FEDERAL
RESERVE
BANK
of ATLANTA

Prior Parameter Uncertainty: Some Implications for Forecasting and Policy Analysis with VAR Models

John C. Robertson and Ellis W. Tallman

Working Paper 99-13
October 1999

Working Paper Series

Prior Parameter Uncertainty: Some Implications for Forecasting and Policy Analysis with VAR Models

John C. Robertson and Ellis W. Tallman

Federal Reserve Bank of Atlanta
Working Paper 99-13
October 1999

Abstract: Models used for policy analysis should generate reliable unconditional forecasts as well as policy simulations (conditional forecasts) that are based on a structural model of the economy. Vector autoregression (VAR) models have been criticized for having inaccurate forecasts as well as being difficult to interpret in the context of an underlying economic model. In this paper, we examine how the treatment of prior uncertainty about parameter values can affect forecasting accuracy and the interpretation of identified structural VAR models.

Typically, VAR models are specified with long lag orders and a diffuse prior about the unrestricted coefficients. We find evidence that alternatives that emphasize nonstationary aspects of the data as well as parsimony in parameterization have better out-of-sample forecast performance and smoother and more persistent responses to a given exogenous monetary policy change than do unrestricted VARs.

JEL classification: E44, C53

Key words: Bayesian inference, prior distributions, out-of-sample forecasting, structural VAR models, impulse responses

---

The authors thank Tao Zha for helpful discussions and the participants of the Atlanta Fed brown bag lunch series for comments. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to John C. Robertson or Ellis W. Tallman, Research Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, NW, Atlanta, Georgia 30303-2713, 404/521-8782 (Robertson), john.c.robertson@atl.frb.org, 404/521-8915 (Tallman), ellis.tallman@atl.frb.org.

INTRODUCTION

Vector autoregression (VAR) models have been used widely for both forecasting and policy analysis. VAR models designed for forecasting purposes utilize stochastic “forecasting priors” (Sims (1987)) on the model’s parameters of the type suggested by Litterman (1979, 1986), Doan, Litterman and Sims (1984), and Sims (1992). In contrast, those working on policy-oriented VAR models, like Gordon and Leeper (1994), and Christiano, Eichenbaum, and Evans (1997), *inter alia*, have largely eschewed the use of prior restrictions beyond the imposition of exact identification restrictions.

This paper documents the forecasting accuracy improvement to a VAR following the application of informative stochastic prior restrictions and then demonstrates how these priors can impact the dynamic relationships implied by the identified impulse responses. We highlight these differences by comparing the dynamics from various alternative specifications of a six-variable VAR subject to a typical set of identifying restrictions to isolate the effects of monetary policy shocks.

We generate finite sample probability distributions of forecasts and impulse responses in each specification allowing us to examine the quantitative and qualitative importance of differences among the model specifications. In forecasting applications, we find that a VAR model that uses an informative stochastic prior is more accurate on average across variables. The next best performing model is one specified in first differences and with the lag length restricted. In no case is the typical flat-prior VAR model in levels of the data the most accurate forecasting model in a mean squared error sense. Moreover, parameter uncertainty is a much larger source of uncertainty about post-sample data values under a flat prior than under more informative priors.

For analysis of the identified dynamics we find that the informative priors reinforce non-stationary aspects of the data displayed graphically in the impulse responses. We show that the impulse responses from a VAR in first-differences with a short lag length are generally similar to those of the stochastic "forecasting-prior" VAR, both of which display relatively smooth and persistent responses to an identified monetary policy shock. In contrast, the flat-prior VAR in levels generally implies only temporary responses.

The paper is organized as follows. The next section provides some additional motivation for our empirical analysis. The third section presents the VAR model we use. The fourth section presents our empirical results, and the final section draws some conclusions from our investigation.

BACKGROUND

Because of the substantial parameter uncertainty inherent to a high-dimensional forecasting model such as a VAR, VAR models designed for forecasting (extrapolation) purposes typically employ stochastic prior restrictions on the model's parameters of the type initially suggested by Litterman (1979, 1986). These forecasting priors usually give higher weight to coefficient values that favor unit roots, small deterministic components, and future time paths that are influenced primarily by recent observations. There is considerable empirical evidence that allowing this type of prior information to influence the shape of the model's likelihood function can substantially improve out-of-sample forecast performance (see for example Litterman (1979, 1986), Doan, Litterman and Sims (1984), Kennedy (1989), Sims (1992), and Robertson and Tallman (1999a,b), *inter alia*).

In contrast, relatively few studies make use of Bayesian forecasting priors as the basis for inference about identified impulse responses. That is, in policy analysis work with VAR models one typically assumes complete prior ignorance about the unknown parameter values, relying instead on the sample of data to determine the shape of the likelihood function. The dichotomy between the treatment of prior information in forecasting and policy VAR models is perhaps not too surprising. Sims' (1980) forceful criticism of the "incredible" nature of the exclusion restrictions traditionally placed on simultaneous equation systems is perhaps one explanation for the general aversion to using anything but a minimal amount of prior information in identified VAR models. It is notable however that while these fitted VAR models are assumed to be adequate descriptions of the distribution of the data, relatively little attention is given to the forecasting performance of the models.

Some studies have used Bayesian forecasting priors in identified VAR models; for example, Sims (1986,1987), Kennedy (1989), and Canova (1991) applied Litterman's form of stochastic priors on the reduced-form VAR coefficients together with a least squares type estimator to approximate the error covariance. However, in those cases the estimated probability distributions could only be viewed as being approximately correct under the assumption that the prior on the coefficients is independent of a diffuse prior about the error covariance (see Zellner (1971, pp. 239-240), and Kadiyala, and Karlsson (1997)). More recently, Highfield, O'Hara and Wood (1991), Leeper, Sims and Zha (1996), Faust (1998), Sims (1998), and Leeper and Zha (1999) have applied modifications of Litterman's setup that permits exact finite sample inference under only slightly stronger assumptions.

Separately, some authors note the apparent sensitivity of results from identified VAR models (with uninformative priors) to changes to the sample (see for example Pagan and

Robertson (1995), Rudebusch (1998)). This sensitivity is a common symptom of an “over-fitting” problem due to estimating many free parameters. If the likelihood surface is relatively flat then small changes in the sample could have a large effect on the maximum likelihood estimates. If the likelihood peak is sharply determined then it is less likely that additional data observations will lead to large revisions in the probability assessment of particular parameter values.

Bayesian VAR models with priors that are designed primarily to improve out-of-sample forecasting performance also impose smoothness on the impulse responses and influence their long-run properties. Sims (1987, 1999) conjectures that doing so allows the responses to be estimated with greater precision than if one ignored the prior information. Geweke (1999) demonstrates a formal link between a model’s adequacy and its out-of-sample prediction record in the context of marginal likelihood analysis. However, even amongst those using Bayesian methods on structural VAR models the perceived importance of the stochastic priors for the results is not always clear. For example, Leeper et al. (1996) use informative stochastic priors because, they argue, they are “essential in order to obtain sensible results…” (p. 62). Yet, the authors also assert that “Our prior tends to make the estimated impulse responses smoother, without changing their overall form.” p. 63. This latter statement conveys the impression that the choice of prior has little impact on the nature of the estimated dynamics, but there is no empirical evidence provided to support this claim. In fact, to the extent that the priors influence the shape of the likelihood function one would expect to observe at least some quantitative differences in the distribution of the impulse responses, and this idea motivates our present investigation.

THE MODELS

The models we examine are based on a *p*th order VAR process

$$y_t = b + B_1 y_{t-1} + \ldots + B_p y_{t-p} + u_t\,, \quad t = 1,\ldots,T \qquad (1)$$

where y_t denotes an $m\times 1$ vector of current dated observations for period *t* on the *m* variables in the VAR; the B_i are $m\times m$ coefficient matrices; and *b* is an $m\times 1$ vector of constant terms. The error term is defined as $u_t = A^{-1}\varepsilon_t$ where ε_t is assumed to be a Normal and independently distributed $m\times 1$ vector such that $E[\varepsilon_t | y_{t-s}, s>0] = 0$, and $E[\varepsilon_t \varepsilon_t' | y_{t-s}, s>0] = I$ for all *t*; and *A* is a non-singular $m\times m$ matrix so that the covariance of u_t is $\Sigma = A^{-1}A^{-1'}$. Imposing sufficient restrictions on *A* allows u_t to be disentangled into interpretable "structural" shocks.

Different values for *p* in (1) correspond to choosing different models in the present context because setting *p* amounts to restricting some coefficients to zero. In some models the coefficients are further restricted to satisfy the constraint $\sum_{l=1}^{p} B_i = I$, so that (1) can be written as a (*p*–1)th order VAR in first differences:

$$y_t = b + y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t\,, \qquad (2)$$

where $\Gamma_l = -\sum_{n=l+1}^{p} B_n$. Notice that this VAR can always be written in the form of (1) by defining $B_1 = I + \Gamma_1$, $B_n = \Gamma_n - \Gamma_{n-1}$, for $n = 2,\ldots,p$–1, $B_p = -\Gamma_{p-1}$.

Models can also differ in terms of the nature of the prior uncertainty assumed about the model's parameters – the prior density. We will work with the so-called natural conjugate (Normal-Wishart) family of prior densities. In most applied work a flat (diffuse or uninformative) prior is specified for $\theta = \text{vec}(B,\Sigma)$ in (1), or $\theta = \text{vec}(\Gamma,\Sigma)$ in (2), where $B = \left[b, B_1, \ldots, B_p\right]'$ and $\Gamma = \left[b, \Gamma_1, \ldots, \Gamma_{p-1}\right]'$, respectively. Our "forecasting prior" specifies an informative prior probability density for vec(B) in (1) conditional on Σ together with a diffuse prior on Σ. The forms of the prior and the posterior distributions are described in the Appendix.

To compute out-of-sample h-period-ahead forecasts and the impulse responses we solve equation (1) forward from period T and express y_{T+h} as

$$y_{T+h} = C_{h-1}b + J\boldsymbol{D}^h\boldsymbol{Y}_T + \sum_{i=0}^{h-1}\Psi_i\varepsilon_{T+h-i}\ , h = 1,2,\ldots \qquad (3)$$

where $C_0 = I$, $C_i = I + \sum_{j=1}^{i} B_j C_{i-j}$, $i = 1,2,\ldots$, with $B_j = 0$ for $j > p$; $\boldsymbol{Y}_T = \begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-p+1} \end{bmatrix}$: $(mp \times 1)$;

$\boldsymbol{D} = \begin{bmatrix} B_1 & \ldots & B_{p-1} & B_p \\ I & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \ldots & I & 0 \end{bmatrix}$: $(mp \times mp)$; $\Psi_i = J\boldsymbol{D}^i J'A^{-1}$: $(m \times m)$ is the i-period-ahead matrix of impulse responses; and $J = \begin{bmatrix} I & 0 & \ldots & 0 \end{bmatrix}$ is the $(m \times mp)$ matrix such that $y_T = J\boldsymbol{Y}_T$ (see Lutkepohl, 1991). Given data, $Y = \left[y_1, \ldots, y_T\right]'$, and parameters, θ, the first two terms in (3) sum to give the average value of y_{T+h}, that is $E[y_{T+h}|Y,\theta]$. The last term in (3) is the forecast error $y_{T+h} - E[y_{T+h}|Y,\theta]$, and has conditional variance equal to $\sum_{i=0}^{h-1}\Psi_i\Psi_i'$.

Having observed the data, equation (3) makes clear that a model's assessment of uncertainty about y_{T+h} emanates from two sources: the unknown parameters, θ, and the post-sample shocks ε_{T+j}, $j = 1,\ldots,h$. In contrast, variation in the impulse responses is due solely to uncertainty about θ. Under a Normal-Wishart prior it is relatively easy to simulate draws on the relevant random variables using the Monte Carlo methods presented in Sims and Zha (1998,1999), and Waggoner and Zha (1999).

The models we examine use data on six variables measured at a monthly frequency beginning in January of 1959, and we generate forecasts taking explicit account of the time lag of the release for different series by using the conditional forecasting technique described in Waggoner and Zha (1998). The data series are the levels of federal funds rate and the unemployment rate, the natural logarithms of a commodity price index, the Consumer price index, real GDP (distributed monthly), and the M2 monetary aggregate.[1] We follow standard practice and specify p = 13 lags as the default lag length. However, we also examine the case when p = 7 lags are used instead.[2]

We also impose sufficient exact exclusion restrictions on A to identify the VAR equations. Traditionally A would be assumed to be triangular for a particular ordering of variables, so that the model is a recursive one (see for example Christiano et.al. (1997)). The triangularity assumption means that there are no restrictions on Σ (the VAR is exactly identified), and so the structural VAR and the reduced-form VAR would imply the same conditional distribution for the data. We

---

[1] We use the Chow and Lin (1971) distribution technique, and it is applied exactly as described in Robertson and Tallman (1999a).
[2] A lag length of p=7 is the typical value chosen by the Akaike information criterion in a VAR estimated by ML.

chose A to be almost, but not quite, triangular.[3] The structural shock that is of primary interest is the one identified with monetary policy – the element of ε_t corresponding to the identified monetary policy rule. This policy equation assumes that the Fed can give weight to the current level of the money stock when determining the Funds rate, but responds to movements observed in real activity and prices with at least a one-month lag.

To summarize, we will analyze the out-of-sample forecast performance and the responses to a specific identified monetary policy shock for the following VAR models:

1. A pth-order VAR specified in levels of the data, a flat prior and p = 13 lags, denoted as model 1. The version of model 1 with p = 7 will be denoted as model 1a.

2. A (p–1)th-order VAR specified in first differences of the data, a flat prior and p = 13 lags, denoted as model 2. The version of model 2 with p = 7 will be denoted as model 2a.

3. A pth-order VAR specified in levels of the data, the informative "forecasting" prior described in the Appendix, and p = 13 lags, denoted as model 3.

EMPIRICAL RESULTS

Out-of-Sample Forecasting

We focus on each VAR model's forecasting performance for the Federal funds rate, unemployment, real output growth, and CPI inflation. Mean forecasts are formed for the current quarter, the next two quarters, as well as for the current calendar year and the subsequent two years over the period 1986 to 1997. Thus, for example, pooling the mean forecasts formed each

---

[3] A contains only 18 free coefficients, so there are 3 less free coefficients in A than in Σ. Consequently, iterative methods are required to estimate A. The restrictions on A are similar to those used in Sims (1987), Leeper and Zha

month of a quarter gives a total of 144 current quarter forecasts for each variable. Each forecast is constructed at the end of each month using real-time (actually available) data and with the posterior for θ updated every three months. Because of the data release lags (up to three months for GDP) we employ the conditional forecasting technique suggested initially by Doan, Litterman and Sims (1984) to allow the forecasts to depend on all the data that are available. The forecasts from each model are the average of 5000 realizations of y_{T+h} simulated from equation (3) and where each draw involves making independent draws on θ and $(\varepsilon_{T+1},\ldots,\varepsilon_{T+h})$.[4]

Table 1 displays forecasting accuracy statistics (root mean squared errors) for each model's mean forecasts.[5] The informative-prior VAR (model 3) is the most accurate in a mean squared error sense. However, the use of the dummy observations to modify the base random walk prior is important, particularly for the funds rate and unemployment series that do not exhibit strong trends (see the Appendix for a description of these dummy observations). For the trending series the "sum of coefficients" dummies appear more important than the "cointegration" dummy observation. A version of model 3 without either of the dummy observations is about as accurate as a VAR in levels with the lag length restricted to 7.

For models 1 and 2 the average forecast accuracy almost always improves when 7 lags are used rather than 13. Model 2a is closest to matching the average performance of the informative-

---

(1999) and is described in detail in Waggoner and Zha (1999).

[4] The data release lags allow us to condition on some elements of $y_{T+1},\ldots y_{T+j}$, $j \le 3$ when making forecasts. For example, at the end of the month we always have the current months funds rate, but only last month's CPI values, and GDP data for the previous quarter. We use the method presented in Waggoner and Zha (1998) for making draws on the constrained shocks $(\varepsilon_{T+1},\ldots,\varepsilon_{T+j})$ over the conditioning period. The mean of these constrained shocks will be non-zero in general. The real-time aspects of the setup tend to be most important for the shorter-run (one- to two-quarter-ahead) forecasts of GDP. Failing to condition on all available data and/or using the 1999 vintage of historical data instead of real-time data vintages has a relatively small effect on the annual forecasts.

[5] The results are essentially the same if we were to evaluate (3) at the posterior mean of the coefficients instead of directly using the posterior mean of (3).

prior VAR, particularly for the trending series. That is, combining the "sum of coefficients" restriction exactly with a shorter lag length is useful for these series. However, for the funds rate and unemployment rate model 2a does not compare quite as favorably. The damping of the intercept terms toward zero and not imposing the "sum of coefficients" restriction exactly in model 3 appears to be important in these cases.

The finding that funds rate forecasts are considerably more accurate in model 3 than in model 1 is also relevant for the debate about the usefulness of VAR models for policy analysis. Rudebusch (1998) cites evidence that VAR model forecasts of the funds rate are disappointingly poor as a possible justification for dismissing VAR models as tools for policy analysis. However, his evidence is based on VARs specified in levels of the data with long lag lengths and a diffuse prior about the coefficient values. Our results indicate that the use of specific informative priors on the VAR coefficients can overturn the negative conclusion about VAR forecast performance.[6]

Summarizing the average accuracy of point forecasts is important but it does not show the effect that the priors have on a model's assessment of the uncertainty about the future. As an illustration, Charts 1 through 4 depicts the forecast distribution for the annual average funds rate, unemployment rate, GDP growth and CPI inflation, respectively, from each of three VAR models. The forecasts are formed at the end of April 1994 (following release of first quarter GDP) for the current and each of the next three calendar years. The first figure in each Chart displays the forecasts from model 3. The second figure gives the forecasts from model 2a. The third figure shows the forecasts from model 1. The actual post-sample realized values for each series are

---

[6] In a related study, Robertson and Tallman (1999b) find that the mean funds rate forecasts from model 3 compare favorably in a mean squared error sense to those from the federal funds futures market at the one and two month horizons. A distinct advantage to the VAR approach in this regard is that it can produce forecasts at horizons much longer than contained in futures market data.

shown as solid lines in each graph. The outer dashed lines are the 16th and 84th percentiles of the conditional distribution of the relevant components of y_{T+h}.[7] The center dashed line is the mean forecast. The inner dotted lines give the 16th and 84th percentiles when we evaluate (3) at the mean of the posterior distribution of the VAR parameters and only let future shocks generate uncertainty about the future time path of the series. That is, the dotted lines correspond to the usually reported case that ignores parameter uncertainty. Not surprisingly, parameter uncertainty is a much smaller source of uncertainty in the forecasts that lead to a more sharply determined likelihood (model's 3 and 2a) than in the forecasts from the 13-lag VAR in levels under a flat prior (model 1).

It is notable that the GDP growth and unemployment forecasts from model 1 are inaccurate – the actual outcomes lie well outside the 68 percent probability bands. The forecast interval of the funds rate contains the actual outcome, but also implies that interest rates at or below zero would have been a better than even chance in 1997. In general, the VAR in first differences (model 2a) has much tighter probability intervals than the VAR in levels (model 1). For GDP growth and inflation these intervals are concentrated around the actual outcomes, and the mean forecasts would have resulted in only small forecast errors. For the funds rate and unemployment rate the outcomes lie near the top and bottom of the 68 percent probability intervals, respectively. The funds rate distribution in particular assigns higher probability to decreasing rather than increasing interest rates over the forecast horizon. The 68 percent probability intervals for GDP growth and inflation from model 3 are about the same width as from model 2a, but are concentrated on slightly lower GDP growth and higher inflation than actually occurred. The intervals for the funds

[7] The bands are constructed point-wise, so that the posterior probability of y_{T+h} being in the band is 68% at each forecast horizon individually.

rate distribution is narrower and would have assigned a slightly higher probability to the actual outcomes than model 2a. The 68 percent probability interval for the unemployment rate is also narrower than from model 2a, but still contains the actual outcomes inside the lower band.

Impulse response analysis

Charts 5 through 8 display the impulse responses to a one-time one standard deviation monetary policy shock in each of the three model specifications discussed above (models 1, 2a and 3) fitted over the full-sample (January 1959 to December 1997). If the policy equation is the *k*th equation in the VAR then the point-wise probability distributions of the impulse responses are constructed by making random draws on the *k*th column of Ψ_i for periods $i = 0,\ldots,h$. The impulse response function shows how much the expectation of Y_{T+h} given Y and θ is revised when one also conditions on the event that $\varepsilon_{k,T+1} = 1$. The solid line in each figure is the mean response while the dashed lines give the 68 percent probability bands. The dotted line is the impulse response evaluated at the posterior mean of θ.

The following short-run features of the mean responses to a monetary policy shock are broadly similar across all three models: the funds rate rises, real output growth and inflation decline, and the rate of unemployment increases. In contrast, the longer-term mean impacts of the shock differ across models, although the main differences distinguish the diffuse prior model (model 1) from the other two models. The mean responses in model 1 display only small and temporary impacts of a policy shock on the inflation rate, the growth rate of real output, and on the federal funds rate. The mean responses for model 2a and model 3 display more persistent effects on the growth rate of real output, the unemployment rate, and inflation than those observed in model 1.

For models 2a and 3 we found that the variability in $E[y_{T+h}|Y,\theta]$ due to parameter uncertainty was much smaller than for model 1, reflecting the more parsimonious nature of models 2a and 3 relative to model 1. An examination of the 68 percent probability bands of the responses suggests that uncertainty about *k*th column of the Ψ_i is quite large even in models 1 and 2a.[8] Nonetheless, the probability bands indicate some significant differences across models. In many instances the mean impulse response outcomes from model 1 would be assigned a low probability of occurring from the perspective of either model 2a or model 3. Model 2a and 3 are more similar in the sense that there is considerable overlap in the probability bands, particularly at longer horizons.

The treatment of prior parameter uncertainty may also influence a model's sensitivity to changes in sample information. As an illustration, we estimate the models (1, 2a, and 3) using data for the period 1959 to 1991 and compare the resulting impulse responses with the full sample estimates discussed above. The effects on the output growth and inflation responses to a monetary shock are displayed in charts 9 and 10, respectively. In each chart, the dashed lines are the 68 percent probability bands based on the full sample, and the dotted lines are the corresponding probability bands based on the shorter sample.

For model 3, the assessment of the likely paths for output growth and inflation following a policy shock are largely the same in either sample period. That is, there is considerable overlap in the probability bands at each horizon. For models 1 and 2a, there appear some notable differences across sample periods at some horizons. Compared to the full sample estimates, model 1 implies that a less persistent reduction in inflation is more likely when the model is estimated over the

[8] Observe that $E[y_{T+h}|Y,\theta]$ depends on *B*, while $E[y_{T+h}|Y,\theta,\varepsilon_{k,T+1}=1]$ depends on both *B* and A^{-1}. It appears

shorter sample. For model 2a, the impulse responses based on the shorter sample period suggest that higher output growth and lower inflation are more likely in the medium run than when the responses are estimated using the full sample.

CONCLUSIONS

In this paper we demonstrate that the choice of prior restrictions on a reduced form VAR can affect forecast performance. We further show that the choice of prior can have an important influence on the nature of structural inferences about the effects of exogenous changes in monetary policy.

The VAR models we examine differ in terms of the lag length used, whether or not the model is specified in levels or first differences, and the extent of prior uncertainty assumed about each models parameters. The empirical results reinforce the conclusions of previous research that for trending data pre-differencing or using a prior that gives weight to a specification in first differences is advantageous for forecasting. There is additional gain from restricting the number of parameters freely estimated in the VAR by either setting coefficients on longer lags to zero or by damping the variation of these coefficients toward zero.

We compute the distribution of the impulse responses to a particular identification of an exogenous monetary policy shock in each model. The longer-run characteristics of the estimated responses differ in ways that can be directly attributed to the nature of the prior uncertainty assumed for each model's reduced form parameters. In particular, the mean responses from a diffuse prior model in levels display small and temporary impacts of a policy shock on the inflation rate, growth rate of real output, and on the federal funds rate. The models that are closer

---

to be the variation in the elements of A^{-1} that is primarily responsible for the relatively wide error bands.

to a parsimonious first differenced form display smoother and more persistent effects on the growth rate of real output, the unemployment rate, and inflation.

Forecasting accuracy and impulse response patterns that can be interpreted in the context of economic theory are both criteria for choosing a model for use in policy analysis. The results in this paper suggest that using certain types of restrictions on the dynamics of a VAR can improve the forecast performance over largely unrestricted VAR specifications of the type typically used in the policy analysis literature. At the same time these restrictions may impact substantially a model's responses to an identified policy shock. Given a particular set of structural restrictions, the effects from using different priors on the interpretation of a VAR should not be ignored.

APPENDIX – The Prior

The VAR in (1) can be written in stacked form as

$$Y = XB + u$$

where $X_t = \left[\iota', y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p}\right]'$, $Y = \left[y_1, \ldots, y_T\right]'$, $X = \left[X_1, \ldots, X_T\right]'$, $u_T = \left[u_1, \ldots, u_T\right]'$ and $B = \left[b, B_1, \ldots, B_p\right]'$. Because the u_t 's are assumed independent and Normally distributed, the ML estimator for (*B*,Σ) is $\hat{B} = \left(X'X\right)^{-1} X'Y$, $\hat{\Sigma} = \frac{1}{T}\left(Y - XB\right)'\left(Y - XB\right)$.

The prior we use is of the Normal-Wishart form (see for example, Drèze and Richard, 1983), Highfield, O'Hara and Wood (1991)). A Normal-Wishart prior distribution is described by a mean coefficient matrix B_0 of size $m \times m(p+1)$, a positive definite mean covariance matrix S_0 of size $m \times m$, as well as an $(p+1) \times (p+1)$ positive definite matrix H_0 and a real number $\nu_0 \geq 0$ to describe the prior uncertainty about (*B*,Σ) around their means. Conditional on Σ, the vectorized form of the coefficient matrix, vec(*B*), follows a Normal distribution $N\left(\mathrm{vec}(B_0), \Sigma \otimes H_0^{-1}\right)$, while Σ^{-1} follows a Wishart distribution $W_m\left(S_0^{-1}/v_0, v_0\right)$ with $E[\Sigma^{-1}] = S_0^{-1}$.

Given the Normality assumption for the u_t 's, an attractive feature of using a Normal-Wishart prior is that the posterior distribution (the product of the likelihood function and the prior distribution) is also Normal-Wishart. In particular, conditional on Σ, vec(*B*) follows the Normal distribution $N\left(\mathrm{vec}(B_T), \Sigma \otimes H_T^{-1}\right)$, while Σ^{-1} is distributed as $W_m\left(S_T^{-1}/v_T, v_T\right)$, where

$H_T = H_0 + X'X$; $v_T = v_0 + T$; $S_T = \frac{v_0}{v_T} S_0 + \frac{T}{v_T}\hat{\Sigma} + \frac{1}{v_T}\left(\hat{B} - B_0\right)' H_0 H_T^{-1} X'X\left(\hat{B} - B_0\right)$; and

$B_T = H_T^{-1}\left(H_0 B_0 + X'X\hat{B}\right)$. Monte Carlo methods are available for making draws from a Normal-Wishart posterior distribution even in the presence of over-identifying restrictions on Σ, (Sims and Zha (1998, 1999) and Waggoner and Zha (1999)).

The nature of the prior distribution is determined by the values assumed for the parameters of the prior distribution. A diffuse prior is characterized by selecting $v_0 = 0$ and $H_0 = 0$. Then $H_T = X'X$, $B_T = \hat{B}$, $v_T = T$, and $S_T = \hat{\Sigma}$. The alternative base prior we use has $H_0 > 0$, so that the prior about Σ is left diffuse and we have an informative prior about *B*. Under this alternative we assume that the prior coefficient mean has the random walk form, $B_0 = [0, I, 0, \ldots, 0]'$, and H_0^{-1} is a diagonal matrix. Hence, up to an unknown scale factor the prior variances of the coefficients are the same in all equations. The first diagonal element of H_0^{-1} is set to 10^6 to reflect an effectively diffuse prior about the intercept terms. The diagonal elements corresponding to the *k*th lag of the *j*th variable is set to $\omega_{jk} = \left(\lambda / ks_j\right)^2$, where the parameter λ controls the overall tightness of the prior distribution of the coefficients about their means.[9] Notably, this is a less complicated form of the Normal-Wishart prior than presented in Sims and Zha (1998), and Robertson and Tallman (1999a). In these papers $v_0 > 0$ so that the prior is informative about Σ. The approach here is similar to that used by Highfield, et.al. (1991).

We then modify the random walk prior by feeding in *m*+1 weighted dummy observations Y_0 and X_0 of the form described in Sims (1992), Sims and Zha (1998), and Robertson and

---

[9] Litterman assumes that the lag coefficient variances have the form $\omega_{jki} = \omega_{jk} s_i \tau^{\delta(i,j)}$, where $\tau \in (0,1]$ is the relative tightness for coefficients on lags of "other" variables, and the function $\delta(i,j)$ is zero for $i = j$ and one otherwise. Allowing the prior variances to differ across equations by more than a single scale factor leads to a less tractable posterior distribution than under the Normal-Wishart assumptions (see Kadiyala and Karlsson (1997)).

Tallman (1999a). These dummy observations are designed to push the model towards a form containing unit roots, while also damping the role of deterministic components. The dummy observations have no effect on the random walk structure of B_0. However, they do interact with H_0 to reduce the size of the prior variation in the constant term around zero as well as inducing negative correlation between the coefficients on the lags of the *j*th variable each equation and between the constant term the lag coefficients in each equation. Thus, if coefficients deviate from their prior means the other coefficients will tend to deviate in an offsetting way. Taken together, the dummy variable modifications re-enforces the view that a “no change” forecast is a reasonable starting point (Sims (1992), Leeper, Sims and Zha (1996)).

The tightness parameter, λ, the s_j factors, and the weights on the dummy observations are numbers that must be pre-specified by the analyst. Following standard practice we choose $\lambda = 0.1$ and set the s_j equal to the standard error from *p*th-order univariate autoregressons fit to each series. As in Sims (1992) the zero/one dummy observations are initially scaled by the pre-sample means of the *p* initial conditions. A weight of $\mu_1 = 5$ is then applied to the *m* “sum of coefficients” dummies and a weight of $\mu_2 = 5$ is attached to the single “cointegration” dummy. To investigate the sensitivity of the results to these choices we also experiment with setting μ_1 and/or μ_2 to zero.

REFERENCES

Canova, Fabio. 1991. "The Sources of Financial Crisis: Pre- and Post-Fed Evidence." *International Economic Review*. 32,3: 689-713.

Christiano, Lawrence, J., Martin Eichenbaum, Charles L. Evans. 1997. "Monetary Shocks: What Have We Learned, and to What End?" Forthcoming in the *Handbook of Macroeconomics.*

Doan, Thomas , Robert B. Litterman and Christopher A. Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*. 3: 1–100.

Drèze, J.H., and Jean-F. Richard. 1983. "Bayesian Analysis of Simultaneous Equation Systems." In Z. Griliches and M.D. Intriligator (eds), *Handbook of Econometrics*, Vol I, Amsterdam.

Faust, Jon. 1998. "The Robustness of Identified VAR Conclusions About Money." *Carnegie-Rochester Conference Series on Public Policy*. 49: 207-244.

Geweke, John. 1999. "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication." *Econometric Reviews*. 18,1: 1-74.

Gordon, David, and Eric M. Leeper. 1994. "The Dynamic Impacts of Monetary Policy: An Exercise in tentative Identification." *Journal of Political Economy* 102,6: 1228-1247.

Highfield, Richard., O'Hara, Maureen, and John Wood. 1991. "Public Ends, Private Means: Central Banking and the Profit Motive 1823-1832" *Journal of Monetary Economics* 28,2: 287-322.

Kadiyala, K. Rao and Sune Karlsson. 1997. "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics* 12:99-132.

Kennedy, J.E. 1989. "The Effect of Bayesian Priors on the Moving-Average Representation of Vector Autoregressions." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series Working Paper 79, June.

Leeper, Eric M., Christopher A. Sims, Tao Zha. 1996. "What Does Monetary Policy Do?" *Brookings Papers of Economic Activity*. 2: 1-78.

Leeper, Eric M., and Tao Zha. 1999. "Identification and Forecasting: Joint Inputs to Policy Analysis" Manuscript. Federal Reserve Bank of Atlanta. May.

Litterman, Robert B. 1979. "Techniques of Forecasting Using Vector Autoregressions." Working Paper 115, Massachusetts Institute of Technology, Department of Economics.

Litterman, Robert B. 1986 "Forecasting with Bayesian Vector Autoregressions – Five Years of Experience." *Journal of Business and Economic Statistics*. (January)4,1:25–38.

Lutkepohl, Helmut. 1991. *Introduction to Multiple Time Series Analysis*. Springer Verlag: New York.

Pagan, Adrian R., and John C. Robertson. 1995. “Resolving the Liquidity Effect.” Federal Reserve Bank of St Louis *Review*. 77,3: 33-54.

Robertson, John C. and Ellis W. Tallman. 1999a. “Vector Autoregressions: Forecasting and Reality.” Federal Reserve Bank of Atlanta *Economic Review,* First Quarter, 4-19.

Robertson, John C. and Ellis W. Tallman. 1999b. “Improving Forecasts of the Federal Funds Rate in a Policy Model.” Federal Reserve Bank of Atlanta Working Paper 99-03. February.

Rudebusch, Glenn D. 1998. “Do Measures of Monetary Policy in a VAR Make Sense?” *International Economic Review*. (November) 39, 4: 907-31.

Sims, Christopher A. 1980. “Macroeconomics and Reality. *Econometrica*. 48,1: 1-48.

Sims, Christopher A. 1986. “Are Forecasting Models Usable For Policy?” Federal Reserve Bank of Minneapolis *Quarterly Review*. Winter: 2-16.

Sims, Christopher A. 1987. “Comment to D. Runkle, ‘Vector Autoregressions and Reality’.” *Journal Business and Economic Statistics*. 5: 443-449.

Sims, Christopher A. 1992. “A Nine-Variable Probabilistic Macroeconomic Forecasting Model.” In *Business Cycles, Indicators, and Forecasting,* Stock, James H. and Mark W. Watson, eds. University of Chicago Press: Chicago. 179–204.

Sims, Christopher A. 1998. “The Role of Interest Rate Policy in the Generation and Propagation of Business Cycles: What Has Changed Since the ‘30s?” In J.C. Fuhrer and S Scott (eds), *Beyond Shocks: What Causes Business Cycles?* Conference Series No. 42. Federal Reserve Bank of Boston. June: 121-160.

Sims, Christopher A. 1999. “Using a Likelihood Perspective to Sharpen Econometric Discourse: Three Examples.” *Journal of Econometrics* (forthcoming).

Sims, Christopher A. and Tao A. Zha. 1998. “Bayesian Methods for Dynamic Multivariate Models.” *International Economic Review*. 39, 4: 949–968.

Sims, Christopher A. and Tao A. Zha. 1999. “Error Bands for Impulse Responses.” *Econometrica*. (forthcoming).

Waggoner, Daniel and Tao A. Zha. 1998. “Conditional Forecasts in Dynamic Multivariate Models.” Federal Reserve Bank of Atlanta Working Paper 98-22. December.

Waggoner, Daniel and Tao A. Zha. 1999. “Does Normalization Matter for Inference?” Manuscript, Federal Reserve Bank of Atlanta. February.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*. John Wiley and Sons. New York.

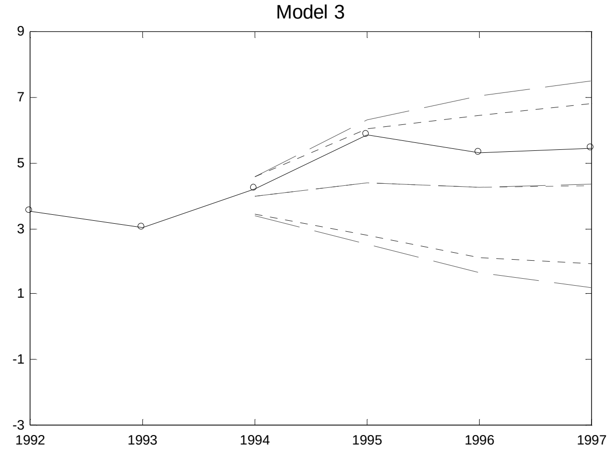
Table 1: RMSE of 6-variable VAR Model Forecasts 1986-1997

	Current Quarter	First Quarter	Second Quarter	Current Year	First Year	Second Year
Federal Funds Rate						
Model 1 (p=13)	0.24	1.05	1.62	0.72	2.69	4.69
Model 1a (p=7)	0.21	0.90	1.35	0.56	2.08	3.33
Model 2 (p=13)	0.20	0.74	1.08	0.43	1.65	2.99
Model 2a (p=7)	0.16	0.63	0.95	0.39	1.56	2.66
Model 3 (p=13, λ=.1, μ1=μ2=5)	0.12	0.46	0.77	0.28	1.23	2.09
Model 3a (p=13, λ=.1, μ1=μ2=0)	0.18	0.84	1.41	0.58	2.30	3.65
Model 3b (p=13, λ=.1, μ1=5, μ2=0)	0.12	0.47	0.81	0.31	1.35	2.19
Model 3c (p=13, λ=.1, μ1=0, μ2=5)	0.18	0.85	1.45	0.60	2.33	3.66
Unemployment						
Model 1 (p=13)	0.19	0.37	0.55	0.26	0.92	1.43
Model 1a (p=7)	0.18	0.34	0.49	0.23	0.84	1.39
Model 2 (p=13)	0.18	0.36	0.57	0.23	0.97	1.46
Model 2a (p=7)	0.17	0.32	0.48	0.20	0.78	1.22
Model 3 (p=13, λ=.1, μ1=μ2=5)	0.15	0.27	0.37	0.16	0.58	0.90
Model 3a (p=13, λ=.1, μ1=μ2=0)	0.16	0.31	0.45	0.21	0.76	1.26
Model 3b (p=13, λ=.1, μ1=5, μ2=0)	0.15	0.28	0.40	0.16	0.62	0.90
Model 3c (p=13, λ=.1, μ1=0, μ2=5)	0.17	0.32	0.47	0.22	0.76	1.26

Table 1 (Cntd).						
	Current Quarter	First Quarter	Second Quarter	Current Year	First Year	Second Year
			CPI Inflation			
Model 1 (p=13)	1.11	2.02	2.18	0.55	1.67	2.53
Model 1a (p=7)	1.06	2.04	2.18	0.52	1.54	1.55
Model 2 (p=13)	1.03	1.69	1.83	0.44	1.17	1.32
Model 2a (p=7)	1.02	1.67	1.58	0.44	1.07	0.98
Model 3 (p=13, λ=.1, μ1=μ2=5)	0.93	1.51	1.55	0.41	1.05	1.06
Model 3a (p=13, λ=.1, μ1=μ2=0)	1.00	1.89	2.04	0.49	1.47	1.47
Model 3b (p=13, λ=.1, μ1=5, μ2=0)	0.93	1.52	1.55	0.41	1.05	1.05
Model 3c (p=13, λ=.1, μ1=0, μ2=5)	1.01	1.89	2.04	0.49	1.47	1.48
			GDP growth			
Model 1 (p=13)	2.82	3.02	2.96	0.95	2.15	2.50
Model 1a (p=7)	2.47	2.48	2.61	0.84	2.05	2.20
Model 2 (p=13)	2.44	2.38	2.50	0.74	1.64	1.92
Model 2a (p=7)	2.23	2.11	2.19	0.71	1.42	1.71
Model 3 (p=13, λ=.1, μ1=μ2=5)	2.14	2.06	2.11	0.71	1.43	1.59
Model 3a (p=13, λ=.1, μ1=μ2=0)	2.45	2.60	2.63	0.85	2.10	2.36
Model 3b (p=13, λ=.1, μ1=5, μ2=0)	2.10	2.00	2.04	0.70	1.40	1.84
Model 3c (p=13, λ=.1, μ1=0, μ2=5)	2.50	2.63	2.64	0.86	2.17	2.54

Shaded cells indicate the smallest RMSE values. Model 1 is a VAR specified in levels as in equation (1) with a diffuse prior. Model 2 is a VAR specified in first differences as in equation (2) and with a diffuse prior. Model 3 is a VAR specified in levels and using a non-diffuse prior of the form described in the Appendix.

Chart 1: Funds rate forecasts



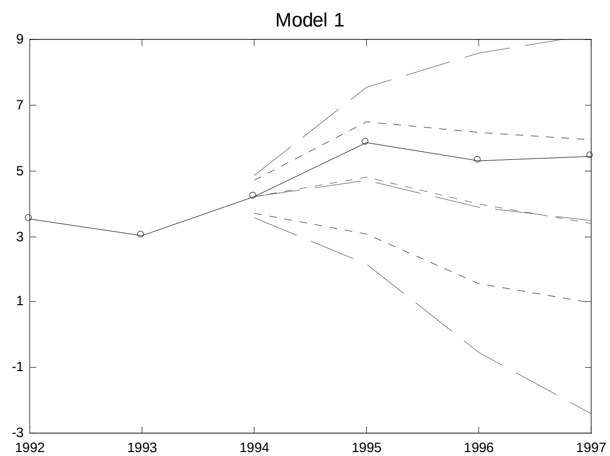
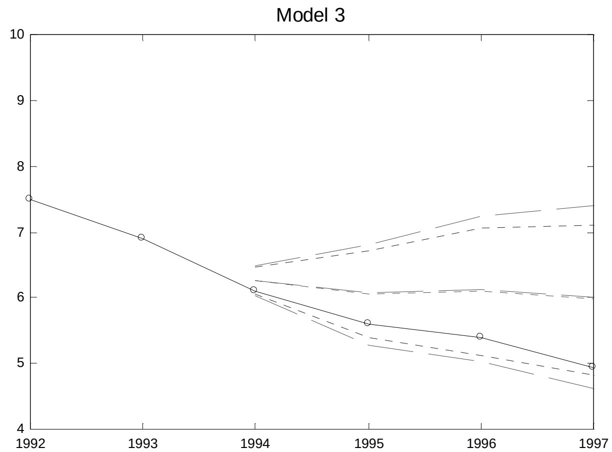


Chart 2: Unemployment forecasts



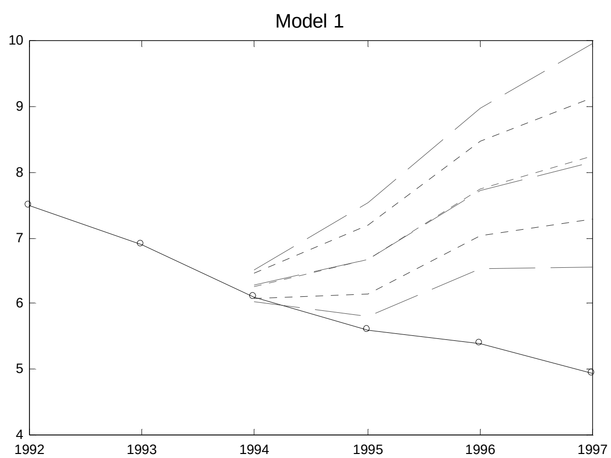
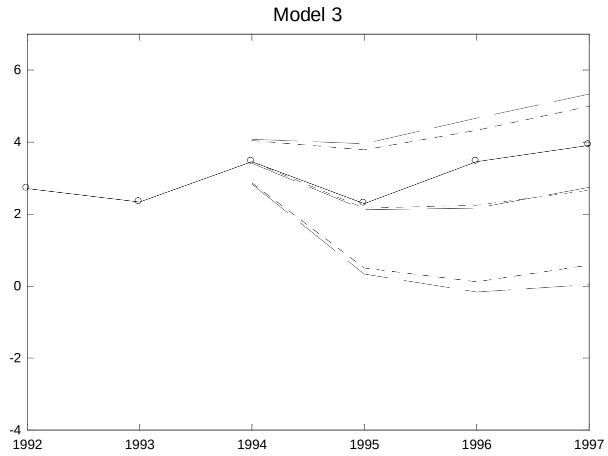


Chart 3: GDP growth forecasts



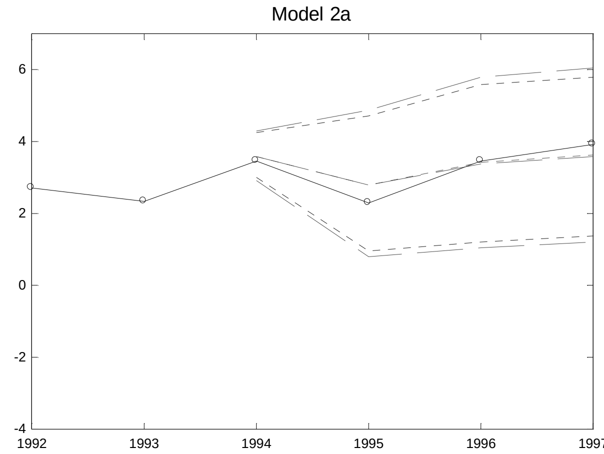


Chart 4: CPI inflation forecasts

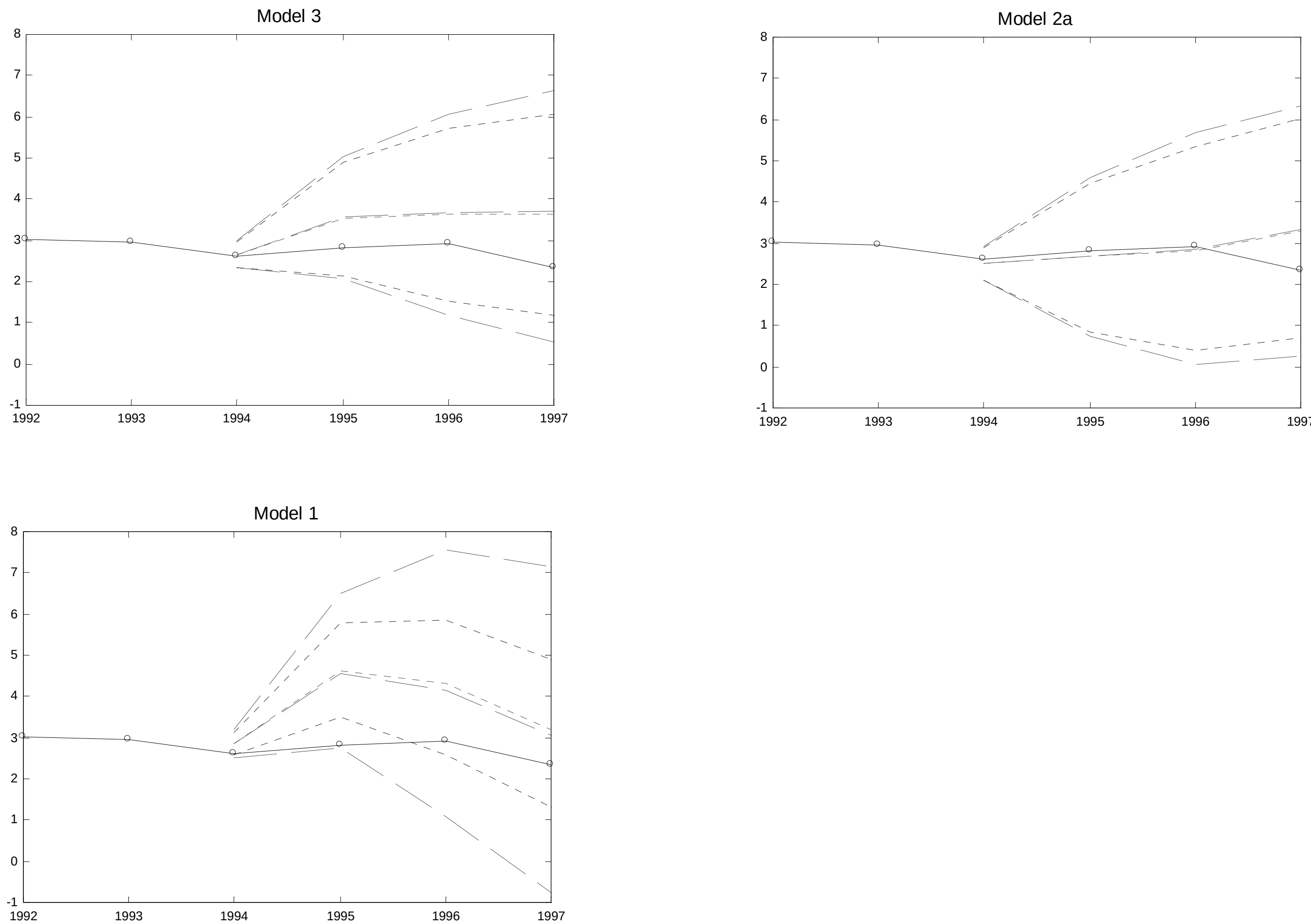


Chart 5: Funds rate response to policy shock (quarterly)

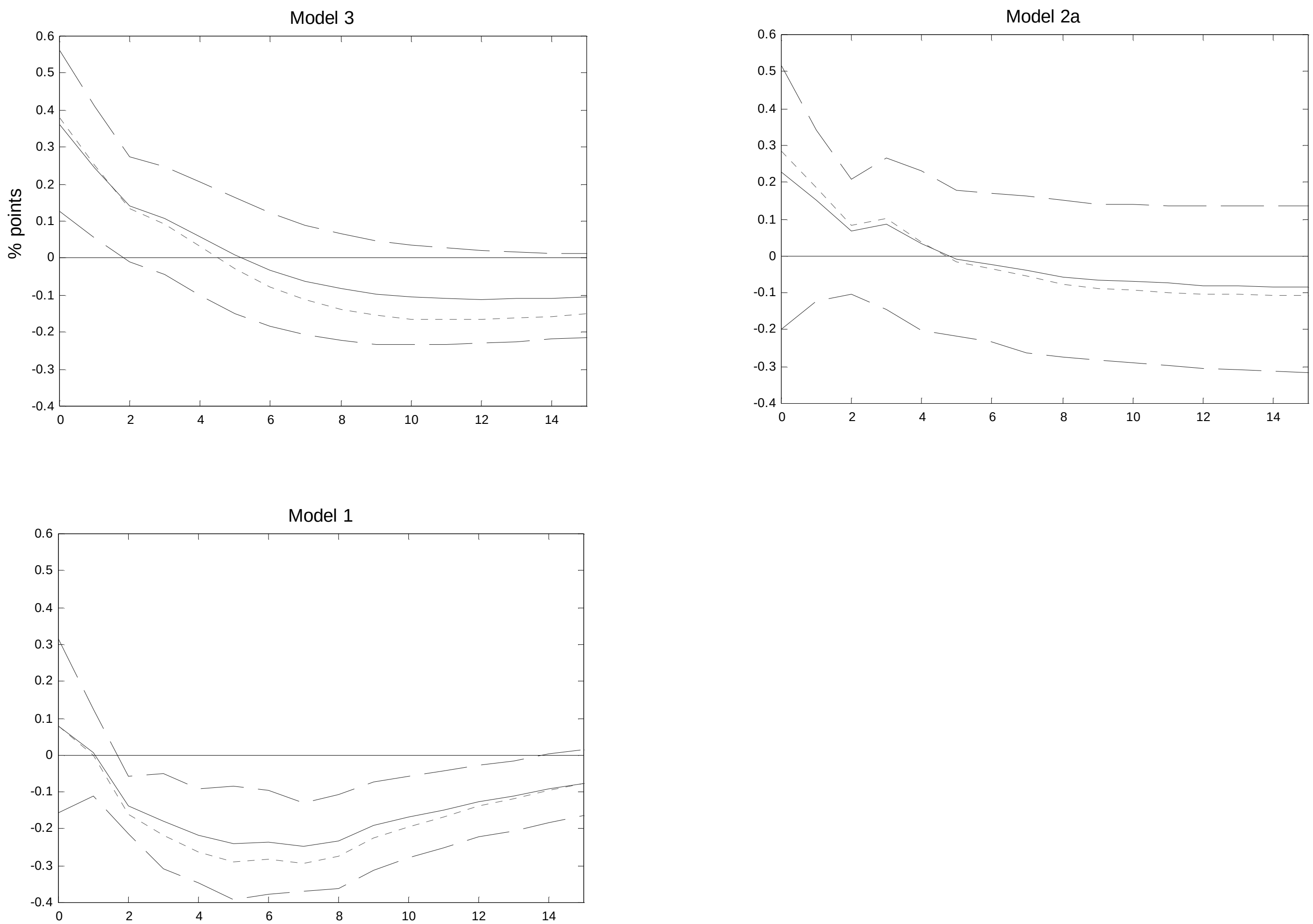


Chart 6: Unemployment response to policy shock (quarterly)

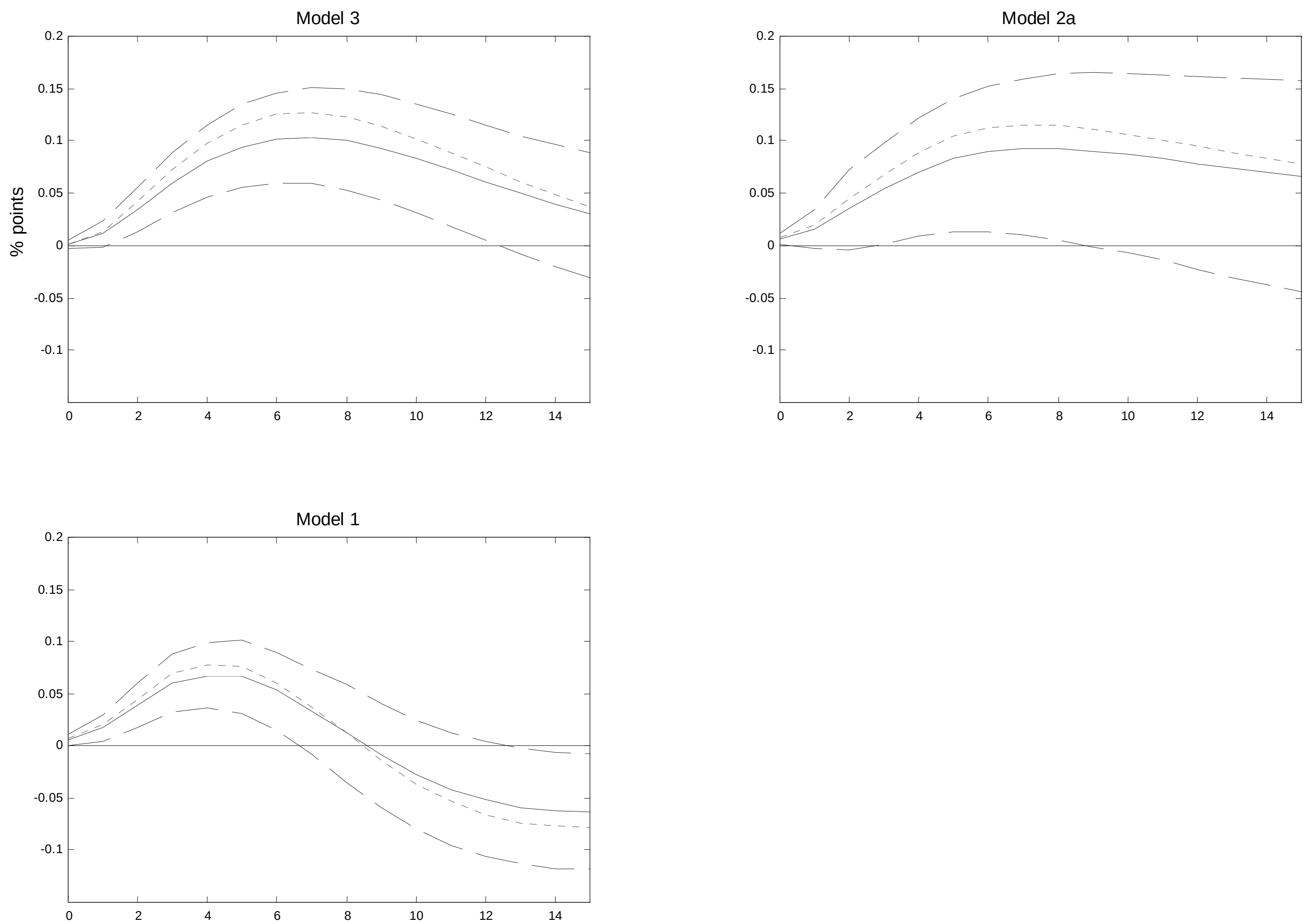


Chart 7: GDP growth response to policy shock (quarterly)

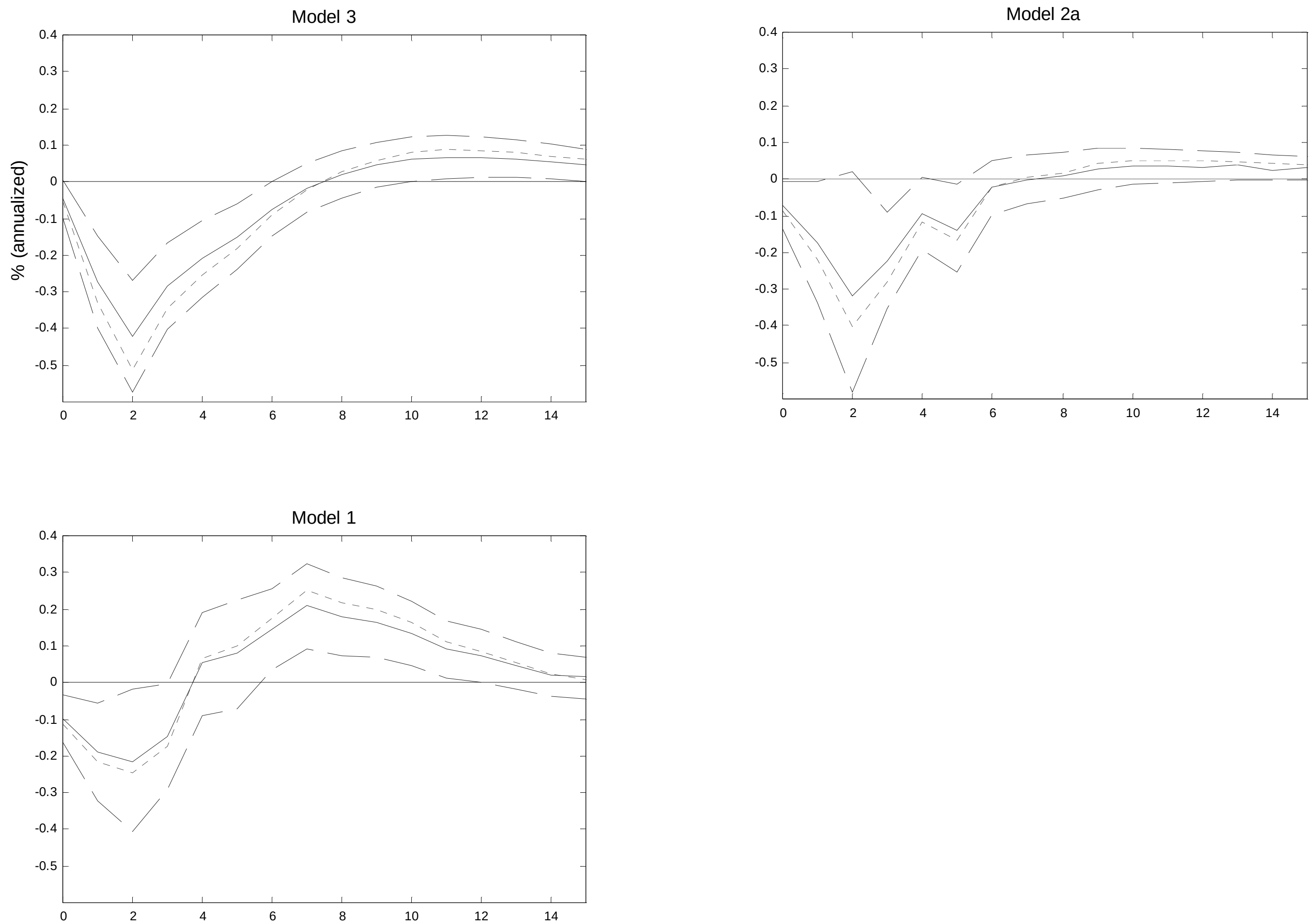


Chart 8: CPI inflation response to policy shock (quarterly)

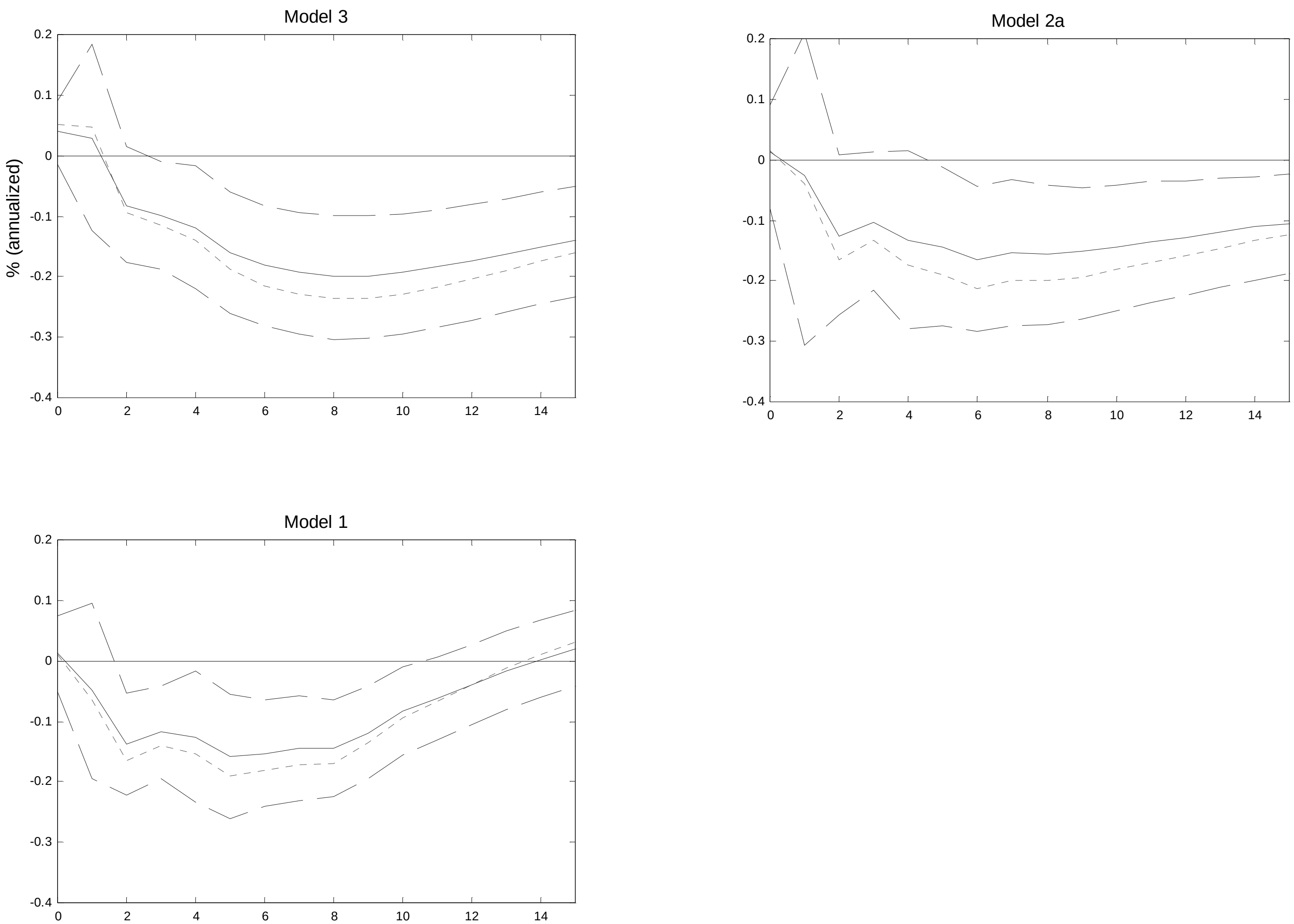


Chart 9: GDP growth response to policy shock (59-91 and 59-97)

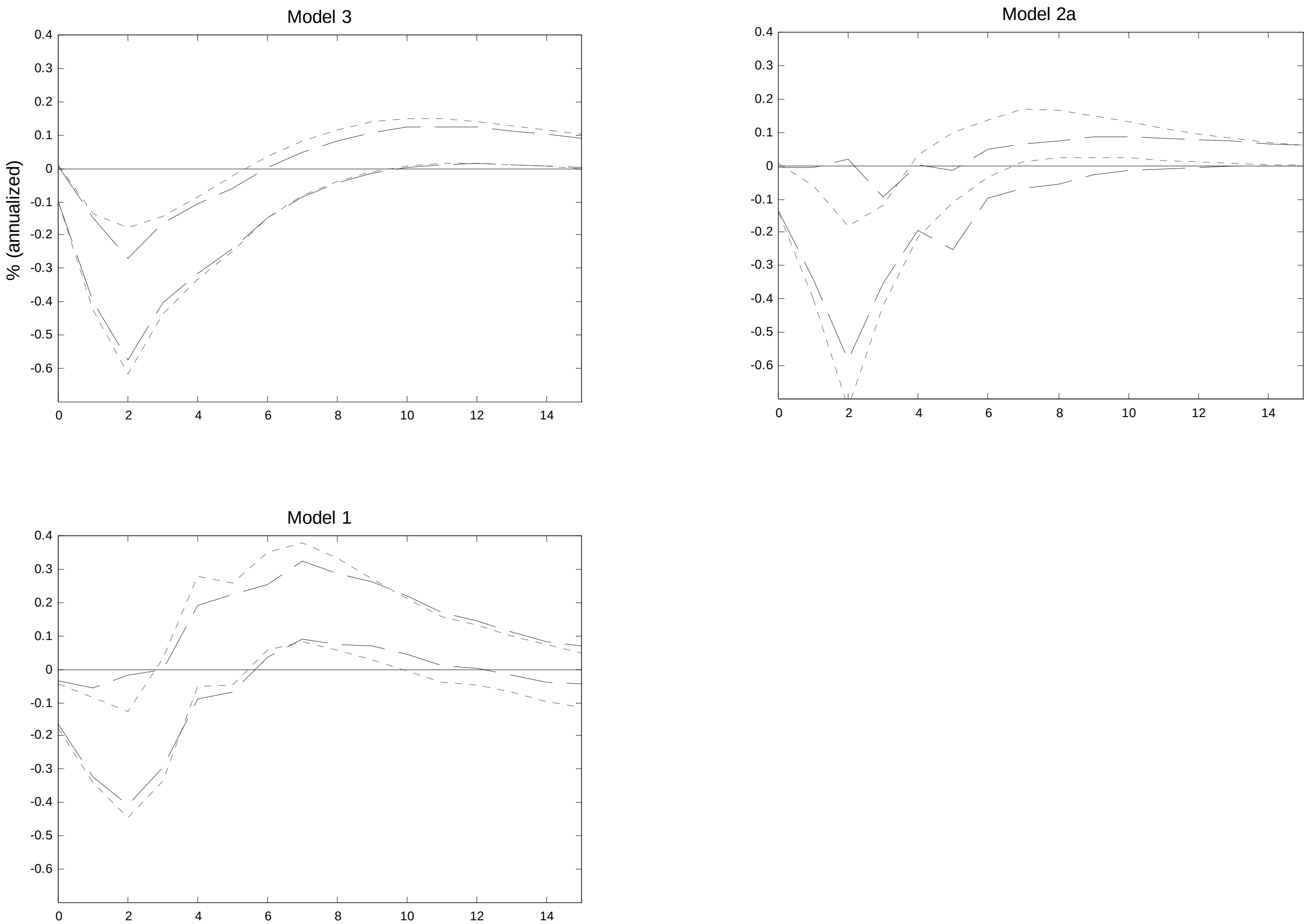


Chart 10: CPI inflation response to policy shock (59-19 and 59-97)