

# Introduction to Bayesian Econometrics Course

Norges Bank

May, 2007

Overheads for Lecture on

Bayesian Analysis of Extensions of AR  
and VAR Models

Gary Koop, University of Strathclyde

# 1 Introduction

- Autoregressive models can be written as regression models (see lecture 2)
- VAR models closely related to SUR model (see lecture 3)
- Hence, we will briefly deal with AR and VARs.
- Extensions of them are increasingly used with empirical work and we will spend more time on these.
- Basic idea: conditional on a parameter (e.g. probability of transition) or a vector of latent data (e.g. a vector of states) such extensions reduce to regression models.
- Thus, use Gibbs sampler

# A Digression: Unit Roots and Cointegration

- We will not have time to discuss issues relating to nonstationary variables.
- Survey paper on cointegration in Palgrave Handbook of Econometrics, Volume 1: Theoretical Econometrics (Koop, Strachan, van Dijk and Villani)
- Non-Bayesians worry since asymptotic distributions of test statistics, etc. become different if unit root present
- Bayesians proceed conditional on data, so are rarely interested in asymptotics.
- Thus, many Bayesians uninterested in unit root issues (e.g. just estimate a VAR, not worry if cointegration is present or not).

- Likelihood function of AR/VAR does not depend on whether variable(s) have unit roots.
- Thus, the Bayesians who do write unit root/cointegration papers typically focus on prior elicitation.

## 2 Autoregressive Models

- As discussed in lecture on “Forecasting in Dynamic Factor Models Using Bayesian Model Averaging” AR(p) model can be written as a linear regression model.
- All our results for Bayesian analysis of regression model can be used.
- Example based on Geweke (1988, JBES)
- $y_t$  is log real GDP, follows AR(3):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \epsilon_t,$$

where  $\epsilon_t$  is i.i.d.  $N(0, h^{-1})$ .

- Properties of  $y_t$  depend on the roots of the polynomial  $1 - \sum_{i=1}^3 \beta_i z^i$  (denote by  $r_i$  for  $i = 1, \dots, p$ )
- Features of interest:  $C = \{\text{Two of } r_i \text{ are complex}\}$  and  $D = \{\min |r_i| < 1\}$
- $C$  and  $D$  are regions which are nonlinear functions of  $\beta_1, \beta_2, \beta_3$ .
- $C$  implies real GDP exhibits an oscillatory response to a shock
- $D$  implies an explosive response to a shock.
- Want  $\Pr(\beta \in C|y)$  and  $\Pr(\beta \in D|y)$ .

## 2.1 Posterior Analysis of the AR(p) Model

- Noninformative prior:

$$p(\beta_0, \dots, \beta_p, h) \propto \frac{1}{h},$$

- Simple treatment of initial conditions:  $y = (y_{p+1}, \dots, y_T)'$  and treat  $y_1, \dots, y_p$  as fixed initial conditions.
- AR(p) can be written as:

$$y = X\beta + \epsilon,$$

where  $\epsilon = (\epsilon_{p+1}, \dots, \epsilon_N)'$  and  $X$  is the  $(T - p) \times (p + 1)$  matrix containing an intercept and  $p$  lags of the dependent variable.

- Can use derivations from Lecture 2, to say

$$\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$$

where parameters of distribution are OLS quantities

- Also as in Lecture 2, marginal posterior is multivariate t:

$$\beta|y \sim t(\bar{\beta}, \bar{s}^2 \bar{V}, \bar{\nu})$$

- But how to get  $\Pr(\beta \in C|y)$  and  $\Pr(\beta \in D|y)$ ?
- If  $C$  and  $D$  were defined by linear combinations of  $\beta$ , then no problem (“linear combinations of t’s are still t”). But they are nonlinear.

- Monte Carlo integration: randomly draw from  $\beta|y$ .
- For each draw, calculate solutions to  $1 - \sum_{i=1}^p \beta_i z^i = 0$  and see if  $C$  and  $D$  are satisfied
- The proportion of draws which satisfy  $C$  (or  $D$ ) will converge to  $\Pr(\beta \in C|y)$  (or  $\Pr(\beta \in D|y)$ )
- Formally, this follows since  $\Pr(\beta \in C|y) = E[I(\beta \in C|y)]$  and  $\Pr(\beta \in D|y) = E[I(\beta \in D|y)]$ , where  $I(\cdot)$  is the indicator function.
- US real GDP from 1947Q1 through 2005Q2,  $\Pr(\beta \in C|y) = 0.021$  and  $\Pr(\beta \in D|y) = 0.132$ .

### 3 Nonlinear Time Series Extensions of AR model

- Many of these have the form “conditional on knowing threshold/state/regime/breakpoint parameter we have a linear regression model”
- E.g. TAR, smooth transition autoregressive, Markov switching models, many structural break models, etc).
- Gibbs sampler (or similar MCMC algorithms) popular
- We will focus on models in the threshold autoregressive (TAR) class
- For TAR analytical results available if we use natural conjugate prior (no Gibbs sampling required)

### 3.1 Threshold Autoregressive (TAR) Models

- 2 regime TAR:

$$\begin{aligned} y_t &= \beta_{10} + \beta_{11}y_{t-1} + \dots + \beta_{1p}y_{t-p} + \epsilon_t \text{ if } y_{t-1} \leq \tau \\ y_t &= \beta_{20} + \beta_{21}y_{t-1} + \dots + \beta_{2p}y_{t-p} + \epsilon_t \text{ if } y_{t-1} > \tau \end{aligned}$$

- If threshold  $\tau$  is known, model can be written as regression model:

$$y = X\beta + \epsilon,$$

- $X$  has  $t^{th}$  row given by

$$\left[ D_t, D_t y_{t-1}, \dots, D_t y_{t-p}, (1 - D_t), (1 - D_t) y_{t-1}, \dots, (1 - D_t) y_{t-p} \right]$$

- $D_t$  is dummy which equals 1 if  $y_{t-1} \leq \tau$  and equals 0 if  $y_{t-1} > \tau$ .
- Can use all our old results for regression model.
- E.g. if prior is  $NG(\underline{\beta}, \underline{Q}, \underline{s}^{-2}, \underline{\nu})$  then posterior is  $NG(\bar{\beta}, \bar{Q}, \bar{s}^{-2}, \bar{\nu})$  (see Lecture 2 for exact definitions of arguments)
- But what if  $\tau$  is an unknown parameter?
- We need the posterior:  $p(\beta, h, \tau|y)$ .
- Could do Gibbs sampling, but an even simpler strategy is possible.
- Rules of conditional probability imply:

$$p(\beta, h, \tau | y) = p(\beta, h | \tau, y) p(\tau | y).$$

- $p(\beta, h | \tau, y)$  use results for regression model.
- But what about  $p(\tau | y)$ ?
- Remember that the marginal likelihood (see Lecture 1, used in calculating posterior model probability, etc.) is  $p(y)$
- For given  $\tau$ , the marginal likelihood can be calculated for regression model (Lecture 2):

$$p(y | \tau) = c \left( \frac{|\bar{V}|}{|\underline{V}|} \right)^{\frac{1}{2}} (\bar{\nu} s^2)^{-\frac{\bar{\nu}}{2}}$$

- Details of formula unimportant (see Lecture 2 for details). Key thing is that we can calculate  $p(y|\tau)$  for every value of  $\tau$ .

- Bayes' theorem implies:

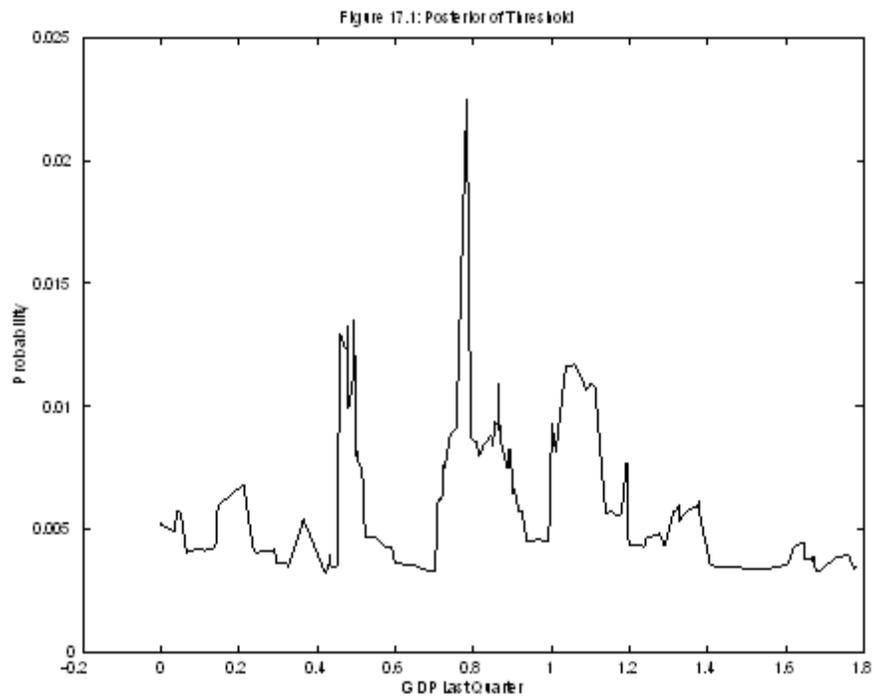
$$p(\tau|y) \propto p(y|\tau)p(\tau),$$

- so combine  $p(y|\tau)$  with a prior for  $\tau$  to get  $p(\tau|y)$ .
- Any prior for  $\tau$  can be used.
- A common choice is a restricted noninformative one.
- E.g. Every value for  $\tau$  which implies each regime contains a minimum number of observations (e.g. 15% of the observations).

- Note: even though  $y_{t-1}$  is a continuous variable,  $\tau$  will be a discrete random variable since there are a finite number of ways of dividing a given data set into two regimes.
- Hence, if  $\tau \in \{\tau_1, \dots, \tau_{T^*}\}$  denotes possible threshold values, then

$$p(\beta, h|y) = \sum_{i=1}^{T^*} p(\beta, h|\tau = \tau_i, y) p(\tau = \tau_i|y).$$

### 3.1.1 Example: US GDP Growth



Posterior Results for TAR Model with Unknown Threshold		
Parameter	Mean	Standard Dev.
$\beta_{10}$	0.542	0.119
$\beta_{11}$	0.269	0.139
$\beta_{12}$	0.373	0.101
$\beta_{20}$	0.249	0.242
$\beta_{21}$	0.384	0.127
$\beta_{22}$	0.196	0.100
$\sigma^2$	0.896	0.083

## 3.2 Extensions of the Basic TAR Model

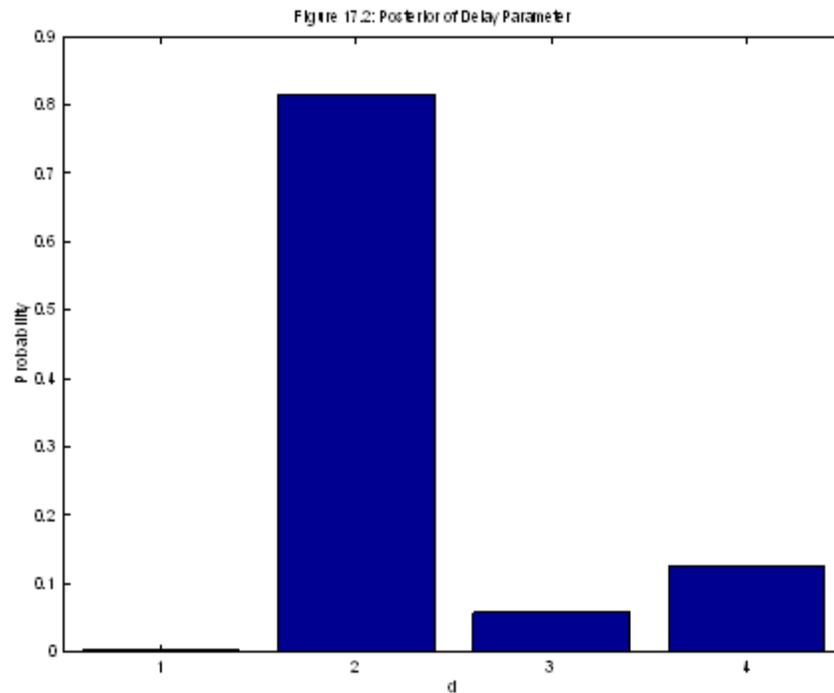
- Many extensions of the TAR can be done in this basic framework
- Basic TAR has last quarter's GDP growth triggering regime switch.
- But it might be another (exogenous or lagged) variable,  $z$ , that is the *threshold trigger*.
- It may take longer than one period to induce the regime switch. Thus, introduce  $d$ : *delay parameter*

$$\begin{aligned} y_t &= \beta_{10} + \beta_{11}y_{t-1} + \dots + \beta_{1p}y_{t-p} + \epsilon_t \text{ if } z_{t-d} \leq \tau \\ y_t &= \beta_{20} + \beta_{21}y_{t-1} + \dots + \beta_{2p}y_{t-p} + \epsilon_t \text{ if } z_{t-d} > \tau \end{aligned} ,$$

- Bayesian inference basically the same as for basic TAR, but now have two unknown parameters:  $\tau$  and  $d$  and so interest centres on:

$$p(\beta, h, \tau, d|y) = p(\beta, h|\tau, d, y) p(\tau, d|y).$$

- Same idea as for basic TAR: evaluate marginal likelihood for every value for  $\tau$  and  $d$  and this can be used to produce  $p(\tau, d|y)$ .
- Extension to more than 2 regimes (with  $\tau_1$  and  $\tau_2$  being two thresholds) same idea (evaluate marginal likelihood for every value for  $\tau_1, \tau_2$  and  $d$  and this can be used to produce  $p(\tau_1, \tau_2, d|y)$ )
- Note: BMA can be used to average over different choices for  $z$  (which may be empirically important).
- Allowing for error variance to differ across regimes may be important (more about this later)



- Example: US real GDP growth data, are provided in Table 17.4.
- Figure plots the posterior of  $d$ .
- Strong support for  $d = 2$  [same as in Potter (1995, JAE)].

## 4 Vector Autoregressive (VAR) Models

- There is a large literature on Bayesian VARs (e.g. work of Sims)
- Lots of different priors tried (priors = shrinkage which seems to help forecasting)
- E.g. Kadiyala and Karlsson (1997, JAE) a good source for different priors.
- Just as AR could be put in form of regression model, VAR can be put in form of multivariate regression model (e.g. like a SUR)
- There is a natural conjugate prior for the VAR model which yields analytical results (comparable to analytical results for AR/TAR done above).

- Bayesian Econometric Methods, Exercise 17.6 does natural conjugate case.
- I will discuss VAR using an independent Normal-Wishart prior (same as we used for SUR model).
- Do this partly to provide a variety of priors, but also because natural conjugate has some restrictive properties in the VAR case.

- VAR( $p$ ) model is:

$$y_t = a_0 + \sum_{j=1}^p A_j y_{t-j} + \varepsilon_t$$

where  $y_t$  for  $t = 1, \dots, T$  is an  $M \times 1$  vector containing observations on  $M$  variables

- $a_0$  is an  $p \times 1$  vector of intercepts and  $A_j$  is an  $M \times M$  matrix of coefficients.
- $\varepsilon_t$  are independent  $N(0, H^{-1})$
- But note (similar to our SUR derivations), we can write the VAR as

$$y_t = Z_t \alpha + \varepsilon_t$$

where  $Z_t$  is an  $M \times m$  matrix of data on explanatory variables (i.e. lags of all dependent variables and an intercept and other deterministic terms) arranged in the same manner as we did for SUR model.

- $\alpha$  is now long vector containing all the VAR coefficients in every equation
- Put in this form, we can do derivations exactly like for SUR model.
- Prior  $p(\alpha, H) = p(\alpha) p(H)$
- Prior for VAR coefficients

$$p(\alpha) = f_N(\alpha | \underline{\alpha}, \underline{V})$$

- Noninformative prior has  $\underline{V}^{-1} = 0$
- Wishart prior for the error precision matrix:

$$H \sim W(\underline{\nu}_H, \underline{H})$$

- Noninformative prior has  $\underline{\nu}_H = 0$  and  $\underline{H}^{-1} = 0$
- As we did with SUR, Gibbs sampler can be set up using  $p(\alpha|y, H)$  and  $p(H|y, \alpha)$ .
- Can show

$$\alpha|y, H \sim N(\bar{\alpha}, \bar{V}),$$

where formula for  $\bar{\alpha}, \bar{V}$  are on page 140 of textbook.

- And the posterior for  $H$  conditional on  $\alpha$  is Wishart:

$$H|y, \alpha \sim W(\bar{\nu}, \bar{H})$$

where formula for  $\bar{\nu}$  and  $\bar{H}$  are on pages 140-141 of textbook.

- So can set up a Gibbs sampler involving drawing from Normal and Wishart distributions.

## 5 Nonlinear VARs

- Many key recent papers use Bayesian methods in nonlinear extensions of VARs
- E.g. Sims and Zha (2006, AER), Primiceri (2005, ReStud) Cogley and Sargent (2001, 2005a, 2005b, various journals), Bernanke, Boivin and Eliasz (2005, QJE)
- To illustrate this literature and as a way of linking many threads together (VARs, state space models, stochastic volatility, etc.) I will present an empirical application based on Primiceri (2005)
- Cogley and Sargent use a similar framework in their work.

## 5.1 Primiceri's TVP-VAR

- Macroeconomic background: evolution of monetary policy
- Extensions of VARs using a few macro variables used.
- Primiceri: US quarterly data on inflation and unemployment rates (“non-policy block”) and interest rates (“policy block”)
- I will not get into macro issues much.
- Suffice it to note: Primiceri has identification restrictions to define “monetary policy shock”, calculates various impulse responses using this (and many other things).

### 5.1.1 Primiceri's Model

- Remember: we wrote the VAR as

$$y_t = Z_t \alpha + \varepsilon_t$$

- $Z_t$  is an  $M \times m$  matrix of data on lags of all dependent variables and an intercept.
- $\varepsilon_t$  are independent  $N(0, \Omega)$
- Note: to be consistent with Primiceri's notation, error covariance matrix is  $\Omega$
- Primiceri extends this in two important ways.

- $\alpha$  becomes  $\alpha_t$  (VAR coefficients can change over time)
- Multivariate stochastic volatility: The error covariance matrix evolves over time.
- Note: Primiceri allows error variances and covariances to evolve over time in a very general way (many other papers more restrictive).
- There is much interest in volatility issues in empirical macro today (“Great Moderation” of business cycle)

## 5.1.2 Evolution of VAR coefficients

- A standard state space model:

$$y_t = Z_t \alpha_t + \varepsilon_t$$

and

$$\alpha_{t+1} = \alpha_t + \eta_t,$$

where  $\alpha_t$  an  $m \times 1$  vector of states and  $\eta_t$  are independent  $N(0, Q)$ .

- Random walk evolution of VAR coefficients
- In Lecture 5 we discussed Bayesian inference (Gibbs sampling) with scalar version of this model (extension to  $M$  dependent variables straightforward)

- I have been using Durbin and Koopman (2002) algorithm for this (but many possible).
- Note: I just found code on James LeSage's Econometrics Toolbox which looks as though it does classical and Bayesian estimation of this model.
- Matlab code tvp.m in <http://www.spatial-econometrics.com/>

### 5.1.3 Multivariate Stochastic Volatility

- Now let  $\Omega$  become  $\Omega_t$ . Many ways to do this.
- Important issue: want error covariances to evolve over time (many specifications do not allow for this).
- Primiceri (2005) uses a triangular reduction  $\Omega_t$ , such that:

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

or

$$\Omega_t = A_t^{-1} \Sigma_t \Sigma_t' (A_t^{-1})'$$

- $\Sigma_t$  is a diagonal matrix with diagonal elements  $\sigma_{j,t}$  (loosely speaking error variances)

- $A_t$  is lower triangular matrix with ones on diagonal (loosely speaking correlations between errors)
- For  $\Sigma_t$  – stochastic volatility.
- to be precise  $h_{i,t} = \ln(\sigma_{i,t})$ ,  $h_t = (h_{1,t}, \dots, h_{p,t})'$  then:

$$h_{t+1} = h_t + u_t,$$

where  $u_t$  is  $N(0, W)$

- Gibbs sampling: draws of  $h = (h'_1, \dots, h'_T)'$  (conditional on  $\alpha$  and the parameters of the model) use algorithm of Kim, Shephard and Chib (1998)
- What about  $A_t$ ?

- Stack into vector as  $a_t = (a_{21,t}, a_{31,t}, a_{32,t}, \dots, a_{p(p-1),t})'$  and use

$$a_t = a_{t-1} + \zeta_t,$$

where  $\zeta_t$  is  $N(0, C)$

- But now we have another state space model
- Durbin and Koopman (2002) algorithm can be used to draw the states (conditional on other model parameters).
- Bottom line: Gibbs sampling algorithm can be set up which draws on off-the-shelf algorithms

### 5.1.4 Extension I am Working On

- Primiceri's model says "all model parameters change every time period" (gradual evolution of coefficients)
- But other structural break models have a small number of more substantive breaks.
- Why not nest these two options?
- Dynamic mixture models (Gerlach, Carter and Kohn, 2000, JASA or Giordani and Kohn, 2006, manuscript).
- Work with Primiceri's model except for the following modifications:

$$\alpha_{t+1} = \alpha_t + K_{1t}\eta_t,$$

$$h_{t+1} = h_t + K_{2t}u_t$$

$$a_t = a_{t-1} + K_{3t}\zeta_t.$$

- $K_{1t}$ ,  $K_{2t}$  and  $K_{3t}$  are 1/0 variables indicating whether a break has/has not taken place
- Hierarchical prior

$$p(K_{jt}) = p_j$$

for  $j = 1, 2, 3$  where  $p_j$  is unknown parameter (probability of break)

- Can use Primiceri's with some extra blocks added to the Gibbs sampler
- Gerlach, Carter and Kohn (2000) is a very efficient algorithm for drawing  $K_{jt}$  (conditional on other model parameters).
- Bottom line: I have set up a Gibbs sampler which combines existing algorithms to carry out posterior inference in this extension of Primiceri's model.
- Much recent Bayesian empirical macro adapts this kind of strategy. Gibbs sampling naturally divides problems into blocks, each block taken from existing literature.

### 5.1.5 Preliminary Empirical Results

- US data on inflation and unemployment rates and interest rates 1947Q1-2006Q3
- With my extension, I am getting results suggesting Primiceri's specification is a good one.
- E.g.  $E(p_1|Data) = 0.92$ ,  $E(p_2|Data) = 0.97$  and  $E(p_3|Data) = 0.62$ .
- Breaks occur in most periods (like in a TVP model).
- Figures 1 through 6 illustrate the kind of things Primiceri presents in his paper (of interest to macro policy)
- Figures 1 through 3 are volatilities

- Figures 4 through 6 are correlations between the errors in the VAR
- These are posterior means (i.e. estimates), but could calculate posterior standard deviations or anything else (e.g. prediction)

Figure 1: Volatility (st dev) in Inflation Equation

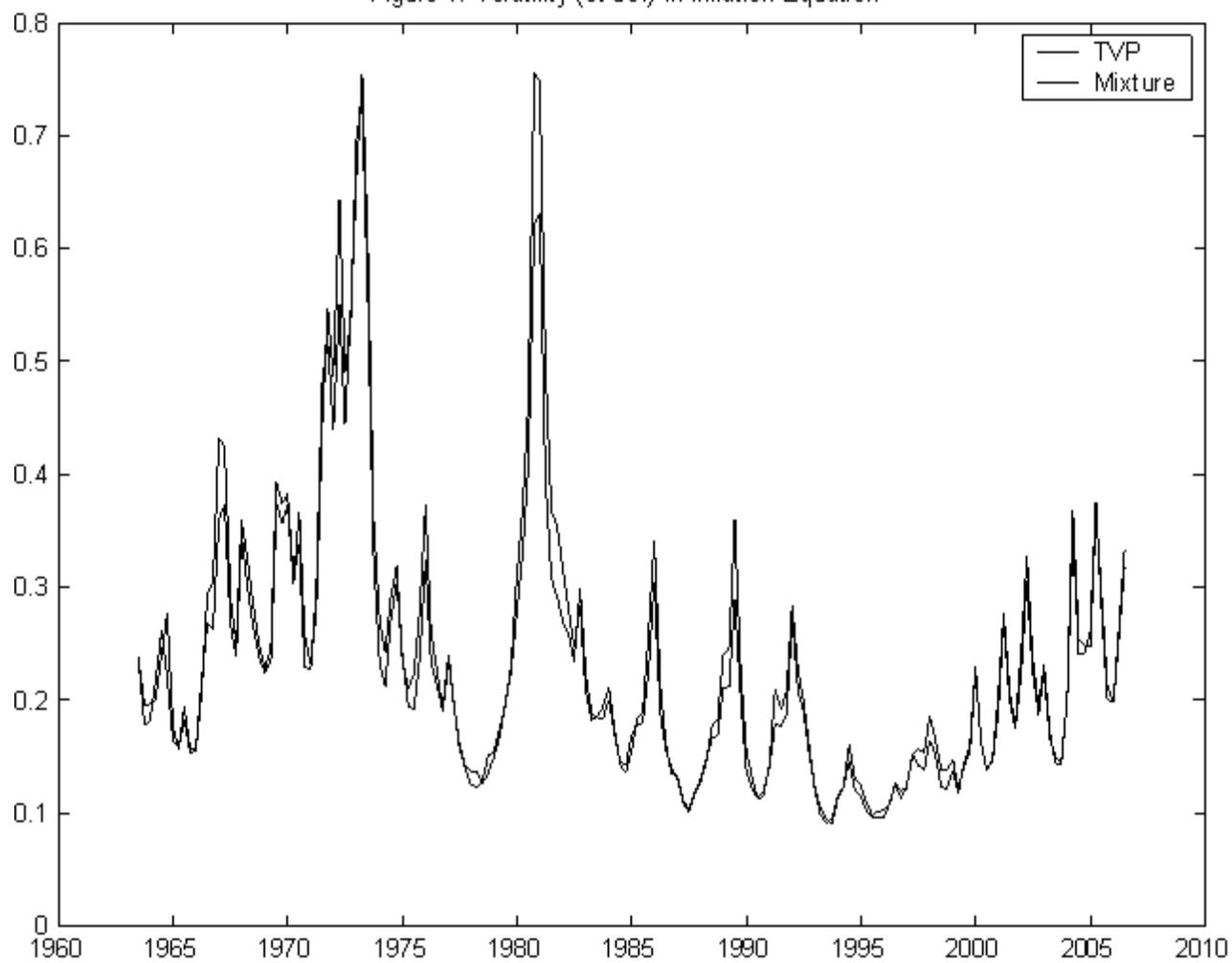


Figure 2: Volatility (st dev) in Unemployment Equation

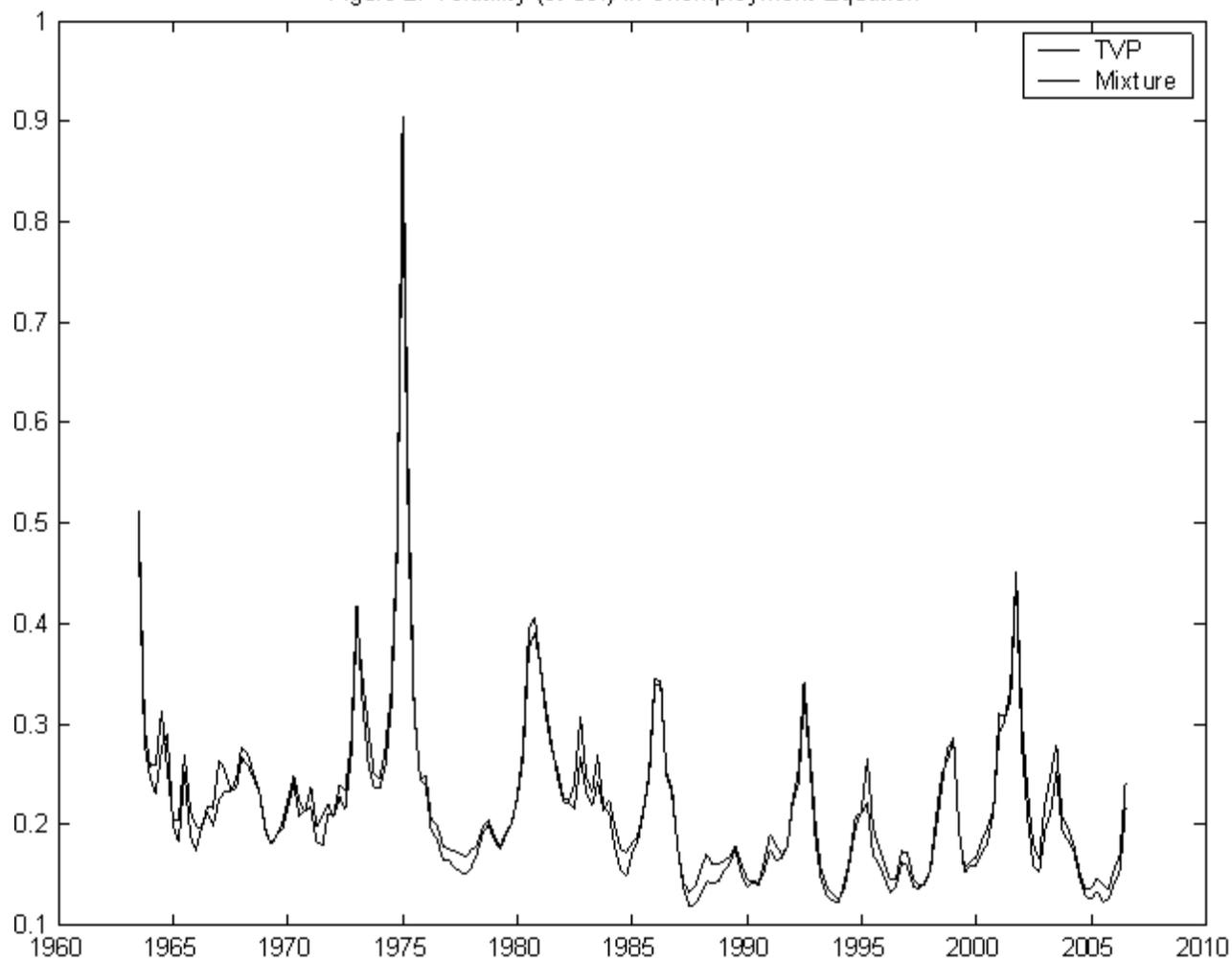


Figure 3: Volatility (st dev) in Interest Rate Equation

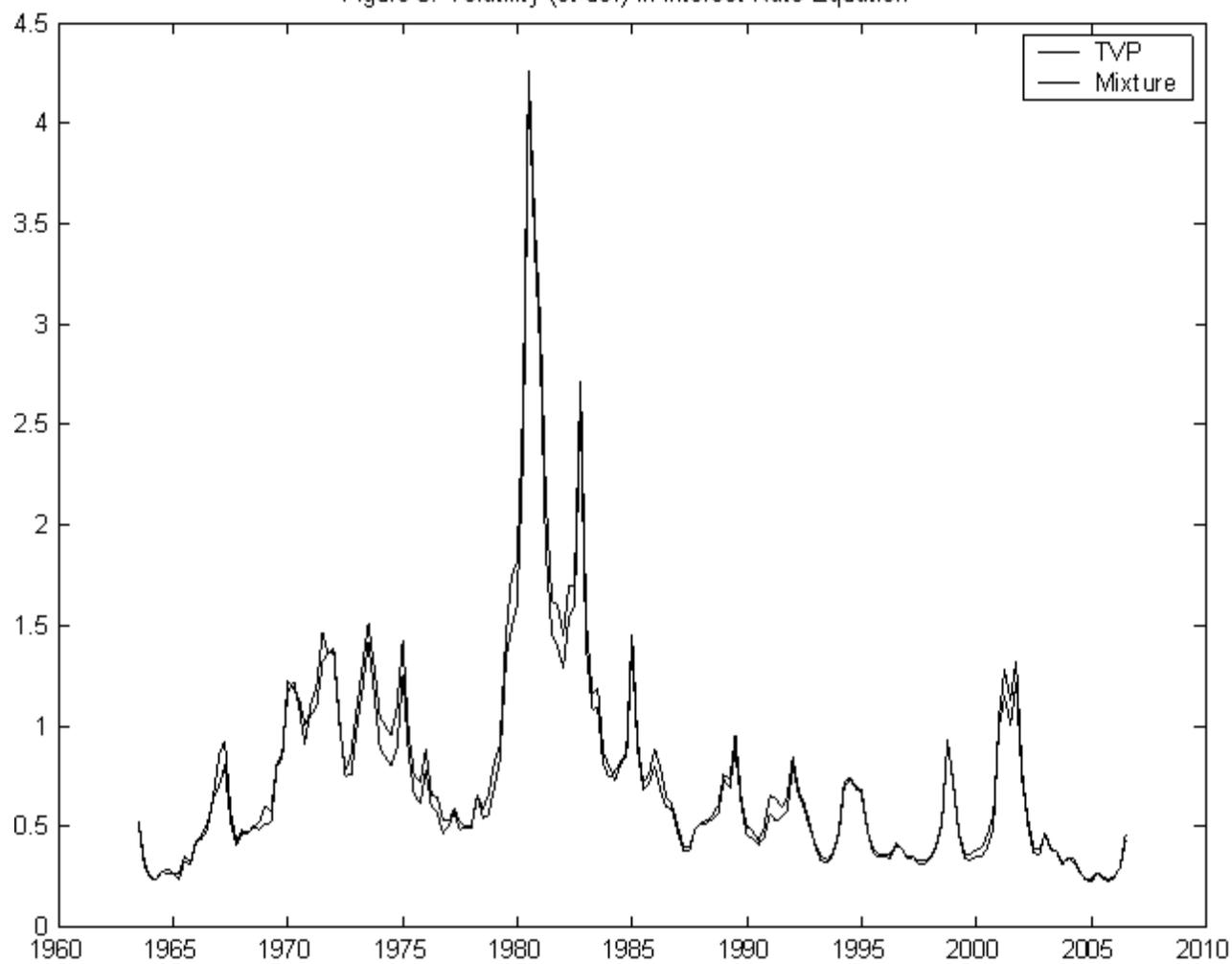


Figure 4: Correlation between Variance in Inflation and unemployment Equation



Figure 5: Correlation between Variance in Inflation and interest rate Equation

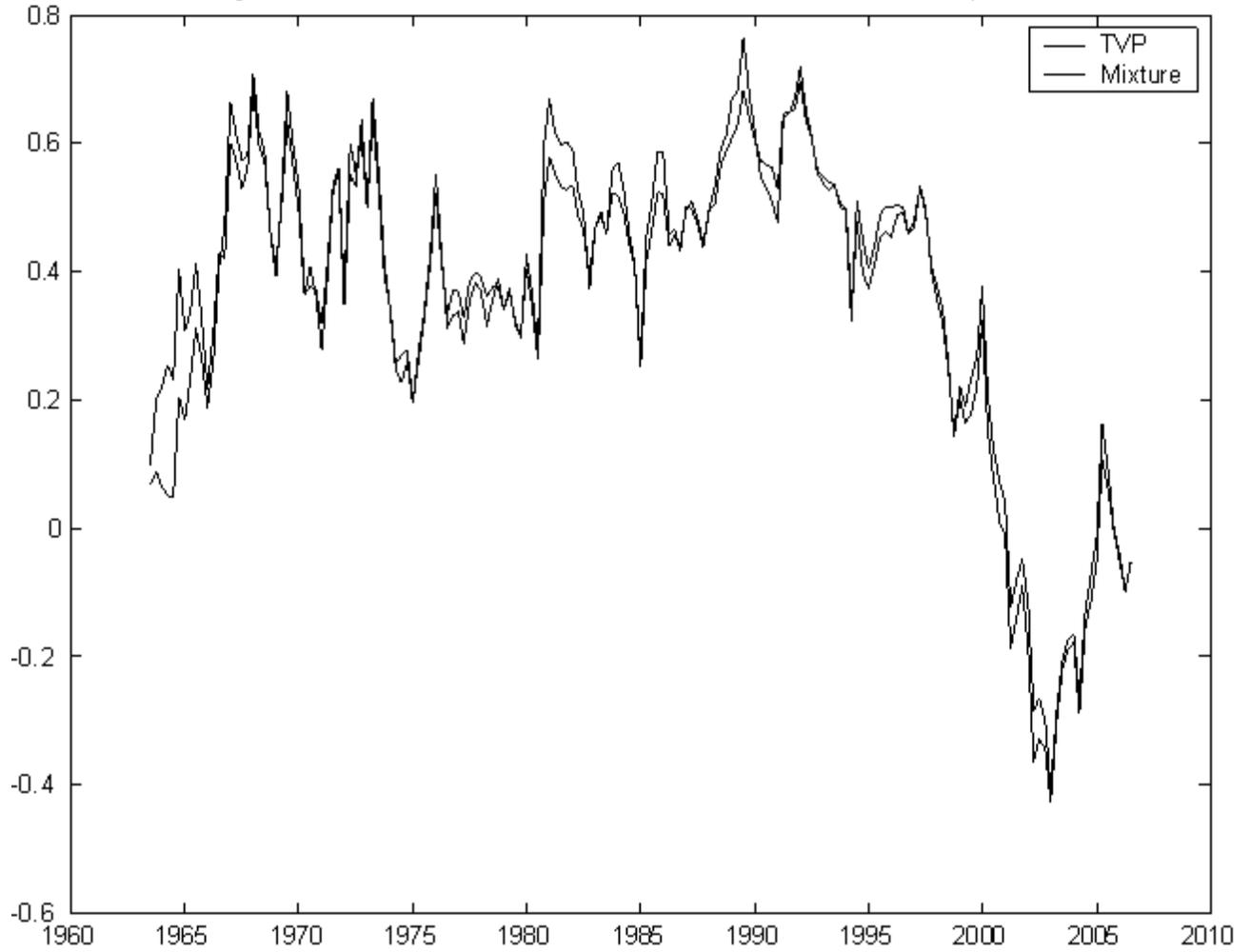


Figure 6: Correlation between Variance in Unemployment and interest rate Equation

