Contemporary Macroeconomic Theory: A Methodological Overview[†]

Nicolas Mäder

1 Introduction

> "Following the advances in microfoundations of macro that followed Kydland and Prescott, it is now fairly accurate to say that all economics is microeconomics."
>
> *Sims et al.* (2018)

Operating on the same fundamental premise — namely that observed outcomes emerge as a result of optimization at the level of microeconomic agents — macroeconomics extends microeconomics by examining the behavior of aggregate economies over time. For purposes of simplicity, however, the aspect of aggregation has been deemphasized in the sense that we have traditionally limited ourselves to representative agent setups in which cross-sectional distributions are degenerate by construction and aggregation is trivial by symmetry.[1] Accordingly, as evidenced by the titles of the most prevalent graduate textbooks (Stokey and Lucas, 1999; Sargent, 1997), the primary historical distinction between microeconomics and macroeconomics has been the latter's dynamic nature.

To generate the desired dynamics, macroeconomists typically study games that are infinitely repeated, but in which the respective circumstances — as given by the economy's state — change from period to period.[2] Since each game's equilibrium yields a new state for the following period,

[†] This note is intended to provide some useful background information for the various solution techniques studied in ECON 8200.

[1] In such settings, examining the effects of government policy on income and wealth inequality is, of course, infeasible. Concerns of this sort are addressed by the recently emerging heterogeneous agent literature in which cross-sectional distributions are viewed — rather than solely acting as integrands in aggregation — as valid macroeconomic targets themselves.

[2] For example, consider the canonical RBC model in which agents enter each period with a certain amount of capital, but that amount often changes across periods.

recursive economies can usually be represented in the form of a dynamical system of equations. In such a setting, one may then principally ask questions that are either intratemporal (how do agents behave/what constitutes equilibrium within the game of a particular period?) or intertemporal (how does our economic system evolve over time?). The primary purpose of this note is to emphasize that even though intratemporal equilibrium is *the* key determinant of a recursive economy's intertemporal behavior, the former only serves as a means to generate the latter.

Before discussing in more detail the various relevant notions of equilibrium, I proceed by illuminating our primary object of interest, the dynamical system. A dynamical system is said to be deterministic if, given some initial state, all of its future states can be predicted with both certainty and arbitrary precision. Conversely, it is said to be stochastic if its evolution cannot be predicted with certainty because it is subject to exogenous, random perturbations. In the context of macroeconomic theory, this distinction may be translated as follows. If all of our system's state variables are controlled within the system (e.g. by agents), meaning that there exist no outside forces that act as exogenous drivers, an initialized vector of states is sufficient to predict the system's state for any future period with arbitrary precision. Conversely, if there exists at least one state variable that is driven by external forces, we have a dynamical system that is random. In turn, the random dynamical system induces a *stochastic process*.[3]

Stochastic processes are often decomposed into two parts. For this, suppose we have univariate stochastic process that we wish to examine. If the process grows over time or features a seasonal component, it is typically decomposed into *trend* and the respective residual fluctuations implied by trend. In turn, the long-run phenomena associated with trend component are studied by the *growth* literature, whereas *DSGE* is typically concerned with providing insights into the cyclical movements surrounding trend.[4] While distinguishing between trend and cycles is relatively straightforward in theory, the corresponding empirical distinction is the subject of heated debate. This is because generating residuals requires a predetermined notion of trend such that the importance of a sound construction of the latter prior to an examination of the former cannot be overstated.[5]

[3]Further notice that the induced stochastic process satisfies the Markov property. See Appendix A for the relation between a random dynamical system and a Markov process.

[4]See King, Plosser, and Rebelo (2002) for an example of a framework that incorporates both trend and cycles. In particular, notice that agents may take growth into account by adjusting their discount rate.

[5]See "Why you should never use the Hodrick-Prescott filter" by Hamilton (2017) for a discussion of issues arising from atheoretical filtering.

2 Equilibrium

> "The idea of an economic equilibrium has undergone a [remarkable] evolution: it no longer carries the connotation of a system at rest."
>
> *Stokey and Lucas* (1999)

The above quote illustrates that once we introduce dynamics, what is meant by 'equilibrium' is no longer self-evident. Following the advances in microfoundations post Kydland and Prescott (KP, 1977), however, macroeconomic equilibrium has, just like its microeconomic counterpart, been viewed as an intratemporal condition in the spirit of Nash (1950). The defining property of this intratemporal condition is that each agent's chosen strategy constitutes a best-response to everyone else's strategy such that, in effect, no agent has an incentive to deviate from their current strategy.

The reason why the equilibrium term underwent a remarkable evolution is that it was used to describe macroeconomic phenomena long before macroeconomists insisted on microfoundations. Specifically, as pointed out by Stokey and Lucas (1999), the intratemporal Nash-type interpretation constitutes a stark deviation from the previous view, which "carried the connotation of a system at rest". Before discussing in more detail how the two notions are related, I proceed by providing an intuitive description of each.

Intratemporal equilibrium: Nash. A strategy profile is said to constitute intratemporal equilibrium of a game if it is "self-countering" in that each agent's current strategy constitutes — given their objectives and constraints — a best-response to everyone else's strategies (see Nash, 1950).

Intertemporal equilibrium: deterministic steady state. A deterministic dynamical system is said to be in steady state if its local position is given by a fixed point (see Frisch, 1936).

First, notice that the two concepts manifest themselves in entirely separate mathematical objects. Intratemporal equilibrium is given by a *set of strategies* within a game (or period), whereas intertemporal equilibrium takes the form of a *fixed point* in the system's state space. However, the two concepts are still related because whenever games are infinitely repeated, as is typically the case in macroeconomic theory, intratemporal equilibrium *induces* a dynamical system, which may or may not feature equilibrium intertemporally. In this sense, as alluded to before, intratemporal equilibrium is only a means to an end, namely to generate a dynamical system.

In face of stochastic innovations, macroeconomists often appeal to two additional types of equilibrium. First, we typically examine the system's *deterministic steady states*, or the steady states of an analogous system in which all exogenous drivers are 'turned off'. Second, rather than fundamentally changing the system at hand, we may account for the stochastic innovations by redefining the relevant notion of intertemporal equilibrium.

Intertemporal equilibrium: stochastic steady state. A stochastic process is said to be in steady state if it is stationary. A stochastic process $X : T \times \Omega \mapsto S$ with $T \subset \mathbb{N}$ is said to be stationary if its *finite-dimensional distributions* are invariant under translations of time, $\Pr(X_t, ..., X_{t+n}) = \Pr(X_{t+\tau}, ..., X_{t+n+\tau})$ for each $t \in T$, $n \leq \max\{T\} - t$, and $\tau < \max\{T\} - t - n$. It is said to be locally stationary in $T^\star \subset T$ if $X : T^\star \times \Omega \mapsto S$ is stationary. If $T = \mathbb{N}$, X is said to be asymptotically stationary if there exists t such that X is locally stationary in $T^{\star\star} = \mathbb{N}^{>t}$.

While asymptotic stationarity tells us whether (long-run) equilibrium exists, it is insufficient to guarantee that actually observed orbits will effectively reveal something about said equilibrium. For this, we need ergodicity.

Ergodic process. A Markov process $X : \Omega \times \mathcal{T} \mapsto S$ with time-invariant transition kernel $K : S \times \mathcal{S} \mapsto [0, 1]$ is said to be ergodic if its t-step transition probability converges in distribution to a unique limiting density π irrespective of the initial condition $x \in S$,

$$\lim_{t \to \infty} K^t(x, E) \xrightarrow{d} \pi(E)$$

for each $x \in S$ and $E \in \mathcal{S}$ (Hordijk, 1992).

In effect, a sufficiently long orbit of an ergodic process is enough to recover the underlying structure of the process. This is because an ergodic process features a limiting marginal distribution that fills out the entire state space except for a set of measure zero. The latter requires that our chain is irreducible such that the ergodic limit is unique and thus invariant to state initialization (see Samuelson's *ergodic hypothesis*, 1968).[6]

[6]Notice that the definition of an ergodic process requires our kernel to be both recurrent and aperiodic. If it were periodic — even if just for one $x \in S$ — $\lim_{t \to \infty} K^t(x, E)$ would not converge. In stark contrast, the canonical notion of ergodicity in the context of deterministic dynamical systems only requires recurrence (this is why some sources distinguish between ergodic and regular Markov chains). See Appendix A for a discussion of the two separate notions of ergodicity.

From theory to a dynamical system of equations

Since our games of interest are infinitely repeated, we require that optimization occurs uniformly across periods.[7] However, even though objectives are time-invariant, each individual agent's circumstances, as given by their *state*, may change from period to period. For example, consider the canonical RBC model in which agents enter each period with a certain amount of capital, but that amount often changes across periods.

To generate non-trivial dynamics over time, we require that at least one optimization problem be intertemporal. This is because we must have at least one of the system's state variables to be controlled within the system (e.g. by agents). If this is not the case, the resulting system would exhibit behavior that is independent and identical across time. Given at least one intertemporal optimization problem, intratemporal equilibrium then gives rise to set of decision rules (one rule per agent), also known as *policy functions*, that control or at least affect one endogenous state. As indicated by their name, policy functions take the form of a function mapping from the system's current state to the controls and must satisfy the condition — in the spirit of Nash (1950) — that no agent has an incentive to deviate from their current plan. In effect, combining the system's current state with the policy functions gives rise to next period's state. In turn, the new state can then be fed into the policy functions to generate the following state, and so on. We have a dynamical system of equations.

After deriving the policy functions from theory, the theorist usually proceeds by parameterizing the resulting dynamical system with reference to empirical data.[8] Finally, the parameterized version of the system is used for simulations and, if deemed appropriate, for purposes of counterfactual policy evaluation. Of the steps described thus far, ECON 8200 covers the ones highlighted in bold.

1. **Choose an appropriate objective for each proposed agent (model primitives)**
2. **Solve for intratemporal equilibrium (= policy function)**
3. Calibrate/estimate the resulting system's parameters
4. Simulate the parameterized version of the dynamical system

[7] This effectively requires that the same types of agents face the same optimization problem each period. It does not, however, categorically preclude agents from aging, for example.

[8] In theory, we often appeal to quantities, called parameters, which are not only determined outside of our system, but that are also time-invariant. Parameterization is often controversial because parameters should not only generate the targeted macro moments, but they should also be in line with microeconomic evidence (see Chari, Kehoe, and McGrattan 2009). To make matters worse, notice that an infinitesimally small parametric change can dramatically alter our system's asymptotic behavior (refer to the logistic map for an illustration).

'Solving for' intratemporal equilibrium: Analytical vs. numerical methods

In contrast to the models analyzed in ECON 8200, contemporary macro models can typically not be solved analytically, meaning that the policy functions are not be available in closed-form and thus need to be approximated numerically. While *global* solution methods solve for equilibrium on a bounded grid of the potentially high-dimensional space of possible states, *local* methods approximate the system's local behavior in the neighborhood of a deterministic steady state. Whether a local approximation of this sort constitutes an adequate approach largely depends on the properties of the original system.[9] We must ask,

Uniqueness. Is the (deterministic) steady state unique?

Stability. Is the (deterministic) steady state stable?

Resilience. If it is stable, but not unique, how resilient is it?

As far as uniqueness is concerned, notice that if a dynamical system is nonlinear, it may feature multiple steady states.[10] In terms of stability, steady states can be characterized as stable fixed points (valleys), unstable fixed points (hills), or saddle points (only stable in certain directions). Finally, to illustrate the notion of resilience, consider the following imagery quote from May's (1977) examination of population dynamics.

> "Is the dynamical behavior described by the multidimensional generalisation of a single valley? Or is the dynamical landscape pockmarked with many different valleys, separated by hills and watersheds? If the latter, [...] the system may return to this state following small perturbations, but large disturbances are likely to carry it into some new region of the dynamical landscape."

While the above quote constitutes a very intuitive description of the canonical non-uniquess issues that can arise in the context of *multistability*, it is worth noting that non-uniqueness also affects equilibria other than (deterministic) steady states.

[9] In particular, the Hartman-Grobman theorem states that linearization is locally effective in predicting qualitative patterns of the original nonlinear system's behavior if our point of interest is hyperbolic. Hyperbolic fixed points include attractors, repellants, and saddles, all of which are "structurally stable" (Strogatz, 2015).

[10] Although the use of the term "nonlinear science" has been described as analogous to labeling zoology the study of non-elephant animals, nonlinearity and its implications in the context of economic dynamical systems have only recently started gaining traction. Accordingly, should you feel inclined to explore a nascent field, you may be interested in investigating nonlinear dynamics (Strogatz, 2015).

Types of non-uniqueness: Indeterminacy, multistability, non-ergodicity

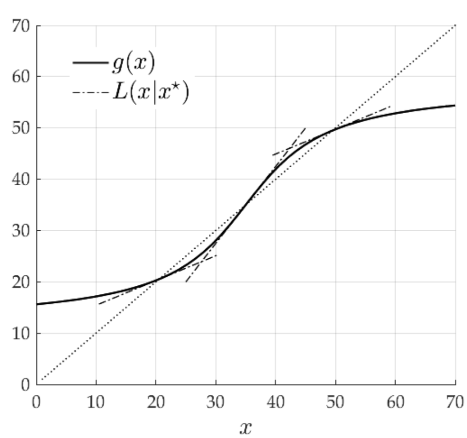
In addition to multistability, multiplicity can also arise when intratemporal equilibrium is *indetermindate* or when a stochastic process is *not ergodic*. For example, consider an economy in which intratemporal equilibrium is not unique. We call such an economy indeterminate because there is no way of knowing which equilibrium will emerge as the observed outcome unless we specify intratemporal initial conditions and define how players update their strategies intratemporally. In turn, indeterminacy often gives rise to self-fulfilling prophecies because expectational initial conditions effectively determine which equilibrium emerges as the observed outcome.

Of the three types of non-uniqueness, indeterminacy is the most problematic because its presence implies that our recursive economy may not have a dynamical-system-representation because the current state is insufficient to determine next period's state. Without a dynamical system, the intertemporal notions of equilibrium effectively lose their meaning. While indeterminacy should thus be viewed separately from multistability and non-ergodicity, the latter two can be intimately linked. To show this, let us consider the following stochastic difference equation,

$$\begin{aligned} x_{t+1} &= g(x_t) + \varepsilon_{t+1} \\ &= 35 + 15 \arctan\left(\frac{x_t - 35}{10}\right) + \varepsilon_{t+1} \end{aligned} \tag{1}$$

where $\varepsilon_{t+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. Notice that the analogous deterministic system — $x_{t+1} = g(x_t)$ — features an unstable fixed point that separates two *basins of attraction* with a corresponding stable attractor each.

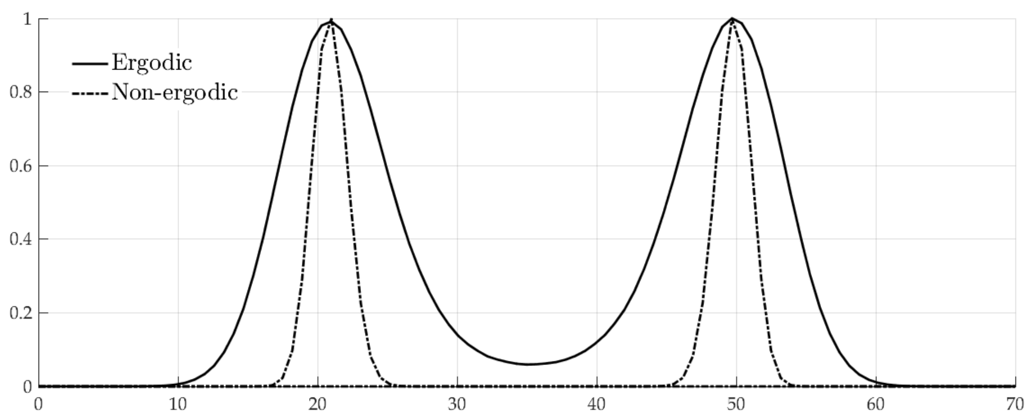
Figure 1: A multistable system



Notes: We can check the stability of our set of fixed points $J^F \approx \{20, 35, 50\}$ by evaluating $g'(x) = 3/2(1 + (\frac{x-35}{10})^2)$ for each $x \in J^F$. As expected, we have $g'(x) < 1$ for both $x \in \{20, 50\}$ and $g'(x) > 1$ for $x = 35$. The corresponding local linearization is then given by $L(x|x^\star) = g(x^\star) + g'(x^\star)(x - x^\star)$ for each $x^\star \in J^F$.

Then, if our system's exogenous innovations are strong and/or persistent enough to propel the state between the two basins, we have an ergodic random dynamical system with a bimodal stationary distribution as shown in Figure 2. Conversely, if the shocks are weak, the two stable steady states are too resilient for the state to ever switch between regimes. In such a case, asymptotic behavior is not invariant to initialization and our random dynamical system is not ergodic (it never 'forgets'). In the latter case, multistability is mirrored by non-ergodicity.

Figure 2: An ergodic and a non-ergodic parameterization of (1)



Notes: Figure 2 illustrates the limiting conditional density implied by the stochastic difference equation (1) for $\sigma \in \{\frac{2}{3}, 2\}$. For $\sigma = 2$, the system is ergodic, whereas for $\sigma = \frac{2}{3}$, it is not ergodic. Generally, notice then that if a deterministic dynamical system is multistable, its corresponding random analogue may or may not be ergodic depending on the structure of the stochastic innovations. For example, initializing (1) system with $x_0 = 20$ will (almost surely) never generate orbits that reach the region above 35 when the shocks are weak ($\sigma = \frac{2}{3}$).

3 Notation

To start with notation, recall that a stochastic process $X : T \times \Omega \mapsto S$ is a collection of random variables defined on a probability space $(W, \mathcal{W}, \mu)$. Here, T denotes an index set usually interpreted as time, whereas Ω and S are called the sample space and state space respectively. While the system's evolution is entirely captured by its state variables, economists are typically also interested in other variables that are functions of the state.

Partitioning the system

Following the logic of King, Plosser, and Rebelo (2002), consider a map $S_{t+1} = f(S_t, Z_t)$ with $f : \mathbb{R}^{n+m} \mapsto \mathbb{R}^n$, where $S_t \in \mathbb{R}^n$ and $Z_t \in \mathbb{R}^m$ denote an endogenous ("controlled") and an exogenous state. The latter evolves according to $Z_{t+1} = g(Z_t, W_{t+1})$, where $g : \mathbb{R}^{m+l} \mapsto \mathbb{R}^m$ and

$W_{t+1} : \Omega \mapsto \mathbb{R}^l$ is a random variable defined on the probability space $(\Omega, \mathcal{W}, \mu)$.[11] Moreover, let $C_t \in \mathbb{R}^k$ be the vector of all variables controlled by agents with intratemporal equilibrium giving rise to the policy function $p : \mathbb{R}^{n+m} \mapsto \mathbb{R}^k$, $C_t = p(S_t, Z_t)$. Controls may then be partitioned into endogenous state variables and non-state variables $C_t = [S_{t+1}, N_t]$ such that the different dimensions of the vector function f must be a subset of the dimensions of p, which implies $k \leq n$. Furthermore, let $h : \mathbb{R}^{n+m} \mapsto \mathbb{R}^{k-n}$ be the mapping from the current state to the non-state controls, $N_t = h(S_t, Z_t)$. We have thus partitioned our system into endogenous states, exogenous states, and non-state controls $[S_t, Z_t, N_t]$ with the three corresponding evolution mappings f, g, h. To describe the evolution of our system's state over time, f, g, and $(\Omega, \mathcal{W}, \mu)$ are sufficient. However, we often care about non-state controls, in which case we also require h. Luckily, the policy function p includes both f and h.

Following Blanchard and Kahn (1980), we can also partition the system's variables according to whether they are predetermined. In our case, the vector of non-predetermined variables P_t consists of — because we do not know their values prior to period t — all exogenous states and, because they are a function of the exogenous states, the non-state controls. We thus have $P_t = [N_t, Z_t] = [h(S_t, Z_t), Z_t]$. While the realization of P_t is not known at $t-1$, its uncertainty boils down to a model-implied conditional density. Oftentimes, we are interested in deriving particular moments of our non-predetermined variables. For example, the conditional expectation of P_t at $t-1$ is given by,

$$\begin{aligned}
\mathbb{E}_t[P_t] &= \mathbb{E}[P_t | S_{t-1}, Z_{t-1}] \\
&= \mathbb{E}[[N_t, Z_t] | S_{t-1}, Z_{t-1}] \\
&= \mathbb{E}[[h(S_t, Z_t), Z_t] | S_{t-1}, Z_{t-1}] \\
&= \mathbb{E}[[h(f(S_{t-1}, Z_{t-1}), g(Z_{t-1}, W_t)), g(Z_{t-1}, W_t)] | S_{t-1}, Z_{t-1}] \\
&= \int_\Omega \left[h\left(f(S_{t-1}, Z_{t-1}), g(Z_{t-1}, W_t(\omega))\right), g(Z_{t-1}, W_t(\omega))\right] \, \mu(d\omega) \qquad (2)
\end{aligned}$$

In contrast, the endogenous state is predetermined because its model-implied conditional density is degenerate.

$$S_t = f(S_{t-1}, Z_{t-1}) \qquad (3)$$

[11] Principally, g could also feature S_{t+1} or other variables determined in period t, but this is not usually the case in macroeconomic theory.

Deriving the system: An example

I now proceed by discussing an example of how a model produces a policy function.

$$
\begin{aligned}
\max_{c(S_t,Z_t)} \quad & \mathbb{E}\left[\sum_{\tau=0}^{\infty} \beta^{\tau} u(c_{t+\tau})\right] \\
\text{s.t.} \quad & c_t = c(S_t, Z_t) \\
& S_{t+1} = f(S_t, Z_t) - c(S_t, Z_t) \\
& Z_{t+1} = g(Z_t, W_{t+1})
\end{aligned} \tag{4}
$$

where $u : \mathbb{R} \mapsto \mathbb{R}$ is the level of utility associated with consumption, f denotes a given natural resource constraint, and $c : \mathbb{R}^{n+m} \mapsto \mathbb{R}$ is the policy function to be chosen by the household. A convenient way to reduce complexity in the above maximization problem was famously introduced by Richard Bellman, who defined (2) as a *value function* and rewrote the infinite sum in the form of a self-referential functional equation,

$$
\begin{aligned}
V(S_t, Z_t) &\equiv \max_{c(S_t,Z_t)} u(c_t) + \mathbb{E}\left[\sum_{\tau=1}^{\infty} \beta^{\tau} u(c_{t+\tau})\right] \\
&= \max_{c(S_t,Z_t)} u(c_t) + \beta \mathbb{E}[V(S_{t+1}, Z_{t+1})]
\end{aligned}
$$

We have thus reduced the original problem to the following self-referential function equation,

$$
\begin{aligned}
V(S_t, Z_t) = \max_{c(S_t,Z_t)} & \; u(c_t) + \beta \int_{\Omega} V(S_{t+1}, g(Z_t, W_{t+1}(\omega))\mu(d\omega) \\
\text{s.t.} \quad & S_{t+1} = f(S_t, Z_t) - c(S_t, Z_t)
\end{aligned} \tag{5}
$$

This functional equation is typically solved numerically using *value function iteration.* For this, we guess that V takes a certain form and optimize given that assumption. The resulting policy function c implies a new form for V, which is again used in optimization. Once the value function has converged, we are done.

4 Expectation formation and internal consistency

As evidenced by the previous example, macroeconomists typically choose objectives that require agents to form 'expectations' regarding the realization of future random variables. In particular, given our traditional objectives, the term expectation is virtually always synonymous with the model-implied (and thus true) conditional expectation of the respective random variable.[12] In such a setting, there's principally two ways to proceed.

Rational expectations. Agents are assumed to fully understand their stochastic environment, in which case individual expectations coincide with the true, *model-implied*, conditional expectation.

Non-rational approaches. Agents are assumed to not fully understand their stochastic environment. They implicitly optimize an objective that is incongruent with their true objective.

In response to the latter, Lucas famously postulates that "if your theory reveals profit opportunities, you have the wrong theory" (2001) while Christiano et al. (2018) poignantly proclaim — with the usual disclaimer that it serves as a "useful modeling strategy" — that "the assumption of rational expectations is *obviously* wrong". While both statements are meant to serve as a defense of contemporary practice, combining the two yields the following insight.

Insight. If humans are "obviously" incapable of evaluating our traditional objectives, then the latter cannot possibly capture the actual quantitative tradeoffs considered by humans.

This is unsatisfactory because the main reason we derive behavior from optimization — rather than imposing it as primitive — is precisely because it forces us to disclose the fundamental economic tradeoffs that are claimed to govern human decisions: "Theory helps keep track of benefits and costs" (Varian, 1993). Therefore, if we cannot realistically assert that human behavior indeed derives from the proposed optimization problem, then the modeler for all intents and purposes imposes behavior as the model primitive and thus invariably obfuscates the *actual tradeoffs* considered by humans. In effect, internal consistency only serves as a 'useful modeling strategy' if we *can* assert that human behavior in fact derives from the proposed optimization problem.

[12] Arguably, this is contrast to the literal definition of an expectation — "a belief that something will happen" — which is closer to the mathematical notion of a mode.

A Ergodicity

> "It was a belief in unique long-run equilibrium independent of initial conditions. I shall call it the *ergodic hypothesis*."
>
> *Samuelson* (1968)

While asymptotic stationarity addresses whether a stochastic process' asymptotic behavior — as given by its limiting joint distribution — is time-invariant, ergodicity ensures that such "equilibrium" is unique in that it is independent of the system's initialization. In effect, ergodicity is important because it ensures that a longitudinal sample from a given stochastic process is sufficient to recover the latter's stochastic structure. Since macroeconomists typically study phenomena that are dynamic, it is thus unsurprising that we often appeal to ergodicity.[13]

The aim of this section is to illuminate the definition of an *ergodic process* by contrasting it with the canonical definition of *ergodicity* from the mathematical branch of ergodic theory. In particular, while both definitions embody the ergodic spirit, which is to determine whether a single, longitudinal sample of data is sufficient to recover the structure of an underlying data generating process, they should be viewed separately. Formally, the two terms may then be described as follows.[14]

Ergodic process. A Markov process $X : \Omega \times \mathcal{T} \mapsto S$ with time-invariant transition kernel $K : S \times \mathcal{S} \mapsto [0, 1]$ is said to be ergodic if its t-step transition probability converges in distribution to a unique limiting density π irrespective of the initial condition $x \in S$.[15] We have,

$$\lim_{t\to\infty} K^t(x, E) \xrightarrow{d} \pi(E)$$

for each $x \in S$ and $E \in \mathcal{S}$ (Hordijk, 1992).

Ergodic dynamical system. A measure-preserving dynamical system $(Y, \mathcal{Y}, \mathbb{P}, T)$ is said to be ergodic if for every $E \in \mathcal{Y}$ with $\mathbb{P}(E) > 0$, we have $\mathbb{P}\left(\cup_{t=1}^{\infty} T^{-t}(E)\right) = 1$.

In words, a measure-preserving dynamical system is said to be ergodic if any initial condition $E \in \mathcal{Y}$ that has positive probability of occurring generates an orbit (or a set of orbits rather) that

[13] Geometric ergodicity is used by Duffie and Singleton (1993) to derive consistency of SMM for example.

[14] To promote intuitive understanding, I oftentimes sacrifice mathematical rigor.

[15] In the following, it is assumed that the index set $\mathcal{T}$, interpreted as time, is given by the set of integers $\mathbb{Z}$.

eventually lead us through the entire state space except for a set of $\mathbb{P}$-measure zero. Crucially, notice that ergodicity is a joint property of $\mathbb{P}$ and T. We say $\mathbb{P}$ is ergodic for T, or vice versa.

To illuminate the relationship between the two definitions, I proceed as follows. First, I present a set of relevant definitions and show that *random dynamical systems* naturally induce *time-homogenous Markov processes*. Second, to apply the canonical definition of ergodicity to Markov processes, I follow Hairer (2018) in rewriting the latter in form of a deterministic dynamical system. Finally, I discuss our principal object of interest — the ergodic process — in light of the previous insights.

Deriving a Markov process X from a random dynamical system

I start by giving a set of relevant definitions.

Stochastic Process. A stochastic process is an ordered set of random variables $X : \Omega \times \mathcal{T} \mapsto S$ defined on a common probability space $(\Omega, \mathcal{W}, \mu)$, where T is the index set, and S is the state space.

Markov Process. A stochastic process $X : \Omega \times \mathcal{T} \mapsto S$ defined on $(\Omega, \mathcal{W}, \mu)$ and adapted to the filtration $\{\mathcal{F}\}$ is said to satisfy the Markov property if $\Pr(X_t|\mathcal{F}_s) = \Pr(X_t|\sigma(X_s))$ for each $t, s \in \mathcal{T}$ with $t > s$, where $\sigma(X_s)$ is the σ-algebra generated by X_s.

Abusing notation, a Markov process' defining property is most easily illustrated via the following conditional independence: $\Pr(X_t|X_s, X_{s-1}, ..., X_0) = \Pr(X_t|X_s)$.

Time homogeneity. A Markov process is said to be time-homogenous if its transition kernel $K_t : \mathcal{T} \times S \times \mathcal{S} \mapsto [0, 1]$, $K_t(x, E) \equiv \Pr(X_{t+1} \in E|X_t = x)$ is time-invariant, $K_t(x, E) = K(x, E)$ for every $t \in T$.

Random dynamical system. A random dynamical system is defined as a tuple $(S, \mathcal{S}, \Gamma, \mathcal{G}, \mathbb{P}_G)$, where $(S, \mathcal{S})$ is a measurable space, Γ is a family of maps from S into itself, and $\mathbb{P}_G : \mathcal{G} \mapsto [0, 1]$ is a probability measure on a σ-algebra of Γ called $\mathcal{G}$ (see Bhattacharya and Majumdar, 2004).

But if random dynamical systems and Markov processes are entirely separate mathematical objects, why is it that "a Markov process may be described informally as a randomized dynamical system" (Kallenberg, **?**)? This is because any stochastic process X generated by way of a random dynamical system as defined above must satisfy the properties of a time-homogenous

Markov process. To show this, I follow Bhattacharya and Majumdar (2004) in constructing X from $(S, \mathcal{S}, \Gamma, \mathcal{G}, \mathbb{P}_G)$ as follows. First, consider another stochastic process $\gamma : \Omega \times \mathcal{T} \mapsto \Gamma$ defined on a probability space $(\Omega, \mathcal{W}, \mathbb{P}_G^{\mathbb{Z}})$, where $\mathbb{P}_G^{\mathbb{Z}} : \mathcal{W} \mapsto [0,1]$ is such that γ is an i.i.d. sequence of maps with each map γ_t being independently and identically drawn according to $\mathbb{P}_G$. In turn, by additionally specifying an initial condition $X_0 : \Omega \mapsto S$ with a corresponding distribution $\iota(E) \equiv \Pr(X_0 \in E)$ for each $E \in \mathcal{S}$, we may then define,

$$X_t(\omega) \equiv \gamma_t(\omega) \circ \gamma_{t-1}(\omega) \circ ... \circ \gamma_1(\omega) \circ X_0(\omega)$$

which constitutes a stochastic process. In addition, because our sequence of maps γ is i.i.d., we must have $K_t(x, E) = K(x, E)$ for every $t \in \mathbb{N}$. To show this, recall from our earlier definition,

$$\begin{aligned} K_t(x, E) &\equiv \Pr(X_{t+1} \in E | X_t = x) \\ &= \mathbb{P}_G^{\mathbb{Z}}(\{\alpha \in \Gamma^{\mathbb{Z}} | \alpha_t(x) \in E\}) \\ &= \mathbb{P}_G(\{\alpha \in \Gamma | \alpha(x) \in E\}) \\ &= K(x, E) \end{aligned}$$

for each $x \in S$, $E \in \mathcal{S}$, and each $t \in \mathbb{N}$. Here, the second step follows from γ being iid. But this is exactly the defining property of a time-homogeonous Markov process. Analogously, it can be shown that the n-step transition probability $K_n^t(x, E) \equiv \Pr(X_{t+n} \in E | X_t = x)$ is also time-invariant $K_t^n(x, E) = K^n(x, E)$ for each $x \in S$, $E \in \mathcal{S}$, and $n \in \mathbb{N}$. Finally, before delving into ergodicity, I define invariance for a particular kernel K. For this, consider the operator $T_K(\iota)$

$$T_K(\iota)(E) = \int_{\mathcal{S}} K(x, E)\iota(dx)$$

***K*-Invariance.** A probability measure $\kappa : \mathcal{S} \mapsto [0,1]$ is said to be invariant for the kernel K, or K-invariant, if it satisfies $\kappa(E) = T_K(\kappa)(E)$

Theorem. The set of K-invariant measures $J^I(K)$ is convex. Formally, $\kappa_1, \kappa_2 \in J(K)$ implies $t\kappa_1 + (1-t)\kappa_2 \in J^I(K)$ (Hairer, 2018).

Remark. Initializing a time-homogenous Markov process with a K-invariant density κ yields a stationarity (Hairer, 2018). Intuitively, this is because while the conditional evolution is fixed by the time-homogeneity of the kernel, K-invariance of π ensures that all marginal distributions are time-invariant as well. In turn, because both marginal and conditional densities are time-invariant, so is the joint density.

Rewriting X as a dynamical system

To apply the canonical definition of ergodicity to our process X, the latter must be rewritten in the form of a dynamical system $(Y, \mathcal{Y}, \mathbb{P}_\iota, T_{\circledcirc})$. For this, I follow Hairer (2018) and let Y be given by $S^{\mathbb{Z}}$, $\mathcal{Y} = \mathcal{S}^{\oplus\mathbb{Z}}$, where each element x in Y represents an orbit X_ω of the entire process X for a given $\omega \in \Omega$. In turn, the measure $\mathbb{P}_\iota : \mathcal{Y} \mapsto \mathbb{R}$ is defined as the joint density $\mathbb{P}_\iota(E) \equiv \Pr(X_0 \in E_0, X_1 \in E_1, ...)$ for each $E \in \mathcal{Y}$. Here, the subscript ι serves as a reminder that our joint density depends on the chain's initialization $\iota(E) = \Pr(X_0 \in E)$. To complete the system, the relevant transition map is then given by the shift map $T_{\circledcirc}(X_t) = X_{t+1}$.

Now, to apply the canonical definition of ergodicity, we must ensure that our constructed system $(Y, \mathcal{Y}, \mathbb{P}_\iota, T_{\circledcirc})$ is measure-preserving. In turn, this requires that ι be invariant for K because $T_{\circledcirc}$ shifts the process forward in time. This means that we are effectively limiting ourselves to processes that are stationary (!). We then define as follows.

Ergodic measure. We say that a K-invariant initial distribution π is *ergodic for* T_K if the implied joint density $\mathbb{P}_\pi$ is ergodic for the shift map $T_{\circledcirc}$ (Hairer, 2018).

As indicated above, since π is chosen among the distributions that are invariant for K, the implied process X — as given by the joint density $\mathbb{P}_\pi$ — is stationary. Now, to understand what it means for π to be ergodic for T_K, we must understand what it means for $\mathbb{P}_\pi$ to be ergodic for $T_{\circledcirc}$.

By definition, we know that or any $E \in \mathcal{Y}$ with $\mathbb{P}_\pi(E) > 0$, we have $\mathbb{P}_\pi\left(\cup_{t=1}^{\infty} T_{\circledcirc}^{-t}(E)\right) = 1$. In the context of our constructed system, this means that if we are given a set of orbits $E \in \mathcal{Y}$ that have a positive probability of occurring then we can generate every other possible orbit $x \in S^{\mathbb{Z}}$ — except for a set of $\mathbb{P}_\pi$-measure zero — simply by continually shifting our existing orbits forward. In effect, we have shown that endowing K with an ergodic measure π generates a process X of which just a few sufficiently long orbits are almost surely enough to recover its joint distribution. However, notice that this is a "local" statement in the sense that it conditions on an initial distribution π. In particular, this means that the set $J^E(T_K)$ of ergodic measures for T_K may not be a singleton. If $J^E(T_K)$ is not a singleton, there exist multiple initial distributions $\{\pi_1, \pi_2, ...\}$ that are all ergodic for T_K, which implies that the asymptotic behavior of the process is not invariant to initial conditions which in turn implies that the resulting process X is, by definition, not ergodic. However, even though π being ergodic for T_K does not imply that the resulting process is ergodic, the set of ergodic measures $J^E(K)$ satisfies a very interesting property.

Theorem. Any two ergodic measures π_1, π_2 for T_K are either identical or they are mutually singular (Hairer, 2018).

If K induces multiple ergodic densities, we can partition the state space into at least two subsets that do not "communicate" with each other. If this is the case, the corresponding process is called *reducible*.

Irreducibility. A time-homogenous Markov process X is said to be irreducible if there exists a non-trivial measure $\phi : \mathcal{S} \mapsto \mathbb{R}$ such that $\phi(E) > 0$ implies $\lim_{t\to\infty} K^t(x, E) > 0$ for each $x \in S$ (Tweedie, 1975)

Intuitively, irreducibility means that there exists a subset of the state space that is reachable from every initial condition. Therefore, irreducibility implies that a single kernel cannot induce multiple ergodic densities.

Corollary. If a process time-homogenous Markov process X is irreducible, the existence of an ergodic density π for T_K implies that X is ergodic.[16]

Remark. The limiting density of an ergodic process is ergodic for T_K.

Remark. An ergodic process need not be initialized with its unique limiting density π, which implies that it need not be stationary.

Stationarity vs. ergodicity

I end by contrasting the two notions of stationarity and ergodicity in the context of Markov processes. First, notice that neither stationarity implies ergodicity nor vice versa. For this, consider the following two illustrative examples.

A stationary, non-ergodic process. Consider a process X^A where $X_0^A \sim \mathcal{N}(0, 1)$ and $X_{t+1}^A = X_t^A$ for each $t \geq 0$.

[16]For necessary and sufficient conditions of ergodicity, geometric ergodicity, and strong ergodicity of a Markov process, see Hordijk (1992).

X^A is stationary because the joint density over X^A is invariant under the shift map. However, since the Markov kernel is degenerate as we have $K(x, E) = 1$ for each $E \in \mathcal{B}$ with $x \in E$ and zero otherwise, any observed orbit is perfectly uninformative. In particular, a single orbit does not yield enough information to recover the marginal density such that — even though the Markov kernel is trivial — it does not yield enough information to infer the process' joint density.

An ergodic, non-stationary process. Consider a process X^B where $X_0^B = a$ and $X_t^B = \varepsilon_t$, $\varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$ for each $t > 0$.

Since the limiting conditional density of X^B is trivially given by $\mathcal{N}(0,1)$ for any $a \in \mathbb{R}$, X^B is ergodic. The process is not stationary, however, because the marginal distributions $\Pr(X_1^B \in E)$ and $\Pr(X_2^B \in E)$ are not equivalent. This means that the joint density of the process is not invariant under a time shift, which means that it is not stationary. It is, however, asymptotically stationary as asymptotic stationarity is implied by ergodicity.

A final remark. The asymptotic nature of ergodicity serves as a temptation to confuse it with asymptotic stationarity: "In calculating asymptotic distributions, [geometric] ergodicity can substitute for stationarity since it means that the process converges [geometrically] to its stationary [ergodic] distribution" (Duffie and Singleton, 1993). While it is true that ergodicity implies asymptotic stationarity, it should be stressed that the former does much more than substitute for the latter. In particular, ergodicity ensures that a single, sufficiently long orbit is enough to recover the structure of an underlying data generating process, whereas (asymptotic) stationarity may produce orbits that are — as evidenced by X^A — perfectly uninformative.

References

Bhattacharya, Rabi and Mukul Majumdar. 2004. "Random Dynamical Systems: A Review." *Economic Theory* 23:13–38.

Blanchard, Olivier Jean and Charles M. Kahn. 1980. "The Solution of Linear Difference Equations under Rational Expectations." *Econometrica* 48 (5):1305–1312.

Chari, V.V., Patrick J. Kehoe, and Ellen R. McGrattan. 2009. "New Keynesian Models: Not Yet Useful for Policy Analysis." *American Economic Journal: Macroeconomics* 1 (1):242–266.

Christiano, Lawrence, Martin S. Eichenbaum, and Benjamin K. Johannsen. 2018. "Does the New Keynesian Model Have a Uniqueness Problem?" Working Paper 24612, National Bureau of Economic Research.

Duffie, Darrell and Kenneth J. Singleton. 1993. "Simulated Moments Estimation of Markov Models of Asset Prices." *Econometrica* 61 (4):929–952.

Frisch, Ragnar. 1936. "On the Notion of Equilibrium and Disequilibrium." *Review of Economic Studies* 3 (2):100–105.

Garín, Julio, Robert Lester, and Eric Sims. 2018. "Intermediate Macroeconomics." https://www3.nd.edu/~esims1/gls_int_macro.pdf. Accessed on 9/13/19.

Hairer, Martin. 2018. "Ergodic Properties of Markov Processes." http://www.hairer.org/notes/Markov.pdf. Accessed on 9/18/19.

Hamilton, James D. 2017. "Why you should never use the Hodrick-Prescott Filter." Working Paper 23429, National Bureau of Economic Research.

Hordijk, Arie and Flora Spieksma. 1992. "On Ergodicity and Recurrence Properties of a Markov Chain with an Application to an Open Jackson Networks." *Advances in Applied Probability* 24 (2):343–376.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo. 2002. "Production, Growth and Business Cycles: Technical Appendix." *Computational Economics* 20:87–116.

Kydland, Finn and Edward C. Prescott. 1977. "Rules rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85 (3):473–492.

Lucas, Robert E. 2001. "Professional Memoir." .

May, Robert M. 1977. "Thresholds and breakpoints in ecosystems with a multiplicity of stable states." *Nature* 269:471–477.

Nash, John. 1950. "Equilibrium points in n-person games." *Proceedings of the National Academy of Sciences* 36 (1):48–49.

Samuelson, Paul A. 1968. "What Classical and Neoclassical Monetary Theory Really was." *The Canadian Journal of Economics* 1 (1):1–15.

Sargent, Thomas. 1997. *Dynamic Macroeconomic Theory*. Harvard University Press.

Stokey, Nancy L., Robert E. Jr. Lucas, and Edward Prescott. 1999. *Recursive Methods in Economic Dynamics*. Harvard University Press.

Strogatz, Steven H. 2015. *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press.

Tweedie, Richard L. 1975. "Sufficient Conditions for Ergodicity and Recurrence of Markov Chains of a General State Space." *Stochastic Processes and their Applications* 3:385–403.

Varian, Hal R. 1993. "What Use is Economic Theory." *University of Michigan CenREST Working Paper Series* 93-14.