# HANDLING LOW FREQUENCIES AND INITIAL CONDITIONS

## 1. Implausible fit of deterministic components

- AR models, particularly VAR models or models with many lags, if estimated by methods that condition on initial observations (like OLS), tend to imply that $E_0[y_t]$, $t = 1, \ldots, T$, where $t = 1$ is the start of the sample on the left-hand-side variable, is an implausibly accurate predictor of the trend or long-run swings in the sample $y_1, \ldots, y_T$
- This happens because the criterion of fit applies no penalty to parameter values that make the initial conditions highly implausible as draws from the model's implied unconditional distribution for $y_t$. The model then attributes the low-frequency behavior of the data to a process, lasting through much or all of the sample, of slow return to "normalcy" from these exotic initial conditions.
- There is no logical contradiction here, but usually parameter values with these characteristics are not plausible.

## 2. Why worse as models get bigger?

- In a univariate, one-lag model, return-to-trend dynamics can only take the exponential form $(y_0 - Ey)\rho^t$.
- With $k$ lags, a univariate model can produce return-to-trend dynamics that are linear combinations of $k$ exponentials. In particular, if all the observations (including the initial $k$ observations) lie on a $k$'th order polynomial, the AR can predict them perfectly.
- A VAR with $k$ lags on $n$ variables has $kn$ roots and can fit perfectly an arbitrary collection of $kn$'th order polynomials.
- So the potential for implausibly precise forecasts from initial conditions grows rapidly with $n$ and $k$, and indeed in practice the problem is clearly worse in larger models.

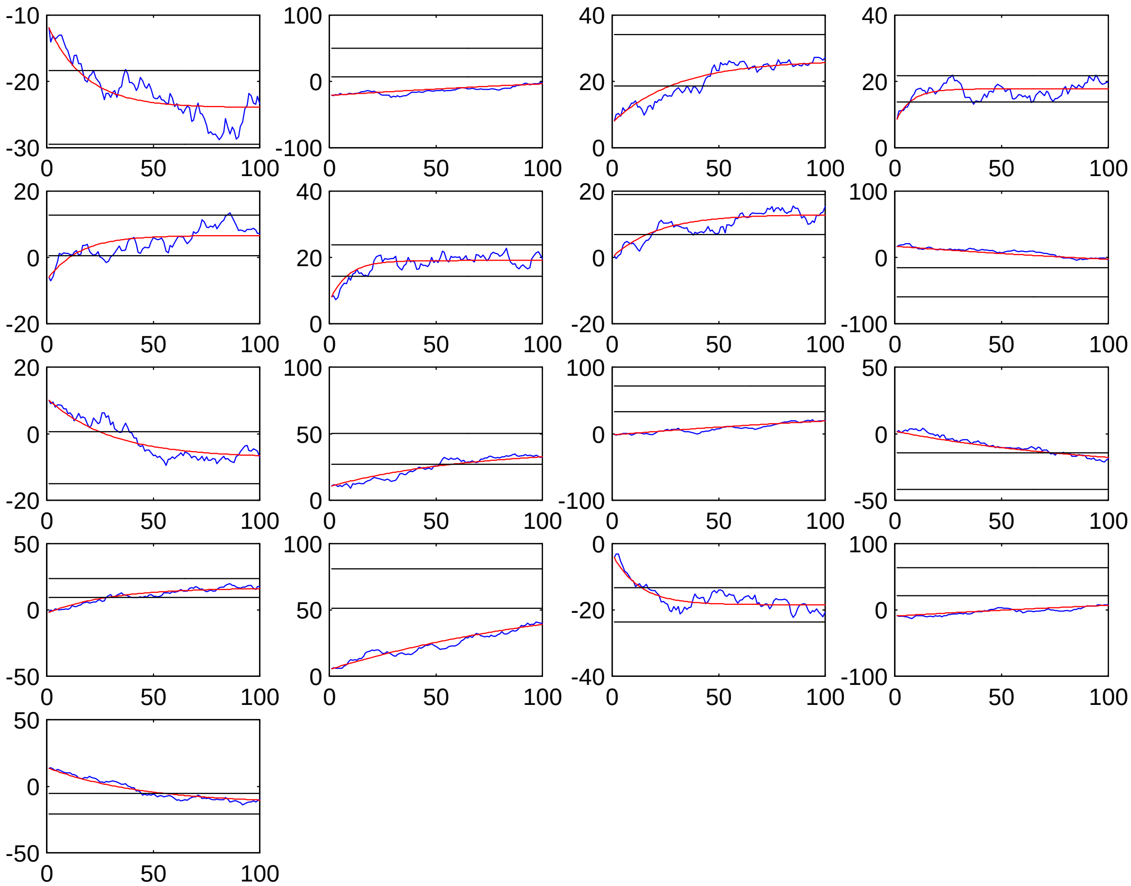
## 3. Remedies

20

- At least check for the problem: Use estimated coefficient values to construct $E_0 y_t$, plot these against actual data to see if the results make sense.
- Use the distribution of initial conditions in estimation.
- Use a prior that captures the idea that implausibly precise long run forecasts have low prior probability.

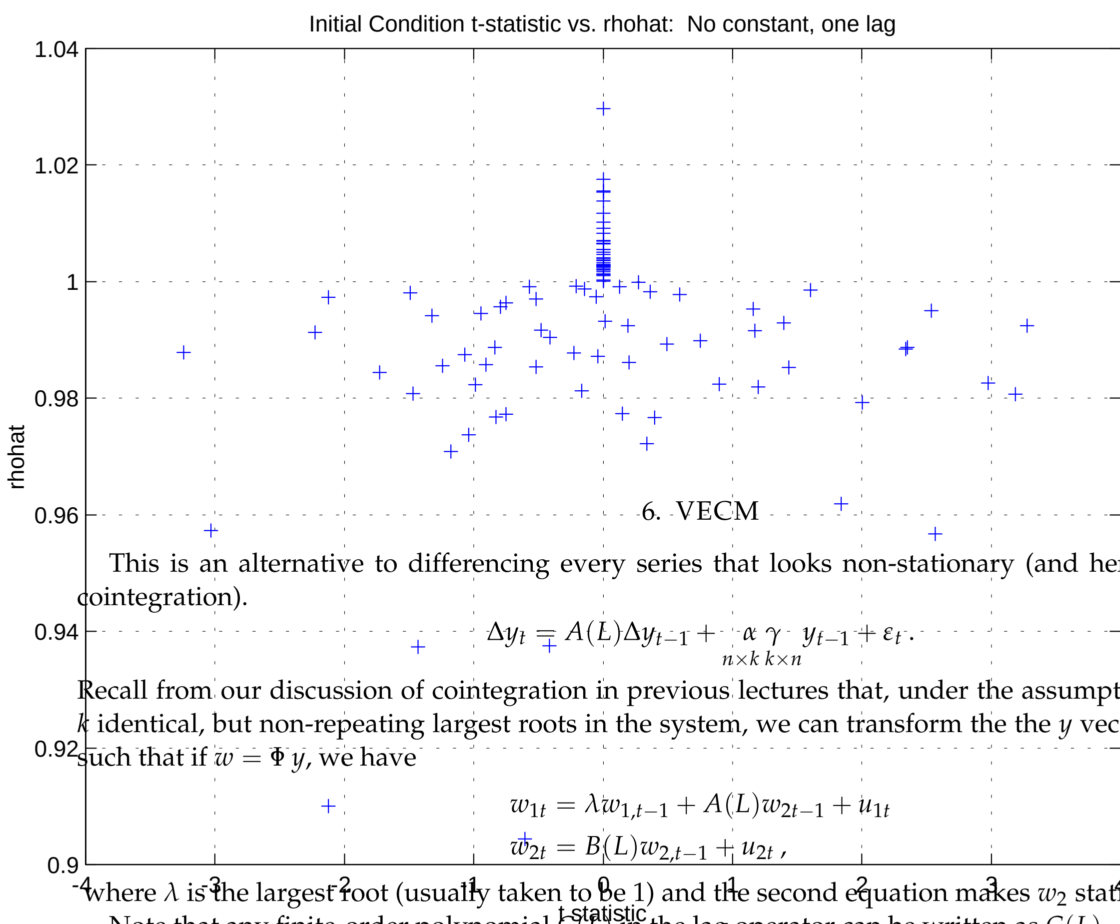
*Date*: January 18, 2009.

4



10 “worst cases” out of 100 random samples of random walks. Estimated unconditional mean and $2\sigma$ bands from fitted first order AR model shown as horizontal lines

5




## 6. VECM

This is an alternative to differencing every series that looks non-stationary (and hence contradicting cointegration).

$$\Delta y_t = A(L)\Delta y_{t-1} + \underset{n\times k}{\alpha}\,\underset{k\times n}{\gamma}\, y_{t-1} + \varepsilon_t\,.$$

Recall from our discussion of cointegration in previous lectures that, under the assumption that there are $k$ identical, but non-repeating largest roots in the system, we can transform the the $y$ vector by a matrix $\Phi$ such that if $w = \Phi\, y$, we have

$$w_{1t} = \lambda w_{1,t-1} + A(L)w_{2t-1} + u_{1t}$$
$$w_{2t} = B(L)w_{2,t-1} + u_{2t}\,,$$

where $\lambda$ is the largest root (usually taken to be 1) and the second equation makes $w_2$ stationary.

Note that any finite-order polynomial $C(L)$ in the lag operator can be written as $C(L) = c(L)(1-L) + F$, where $c(L)$ is a polynomial in the log operator of order one lower than the order of $C(L)$ and $F$ is a constant matrix. (It also can be written as $c^*(L)(1-L) + F^*L$, for example. The general point is that it can be written in differences, except for one term in levels.) Thus with $\lambda = 1$ the system above can also be written

$$\Phi_{\cdot 1}\Delta y_t = a(L)\Phi_{\cdot 2}\Delta y_t + F\Phi_{\cdot 2}y_t + u_{1t}$$
$$\Phi_{\cdot 2}\Delta y_t = b(L)\Phi_{\cdot 2}\Delta y_t + G\Phi_{\cdot 2}y_t + u_{2t}\,.$$

If we then multiply through by $\Phi^{-1}$, we get a system in VECM form with $\Phi_2$ playing thin previous lecturese role of $\gamma$ and $\Phi^{-1}\begin{bmatrix} F \\ G \end{bmatrix}$ playing the role of $\alpha$. [You might try writing out a small VAR system with one unit root and translating it back and forth between levels and VECM format.]

## 7. USING VECM

- The form requires a normalization, since if $\alpha$ is $n \times k$, $\alpha^* = \alpha w$ paired with $\gamma^* = w^{-1}\gamma$ produce the same model as $\alpha, \gamma$.
- It can be estimated by maximum likelihood, but the system must be considered jointly, since there are constraints running across equations.
- It constrains $k$ roots to be one, and non-repeating, but does not constrain the cointegrating vectors. This is appealing only if (as is rare in practice) we have firm reasons to believe we know that roots are either exactly one or stationary, and that we can specify the number of unit roots in the system in advance.
- We may sometimes have substantive reasons to know, or at least hypothesize, the number of unit roots and a likely form for the cointegrating vectors, as in the case of a collection of log price series that we think have one general-inflation unit root but have stationary relative prices. Then VECM is handy, and because $\gamma$ is known, the model is linear and satisfies the conditions for equation-by-equation estimation (same variables on the right-hand-side in every equation).
- More commonly it probably makes sense to estimate in levels, without the VECM constraints, and check for cointegration and cointegrating vectors from the estimated system matrix, as described in the notes on cointegration.