



# Working Paper Series

## Efficient VAR Discretization

**WP 20-06**

Grey Gordon  
Federal Reserve Bank of Richmond

This paper can be downloaded without charge  
from: <http://www.richmondfed.org/publications/>



Richmond • Baltimore • Charlotte

# Efficient VAR Discretization

Grey Gordon\*

Research Department

Federal Reserve Bank of Richmond

June 5, 2020

First version: February, 2019

## Abstract

The standard approach to discretizing VARs uses tensor grids. However, when the VAR components exhibit significant unconditional correlations or when there are more than a few variables, this approach creates large inefficiencies because some discretized states will be visited with only vanishingly small probability. I propose pruning these low-probability states, thereby constructing an efficient grid. I investigate how much an efficient grid improves accuracy in the context of an AR(2) model and a small-scale New Keynesian model featuring four shocks. In both contexts, the efficient grid vastly increases accuracy.

**JEL Codes:** C32, C63, E32, E52

**Keywords:** VAR, Autoregressive, Discretization, New Keynesian

## 1 Introduction

VAR(1) models provide a flexible way of modeling dynamics that encompasses both AR( $p$ ) models and VAR( $p$ ) models. This note shows how to efficiently discretize VARs building on work by [Tauchen \(1986\)](#). Specifically, Tauchen proposed a tensor-grid-based method for discretizing VARs. This approach, however, is inefficient because the VAR variables will in many cases exhibit large unconditional correlations, which means some of the discretized tensor-grid states will only occur with vanishingly small probability. I propose dropping low-probability states to gain efficiency. Specifically, given some method of generating a Markov chain with  $N$  states and some target number of states  $\bar{N}$ , I propose progressively increasing  $N$  and dropping low-probability states until the number of states left,  $N^*$ , is close to  $\bar{N}$ , producing an *efficient grid*.

I assess how large the efficiency gains are, from both statistical and computational perspectives, in two contexts. First, I consider the discretization of an AR(2) model (cast into VAR(1) form). Statistically, the estimated persistence matrix can be up to four orders of magnitude more accurate with

---

\*Email: [greygordon@gmail.com](mailto:greygordon@gmail.com). I thank Christian Matthes, Gabriel Mihalache, and Alex Wolman for helpful comments. I also thank Emma Yeager for excellent research assistance. Any mistakes are my own. The views expressed are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

efficient grids than with the [Tauchen \(1986\)](#) discretization (which uses tensor grids). The estimated variance (kurtosis) of the innovations achieves smaller gains of up to one (three) orders of magnitude, but are still very large. The largest gains occur at high degrees of autocorrelation. Computationally, I assess the accuracy using Euler-equation errors in an overlapping-generations model, which has aggregate shocks following an AR(2), that admits an analytic solution. The Euler errors are reduced by up to an order of magnitude, and again the gains are largest at high levels of autocorrelation. Consistent with these findings, the gains achieved for an AR(2) fitted to Spanish GDP are very large.

The second context is in a small-scale New Keynesian (NK) model with and without a zero lower bound (ZLB). The model has four shocks (productivity, demand, monetary policy, and government spending), and I consider the case of uncorrelated shocks and the case of correlated shocks, the latter coming from an estimation. Even in the uncorrelated case, efficient grids drastically improve on tensor grids because the latter features individual states that are not unlikely but whose joint probability is extremely low. A 1% probability demand shock paired with a 1% probability productivity shock occurs with only 0.01% probability. With four shocks, there are many of these low probability combinations. I show the nonlinearly solved NK model with an efficient-grid rivals the performance of a third-order perturbation, while the tensor-grid solution is terribly inaccurate.

While a fair amount of work has investigated AR(1) discretization, much less attention has been given to VAR discretization. In particular, [Tauchen \(1986\)](#), [Tauchen and Hussey \(1991\)](#), [Adda and Cooper \(2003\)](#), [Flodén \(2008\)](#), and [Kopecky and Suen \(2010\)](#) all discuss AR(1) discretization, but of these, only [Tauchen \(1986\)](#) and [Tauchen and Hussey \(1991\)](#) discuss how to discretize a VAR. [Terry and Knotek II \(2011\)](#) is one of the few papers that investigate VAR approximations, and they show one can work directly with correlated shocks using quadrature techniques (specifically, by using quadrature methods from [Genz, 1992](#) and [Genz and Kwong, 2000](#)) rather than first doing linear transformations to eliminate correlation in the shocks (as is required in the [Tauchen, 1986](#) procedure). While I have used the [Tauchen \(1986\)](#) procedure as a baseline approach for discretizing VARs, the methodology proposed in this paper applies to [Tauchen and Hussey \(1991\)](#), [Terry and Knotek II \(2011\)](#), or any other method.

My proposed approach is tangentially related to the grid-selection method in [Maliar and Maliar \(2015\)](#), which uses a first-stage simulation of points to identify some “representative” points. Here, the unconditional density of the VAR states allows high-probability states to be identified a priori without simulation. However, the similarities between [Maliar and Maliar \(2015\)](#) and this paper end at choosing states because they do not discuss discretizing VARs nor the construction of transition probabilities.

The paper is organized as follows. Section 2 gives the standard, tensor-grid approach to discretizing a VAR and briefly discusses the key inefficiencies. Section 3 discusses how to use efficient grids. The performance of the tensor-grid and efficient approaches is compared in the context of an AR(2) in section 4 and a New Keynesian model in section 5. Section 6 concludes.

## 2 The standard approach

Consider discretizing a VAR of the form

$$z_t = c + Az_{t-1} + \eta_t \quad (1)$$

with  $\eta_t \stackrel{i.i.d.}{\sim} N(0, \Sigma)$ , where  $z_t$  is a  $D \times 1$  vector.

In the standard approach due to [Tauchen \(1986\)](#), one applies a linear transformation to the VAR in (1) so that the innovations have a diagonal variance-covariance matrix. Specifically, since  $\Sigma$  is a real, symmetric matrix, it can be decomposed as  $\Sigma = L\Lambda L'$ , where  $L$  is an orthogonal matrix (i.e.,  $L'L = I$ ) and  $\Lambda$  is diagonal.<sup>1</sup> Then defining  $\tilde{z} = L'z$ ,  $\tilde{c} = L'c$ ,  $\tilde{A} = L'AL$ , and  $\tilde{\eta}_t = L'\eta_t$ , one has

$$\tilde{z}_t = \tilde{c} + \tilde{A}\tilde{z}_{t-1} + \tilde{\eta}_t \quad (2)$$

for  $\tilde{\eta}_t \sim N(0, \Lambda)$ .<sup>2</sup> The benefit of this transformation is that the conditional distribution  $\tilde{z}_{t,d}|\tilde{z}_{t-1}$  is simply  $N(\tilde{c}_d + \tilde{A}_{(d,\cdot)}\tilde{z}_{t-1}, \Lambda_d)$ , where  $\tilde{A}_{(d,\cdot)}$  is the  $d$ th row of  $\tilde{A}$ , which can be approximated using the logic from the univariate Tauchen method.

Given the VAR with a diagonal variance-covariance matrix as in (2), the standard approach chooses for each dimension  $d$  a number of grid points  $N_d$  and a corresponding grid  $\tilde{\mathcal{Z}}_d$ . Tauchen suggested using a linearly spaced grid (though any grid can be used)

$$\tilde{\mathcal{Z}}_d = \left\{ \tilde{E}_d + \kappa \left( 2 \frac{i-1}{N_d-1} - 1 \right) \sqrt{\tilde{V}_{d,d}} \right\}_{i=1}^{N_d} \quad (3)$$

in each dimension  $d$  where  $\tilde{E} = \mathbb{E}(\tilde{z}_t)$  and  $\tilde{V} = \mathbb{V}(\tilde{z}_t)$ , which is what I will use in the applications.<sup>3</sup> This covers  $\pm\kappa$  of the unconditional standard deviation in each dimension, and  $\kappa$  is commonly referred to as the *coverage*. Then a tensor grid of points  $\tilde{\mathcal{Z}} = \times_{d=1}^D \tilde{\mathcal{Z}}_d$  is formed, which has cardinality  $N = \prod_{d=1}^D N_d$ . This tensor grid  $\tilde{\mathcal{Z}}$  for the transformed system (2) implies a grid  $\mathcal{Z} = \{L\tilde{z} | \tilde{z} \in \tilde{\mathcal{Z}}\}$  for the untransformed system (1). Finally, the standard approach constructs the probability of transitioning from  $z_i \in \mathcal{Z}$  to  $z_j \in \mathcal{Z}$ , which I denote  $\pi_{ji}$ , using the conditional distribution  $N(\tilde{c}_d + \tilde{A}_{(d,\cdot)}\tilde{z}_i, \Lambda_d)$  as in the univariate case.<sup>4</sup> The appendix gives complete details on how these transition probabilities are constructed in the [Tauchen \(1986\)](#) approach.

The inefficiency in the standard approach is that some of the discretized states will be visited with

<sup>1</sup>To see this, note that  $\Sigma = HH'$  for some  $H$  because it is positive semidefinite ([Strang, 2014](#), p. 398). The singular value decomposition gives  $H = L\Gamma U'$  for  $L$  and  $U$  orthogonal and  $\Gamma$  diagonal. (An orthogonal matrix  $L$  by definition has  $L'L = I$ .) Consequently,  $\Sigma = HH' = (L\Gamma U')(U\Gamma L') = L\Gamma^2 L' = L\Lambda L'$  for  $\Lambda := \Gamma^2$ . This decomposition should be preferred to the Cholesky because it handles the case of  $\Sigma$  being only positive semidefinite as it is in (8). For a positive definite  $\Sigma$ , one can do a Cholesky followed by the SVD. For the positive semidefinite, the procedure is more involved and involves computing and the eigenvalues and eigenvectors of  $\Sigma$ , and it is described in the appendix.

<sup>2</sup>Specifically, left multiplying (1) by  $L'$  and using  $LL' = I$  gives  $L'z_t = L'c + L'ALL'L'z_{t-1} + L'\eta_t$ . The variance of  $\tilde{\eta}_t = \Lambda$  because the variance of  $L'\eta_t$  is  $L'\Sigma L = L'L\Lambda L'L = I\Lambda I = \Lambda$ .

<sup>3</sup>The mean is  $\tilde{E} = (I - \tilde{A})^{-1}\tilde{c}$ , and  $\tilde{V}$  can be found either iteratively using  $T \circ \tilde{V} = \tilde{A}\tilde{V}\tilde{A}' + \Lambda$  or directly via  $\text{vec}(\tilde{V}) = (I - \tilde{A} \otimes \tilde{A})^{-1} \text{vec}(\Lambda)$  ([Lütkepohl, 2006](#), p. 27).

<sup>4</sup>Here, one can think of  $\mathbf{i}$  as an index in  $\mathbb{Z}^{++}$ . The appendix uses additional structure to formalize the discretization procedure.

vanishingly small probability. These low probability states happen because of two forces. The first is when the VAR components exhibit strong correlation. For an example of this, consider the application in section 4 of quarterly, log, real, Spanish GDP data  $y_t$  modeled with an AR(2). This AR(2) can be mapped into a VAR(1) with  $z_t = [y_t, y_{t-1}]'$ .<sup>5</sup> Because of the high autocorrelation of  $y_t$ , the components of  $z_t$  exhibit a high degree of correlation. This can be seen in figure 1, where in a simulation of  $y_t$  against  $y_{t-1}$ , the points cluster along the 45-degree line. The inefficiency is immediately apparent when comparing the realized  $(y_t, y_{t-1})$  pairs with the tensor grid: Of the 49 tensor grid points in this example, only 7 have simulated values close to them.

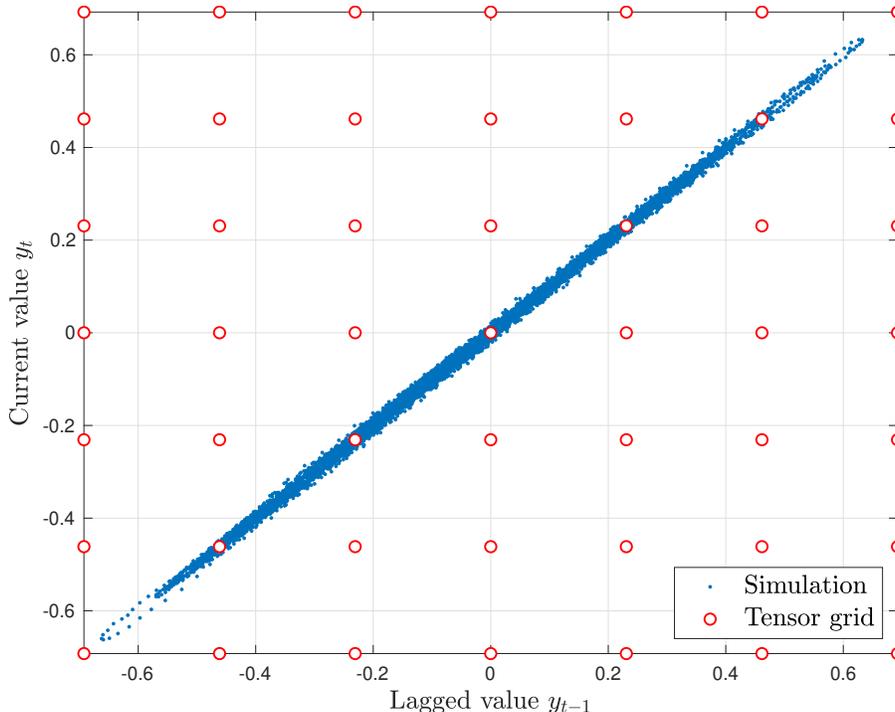


Figure 1: Simulated AR(2) process using estimates from Spanish real GDP data

The second force creating inefficiency occurs even when the the components of  $z_t$  are uncorrelated. Specifically, the force is that the joint probability of multiple individually unlikely states occurring simultaneously is extremely small. For a concrete example, suppose the unconditional distribution of  $z_t$  was  $N(0, I)$ . Then the probability  $\mathbb{P}(z_{t,d} \leq -2)$  is roughly 0.025 (for each  $d$ ), which is non-negligible. Consequently, one might reasonably want a coverage of at least  $\kappa = 2$  to capture this. But the joint probability  $\mathbb{P}(z_{t,d} \leq -2 \forall d)$  is approximately  $0.025^D$ , which goes to zero exponentially fast in the dimensionality and is already  $4 \times 10^{-7}$  for  $D = 4$ . Despite this, the tensor grid will place a significant number of points in these extremely low probability regions, and ever more so as either  $D$  or  $\kappa$  increases. In section 5, I will use four uncorrelated shocks from a standard *small*-scale New Keynesian model to highlight this inefficiency.

<sup>5</sup>For this series, the data vastly prefer an AR(2) to an AR(1): As shown in table 3, the AIC for the AR(1) is -626, while the AIC for AR(2) is -824. For both estimated processes, the autocorrelation is on the order of 0.999.

### 3 Gaining efficiency by dropping low-probability states

A simple way to gain efficiency is to increase the tensor-grid size  $N$  while simultaneously dropping states from  $\mathcal{Z}$  so that the number of points left (after dropping) is not larger than when one was using a tensor grid. I now formalize how to do this.

To begin, one must determine which states occur with low probability. One way to do this is by computing the invariant distribution  $\pi_{\mathbf{i}}$  associated with  $\pi_{\mathbf{j}|\mathbf{i}}$  such that  $\pi_{\mathbf{j}} = \sum_{\mathbf{i}} \pi_{\mathbf{j}|\mathbf{i}} \pi_{\mathbf{i}}$  for all  $\mathbf{j}$ . However, if the underlying approximation is poor,  $\pi_{\mathbf{j}|\mathbf{i}}$  may admit multiple invariant distributions or artificially slow convergence.<sup>6</sup> Consequently, a numerically more attractive approach is to exploit the VAR unconditional distribution  $z_t \sim N(E, V)$  for  $E = \mathbb{E}(z_t)$  and  $V = \mathbb{V}(z_t)$  to obtain the normal density  $\phi(z_t; E, V)$ . One can then use  $\hat{\pi}_{\mathbf{i}} \propto \phi(z_{\mathbf{i}})$ , scaled to sum to one, as an approximate invariant distribution to determine which states are low-probability.<sup>7</sup>

Having determined  $\hat{\pi}_{\mathbf{i}}$ , one must then choose a threshold  $\underline{\pi} \geq 0$  and drop states with  $\hat{\pi}_{\mathbf{i}} \leq \underline{\pi}$ . Formally, let

$$I := \{\mathbf{i} | \hat{\pi}_{\mathbf{i}} > \underline{\pi}\}, \quad (4)$$

define a new set

$$\mathcal{Z}^* = \{z_{\mathbf{i}} | \mathbf{i} \in I\} \subset \mathcal{Z} \quad (5)$$

and new transition matrix from states  $\mathbf{i} \in I$  to  $\mathbf{j} \in I$  as

$$\pi_{\mathbf{j}|\mathbf{i}}^* := \frac{\pi_{\mathbf{j}|\mathbf{i}}}{1 - \sum_{\mathbf{j} \notin I} \pi_{\mathbf{j}|\mathbf{i}}} = \frac{\pi_{\mathbf{j}|\mathbf{i}}}{\sum_{\mathbf{j} \in I} \pi_{\mathbf{j}|\mathbf{i}}}. \quad (6)$$

In the end, this procedure produces a set  $\mathcal{Z}^* \subset \mathcal{Z}$  with cardinality  $N^* \leq N$  and a transition matrix  $\pi_{\mathbf{j}|\mathbf{i}}^*$ .

As is evident in figure 1, this procedure will drop *most* of the points when components of the VAR have large correlations. Hence,  $N^*$  may be small and the resulting approximation poor if no further steps are taken. Consequently, in most cases it will be desirable to progressively increase  $N$  until the final number of points  $N^*$  is close to a target value, say  $\bar{N}$ .

This suggests the following algorithm:

1. Choose an  $\bar{N} \geq 2^D$ , a coverage  $\kappa > 0$ , and a threshold for near-zero values  $\underline{\pi} \geq 0$ . Set a flag  $f := 0$ , and set  $N = \bar{N}$ .
2. Setting  $N_d := \lfloor N^{1/D} \rfloor$ , use a tensor-grid method to construct  $\mathcal{Z}$  and  $\pi_{\mathbf{j}|\mathbf{i}}$ .
3. Compute  $\hat{\pi}_{\mathbf{i}}$ , calculate  $I$  as in (4), and obtain  $\mathcal{Z}^*$  as in (5) and  $\pi_{\mathbf{j}|\mathbf{i}}^*$  from (6).
4. If  $f = 1$  (where  $f$  is the flag) and  $N^*$  is less than or equal to  $\bar{N}$  or if  $f = 0$  and  $N^* = \bar{N}$ , STOP. Otherwise, continue.

---

<sup>6</sup>Multiple invariant distributions can arise because the transition probabilities can be zero to numerical precision. In the extreme, the *computed* transition probabilities can be the identity matrix.

<sup>7</sup>Using the normal density at each point gives a good approximation here because the equally spaced tensor grids give a tight connection between the density at the particular point and the mass of points closest to it. If one were using a different discretization procedure, the first approach could be preferable.

5. Here,  $N^* \neq \bar{N}$ . Proceed as follows:

- (a) If  $N^* < \bar{N}$ , replace  $N := (\lfloor N^{1/D} \rfloor + 1)^D$  and go to step 2.
- (b) If  $N^* > \bar{N}$ , set  $f := 1$ , replace  $N := (N^{1/D} - 1)^D$ , and go to step 2.<sup>8</sup>

This procedure produces a  $\mathcal{Z}^*$  (and  $\pi_{j|i}^*$ ) that has cardinality  $N^* \leq \bar{N}$ .

The preceding algorithm treated every component of  $z$  the same in that  $N_d$  is equated for all  $d$ . This is natural in some cases, but in others it may be desirable to have an uneven number of grid points in various dimensions, and the algorithm can be readily adapted for this. For instance, one can choose weights  $\omega_d \geq 0$  and then, in step 2, use  $N_d = 2 + \lfloor e^{\omega_d - \sum_{\bar{a}} \omega_{\bar{a}}/D} (N - 2^D)^{1/D} \rfloor$  instead of  $N_d = \lfloor N^{1/D} \rfloor$ . (The unusual incrementing of  $N$  in step 5(a) was to handle this case of unequal  $N_d$ .)

## 4 Accuracy in an AR(2)

I now compare the accuracy of the standard approach with the efficient approach in the context of an AR(2).

### 4.1 Mapping the AR(2) into a VAR(1)

Consider an AR(2) process

$$y_t = (1 - \rho_1 - \rho_2)\mu + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t, \quad (7)$$

with  $\varepsilon_t \sim N(0, \sigma^2)$ . Defining  $z_t := [y_t, y_{t-1}]'$ , this AR(2) can be written as

$$z_t = \begin{bmatrix} (1 - \rho_1 - \rho_2)\mu \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix} z_{t-1} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}. \quad (8)$$

Evidently,  $z_t$  follows a VAR(1) process. As is well-known, this simple mapping procedure can be extended to map any AR( $p$ ) or even VAR( $p$ ) process into a VAR(1).

### 4.2 Statistical efficiency

To begin the analysis, I discretize the estimated values  $(\rho_1, \rho_2, \sigma) = (1.936, -.938, .0029)$  with both methods and then—without simulation—recover the discretization-implied values for a number of key statistics reported in table 1.<sup>9</sup> First, notice that the tensor-grid approach—despite 961 discrete states—fails to accurately reproduce the statistics, predicting that the autocorrelation is unity (to five decimal places) and that innovations are essentially non-existent.<sup>10</sup> The tensor grid does not even get

<sup>8</sup>No rounding needs to be done here because, by virtue of  $f = 1$  and not having stopped, the previous step did  $N := (\lfloor N^{1/D} \rfloor + 1)^D$ , so inverting this mapping via  $N := (N^{1/D} - 1)^D$  gives an integer.

<sup>9</sup>This is the approach Tauchen advocated, and, for details on how this is done, see the appendix.

<sup>10</sup>In fact, to ensure there was at least some mixing, I had to restrict the largest transition probability in the Tauchen probability to be  $1 - 10^{-8}$ . This is binding for the tensor grid, which means with probability  $1 - 10^{-8}$  there is no transition and with probability  $10^{-8}$  the transition deterministically goes to an adjacent grid point. Capping these transition probabilities prevents multicollinearity in the estimation of  $c$  and  $A$ .

the mean correct. This is because the process is essentially perfectly autocorrelated and the guess on the invariant distribution is biased toward higher values (to highlight this extreme inaccuracy). While the performance of the efficient grid is not stellar, it is significantly better along almost every dimension despite having the same number of grid points.

	Discretization procedure		
	Actual	Efficient	Tensor
Persistence $\rho_1$	1.936	1.964	1.966
Persistence $\rho_2$	-0.938	-0.965	-0.966
Autocorrelation (lag 1)	0.99935	0.99959	1.00000
Autocorrelation (lag 2)	0.99747	0.99840	1.00000
Innovation size $\sigma$	0.0029	0.0027	0.00001
Innovation kurtosis $K$	3.000	9.568	0.001
Unconditional s.d. $\mathbb{V}(y_t)^{1/2}$	0.23	0.36	0.59
Unconditional mean $\mu$	1.0000	1.0000	1.3415
Mean Euler-equation error		-3.611	-3.132
Number of grid points		941	961
Number of grid points with $\mathbb{P}(z_1) > 10^{-9}$		933	77

Table 1: Discretization performance for an estimated AR(2)

Of course, the relative performance of the efficient grid is impacted by the degree of correlation in components of  $z_t$ . So now I will vary this degree of correlation by varying  $\rho_1, \rho_2$ , and hence  $\mathbb{V}(z_t)$ , while holding the unconditional mean and variance fixed at 1 and 0.01, respectively. Letting  $\hat{A}$  ( $\hat{A}^*$ ) denote the estimates of  $A$  from the tensor (efficient) grid, and similarly for  $\Sigma$ ,  $K$ , and  $V$ , one can compare the relative accuracy of the efficient grid by using

$$\epsilon_A := \log_{10} \frac{\|\hat{A} - A\|_\infty}{\|\hat{A}^* - A\|_\infty}, \epsilon_\Sigma := \log_{10} \frac{\|\hat{\Sigma} - \Sigma\|_\infty}{\|\hat{\Sigma}^* - \Sigma\|_\infty}, \epsilon_V := \log_{10} \frac{\|\hat{V} - V\|_\infty}{\|\hat{V}^* - V\|_\infty}, \text{ and } \epsilon_K := \log_{10} \frac{|\hat{K} - K|}{|\hat{K}^* - K|}. \quad (9)$$

Then, e.g., if  $\epsilon_A$  is 2, the tensor-grid approach has a maximal error 100 ( $10^{\epsilon_A}$ ) times larger than the efficient-grid approach.

The top four plots of figure 2 give contours of the errors for varying levels of first- and second-order autocorrelations.<sup>11</sup> (The white area in the graph corresponds to the non-stationary region.) To appropriately show the scale, the range varies from plot to plot. The graph reveals that accuracy gains of up to four orders of magnitude are possible from using efficient grids. These occur at high levels of autocorrelation, which is when the VAR components exhibit a large degree of covariance. While tensor grids can be better for some of the error measures in some regions of the parameter space, efficient grids almost always improve accuracy and often do so by orders of magnitude.

<sup>11</sup>In all cases, I choose  $\mu$  and  $\sigma$  so that the unconditional mean (variance) is 1 (0.01). The coverage  $\kappa$  is 5, and—for the tensor grid—21 points are used in each dimension (so the spacing between points is 1/2 an unconditional standard deviation).

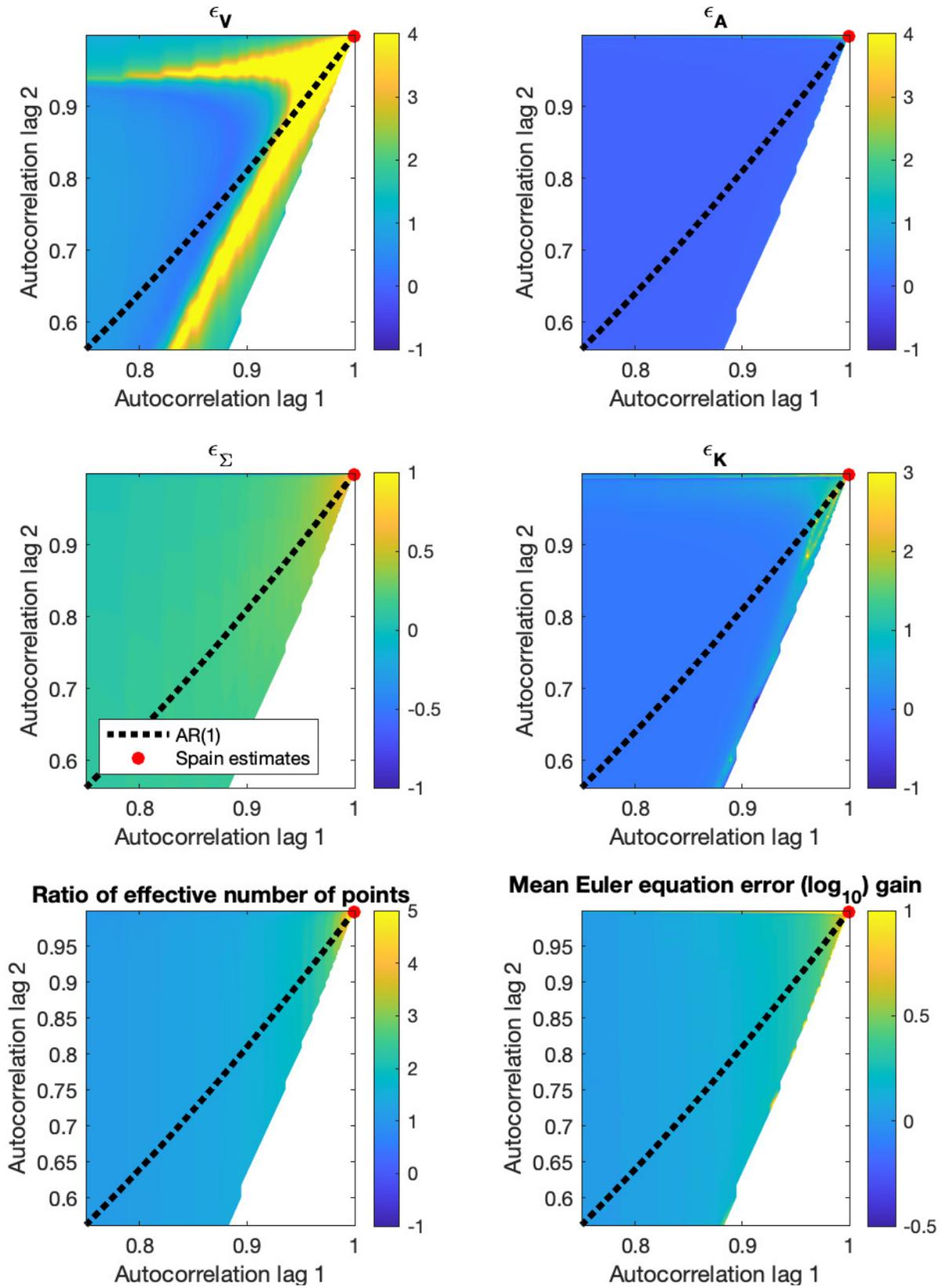


Figure 2: Efficiency gains by 1st and 2nd autocorrelation

### 4.3 Computational efficiency

Thus far, the measures of efficiency have been purely statistical. However, the primary reason to discretize shocks is as an input into a computational model. To quantify the computational efficiency gains in a simple way, consider the problem of an OLG model where households live for only two periods  $t$  and  $t + 1$ . Assume aggregate TFP follows the AR(2) in (7) and that households supply a unit of labor inelastically. Additionally, let households have access to a risk-free asset  $b_{t+1}$  at an exogenous price  $q$ . Then generation  $t$ 's problem is

$$\max_{b_{t+1}} u(y_t - qb_{t+1}) + \beta \mathbb{E}_{y_{t+1}|y_t, y_{t-1}} u(y_{t+1} + b_{t+1}). \quad (10)$$

Note that the optimal bond choice depends on the expectation of  $y_{t+1}$  conditional on  $y_t$  and  $y_{t-1}$ .

A primary way of assessing numerical errors, due to Judd and Guu (1997), is Euler-equation errors. These convert mistakes in policy choices into units of consumption via

$$EEE(b_{t+1}; y_t, y_{t-1}) := \log_{10} \left| 1 - \frac{u'^{-1} \left( \frac{\beta}{q} \mathbb{E}_{y_{t+1}|y_t, y_{t-1}} u'(y_{t+1} + b_{t+1}) \right)}{y_t - qb_{t+1}} \right|. \quad (11)$$

The interpretation is that if  $EEE(b_{t+1}; y_t, y_{t-1}) = -X$ , then a one-dollar mistake in consumption is made for every  $10^X$  spent. Since we are testing the accuracy, essentially, of the conditional expectation operator, we need to accurately obtain  $\mathbb{E}_{y_{t+1}|y_t, y_{t-1}} u'(y_{t+1} + b_{t+1})$ . To do this without loss, we assume CARA utility  $u(c) = \frac{1}{-\alpha} e^{-\alpha c}$ , which with well-known simplifications (shown in the appendix) gives an analytic expression for this expectation. Then, after finding the optimal policy  $\hat{b}_{t+1}(y_t, y_{t-1})$  ( $b_{t+1}^*(y_t, y_{t-1})$ ) using the tensor (efficient) grid, we can define the error

$$\epsilon_{EEE} := \hat{\mathbb{E}}(EEE(\hat{b}_{t+1}(y_t, y_{t-1}); y_t, y_{t-1})) - \mathbb{E}^*(EEE(b_{t+1}^*(y_t, y_{t-1}); y_t, y_{t-1})), \quad (12)$$

where  $\hat{\mathbb{E}}$  ( $\mathbb{E}^*$ ) uses the invariant distribution from the tensor (efficient) grid.<sup>12</sup>

The bottom right panel of figure 2 plots the contours of  $\epsilon_{EEE}$  at differing values of first and second autocorrelations. The values are almost always positive, and for more autocorrelated processes, tend to be around one. Such a value implies the average Euler-equation error using a tensor grid is an order of magnitude larger than with the efficient grid. Similarly, table 1 gives the Euler errors specifically using Spain's GDP process. As was the case with statistical efficiency, the computation efficiency of the efficient-grid method is not perfect. In particular, the Euler error implies a \$1 mistake is made for every \$4,200 ( $\approx 10^{3.265}$ ) spent. Nevertheless, it is significantly better than with the tensor grid, where a \$1 mistake is made for every \$750 spent.

<sup>12</sup>To find the optimal policy, I used Brent's method with an extremely tight tolerance of  $10^{-10}$ . Since the objective function is concave, this is guaranteed to find the optimal policy. For the numerical example, I use  $\alpha = 1.2861$ , which reproduced as closely as possible a constant relative risk aversion utility of 2 over the range  $c \in [.5, 1.5]$ . I also took  $\beta = .9$  and  $q = .96$  to ensure  $b_{t+1} = 0$  was not generally optimal.

## 5 Accuracy in a multiple-shock NK model

The previous section considered one important type of VAR, a suitably rewritten  $AR(p)$ . With a highly autocorrelated series, implying a high degree of correlation among the VAR components, the efficient grid improved considerably on the tensor grid. Now I consider another common type of VAR, a collection of several, possibly uncorrelated  $AR(1)$  shocks, and I will embed these in a small-scale New Keynesian model. First, I will consider the worst case for the efficient grid, where the VAR states are uncorrelated. Even there, the efficient grid will be far more efficient. Second, I will use an estimated VAR, which exhibits moderate correlation that makes the efficient grid perform even better relative to the tensor grid.

### 5.1 The shock structure

The NK model has a demand shock  $\beta_t$ , a productivity shock  $A_t$ , a government-spending-share shock  $s_{g,t}$ , and a monetary-policy-surprise shock  $m_t$ . Defining  $z_t = [\log \beta_t, \log A_t, \log s_{g,t}, \log m_t]'$ , the shocks evolve according to a VAR.

In the uncorrelated case, I take

$$z_t = \begin{bmatrix} (1 - \rho_b) \log \beta \\ (1 - \rho_a) \log A \\ (1 - \rho_g) \log s_g \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_b & 0 & 0 & 0 \\ 0 & \rho_a & 0 & 0 \\ 0 & 0 & \rho_g & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} z_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} N \left( 0, \begin{bmatrix} \sigma_b^2 & 0 & 0 & 0 \\ 0 & \sigma_a^2 & 0 & 0 \\ 0 & 0 & \sigma_g^2 & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{bmatrix} \right).$$

In the correlated case, I construct empirical measures for the four variables, logged, demeaned, and linearly detrended (as described in the appendix), estimate the process in deviations, and then add the desired means back. Defining  $\hat{z}_t$  as the demeaned values, the estimates give

$$\hat{z}_t = A\hat{z}_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \Sigma). \quad (13)$$

To deliver an unconditional mean of  $\mu = [\log \beta, \log A, \log s_g, 0]'$ , one takes  $c = (I - A)\mu$  so that

$$z_t = c + Az_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \Sigma) \quad (14)$$

represents the same process as (13).<sup>13</sup> The unconditional correlation matrix of the components of  $z_t$  is

$$\begin{bmatrix} 1.00 & & & \\ 0.05 & 1.00 & & \\ -0.48 & 0.83 & 1.00 & \\ -0.18 & 0.66 & 0.68 & 1.00 \end{bmatrix},$$

which exhibits a considerable amount of correlation. The estimates of  $A$  and  $\Sigma$  are in the appendix.

<sup>13</sup>To see this, note  $z_t = (I - A)\mu + Az_{t-1} + \varepsilon_t, \Leftrightarrow z_t - \mu = A(z_{t-1} - \mu) + \varepsilon_t, \Leftrightarrow \hat{z}_t = A\hat{z}_{t-1} + \varepsilon_t$ .

## 5.2 A NK model

The model nests [Fernández-Villaverde, Gordon, Guerrón-Quintana and Rubio-Ramírez \(2015\)](#) (FGGR), except with [Rotemberg \(1982\)](#) pricing instead of [Calvo \(1983\)](#). (Rotemberg pricing allows one to express the solution entirely as a function of the discretized exogenous states, whereas Calvo requires price dispersion as an endogenous state variable.) The model is given by the representative agent’s optimality conditions (the odd-looking Euler equation will be discussed below),

$$\frac{1}{c_t} = R_t \min \left\{ \mathbb{E}_t \frac{\beta_{t+1}}{c_{t+1}} \frac{1}{\pi_{t+1}}, \frac{1}{\gamma c} \right\},$$

$$\psi l_t^\theta c_t = w_t,$$

monetary policy,<sup>14</sup> constrained by the effective lower bound  $\underline{R}$  (which will be either 0 or 1),

$$R_t = \max\{Z_t, \underline{R}\},$$

$$Z_t = R \left( \frac{\Pi_t}{\bar{\Pi}} \right)^{\phi_\pi} \left( \frac{y_t}{\bar{y}} \right)^{\phi_y} m_t,$$

production, goods market clearing, and government spending,

$$y_t = A_t l_t,$$

$$y_t = c_t + g_t + \frac{\varphi \left( \frac{\pi_t}{\pi} - 1 \right)^2}{2} y_t,$$

$$g_t = s_{g,t} y_t,$$

and inflation behavior (from optimal price setting by firms),

$$\varphi \left( \frac{\pi_t}{\pi} - 1 \right) \frac{\pi_t}{\pi} = (1 - \epsilon) + \epsilon \frac{w_t}{A_t} + \varphi \mathbb{E}_t \frac{\beta_{t+1} c_t}{c_{t+1}} \left( \frac{\pi_{t+1}}{\pi} - 1 \right) \frac{\pi_{t+1}}{\pi} \frac{y_{t+1}}{y_t},$$

together with the shocks. Variables without time-subscripts denote steady state values.

The benchmark in FGGR corresponds to “ $\gamma = 0$ ” (i.e., no min term in the Euler equation) and  $\underline{R} = 1$ . As discussed in FGGR and elsewhere, the standard NK model with a ZLB features a “death spiral,” where, if the ZLB is expected to bind frequently enough, consumption and labor fall dramatically. A savings tax, which is used to micro-found  $\gamma > 0$ , eliminates this spiral (see section 6 of FGGR for details).<sup>15</sup>

The model calibration follows FGGR with three exceptions. First, to increase time at the ZLB, the demand shock persistence  $\rho_b$  is set to 0.9 in the uncorrelated shock case. Second, the Rotemberg adjustment cost parameter  $\varphi$  is set to mimic FGGR’s Calvo parameter of 0.75.<sup>16</sup> Last, with correlated

<sup>14</sup>The monetary policy rule does not exhibit inertia (i.e., include a lagged interest rate term) because doing so adds an endogenous, continuous state variable and so introduces extra approximation error. The aim here is to test only the error arising from discretization of the VAR.

<sup>15</sup>The death spiral occurs when the nominal interest rate is at the ZLB with sufficient deflation to make the real interest rate large. Then, the only way for the Euler equation to hold—for given  $t + 1$  values—is for consumption, output, and labor to fall, which only increases the desire to save at  $t - 1$ . The savings tax breaks this cycle.

<sup>16</sup>Rotemberg and Calvo pricing is observationally equivalent up to a first-order approximation ([Rotemberg, 1987](#), pp.

shocks, the inflation-response parameter in the Taylor rule is increased from 1.5 to 2.5.<sup>17</sup> The solution method is described in the appendix.

### 5.3 The discretized states

Figures 3 and 4 present the discretized tensor and efficient-grid states in three of the four dimensions for the uncorrelated and correlated case, respectively. The tensor grid has seven points in each dimension, for a total of  $7^4$ , while the efficient grid has fewer than  $7^4$ . For both grids, a coverage of  $\kappa = 3$  was used. To visualize the probability of each state occurring, the color map presents the unconditional marginal density of the states in  $\log_{10}$ . (Note the scales are not the same in the top and bottom panels.)

Consider first the uncorrelated case in figure 3. With efficient grids (in the top panel), the points form an ellipsoid, and the marginal density of the points ranges from  $10^{-5}$  to  $10^{-2.5}$ . In contrast, with tensor grids, the points necessarily form a box, and the marginal density ranges from  $10^{-7.5}$  to  $10^{-2.5}$ . As can be seen, the lowest probability points of the tensor grid occur in the corners of the box. These corners represent values where  $\beta$ ,  $A$ , and  $m$  are all unusually small or large, making the joint probability of all three of those events occurring exceedingly small. It is these low joint-probability events that are dropped by the efficient grid, resulting in the elliptical shape. What is not visible in the graph is  $s_g$  states. For the tensor-grid case, there are seven repeated values of each point. Consequently, dropping one corner of the box actually drops seven grid points.

Now consider the correlated case presented in figure 4. While there are the same number of points ( $7^4$ ) in figures 3 and 4, there are more distinct combinations of  $(\beta, A, m)$  in the latter. This is because Tauchen’s method forms a box with repeated values in the linearly transformed space  $Lz$  rather than  $z$ ; mapping the box back from  $Lz$  space to  $z$  space translates the states, making them visible (in the previous graph,  $L$  was diagonal). Crucially, note that the marginal density for the tensor grid now ranges from less than  $10^{-16}$  to  $10^{-2}$ . Consequently, in a simulation, areas close to the dark blue points will only be visited every few billion periods or less. Clearly, it is wasteful to include such states, and the efficient grid drops them, focusing just on the comparatively high probability  $10^{-5}$  to  $10^{-2.5}$  density region.

### 5.4 Formal accuracy tests

The preceding graphs suggest that the efficient-grid discretization is far more efficient than the tensor-grid approach. However, to formally test the accuracy of the efficient and tensor grids, while avoiding any interpolation or error-prone numerical integration, I use the approach of [den Haan and Marcet \(1994\)](#) (dHM) to test the model accuracy. The dHM accuracy test is very general and can be applied to multiple model equations using different weighting schemes. Here, I will focus on just the two

---

92-93).

<sup>17</sup>For smaller values, backwards iteration led to consumption collapsing.

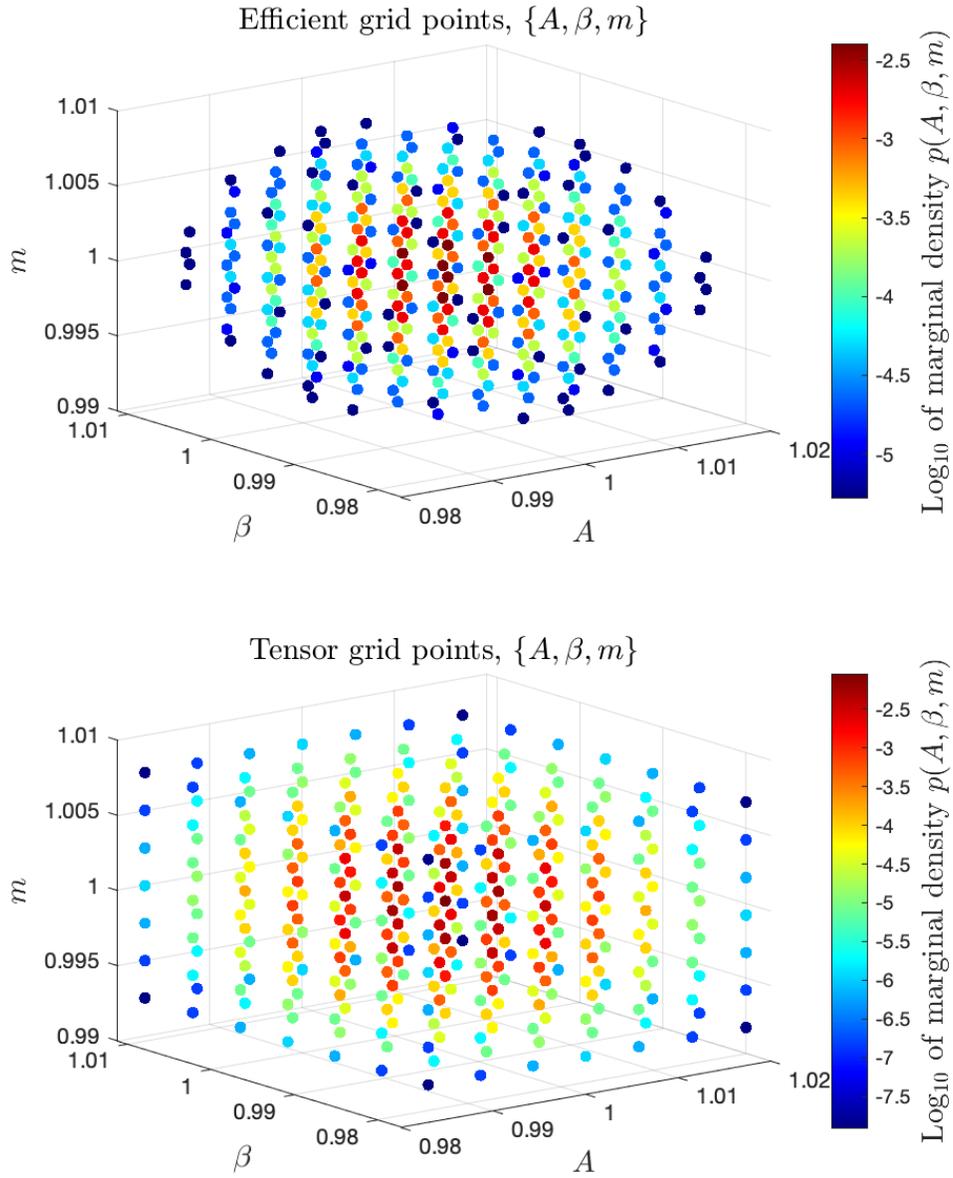


Figure 3: Efficient versus tensor grids in three of the four dimensions: Uncorrelated states

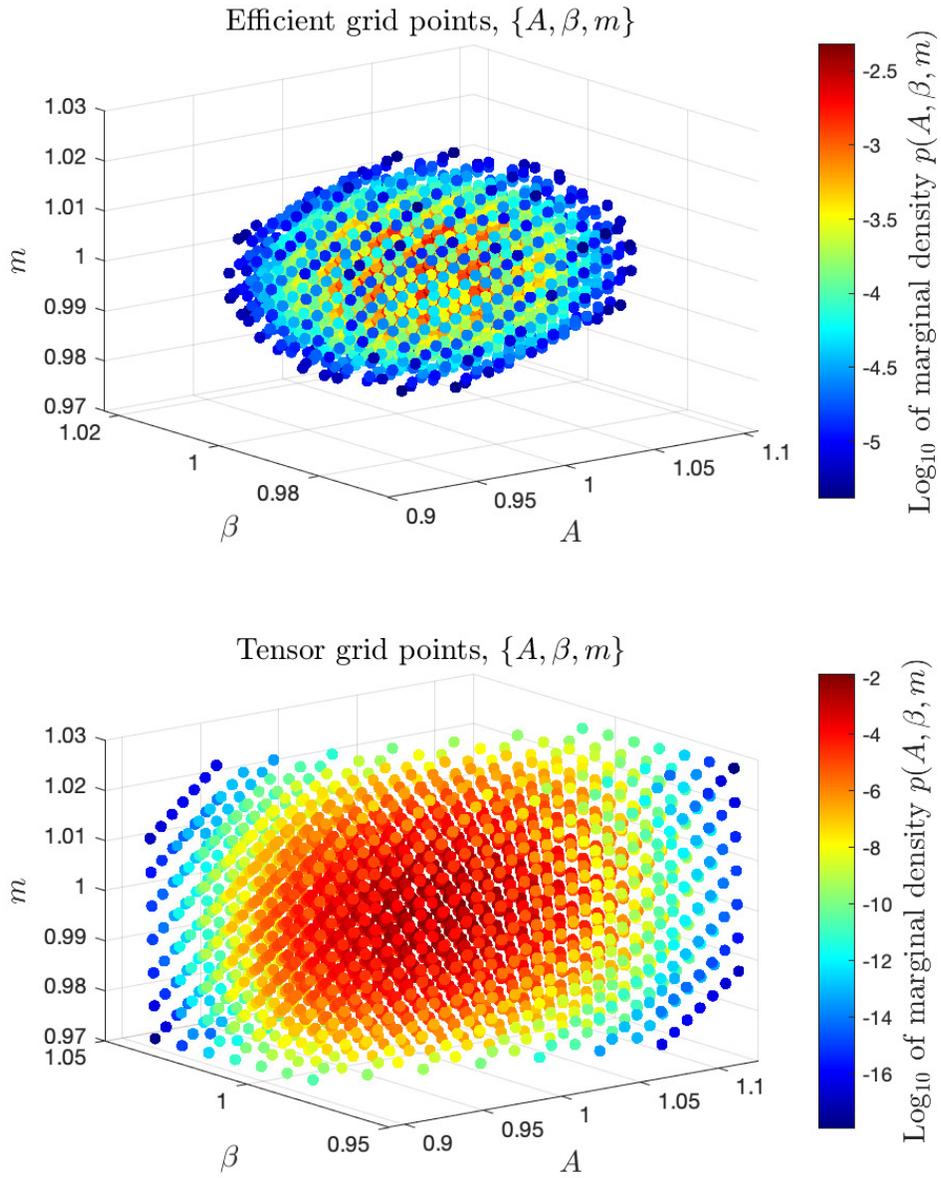


Figure 4: Efficient versus tensor grids in three of the four dimensions: Correlated states

intertemporal model equations governing consumption and price-level optimality. Specifically, define

$$f_t^1 = -\frac{1}{c_t} + \frac{\beta_{t+1}}{c_{t+1}} \frac{R_t}{\pi_{t+1}},$$

$$f_t^2 = -\varphi\left(\frac{\pi_t}{\pi} - 1\right) \frac{\pi_t}{\pi} + (1 - \epsilon) + \epsilon \frac{w_t}{A_t} + \varphi \frac{\beta_{t+1} c_t}{c_{t+1}} \left(\frac{\pi_{t+1}}{\pi} - 1\right) \frac{\pi_{t+1}}{\pi} \frac{y_{t+1}}{y_t},$$

and  $f_t = [f_t^1, f_t^2]'$ . Without a consumption floor ( $\gamma \approx 0$ ), optimality requires  $\mathbb{E}_t f_t = 0$  and consequently  $\mathbb{E} f_t = 0$ . (Here, I focus on the no-consumption-floor case to avoid computing an error-prone expectation.) For a simulation of length  $T$ , define  $M_T = T^{-1} \sum f_t$  and  $W_T = T^{-1} \sum f_t f_t'$ . The dHM result applied here gives that—under the null hypothesis of no error— $J_T = T M_T' W_T^{-1} M_T$  is distributed according to a chi-squared distribution with two degrees of freedom (since there are two equations). Consequently, just from the simulated data, we can test the null hypothesis of no error by computing the “ $J$ -stat” and the associated  $p$ -value.

Table 2 reports the number of discretized VAR states, the  $J$ -stat, and the  $p$ -value for the efficient- and tensor-based methods for the NK model without a ZLB ( $\underline{R} = 0, \gamma = 0$ ). Because the dHM statistic is rather noisy, I report boot-strapped errors for all the statistics. For comparison, I also include the values for perturbations of different order (first order, which is a linearized solution, second, and third). (For the perturbations, the shocks are not discretized.)

In each row, the efficient grid has fewer discretized states than the tensor grid, but, despite this, much less error. Specifically, the  $J$ -stat is orders of magnitude smaller for the efficient grid than for the tensor grid in each case. The  $p$ -value, which is the probability of observing the  $J$ -stat or a larger value under the null of no error, is zero for the tensor grid—meaning one would reject the null of no solution error. In contrast, with uncorrelated shocks and an efficient grid of 1257 or more states, one cannot reject the null of no error at the 10% confidence level. Consider also that the smallest efficient grid considered of 176 states is far more accurate than the 2401 tensor grid. Moreover, the  $J$ -stats of the efficient grid and a highly accurate third-order perturbation are statistically indistinguishable. With correlated shocks, the tensor grid performs worse, while the efficient grid and perturbations perform similarly.

Crucially, these numerical errors spill over into statistics researchers care about. For example, using the finest discretization in table 2 under a ZLB ( $\underline{R} = 1$ ) and consumption floor of  $\gamma = 0.99$ , the distributions of consumption and average durations of ZLB events are strongly influenced by the numerical error, as can be seen in figure 5. With tensor grids, the consumption distribution is shifted left, and noticeably more time is spent at the ZLB. Note that this model cannot be solved at all with perturbation (due to the ZLB), and it can be solved only extremely inaccurately with a tensor-grid approach. In contrast, the efficient-grid approach solves the model accurately.

## 6 Conclusion

This paper has proposed using efficient grids to improve the statistical and computational efficiency of discretized VARs. Numerical evaluations showed the resulting approximations are far more accurate in a number of dimensions. The efficient discretization of VARs proposed in this paper should significantly

<b>Uncorrelated shocks</b>								
Efficient			Tensor			Perturbation		
# states	$J$ -stat	$p$ -value	# states	$J$ -stat	$p$ -value	Order	$J$ -stat	$p$ -value
176	1.96 (3.82)	0.38 (0.30)	256	4733.12 (45.68)	0.00 (0.00)	1	7.87 (7.66)	0.02 (0.11)
321	1.35 (2.86)	0.51 (0.27)	625	2808.26 (82.96)	0.00 (0.00)	2	0.48 (1.98)	0.79 (0.28)
1257	0.03 (2.58)	0.98 (0.30)	1296	1731.62 (55.23)	0.00 (0.00)	3	0.44 (1.98)	0.80 (0.28)
1920	0.19 (2.20)	0.91 (0.31)	2401	1074.49 (50.52)	0.00 (0.00)			
<b>Correlated shocks</b>								
Efficient			Tensor			Perturbation		
# states	$J$ -stat	$p$ -value	# states	$J$ -stat	$p$ -value	Order	$J$ -stat	$p$ -value
243	1.55 (3.59)	0.46 (0.31)	256	6632.28 (55.49)	0.00 (0.00)	1	0.00 (1.77)	1.00 (0.28)
478	1.90 (2.79)	0.39 (0.27)	625	14992.6 (0.08)	0.00 (0.00)	2	0.25 (1.82)	0.88 (0.28)
847	0.03 (2.04)	0.99 (0.31)	1296	7623.97 (63.99)	0.00 (0.00)	3	0.24 (1.82)	0.89 (0.28)
2029	2.24 (3.71)	0.33 (0.28)	2401	4043.66 (66.16)	0.00 (0.00)			

Note: The simulation length is 15,000 periods; the model is no effective lower bound ( $\underline{R} = 0$ ), and no consumption floor ( $\gamma \approx 0$ ); bootstrapped standard errors are in parentheses; a higher  $J$ -stat (lower  $p$ -value) means one is more likely to reject the null of no error.

Table 2: den Haan and Marcet Hypothesis Testing of No Error

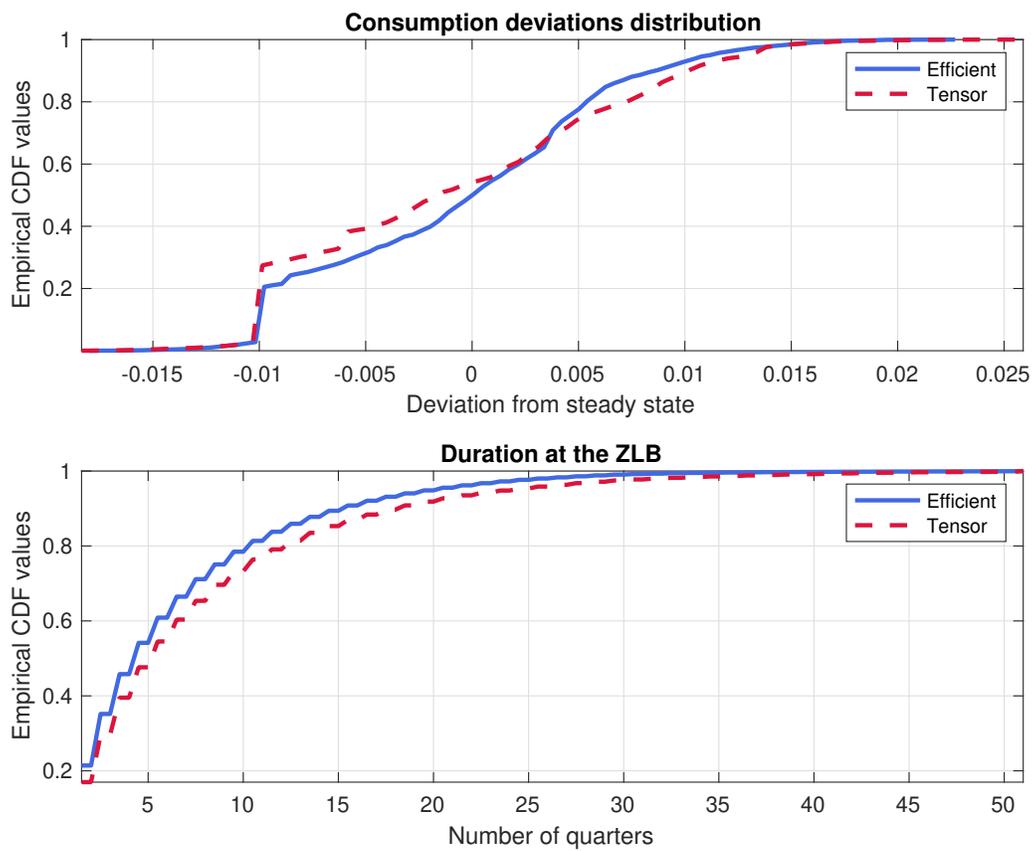


Figure 5: Distributions of consumption and nominal interest rates,  $\underline{R} = 1, \gamma = .99$

expand the ability of researchers to solve and estimate economic models.

## References

- Adda, J. and Cooper, R. W. (2003), *Dynamic Economics: Quantitative Methods and Applications*, The MIT Press, Cambridge, MA.
- Calvo, G. A. (1983), ‘Staggered prices in a utility-maximizing framework’, *Journal of Monetary Economics* **12**(3), 383 – 398.
- den Haan, W. J. and Marcet, A. (1994), ‘Accuracy in simulations’, *The Review of Economic Studies* **61**(1), 3–17.
- Fernández-Villaverde, J., Gordon, G., Guerrón-Quintana, P. and Rubio-Ramírez, J. F. (2015), ‘Nonlinear adventures at the zero lower bound’, *Journal of Economic Dynamics and Control* **57**, 182–204.
- Flodén, M. (2008), ‘A note on the accuracy of Markov-chain approximations to highly persistent AR(1) processes’, *Economics Letters* **99**, 516–520.
- Freund, R. M. (2014), ‘Symmetric matrices and eigendecomposition’, <http://s3.amazonaws.com/mitsloan-php/wp-faculty/sites/30/2016/12/15032137/Symmetric-Matrices-and-Eigendecomposition.pdf>. Accessed: 2020-05-19.
- Genz, A. (1992), ‘Numerical computation of multivariate normal probabilities’, *Journal of Computational and Graphical Statistics* **1**, 141–149.
- Genz, A. and Kwong, K. (2000), ‘Numerical evaluation of singular multivariate normal distributions’, *Journal of Statistical Computation and Simulation* **68**, 1–21.
- Judd, K. L. and Guu, S.-M. (1997), ‘Asymptotic methods for aggregate growth models’, *Journal of Economic Dynamics and Control* **21**, 1025–1042.
- Kopecky, K. and Suen, R. (2010), ‘Finite state Markov-chain approximations to highly persistent processes’, *Review of Economic Dynamics* **13**(3), 701–714.
- Lütkepohl, H. (2006), *New Introduction to Multiple Time Series Analysis*, Springer.
- Maliar, L. and Maliar, S. (2015), ‘Merging simulation and projection approaches to solve high-dimensional problems with an application to a New Keynesian model’, *Quantitative Economics* **6**, 1–47.
- Rotemberg, J. (1987), The New Keynesian microfoundations, in S. Fischer, ed., ‘NBER Macroeconomics Annual 1987, Volume 2’, The MIT Press, pp. 69–116.
- Rotemberg, J. J. (1982), ‘Sticky prices in the United States’, *Journal of Political Economy* **90**(6), 1187–1211.

Strang, G. (2014), *Differential Equations and Linear Algebra*, Wellesley-Cambridge, UK.

Tauchen, G. (1986), ‘Finite state Markov-chain approximations to univariate and vector autoregressions’, *Economics Letters* **20**(2), 177–181.

Tauchen, G. and Hussey, R. (1991), ‘Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models’, *Econometrica* **59**(2), 371–396.

Terry, S. J. and Knotek II, E. S. (2011), ‘Markov-chain approximations of vector autoregressions: Application of general multivariate-normal integration techniques’, *Economics Letters* **110**, 4–6.

## A Benchmark method of discretization using tensor-grid methods [Not for publication]

Tauchen (1986) provides a way to discretize a VAR of the form in (2) where the variance-covariance matrix is diagonal (and he discusses how suitable linear transformations can turn a VAR with a non-diagonal covariance matrix into a VAR with a diagonal one). Specifically, the approach is as follows. For each dimension  $d$ , choose a number of grid points  $N_d$  in each dimension and a corresponding grid  $\tilde{\mathcal{Z}}_d = \{\tilde{z}_{i,d}\}_{i=1}^{N_d}$  with  $\tilde{z}_{i,d} < \tilde{z}_{i+1,d}$ . Define the tensor grid as  $\tilde{\mathcal{Z}} = \times_d \tilde{\mathcal{Z}}_d$ . Use the multi-index  $\mathbf{i} = (i_1, \dots, i_D)$  to select elements of  $\tilde{\mathcal{Z}}$  such that  $\tilde{z}_{\mathbf{i}} = [\tilde{z}_{i_1,1}, \dots, \tilde{z}_{i_D,D}]'$ . Then, letting  $\tilde{A}_{(d,\cdot)}$  denote the  $d$ th row of  $\tilde{A}$ , one has

$$\mathbb{P}(\tilde{z}_d \leq x | \tilde{z}_{\mathbf{i}}) = \Phi(x; \tilde{A}_{(d,\cdot)} \tilde{z}_{\mathbf{i}}, \Lambda_d) \quad (15)$$

for any  $x$  where  $\tilde{z}_d$  is the random variable associated with the  $d$ th equation in (2) and where  $\Phi$  is the normal cdf. So the transition probability from  $\tilde{z}_{\mathbf{i}} \in \tilde{\mathcal{Z}}$  to a point  $\tilde{z}_{j,d} \in \mathcal{Z}_d$  is approximated by

$$\hat{\mathbb{P}}(\tilde{z}_{j,d} | \tilde{z}_{\mathbf{i}}) := \begin{cases} \Phi(m_{1,d}; \tilde{A}_{(d,\cdot)} \tilde{z}_{\mathbf{i}}, \Lambda_d) & \text{if } j = 1 \\ \Phi(m_{j,d}; \tilde{A}_{(d,\cdot)} \tilde{z}_{\mathbf{i}}, \Lambda_d) - \Phi(m_{j-1,d}; \tilde{A}_{(d,\cdot)} \tilde{z}_{\mathbf{i}}, \Lambda_d) & \text{if } j \in \{2, \dots, N_d - 1\} \\ 1 - \Phi(m_{N_d-1,d}; \tilde{A}_{(d,\cdot)} \tilde{z}_{\mathbf{i}}, \Lambda_d) & \text{if } j = N_d \end{cases} \quad (16)$$

where  $m_{j,d}$  is the midpoint  $\frac{1}{2}(\tilde{z}_{j+1,d} + \tilde{z}_{j,d})$ . So the joint probability of a transition to  $\tilde{z}_{\mathbf{j}}$  is

$$\pi_{\mathbf{j}|\mathbf{i}} := \hat{\mathbb{P}}(\tilde{z}_{\mathbf{j}} | \tilde{z}_{\mathbf{i}}) = \prod_d \hat{\mathbb{P}}(\tilde{z}_{j,d} | \tilde{z}_{\mathbf{i}}).$$

Given the transformed set  $\tilde{\mathcal{Z}}$ , one can recover the states in the untransformed space by reversing the transformation:

$$\mathcal{Z} := \{z_{\mathbf{i}} | z_{\mathbf{i}} = L \tilde{z}_{\mathbf{i}}, \tilde{z}_{\mathbf{i}} \in \tilde{\mathcal{Z}}\}. \quad (17)$$

The transition probabilities should remain the same because  $L$  is invertible and, therefore,  $\mathbb{P}(\tilde{z}_{\mathbf{j}} | \tilde{z}_{\mathbf{i}}) = \mathbb{P}(L \tilde{z}_{\mathbf{j}} | L \tilde{z}_{\mathbf{i}}) = \mathbb{P}(z_{\mathbf{j}} | z_{\mathbf{i}})$ .

## B Decomposition of real, symmetric, positive semidefinite matrices [Not for publication]

As part of the Tauchen procedure, I decompose  $\Sigma = L\Lambda L'$  for  $L$  orthogonal and  $\Lambda$  diagonal. In the case of a positive definite  $\Sigma$ , one can obtain this decomposition by doing a Cholesky decomposition of  $\Sigma$  into  $HH'$  and then a singular value decomposition of  $H$  into  $L\Gamma U'$ , and lastly defining  $\Lambda = \Gamma^2$ :

$$\Sigma = HH' = (L\Gamma U')(L\Gamma U')' = L\Gamma U'U\Gamma L' = L\Gamma^2 L' = L\Lambda L'.$$

For a positive semidefinite  $\Sigma$ , like with the AR(2), one cannot do a Cholesky decomposition.<sup>18</sup>

To do the decomposition for all positive semidefinite  $\Sigma$ , I build an algorithm based on the following result from linear algebra:

**Proposition 1** (Proposition 7 in Freund, 2014). *If  $Q$  is a real, symmetric matrix, then  $Q = RDR'$  for some orthonormal matrix  $R$  and diagonal matrix  $D$ , where the columns of  $R$  constitute an orthonormal basis of eigenvectors of  $Q$ , and the diagonal matrix  $D$  is comprised of the corresponding eigenvalues of  $Q$ .*

Additionally, proposition 6 in Freund (2014) states that the eigenvalues of a symmetric positive semidefinite (definite) matrix are nonnegative (positive), implying  $D$  has only nonnegative elements. Hence, the  $Q = RDR'$  decomposition, which is an eigendecomposition of  $Q$ , is suitable for transforming the VAR from (1) into (2).

**Algorithm 1.** *Algorithm for constructing an orthonormal basis of eigenvectors  $Q$ , with their eigenvalues  $\lambda$ , for a real, symmetric matrix  $Q$ .*

*Procedure.* To compute the decomposition, I follow the mostly constructive proof of proposition 5 of Freund (2014) (which is crucial in the proof of proposition 7).

First of all, if the eigenvalues of  $Q$  are distinct, then the matrix comprised of columns of eigenvectors, call it  $U$ , is orthogonal (which follows from proposition 4 in Freund). And, this can be made orthonormal by rescaling each column of  $U$  to have a norm of unity.

If the eigenvalues of  $Q$  are not distinct, one can construct the matrix  $U$  as follows:

1. Let  $u_1, \gamma_1$  be an eigenvector/value pair of  $Q$  with  $\|u_1\| = 1$ . Define  $k = 1$ .
2. Here,  $U = [u_1, \dots, u_k] \in \mathbb{R}^{n \times k}$  are eigenvectors with  $\|u_j\| = 1 \forall j$  and  $\gamma = [\gamma_1, \dots, \gamma_k]$  are eigenvalues so that  $QU = [\gamma_1 u_1, \dots, \gamma_k u_k]$ .
3. Now, construct a matrix  $V = [v_{k+1}, \dots, v_n] \in \mathbb{R}^{n \times (n-k)}$  with  $V$  orthogonal such that  $[U \ V] \in \mathbb{R}^{n \times n}$  is an orthonormal basis for  $\mathbb{R}^n$  by doing the following:
  - (a) Define  $X = U$  and  $j = 0$ . Here, the rank of  $X$  is  $k$ . Define  $r = k$ .

---

<sup>18</sup>There are alternative approaches that can be used, however, such as pivoting.

- (b) If  $r = n$ , go to the step 4. Otherwise, let  $j = j + 1$ . If  $\text{rank}([Xe_j]) > r$  where  $e_j$  is a vector of zeros except for a one in the  $j$ th element, then redefine  $X := [X \ e_j]$  and  $r := r + 1$ . Repeat this step (step 3(b)).
- (c) Here,  $\text{rank}(X) = n$  and so  $X$  forms a basis for  $\mathbb{R}^n$ . However, it is not necessarily orthogonal. Orthonormalize  $X$  using the Gram-Schmidt process. Note that because  $U$  was orthonormal with  $\|u_i\| = 1$  for  $i = 1, \dots, k$ , this leaves  $U$  unmodified but transforms the other columns of  $X$ . Call the new matrix  $Y$ , and define  $V$  as columns  $k + 1$  to  $n$  (so that  $Y = [U \ V]$ ). Now,  $Y = [U \ V]$  forms an orthonormal basis for  $\mathbb{R}^n$ .
4. Here,  $U'V = 0$  and  $V'QU = V'[\gamma_1 u_1, \dots, \gamma_k u_k] = 0$ . Compute an eigenvector  $w$  of  $V'QV \in \mathbb{R}^{(n-k) \times (n-k)}$  with eigenvalue  $\gamma_{k+1}$ , scaling  $w$  so that  $\|Vw\| = 1$ . Define  $u_{k+1} = Vw$ . Then
- $U'u_{k+1} = 0$  since  $U'u_{k+1} = U'Vw = 0w = 0$ , and
  - $Qu_{k+1} = \lambda_{k+1}u_{k+1}$  (i.e.,  $\lambda_{k+1}$  and  $u_{k+1}$  form an eigenpair of  $Q$ ).<sup>19</sup>

At this point, we have found a new  $u_{k+1}$  and  $\gamma_{k+1}$  that form an eigenpair of  $Q$  and are orthonormal. If  $k + 1 = n$ , we are done. Otherwise, go to step 2. □

Now consider a real, symmetric matrix  $Q$ . Using the above algorithm, one can compute orthonormal eigenvectors  $U$  with eigenvalues  $\lambda$ . Then note  $U'U = I$ . Defining  $D$  as the diagonal matrix having  $\lambda$  on its diagonal,

$$U'QU = U'[\lambda_1 u_1 \dots \lambda_n u_n] = \begin{bmatrix} \lambda_1 u_1 \cdot u_1 & \lambda_2 u_2 \cdot u_1 & \dots & \lambda_n u_n \cdot u_1 \\ \lambda_1 u_1 \cdot u_2 & \lambda_2 u_2 \cdot u_2 & \dots & \lambda_n u_n \cdot u_2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 u_1 \cdot u_n & \lambda_2 u_2 \cdot u_n & \dots & \lambda_n u_n \cdot u_n \end{bmatrix} = D.$$

Then,  $Q = IQI = (UU')Q(UU') = U(U'QU)U' = UDU'$ . Consequently, the  $U$  produced by algorithm 1 with the associated diagonal matrix of eigenvalues  $D$  forms the desired decomposition (the eigendecomposition) of the real, symmetric positive semidefinite matrix  $Q$ .

## C Solution method for the NK model [Not for publication]

In solving the NK model nonlinearly, I do the following:

1. Guess on  $c_{t+1}(z), \pi_{t+1}(z), y_{t+1}(z)$  for all  $z$  where  $z$  is a discretized shock vector. In practice, I use the steady-state values as the initial guesses.
2. For each  $z$ ,

---

<sup>19</sup>The proof is that  $d := Qu_{k+1} - \gamma u_{k+1}$  has  $d = QVw - \gamma Vw$ , hence  $V'd = V'QVw - \gamma V'Vw = V'QVw - \gamma w = 0$ . Since  $d$  is in the left null space of  $V$ , it is orthogonal to the column space of  $V$ . Since the column space of  $V$  is orthogonal to  $U$ , and the rank of  $[U \ V]$  is  $n$ ,  $d$  must be in the column space of  $U$ . That is, there exists an  $r$  such that  $d = Ur$ . Then note  $r = U'Ur = U'd = U'QVw - \gamma U'Vw = 0 - 0 = 0$ . Therefore,  $d = 0$  and hence  $Qu_{k+1} = \gamma u_{k+1}$ .

- (a) define an error function  $e(x)$  for  $x = [c_t(z), \pi_t(z), R_t(z), l_t(z)]$  as follows:
- i. define  $e_1(x)$  as the error in the Euler equation,
  - ii. recover  $w_t(z)$  assuming labor optimality holds,
  - iii. recover  $y_t(z)$  from the production equation,
  - iv. recover an implied nominal interest rate and define  $e_2(x)$  as the error between it and the guess,
  - v. recover  $g_t(z)$  as a share of output,
  - vi. define  $e_3(x)$  as the goods market clearing error,
  - vii. define  $e_4(x)$  as the error in the Rotemberg inflation equation,
  - viii. stack  $[e_1, e_2, e_3, e_4]$  into a vector  $e$ .
- (b) use the Levenberg-Marquardt algorithm to solve for an  $x^*$  such that  $e(x^*) \approx 0$ .
- (c) Recover  $c_t(z), \pi_t(z), y_t(z)$  at  $x^*$ .
3. Check if  $\max\{\|c_t - c_{t+1}\|_\infty, \|\pi_t - \pi_{t+1}\|_\infty, \|y_t - y_{t+1}\|_\infty\} < 10^{-6}$ . If so, STOP; otherwise, replace the  $t + 1$  guesses with the newly computed  $c_t(z), \pi_t(z), y_t(z)$  values and go to step 2.

## D Estimation and derivations [Not for publication]

This section provides some additional estimation details and derivations.

### D.1 Spanish GDP estimates

Table 3 reports the AR(1) and AR(2) estimates for log, real, Spanish GDP data.

	AR(1)	AR(2)
ARMA		
L.ar	0.999 (168.09)	1.936 (49.50)
L2.ar		-0.938 (-24.54)
Observations	95	95
<i>AIC</i>	-626.1	-823.8
Innovation size $\sigma$	0.00838	0.00287

*t* statistics in parentheses

A constant has been included in all regressions.

Table 3: AR(1)-AR(2) models for Spanish quarterly, log, real GDP data

### D.2 NK shock estimation

To construct empirical counterparts of the shocks in the NK model, I proceed as follows.

1. Construct an approximation of  $\beta_t$  via
  - (a) Define  $c_t$  using real personal consumption expenditures (FRED series DPCERA3Q086SBEA)
  - (b) Define the quarterly, gross nominal rate  $R_t$  as the gross, 3-month T-bill secondary market rate at a quarterly rate (FRED series DTB3 modified as  $(1 + DTB3/100)^{1/4}$ ).
  - (c) Define the quarterly, gross inflation rate  $\pi_t$  as the change  $p_t/p_{t-1}$ , where  $p_t$  is measured as the Core PCE (FRED series PCEPILFE).
  - (d) Define the measured  $\beta_t$  as  $\beta_t := \frac{c_{t+1}}{c_t} \frac{\pi_{t+1}}{R_t}$ .
2. Define  $s_{g,t}$  as the ratio of real government consumption expenditures and gross investment (FRED series GCEC1) to real GDP (FRED series GDPC1).
3. Define  $A_t$  as real GDP per worker by taking the ratio of GDPC1 and PAYEMS.
4. Construct  $m_t$  through the following steps:
  - (a) After logging and linearly detrending  $R_t$ ,  $\pi_t$ , and  $Y_t$  (FRED series GDPC1), regress without a constant the nominal rate deviations on inflation and GDP deviations for the years 1980 through 2007;
  - (b) Extract  $m_t$  as the residual from the regression coefficients applied to the actual time series.

This gives the series in levels for  $\beta_t$ ,  $s_{g,t}$ , and  $A_t$ , which I log and linearly detrend. For  $m_t$ , the series is in logs and is already zero mean.

This procedure is by no means ideal. It ignores the expectation in construction  $\beta_t$  and ignores the ZLB in the construction of  $m_t$ . A proper estimation would require a particle-filter approach. However, the point of this exercise is just to construct a rough approximation of the series for the purpose of producing correlation in the VAR states.

Keeping this in mind, the fitted process applied to these series is

$$\hat{z}_t = \begin{bmatrix} 0.370 & 0.039 & 0.014 & -0.112 \\ 0.434 & 0.928 & 0.031 & 0.193 \\ -0.614 & 0.028 & 0.976 & 0.014 \\ -0.052 & -0.006 & 0.004 & 0.826 \end{bmatrix} \hat{z}_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} N(0, CC'),$$

where

$$C = \begin{bmatrix} 0.0071 & & & \\ 0.0003 & 0.0056 & & \\ 0.0001 & -0.0018 & 0.0098 & \\ -0.0002 & 0.0001 & -0.0004 & 0.0032 \end{bmatrix}.$$

### D.3 VAR parameter estimation using discretized states and transition probabilities

To estimate the discretization-implied VAR without simulation, rewrite the system as

$$z_t = \beta x_t + \eta_t \text{ for } \beta = \begin{bmatrix} c & A \end{bmatrix} \text{ and } x_t = \begin{bmatrix} 1 \\ z_{t-1} \end{bmatrix}.$$

Then one can estimate  $c$  and  $A$  jointly using OLS with population moments:

$$z_t x_t' = \beta x_t x_t' + \eta_t x_t' \Rightarrow \beta = \mathbb{E}(z_t x_t') (\mathbb{E}(x_t x_t'))^{-1}.$$

Similarly, one can estimate the variance-covariance of the innovations via

$$\Sigma = \mathbb{E}(z_t - \beta x_t)(z_t - \beta x_t)'$$

The expectations are taken over  $z_{t-1}$  and  $z_t$  values, and so to do this without error, one needs the joint distribution of  $F(z_{t-1}, z_t)$ . This is obtained by multiplying the Markov transition probabilities  $\pi_{j|i}$  with the invariant probabilities  $\pi_i$  to arrive at the joint distribution.

### D.4 Derivation of the closed-form solution in the CARA-Normal framework

This subsection derives a closed-form expression for the marginal, continuation utility in the CARA-Normal framework.

Define  $\tilde{\mu} := (1 - \rho_1 - \rho_2)\mu + \rho_1 y_t + \rho_2 y_{t-1}$  (the conditional mean) and  $\xi := (2\sigma^2)^{-1}$ .

$$\begin{aligned} & \mathbb{E}_{y_{t+1}|y_t, y_{t-1}} u(y_{t+1} + b_{t+1}) \\ &= \int \frac{1}{-\alpha} e^{-\alpha(y_{t+1} + b_{t+1})} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_{t+1} - \tilde{\mu})^2 \xi} dy_{t+1} \\ &= u(b_{t+1}) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp[-\alpha y_{t+1} - (y_{t+1} - \tilde{\mu})^2 \xi] dy_{t+1} \\ &= u(b_{t+1}) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left[-\xi\left(\frac{\alpha}{\xi} y_{t+1} + (y_{t+1} - \tilde{\mu})^2\right)\right] dy_{t+1} \\ &= u(b_{t+1}) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left[-\xi\left(\frac{\alpha}{\xi} y_{t+1} + y_{t+1}^2 - 2\tilde{\mu} y_{t+1} + \tilde{\mu}^2\right)\right] dy_{t+1} \\ &= u(b_{t+1}) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left[-\xi(y_{t+1}^2 + \left(\frac{\alpha}{\xi} - 2\tilde{\mu}\right)y_{t+1} + \tilde{\mu}^2)\right] dy_{t+1} \\ &= u(b_{t+1}) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left[-\xi\left(\left(y_{t+1} + \frac{\frac{\alpha}{\xi} - 2\tilde{\mu}}{2}\right)^2 - \left(\frac{\frac{\alpha}{\xi} - 2\tilde{\mu}}{2}\right)^2 + \tilde{\mu}^2\right)\right] dy_{t+1} \\ &= u(b_{t+1}) \exp\left[-\xi\left(-\left(\frac{\frac{\alpha}{\xi} - 2\tilde{\mu}}{2}\right)^2 + \tilde{\mu}^2\right)\right] \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left[-\xi\left(\left(y_{t+1} + \frac{\frac{\alpha}{\xi} - 2\tilde{\mu}}{2}\right)^2\right)\right] dy_{t+1} \end{aligned}$$

$$\begin{aligned}
&= u(b_{t+1}) \exp \left[ -\xi \left( -\left( \frac{\alpha}{\xi} - 2\tilde{\mu} \right)^2 + \tilde{\mu}^2 \right) \right] \\
&= u(b_{t+1}) \exp \left[ -\xi \left( -\frac{1}{4} \left( \frac{\alpha}{\xi} - 2\tilde{\mu} \right)^2 + \tilde{\mu}^2 \right) \right] \\
&= u(b_{t+1}) \exp \left[ -\xi \left( -\frac{1}{4} \left( \frac{\alpha^2}{\xi^2} - 4\frac{\alpha\tilde{\mu}}{\xi} + 4\tilde{\mu}^2 \right) + \tilde{\mu}^2 \right) \right] \\
&= u(b_{t+1}) \exp \left[ -\xi \left( -\frac{\alpha^2}{(2\xi)^2} + \frac{\alpha\tilde{\mu}}{\xi} - \tilde{\mu}^2 + \tilde{\mu}^2 \right) \right] \\
&= u(b_{t+1}) \exp \left[ -\alpha \left( -\frac{\alpha}{4\xi} + \tilde{\mu} \right) \right] \\
&= u \left( b_{t+1} - \frac{\alpha}{4\xi} + \tilde{\mu} \right) \\
&= u \left( b_{t+1} + (1 - \rho_1 - \rho_2)\mu + \rho_1 y_t + \rho_2 y_{t-1} - \alpha \frac{\sigma^2}{2} \right).
\end{aligned}$$

Taking the derivative, one has

$$\mathbb{E}_{y_{t+1}|y_t, y_{t-1}} u'(y_{t+1} + b_{t+1}) = u' \left( b_{t+1} + (1 - \rho_1 - \rho_2)\mu + \rho_1 y_t + \rho_2 y_{t-1} - \alpha \frac{\sigma^2}{2} \right).$$