Notes on Ross’s Recovery Theorem*
(Started: September 20, 2014; Revised: July 4, 2015)

Some thoughts about Steve Ross’s “recovery theorem” triggered by Mike’s discussion of Borovicka, Hansen, and Scheinkman (“Misspecified recovery”) at Carnegie Mellon in September 2014. The origin is Ross (JF, 2013), but our primary sources are Ian Martin’s lucid slides (Martin and Ross, “The long bond,” 2013), the Hansen-Scheinkman collection (starting with Econometrica, 2009), and conversations with the always helpful Jarda Borovicka.

The issue is identification: Can we identify true probabilities and the pricing kernel from state prices. Ross states (abstract): “The Recovery Theorem enables us to ... determine the market’s forecast of returns and [the pricing kernel] from state prices alone.”

We would say instead that we generally need both cross section and time series data to identify a pricing model — the distribution of a pricing kernel, for example. A shortcut based on (say) cross-section information alone is bound to lead to mistakes somewhere in any model capable of accounting for the salient features of asset prices and returns: the equity premium, the term structure, and so on. But it’s like whack-a-mole: If you fix one thing, another pops up. A more subtle question is whether there are situations in which these mistakes are innocuous. We’re not sure.

1 The question

We start with an ergodic Markov environment with a state variable $x_t$ and transition probabilities $p(x_{t+1}|x_t)$. In this setting, let $q(x_t, x_{t+1})$ be the price now in state $x_t$ of one unit paid in state $x_{t+1}$ tomorrow. Under the conditions of the no-arbitrage theorem, such state prices exist and are strictly positive.

The question is whether we can find a unique factorization,

$$q(x_t, x_{t+1}) = p(x_{t+1}|x_t)m(x_t, x_{t+1}),$$

and compute the probabilities $p$ and pricing kernel $m$ from state prices $q$ alone.

We might, for example, compute state prices from observations of a broad enough cross-section of asset prices, perhaps through options via the Breeden-Litzenberger method (J Business, 1978). But what do we need to deduce $p$ and $m$? You might guess that observing one thing (state prices $q$) will not generally be enough to deduce two (the components $p$ and $m$). Or to put it differently: we’ll probably need some structure on the problem.

There’s a similar issue with risk-adjusted probabilities. If $b^1(x_t)$ is the price of a one-period riskfree bond in state $x_t$ — the sum or integral of $q(x_t, x_{t+1})$ over $x_{t+1}$ — then the risk-adjusted probabilities $p^*$ are defined by

$$q(x_t, x_{t+1}) = b^1(x_t)p^*(x_{t+1}|x_t).$$

*Working notes, no guarantee of accuracy or sense.
If we know the state prices, then we know the risk-adjusted probabilities, but knowing the risk-adjusted probabilities is generally not enough to compute the true probabilities. And we need the true probabilities to compute “the market’s forecast of returns.”

2 The iid case

The simplest example is one in which the state is iid. Or more precisely: suppose neither state prices nor conditional probabilities depend on the current state $x_t$. By construction, neither does the pricing kernel. The factorization is then $q(x_{t+1}) = p(x_{t+1})m(x_{t+1})$.

Consider the special case in which the set of states $x_{t+1}$ has finite dimension $n$. Then we have $n$ state prices and would like to deduce $n-1$ probabilities (they sum to one) and $n$ values of the pricing kernel. There are obviously lots of ways to do this: we have $n + (n - 1) - n = n - 1$ degrees of freedom. Evidently the cross-section of asset prices isn’t enough. Moving to a more complex state space won’t change this.

What we would typically do in applied work is estimate the probabilities from time series observations. Given some regularity, we can do this arbitrarily well if we have enough data, which allows us to deduce the pricing kernel from the ratio: $m(x_{t+1}) = q(x_{t+1})/p(x_{t+1})$. It seems clear, though, that we need the time series evidence: the cross section of state prices isn’t enough on its own.

3 Markov chain recovery

Here’s how the recovery theorem works. We extend the finite-state iid setup to a Markov chain of dimension $n$, as Ross and Martin-Ross do. Then $Q = [q_{ij}]$ and the factorization is $q_{ij} = p_{ij}m_{ij}$ for $i, j = 1, \ldots, n$. Counting knowns and unknowns again, we have $n^2$ knowns (the state prices $q_{ij}$) and $n^2 - n$ (the $p_{ij}$’s, whose rows sum to one) plus $n^2$ (the $m_{ij}$’s) unknowns. Obviously we’re short of knowns.

The recovery theorem reduces $m$ to dimension $n$. Think of $m$, for example, as the (scaled) ratio of marginal utilities in a model in which consumption is stationary: we have one marginal utility for each state, hence $n$ of them. More formally, the Perron-Frobenius theorem tells us that the matrix $Q$ has a dominant positive eigenvalue $\lambda$ and positive eigenvector $v$:

$$Qv = \lambda v. \quad (2)$$

If we define $D = \text{diag}(v)$ and $e$ is a vector of ones, then $v = De$ and $\lambda^{-1}D^{-1}QDe = e$. Which is to say: the matrix $\Pi = \lambda^{-1}D^{-1}QD$ is a stochastic matrix: its elements are positive and its rows sum to one.

If $P = [p_{ij}] = \Pi$, then we’ve recovered the probabilities from state prices alone. And given the probabilities, we can compute “the market’s forecast of returns.” We can also compute the pricing kernel from the ratio: $m_{ij} = q_{ij}/\pi_{ij}$. It’s almost magical how this works. The question, though, is what leads us to believe that $\Pi$ and $P$ are the same. Ian’s slides refer to this as the “maintained hypothesis.”

The iid case is illustrative. Here every row of $Q$ is the same. The Perron-Frobenius theorem
gives us $Qe = b^1 e$ where $b^1$ is the scalar constant one-period bond price and (again) $e$ is a vector of ones. $D$ is then the identity and $\Pi = (b^1)^{-1}Q$ gives us the risk-adjusted probabilities. Are the true and risk-adjusted probabilities the same? If so, the pricing kernel $m = b^1$ is constant and risk premiums are zero on all assets.

Ross (Theorem 2) has a nice variant of this result. He uses the weaker assumption that the bond price is the same in all states. Then $Qe = b^1 e$ and we’re back where we were, except that the rows of $Q$ (hence $\Pi$) can be different. This shares a feature with Campbell and Cochrane: We know that iid is sufficient for getting a flat constant yield curve, but these two examples show that it’s not necessary.

The interpretation, again, is that $\Pi$ gives us the true probabilities $P$. That’s not really an identification scheme, it’s simply a claim that there’s no difference between true and risk-adjusted probabilities so no identification scheme is needed. That seems to us a strange assumption (or “hypothesis”) in the world of asset pricing, where risk premiums are the subject of interest, but does it work? Our plan is to apply the recovery theorem to realistic theoretical environments and see what happens. But first some theory.

4 The Alvarez-Jermann/Hansen-Scheinkman decomposition

Hansen and Scheinkman (Econometrica, 2009), and before them Alvarez and Jermann (Econometrica, 2005), show us explicitly what Ross is recovering. This is done at a higher level of generality than Ross, but the ideas have the same Perron-Frobenius flavor. Some of what follows is cannibalized from our paper with Stan (“Sources of entropy,” JF, 2014).

Hansen and Scheinkman (2009, Section 6) show that since asset pricing is linear, Perron-Frobenius logic suggests we have a positive eigenvalue $\lambda$ and associated positive eigenfunction $v$ that solve

$$E_t(m_{t,t+1}v_{t+1}) = \lambda v_t. \quad (3)$$

Existence and uniqueness are hard to establish in general, but we can compute these objects in lots of models, including the popular class of exponential-affine models. Here a subscript $t$ denotes dependence on the state at date $t$; $v_t$, for example, stands for $v(x_t)$ and $m_{t,t+1}$ stands for $m(x_t,x_{t+1})$. Probabilities $p(x_{t+1}|x_t)$ are implicit in the expectation $E_t$.

One consequence is Alvarez and Jermann’s (2005) decomposition of the pricing kernel into permanent and transitory components. If we divide (3) by its rhs, we have

$$E_t [m_{t,t+1}v_{t+1}/(v_t \lambda)] = E_t(m^1_{t,t+1}) = 1.$$

where $m^1_{t,t+1} = m_{t,t+1}v_{t+1}/(v_t \lambda)$. Alvarez and Jermann call $m^1$ the permanent component of the pricing kernel. Hansen and Scheinkman refer to it as the martingale component. Whatever we call it, we can now factor the pricing kernel into $m^1$ and whatever’s left, which we label $m^2$ and refer to as the transitory component:

$$m_{t,t+1} = m^1_{t,t+1}m^2_{t,t+1},$$

$$m^2_{t,t+1} = \lambda v_t/v_{t+1}.$$
The second component is the (scaled) ratio of the stationary object $v_t$. In logs, it’s overdifferenced.

There’s an obvious connection between (2) and (3) that we can make clearer if we rewrite the latter as

$$\int_{x_{t+1}} q(x_t, x_{t+1}) v(x_{t+1}) \, dx_{t+1} = \int_{x_{t+1}} m(x_t, x_{t+1}) p(x_{t+1}|x_t) v(x_{t+1}) \, dx_{t+1} = \lambda v(x_t).$$

This is clunky notation, but you get the idea. Now take the middle expression and substitute the decomposition $m = m^1 m^2$:

$$m(x_t, x_{t+1}) p(x_{t+1}|x_t) = m^1(x_t, x_{t+1}) m^2(x_t, x_{t+1}) p(x_{t+1}|x_t) = m^2(x_t, x_{t+1}) \pi(x_{t+1}|x_t),$$

where $\pi = m^1 p$. Note that $\pi$ is a probability density function, or whatever flavor of that object we need, because $E_t(m^1_{t,t+1}) = 1$. Here $m^1$ serves as a change of measure from $p$ to $\pi$ — Sargent would call it a Radon-Nikodym derivative.

One way to express this result is to say that $p$ characterizes the distribution of $m$ and $\pi$ characterizes the distribution of $m^2$. In the terms of equation (1), the recovery theorem identifies the product $m^2 \pi$, not the product $mp$. The two products are the same, but their components need not be.

So when are the true probabilities $p$ and the recovered probabilities $\pi$ the same? Evidently when $m^1 = 1$. Is that reasonable? Well, the point of the Alvarez-Jermann paper was that there was more variability in $m^1$ than in $m^2$. In that sense, reducing variability in $m^1$ to zero seems like a bad idea. They also show that when $m^1 = 1$ the high return asset is the long bond: the bond with infinite maturity. We don’t see bonds of infinite maturity, but the bonds we do see have significantly lower average returns than equity indexes and a number of other assets. There’s some sampling variability to worry about, but the $m^1 = 1$ world doesn’t seem to be the one we live in.

To summarize:

- The AJ/HS decomposition tells us that Ross’s recovered probabilities $\pi$ aren’t the true probabilities $p$ unless $m^1 = 1$.
- When $m^1 = 1$ we get unrealistic risk premiums.

We’ll use examples to flesh out the connection between $p$ and $\pi$ in specific situations.

5 Example 1: representative agent models

The simplest case is an environment in which (the level of) consumption is stationary — say, $c(x_t)$ — and utility is additive with period utility $u(c)$ and discount factor $\beta$. A Lucas economy, in other words. Then

$$m(x_t, x_{t+1}) = \beta u'[c(x_{t+1})]/u'[c(x_t)].$$

Comparing to (3), we see that the eigenfunction is $v(x_t) = 1/u'[c(x_t)]$ and the eigenvalue is $\lambda = \beta$. The transitory component is

$$m^2(x_t, x_{t+1}) = \lambda v_t/v_{t+1} = \beta u'[c(x_{t+1})]/u'[c(x_t)] = m(x_t, x_{t+1}).$$
Evidently \( m^2 = 1 \) and the recovery theorem works as claimed. Ross describes the Markov chain version of this example in Section II.

This success, so to speak, comes with the counterfactual byproducts described by Alvarez and Jermann. It’s no accident that macro-finance models of asset prices start with stationary processes for consumption growth. If we put a unit root in (say) log consumption, and preferences are homothetic, the situation changes. The eigenfunction-eigenvalue solution will generally depend on both preferences and the stochastic process for consumption growth and we will not generally find that \( m^1 = 1 \).

Ross hints at this in Section IV, where he describes problems with his approach when consumption growth is iid. Suppose \( g_{t+1} = c_{t+1}/c_t \) is iid and preferences are recursive and homothetic. Then \( m_{t,t+1} \) is iid, too, and the solution to (3) is \( \lambda = b^1 \) (a constant) and \( v(x_t) = 1 \). In this case it’s the second component that’s constant: \( m_{t,t+1}^2 = b^1 \). All of the action is in \( m^1 \), which is then folded into the recovered probabilities: \( \pi = m^1 p \). In this case it’s impossible to get \( \pi = p \) unless (overall) \( m \) is constant, which eliminates risk premiums from the model. You might recognize most of this from Section 2.

6 Example 2: Vasicek model

Bond pricing zeroes in explicitly on the dynamics of the pricing kernel. The canonical model starts with

\[
\log m_{t-1,t} = \log \beta + \sum_{j=0}^{\infty} a_j w_{t-j} = \log \beta + a(B)w_t, \tag{4}
\]

where \( a_0 > 0 \), \( \sum_j a_j^2 < \infty \), and \( B \) is the lag or backshift operator. The innovations \( w \) are iid with mean zero, variance one, and (otherwise arbitrary) cumulant generating function \( k(s) = \log E(e^{sw}) \). By construction, \( k(0) = 0 \). If \( w \) is normal, \( k(s) = s^2/2 \).

In this model, forward rates are

\[
-f^0_n = \log \beta + k(A_n) = [a(B)/B^n] w_t
\]

for any \( n \geq 0 \) and \( A_n = \sum_{j=0}^{n} a_j \). Mean forward rates are therefore \( -E(f^0_n) = \log \beta + k(A_n) \) and mean forward spreads are \( E(f^0_n - f^0_0) = k(A_0) - k(A_n) \). If we choose parameters to approximate (i) mean forward rates (cross section) and (ii) the persistence and variability of interest rates (time series), we get \( a_0 \) large and \( a_j \) small and negative for \( j \geq 1 \). These are robust qualitative features needed to mimic what we see in bond yields. (We have lots of versions of this, including “Sources of entropy,” Section I.E.) A typical result is Figure 1; the Python code that generated it is attached at the end.

But what does this model suggest for the recovery theorem? If \( A_\infty \) exists, the eigenvalue \( \lambda \) and eigenfunction \( v_t \) are

\[
\log \lambda = \log \beta + k(A_\infty)
\]

\[
\log v_t = \sum_{j=0}^{\infty} (A_\infty - A_j)w_{t-j}.
\]
The permanent component and transitory components are therefore

\begin{align*}
\log m^1_{t,t+1} &= -k(A_\infty) + A_\infty w_{t+1} \\
\log m^2_{t,t+1} &= \log \beta + k(A_\infty) + (A_0 - A_\infty) w_{t+1} + [a(B)/B]_+ w_t.
\end{align*}

The restriction \( m^1 = 1 \) corresponds to \( A_\infty = 0 \). With our numbers, \( A_\infty \) is big and \( A_0 - A_\infty \) is small, so the first component is much more variable than the second. This is, of course, Alvarez and Jermann again: the action is in the first component.

As a result, the true distribution \( p \) and the recovered distribution \( \pi \) are wildly different. We see a vivid illustration of this in Figure 2. The true distribution here of \( \log m^1 \) is \( N(0, A_\infty^2) \). (We’ve set the mean equal to zero, which corresponds to setting \( \log \beta + [a(B)/B]_+ w_t = 0 \), a normalization.) The recovered distribution is that of \( \log m^2 \), which (with the same normalization) is \( N[k(A_\infty), (A_0 - A_\infty)^2] \). With our numbers, these are very different. There’s a small shift in the mean, the \( k(A_\infty) \) term. But the big difference is in the variances. With \( A_0 \) large and \( A_0 - A_\infty \) small, there’s a lot less dispersion in the recovered distribution, which is obvious in the figure.

Could we fix this up? Sure, but something else always seems to go wrong. If we want to make \( \pi \) more similar to \( p \), we need to drive \( A_0 - A_\infty \) closer to \( A_0 \). We can do that by making \( A_\infty \) small, but if we hold \( A_0 \) constant that requires larger (negative) values of the \( a_j \)’s, which generates excessive variability in interest rates. Martin and Ross find the same in their example. Ian notes in his slides: “Confronted with high risk premia in equity markets, the framework is forced to conclude that the riskless rate varies wildly.”

The larger issue is that we need both cross section and time series evidence to identify the parameters of the model. Cross section evidence tells us that \( a_0 \) is large: otherwise risk premiums are too small. Time series evidence tells us the \( a_j \)’s are small for \( j \geq 1 \), otherwise interest rates are too variable. We don’t see any way to avoid that. The recovery theorem produces probabilities, but they’re the probabilities of the transitory component of the pricing kernel, not the probabilities that show up in (1).

7 Example 3: more representative agent models

See Mike’s slides.

8 Python code

Code for Vasicek calibration and related figures:
Figure 1. Estimated moving average coefficients in the Vasicek model.
Figure 2. True and recovered distributions of log pricing kernel.