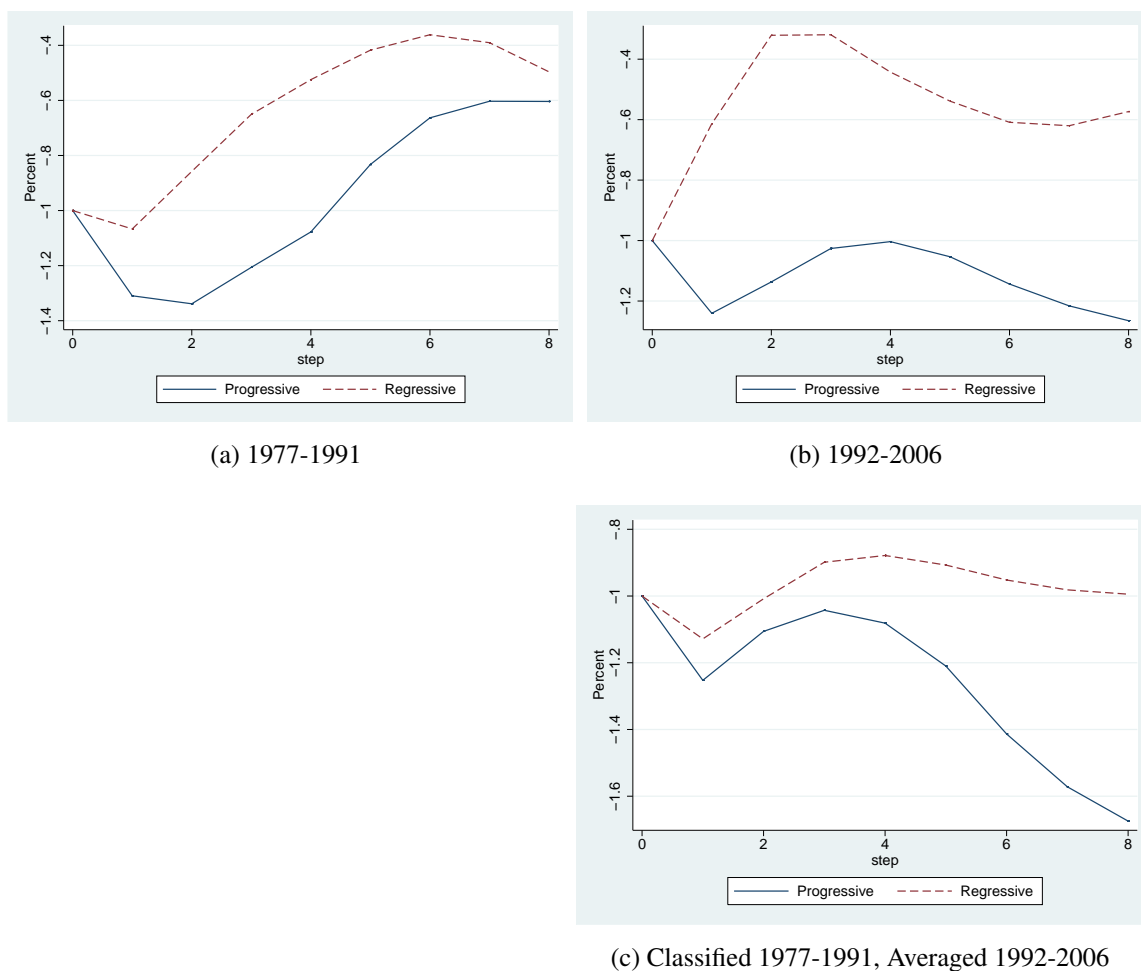
Figure 3.8: Total budget: Impulse-response function of personal income to its own shock
Note: Response to a -1% orthogonal shock to personal income. Figure (c) uses the same partition as figure (a).
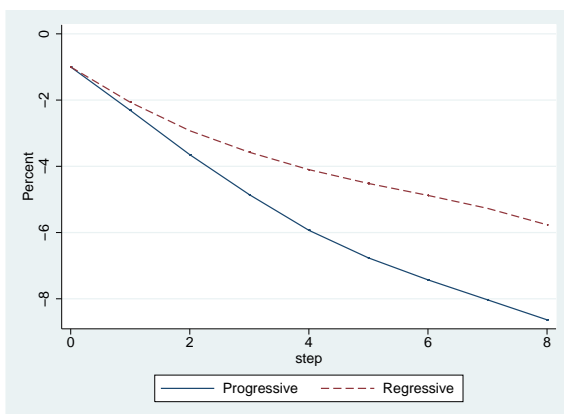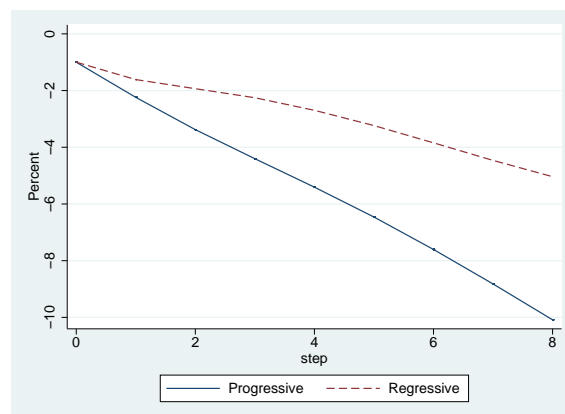
(a) 1977-1991
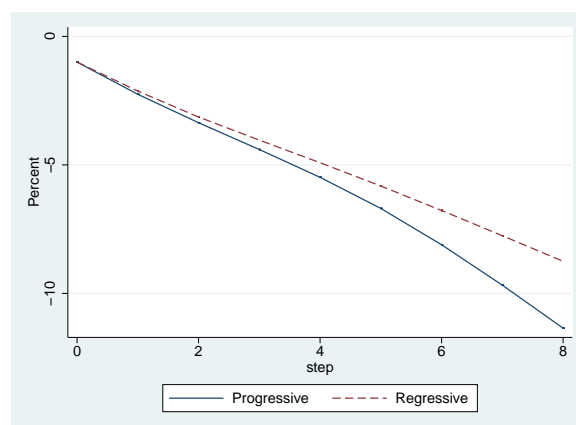
(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

Figure 3.9: Total budget: Cumulative impulse-response function of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income. Figure (c) uses the same partition as figure (a).

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

Figure 3.10: Total budget: Impulse-response function of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

Figure 3.11: Total budget: Cumulative impulse-response of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

Table 3.4: WLS regression of expenditure elasticity on revenue elasticity

	η_0^E (1977-1991)	η_0^E (1992-2006)
η_0^R (1977-1991)	0.551** (0.189)	
constant (1977-1991)	-0.405** (0.144)	
η_0^R (1992-2006)		-0.029 (0.067)
constant (1992-2006)		-0.832*** (0.148)
R^2	0.156	0.004

Notes: η_0^R and η_0^E from equation 3.2.

* p<0.05, ** p<0.01, *** p<0.001

average of 42/48 states in the second half of the sample. In 1977-1991, regressive states have a small surplus for the first four years before expenditures explode. In 1992-2006, the regressive states have a large persistent surplus which is at odds with the reported budget disasters following the dot-com bubble bursting in 2001. If we use the same partition as the first half, figure 3.11.C shows progressive and regressive states have the same initial budget gap, but regressive states close their gap one year earlier and the income shock is less persistent.

Table 3.4 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. The weights are the asymptotic standard errors of the IRF estimates. For 1977-1991, this says that a 1% increase in revenue elasticity is associated with a 0.55% increase in expenditure elasticity. For 1992-2006, it is a 0.03% drop in expenditure elasticity, although it is not statistically significant. Looking at figure 3.10.B we see that the two groups have nearly identical impact expenditure elasticities, although the subsequent dynamics are much different. This confirms that the average result holds in the cross-section.

3.6.1.2   General Revenues and Expenditures: Exclude LUSI

For the total budget, we find that 30/48 (62.5%) of the states have progressive revenues in 1977-1991 and 27/48 (56.25%) in 1992-2006. 17/30 (~57%) of the states overlap.

We average the impulse-response functions, subject to a -1% shock to personal income, over these partitions for each time period.

First, we check whether personal income behaved similarly in the two groups. Figures 3.12 and 3.13 show the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income for each time period. In figure 3.12.A we see regressive states begin to recover a year earlier, and more rapidly, than progressive states and the two series appear to be converging in the first half of the sample. In figure 3.12.B, for the second half of the sample, we see that regressive states follow the same path but progressive states never recover. This is due to three outliers; Indiana, Louisiana, and Mississippi. In figures 3.12.C, we maintain the same partition as the first half but average the IRFs over the second half. Regressive states still outperform. They track each other fairly closely for the first four years before diverging due to the three outliers. Figure 3.12.D removes those outliers and we see that the two groups perform remarkably similarly in the second half. In figure 3.13 we see the cumulative gap disappear in the second half after we remove the outliers.

Figures 3.14 and 3.15 display the IRF and CIRF for revenues and expenditures both sub-periods. These figures are much closer to the total budget than we found using the full sample in section 3.5. In figure 3.14.A, for 1977-1991, we see regressive states have essentially no initial budget gap. Furthermore, there is a permanent -0.5% response in revenues but expenditures rise. Progressive states have an initial budget gap and correct using both revenues and expenditures which converge at about 3 years. In figure 3.14.B, for 1992-2006, regressive states begin with a negligible budget gap that is rapidly closed, whereas progressive states begin with a larger gap. The divergence between revenue and expenditures after four years is driven by the outliers. Figure 3.14.D shows that the initial revenue response is the same if we use the first period classification, however regressive states recover more quickly. The initially classified progressive states have a lower impact expenditure response (implying larger deficits) however they perform worse than regressive states from 1 year on.

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

(d) Classified 1977-1991, Averaged 1992-2006, no outliers
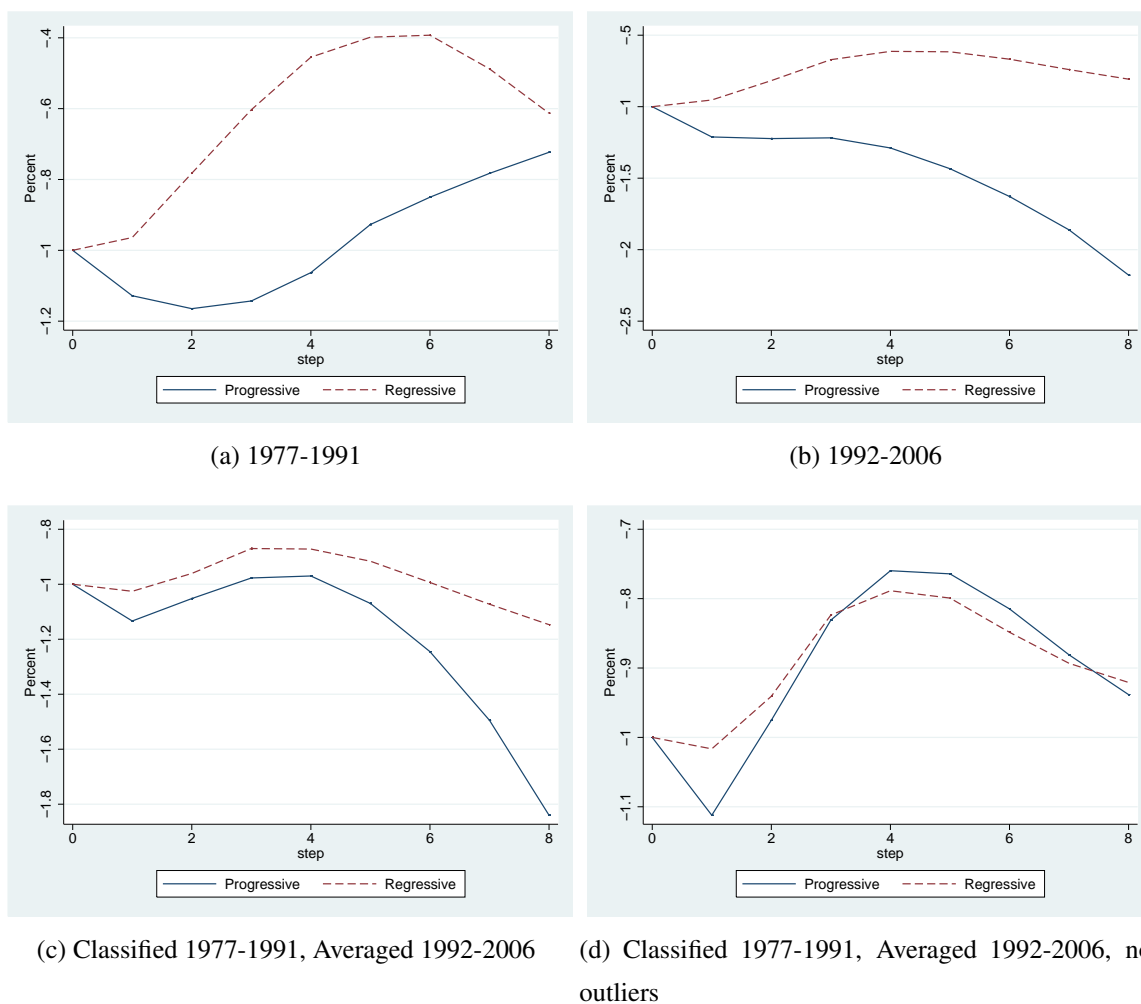
Figure 3.12: Non-LUSI budget: Impulse-response function of personal income to its own shock

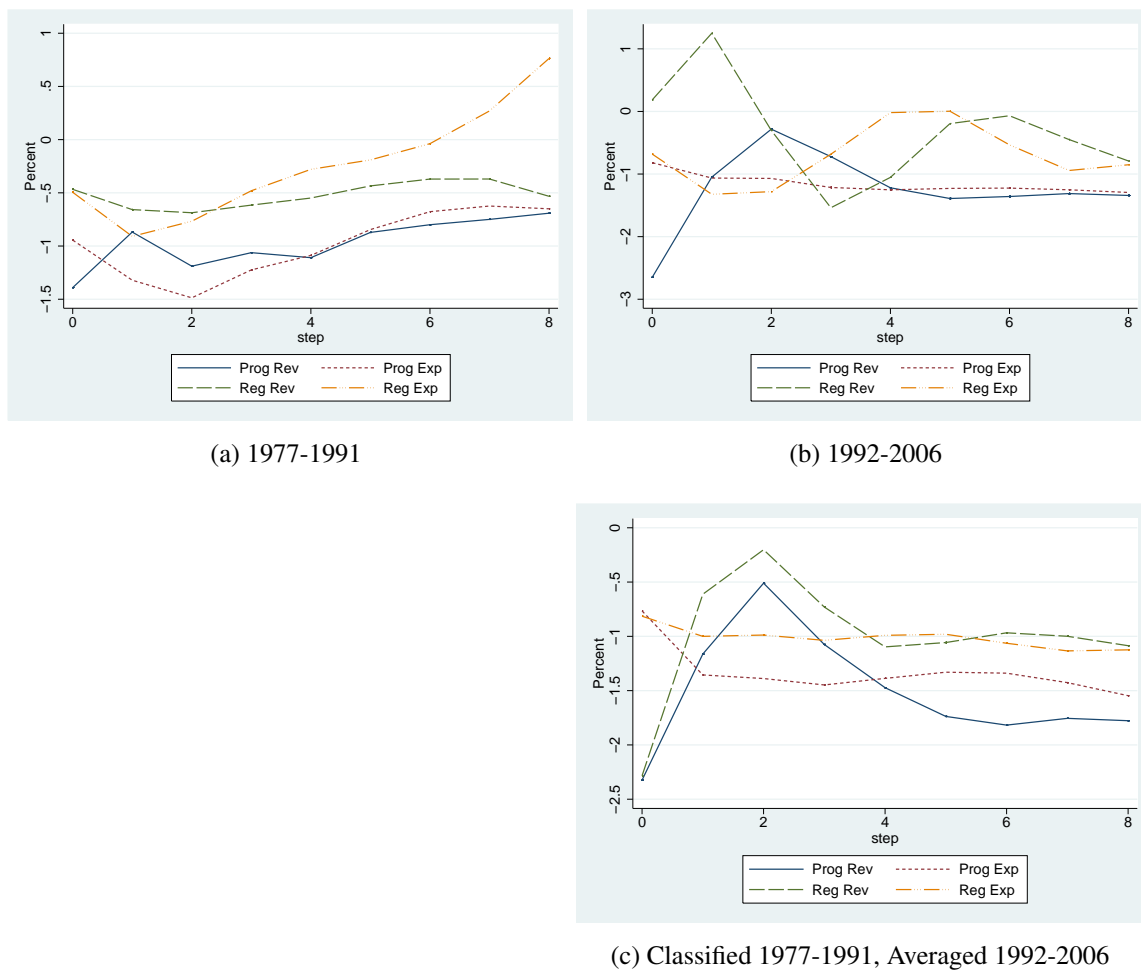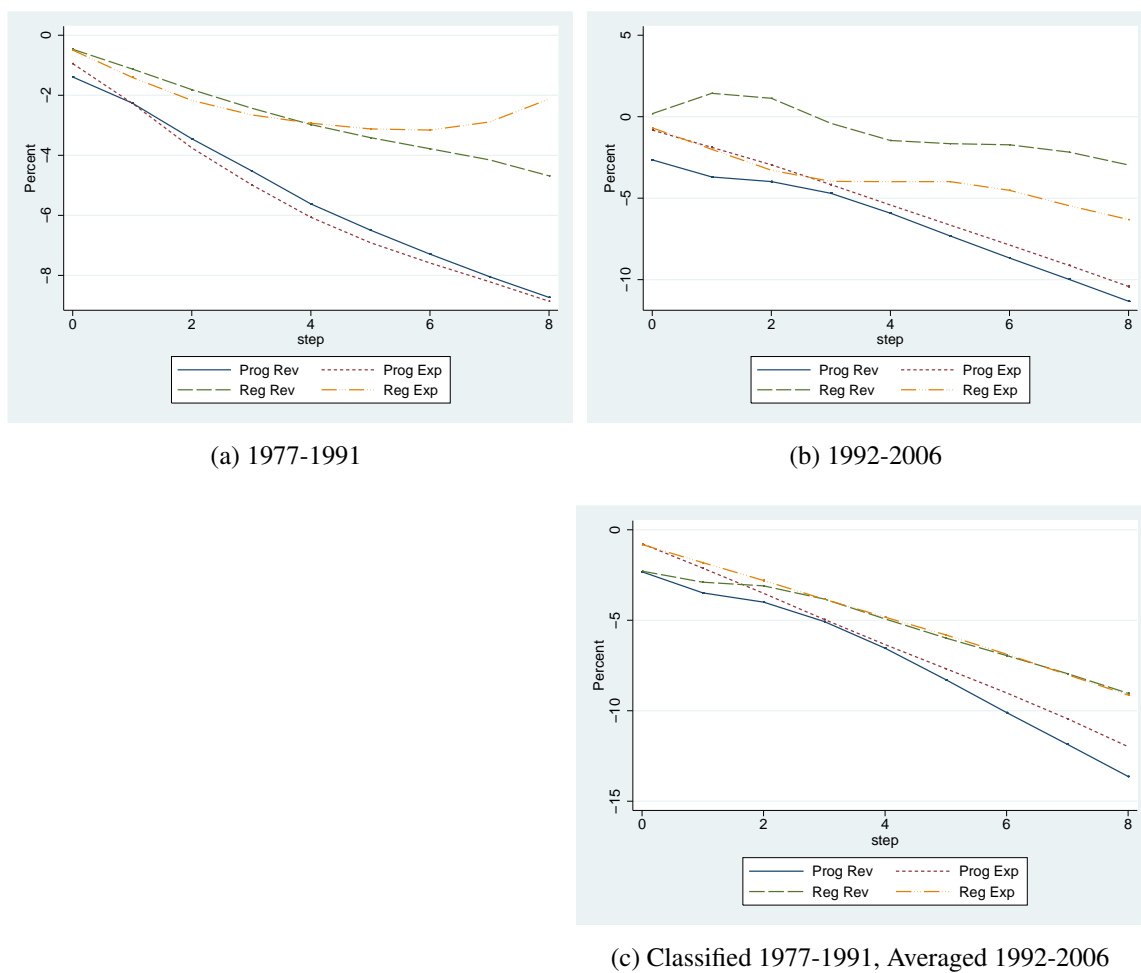Note: Response to a -1% orthogonal shock to personal income. Figures (c) & (d) uses the same partition as figure (a). Outliers are Indiana, Louisiana, and Mississippi.

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

(d) Classified 1977-1991, Averaged 1992-2006, no outliers

Figure 3.13: Non-LUSI budget: Cumulative impulse-response function of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income. Figures (c) & (d) uses the same partition as figure (a). Outliers are Indiana, Louisiana, and Mississippi.

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

(d) Classified 1977-1991, Averaged 1992-2006, no outliers

Figure 3.14: Non-LUSI budget: Impulse-response function of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income. Figures (c) & (d) uses the same partition as figure (a). Outliers are Indiana, Louisiana, and Mississippi.

(a) 1977-1991

(b) 1992-2006

(c) Classified 1977-1991, Averaged 1992-2006

(d) Classified 1977-1991, Averaged 1992-2006, no outliers
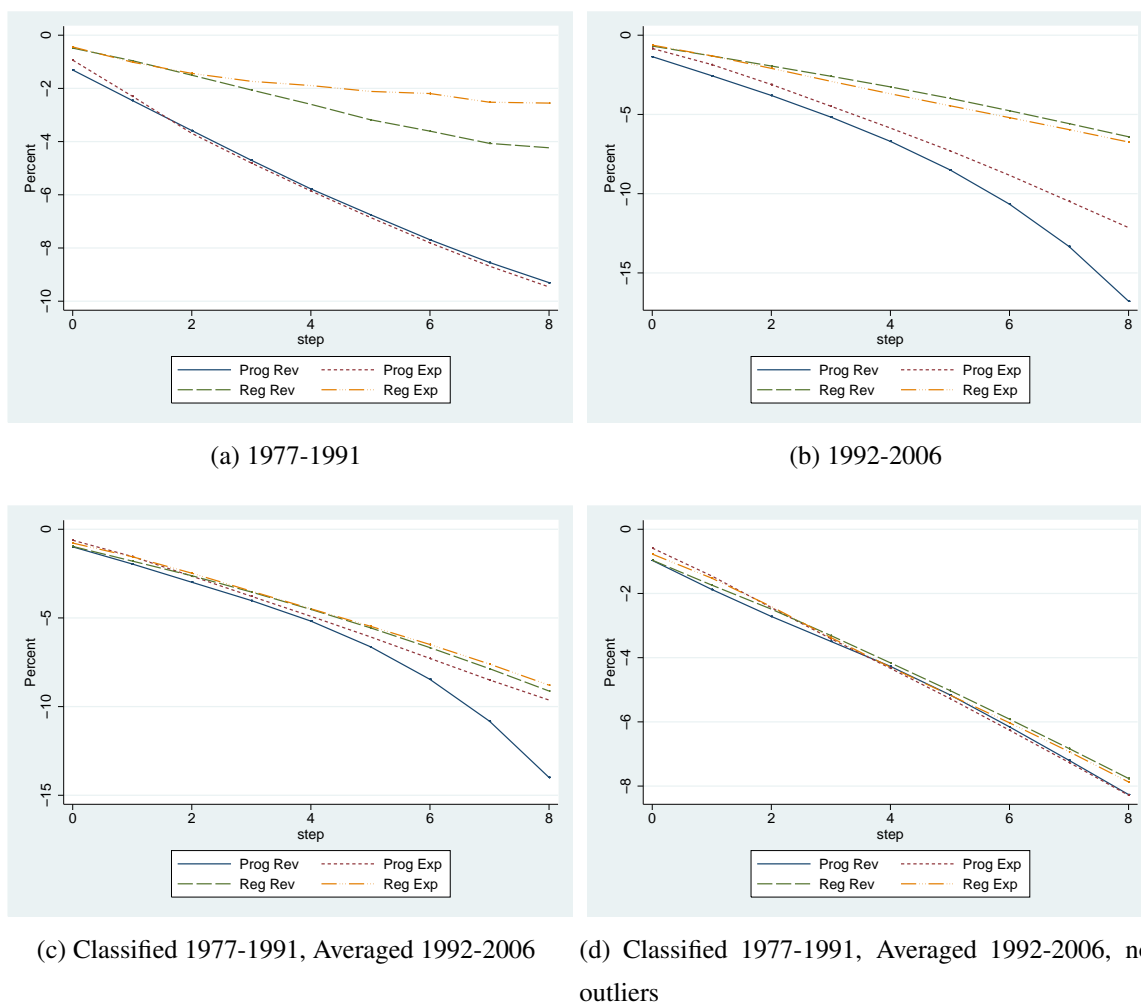
Figure 3.15: Non-LUSI budget: Cumulative impulse-response of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income. Figures (c) & (d) uses the same partition as figure (a). Outliers are Indiana, Louisiana, and Mississippi.
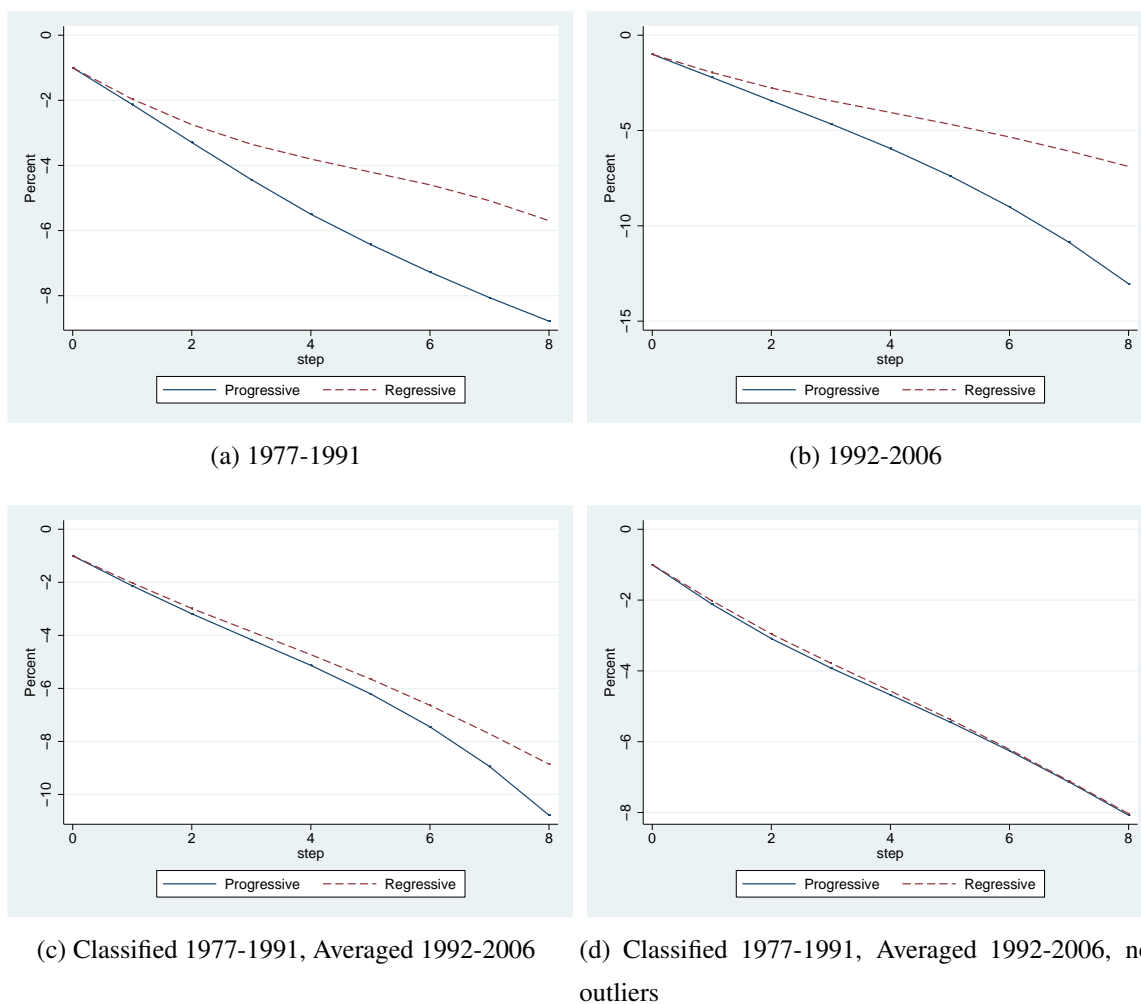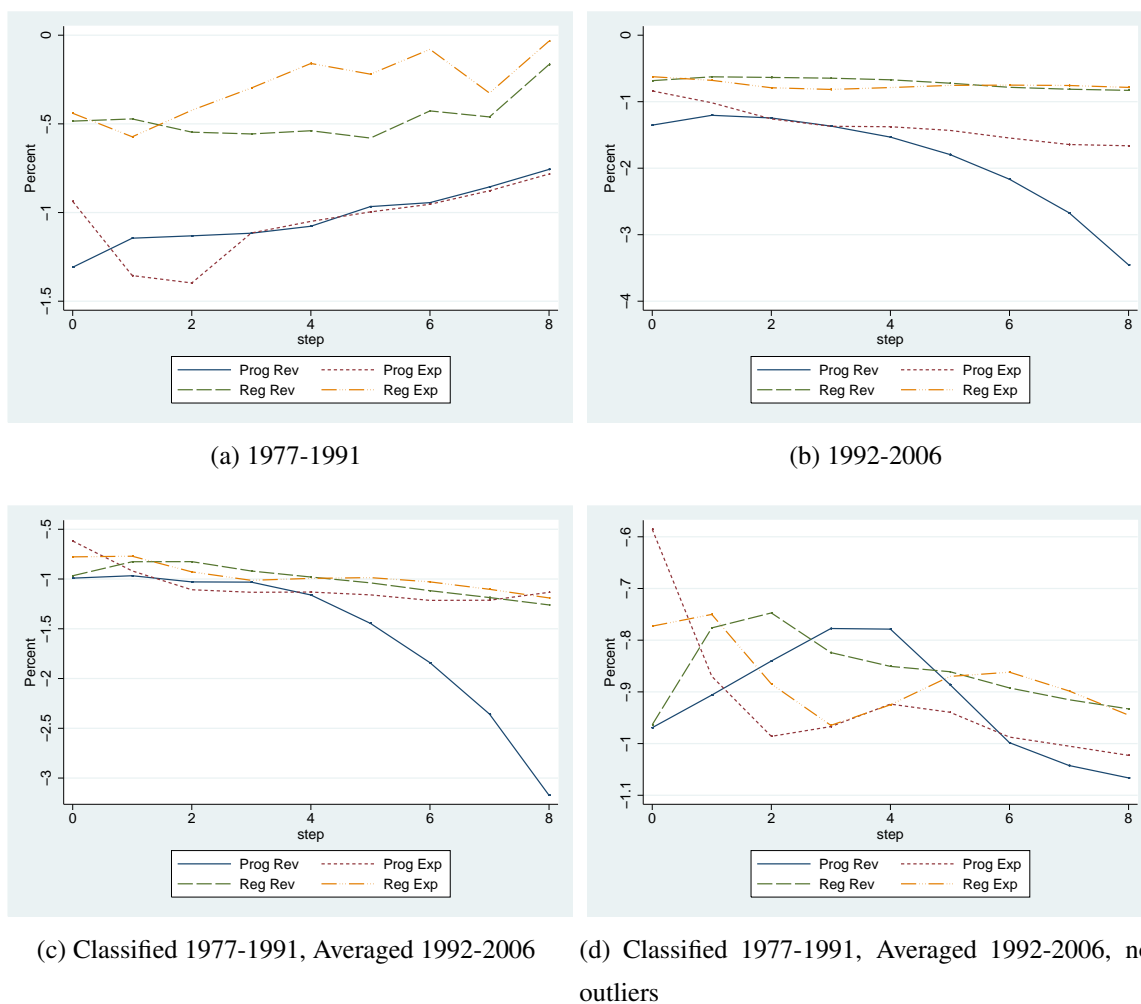
Table 3.5: WLS regression of expenditure elasticity on revenue elasticity: Outliers excluded

	η_0^E (1977-1991)	η_0^E (1992-2006)
η_0^R (1977-1991)	0.861***	
	(0.139)	
constant (1977-1991)	-0.009	
	(0.117)	
η_0^R (1992-2006)		0.236
		(0.121)
constant (1992-2006)		-0.409***
		(0.110)
R^2	0.470	0.082

Notes: η_0^R and η_0^E from equation 3.2. Outliers are Indiana, Louisiana, and Mississippi.
* p<0.05, ** p<0.01, *** p<0.001

Comparing figures 3.15.A and 3.15.D we see that the progressive and regressive states converge over the second half the sample, excluding outliers.

Table 3.5 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. For 1977-1991, this says that a 1% increase in revenue elasticity is associated with a 0.86% increase in expenditure elasticity. For 1992-2006, it is a 0.24% increase in expenditure elasticity, although it is not statistically significant (p-value = 5.6%). Compared to the total budget, the higher regression coefficient implies more volatile expenditures but smaller deficits. This supports the view that balanced budget requirements are more binding for the non-LUSI budget.

3.6.2   Level VAR With Trend

Our empirical methodology in section 3.4 is consistent with stationary or cointegrated series. Alternatively, the three series may be stationary around a common trend. One possibility is that the majority of expenditures increase at a statutory rate, e.g. 3%, with discretionary spending representing stationary noise around this trend. Similarly, the legislature strives for the same growth in revenues and the stationary shocks are policy errors and/or non-linear feedbacks from personal income shocks through the tax code. Lastly,

we assume real per-capita personal income grows according to trend representing techno-logical progress. In this scenario, BBRs are easily evaded in the short term and the common trend closes the budget. In the interest of space, we only show the results for the total budget and non-LUSI budget.

To be precise, equation 3.2 is augmented with a time trend for each state.

$$x_{i,t} = \alpha_i + \beta_i t + \sum_{j=1}^{2} \Phi_{i,j} x_{i,t-j} + \varepsilon_{it} \tag{3.3}$$

where

$$x_{i,t} = \begin{pmatrix} y_{i,t} \\ r_{i,t} \\ e_{i,t} \end{pmatrix}$$

$y_{i,t}$ is personal income in state i in year t; $r_{i,t}$ is government revenue; and $e_{i,t}$ is government expenditure, all in logs. Two lags are chosen based upon the Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC).

3.6.2.1 Total Revenues and Expenditures

For the total budget, we find that 36/48 (75%) of the states have progressive rev-enues. We next average the impulse-response functions, subject to a -1% shock to personal income, over this partition. First, we check whether personal income behaved similarly in the two groups. Figure 3.16 shows the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income. In figure 3.16.A we see that the progressive states have a more volatile response. The response of personal income for progressive states initially increases above and then overshoots below that of regressive states. This is in contrast to the results in section 3.5.1, where the two categories had sim-ilar responses in the first year and then diverged. The upward trend manifests as a quicker recovery. In figure 3.16.B we see that the volatility does not average out, but progressive states have an excess persistence of about (4.9% - 4.7%) = 0.2% in the point estimates. In contrast to the level-VAR results, the common trend ameliorates the impact of the personal

(a) Impulse-Response   (b) Cumulative Impulse-Response

Figure 3.16: Total budget: IRF and CIRF of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income.

income shock such that the cumulative IRF is nearly three times smaller. Furthermore, progressive revenues are more persistent than regressive revenues, contrary to what was found in section 3.5.1.

Figure 3.17 displays the IRF and CIRF for revenues and expenditures, for both progressive and regressive revenues. In figure 3.17.A we see that progressive revenues initially respond much more strongly than expenditures. Compared to section 3.5.1, the curves are parallel shifted up and progressive states are rotated counter-clockwise, however their revenue is flatter and does not exceed that in regressive states. This results in less overall elasticity but a larger gap between initial progressive and regressive responses that narrows over time.

In figure 3.17, we see the cumulative response. The progressive states do not overcome the initial disparity and narrow the budget gap in the first three years before widening again. Regressive states close their budget gap in less than a year. Without the time trend, the regressive states also recovered more quickly however they proceeded to accumulate a persistent budget *surplus*.

Table 3.6 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. The weights are the asymptotic standard errors of the IRF estimates. This says that a 1% increase in revenue elasticity is associated with a 0.25% increase in expenditure elasticity. Although the progressive
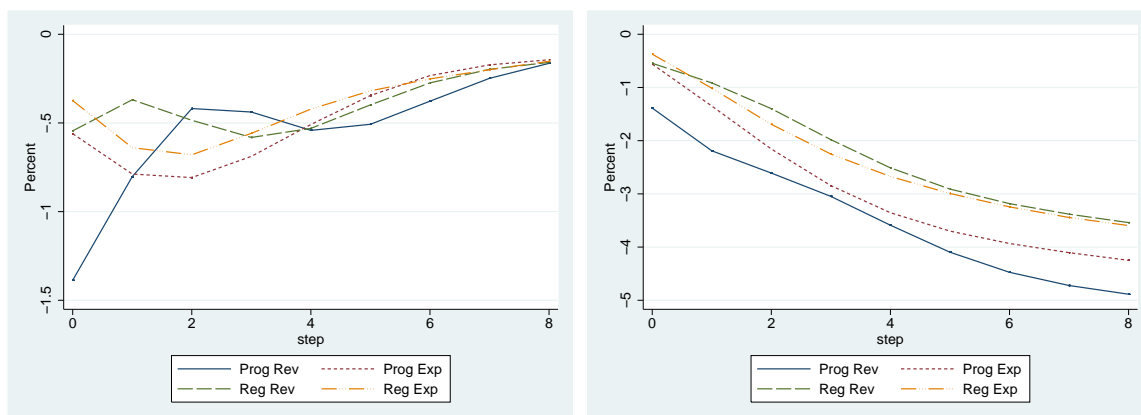
(a) Impulse-Response

(b) Cumulative Impulse-Response

Figure 3.17: Total budget: IRF and CIRF of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

Table 3.6: WLS regression of expenditure elasticity on revenue elasticity

	η_0^E
η_0^R	0.249**
	(0.084)
constant	-0.234*
	(0.101)
R^2	0.16

Notes: η_0^R and η_0^E from equation 3.3.

* p<0.05, ** p<0.01, *** p<0.001

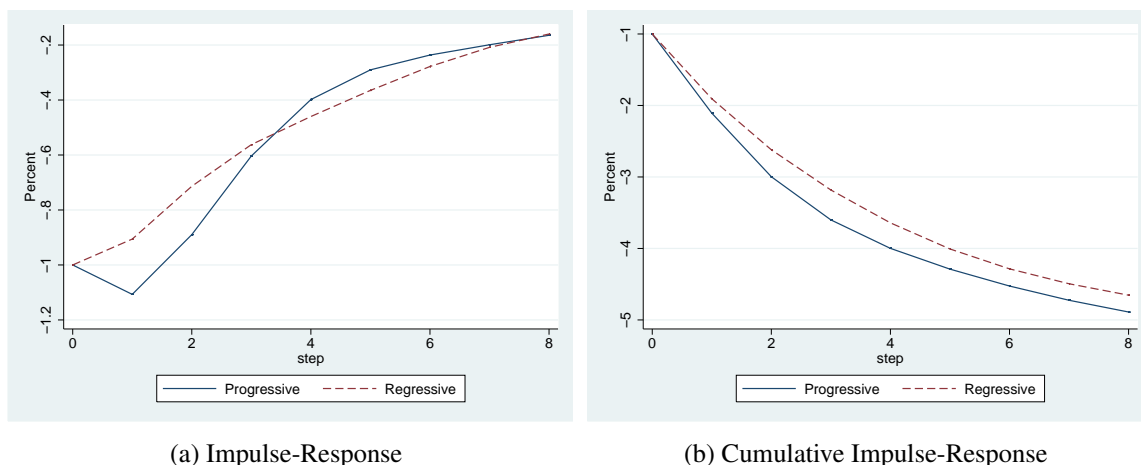(a) Impulse-Response   (b) Cumulative Impulse-Response
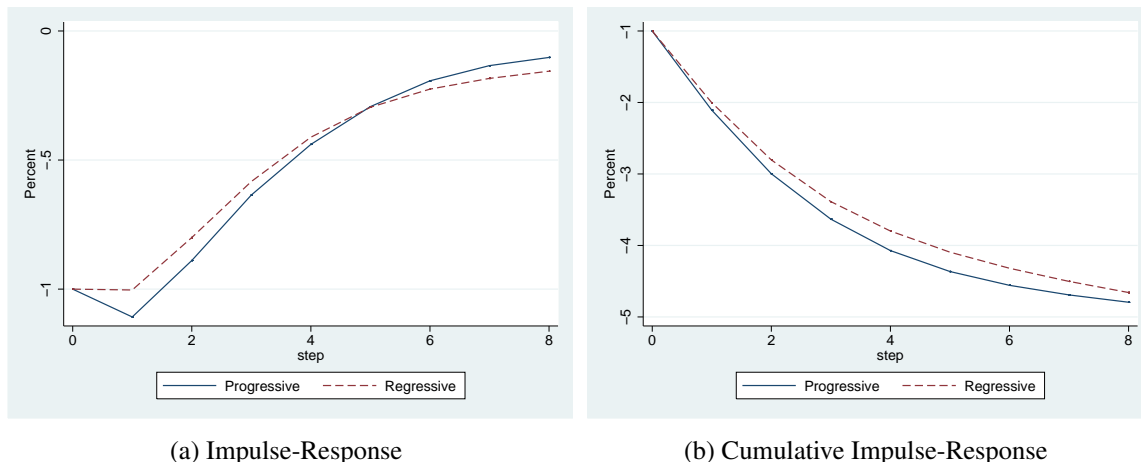
Figure 3.18: Excluding LUSI: IRF and CIRF of personal income to its own shock

Note: Response to a -1% orthogonal shock to personal income.

and regressive states exhibit similar dynamics with and without a trend, the larger initial difference in expenditures translates into a statistically significant relationship.

3.6.2.2 General Revenues and Expenditures: Exclude LUSI

We repeat the analysis for the non-LUSI budget. In section 3.5.2 we found that the non-LUSI budget had a tighter link between revenues and expenditures. That result is repeated here.

14/48 (~29%) of states have progressive revenues once LUSI are excluded. Figure 3.18 shows the impulse-response and cumulative IR of personal income for a -1% orthogonal shock to personal income. In figure 3.18.A we see that, in contrast to the total budget, the excess response of personal income for progressive over regressive states persists for several more years before reverting. However, the progressive states are not as extreme. Consequently, in figure 3.18.B we see that the excess response is less than for the total budget, and it is due to the more subdued response in the progressive states. The progressive states have an excess response of about (4.8% - 4.7%) = 0.1% in the point estimates after 8 years. This is in sharp contrast to the results without a trend. There we found that the gap in persistence between progressive and regressive states *expanded* when we considered the non-LUSI budget, rather than narrowing.

Figure 3.19 displays the IRF and CIRF for revenues and expenditures, for both

(a) Impulse-Response

(b) Cumulative Impulse-Response
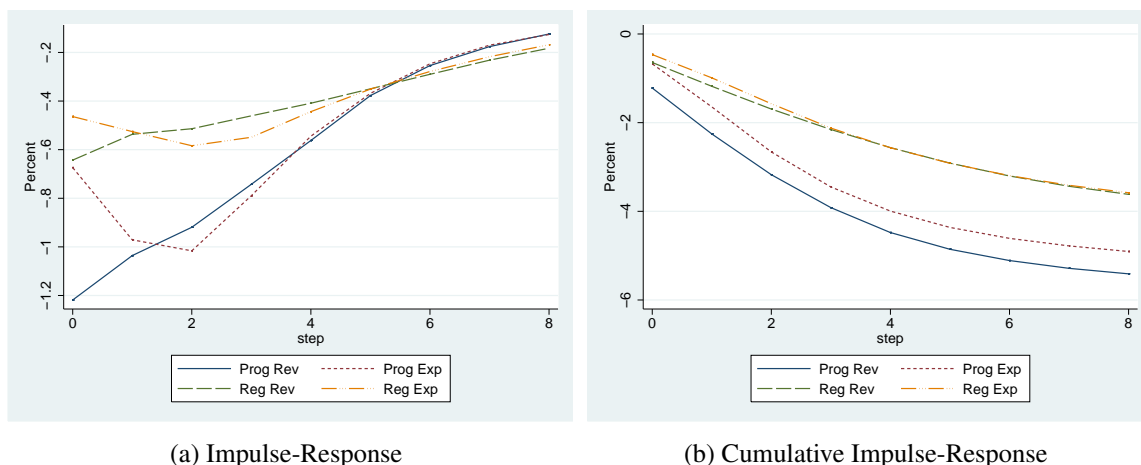
Figure 3.19: Excluding LUSI: IRF and CIRF of revenue and expenditures to personal income shock

Note: Response to a -1% orthogonal shock to personal income.

progressive and regressive revenues. In figure 3.19.A we see that progressive revenues respond much more strongly than expenditures. However, they reach response parity in about a year and then track each other thereafter. For regressive revenues, we see that the budget imbalance is initially much smaller and response parity is reached in about a year. As we found in the total budget, the progressive states appear to be rotated counter-clockwise relative to the no-trend VAR. In section 3.5.2 we observed a similar separation in the initial elasticities, however they remained separated at all horizons although revenues and expenditures converge for both classifications in both methods.

In figure 3.19, we see the cumulative response. For progressive revenues, the budget gap point estimates never tighten. This is due to the correction being insufficient to overcome the initial budget gap. In the no-trend VAR, we find the subsequent correction to be closer to the intial budget gap and the budget gap closes in the point estimates. The regressive states closes their budget gaps in both methods. The final gap in revenues is about 2%, half of that found in the no-trend VAR.

Table 3.7 displays the results of a weighted least-squares (WLS) regression of the impact expenditure elasticity on impact revenue elasticity. This says that a 1% increase in revenue elasticity is associated with a 0.55% increase in expenditure elasticity. This is statistically the same as the 0.57% found in section 3.5.2 and reinforces the Shoag (2010)

Table 3.7: WLS regression of expenditure elasticity on revenue elasticity: Excluding LUSI

	η_0^E
η_0^R	0.545***
	(0.106)
constant	-0.071*
	(0.085)
R^2	0.36

Notes: η_0^R and η_0^E from equation 3.2. LUSI revenues and expenditures are excluded.
* p<0.05, ** p<0.01, *** p<0.001

intuition that state legislatures are able to use LUSI budget items to evade BBRs. Further-more, the difference in dynamics is in terms of magnitude rather than relative values. In other words, both methods find progressive states have stronger and persistent response in both revenues and expenditures to a shock to personal income and thus have more volatile budget.

3.7 Conclusion

Balanced budget requirements (BBRs) induce a short-term relationship between state-level revenues and expenditures. Previous research has found that BBRs accelerate the closing of budget gaps and lower macroeconomic volatility by limiting state legis-lature's discretion. We find that more elastic revenues are associated with more elastic expenditures when we consider the most transparent budget categories in a VAR in lev-els. Thus, state income volatility is transmitted to state expenditure via the BBRs. For the budget excluding LUSI (liquor stores, utilities, and insurance trusts), a 1% increase in the income elasticity of revenue is associated with a 0.56% increase in expenditure elas-ticity. No statistically significant relationship is found when considering the total budget or excluding capital expenditures. This supports the Shoag (2010) evidence that the LUSI budget is used to bypass BBRs.

We consider two robustness checks. If we split the sample into two halves, the results hold for 1977-1991 but are subdued for 1992-2006. In addition, if we use the first-half partition to average the second-half impulse-response functions, we find that regressive

states close their budget gap more quickly and the income shock is less persistent. We also consider a deterministic trend in our VAR. The non-LUSI budget exhibits the same point estimate for the relationship between revenue and expenditure elasticity. In addition, the total budget now becomes statistically significant. A 1% increase in the income elasticity of revenue is associated with a 0.25% increase in expenditure elasticity.

With respect to tax revenue, at the federal level procyclical tax revenue has long been considered a virtue of the tax system. Going back to Musgrave and Miller's (1948) "built-in-flexibility", and Groves and Kahn (1952) (emphasis added) "..... the federal government whose special and strategic position in the economy makes *deficits and surpluses* from fluctuating revenue a blessing and hence instability of taxes (built-in flexibility) a virtue....." (p. 88).[20] However, states' balanced budget requirements preclude deficit spending. Adjustment is not instantaneous and transient deficits exist. Fatas and Mihov (2006) results suggest that this procyclical discretionary spending increases macroeconomic volatility. Since our results suggest more procyclical expenditures are associated with more procyclical revenues, the beneficial effects of procyclical revenues is outweighed by the negative impact of procyclical expenditures.

---

[20]See further discussion in the conclusions to Ahsan (2011)

A Converting GDP and Hours Estimates to Productivity Estimates

Let y_t be output and ℓ_t be hours, both in logs. Then log productivity is given by $x_t = y_t - \ell_t$. Given variances and covariances of output and hours, we can determine variances and covariances with productivity by simple identities.

The variance of productivity is given by

$$
\begin{aligned}
Var(x_t) &= Var(y_t - \ell_t) \\
&= Var(y_t) + Var(\ell_t) - 2Cov(y_t, \ell_t)
\end{aligned}
$$

The covariance between productivity and output is derived as follows

$$
\begin{aligned}
Cov(x_t, y_t) &= Cov(y_t - \ell_t, y_t) \\
&= Cov(y_t, y_t) - Cov(y_t, \ell_t) \\
&= Var(y_t) - Cov(y_t, \ell_t)
\end{aligned}
$$

The covariance between productivity and hours is derived as follows

$$
\begin{aligned}
Cov(x_t, \ell_t) &= Cov(y_t - \ell_t, \ell_t) \\
&= Cov(y_t, \ell_t) - Cov(\ell_t, \ell_t) \\
&= Cov(y_t, \ell_t) - Var(\ell_t)
\end{aligned}
$$

B Time-varying parameter VAR

B.1 Priors

The prior assumption that the VAR is stationary is imposed by multiplying the conditional density of the coefficients by an indicator function that is one if the roots of the VAR polynomial are all outside the unit circle. To be precise, let z^T denote a sequence of $z's$ up to time T. The conditional prior density is given by

$$p(\theta^T|\alpha^T,\sigma^T,V) \propto I(\theta^T)f(\theta^T|\alpha^T,\sigma^T,V)$$

where $I(\theta^T) = \prod_{t=0}^{T} I(\theta_t)$ is the product of indicators for each time time t. The conditional distribution further be decomposed by using the Markov property of the random walk on the coefficients.

$$f(\theta^T|\alpha^T,\sigma^T,V) = f(\theta_0)\prod_{t=1}^{T}f(\theta_t|\theta_{t-1},\alpha^T,\sigma^T,V) \tag{B.1}$$

The prior densities are calibrated by estimating a time-invariant VAR using the first ten years of data for each country. The prior densities and calibrations follow from Benati and Mumtaz (2007) and Primiceri (2005).

$$
\begin{aligned}
p(\theta_0) &\propto I(\theta_0)N(\hat{\theta}_{OLS}, \hat{V}(\hat{\theta}_{OLS})) \\
p(2\log\sigma_0) &= N(2\log\hat{\sigma}_{OLS}, 10 \times I) \\
p(\alpha_0) &= N(\hat{\alpha}_{OLS}, |\hat{\alpha}_{OLS}|) \\
p(Q) &= IW\left(\bar{Q}^{-1}, T_0\right) \\
p(S_i) &= IW\left(\bar{S}^{-1}, i+1\right) \\
p(W_{i,i}) &= IG\left(\frac{1}{2}, \frac{k_W}{2}\right)
\end{aligned}
$$

where $\hat{\theta}_{OLS}$ is the vector of OLS estimates of the VAR coefficients and $\hat{V}(\hat{\theta}_{OLS})$ is the associated covariance matrix using the initial sample of size T_0. $\hat{\alpha}_{OLS}$ and $\hat{\sigma}_{OLS}$ are the corresponding vectors from the decomposition of the OLS residual covariance matrix, $\hat{A}\hat{\Omega}_{OLS}\hat{A}' = \hat{\Sigma}\hat{\Sigma}'$.

$\bar{Q} = k_Q \times V(\hat{\theta}_{OLS})$ and $\bar{S} = k_S \times |\hat{\alpha}_{OLS}|$. $(k_Q, k_S, k_W) = (0.005, 0.00001, 0.0001)$ denote tuning parameters on the prior variances. Primiceri (2005) p.841-843 discusses how these tuning parameters determine the prior probability of time variation and provides details on a reversible jump MCMC method for choosing them.

B.2 Estimation

The Gibbs sampler proceeds by iteratively drawing from the conditional distribution of the current parameter, conditioning on past values and any realizations from the current iteration. Under regularity conditions, the iterations will eventually draw realizations from the true joint distribution. The following discussion maintains the assumption from the paper that the state equations follow a random walk. More general forms of the estimation can be found in Primiceri (2005).

Step 1: $p(\theta^T | z^T, \alpha^T, \sigma^T, V)$

The conditional distribution is Normal. Draws from the posterior are obtained via the Carter and Kohn (1994) algorithm. First, the coefficients, $\theta_{t|t}$, and precision matrices, $P_{t|t}$, are estimated by the Kalman Filter. The state-space is given by

$$z_t = G_t'\theta_t + A_t^{-1}\Sigma_t\varepsilon_t$$
$$\theta_t = \theta_{t-1} + v_t$$

and the familiar Kalman Filter recursion equations are employed. The final period coefficient is then drawn from the Normal distribution centered at the Kalman Filter estimates, $\hat{\theta}_T \sim N(\theta_{T|T}, P_{T,T})$. The mean and variance of the remaining coefficients follows the backward recursion

$$\theta_{t|t+1} = \theta_{t|t} + P_{t|t}P_{t|t+1}^{-1}(\hat{\theta}_{t+1} - \theta_{t|t})$$
$$P_{t|t+1} = P_{t|t}P_{t+1|t}^{-1}P_{t|t}$$

Where the coefficients are drawn according to $\hat{\theta}_t \sim N(\theta_{t|t+1}, P_{t|t+1})$.

Step 2: $p(\alpha^T|z^T, \theta^T, \sigma^T, V)$

This is done following the same procedure as in Primiceri (2005). This involves transforming the the measurment equation such that the Carter-Kohn algorithm can be employed. Specifically, we can rewrite the the state-space as

$$A_t\hat{z}_t = \Sigma_t\varepsilon_t$$
$$\alpha_t = \alpha_{t-1} + \zeta_t$$

where $\hat{z}_t = G_t'\theta_t$ is known given θ^T. The lower-triangular structure of A_t and diagonal Σ_t allows the Carter-Kohn algorithm to be employed equation by equation.

Step 3: $p(\sigma^T|z^T, \theta^T, \alpha^T, V)$

Application of the Carter-Kohn algorithm is no longer simple. The state-space is now

$$\hat{z}^* = \Sigma_t\varepsilon_t$$
$$\log\sigma_t = \log\sigma_{t-1} + \eta_t$$

which is non-linear in Σ_t. We employ the univariate algorithm by Jacquier, Polson, and Rossi (2004) for each element $\sigma_{i,t}$. Details for this application can be found in Cogley and Sargent (2005).

Note that Primiceri (2005) uses a different algorithm. He log-linearizes the measurement equation and invokes the Carter-Kohn algorithm. However the error term is no longer Normal. Therefore there is an additional step that uses Kim et al. (1998)'s mixture of Normals approximation to $\log \varepsilon_t$. For those interested in this approach, a more accurate approximation to $\log \varepsilon_t$ can be found in Omori et al. (2007).

Step 4: $p(V|z^T, \theta^T, \alpha^T, \sigma^T)$

Under the assumption that the block-diagonal elements of V are independent and the conjugate prior specification of our covariance matrices, we can draw separately from each conditional distribution using standard techniques. These are standard draws from Normal-Inverse Wishart and Normal-Inverse Gamma setups.

C Labor Friction Model

There are two labor frictions in the model; endogenous effort choice and convex labor adjustment costs. Endogenous effort choice provides an intensive margin that is not subject to the adjustment cost. The two shocks in the economy are technology shocks and consumption preference shocks. Both shocks follow AR(1) processes.

First start with a completely flexible labor market. Assuming effort has stronger diminishing returns in production and higher marginal disutility than employment, then the intensive margin is never used to adjust. Thus in a completely flexible market only employment adjusts. This leads to the following equilibrium equations

$$
\begin{aligned}
n_t &= (1-\eta)a_t + z_t \\
y_t &= a_t + (1-\alpha)z_t
\end{aligned}
$$

where n_t is employment, y_t is output, a_t is the technology shock, and z_t is the preference shock, all in logs. The parameters $\eta \in [0,1]$ and $\alpha \in (0,1)$ are the inverse of the intertemporal elasticity of substitution and diminishing returns to total labor, respectively. This leads to the following covariances

$$
\begin{aligned}
cov(y_t - n_t, y_t) &= \eta\,var(a_t) - \alpha(1-\alpha)var(z_t) \\
cov(y_t - n_t, n_t) &= \eta(1-\eta)var(a_t) - \alpha\,var(z_t)
\end{aligned}
$$

We are interested in how these change in response to a change in labor market frictions. However, for logarithmic utility over consumption ($\eta = 1$) we can unambiguously sign labor productivity as being countercyclical with respect to employment. The last moment we are interested in is the volatility of labor input relative to the volatility of output

$$\frac{var(n_t)}{var(y_t)} = \frac{(1-\eta)^2 var(a_t) + var(z_t)}{var(a_t) + (1-\alpha)^2 var(z_t)}$$

For the case of infinite labor market frictions, no new workers will be hired. Therefore, all labor adjustment will occur along the intensive margin. This leads to the following equilibrium equations

$$
\begin{aligned}
e_t &= (1-\eta)a_t + z_t \\
y_t &= (1+\phi)a_t + (1-\alpha)\psi z_t
\end{aligned}
$$

where the parameters $\phi \geq 0$ and $\psi \in [0,1]$ represent the marginal disutility of effort and the diminishing return to effort in production, respectively. Output then responds more aggressively to technology shocks but less to preference shocks. This leads to the following covariances

$$
\begin{aligned}
cov(y_t - e_t, y_t) &= (1+\phi)(\phi+\eta)var(a_t) + (1-\alpha)[(1-\alpha)\psi - 1]var(z_t) \\
cov(y_t - e_t, e_t) &= (\phi+\eta)(1-\eta)var(a_t) + [(1-\alpha)\psi - 1]var(z_t)
\end{aligned}
$$

Notice that the response to preference shocks are equivalent if $\psi = 1$. Thus ψ leads to a more negative response to preference shocks. However this is offset by the increased response to technology shocks.

Galí and van Rens (2010) show in their calibration that the larger technology shocks relative to preference shocks drive the US results.

D Labor Market Institutions Data

The labor market institution (LMI) data are drawn from the OECD[1], Nickell (2006)[2], and AIAS[3]. Table D.1 lists the variables investigated in our specification search. It also includes the source, source's variable name, and maximum dates available. This lists all variables used in both Gnocchi and Pappa (2011) and Rumler and Scharler (2011). Our final variable selection can be found in section 2.5.

Table D.1: Labor Market Institution Data Sources

Name	Variable	Source	Dates	Notes
Employment Protection Legislation (EPL)	epl_allard	Nickell	1960-2003	Details in Allard (2005a)
EPL: Regular Employment	epl_r	OECD	1985-2008	Version 1
EPL: Temporary Employment	epl_t	OECD	1985-2008	Version 1
EPL: Collective Dismissals	epl_cd	OECD	1998-2008	
Wage Coordination	wcoord	AIAS	1960-2010	
Wage Coordination	cowint	Nickell	1960-2000	Used in Rumler and Scharler (2011)
Union Density	ud	AIAS	1960-2010	
Union Density	udnet_vis	Nickell	1960-2004	Used in Rumler and Scharler (2011)
Union Coordination	uc	Nickell	1960-2000	
Unemployment: Wage Replacement Rate	brr_oecd	Nickell	1960-2003	
Unemployment: Wage Replacement Rate	nrw	Nickell	1960-2003	Details in Allard (2005b)[1]
Level of wage bargaining	level	AIAS	1960-2010	
Government intervention in wage setting	govtint	AIAS	1960-2010	
Union concentration	conc	AIAS	1960-2010	
Union centralization	cent	AIAS	1960-2010	
Mandatory extension of union contracts to non-unions	ext	AIAS	1960-2010	
Minimal Wage Setting	mws	AIAS	1960-2010	

Notes:

1) Includes the effective taxes and transfers in unemployment benefits' measure.

---

[1]Labour/Employment Protection at http://stats.oecd.org

[2]The dataset and definitions can be found at: http://eprints.lse.ac.uk/19789/

[3]Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts (ICTWSS): http://www.uva-aias.net/208. Data definitions can be found at: http://www.uva-aias.net/uploaded_files/regular/ICTWSScodebook30(2).pdf.

E Impulse-Response Functions[1]

Our primary forecasting equation in section 3.4is a three variable VAR for each state i.

$$x_{i,t} \;\; = \;\; \alpha_i + \sum_{j=1}^{2} \Phi_{i,j} x_{i,t-j} + \varepsilon_{it} \qquad\qquad \text{(E.1)}$$

where

$$x_{i,t} \;\; = \;\; \begin{pmatrix} y_{i,t} \\ r_{i,t} \\ e_{i,t} \end{pmatrix}$$

$y_{i,t}$ is personal income in state i in year t; $r_{i,t}$ is government revenue; and $e_{i,t}$ is government expenditure, all in logs.

Equation E.1 can be written in $MA(\infty)$ form as

$$x_{i,t} = \mu_i + \varepsilon_{it} + \Psi_{i1} \varepsilon_{i,t-1} + \Psi_{i2} \varepsilon_{i,t-2} + \dots$$

Thus, the matrix Ψ_s has the interpretation as

$$\frac{\partial x_{i,t+s}}{\partial \varepsilon_{i,t}'} = \Psi_s$$

We are interested in the income elasticity of revenues and expenditures. This is investigated by considering how, for example, our revenue forecast is updated given new information in income and holding current and past revenue and expenditure constant

$$\frac{\partial \hat{E}(r_{i,t+s}|y_{i,t}, x_{i,t-1})}{\partial y_{i,t}} \qquad\qquad \text{(E.2)}$$

---

[1]This section is adapted from Hamilton (1994) Chapter 11.4.

This new information is found in $\varepsilon_{i,t}^y$ holding the other elements constant. However, holding the other shocks constant is only meaningful if $\Omega_i = E(\varepsilon_{it}\varepsilon_{it}')$ is diagonal, otherwise there is contemporaneous correlation.

In order to isolate the impact of a single variable, we construct uncorrelated innovations $u_{it} \equiv A_i^{-1}\varepsilon_{it}$ where A_i is a lower triangular matrix with 1s along the principal diagonal in the *triangular decomposition*

$$\Omega = A_i D_i A_i'$$

and D_i is a diagonal matrix with positive entries along the diagonal. This gives a system of equations

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{it}^y \\ u_{it}^r \\ u_{it}^e \end{bmatrix} = \begin{bmatrix} \varepsilon_{it}^y \\ \varepsilon_{it}^r \\ \varepsilon_{it}^e \end{bmatrix}$$

Thus u_{it}^y is ε_{it}^y, and u_{it}^r is the residual from the projection of ε_{it}^r on (u_{it}^y, u_{it}^r) since the u_{jt}'s are orthogonal

$$\hat{E}(\varepsilon_{it}^r | u_{1t}) = a_{21} u_{it}^y$$

Similarly for ε_{it}^e

$$\hat{E}(\varepsilon_{it}^e | u_{1t}) = a_{31} u_{it}^y + a_{32} u_{it}^r$$

In this way, we recursively control for the information set. We order personal income first so that the shocks to revenue and expenditure account for the information already found in personal income. Returning to our object of interest, equation E.2, we see that this is comprised of two components, the direct effect of the shock and the indirect effect via the correlation of the innovations.

$$\begin{aligned} \frac{\partial \hat{E}(r_{i,t+s}|y_{i,t},x_{i,t-1})}{\partial y_{i,t}} &= \frac{\partial \hat{E}(r_{i,t+s}|y_{i,t},x_{i,t-1})}{\partial \varepsilon_{i,t}^r} \frac{\partial \hat{E}(\varepsilon_{i,t}^r|y_{i,t},x_{i,t-1})}{\partial y_{i,t}} \\ &= \psi_s^{r,y} a_{21} \end{aligned} \tag{E.3}$$

Where $\psi_s^{r,y}$ is the element of Ψ_s corresponding to the revenue-row, income-column, or, given our ordering in $x_{i,t}$, the 2nd row, 1st column element. A plot of equation E.3 with respect to s is the *orthogonalized impulse-response function* (IRF).

F Cross-sectional elasticity: Total budget

	Total Revenue and Expenditure			
	Impact Elasticity		2yr Elasticity	
	Revenue	Expenditure	Revenue	Expenditure
Alabama	1.56	0.78	1.05	1.71
Arizona	0.99	1.27	1.73	2.08
Arkansas	0.98	0.47	1.20	1.26
California	1.48	0.51	1.12	1.80
Colorado	1.16	0.41	0.19	1.12
Connecticut	0.98	0.56	0.74	1.10
Delaware	1.03	1.55	1.35	2.25
Florida	1.41	0.71	1.53	2.03
Georgia	0.76	0.55	0.60	1.67
Idaho	0.40	0.70	0.79	1.50
Illinois	1.05	0.62	0.45	0.98
Indiana	1.14	0.78	0.93	1.62
Iowa	1.19	0.85	0.99	1.48
Kansas	0.88	0.73	0.82	1.50
Kentucky	1.12	1.02	0.81	0.97
Louisiana	1.32	0.59	2.45	2.49
Maine	1.37	0.77	0.36	0.77
Maryland	1.49	0.85	0.82	1.76

Massachusetts	1.68	0.83	1.05	1.19
Michigan	1.27	0.38	0.62	1.03
Minnesota	0.93	0.51	0.56	1.22
Mississippi	1.23	1.13	0.86	1.21
Missouri	1.52	1.10	1.21	1.31
Montana	0.62	0.59	1.31	1.20
Nebraska	1.38	0.63	0.47	0.53
Nevada	1.74	0.59	1.33	1.05
New Hampshire	0.58	0.90	1.04	1.53
New Jersey	0.94	0.58	0.79	1.18
New Mexico	0.68	0.85	0.30	1.26
New York	0.40	0.73	2.09	2.21
North Carolina	0.81	0.50	1.43	1.69
North Dakota	0.94	0.95	1.32	1.53
Ohio	0.83	0.73	0.54	1.72
Oklahoma	1.30	0.63	1.45	1.77
Oregon	2.48	0.67	1.05	1.97
Pennsylvania	1.00	0.68	1.01	1.77
Rhode Island	1.25	0.91	1.39	1.66
South Carolina	1.02	1.02	1.07	1.95
South Dakota	0.33	0.76	1.28	1.38
Tennessee	0.69	1.38	1.28	1.78
Texas	0.95	0.34	0.69	0.90
Utah	2.17	0.63	0.40	1.07
Vermont	1.63	1.11	0.77	1.30
Virginia	1.34	0.90	1.22	1.90
Washington	0.97	0.60	0.81	1.83
West Virginia	0.92	0.84	1.20	1.76
Wisconsin	0.71	0.42	1.21	1.73
Wyoming	0.84	0.25	1.02	1.26

Bibliography

[1] **Afonso, Antonio, Luca Agnello, and Davide Furceri**, "Fiscal policy responsiveness, persistence, and discretion," *Public Choice*, 2010, *145*, 503–530.

[2] **Ahmed, Shaghil, Andrew Levin, and Beth Anne Wilson**, "Recent U.S. Macroeconomic Stability: Good Policies, Good Practices, or Good Luck?," *The Review of Economics and Statistics*, August 2004, *86* (3), 824–832.

[3] **Ahsan, Syed M.**, "The Instability of Tax Revenue and the Missing Risk Premium," *Journal of Business and Economic Research*, February 2011, *9* (2), 61–72.

[4] **Alexander, Gordon, Amy K. Edwards, and Michael G. Ferri**, "Trading volume and liquidity in Nasdaq's high-yield bond market," Working Paper, Securities and Exchange Commission 1998.

[5] **Allard, Gayle**, "Measuring Job Security Over Time: In Search of a Historical Indicator for EPL (Employment Protection Legislation)," Working Paper: WP05-17, Instituto de Empresa August 2005.

[6] ____ , "Measuring the Changing Generosity of Unemployment Benefits: Beyond Existing Indicators," Working Paper: WP05-18, Instituto de Empresa November 2005.

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

[8] ____ **and W. Ploberger**, "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 1994, *62* (6), 1383–1414.

[9] **Baker, Dean**, "The Origins and Severity of the Public Pension Crisis," Working Paper, Center for Economic and Policy Research February 2011.

[10] **Bao, Jack, Jun Pan, and Jiang Wang**, "The illiquidity of Corporate Bonds," Forthcoming, Journal of Finance May 2009.

[11] **Bauwens, Luc, Sébastien Laurent, and Jeroen V.K. Rombouts**, "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 2006, *21*, 79–109.

[12] **Baxter, Marianne and Robert G. King**, "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *The Review of Economics and Statistics*, November 1999, *81* (4), 575–593.

[13] **Becsi, Zsolt**, "Do State and Local Taxes Affect Relative State Growth?," *Federal Reserve Bank of Atlanta Economic Review*, March/April 1996, pp. 18–36.

[14] **Benati, Luca**, "Drift and Breaks in Labor Productivity," *Journal of Economic Dynamics and Control*, 2007, *31*, 2847–2877.

[15] ____ , "The "Great Moderation" in the United Kingdom," *Journal of Money, Credit and Banking*, February 2008, *40* (1), 121–147.

[16] ____ **and Haroon Mumtaz**, "US Evolving Macroeconomic Dynamics: A Structural Investigation," Working Paper 746, European Central Bank April 2007.

[17] ____ **and Paolo Surico**, "VAR Analysis and the Great Moderation," *American Economic Review*, 2009, *99* (4), 1636–1652.

[18] **Bernanke, Ben S. and Mark Gertler**, "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, March 1989, *79* (1), 14–31.

[19] ____ , ____ , **and Simon Gilchrist**, "The Financial Accelerator in a Quantitative Business Cycle Framework," in J.B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, Vol. 1, Elsevier Science B.V., 1999, chapter 21, pp. 1341–1393.

[20] **Biggs, Andrew G.**, "The Market Value of Public-Sector Pension Deficits," *American Enterprise Institute for Public Policy Research*, April 2010, *1*, 1–8.

[21] **Black, David C., Paul R. Corrigan, and Michael R. Dowd**, "New dogs and old tricks: do money and interest rates still provide information content for forecasts of output and prices?," *International Journal of Forecasting*, 2000, *16*, 191–205.

[22] **Blanchard, Olivier and John Simon**, "The Long and Large Decline in U.S. Output Volatility," *Brookings Papers on Economic Activity*, 2001, *1*, 135–174.

[23] **Blanco, Roberto, Simon Brennan, and Ian W. Marsh**, "An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps," *The Journal of Finance*, October 2005, *60* (5), 2255–2281.

[24] **Boehmer, Ekkehart and Julie Wu**, "Short selling and the price discovery process," May 2010. Working Paper.

[25] **Bohn, H and R.P. Inman**, "Balanced-budget Rules and Public Deficits: Evidence from the States," 1996. Carnegie-Rochester Conference Series on Public Policy.

[26] **Boivin, Jean and Marc P. Giannoni**, "Has Monetary Policy Become More Effective," *The Review of Economics and Statistics*, 2006, *88* (3), 445–462.

[27] ____ , **Michael T. Kiley, and Frederic S. Mishkin**, "How has the monetary transmission mechanism evolved over time?," in "Finance and Economics Discussion Series" number 2010-26 Board of Governors of the Federal Reserve System (U.S.) 2010.

[28] **Bollerslev, Time, Robert F. Engle, and Jeffrey M. Wooldridge**, "A Capital Asset Pricing Model with Time-Varying Covariances," *The Journal of Political Economy*, February 1988, *96* (1), 116–131.

[29] **Bruce, Donald, William F. Fox, and M. H. Tuttle**, "A Multi-State Analysis of Long-Run and Short-Run Dynamics," *Southern Economic Journal*, October 2006, *73* (2), 315–341.

[30] **Brügemann, Björn, Marcus Hagedorn, and Iourii Manovskii**, "Labor Productivity during Recession of 2008: Basic Data, Facts, and Figures," September 2010. Working Paper.

[31] **Canova, Fabio**, "Monetary Policy and the Evolution of the US Economy," Working Paper, ICREA, Universitat Pompeu Fabra, CREI and CEPR January 2006.

[32] ____ , "You can use VARs for structural analyses. A comment to VARs and the Great Moderation," Working Paper, ICREA, Universitat Pompeu Fabra, CREI and CEPR November 2006.

[33] ____ **and Luca Sala**, "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 2009, *56*, 431–449.

[34] ____ **, Luca Gambetti, and Evi Pappa**, "The Structural Dynamics of Output Growth and Inflation: Some International Evidence," *The Economic Journal*, March 2007, *117*, 167–191.

[35] **Carter, Chris K. and Robert Kohn**, "On Gibbs Sampling for State Space Models," *Biometrika*, 1994, *81* (3), 541–553.

[36] **Cecchetti, Stephen G., Alfonso Flores-Lagunes, and Stefan Krause**, "Has Monetary Policy Become More Efficient? A Cross-Country Analysis," *The Economic Journal*, April 2006, *116*, 408–433.

[37] **Clarida, Richard, Jordi Galí, and Mark Gertler**, "Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, February 2000, *115* (1), 147–180.

[38] **Cogley, Timothy and Thomas J. Sargent**, "Drifts and Volatilities: Monetary Policies and Outcomes in the post WWII US," *Review of Economic Dynamics*, 2005, *8*, 262–302.

[39] **Dadayan, Lucy and Donald J. Boyd**, "State Tax Revenues Show Record Drop, For Second Consecutive Quarter," October 2009.

[40] **den Haan, Wouter J.**, "The Comovement between output and prices," *Journal of Monetary Economics*, 2000, *46*, 3–30.

[41] **Duca, John V.**, "What Credit Market Indicators Tell Us," *Economic and Financial Review*, 1999, *(Q3)*, 2–13. Federal Reserve Bank of Dallas.

[42] **Engle, Robert**, "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, July 2002, *20* (3), 339–350.

[43] **Engle, Robert F. and C.W.J. Granger**, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 1987, *55* (2), 251–276.

[44] _____ **, Victor Ng, and Michael Rothschild**, "A multi-dynamic-factor model for stock returns," *Journal of Econometrics*, 1992, *52*, 245–266.

[45] **Fatas, A. and I. Mihov**, "The Macroeconomic Effects of Fiscal Rules in the US States," *Journal of Public Economics*, 2006, *90*, 101–117.

[46] **Faust, Jon, Simon Gilchrist, Jonathan H. Wright, and Egon Zakrajsek**, "Credit Spreads as Predictors of Real-Time Economic Activity: A Bayesian Model-Averaging Approach," Working Paper 16725, National Bureau of Economic Research January 2011.

[47] **Fernald, John G.**, "Trend breaks, long-run restrictions, and contractionary technology improvements," *Journal of Monetary Economics*, 2007, *54*, 2467–2485.

[48] **Fernandez, Carmen, Eduardo Ley, and Mark F.J. Steel**, "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 2001, *100*, 381–427.

[49] **Francis, Neville and Valerie A. Ramey**, "Measures of per Capita Hours and Their Implications for the Technology-Hours Debate," *Journal of Money, Credit and Banking*, September 2009, *41* (6), 1071–1097.

[50] **Fricke, Hans and Bernd Süssmuth**, "Growth and Volatility of Tax Revenues in Latin America," Working Paper 3312, Center for Economic Studies, the Ifo Institute, and the CESifo GmbH January 2011.

[51] **Galí, Jordi**, "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, March 1999, *89* (1), 249–271.

[52] _____ **and Luca Gambetti**, "On the Sources of the Great Moderation," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 26–57.

[53] _____ **and Thijs van Rens**, "The Vanishing Procyclicality of Labor Productivity," Working Paper, CREI, Universitat Pompeu Fabra and Barcelona GSE July 2010.

[54] **Gambetti, Luca, Evi Pappa, and Fabio Canova**, "The Structural Dynamics of U.S. Output and Inflation: What Explains the Changes?," *Journal of Money, Credit and Banking*, March-April 2008, *40* (2-3), 369–388.

[55] **Gertler, Mark and Cara S. Lown**, "The Information in the High-Yield Bond Spread for the Business Cycle: Evidence and Some Implications," *Oxford Review of Economic Policy*, 1999, *15* (3), 132–150.

[56] **Gilchrist, Simon, Vladimir Yankov, and Egon Zakrajsek**, "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets," *Journal of Monetary Economics*, 2009, *56*, 471–493.

[57] **Gollin, Douglas**, "Getting Income Shares Right," *Journal of Political Economy*, 2002, *110* (2), 458–474.

[58] **Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright**, "The U.S. Treasury Yield Curve: 1961 to the Present," Finance and Economics Discussion Series Paper 28, Federal Reserve Board June 2006.

[59] **Hagedorn, Marcus and Iourii Manovskii**, "Productivity and the Labor Market: Co-Movement over the Business Cycle," *International Economic Review*, August 2011, *52* (3), 603–619.

[60] **Hamilton, James D.**, *Time Series Analysis*, Princeton, New Jersey: Princeton University Press, 1994.

[61] **Hansen, Bruce E.**, "Approximate Asymptotic P Values for Structural-Change Tests," *Journal of Business & Economic Statistics*, January 1997, *15* (1), 60–67.

[62] **Hansen, Peter R. and Asger Lunde**, "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?," *Journal of Applied Econometrics*, December 2005, *20* (7), 873–889.

[63] **Hatzius, Jan, Peter Hooper, Frederic S. Mishkin, Kermit L. Schoenholtz, and Mark W. Watson**, "Financial Conditions Indexes: A Fresh Look after the Financial Crisis," Working Paper 16150, National Bureau of Economic Research July 2010.

[64] **Hodrick, Robert J. and Edward C. Prescott**, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, February 1997, *29* (1), 1–16.

[65] **Hou, Yilin and Daniel L. Smith (2006)**, "A Framework for Understanding State Balanced Budget Systems: Reexamining Distinctive Features and Operational Definitions," *Public Budgeting and Finance*, 2006, *26* (3), 22–45.

[66] **Hull, John C.**, *Options Futures and Other Derivatives*, 5th ed., Prentice Hall, 2002.

[67] **Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi**, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 1994, *12* (4), 69–87.

[68] **Justiniano, Alejandro and Giorgio E. Primiceri**, "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, June 2008, *98* (3), 604–641.

[69] **Kim, Sangjoon, Neil Shephard, and Siddhartha Chib**, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 1998, *65*, 361–393.

[70] **Koester, Reinhard B. and Roger C. Kormendi**, "Taxation, Aggregate Activity and Economic Growth: Cross-Country Evidence on Some Supply-Side Hypotheses," *Economic Inquiry*, July 1989, *27*, 367–386.

[71] **Krol, Robert and Shirley Svorny**, "Budget Rules and State Business Cycles," *Public Finance Review*, July 2007, *35* (4), 530–544.

[72] **Kydland, Finn E. and Edward C. Prescott**, "Time to Build and Aggregate Fluctuations," *Econometrica*, November 1982, *50* (6), 1345–1370.

[73] **Lane, Philip R.**, "The cyclical behavior of fiscal policy: evidence from the OECD," *Journal of Public Economics*, 2003, *87*, 2661–2675.

[74] **Levinson, Arik**, "Balanced budgets and business cycles: Evidence from the states," *National Tax Journal*, 1998, *51* (4), 715–732.

[75] **Litan, Robert E.**, "In Defense of Much, But Not All, Financial Innovation," Working Paper., The Brookings Institution February 2010.

[76] **Longstaff, Francis A., Sanjay Mithal, and Eric Neis**, "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market," *The Journal of Finance*, October 2005, *60* (5), 2213–2253.

[77] **Lubik, Thomas A. and Frank Schorfheide**, "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, March 2004, *94* (1), 190–217.

[78] **Lucas, Robert E.**, "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy*, 1976, *1* (1), 19–46.

[79] **Maccini, Louis J. and Adrian Pagan**, "Exploring the Role of Inventories in the Business Cycle," Working Paper, Johns Hopkins University and Australian National University 2003.

[80] **Mahdavi, Saeid and Joakim Westerlund**, "Fiscal stringency and fiscal sustainability: Panel evidence from the American state and local governments," *Journal of Policy Modeling*, October 2011, *33*, 953–969. Working Paper. The University of Texas at San Antonio, College of Business.

[81] **Mankiw, N. Gregory**, "Real Business Cycles: A New Keynesian Perspective," *Journal of Economic Perspectives*, Summer 1989, *3* (3), 79–90.

[82] **Merton, Robert C.**, "Theory of Rational Option Pricing," *The Bell Journal of Economics and Management Science*, Spring 1973, *4* (1), 141–183.

[83] ———, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *The Journal of Finance*, May 1974, *29* (2), 449–470.

[84] **Morris, Stephen D.**, "Linear Canonical Analysis of Nonlinear DSGE Models," Working Paper, UCSD 2011.

[85] **Nickell, Stephen**, "Unemployment and Labor Market Rigidities: Europe versus North America," *Journal of Economic Perspectives*, Summer 1997, *11* (3), 55–74.

[86] **Nickell, William**, "The CEP-OECD institutions data set (1960-2004)," November 2006. V1.

[87] **Ochel, Wolfgang**, "Collective Bargaining Coverage in the OECD from the 1960s to the 1990s," in "CESifo Forum" Winter 2001, pp. 62–65.

[88] **OECD**, *Collective Bargaining: Levels and Coverage*, Paris: OECD,

[89] ———, *Economic Performance and the Structure of Collective Bargaining*, Paris: OECD,

[90] **Ohanian, Lee E. and Andrea Raffo**, "Aggregate Hours Worked in OECD Countries: New Measurement and Implications for Business Cycles," Working Paper 17420, National Bureau of Economic Research September 2011.

[91] **Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima**, "Stochastic volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 2007, *140*, 425–449.

[92] **on Intergovernmental Relations, Advisory Commission**, "Fiscal Discipline in the Federal System: Experience of the States," 1987. Washington, D.C.

[93] **Phillips, P.C.B. and S. Ouliaris**, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, January 1990, *58* (1), 165–193.

[94] **Plosser, Charles I.**, "Understanding Real Business Cycles," *Journal of Economic Perspectives*, Summer 1989, *3* (3), 51–77.

[95] **Poterba, J. M.**, "State responses to fiscal crises: the effects of budgetary institutions," *Journal of Political Economy*, 1994, *102* (4), 799–821.

[96] **Poterba, James M.**, "Budget Institutions and Fiscal Policy in the U.S. States," *The American Economic Review*, May 1996, *86* (2), 395–400.

[97] **Primiceri, Giorgio E.**, "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 2005, *72*, 821–852.

[98] **Primo, David M.**, "Stop Us Before We Spend Again: Institutional Constraints on Government Spending," *Economics and Politics*, November 2006, *18* (3), 269–312.

[99] **Quintos, C.**, "Sustainability of the Deficit Process with Structural Shifts," *Journal of Business and Economic Statistics*, 1995, *13*, 409–417.

[100] **Ramey, Valerie A.**, "A Brief History of the Real Business Cycle - Search & Matching Literature," Note, UCSD February 2011.

[101] ____ **and Daniel J. Vine**, "Declining Volatility in the U.S. Automobile Industry," *American Economic Review*, December 2006, *96* (5), 1876–1889.

[102] **Ravn, Morten O. and Harald Uhlig**, "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *The Review of Economics and Statistics*, May 2002, *84* (2), 371–380.

[103] **Reed, W. Robert, Cynthia L. Rogers, and Mark Skidmore**, "On Estimating Marginal Tax Rates for U.S. States," *National Tax Journal*, 2011, *64* (1), 59–84.

[104] **Rumler, Fabio and Johann Scharler**, "Labor Market Institutions and Macroeconomic Volatility in a Panel of OECD Countries," *Scottish Journal of Political Economy*, July 2011, *3* (58), 396–413.

[105] **Saffi, Pedro A. C. and Kari Sigurdsson**, "Price Efficiency and Short Selling," *The Review of Financial Studies*, 2011, *24* (3), 821–852.

[106] **Schultz, Paul**, "Corporate Bond Trading Costs: A Peek Behind the Curtain," *The Journal of Finance*, April 2001, *56* (2), 677–698.

[107] **Seidl, Christian**, "Measurement of Tax Progression with Different Income Distributions," Working Paper, University of Kiel, Germany 2009.

[108] **Sensier, Marianne and Dick van Dijk**, "Testing for Volatility Changes in U.S. Macroeconomic Time Series," *The Review of Economics and Statistics*, August 2004, *86* (3), 833–839.

[109] **Shoag, Daniel**, "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns," Job Market Paper, Harvard University 2010.

[110] **Sims, Christopher A.**, "Macroeconomics and Reality," *Econometrica*, January 1980, *48* (1), 1–48.

[111] ____ **and Tao Zha**, "Were There Regime Switches in U.S. Monetary Policy?," *The American Economic Review*, March 2006, *96* (1), 54–81.

[112] **Sorensen, Bent E., Lisa Wu, and Oved Yosha**, "Output fluctuations and fiscal policy: U.S. state and local governments 1978-1994," *European Economic Review*, 2001, *45*, 1271–1310.

[113] **Stiroh, Kevin J.**, "Volatility Accounting: A Production Perspective on Increased Economic Stability," *Journal of the European Economic Association*, June 2009, *7* (4), 671–696.

[114] **Stock, James H. and Mark W. Watson**, "Has the Business Cycle Changed and Why?," *NBER Macroeconomics Annual*, 2002, *17*, 159–218.

[115] ____ **and** ____ , "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, September 2003, *41*, 788–829.

[116] ____ **and** ____ , "Understanding Changes in International Business Cycle Dynamics," *Journal of the European Economic Association*, September 2005, *3* (5), 968–1006.

[117] ____ **and** ____ , "Forecasting with Many Predictors," *Handbook of Economic Forecasting*, 2006, *1*, 515–554.

[118] **Taylor, John B.**, "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, 195–214.

[119] **Thoma, Mark A. and Jo Anna Gray**, "Financial market variables do not predict real activity," *Economic Inquiry*, 1998, *36*, 522–539.

[120] **Tse, Y. K. and Albert K. C. Tsui**, "A Multivariate Generalized Autoregressive Conditional Heteroskedasticity Model with Time-Varying Correlations," *Journal of Business & Economic Statistics*, July 2002, *20* (3), 351–362.

[121] **Uhlig, Harald**, "Economics and Reality," w16416, National Bureau of Economic Research September 2010.

[122] **White, Fred C.**, "Trade-Off in Growth and Stability in State Taxes," *National Tax Journal*, 1983, *36* (1), 103–114.

[123] **Wolswijk, Guido**, "The short- and long-run tax revenue response to changes in tax bases," *Economics Bulletin*, August 2009, *29* (3), 1960–1970.

[124] **Zellner, Arnold**, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in P.K. Goel and A. Zellner, eds., *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, North-Holland, Amsterdam, 1986, pp. 233–243.