Lecture 6
Estimating HANK

Adrien Auclert and Matthew Rognlie

Goethe Heterogeneous-Agent Macro Workshop
June 2024

This lecure

Workshop so far: positive analysis for given shocks (to fiscal, monetary, TFP, etc)

Now: Estimation. What parameters best fit the aggregate data? What shocks rationalize the observed time series?

This lecure

Workshop so far: positive analysis for given shocks (to fiscal, monetary, TFP, etc)

Now: Estimation. What parameters best fit the aggregate data? What shocks rationalize the observed time series?

Main reference: Auclert et al. (2021) for estimation in the sequence space

See also Herbst and Schorfheide (2015), Fernández-Villaverde et al. (2016) for standard (state-space) DSGE estimation

Roadmap

1. Quick recap on estimating state-space models
2. Estimating models in the sequence space
3. Estimating HANK

Quick recap on estimating state-space models

Estimation of state-space model

- Suppose you have a model in the "state space"

$$s_t = \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta) \epsilon_t \quad \epsilon_t \sim_{iid} \mathcal{N}(0, I) \tag{1}$$

$$y_t = \Psi_0(\theta) + \Psi_1(\theta) s_t \tag{2}$$

  - θ are model **parameters** (including s.s. and shock process parameters)
  - s_t is a $(n_s \times 1)$ vector of **states** (e.g. capital, the distribution of agents)
  - y_t is a $(n_y \times 1)$ vector of **observables** (e.g. aggregate Y, C, etc)
  - ϵ_t is an $(n_\epsilon \times 1)$ vector of shock **innovations**

What do we get from the state space representation?

For **given** θ (i.e., for a given calibration), standard outcomes of interest are:

1. Impulse responses
2. Second moments of observables
3. Forecast error covariance decomposition into shocks at any horizon h
4. Historical decompositions into likely shocks
5. The likelihood $p\left(Y_{1:T}|\theta\right)$ given data $Y_{1:T}$ (obtained via Kalman filter)

Estimating using the state space representation

We can also **look for** θ using one of the following estimation procedures:

1. Impulse response matching
2. Minimum distance estimation/simulated method of moments
3. Maximum likelihood estimation
4. Bayesian estimation
    - Posterior mode estimation
    - Description of the posterior distribution using MCMC

Estimating het-agent models in the state space

- All these procedures are entirely standard, coded up in Dynare
- Estimation is fast provided that you have a fast way of getting to the state-space solution (1)–(2) given your parameters θ
- Key problem with het agent models: the state space is very large!

Estimating het-agent models in the state space

- All these procedures are entirely standard, coded up in Dynare
- Estimation is fast provided that you have a fast way of getting to the state-space solution (1)–(2) given your parameters θ
- Key problem with het agent models: the state space is very large!
- Specifically, when using the Reiter method, you need a Schur decomposition to get to (1)–(2). This is a huge bottleneck when n_s is big

Estimating het-agent models in the state space

- All these procedures are entirely standard, coded up in Dynare
- Estimation is fast provided that you have a fast way of getting to the state-space solution (1)–(2) given your parameters θ
- Key problem with het agent models: the state space is very large!
- Specifically, when using the Reiter method, you need a Schur decomposition to get to (1)–(2). This is a huge bottleneck when n_s is big
- There is a large "model reduction" literature that proposes to approximate the distribution to get around that problem. eg, Reiter (2010), Ahn et al. (2018), Winberry (2018), Bayer and Luetticke (2020)

Estimating het-agent models in the state space

- All these procedures are entirely standard, coded up in Dynare
- Estimation is fast provided that you have a fast way of getting to the state-space solution (1)–(2) given your parameters θ
- Key problem with het agent models: the state space is very large!
- Specifically, when using the Reiter method, you need a Schur decomposition to get to (1)–(2). This is a huge bottleneck when n_s is big
- There is a large "model reduction" literature that proposes to approximate the distribution to get around that problem. eg, Reiter (2010), Ahn et al. (2018), Winberry (2018), Bayer and Luetticke (2020)
- **Next**: an alternative way of doing all of this with the sequence space

Estimating models in the sequence space

Alternative route to estimation

- **Key idea**: use the MA representation implied by the sequence space solution
- From this we can get immediately:

1. Impulse responses
2. Second moments of observables
3. Forecast error covariance decomposition into shocks
4. Historical decompositions into likely shocks
5. The likelihood $p\left(Y_{1:T}|\theta\right)$ given data $Y_{1:T}$

Alternative route to estimation

- **Key idea**: use the MA representation implied by the sequence space solution
- From this we can get immediately:

    1. Impulse responses
    2. Second moments of observables
    3. Forecast error covariance decomposition into shocks
    4. Historical decompositions into likely shocks
    5. The likelihood $p\left(Y_{1:T}|\theta\right)$ given data $Y_{1:T}$

  and so we can directly do impulse matching, minimium-distance estimation, and likelihood-based estimation.

Impulse response matching

- Let's start simple: **impulse response matching**
- Given θ, we know how to get the impulse response $IRF(\theta)$
  - Just use `solve_steady_state`, then `solve_impulse_linear` !
- How do we confront this to an identified $\widehat{irf}$ from the data?

- Let's start simple: **impulse response matching**
- Given θ, we know how to get the impulse response $IRF(\theta)$
  - Just use `solve_steady_state`, then `solve_impulse_linear` !
- How do we confront this to an identified $\hat{irf}$ from the data?
- Natural answer: find θ to minimize distance, eg

$$\widehat{\theta} = \arg\min \left(IRF(\theta) - \hat{irf}\right)' V^{-1} \left(IRF(\theta) - \hat{irf}\right)'$$

where V contains, e.g., sample variances of $\hat{irf}$ (eg, Christiano et al. 2005).

Impulse response matching

- Let's start simple: **impulse response matching**
- Given θ, we know how to get the impulse response $IRF(\theta)$
  - Just use `solve_steady_state`, then `solve_impulse_linear` !
- How do we confront this to an identified $\hat{irf}$ from the data?
- Natural answer: find θ to minimize distance, eg

$$\widehat{\theta} = \arg\min \left(IRF(\theta) - \hat{irf}\right)' V^{-1} \left(IRF(\theta) - \hat{irf}\right)'$$

where V contains, e.g., sample variances of $\hat{irf}$ (eg, Christiano et al. 2005).

- Especially fast if, when θ changes...
  - You don't have to recalibrate the steady state
  - (even better) You don't have to recompute all Jacobians

Second moments, minimum distance estimation

- Given a θ that includes a specification of aggregate shock processes, we can get *moments* $M(\theta)$ of the data generated by our model
  - Classic case: second moments of aggregates $\text{Cov}(dX_t, dY_{t'})$
  - Also: OLS regression coefficients, HP filtered moments, etc.
  - A naive way would be to simulate (a better way just up next!)
- How do we confront this to empirical counterpart moments $\hat{m}$?

Second moments, minimum distance estimation

- Given a θ that includes a specification of aggregate shock processes, we can get *moments* $M(\theta)$ of the data generated by our model
    - Classic case: second moments of aggregates $\text{Cov}(dX_t, dY_{t'})$
    - Also: OLS regression coefficients, HP filtered moments, etc.
    - A naive way would be to simulate (a better way just up next!)
- How do we confront this to empirical counterpart moments $\hat{m}$?
- **Minimum distance estimation**: find θ to minimize distance, eg

$$\widehat{\theta} = \arg\min \left(M(\theta) - \hat{m}\right)' V^{-1} \left(M(\theta) - \hat{m}\right)'$$

  for V, can take the model's $\mathbb{E}\left[M(\theta) M(\theta)'\right]$

- Sometimes also called **simulated method of moments**

Obtaining second moments

- Suppose shocks $\{dZ_t\}$ are $MA(\infty)$ in iid structural innovation vectors $\{\epsilon_t\}$:

$$dZ_t = \sum_{s=0}^{\infty} M_s^Z \epsilon_{t-s}$$

Obtaining second moments

- Suppose shocks $\{dZ_t\}$ are $MA(\infty)$ in iid structural innovation vectors $\{\epsilon_t\}$:

$$dZ_t = \sum_{s=0}^{\infty} M_s^Z \epsilon_{t-s}$$

- Any endogenous $\{dX_t\}$ is also $MA(\infty)$:

$$dX_t = \sum_{s=0}^{\infty} M_s^{X|Z} \epsilon_{t-s}$$

where, if G denotes the GE Jacobian mapping shock Z to endogenous X,

$$M^{X|Z} = G^{X,Z} M^Z \tag{3}$$

Obtaining second moments

- Suppose shocks $\{dZ_t\}$ are $MA(\infty)$ in iid structural innovation vectors $\{\epsilon_t\}$:

$$dZ_t = \sum_{s=0}^{\infty} M_s^Z \epsilon_{t-s}$$

- Any endogenous $\{dX_t\}$ is also $MA(\infty)$:

$$dX_t = \sum_{s=0}^{\infty} M_s^{X|Z} \epsilon_{t-s}$$

where, if G denotes the GE Jacobian mapping shock Z to endogenous X,

$$M^{X|Z} = G^{X,Z} M^Z \tag{3}$$

- All **second moments** given Z follow directly from the standard expression

$$\text{Cov}(dX_t, dY_{t'})_Z = \sum_{s=0}^{\infty} M_s^{X|Z} M_{s+t'-t}^{Y|Z}$$

Covariance decomposition

- What if we have many shocks? Total covariance

$$\text{Cov}(dX_t, dY_{t'}) = \sum_Z \sum_{s=0}^{\infty} M_s^{X|Z} M_{s+t'-t}^{Y|Z}$$

- Immediate to then decompose this into the contribution of each shock:

$$\text{Contribution of Z to covariance} = \frac{\text{Cov}(dX_t, dY_{t'})_Z}{\text{Cov}(dX_t, dY_{t'})}$$

Covariance decomposition

- What if we have many shocks? Total covariance

$$\text{Cov}(dX_t, dY_{t'}) = \sum_Z \sum_{s=0}^{\infty} M_s^{X|Z} M_{s+t'-t}^{Y|Z}$$

- Immediate to then decompose this into the contribution of each shock:

$$\text{Contribution of Z to covariance} = \frac{\text{Cov}(dX_t, dY_{t'})_Z}{\text{Cov}(dX_t, dY_{t'})}$$

- Generalization: "forecast error covariance decomposition" at horizon h
- Instead of $\text{Cov}(dX_t, dY_{t'})_Z$, focus on $\text{Cov}(e^X_{t|t-h}, e^Y_{t|t-h})_Z$, where

$$e^X_{t|t-h} \equiv dX_t - \mathbb{E}_{t-h}[dX_t] = \sum_{s=0}^{h-1} M_s^X \epsilon_{t-s}$$

- Same, just replacing ∞ with h!

- The model can have any n_y and n_ϵ in principle
  - Think RBC model with just TFP shocks: $n_\epsilon = 1$, v.s. n_y typically at least 8
- If $n_\epsilon < n_y$, the F.E. covariance matrix $\mathbb{E}\left[e_{t|t-h}e'_{t|t-h}\right]$ cannot have full rank
- From an estimation perspective this represents "obvious misspecification"
- The standard solution is to add primitive shocks and/or add measurement error to observables, until $n_\epsilon \geq n_y$. Common case is $n_\epsilon = n_y$.

Modeling shocks, historical decompositions

- The model can have any n_y and n_ϵ in principle
  - Think RBC model with just TFP shocks: $n_\epsilon = 1$, v.s. n_y typically at least 8
- If $n_\epsilon < n_y$, the F.E. covariance matrix $\mathbb{E}\left[e_{t|t-h}e'_{t|t-h}\right]$ cannot have full rank
- From an estimation perspective this represents "obvious misspecification"
- The standard solution is to add primitive shocks and/or add measurement error to observables, until $n_\epsilon \geq n_y$. Common case is $n_\epsilon = n_y$.
- Then, given θ, can recursively back out most likely shocks $\epsilon_0, \epsilon_1, \ldots$ that solve

$$\min_{\epsilon_t} \left\| y_t^{data} - e_{t|0}\left(\theta\right)\left(\epsilon_t\right) \right\|^2$$

- Feeding in one shock at a time, can then construct a **historical decomposition** of the time series into shocks that explain it

Likelihood function

- So far, we've learned how to do:
  - Estimation: impulse response matching, "simulated" method of moments
    (Sometimes called "limited information estimation")
  - Post-estimation: covariance decomposition, historical decomposition

Likelihood function

- So far, we've learned how to do:
  - Estimation: impulse response matching, "simulated" method of moments
    (Sometimes called "limited information estimation")
  - Post-estimation: covariance decomposition, historical decomposition
- Suppose now that we can compute the likelihood $p(Y_{1:T}, \theta)$ given θ
  - Then, we can also do "full information estimation"

Likelihood-based estimation methods

- Likelihood-based estimation: use $p\left(Y_{1:T}, \theta\right)$ to estimate θ given data $Y_{1:T}$

  1. Maximum likelihood estimator:

  $$\widehat{\theta} = \arg\max p\left(Y_{1:T}, \theta\right)$$

  2. Bayesian estimation given prior $p\left(\theta\right)$: find the distribution of the posterior

  $$p\left(\theta | Y_{1:T}\right) = \frac{p\left(Y_{1:T} | \theta\right) p\left(\theta\right)}{p\left(Y_{1:T}\right)} \propto p\left(Y_{1:T}, \theta\right) p\left(\theta\right)$$

Likelihood-based estimation methods

- Likelihood-based estimation: use $p(Y_{1:T}, \theta)$ to estimate θ given data $Y_{1:T}$
    1. Maximum likelihood estimator:
    $$\widehat{\theta} = \arg\max p(Y_{1:T}, \theta)$$
    2. Bayesian estimation given prior $p(\theta)$: find the distribution of the posterior
    $$p(\theta|Y_{1:T}) = \frac{p(Y_{1:T}|\theta)\, p(\theta)}{p(Y_{1:T})} \propto p(Y_{1:T}, \theta)\, p(\theta)$$
- Finding the **posterior mode** is just like ML, except with p added
$$\widehat{\theta} = \arg\max p(Y_{1:T}, \theta)\, p(\theta)$$

Likelihood-based estimation methods

- Likelihood-based estimation: use $p\left(Y_{1:T},\theta\right)$ to estimate θ given data $Y_{1:T}$

1. Maximum likelihood estimator:

$$\widehat{\theta}=\arg\max p\left(Y_{1:T},\theta\right)$$

2. Bayesian estimation given prior $p\left(\theta\right)$: find the distribution of the posterior

$$p\left(\theta|Y_{1:T}\right)=\frac{p\left(Y_{1:T}|\theta\right)p\left(\theta\right)}{p\left(Y_{1:T}\right)}\propto p\left(Y_{1:T},\theta\right)p\left(\theta\right)$$

- Finding the **posterior mode** is just like ML, except with p added

$$\widehat{\theta}=\arg\max p\left(Y_{1:T},\theta\right)p\left(\theta\right)$$

- Tracing out the full distribution is more complicated.

Likelihood-based estimation methods

- Likelihood-based estimation: use $p\left(Y_{1:T},\theta\right)$ to estimate θ given data $Y_{1:T}$
  1. Maximum likelihood estimator:
  $$\widehat{\theta} = \arg\max p\left(Y_{1:T},\theta\right)$$
  2. Bayesian estimation given prior $p\left(\theta\right)$: find the distribution of the posterior
  $$p\left(\theta|Y_{1:T}\right) = \frac{p\left(Y_{1:T}|\theta\right)p\left(\theta\right)}{p\left(Y_{1:T}\right)} \propto p\left(Y_{1:T},\theta\right)p\left(\theta\right)$$
- Finding the **posterior mode** is just like ML, except with p added
  $$\widehat{\theta} = \arg\max p\left(Y_{1:T},\theta\right)p\left(\theta\right)$$
- Tracing out the full distribution is more complicated. A typical procedure is to construct a Markov chain whose stationary distribution is the posterior. This is called MCMC. The most common procedure for how to move from θ_t to θ_{t+1} is known as Metropolis Hastings. See Herbst and Schorfheide (2015).

Obtaining the likelihood function

- How do we get the likelihood? From the model's covariances!
- Say observed data is $\mathbf{Y}$. Stacking the covariances at all lags in $\mathbf{V}(\theta)$, can calculate log-likelihood of θ and $\mathbf{Y}$, assuming Gaussian innovations, as:

$$\mathcal{L}(\mathbf{Y};\theta) = -\frac{1}{2}\log\det\mathbf{V}(\theta) - \frac{1}{2}\mathbf{Y}'\mathbf{V}(\theta)^{-1}\mathbf{Y}$$

- No Kalman filter needed here!
- In practice, use a Cholesky decomposition of $\mathbf{V}$ to then quickly calculate $\log\det\mathbf{V}$ and $\mathbf{Y}'\mathbf{V}(\theta)^{-1}\mathbf{Y}$
- (Alternatives: Levinson recursion for $\mathbf{V}$, Whittle approximation)
- See Auclert et al. (2021) for details

Advantages of this approach

- Sequence space approach let us to do everything we can do in state space!
- For instance, you get the exact same results that Smets and Wouters got with Dynare on their data, see Auclert et al. (2021)
- In addition, there are important speed benefits whenever estimating:
  1. Shock processes: calculate $G^{X,Z}$ once, reuse to calculate M^X many times
  2. "Transition-relevant" parameters that do not affect the model steady state. These do not require recalculating the Jacobian of the HA model. Just reuse the same Jacobian and change the way in which these are combined into G!
  3. (In some behavioral models, behavioral parameters also work.)
- Much more complex to estimate parameters that do affect the steady state
  - See progress by Acharya et al. (2020) in this dimension

Estimating HANK

Application: estimating HANK

- What does an estimated HANK model on US data look like?
- What does it tell us about the importance of price rigidities? the shocks that drive the business cycle?
- Let's now see for ourselves!
- We are going to estimate the canonical HANK model
  - 3 shocks: TFP, government spending, and monetary policy
  - 3 observables: output, inflation, interest rates

- Canonical HANK model:

$$\begin{aligned} V_t(a_-, e) &= \max \frac{c^{1-\sigma}}{1-\sigma} + \beta \mathbb{E}\left[V_{t+1}(a, e') \,|e\right] \\ c + a &= (1 + r_{t-1})\, a_- + Z_t \frac{e^{1-\lambda}}{\mathbb{E}\left[e^{1-\lambda}\right]} \\ a &\geq \underline{a} \end{aligned} \tag{4}$$

- Post-tax income $Z_t = Y_t - T_t$
- Output $Y_t = X_t N_t$ [$X_t \equiv$TFP shock]
- Market clearing [$G_t \equiv$Gov spending shock]

$$C_t + G_t = Y_t = X_t N_t$$

Model setup 2/2

- Sticky wages,

$$\pi_t^w = \kappa_w \left(\frac{v'(N_t)/u'(C_t)}{(1-\lambda) Z_t/N_t} - \mu \right) + \beta \pi_{t+1}^w$$

- Flexible prices, $P_t = W_t/X_t$, so inflation is

$$1 + \pi_t = (1 + \pi_{wt}) \frac{X_{t-1}}{X_t}$$

- Taylor rule for monetary policy [$\epsilon_t \equiv$Mon policy shock]

$$i_t = i + \phi_\pi \pi_t + \epsilon_t$$

- Gov budget constraint and fiscal rule:

$$B_t = (1 + r_{t-1}) B_{t-1} + G_t - T_t$$
$$T_t = T + \phi_T (B_{t-1} - B)$$

- Shock processes and priors as in Herbst and Schorfheide (2015)

$$\log G_t = \rho_G \log G_{t-1} + \eta_t^G$$
$$\log X_t = \rho_X \log X_{t-1} + \eta_t^X$$
$$\epsilon_t = \rho_\epsilon \epsilon_{t-1} + \eta_t^\epsilon$$

Shock process parameters		Model parameters	
Param.	Prior Distribution	Param.	Prior Distribution
ρ_G	Uniform(0,1)	κ_w	Uniform(0,1)
ρ_ϵ	Uniform(0,1)	ϕ_π	Gamma(1.5,0.25)
ρ_X	Uniform(0,1)	ϕ_T	Uniform(0,1)
σ_G	InvGamma(1,4)		
σ_ϵ	InvGamma(0.4,4)		
σ_X	InvGamma(0.5,4)		

Let's see this in practice!

- Our plan:
  1. Set up the model and calibrate it with arbitrary parameters (θ)
  2. Check: simulate data from the model, then estimate on simulated data
  3. Take data from US time series and
     - Estimate shock processes
     - Estimate all parameters
  4. Post-estimation:
     - Report the impulse responses at the posterior mode
     - Do a historical decomposition of the observed time series into our shocks.
     - Do a forecast error variance decomposition of Y, π, i into our shocks at various horizons.
- See the lecture notebook!

References

Acharya, S., Cai, M. D., Del Negro, M., Dogra, K., Matlin, E., and Sarfati, R. (2020). Estimating HANK: Macro Time Series and Micro Moments. *Manuscript*.

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2018). When Inequality Matters for Macro and Macro Matters for Inequality. *NBER Macroeconomics Annual*, 32(1):1–75.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021). Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models. *Econometrica*, 89(5):2375–2408.

Bayer, C. and Luetticke, R. (2020). Solving Discrete Time Heterogeneous Agent Models with Aggregate Risk and Many Idiosyncratic States by Perturbation. *Quantitative Economics*, 11(4):1253–1288–1288.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, 113(1):1–45.

Fernández-Villaverde, J., Rubio-Ramírez, J., and Schorfheide, F. (2016). Chapter 9 - Solution and Estimation Methods for DSGE Models. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 527–724. Elsevier.

Herbst, E. P. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.

Reiter, M. (2010). Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models. *IHS Working Paper 258*.

Winberry, T. (2018). A Method for Solving and Estimating Heterogeneous Agent Macro Models. *Quantitative Economics*, 9(3):1123–1151–1151.