TAKEHOME FINAL EXAM
PART 2: 135 POINTS

There are four questions, numbered 2 to 5. Answer all questions, and keep your answers for these questions on separate sheets from the answers to question 1 (from part 1 of the exam).

(2) (45 points) Consider the following simple macro model:

$$\text{Taylor rule:} \qquad r_t = \alpha_0 + \alpha_1 \pi_{t-1} + \alpha_2 y_{t-1} + \varepsilon_t \tag{2.1}$$

$$\text{Phillips curve:} \qquad \pi_t = \phi_0 + \phi_1 \pi_{t-1} + \phi_2 y_{t-1} + \nu_t \tag{2.2}$$

$$\text{Fisher relation:} \qquad r_t = E_t \pi_{t+1} + \rho_0 + \xi_t\,. \tag{2.3}$$

In the Phillips curve and the Fisher relation the exogenous disturbances ν_t and ξ_t are assumed to be i.i.d. normal, with mean zero. In the Taylor rule, it is thought that the exogenous disturbance ε_t term might be serially correlated, though it is stationary and Gaussian. The three disturbances are mutually independent at all leads and lags. Their variances are unknown. The E_t operator is expectation conditioned on the values of all random variables in the system dated t and earlier.

(a) Show how to check, in the general case, whether this model implies a stationary process for r, π and y. [Hint: It may help to begin by using the Phillips curve to eliminate the expected inflation term, then reduce the system to one in π and y alone.]

(10 points)

Using the Phillips Curve to get $E_t[\pi_{t+1}]$, we can replace (2.3) by

$$r_t = \phi_0 + \phi_1 \pi_t + \phi_2 y_t + \rho_0 + \xi_t\,.$$

With this replacement, the system in matrix form is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -\phi_1 & -\phi_2 \end{bmatrix} \begin{bmatrix} r_t \\ \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \phi_0 \\ \phi_0 + \rho_0 \end{bmatrix} + \begin{bmatrix} 0 & \alpha_1 & \alpha_2 \\ 0 & \phi_1 & \phi_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \pi_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \nu_t \\ \xi_t \end{bmatrix}.$$

If we call the 3×3 matrices on the left and right sides of this equation Γ_0 and Γ_1, respectively, then stability of the system will hold if and only if

$$\Gamma_0^{-1}\Gamma_1$$

has all its eigenvalues less than one in absolute value.

(b) Check whether the system is stationary with $\phi_1 = .6$, $\phi_2 = .3$, $\alpha_1 = 1.2$, and $\alpha_2 = .2$.

(5 points)

$$\Gamma_0^{-1}\Gamma_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 10/3 & -2 & -10/3 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1.2 & .2 \\ 0 & .6 & .3 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1.2 & .2 \\ 0 & .6 & .3 \\ 0 & 2.8 & .067 \end{bmatrix}$$

The eigenvalues of this matrix are 1.29 and -.62, so the system is not stable at these parameter values. [With a purely backward-looking Phillips curve like this, the usual prescription that α_1 should exceed 1 tends to produce non-existence of a stable equilibrium. Here an α_1 around .7 is small enough to produce a stable equilibrium solution.]

(c) Can the Taylor rule coefficients be estimated consistently by applying GLS to the single Taylor rule equation? Either prove that this is always or never possible, or else give conditions on the parameter values that make it possible or impossible.

(10 points)

This question was trickier than I meant it to be. In lectures I emphasized that strict exogeneity is generally a precondition for applying GLS to obtain efficient estimates of a regression. This is true, however, when we start with a regression equation that can be consistently estimated by OLS and simply want to get efficient estimates by accounting for serial correlation. For example, if our equation is $y_t = \psi_0 + \psi_1 f_{t-4} + \upsilon_t$, with f_{t-4} a four-period ahead forecast of y_t, then we expect the errors to be serially correlated and rationality of the forecasts implies that $E[\upsilon_t \mid f_{t-4}] = 0$, so OLS is consistent, and GLS is not.

The model specified here implies that OLS applied to the reaction function is *not* consistent, and it also implies that GLS will be consistent. The filtering to eliminate serial correlation that is implicit in GLS will make the right-hand-side variables predetermined (not strictly exogenous) in the filtered equation.

To make matters still more confusing, in this model y and π are GCP to r. This means there is an equation with r on the left and *current* and lagged y's and π's on the right in which the right-hand-side variables are exogenous. This equation is not the Taylor rule equation, however, because the Taylor rule excludes current y's and π's from the right-hand side, according to the model. Of course the fact that this other regression equation, with exogenous right-hand side variables, exists might be a reason to question whether we are sure that it is appropriate to exclude current π and y from the Taylor rule.

To understand these points, consider the reduced form of the original system obtained by multiplying it through by Γ_0^{-1}:

$$\begin{bmatrix} r_t \\ \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \phi_0 \\ \dfrac{\alpha_0 - \phi_1\phi_0 - \rho_0 - \phi_0}{\phi_2} \end{bmatrix} + \begin{bmatrix} 0 & \alpha_1 & \alpha_2 \\ 0 & \phi_1 & \phi_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \pi_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \nu_t \\ \dfrac{\varepsilon_t - \xi_t - \phi_1\nu_t}{\phi_2} \end{bmatrix} . \qquad (*)$$

So long as the system is stable, this AR representation can be converted to an MA representation in which all variables, including y_t and π_t, are functions of the disturbances dated t and earlier. In particular, y_{t-1} will contain ε_{t-1} as a component. If ε_t is serially correlated, therefore, in general the Taylor rule residual is correlated with at least one of the right-hand-side variables so OLS estimates of it will not be consistent. There are probably complicated conditions on the model parameters that would eliminate the correlation in some hairline special cases, but you were not expected to derive them.

To see that GLS will give consistent estimates, consider the case where serial correlation is first-order, i.e. $\varepsilon_t = \theta\varepsilon_{t-1} + \zeta_t$. Multiplying the Taylor rule equation through by $(1-\theta L)$ produces

$$r_t = \theta r_{t-1} + (1-\theta)\alpha_0 + \alpha_1\pi_{t-1} - \theta\alpha_1\pi_{t-2} + \alpha_2 y_{t-1} - \theta\alpha_2 y_{t-2} + \zeta_t .$$

In this equation all the right-hand-side variables are uncorrelated with the residual ζ_t, because ζ_t itself is uncorrelated with lagged values of any of the disturbances. Being a linear combination of ε_t and ε_{t-1}, it is uncorrelated with ν_t and ξ_t by assumption in the problem statement, and it is uncorrelated with lagged ε's by our assumption on the form of serial correlation. So this equation can be estimated consistently by nonlinear least squares. GLS is asymptotically equivalent to estimating such a "filtered" equation by OLS, if the estimated serial correlation coefficient used to form the covariance matrix for GLS is iteratively updated until convergence.

To see that π and y are GCP to r, note that to get the system $(*)$ into its full AR form, with innovations as residuals, we have to filter the residuals in the first and third equations. The first has an AR residual with parameter θ, and the third has an ARMA(1,1) residual with AR parameter θ and some MA parameter γ. The full AR representation, then, is

$$\begin{bmatrix} 1-\theta L & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & \dfrac{1-\theta L}{1-\gamma L} \end{bmatrix} (I - \Phi L)x_t = \eta_t , \qquad (\ddagger)$$

where $x = \begin{bmatrix} r & \pi & y \end{bmatrix}'$, $\Phi = \Gamma_0^{-1}\Gamma_1$, and η is the innovation vector. Because Φ is block triangular, with π and y in one block and r in the other, the full

AR system is block triangular in the same pattern, which we know is the condition for π and y to be jointly GCP to r.

The GCP condition and strict exogeneity are related: GCP $\Rightarrow$ strict exogeneity. But they are not equivalent. GCP of π and y here implies there is some equation in which these two variables are strictly exogeneous and r is on the left, but it is not this system's Taylor rule, as we have already verified.

If one wanted to find the equation in which π and y are exogenous (which was not at all expected for the exam), one would find the coefficients in the regression of $\zeta_t = \eta_{1t}$ on η_{2t} and η_{3t}, then take the corresponding linear combination of the equations of the AR representation (‡).

(d) Suggest a way to transform the system into one that has no serial correlation in disturbances. You will have to propose a model for the serial correlation in the Taylor rule disturbance.

(5 points)

This was already done above in the answer to part (2c)

(e) Are all the parameters in the equation system identified? If not, which are and which aren't?

(8 points)

Since all three ϕ_i's can be estimated by applying OLS to the Phillips curve, they are identified. As we noted above, the Taylor rule can be estimated by GLS or by NLLS applied to the equation transformed to get rid of serial correlation. The innovation variances in the first two equations can then be estimated as sample averages of the squared equation residuals. Finally, once the Phillips curve is estimated we can substitute its right-hand-side into the Fisher equation, allowing us to estimate ρ_0 and $\mathrm{Var}(\xi_t)$ as the mean and variance of $r_t - E_t\pi_{t+1}$. So all the parameters of the system are identified.

(f) Suggest an optimal method for estimating the identified parameters of the system. Describe your method in as much detail as you can, taking account of the specific nature of this model.

(7 points)

The best idea is to apply maximum likelihood, or else take a posterior mean with a conjugate prior. In either case, one needs to form the log likelihood function, which has the form

$$T\log|\Gamma_0| - T(\log\sigma_\xi + \log\sigma_\nu + \log\sigma_\zeta) - \tfrac{1}{2}\,\mathrm{trace}(\Delta^{-1}\Gamma_0 S_T \Gamma_0' \Delta^{-1})\,,$$

where Δ is the diagonal matrix with σ_ζ, σ_ν, σ_ξ down the diagonal and S_T is the sample cross product of residuals matrix. Of course S_T is a function of the equation parameters, as is Γ_0. This assumes that the system has been transformed to the full AR form (‡), eliminating serial correlation. The model has restrictions on coefficients on lagged values and is overidentified, with fewer parameters than there are free coefficients in the reduced form. So

OLS will not give efficient estimators, and a direct maximization of the likelihood by numerical methods is probably the best course for MLE estimation.

(3) (45 points) It has been proposed that the US underwent a sudden slowdown its long run growth rate, perhaps in the late 70's, and also that the volatility of economic growth — the variance of its one-step-ahead forecast errors — under went a sharp reduction, probably some time in the 80's. (Though it is sometimes also proposed that in the late 90's the trend growth rate rose again, there is less consensus on this and we are going to ignore it here.)

(a) Assume the output growth rate y_t satisfies

$$y_t = \rho y_{t-1} + c_t + \sigma_t \varepsilon_t \,, \tag{3.1}$$

where $\varepsilon_t \sim N(0,1)$ is independent of $\{y_s,\, s<t; c_v, \sigma_v, \text{ all } v\}$. Formulate a hidden Markov chain model that captures the idea of single, randomly timed, transitions in the trend growth rate (indexed by c_t) and the volatility (indexed by σ_t). The model should imply that the timing of the two transitions is independent. Both c and σ should be modeled as constant except at their respective transition dates. The uniqueness and independence of the two transition dates can be enforced by appropriate restrictions on the form of the Markov transition matrix.

(12 points)
We need four states, in which the parameter values are as follows:

	ρ	σ	c
1	ρ	$\bar{\sigma}$	$\bar{c}$
2	ρ	$\bar{\sigma}$	$\hat{c}$
3	ρ	$\hat{\sigma}$	$\bar{c}$
4	ρ	$\hat{\sigma}$	$\hat{c}$

The transition matrix is

$$\begin{bmatrix} (1-\pi_1)(1-\pi_2) & (1-\pi_1)\pi_2 & \pi_1(1-\pi_2) & \pi_1\pi_2 \\ 0 & (1-\pi_1) & 0 & \pi_1 \\ 0 & 0 & (1-\pi_2) & \pi_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The block triangularity of this transition matrix insures that the state index always increases. The states that precede the change in σ are 1 and 2. Conditional on either of these states as initial state, the probability of going to one of the states (3 or 4) with the changed σ is π_1. Similarly, the conditional probability of going from state 1 or 3 to state 2 or 4 is π_2. In other words, the probability of a change in σ is independent of whether the change in c has already occurred, and vice versa. The 4×4 transition matrix has only two free parameters.

(b) Does the fact that each transition is modeled as occurring only once imply that the likelihood is ill-behaved? Why or why not?

(9 points)

Depends what you mean by ill-behaved. The joint pdf of the data and the sequence of states $\{S_t\}$ has the π's entering only in a factor of the form

$$(1-\pi_1)^{T_1}\pi_1(1-\pi_2)^{T_2}\pi_2\,, \tag{†}$$

where T_1 is the number of periods before the jump in σ and T_2 is the number of periods before the jump in c. These terms are log-concave and have unique maxima at $1/(T_i+1)$. Integrating them over the uncertainty in the states will not introduce any singularities. However, there will be no tendency for the likelihood to concentrate increasingly sharply on unique values of T_1 and T_2 as sample size increases. This is because once the transitions occur, further observations distant in time from the transition date contain little further information about the transition date. Mathematically, we can see that if the transition dates were known with precision, there would be little uncertainty about the sequence of states, and therefore the likelihood as a function of T_1 and T_2 would be almost exactly of the form (†). But this form is does *not* grow more sharply peaked in the T_i as sample size increases, contradicting the initial assumption that the dates are known with precision.

(c) Suggest a reasonable prior for the parameters in your model and describe in detail how you would compute the peak of the posterior pdf, taking account of the specific nature of this model. Assume that the two initial observations are uninformative (i.e., condition on the initial observations).

(12 points)

It will be convenient to make the prior on the regression parameters, conditional on the state, of the conjugate normal-inverse-gamma form. The transition probabilities π_1 and π_2 could naturally be given independent Beta priors. It would be reasonable to use a prior that keeps these parameters away from 0 and 1, where the model ceases to make sense, so a form proportional to $\pi_i(1-\pi_i)$ (a $\mathrm{Beta}(2,2)$ prior) would be OK.

This is a hidden Markov Chain model, so it is possible to compute the likelihood of the sample, with the state sequence integrated out, by the usual recursive methods. That is, one uses the facts (with $\beta=(\rho,\{c(S),\sigma(S),S=1,\dots,4\})$:

$$\begin{aligned}
p(y_t,S_t\mid Y_{t-1};\beta) &= p(y_t\mid S_t,Y_{t-1};\beta)\,p(S_t\mid Y_{t-1};\beta)\\
p(y_t\mid Y_{t-1};\beta) &= \sum_{S_t=1,\dots,4} p(y_t,S_t\mid Y_{t-1};\beta)\\
p(S_t\mid Y_{t-1};\beta) &= \sum_{S_{t-1}=1,\dots,4} p(S_t\mid S_{t-1})P(S_{t-1}\mid Y_{t-1};\beta)\\
p(S_t\mid Y_t;\beta) &= \frac{p(y_t,S_t\mid Y_{t-1};\beta)}{\sum_{S_t=1,\dots,4} p(y_t,S_t\mid Y_{t-1};\beta)}\\
p(y_t\mid S_t,Y_{t-1};\beta) &= \phi(y_t;c(S_t)+\rho y_{t-1},\sigma(S_t)^2)\,.
\end{aligned}$$

These formulas let one recursively compute the likelihood as a function of β, $\prod_t p(y_t \mid Y_{t-1};\beta)$. Then to maximize the likelihood one would use a numerical optimization program.

(d) Explain how you would compare this model with one that enforces constancy of c_t and σ_t, assuming that you need posterior odds on the two models in order to guide thinking about economic policy. Give as much detail as you can about the computation, taking account of the specific nature of this model.

(12 points) Even though the constant-parameter model is nested within the hidden Markov chain model, the Schwarz criterion or a traditional likelihood ratio test will not work here. The problem is that the second derivative matrix of the likelihood of the unrestricted model is singular at points in the restricted model. (The parameters of the different states have no effect on fit if there are no transitions.) A direct Bayesian posterior odds calculation is the only practical alternative. This will require careful attention to priors and Monte Carlo methods. The likelihood for the restricted model is easy to handle, because it is just a Gaussian regression model. For the unrestricted model, there are recursive methods for Gibbs sampling that are closely related to the recursive methods for likelihood evaluation. Answers that give detail about the application of Gibbs sampling methods to this model, which are covered in the notes, get credit for doing so.

(4) (15 points) Discuss the following assertions, explaining why they are correct or incorrect:

(a) *The Schwarz criterion can be used to obtain consistent model selection as $T \to \infty$ when comparing models with different numbers of parameters. However in the special case where one of the models is a point, it does not work, because when there are no free parameters the Gaussian approximation to the likelihood, on which the Schwarz criterion is based, does not apply.*

(7 points)

The only reason we need the Gaussian approximation to the likelihood is to integrate out the free parameters. When there are no free parameters, the likelihood of the model is available directly, without any integration, so the Schwarz criterion's "correction" of the likelihood, which is $k \log T$, which comes out zero in this $k = 0$ case, is just right.

(b) *When comparing regression models with dummy-observation priors, posterior odds ratios are determined entirely by the sums of squared residuals calculated by OLS across actual and dummy-observation data.*

(8 points)

While the posterior pdf for the regression parameters is determined entirely by this sum of squared residuals, the posterior odds on the model are not, because the prior pdf corresponding to the dummy observations has a "constant term" that drops out of the posterior pdf for the single model, but affects comparisons between models.

(5) (30 points) We would like to estimate the univariate AR model with constant term

$$y_t = c + \rho y_{t-1} + \varepsilon_t \,,\; t = 1, \ldots, 20\,. \tag{5.1}$$

ε_t is the innovation in y_t and is i.i.d. $N(0, \sigma^2)$, where σ^2, c, and ρ are unknown. Recognizing that use of OLS on such a model tends to give unreasonably large weight to models in which initial conditions "explain too much", we add to the actual 20 data points (21 counting the initial condition) a pair of dummy observations that append to the sample y and $X = [\mathbf{1}\; y_{-1}]$ matrices

$$\tilde{y} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} 1 & 10 \\ 0 & 10 \end{bmatrix}. \tag{5.2}$$

(a) If the moments of the actual data are

$$X'X = \begin{bmatrix} 20 & 235 \\ 235 & 3000 \end{bmatrix}, \quad \sum_{t=1}^{20} y_t y_{t-1} = 3035$$

$$\sum_{t=1}^{20} y_t^2 = 3100, \quad \sum_{t=1}^{20} y_t = 240\,,$$

calculate the OLS estimators with and without the dummy observations.

(10 points)

With the dummy observations, the moment matrices become

$$X'X = \begin{bmatrix} 21 & 245 \\ 245 & 3200 \end{bmatrix} \qquad X'y = \begin{bmatrix} 250 \\ 3235 \end{bmatrix} .$$

The OLS estimates without the dummy observations are $\hat{c} = 1.4189$, $\hat{\rho} = .9005$. With the dummy observations they are $\hat{c} = 1.0348$, $\hat{\rho} = .9317$.

(b) What prior pdf would give the same shape to the posterior as would be obtained by treating the likelihood for the expanded data set as if it were the posterior? Is this prior pdf a proper prior over σ^2, ρ, c? Explain your answer.

(20 points)

The prior mean is

$$(\bar{X})^{-1}\bar{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

The prior variance is

$$\sigma^2(\bar{X}'\bar{X})^{-1} = \sigma^2 \begin{bmatrix} 2 & -.1 \\ -.1 & .01 \end{bmatrix} .$$

If we multiply the likelihood by the pdf of a normal with this mean and variance, we get as log posterior pdf

$$-(T+2)\log\sigma - \frac{T+2}{2}\log(2\pi) + \log|\bar{X}| - \frac{1}{2\sigma^2}(y^* - X^*\beta)'(y^* - X^*\beta) .$$

where y^* and X^* are the expanded data matrices, with dummy observations at the bottom, and $\beta = \begin{bmatrix} c & \rho \end{bmatrix}'$.

Since what we are multiplying by is a pdf for β for every σ value, it integrates to one for every σ and implies a flat prior on σ or σ^2. It is therefore not a proper prior in a model where σ^2 is unknown.