

## Assignment 3

### Problem 1.

The question is motivated by the original work of [Hansen and Singleton](#), contrasting the assumptions and empirical evidence for consumption-based asset pricing obtained via GMM versus maximum likelihood.

I have posted quarterly data in a file labeled `HansenSingletondata.xlsx`, containing time series on U.S. aggregate real private consumption and real value of the S&P500 index covering 1958:Q1 to 2016:Q4.

Consider a representative agent who maximizes utility,

$$\sum_{t=0}^{\infty} \beta^t u(C_t)$$

subject to the budget constraint:

$$W_{t+1} = (W_t - C_t) \cdot R_{t+1}^m$$

where  $C_t$  is real consumption,  $W_t$  is real wealth,  $R_{t+1}^m$  is the gross real return on the market portfolio, which is a weighted sum of individual asset returns. That is,  $R_{t+1}^m = \sum_i \alpha^i R_{t+1}^i$ ,  $\alpha^i$  is the weight of asset  $i$  in the market portfolio, and  $R_{t+1}^i$  is gross real return on asset  $i$ .

Optimization on the part of the agent implies that,

$$u'(C_t) = \mathbb{E}_t(\beta R_{t+1}^m u'(C_{t+1}))$$

The left hand side of the above Euler equation is the increased utility from consuming one more unit of goods today. The right hand side is equal to the expectation of the present value of the increased utility, if the consumer invests one unit in the market portfolio today and consumes the gross return tomorrow. In equilibrium, the equality holds.

Let the utility function be given in terms of power utility (CRRA),

$$u(C_t) = \frac{C_t^{1-\gamma} - 1}{1 - \gamma}$$

Then the above Euler equation becomes,

$$1 = \mathbb{E}_t \left[ R_{t+1}^m \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \right]$$

Rewrite the above equation, letting  $F_t$  denote all information available at time  $t$ ,

$$\mathbb{E}\left(R_{t+1}^m \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} - 1 \middle| F_t\right) = 0 \quad (1)$$

- A) First, we do something unusual, partially to display the versatility of GMM, but also to guide us towards good starting values for a traditional GMM procedure. Equation 1 contains two parameters and translates into one (unconditional) moment condition. You will impose specific values on these parameters and check if the resulting moment condition is close to be satisfied – without estimating any parameters. Simply fix  $\beta = 0.99$  and, for a grid of likely  $\gamma$  values, compute the GMM criterion function. That is, for each value of  $\gamma$ , compute the J-criterion using an efficient GMM procedure. Estimate the spectral density for the moment condition at frequency zero (the long-run variance) using a Bartlett (Newey-West) kernel with a cut-off' parameter that is appropriate given the persistence in the data. Plot the value of the criterion function for the different values of  $\gamma$ . Discuss the results.
- B) Repeat the procedure in question A), but letting  $F_t$  be represented by the instrument set:  $Z_t = (1, C_{t-1}, R_{t-1}^m)$ , i.e., a constant and one-quarter lags of consumption growth  $C_{t-1}$  and real returns. You are now exploiting over-identifying restrictions generated by the model.
- C) Let us now try to perform actual parameter estimation with GMM. You have two free parameters, so you have to include moment conditions generated from (non-constant) instruments. Initially, estimate the system using the instruments in question B). You should now have a sense of reasonable values for the parameter vector, so you should have a good candidate for the original weighting matrix and do not have to estimate it using, say, the identity matrix. Report the parameter estimates, standard errors, and the J-test for the fit of the over-identifying restriction(s).
- D) Expand the instrument set in question C) to include two, three, and four lags of the consumption growth and market return. For each set of (two) additional instruments, repeat the GMM estimation and report the results. In this case, you must obtain estimates for the weighting matrix in an initial step.

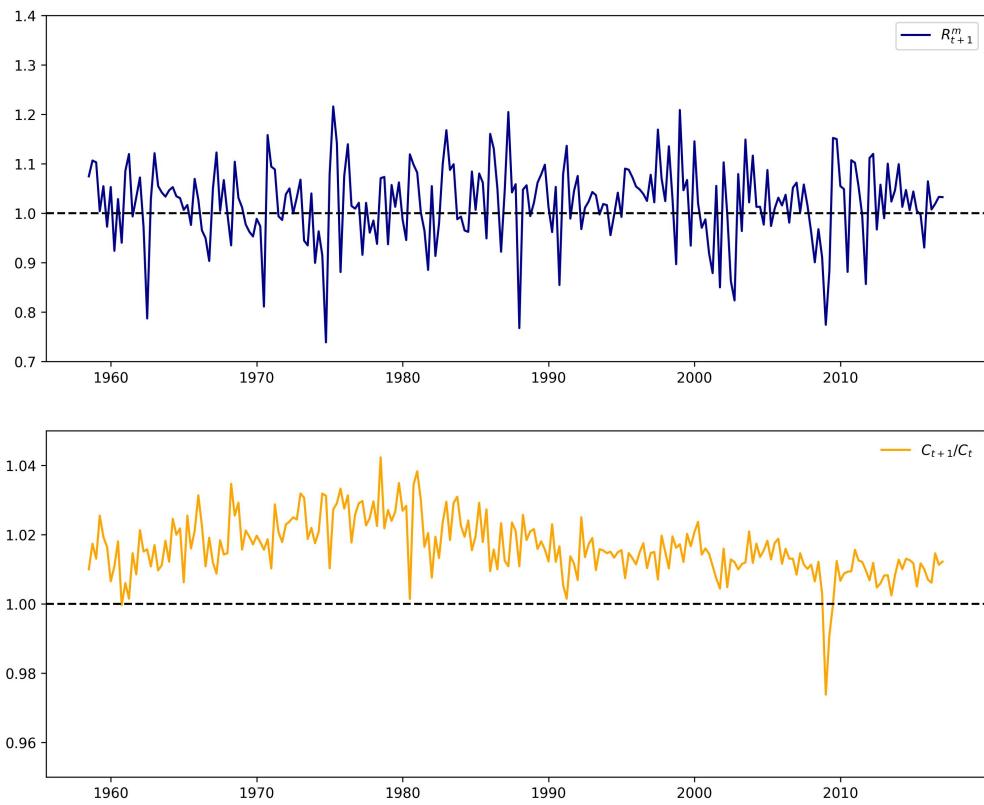
### *Solution.*

- A) The returns are shown in Figure 1: We start by defining some notation for this question. Following the lecture notes we denote

$$m(z_t, \theta) = m((R_{t+1}^m, C_t, C_{t+1}), (\beta, \gamma)) = R_{t+1}^m \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} - 1$$

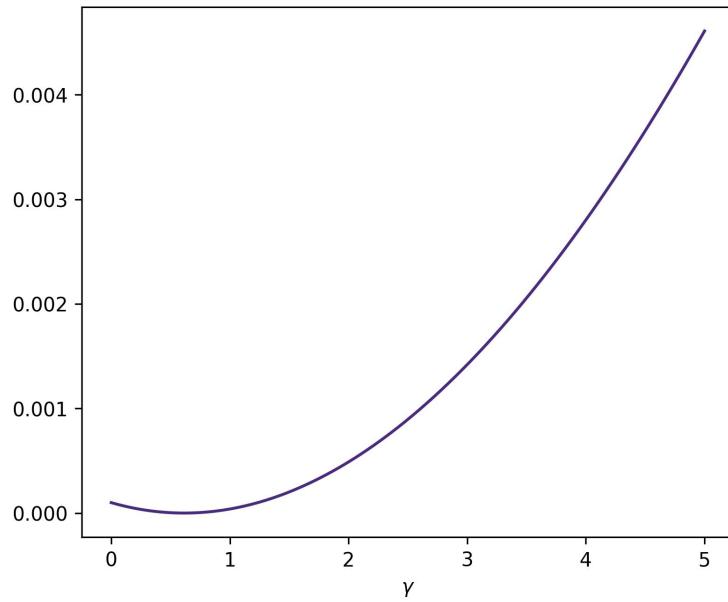
and  $g_t(\theta) = T^{-1} \sum_t m(z_t, \theta)$  and  $\widehat{S}_T(\theta) = T^{-1} \sum_t m(z_t, \theta)^2$ . Therefore the J-criterion of the GMM estimator is given by

$$J_T(\theta) = g_T(\theta)' \widehat{S}_T(\theta)^{-1} g_T(\theta)$$

Figure 1: Series of  $R_t$  and  $C_t$  from 1958:Q1 to 2016:Q4

I then calculate the J-criterion for a grid of  $\gamma$  from 0 to 10 with length 1000. I used both the efficient weighting matrix ( $\widehat{S}_T$ ) and the identity matrix. The results are shown in Figure 2 and the optimal value of gamma seems to be close to 1 for both criteria.

Figure 2: J-criterion for the grid of  $\gamma$



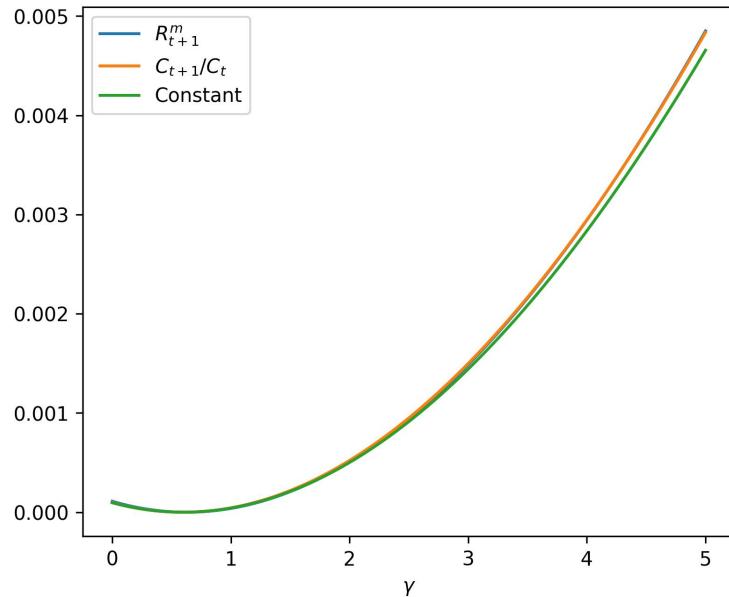
B) Using the conditional moment from Equation 1, we can use the instruments  $Z_t$  to write

$$\mathbb{E}(m(z_t, \theta)Z_t) = 0$$

so letting  $\tilde{m}(z_t, \theta, Z_t) = m(z_t, \theta)Z_t$  we can rewrite the J-criterion and the efficient matrix as

$$\begin{aligned}\widetilde{S}_t(\theta) &= T^{-1} \sum_t \widetilde{m}(z_t, \theta, Z_t) \widetilde{m}(z_t, \theta, Z_t)' = T^{-1} \sum_t m(z_t, \theta)^2 Z_t Z_T' \\ \widetilde{J}_T(\theta) &= g_T(\theta)' \widetilde{S}_T(\theta)^{-1} g_T(\theta)\end{aligned}$$

I then calculate this criteria as in the previous question and display it in Figure 3. The results remain the same.

Figure 3: J-criterion for the grid of  $\gamma$  using the instruments  $Z_t$ 

- C) I estimate the parameters using a the first lag of  $R_t$  and  $C_t$  as instruments as well as a constant. The model is estimated using the minimization of the moment conditions. For the minimization algorithm, I used as initial guess the value of  $\gamma$  that minimized the J-criterion in (a) with  $\beta = 0.99$ . Since these parameters are restricted ( $\gamma > 0, \beta \in (0, 1)$ ) I reparametrized the model to allow the algorithm to search over the entire real line which facilitates the minimization. Therefore the model is taken with

$$\gamma = \frac{1}{1 + \exp^{-\tilde{\gamma}}} \quad \beta = \tilde{\beta}^2$$

Initially I tried to perform a two-step approach to estimate the parameters using the identity weighting matrix in the first step and using these estimates to calculate the efficient weighting matrix by calculating  $\hat{S}(\beta, \gamma)$ . This however did not turn out so well as the determinant of the estimated  $S$  was nearly zero, making the second-step estimate very volatile. I then kept the first stage as my final results which are reported below with the standard deviations in parenthesis:

$$\begin{array}{ll} \beta = & 0.9900 \\ & (0.0154) \end{array} \quad \begin{array}{ll} \gamma = & 0.6084 \\ & (0.9286) \end{array}$$

As we can see, the estimates of  $\gamma$  are close to the ones guessed on item (a) and  $\beta$  is really close to 0. We also note that the variance of  $\gamma$  is really high.

To understand the stability of the estimates, I reestimate the model using a set of 10.000 random initial guesses and store the solutions for each one of them. The distribution of these solutions are shown in Figure 4. The results indicate that there may be two local solutions based on the modes of the marginal distribution in Figure 4. This presents a challenge to the estimation since it

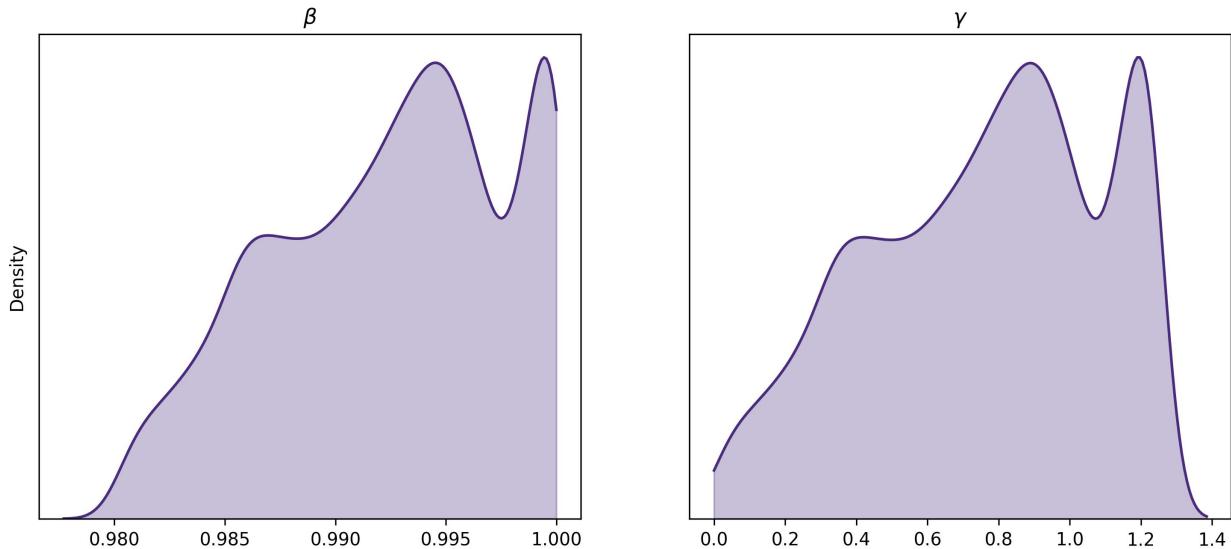


Figure 4: Solutions for 10,000 random initial guesses

I now perform the Hausman test of overidentified restrictions whose null hypothesis is that a subset of the instruments is exogenous. If that is the case, then excluding (some) of the instruments would lead to a more efficient estimator. We analyze this by comparing the moment condition of both estimates. Under the null, a subset of the instruments can be used in a just-identified model to efficiently estimate the parameters. Under a just-identified model, the moment condition is perfectly matched and has zero average in the sample. Also, using the results of the previous problem set, we can compare an efficient and a consistent (but not efficient) estimator using the Hausman test. The test statistic is given by:

$$\mathcal{H} = T \times \left[ \frac{1}{T} \sum_{t=1}^T m(z_t, \hat{\theta}) \right]' A \left[ \frac{1}{T} \sum_{t=1}^T m(z_t, \hat{\theta}) \right]$$

The results are shown below:

$$\chi^2 = 4.5339 \quad \text{p-value} = 0.9668$$

As we can see, the test is not able to reject the null so there may be exogenous instruments.

- D) I calculate the model using 1, 2, 3 and 4 lags of the instruments. The estimates are shown below:

|               | Number of Lags   |                  |                  |                  |
|---------------|------------------|------------------|------------------|------------------|
|               | (1)              | (2)              | (3)              | (4)              |
| $\beta$       | 0.990<br>(0.016) | 0.990<br>(0.014) | 0.990<br>(0.014) | 0.990<br>(0.013) |
| $\gamma$      | 0.575<br>(0.941) | 0.566<br>(0.835) | 0.563<br>(0.846) | 0.563<br>(0.806) |
| N             | 231              | 231              | 231              | 231              |
| $\mathcal{H}$ | 1.922            | 6.366            | 6.493            | 7.137            |
| p-value       | 0.834            | 0.988            | 0.989            | 0.992            |

The results are pretty similar all the set of instruments used and the interpretations of all the tests remain the same. We note that, as we increase the number of instruments, the estimation of  $\gamma$  gets more precise, judging by the decay in the standard deviations.

□

**Problem 2.**

(Stambaugh, 1999). You will examine the bias in the slope coefficient in the regression of returns on past dividend yields, when the dividend yield process is highly persistent. Let  $y_t$  denote the return and  $x_t$  the dividend-price ratio (dividend yield) at date  $t$ . Consider the following regression equations,

$$\begin{aligned}y_t &= \alpha + \beta x_{t-1} + u_t \\x_t &= \theta + \rho x_{t-1} + v_t\end{aligned}$$

for  $t = 1, \dots, T$  and where,

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N(\mathbf{0}_2, \Sigma)$$

We can write the above equations in matrix form as follows,

$$\begin{aligned}\mathbf{y} &= \alpha + \beta \mathbf{x} + \mathbf{u} \\ \mathbf{x}^+ &= \theta + \rho \mathbf{x} + \mathbf{v}\end{aligned}$$

where,

$$\begin{aligned}\mathbf{y} &= [y_1 \ \dots \ y_T]' \\ \mathbf{x} &= [x_0 \ \dots \ x_{T-1}]' \\ \mathbf{x}^+ &= [x_1 \ \dots \ x_T]' \\ \mathbf{u} &= [u_1 \ \dots \ u_T]' \\ \mathbf{v} &= [v_1 \ \dots \ v_T]'\end{aligned}$$

(a) Show that

$$\hat{\beta} = \beta + [\mathbf{0} \ \mathbf{1}] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}$$

where  $\mathbf{X} = [\mathbf{1}_T \ \mathbf{x}]$

(b) Show that  $\mathbf{u}$  can be decomposed as  $\mathbf{u} = \frac{\sigma_{uv}}{\sigma_v^2} \mathbf{v} + \mathbf{e}$  where  $\mathbf{e}$  is uncorrelated with  $v$ , and,

$$\hat{\beta} = \beta + \frac{\sigma_{uv}}{\sigma_v^2} [\mathbf{0} \ \mathbf{1}] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v} + [\mathbf{0} \ \mathbf{1}] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{e}$$

(c) Show that,

$$\hat{\rho} = \rho + [\mathbf{0} \ \mathbf{1}] (\mathbf{X}\mathbf{X})^{-1} \mathbf{X}\mathbf{v}$$

(d) Show that

$$\mathbb{E}(\hat{\beta} - \beta) = \frac{\sigma_{uv}}{\sigma_v^2} \mathbb{E}(\hat{\rho} - \rho)$$

and, using the expression  $\mathbb{E}(\hat{\rho} - \rho) \approx -\frac{1+3\rho}{T}$ , for the finite sample bias of the estimated AR(1) coefficient from Kendall (1954), discuss the potential magnitude of the bias in  $\hat{\beta}$  above, when  $T$  is 50 years,  $\rho = 0.94$  and  $\frac{\sigma_{uv}}{\sigma_v^2} = -1$ .

- (e) Try to interpret the implications of your results for predictive regressions more generally. The issues highlighted here have generated a large literature!
- (f) Design a small Monte Carlo study to explore the size of bias in finite samples by experimenting with different values for sample size,  $T$ , the degree of persistence in the regressor,  $\rho$ , and the (normalized) covariance between the innovations,  $\sigma_{uv}$  (or  $\sigma_{uv}/\sigma_v^2$ ). Please report some relevant findings.

**Solution.**

- (a) Since

$$\mathbf{y}_t = \alpha + \beta \mathbf{x} + \mathbf{u} = \theta \mathbf{X} + \mathbf{u}$$

Where  $\mathbf{X} = [\mathbf{1}_T \quad \mathbf{x}]$  and  $\theta = (\alpha, \beta)$ . Then the usual OLS estimator of  $\theta$  is given by

$$\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Therefore  $\hat{\beta} = [0 \quad 1] \hat{\theta}$  and

$$\begin{aligned}\hat{\beta} &= [0 \quad 1] \hat{\theta} \\ &= [0 \quad 1] [\theta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}\mathbf{u}] \\ &= \beta + [0 \quad 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}\end{aligned}$$

- (b) Since  $u_t$  and  $v_t$  are jointly normal, then we can write that

$$u_t = \gamma v_t + e_t$$

where  $e_t$  is uncorrelated with  $v_t$  and  $\gamma$  is given by the OLS coefficient  $\gamma = \frac{\sigma_{uv}}{\sigma_v^2}$ . Therefore we can rewrite the equation from the previous item as:

$$\begin{aligned}\hat{\beta} &= \beta + [0 \quad 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} \\ &= \beta + [0 \quad 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' (\gamma \mathbf{v} + \mathbf{e}) \\ &= \beta + \frac{\sigma_{uv}}{\sigma_v^2} [0 \quad 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v} + [0 \quad 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{e}\end{aligned}$$

- (c) Follows from the same logic as in the first item.  
(d) Combining item (b) and (c) we can write that

$$\hat{\beta} - \beta = \frac{\sigma_{uv}}{\sigma_v^2} [0 \quad 1] (\hat{\rho} - \rho) + [0 \quad 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}\mathbf{e}$$

Then using the assumption that  $e_t$  is independent from  $\mathcal{F}_{t-1}$  (the information from the previous period, which contains the lagged  $\mathbf{X}$ ), we can take the conditional expectation

$$\begin{aligned}\mathbb{E}_{t-1} [\hat{\beta} - \beta] &= \frac{\sigma_{uv}}{\sigma_v^2} [0 \ 1] \mathbb{E}_{t-1} [\hat{\rho} - \rho] + [0 \ 1] (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X} \mathbb{E}_{t-1} [\mathbf{e}] \\ \mathbb{E}_{t-1} [\hat{\beta} - \beta] &= \frac{\sigma_{uv}}{\sigma_v^2} [0 \ 1] \mathbb{E}_{t-1} [\hat{\rho} - \rho] \\ \mathbb{E}(\hat{\beta} - \beta) &= \frac{\sigma_{uv}}{\sigma_v^2} \mathbb{E}(\hat{\rho} - \rho)\end{aligned}$$

Using the expression from Kendall (1954) that implies

$$\mathbb{E}(\hat{\beta} - \beta) \approx -\frac{\sigma_{uv}}{\sigma_v^2} \frac{1 + 3\rho}{T}$$

Using the proposed values of  $T = 50$ ,  $\rho = 0.94$  and  $\frac{\sigma_{uv}}{\sigma_v^2} = -1$  we would have a bias of  $\hat{\beta}$  equal to 0.0764. Therefore, just from the persistence of the regressor, our estimator would be biased.

- (e) This results suggest that the persistence of the regressors can be a real enemy when estimating a predictive regression. As we know, the more persistent the  $x$  ( $\rho \rightarrow 1$ ) the harder it gets to estimate this parameter and the higher the estimation bias of the AR coefficients. This feeds back to the predictive equation by also increasing the bias of the estimated  $\beta$ . This is a problem because we are interested in the predictive power of the regressor, but the more persistent it is, the harder it gets to estimate it and the more biased the estimated  $\beta$  will be. The interpretation behind this is that, as  $x$  gets more persistent, shocks in the predictor tend to last longer, which makes it harder to estimate the real impact of  $x$  on  $y$ .
- (f) To start, I fix  $\beta = 1$  and  $\alpha = 0$ . The errors  $u$  and  $v$  are set to have unit variance. Then I simulate the model varying the levels of  $T, \sigma_{uv}, \rho$ :

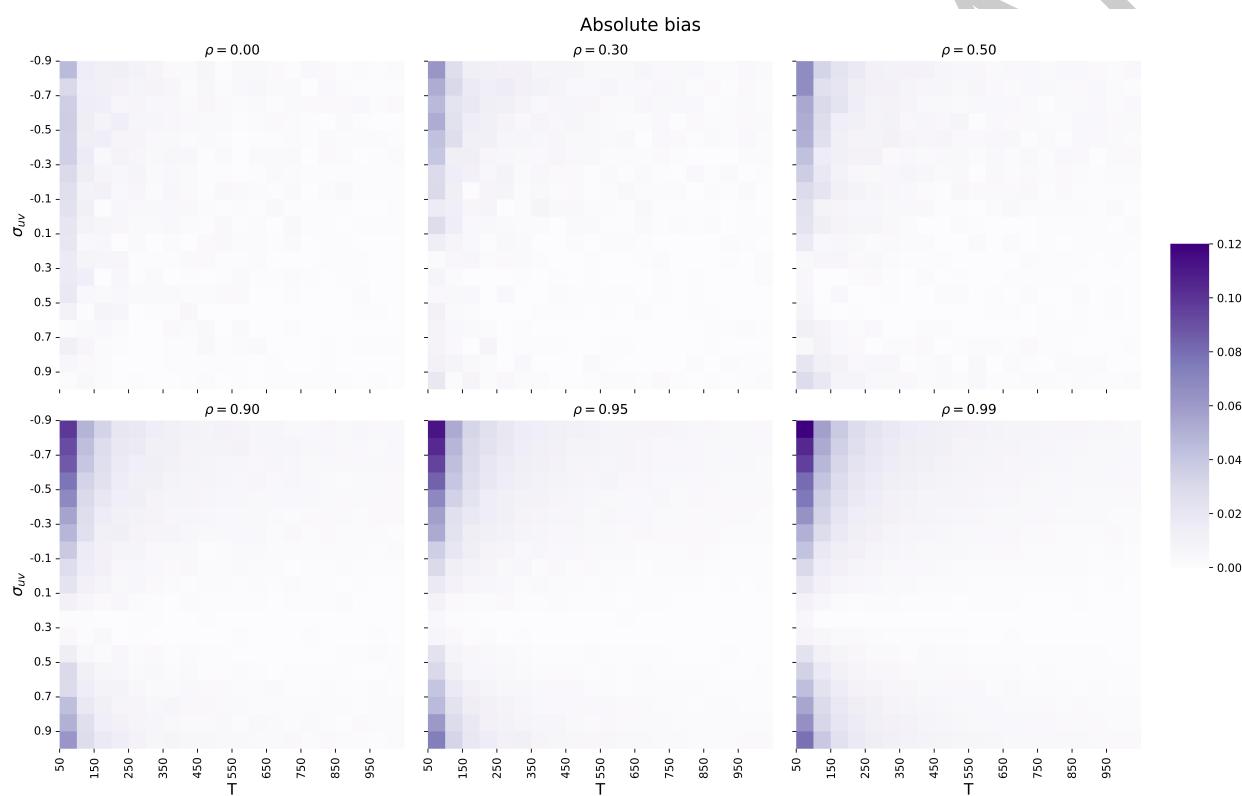
$$T \in \{50, 100, 150, \dots, 1000\}$$

$$\sigma_{uv} \in \{-0.9, -0.8, \dots, 0.8, 0.9\}$$

$$\rho \in \{0, 0.3, 0.5, 0.9, 0.95, 0.99\}$$

Every model is simulated 10.000 times and I calculate the average bias of  $\hat{\beta}$  among these simulations. The results are shown in Figure 5 which displays the absolute level of the bias for each level of  $\rho, \sigma_{uv}$  and  $T$ . As we can see, the bias dies out quickly with  $T$ , but it is very sensitive to the level of persistence of the regressor. The higher the persistence, the higher the bias and the longer it takes to die out. The same happens with the covariance between the errors. The higher the covariance, the higher the bias and the longer it takes to die out. For higher levels of  $\rho$  such as 0.99, we see that the bias is still big for sample sizes around 150, which can sometimes be considered big.

□

Figure 5: Bias of the OLS estimator under different levels of  $\rho$ ,  $\sigma_{uv}$  and  $T$

**References**

- L. P. Hansen and K. J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, pages 1269–1286, 1982.
- M. G. Kendall. Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3-4): 403–404, 1954.
- R. F. Stambaugh. Predictive regressions. *Journal of financial economics*, 54(3):375–421, 1999.