Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions

Richard Startz

Department of Economics

University of Washington

Working Paper No. 56

Center for Statistics in the Social Sciences

University of Washington

October 20, 2006

Acknowledgements: Advice from Peter Hoff, Chang-Jin Kim, Adrian Raftery, Shelly Lundberg, the associate editor, and an anonymous referee is gratefully acknowledged, as is research support from the Cecil and Jane Castor Professorship at the University of Washington. Author address: Department of Economics, University of Washington, Seattle, WA 98195, USA, email:startz@u.washington.edu.

Abstract

Binary Autoregressive Moving Average (BARMA) models provide a modeling technology for binary time series analogous to the classic Gaussian ARMA models used for continuous data. BARMA models mitigate the curse of dimensionality found in long lag Markov models and allow for non-Markovian persistence. The autopersistence function (APF) and autopersistence graph (APG) provide analogs to the autocorrelation function and correlogram. Parameters of the BARMA model may be estimated by either maximum likelihood or MCMC methods. Application of the BARMA model to U.S. recession data suggests that a BARMA(2,2) model is superior to traditional Markov models.

JEL Codes: C220,C250, C110.

1 Introduction

Binary time series are typically modeled in economics as Markov processes, most often as first-order Markov processes. In contrast, for continuous-valued time series the Gaussian autoregressive-moving average model is widely used. This situation persists despite the introduction in the statistics literature of several ARMA models for discrete variables. (An early paper is Jacobs and Lewis (1978). See Benjamin et. al. (2003) for the introduction of GARMA models, as well as a review of the literature.) In this paper I suggest a new practical tool for analysis of binary series: the autopersistence function and autopersistence graph, analogous to the standard autocorrelation function and correlogram. I then turn to remarks on Li's (1994) elegant, but too little used, binary autoregressive moving average model. Parameters of the BARMA model may be estimated by either maximum likelihood or, as I show below, by MCMC methods. These tools are used to analyze quarterly data on U.S. recessions, which are seen to be non-Markovian.

While an obviously valuable tool for the study of binary time series, Markov models suffer from two practical shortcomings. First, they do not fit well when data have strong moving average components. Second, when there are long lags, Markov models face the curse of dimensionality. While the models discussed below can include Markov models as special cases, they would typically be more general in that they add in a moving average component. These models also provide a convenient way to place restrictions on the pure Markov models so as to eliminate the curse of dimensionality.

The study of Gaussian ARMA models traditionally starts with an identification step in which the correlogram is examined to suggest a model whose ARMA representation is estimated in the next step. Correlation is a natural metric for a Gaussian series, but much less so for a binary series. Obversely, looking at a conditional probability is more natural for a binary than for a continuous series, since a binary series has only two discrete values on which it is necessary to condition. After defining tools for the identification step for binary data, I apply them to U.S. recessions. The Binary Autoregressive (BAR) model is discussed as a way to connect BARMA models with Markov models. After discussing several practical difficulties, the full BARMA model is then applied to the recession data.

2 The Autopersistence Function

The autocorrelation function and correlogram provide useful information about the behavior of theoretical and empirical continuous time series. Analogously, the autopersistence function and autopersistence graph provide useful information about the behavior of theoretical and empirical binary time series.

2.1 ACF, Correlogram, APF, and APG

For a stationary, ergodic, *Gaussian* $ARMA(p,q)$ process, the joint distribution of the observations is completely described by the autocorrelation function, ACF, (together with the unconditional mean and variance). Similarly, an observed series is described by its correlogram. The ACF and correlogram are useful for continuous data even when the time series is not Gaussian. The shape of the correlogram sometimes provides a hint as to the order of the underlying ARMA process, while $ACF(k)$ is informative about how quickly information in the current observation fades in a given theoretical process. The ACF or correlogram provides the information necessary for making a $k-$ahead linear forecast from the current observation on the series. For a first-order autoregressive series the $AR(1)$ parameter is estimated by $ACF(1)$ and the shape of the ACF follows a familiar geometric decline asymptoting to zero.

Looking at correlations is less useful as a summary statistic for a binary series than it is for continuous series. However, looking at $k-$ahead conditional probabilities is useful and is feasible since one need only condition on two values rather than on a continuum.

For an ergodic, binary time series y, where w.o.l.g. y takes the values 0 and 1, the appropriate analog to the ACF is the pair of *autopersistence functions*, $APF^0\left(k\right) \equiv \Pr\left(y_{t+k}=1\,|y_t=0\right)$

and $APF^1(k) \equiv \Pr(y_{t+k} = 1 \,|y_t = 1)$. The *autopersistence graphs* APG^0 and APG^1 are, by analogy to the correlogram, the empirical counterparts to the APF and may be estimated by the appropriate sample conditional means. While the APF does not completely describe the joint distribution of an ergodic series (nor does the ACF for a continuous series except in the Gaussian case), the shape of the APG may provide a hint about the order of an appropriate BARMA process. $APF(k)$ is informative about how quickly information in the current observation fades. The APF or APG provides the information necessary for making a $k-$ahead forecast from the current observation on the series (although the APF is more limited than the Gaussian ACF in that the APF can be used to forecast conditional only on the current observation, where the in the Gaussian case the ACF can be used to predict conditionally on any set of lags). For a first-order Markov process the two transition probabilities are estimated by $APF^0(1)$ and $APF^1(1)$, and the shape of the APF follows a familiar geometric decline asymptoting to the unconditional mean.

2.2 U.S. Recession Data

Figure 1 shows the APG for U.S. recessions. The oscillating nature of the APG, being very unlike a geometric decline, suggests that a first-order Markov is not a good model for this data. With this as a motivating example, we begin with theory and then return to an empirical examination of recession data in Section 8.

3 Binary Autoregressive Models

For what follows, it is useful to recast the p^{th}-order Markov model in an autoregressive framework. The Binary Autoregressive with Cross-terms model of order p, $BARX(p)$, suggested

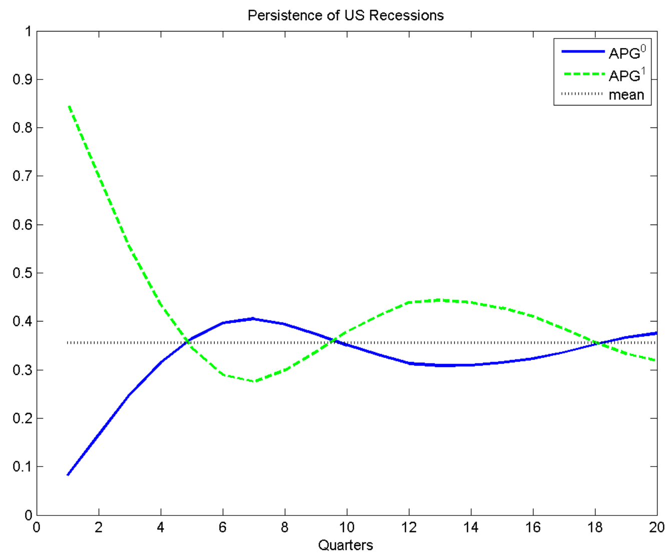


Figure 1:

by Zeger and Qaqish (1988) can be written

$$
\begin{aligned}
\eta_t &= \beta_0 + I_{p\geq 1}\sum_{i=1}^{p}\phi_i y_{t-i} + I_{p\geq 2}\sum_{i=1}^{p}\sum_{j=i+1}^{p}\phi_{i,j}\left(y_{t-i}\times y_{t-j}\right) \\
&\quad + \ldots + \sum_{i_1=1}^{p}\sum_{i_2=i_1+1}^{p}\cdots\sum_{i_p=i_{p-1}+1}^{p}\left(\phi_{i_1,i_2,\ldots i_p}\prod_{k=i_1}^{i_p} y_{t-k}\right) \qquad (1) \\
\mu_t &= \frac{e^{\eta_t}}{1+e^{\eta_t}} \\
\Pr\left(y_t\left|\mu_t, y_{t-1}, y_{t-2}, \ldots y_{t-p}\right.\right) &= \mu_t
\end{aligned}
$$

where $I_{p\geq i}$ is the indicator function. In other words, the model includes all the unique lags and lagged cross-terms through lag p. The log-likelihood equals

$$
\pounds = \sum_{t=1}^{T}\left(y_t \log \mu_t + (1-y_t)\log\left(1-\mu_t\right)\right) \qquad (2)
$$

The $BARX(p)$ model is an alternative representation of the p^{th}-order Markov model and one can map back and forth between the parameters of the BARX and the Markov representations. The BARX approach has two minor disadvantages relative to the familiar

$Markov(p)$ representation: transition probabilities on the edge of the parameter space, 0 or 1, require infinite values for the BARX parameters, and the interpretation of β_0 and $\overrightarrow{\phi}$ is less familiar than direct statement of the transition probabilities. The advantage of the BARX representation is that it provides a natural starting point for moving away from unrestricted Markov models.

The difficulty with application of the p^{th}-order Markov model is that it requires 2^p parameters to capture p lags of behavior, which is impractical for even modest sizes of p. As a remedy, Raftery (1985) suggested the MTD model to impose linear restrictions on the Markov transition probabilities to reduce the size of the parameter space from 2^p to p. Similarly, the Binary Autoregressive model of order p (first suggested by Cox (1981)), $BAR(p)$, imposes linear restrictions on the $BARX(p)$ model in the form of zero restrictions on cross-terms, substituting

$$\eta_t = \beta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \tag{3}$$

In the BAR model, restrictions are linear in logits of the transition probabilities rather than in the transition probabilities themselves. Models intermediate between $BAR(p)$ and $BARX(p)$ may be specified in a natural way, for example by including cross-pairs but not cross-triples or higher.

Use of the logit link function, $\mu_t = \frac{e^{\eta_t}}{1+e^{\eta_t}}$, is convenient but a different link function could also be used. For example, a standard normal CDF would lead to a probit-based model. Eichengreen et. al. (1985) present a dynamic ordered-probit model for trinary rather than binary outcomes with a somewhat different stochastic specification. de Jong and Woutersen (2005) examine the asymptotic properties of estimates of related models.

4 Binary Autoregressive-Moving Average Models

Markov models do not give an adequate representation of the persistence of recessions. The APG in Figure 1 crosses the unconditional mean approximately one year out, and then shows damped, but considerable, oscillation. While a second-order Markov model could in principle produce oscillations in the APF, that does not happen at our estimated parameters (see below). This suggests considering non-Markovian models.

Li (1994) suggested formulating the $BARMA\,(p,q)$ model as

$$\eta_t = \beta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \left(y_{t-i} - \mu_{t-i}\right) \tag{4}$$

where $y_{t-i} - \mu_{t-i}$ plays a role analogous to the innovation in a continuous ARMA model. The BARMA model can be extended by adding cross-terms as in the BARX model above and/or by Li's suggestion of replacing β_0 with a covariates term $X_t\beta$. The moving average component is in the class described by Cox (1981) as observation-driven. Note that the analogy with the continuous ARMA model is not perfect, because it is the memory of prediction errors $(y_{t-i} - E\,(y_{t-i}))$ rather than shocks that is carried forward.

The $BARMA\,(p,q)$ model can be estimated by quasi-maximum likelihood. Li suggests setting the initial q values of μ_t to zero or to the sample mean of y. One could also set initial values of μ_t to 0.5 or initial values of $y_t - \mu_t$ to zero. (The estimates in this paper use the sample mean of y for initial values of μ_t.)

5 Practical Considerations

We turn now to some practical considerations in use of the BARMA model, as illustrated with our recession data.

5.1 Three practical considerations for the BAR model

Parameterization of the BAR in terms of logits on transition probabilities raises three practical considerations, each of which arises with our sample data. The first issue is what happens when the estimated parameters are on the edge of the permissible space. For our recession data, the transition probabilities for a second-order Markov are $pr\left(y_t = 1 \,|y_{t-2}, y_{t-1}\right) = \begin{bmatrix} 0.090141 & 0 \\ 1 & 0.824176 \end{bmatrix}$, so two of the four parameters take limiting values. The $BARX(2)$ representation is $\beta_0 = -2.49268$, $\phi_1 = \infty$, $\phi_2 = -\infty$, $\phi_{12} = 4.03758$. Estimation of the $BARX(2)$ represents no difficulty, as large values of ϕ are effectively infinite, so long as the user remembers that $\phi = 25$ and $\phi = 2500$ mean the same thing. Likelihood values are computed correctly.

The second practical issue that can arise is dealing with empty cell counts. For example, despite having 600 observations for our recession data, the eight sample transition probabilities for a third-order Markov process produce two empty cells (plus, as it happens, four cells with 0 or 1 probabilities and two interior values.) Therefore, some parameters in the $Markov(3)/BARX(3)$ representation are unidentified. Although this does not prevent calculation of the likelihood function, use of a likelihood value based on unidentified parameters for testing may be unwise. Further, having unidentified parameters is problematic in analysis and simulation of the estimated process, since these require values for the cells that were unobserved in the sample. The linear restrictions implicit in the BAR model reduce the information required for parameter identification so that the BAR model is generally unaffected by the presence of empty cells. As it happens, our data has a $BAR(3)$ representation with an identical likelihood value as the third-order Markov.

5.1.1 BAR 2^{nd} partials

The third practical consideration regards both calculation of Wald statistics and the behavior of search algorithms when parameters are on the edge of the permissible space. Both issues require looking at the second partials of the log-likelihood function, computation of which can be problematic. For the $BAR\,(p)$ model, the observation-by-observation contributions to the second partials for β_0, ϕ_1, and ϕ_2 are

$$-\mu_t\,(1-\mu_t)\begin{bmatrix} 1 & y_{t-1} & y_{t-2} \\ & y_{t-1}^2 & y_{t-1}y_{t-2} \\ & & y_{t-2}^2 \end{bmatrix} \tag{5}$$

Consider what happens when ϕ_1 is large. When $y_{t-1} = 1$, then $\mu_t \approx 1$ and $(1-\mu_t) \approx 0$ (except possibly in special cases where offsetting values of ϕ_i, ϕ_j lead to interior values of μ_t). Since $(1-\mu_t) \approx 0$, the contribution to the second diagonal element in the second-partials matrix equals zero. When $y_{t-1} = 0$, the second diagonal element equals zero as well. As a result, the estimated information matrix is singular. It follows that the traditional estimates of the variance-covariance matrix of the parameters is unavailable, as are the associated Wald tests.

Having a singular information matrix when the maximum likelihood estimate of ϕ_1 is large may be regarded as a desirable feature. Since the mle parameter estimates do not follow the standard distributions at the edge of the permissible parameter space, variance estimates and Wald tests may well be misleading. However, the estimated log-likelihood is also flat for extreme values of ϕ examined in the search process even though these values are very far from the mle. As a result, standard search algorithms which rely on second partials can become "stuck" in areas of the parameter space far from the optimum. Modification of such algorithms or manual intervention in the search process may be needed.[1] Alternatively, use of the Gibbs sampler proposed in section 7 avoids numerical problems with the likelihood function entirely.

[1] Choice of an appropriate intervention depends on whether large ϕ_1 is found in the vicinity of a local or a global maximum of the likelihood function. In the former case, the search algorithm needs to be restarted elsewhere. When ϕ_1 is close to a global maximum this suggests that an arbitrary scaling can be applied to the parameter set, so ϕ_1 should be set to large constant (e.g., $\phi_1 = 100$) and the search should continue for the remaining parameters. (I am grateful to the associate editor for this suggestion.)

5.2 Practical analysis of the BARMA and BMA models

Unlike the Gaussian ARMA model, the BARMA model is inherently nonlinear, and does not directly translate between AR and MA representations. Because of the logit link, there are no pleasant analytic solutions for the APF, autocorrelations, or even the unconditional mean.

While a $BAR(p)$ model always has a $p^{th}-$order Markov representation, for which there are a variety of tools available, a model with a BMA component does not. Fortunately, given the recursive nature of the BARMA specification the APF, etc., can be drawn by straightforward numerical simulation, starting at arbitrary initial values, discarding the first few draws, and then using simulation sample averages for the desired statistic.[2]

Interpretation of magnitudes for BARMA coefficients is less neat than for Gaussian ARMA models, but some examples provide intuition. A BARMA coefficient gives the change in the log odds ratio when the corresponding data lag equals 1 rather than 0. For a $BAR(1)$ with $\beta_0 = 0$, for example, observing $y_{t-1} = 0$ implies $\mu = 0.5$, while observing $y_{t-1} = 1$ implies $\mu = .73$ for $\phi_1 = 1.0$ and $\mu = .9$ for $\phi_1 = 2.2$. (As additional examples, $\mu = .95$ for $\phi_1 = 2.95$ and $\mu = .99$ for $\phi_1 = 4.60$.) For $\beta_0 = -2.2$, $\phi_1 = 4.4$ switches the conditional mean from 0.1 to 0.9. For $\beta_0 = 2.2$, $\phi_1 = 4.4$ switches the conditional mean from 0.9 to 0.999. This suggests as a rule of thumb that BARMA coefficients above 1 are "large" and coefficients in the high single digits are very large.

The APF for a $BMA(q)$ model returns to the unconditional mean (and the autocorrelation function goes to zero after q lags)–almost. Because of the curvature of the logit function, the $APF(k)$ for $k > q$ can differ very slightly from the unconditional mean. Consider Li's simulation of a $BMA(1)$ with parameters $\beta_0 = 1$ and $\theta_1 = 0.8$, for which he states "insignificant autocorrelations after lag one...are typical." The left hand panel of Figure 2 shows the

[2] Such a simulation assumes that the process does not have an absorbing state. In economics this is not an issue as the usual assumption is that we are sampling from a time series process with a very long history, implying that if the process has an absorbing state our entire sample will be in the absorbing state. In other areas the issue of an absorbing state may be more problematic.

APF and autocorrelation function (for 2,000 simulations) for Li's parameters, confirming his claims. As a contrast, the right hand panel shows the APF and autocorrelation function for $\beta_0 = -2.2$ and $\theta_1 = 4.4$. The unconditional expectation of y is 0.136. The first two values of APF^1 are 0.539 and 0.160; so $APF^1(2)$ is measurably above the unconditional expectation. Similarly $ACF(2) = 0.028$, which is not quite zero. Thus, while pure BMA models do not formally have the same finite autocorrelation function property found for Gaussian models, the deviation is so small as to be unlikely to have much practical consequence.

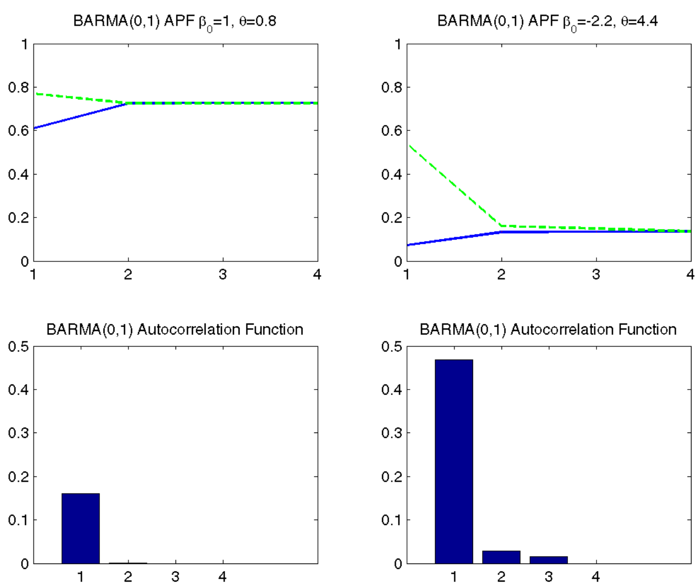


Figure 2:

6 Goodness of Fit Measures

Comparison of a model's APF with the empirical APG is one way to evaluate model adequacy. Scalar goodness of fit measures are also useful. One obvious measure is McFadden's (1974) $R_M^2 \equiv 1 - \frac{\widehat{\pounds}}{\pounds^0}$, where $\widehat{\pounds}$ is the maximized model log-likelihood and $\pounds^0$ is the restricted log-likelihood from maximizing $\eta_t = \beta_0$, in other words from simply using the sample mean. Another measure is the *predictive* $R_p^2 \equiv 1 - \frac{\sum(y_t - \mu_t)^2}{\sum(y_t - \bar{\mu})^2}$, where $\bar{\mu}$ is the sample mean, due to

Efron (1978). If one's interest is forecasting and one has a mean square error loss function, then R_p^2 is the appropriate in-sample goodness of fit measure.

7 Gibbs Sampling

Estimation of the BARMA model by Gibbs sampling makes available the set of tools associated with MCMC methods. Additionally Gibbs sampling avoids the computational issues described above. The approach here is similar to Gibbs sampling for probits. (See Albert and Chib (1993) or the expository presentation in Koop (2003).) The model is augmented with a latent variable η^*, and sampling proceeds in three blocks. In the first block, η^* is effectively regressed on the right-hand side of the BARMA model to draw the BARMA parameters. Here, a diffuse prior is assumed. Any prior applicable to a regression could be used. In the second block, values of μ^* are drawn (conditionally deterministically) by evaluating the BARMA model. In the third block, the latent η^* are drawn from truncated logits.

To motivate the latent variable model, assume that nature draws $z_t \sim \text{uniform}\,(0,1)$ and $y_t = 1$ *iff* $\mu_t > z_t$. This is equivalent to $g^{-1}(\mu_t) > g^{-1}(z_t)$, where $g = \frac{e^{\eta_t}}{1+e^{\eta_t}}$. Define the latent variable $\eta_t^* = g^{-1}(\mu_t) - g^{-1}(z_t)$, so that $y_t = 1$ *iff* $\eta_t^* > 0$. We can then rewrite, the BARMA equation as a linear regression

$$\eta_t^* = \beta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \left(y_{t-i} - \mu_{t-i}^*\right) - g^{-1}(z_t) \tag{6}$$

The Gibbs sampler consists of an initialization block followed by iteration between drawing regression coefficients, drawing μ^*, and drawing η^*.

7.1 Initialization

$$
\begin{aligned}
\eta_t^* &= y_t, \text{ for } t > \max(p,q) \\
\eta_t^* &= mean(y), \text{ for } t \leq \max(p,q) \\
\mu_t^* &\sim \text{uniform}(0,1), \text{ for } t > \max(p,q) \\
\mu_t^* &= mean(y), \text{ for } t \leq \max(p,q) \\
\bar{\eta}_t &= -\log \frac{1-\mu_t}{\mu_t}, \text{ for } t \leq \max(p,q)
\end{aligned} \tag{7}
$$

7.2 Draw BARMA parameters

Discarding the first $\max(p,q)$ observations, create X where the first column equals 1.0, followed by p columns of lags of y, followed by q columns of $y - \mu^*$.

$$
\begin{aligned}
\tilde{b} &\sim N\left((X'X)^{-1} X'\eta^*, \frac{\pi^2}{3}(X'X)^{-1}\right) \\
\beta_0 &= \tilde{b}_0 \\
\phi &= \tilde{b}_{1\ldots p} \\
\theta &= \tilde{b}_{p+1\ldots q}
\end{aligned} \tag{8}
$$

We treat the posterior for the regression parameters as multivariate normal, even though the errors are logistic rather than normal.[3]

Note that equation (8) assumes a diffuse prior for the BARMA parameters. Since the variance of the logistic distribution is $\frac{\pi^2}{3}$, no priors are needed for the regression variance.

[3] I am grateful to the anonymous referee for pointing out that this can be regarded as a draw from a proposal density in a Metropolis-Hastings step.

7.3 Draw μ^*

Starting at $t = 1 + \max(p, q)$ draw $\bar{\eta}_t = \beta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \left(y_{t-i} - \mu_{t-i}^*\right)$, compute $\mu_t^* = \frac{e^{\bar{\eta}_t}}{1+e^{\bar{\eta}_t}}$, and proceed iteratively. Because of the "observation-driven" nature of the BARMA model, this draw is, conditional on the most recent draw of the BARMA parameters, deterministic.

7.4 Draw latent η^*

Let $F^R(\bar{\eta})$ be the logistic distribution with mean $\bar{\eta}$ right-truncated at zero and let $F^L(\bar{\eta})$ be the corresponding left-truncated distribution. Draw the latent η_t^* according to

$$\begin{aligned} \eta_t^* &\sim F^R(\bar{\eta}_t), \text{ if } y_t = 0 \\ \eta_t^* &\sim F^L(\bar{\eta}_t), \text{ if } y_t = 1 \end{aligned} \tag{9}$$

The regression draw, calculation of μ^*, and latent draw blocks are repeated until a sufficient size sample is collected.

8 Application to U.S. Recessions

Recessions in the United States are identified by the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER). "A recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales."[4] The NBER identifies 32 recessions since 1854, the shortest being 6 months in length and the longest being 65 months.

Figure 3 shows the 602 quarterly observations given in the NBER recession chronology, with recession periods shown by shaded areas. In addition, horizontal lines are drawn show-

[4]Business Cycle Dating Committee, NBER, October 21, 2003 statement.

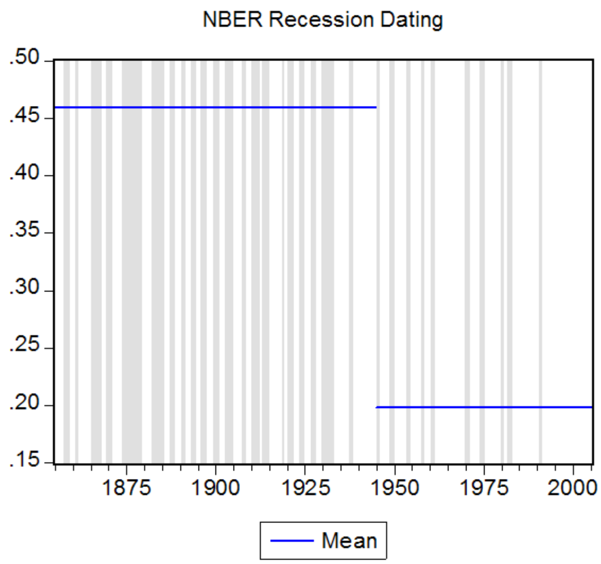


Figure 3:

ing the mean probability of the United States being in recession through 1945, and then, separately, in the post-War period. While the NBER dates recessions on a monthly basis, quarterly data is used here. (By convention, a quarter is coded with a 1 for recession is any month in the quarter is identified by the NBER as being in a recession.) This is done for two reasons. First, as most national income accounting data is quarterly, particularly GDP, much statistical modeling of recessions is quarterly. Second, the notion that recessions last "more than a few months" is typically interpreted as a six month minimum, so a monthly series is necessarily difficult to fit well with a low order Markov model.

Starting from a standard 1^{st}-order Markov model as a benchmark, we see what extra light can be shed by turning to BAR and BARMA models of recessions.

8.1 Autoregressive Recession Models

The natural starting point for analysis of the time series of U.S. recessions is with a Markov model. Table 1 shows the transition probabilities for $y_t = 1$ conditional on lagged y for Markov models of order 0 through 3.

$P\left(y_t \mid (y_{t-1}, y_{t-2}, y_{t-3})\right)$	$(0,0,0)$	$(0,0,1)$	$(0,1,0)$	$(0,1,1)$	$(1,0,0)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$	$logL$	R_M^2	R_p^2
3^{rd}-order Markov	0.0994	0	NA	0	1.0	NA	1.0	0.7867	-181.99	.53	.61
2^{nd}-order Markov	0.0901		0		1.0		0.8242		-192.05	.51	.60
1^{st}-order Markov	0.0829				0.8505				-200.60	.49	.59
mean	0.3567								-390.89	0	0

Markov Models of U.S. Recessions - Table 1

The 1^{st}-order Markov order model is clearly preferred to a constant mean. The 2^{nd}-order Markov model has a much higher log-likelihood than does the 1^{st}-order Markov. Note that two of the parameters in the 2^{nd}-order model are on the edge of the parameter space. The 3^{rd}-order Markov model has a yet higher log-likelihood. Note that four of the eight parameters are 0 or 1 and, more problematically, two of the parameters are not identified.

Moving from low-order to higher-order Markov models improves the log-likelihood function. However, neither of the R^2 goodness of fit measures is very much improved. What's more, the APF for the 1^{st} and 2^{nd}-order models are quite similar to one another (Figure 4) and not at all like the empirical APG shown in Figure 1.[5][6]

The results for the Markov models hint that longer lags matter, but that 600 observations is insufficient to estimate a high-order Markov model. Figure 5 shows the APFs for $BAR(1)$, $BAR(2)$, and $BAR(3)$ models. $BAR(1)$ and 1^{st}-order Markov models are necessarily the same. Coincidentally, the four parameter 2^{nd}-order Markov can be represented exactly by a three parameter $BAR(2)$ ($\beta_0 = -2.31$, $\phi_1 = \alpha$, $\phi_2 = -\alpha + 3.86$, for any very large value of α.), so for this data the two are equivalent. Serendipitously (all its parameters are identified), the four parameter $BAR(3)$ ($\beta_0 = -2.20$, $\phi_1 = \alpha$, $\phi_2 = 0$, $\phi_3 = -\alpha + 3.51$, for large α.) has the same log likelihood value as the eight parameter 3^{rd}-order Markov model. The $BAR(3)$

[5]The APF for the 3^{rd}-order Markov model cannot be calculated since some of the parameters are unknown.

[6]Figures 4-7 and 9 show APF^1 with a dashed line, APF^0 with a solid line, and the steady-state mean with a dotted line.

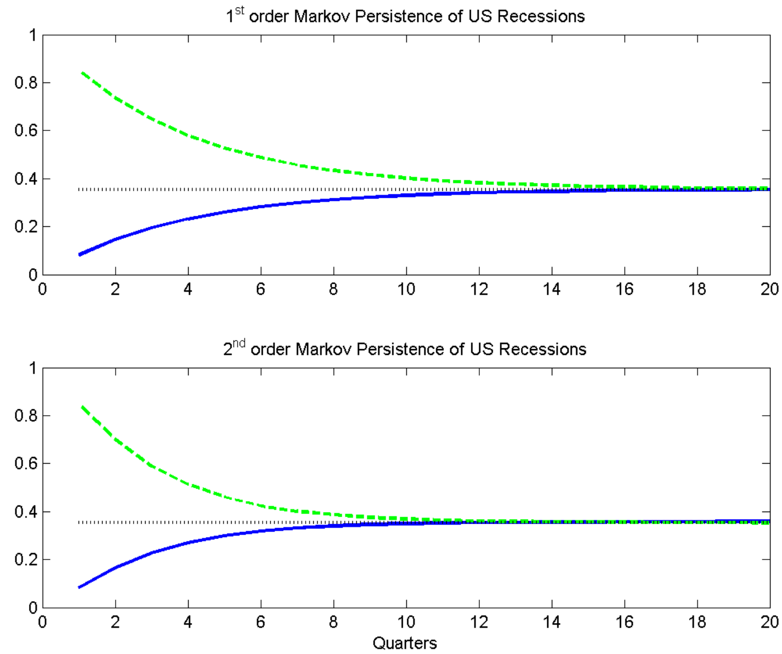


Figure 4:

APF returns to the unconditional mean somewhat faster than does the lower-order models and shows a shade of the APF crossing property that is prominent in the empirical APG.

8.2 BARMA Recession Models

Can we find a parsimonious BARMA model for understanding recessions which improves on the Markov models? Since it is clear from the APG that some autoregressive component exists, pure BMA models are not useful candidates. We present three low-order BARMA models here, as shown in Table 2. The log-likelihood of the $BARMA(1,1)$ model is noticeably larger than the log-likelihood of the nested $BARMA(1,0)$ model, i.e. the 1^{st}-order Markov order model shown in Table 1. The same is true in comparing $BARMA(2,1)$ to $BARMA(2,0)$. Figure 6 displays two APFs. The $BARMA(1,1)$ APF looks pretty much like the $BARMA(1,0)$ APF. However, while R^2_M shows little difference between the BAR and BARMA models, R^2_p is notably lower for the BARMA models.

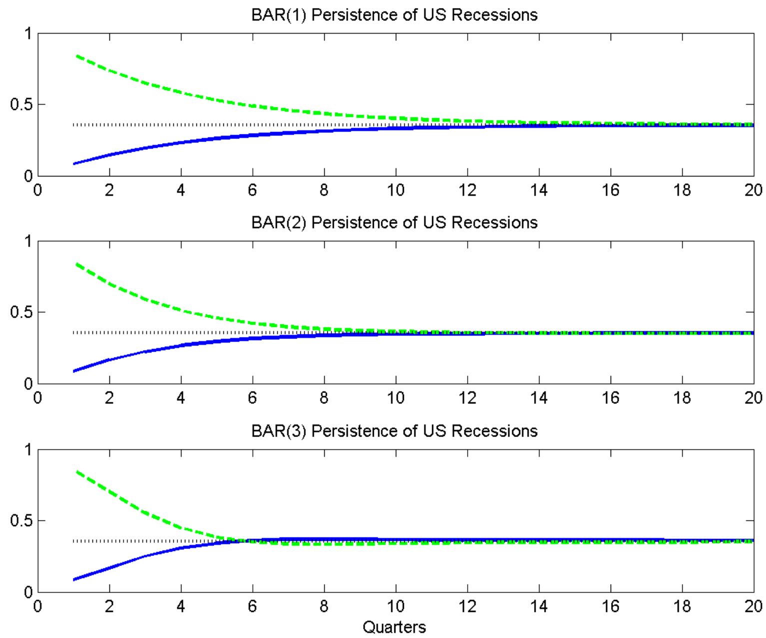


Figure 5:

	β_0	ϕ_1	ϕ_2	θ_1	θ_2	$logL$	R^2_M	R^2_p
$BARMA(2,2)$	-2.62	42.54	-37.58	-9.53	5.20	-180.78	.54	.61
$BARMA(2,1)$	-2.67	18.84	-13.95	-4.51		-187.53	.52	.27
$BARMA(1,1)$	-2.183	3.53		2.13		-195.94	.50	.24

BARMA Models of U.S. Recessions - Table 2

The $BARMA(2,2)$ model has essentially the same likelihood value, R^2_M, and R^2_p as the 3^{rd}-order Markov/$BAR(3)$ model. Figure 7 shows the $BARMA(2,2)$ APF next to the empirical APG. The match is closer than for earlier models. Based on the slightly higher likelihood value for the $BARMA(2,2)$ model and the better match of the APF to APG, the BARMA model is clearly preferred to an unrestricted Markov model, and arguably to the restricted $BAR(3)$ as well.

It is unknown whether methods of lag length determination employed for Gaussian ARMA models work well for BARMA models. For the record, the likelihood ratio statistic

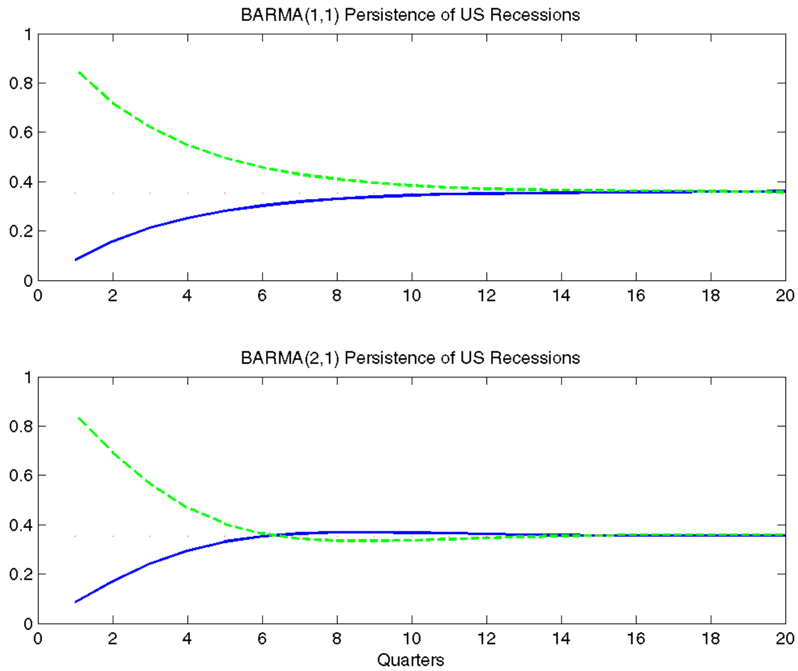


Figure 6:

for a $BARMA(2,2)$ versus $BARMA(2,4)$ equals 4.97, which corresponds to a p-value of 0.08.

8.3 Gibbs Sampling

A $BARMA(2,2)$ model was estimated by Gibbs sampling, discarding 1,000 draws and retaining 10,000. Note that the maximum likelihood estimates in Table 2 show effectively infinite values for both the BAR coefficients with $\phi_1 + \phi_2 \approx 5.0$. Figure 8 presents posterior medians and histograms. The results of the Gibbs sampler are quite close to the mle results. The BAR coefficients are effectively infinite. Note that the median of $\phi_1 + \phi_2$ is close to the mle estimate and clearly positive. The distribution of θ_1 is almost entirely to the left of zero and the distribution of θ_2 is almost entirely to the right. The posterior for $\theta_1 + \theta_2$ has greater spread and the median is somewhat farther from the mle.

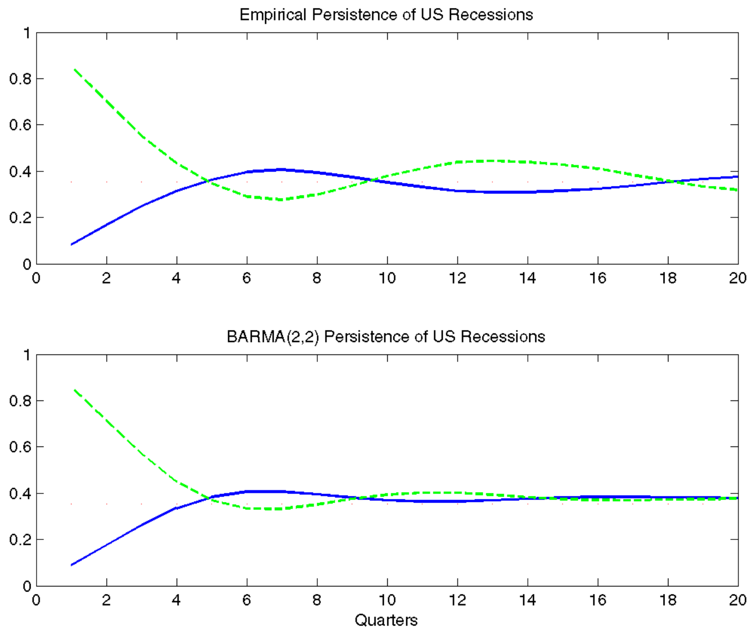


Figure 7:

8.4 Structural Break in the Recession Process

From visual inspection of Figure 3 it appears that pre-War and post-War business cycles are different. (The choice of 1945 as a break date reflects the NBER's use of 1945 as a break in presenting summary statistics.) Since the end of World War II, the U.S. economy has spent a lower proportion of time in recessions. Contractions have been shorter and expansions have been longer.

$BARMA(2,2)$	β_0	ϕ_1	ϕ_2	θ_1	θ_2	$logL$	R_M^2	R_p^2
all	-2.62	42.54	-37.58	-9.53	5.20	-180.78	.54	.61
pre-1945	-2.64	24.30	-19.06	-10.46	4.50	-114.52	.54	.55
1945-	-2.97	24.65	-20.10	-7.12	3.30	-56.20	.54	.61

$BARMA(2,2)$ Models of U.S. Recessions - Table 3

The likelihood ratio statistic on the null of no break equals 20.12, with an associated p-value of 0.001. Figure 9 shows the separate APFs. The shapes change modestly, with the primary difference being the the lower post-War mean reflecting the lower value of β_0.

9 Conclusion

The $BARMA(2,2)$ model is a substantial improvement over the traditional Markov model for U.S. recession data. More generally, the BARMA model is a useful extension to the statistical toolbox for modeling binary series over time. Its principal advantages are the ability to estimate restricted Markov models to circumvent the curse of dimensionality, and the ability to model non-Markovian processes. The autopersistence function and autopersistence graph provide graphical tools analogous to the autocorrelation function and correlogram used for Gaussian ARMA models.

10 References

Albert. J.H and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Society*, 89, 669-679.

Benjamin, M.A., Rigby, R.A., and Stasinopoulos, D.M. (2003), "Generalized Autoregressive Moving Average Models," *Journal of the American Statistical Society*, 98, 214-223.

Cox, D.R. (1981), "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93-115.

de Jong, R.M. and Woutersen, T.M. (2005), "Dynamic Time Series Binary Choice," Department of Economics, Ohio State University working paper.

Efron, B. (1978), "Regression and ANOVA with Zero-One Data: Measures of Residual Variation," *Journal of the American Statistical Association*, 73, 113-212.

Eichengreen, B., Watson, M.W., and Grossman, R.S. (1985), "Bank Rate Policy Under the Interwar Gold Standard," *The Economic Journal*, 95, 725-745.

Hamilton, J.D., (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Jacobs, P.A. and Lewis, P.A.W. (1978), "Discrete Time Series Generated By Mixture, I: Correlational and runs properties," *Journal of the Royal Statistical Society, Series B* 40, 94-105.

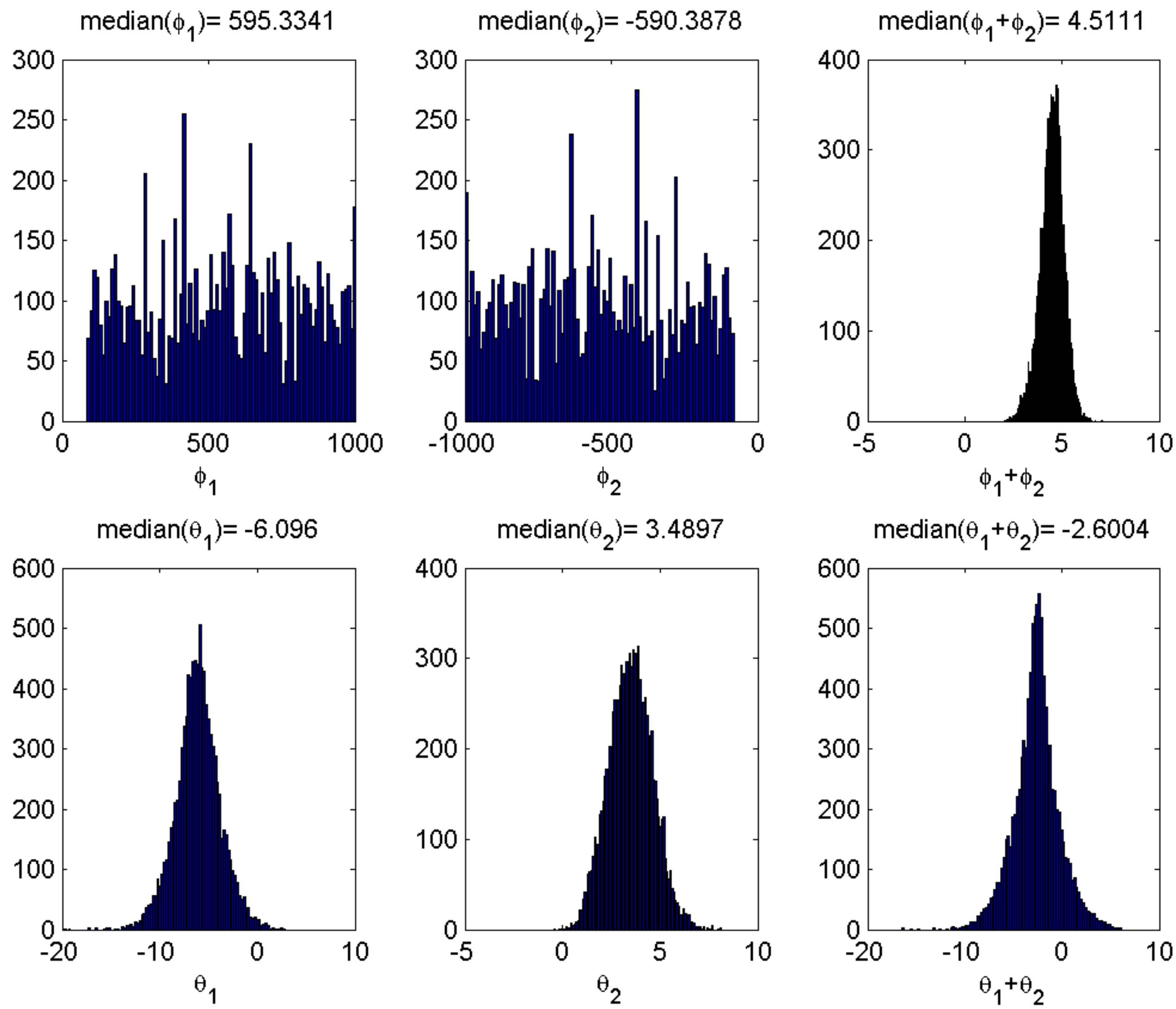
Koop, Gary (2003), *Bayesian Econometrics*, John Wiley & Sons.

Li, W.K. (1994), "Time Series Models Based on Generalized Linear Models: Some Further Results," *Biometrics*, 50, 506-511.

McFadden, D.(1974), "The Measurement of Urban Demand Travel," *Journal of Public Economics*, 3, 303-328.

Raftery, A. E. (1985). "A Model for High-Order Markov Chains," *J. Roy. Statist. Soc. Ser. B*, 47, 528–539.

Zeger, S.L. and Qaqish, B. (1988), "Markov Regression Models for time series: A quasi-likelihood approach," *Biometrics*, 44, 1019-1031.



median(ϕ1)= 595.3341
median(ϕ2)= -590.3878
median(ϕ1+ϕ2)= 4.5111
ϕ1
ϕ2
ϕ1+ϕ2
median(θ1)= -6.096
median(θ2)= 3.4897
median(θ1+θ2)= -2.6004
θ1
θ2
θ1+θ2

Figure 8:

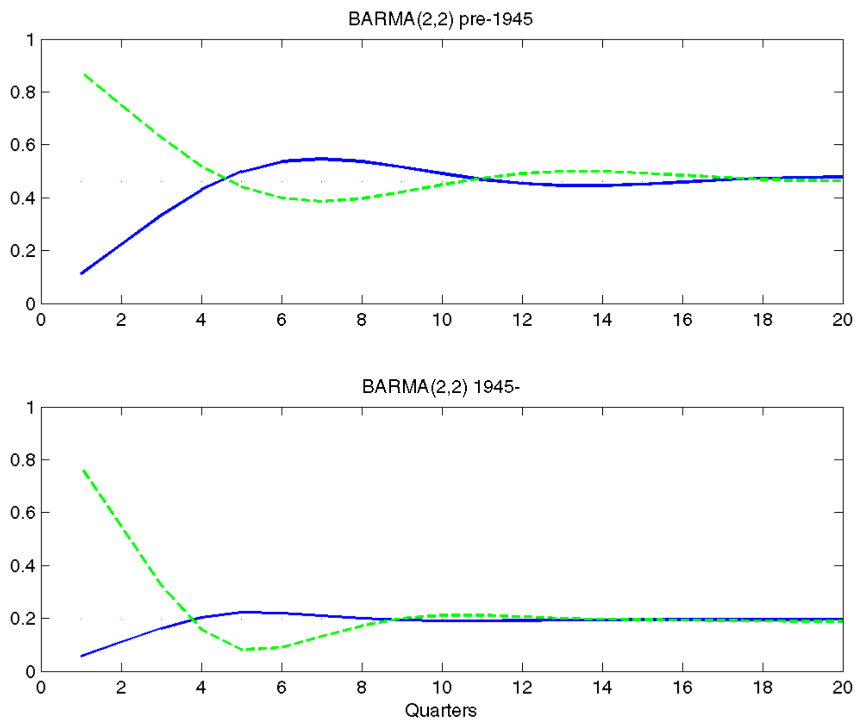

BARMA(2,2) pre-1945
BARMA(2,2) 1945-
Quarters
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10
12
14
16
18
20

Figure 9: