Financial Econometrics
Forecasting with Many Predictors

Nour Meddahi
Toulouse School of Economics

Winter 2025

Why Might You Want To Use Hundreds of Series?

One of the major challenges of empirical macro is that there is limited information and limited historical experience. But thousands of economic time series are available on line in real time. Can these be used to expand our information for economic monitoring and forecasting? For estimation of single and multiple equation models?
VARs with 6 variables and 4 lags have 144 coefficients (plus variances)

The curse part:

- VARs with 6 variables and 4 lags have $4 \times 62 = 144$ coefficients (plus variances)
- A VAR with 200 variables and 6 lags has 240,000 coefficients, and another 20,100 variance parameters.
- OLS is a bad idea with many regressors: in theory, if regressors are proportional to sample size, consistency is lost; in practice, the problem is introducing large estimation error.

The blessing part (one view)

In some models - dynamic factor models in particular - many series helps to identify the statistical object of interest - inference can be improved when there are many series.

From curse to blessing: dynamic factor models

Geweke (1977), Sargent and Sims (1977):

Suppose the n variables in X_t are related to q unobserved factors f_t, which evolve according to a time series process:

$$
\begin{aligned}
& X_{it} = \lambda_i(L) f_t + e_{it}, \quad i = 1, \cdots, n \\
& \Psi(L) f_t = \eta_t.
\end{aligned}
$$

- If the factors were observed they could be very useful for forecasting, but they aren't observed.
- The original approach to this problem (Engle and Watson (1981), Stock and Watson (1989, 1991), Sargent (1989), Quah and Sargent (1993)) was to fit the two equations above by ML using the Kalman filter.
- But the proliferation of parameters and computational limitations of ML in high dimensions limited this approach to small n.

An example following Forni and Reichlin (1998)

Suppose f_t is scalar and $\lambda_i(L) = \lambda_i$ ("no lags in the factor loadings"), so

$$X_{it} = \lambda_i f_t + e_{it}.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X_{it} = \left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\right) f_t + \frac{1}{n}\sum_{i=1}^{n} e_{it}.$$

If the errors e_{it} have limited dependence across series, then as n gets large,

$$\frac{1}{n}\sum_{i=1}^{n} X_{it} \xrightarrow{P} \bar{\lambda} f_t.$$

In this special case, a very easy nonparametric estimate (the cross-sectional average) is able to recover f_t as long as n is large!

- All the procedures below are justified using asymptotic theory for large n by assuming that $n \to +\infty$, usually at some rate relative to T . Often n^2/T is treated as large in the asymptotics; this makes sense in an application with $T = 160$ and $n = 130$, say.

- By having large n, procedures (more sophisticated than the simple average in the previous example) are available for consistent estimation of tuning priors (prior hyperparameters) in forecasting and for factors in DFMs.

- Most of the theory, and all of the empirical work, has been developed within the past 20 years.

3) Dynamic Factor Models: Specification and Estimation

(A) Specification: The DFM, the Static Form, and the Approximate DFM

The idea (conjecture) behind DFMs is that small number of factors captures the covariation in macro time series (Geweke (1977), Sargent and Sims (1977)).

The exact DFM:

$$
\begin{aligned}
X_{it} &= \lambda_i(L) f_t + e_{it}, \quad i = 1, \cdots, n \\
\Psi(L) f_t &= \eta_t,
\end{aligned}
$$

where $f_t =$ q unobserved \dynamic factors"

$\lambda_i(L) f_t$ *common component*"

$\lambda_i(L)$ \dynamic factor loadings" lag polynomial

e_{it} idiosyncratic disturbance

$Cov(f_t, e_{is}) = 0$, for all i, s

$E(e_{it} e_{js}) = 0, \imath \neq j$, for all t, s, (exact DFM)

DFM in vector notation: $\underset{n\times 1}{X_t} = \underset{n\times q}{\lambda(L)}\,\underset{q\times 1}{f_t} + \underset{n\times 1}{e_t}$

Identification of the factors: $\lambda(L)$ and f_t are only identified up to a normalization:

$$\lambda(L)f_t = \lambda(L)H \quad H^{-1}f_t$$

for any square matrix H. This is unimportant if you are only interested in the space spanned by the f's but it will come up in our discussion of FAVAR.

Forecasting in the exact DFM

Consider forecasting $X_{i,t+1}$ using all the data in X_t, and treat ft as observed. If e_{it} follows an autoregression and the errors are Gaussian, then

$$
\begin{aligned}
&E[X_{i,t+1} \mid X_t, f_t, X_{t-1}, f_{t-1}, \cdots] \\
&\qquad = E[\lambda_i(L) f_{t+1} \mid X_t, f_t, X_{t-1}, f_{t-1}, \cdots] + E[e_{i,t+1} \mid X_t, f_t, X_{t-1}, f_{t-1}, \cdots] \\
&\qquad = E[\lambda_i(L) f_{t+1} \mid f_t, f_{t-1}, \cdots] + E[e_{i,t+1} \mid e_{i,t}, e_{i,t-1} \cdots] \\
&\qquad = \alpha(L) f_t + \delta(L) X_{i,t}.
\end{aligned}
$$

- The f's contain all the relevant information from the other X's.
- The dimension reduction is from np parameters, to (q+1)p, where p is the number of lags.
- Under the DFM, the OLS dimension problem is eliminated and the forecast using the f's will be first order efficient.

The Static Form of the DFM ("little f and big F")

$$X_t = \lambda(L) f_t + e_t, \qquad \Psi(L) f_t = \eta_t.$$

Suppose that $\lambda(L)$ has at most p_f lags. Then the DFM can be written,

$$\begin{pmatrix} X_{1t} \\ \vdots \\ X_{nt} \end{pmatrix} = \begin{pmatrix} \lambda_{10} & \cdots & \lambda_{1p_f} \\ \vdots & \ddots & \vdots \\ \lambda_{n0} & \cdots & \lambda_{np_f} \end{pmatrix} \begin{pmatrix} f_t \\ \vdots \\ f_{t-p_f} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ \vdots \\ e_{nt} \end{pmatrix}$$

or

$$\underset{n\times 1}{X_t} = \underset{n\times r}{\Lambda} \; \underset{r\times 1}{F_t} + \underset{n\times 1}{e_t}$$

where the number of static factors, r, could be as much as qp_f.
F_t is the vector of **static factors**. The VAR for f_t implies that there is a VAR for F_t:

$$\Phi(L) F_t = G\eta_t$$

where G is a matrix of 1's and zeros and Φ consists of 1's, 0's, and Ψ's.

(B) Estimation: Principal Components, Generalized PC, and MLE

(i) Estimation by Principal Components DFM in static form:

$$X_t = \Lambda F_t + e_t$$
$$\Phi(L) F_t = G\eta_t.$$

By analogy to regression, consider estimating Λ and $\{F_t\}$ by least squares:

$$\underset{F_1,\cdots,F_T,\Lambda}{Min} \frac{1}{T}\sum_{t=1}^{T}(X_t - \Lambda F_t)'(X_t - \Lambda F_t),$$

subject to $\Lambda'\Lambda = I_r$ (identification). Given Λ, the (infeasible) OLS estimator of F_t is:

$$\hat{F}_t(\Lambda) = (\Lambda'\Lambda)^{-1}\Lambda' X_t.$$

Now substitute $\hat{F}_t(\Lambda)$ into the minimization problem to concentrate out $\{F_t\}$:

$$\underset{\Lambda}{Min} \frac{1}{T}\sum_{t=1}^{T} X_t'[I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda]X_t.$$

$$\underset{\Lambda}{Min}\frac{1}{T}\sum_{t=1}^{T} X_t'[I-\Lambda(\Lambda'\Lambda)^{-1}\Lambda]X_t$$

$$\Longleftrightarrow \underset{\Lambda}{Max}\frac{1}{T}\sum_{t=1}^{T} X_t'\Lambda(\Lambda'\Lambda)^{-1}\Lambda X_t$$

$$\Longleftrightarrow \underset{\Lambda}{Max}\ Tr\left\{(\Lambda'\Lambda)^{-1/2'}\Lambda'\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)\Lambda(\Lambda'\Lambda)^{-1/2}\right\}$$

$$\Longleftrightarrow \underset{\Lambda}{Max}\ \Lambda'\widehat{\Sigma}_{XX}\Lambda\ \ s.t.\ \ \Lambda'\Lambda = I_r, \quad \text{where} \quad \widehat{\Sigma}_{XX} = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$$

$$\Longrightarrow \widehat{\Lambda} = \text{first r eigenvectors of } \widehat{\Sigma}_{XX}.$$

Given that $\hat{F}_t(\Lambda) = (\Lambda'\Lambda)^{-1}\Lambda' X_t$, one gets

$$\begin{aligned}\hat{F}_t(\hat{\Lambda}) &= (\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\Lambda}' X_t = \hat{\Lambda}' X_t \\ &= \text{first r principal components of } X_t.\end{aligned}$$

Distribution theory for PC as factor estimator

Results for the exact static factor model:
Connor and Korajczyk (1986): consistency in the exact static FM with T fixed, $n \to +\infty$.

Selected results for the approximate DFM: $X_t = \Lambda F_t + e_t$
Typical conditions (Stock-Watson (2002), Bai-Ng (2002, 2006),...):

- (a) $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{P} \Sigma_F$ (stationary factors)
- (b) $\Lambda'\Lambda/n \to$ (or $\xrightarrow{P}$)Σ_Δ: Full rank factor loadings
- (c) e_{it} are weakly dependent over time and across series (approximate DFM)
- (d) F, e are uncorrelated at all leads and lags
- (e) n, T $\to \infty$, with a relative rate condition.

- Stock and Watson (2002a):
  - consistency in the approximate DFM, n, T $\rightarrow \infty$, no n/T restrictions
  - justify using $\hat{F}_t$ as a regressor without adjustment

- Bai and Ng (2006):
  - $n^2/T \rightarrow \infty$ (Not the principle of parsimony!)
  - asymptotic normality of PCA estimator of the common component at rate $\min(n^{1/2}, T^{1/2})$ in approximate DFM
  - improve upon Stock-Watson (2002a) rate for using $\hat{F}_t$ as a regressor
  - Method for constructing confidence bands for predicted value (these are for predicted value – not forecast confidence bands)

- Data irregularities probably are best handled parametrically in the State-Space (SS) setup using the KF.
- However the PC algorithm can be modified for data irregularities including mixed frequency data, see Stock and Watson (2002b).

Forecasting with estimated factors

The basic idea - using factors as predictors. Suppose the object is to forecast $X_{i,t}$ using estimated factors. According to the exact DFM theory, the (first order) optimal forecast is obtained from the regression implied by the model. The dynamic factors aren't observed, so this leads to the regression,

$$X_{i,t+1} = \alpha(L)\hat{F}_t + \delta(L)X_{i,t} + \xi_{i,t+1}.$$

In some cases you might think some other variables X_t are good predictors so you could augment this: ˆ

$$X_{i,t+1} = \alpha(L)\hat{F}_t + \delta(L)X_{i,t} + \xi_{i,t+1}.$$

If the number of regressors is small, this will yield first-order optimal forecasts.

Multiple horizon forecasts: Two choices for h-step ahead forecasting:

1. Direct forecasts:

$$X_{i,t+h} = \alpha_h(L)\hat{F}_t + \delta_h(L)X_{i,t} + \xi_{i,t+h}^{(h)}.$$

2. Iterated forecasts: Use the VAR structure of $(X_{i,t+1}, \hat{F}_{t+1})$:

$$\begin{aligned} X_{i,t+1} &= \alpha(L)\hat{F}_t + \delta(L)X_{i,t} + \xi_{i,t+1} \\ \Phi(L)\hat{F}_{t+1} &= G\eta_{t+1} = w_{t+1}. \end{aligned}$$

   Alternatively, the iterated forecasts can be implemented in the SS setup using the KF. The advantages and disadvantages of iterated versus direct are an empirical matter (see Marcellino, Stock, & Watson (2006), Pesaran, Pick, and Timmermann (2009)).

Forecast evaluation: out-of-sample methods.

Selecting the number of factors

DFM in static form: $X_t = \Lambda \underset{r\times 1}{F}_t + e_t$.

Big Question: What is r?

Will discuss:

1. Informal data analysis
2. Estimating the number of static factors
   a. Estimation of r.
   b. Testing $r = r_0$ against $r > r_0$.
3. Estimating the number of dynamic factors, q

Informal data analysis:

- Largest eigenvalues.
- scree plots: plots of ordered eigenvalues of $X'X/T$.
- fraction of trace R^2 explained.

Estimation Approach

Bai and Ng (2002) propose an estimator of r based on an information criterion; their main result is $\hat{r} \xrightarrow{P} r_0$ for the approximate DFM.

Digression on information criteria (IC) for lag length selection in an AR:

Consider the AR(p): $y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} +_t$

- Why not just maximize the R^2?
- IC trades off estimator bias (too few lags) vs. estimator variance (too many lags), from the perspective of fit of the regression:

Bayes Information Criterion: $BIC(p) = \ln\left(\frac{SSR(P)}{T}\right) + p\frac{\ln(T)}{T}.$

Akaike Information Criterion: $AIC(p) = \ln\left(\frac{SSR(P)}{T}\right) + p\frac{2}{T}$

Estimating the number of static factors

The Bai and Ng (2002) information criteria have the same form:

$$IC(r) = \ln\left(\frac{SSR(r)}{T}\right) + penalty(N, T, r).$$

Bai and Ng (2002) propose several IC's with different penalty factors that all produce consistent estimators of k. Here is the one that seems to work best in simulations (and is the most widely used in empirical work):

$$IC_{p2}(r) = \ln\left(V(r, \hat{F}^r)\right) + r\left(\frac{N+T}{NT}\right) \ \ln[min(N, T)],$$

where

$$\begin{aligned} V(r, \hat{F}^r) &= \underset{\Lambda}{Min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \lambda_i^{r\prime} \hat{F}^r\right)^2 \\ &= \underset{F_1, \cdots, F_T, \Lambda}{Min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t) \end{aligned}$$

and $\hat{F}_t^r$ are the PC estimates of r factors.

$$IC_{p2}(r) = \ln\left(V(r,\hat{F}^r)\right) + r\left(\frac{N+T}{NT}\right)\ \ln[min(N,T)],$$

$$V(r,\hat{F}^r) = \underset{\Lambda}{Min}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \lambda_i^{r'}\hat{F}^r\right)^2.$$

Comments:

- $\ln\left(V(r,\hat{F}^r)\right)$ is a measure of (trace) fit – generalizes ln(SSR/T) in the BIC.
- If $N = T$, then $r\left(\frac{N+T}{NT}\right)\ \ln[min(N,T)] = 2r\frac{\ln(T)}{T}$, which is 2 $\times$ the usual BIC penalty factor.
- Both N and T are in the penalty factor: you need $N, T \to \infty$.
- Bai and Ng's (2002) main result: $\hat{r} \xrightarrow{P} r_0$.
- Logic of proof is same as for BIC.

Stock and Watson (JBES, 2002)

$$y_{t+h}^{(h)} = (1200/h)\ln(IP_{t+h}/IP_t), \quad y_{t+h}^{(h)} = \alpha_h + \beta_h' F_t + \gamma_h(L) y_t + \varepsilon_{t+h}^{(h)}$$

Diffusion Index Forecast:
$\hat{y}_{T+h|T} = \hat{\alpha}_h + \sum_{j=1}^{m} \hat{\beta}_{hj}' \hat{F}_{T-j+1} + \sum_{i=1}^{p} \hat{\gamma}_{hi} y_{T-j+1}$

Autoregressive Forecast $\hat{y}_{T+h|T} = \hat{\alpha}_h + \sum_{i=1}^{p} \hat{\gamma}_{hi} y_{T-j+1}$

VAR Forecast: 3 variables: monthly growth in real activity; change in monthly inflation; change in the 90-day U.S. treasury bill rate.

Mutivariate Leading Indicator Forecast: 11 indicators:
$\hat{y}_{T+h|T} = \hat{\delta}_{0h} + \sum_{i=1}^{p} \hat{\delta}_{hi} W_{T-j+1} + \sum_{i=1}^{p} \hat{\gamma}_{hi} y_{T-j+1}$

Phillips Curve Forecast: indicators are unemployment rate and $m-1$ lags, relative price of food and energy (plus one lagg) and Gordon's (1982) variable that controls for the imposition and removal of Nixon wage and price controls.
Forecast comparison: Relative MSE to AR forecast and

$$y_{t+h}^{(h)} = \alpha \hat{y}_{t+h|t} + (1-\alpha)\hat{y}_{t+h|t}^{AR} + u_{t+h}^{(h)}.$$

Table 1. *Simulated Out-of-Sample Forecasting Results: Real Variables, 12-Month Horizon*

Forecast method	Industrial production Rel. MSE	Industrial production $\hat{\alpha}$	Personal income Rel. MSE	Personal income $\hat{\alpha}$	Mfg & trade sales Rel. MSE	Mfg & trade sales $\hat{\alpha}$	Nonag. employment Rel. MSE	Nonag. employment $\hat{\alpha}$
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	.86 (.27)	.57 (.13)	.97 (.21)	.52 (.15)	.82 (.25)	.63 (.17)	.89 (.23)	.56 (.14)
VAR	.97 (.07)	.75 (.68)	.98 (.05)	.68 (.34)	.98 (.04)	.73 (.58)	1.05 (.09)	.22 (.41)
Full dataset ($N = 215$)								
DI-AR, Lag	.57 (.27)	.76 (.13)	.77 (.14)	.76 (.13)	.48 (.25)	.99 (.15)	.91 (.13)	.63 (.18)
DI-AR	.63 (.25)	.71 (.12)	.86 (.16)	.61 (.12)	.57 (.24)	.84 (.18)	.99 (.31)	.51 (.20)
DI	.52 (.26)	.88 (.17)	.86 (.16)	.61 (.12)	.56 (.23)	.94 (.20)	.92 (.26)	.55 (.20)
Balanced panel ($N = 149$)								
DI-AR, Lag	.67 (.25)	.70 (.13)	.82 (.15)	.70 (.13)	.56 (.23)	.91 (.16)	.88 (.14)	.68 (.18)
DI-AR	.67 (.25)	.70 (.12)	.92 (.14)	.57 (.12)	.61 (.23)	.80 (.17)	.88 (.22)	.58 (.17)
DI	.59 (.25)	.81 (.17)	.92 (.14)	.57 (.12)	.57 (.23)	.91 (.18)	.84 (.21)	.62 (.16)
Stacked balance panel								
DI-AR	.65 (.25)	.70 (.12)	.93 (.15)	.56 (.12)	.61 (.22)	.89 (.19)	1.02 (.30)	.49 (.14)
DI	.62 (.25)	.81 (.18)	.93 (.15)	.56 (.12)	.66 (.21)	.85 (.20)	.95 (.24)	.53 (.14)
Full dataset; $m = 1$, $p = BIC$, k *fixed*								
DI-AR, $k = 1$	1.06 (.11)	.27 (.34)	1.03 (.08)	.34 (.41)	.98 (.06)	.63 (.46)	1.01 (.09)	.49 (.24)
DI-AR, $k = 2$	.63 (.25)	.76 (.14)	.78 (.14)	.77 (.14)	.53 (.24)	.93 (.15)	.77 (.13)	.82 (.15)
DI-AR, $k = 3$	.56 (.26)	.84 (.14)	.77 (.15)	.77 (.13)	.52 (.23)	.99 (.16)	.84 (.14)	.75 (.20)
DI-AR, $k = 4$	.54 (.26)	.85 (.14)	.76 (.15)	.78 (.14)	.51 (.23)	1.01 (.16)	.83 (.15)	.73 (.19)
Full dataset; $m = 1$, $p = 0$, k *fixed*								
DI, $k = 1$	1.03 (.07)	.30 (.49)	1.01 (.09)	.46 (.34)	.98 (.05)	.67 (.49)	1.01 (.09)	.48 (.24)
DI, $k = 2$	.55 (.25)	.89 (.15)	.78 (.14)	.76 (.13)	.57 (.24)	.95 (.17)	.78 (.13)	.83 (.16)
DI, $k = 3$	.51 (.25)	1.00 (.16)	.77 (.15)	.77 (.13)	.60 (.21)	1.02 (.19)	.84 (.14)	.76 (.19)
DI, $k = 4$	.49 (.25)	1.00 (.16)	.76 (.15)	.78 (.14)	.59 (.22)	1.03 (.20)	.82 (.15)	.75 (.18)
RMSE, AR Model	.049		.027		.045		.017	

Table 2. *Simulated Out-of-Sample Forecasting Results: Real Variables, 6- and 24-Month Horizons*

Forecast method	Industrial production		Personal income		Mfg & trade sales		Nonag. employment	
	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$
A. Horizon = 6 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	.70 (.25)	.68 (.13)	.83 (.15)	.64 (.11)	.77 (.19)	.68 (.14)	.75 (.19)	.67 (.12)
VAR	1.01 (.05)	.43 (.39)	.99 (.03)	.63 (.43)	.99 (.04)	.64 (.45)	1.06 (.07)	.12 (.34)
Full dataset ($N = 215$)								
DI-AR, Lag	.69 (.25)	.69 (.14)	.77 (.12)	.86 (.15)	.63 (.18)	.89 (.17)	.94 (.16)	.56 (.18)
DI-AR	.77 (.30)	.62 (.16)	.81 (.16)	.66 (.13)	.70 (.20)	.76 (.17)	1.02 (.32)	.49 (.19)
DI	.74 (.25)	.68 (.17)	.81 (.16)	.65 (.13)	.67 (.20)	.79 (.18)	.96 (.28)	.52 (.19)
Balanced panel ($N = 149$)								
DI-AR, Lag	.73 (.25)	.68 (.16)	.79 (.13)	.78 (.13)	.66 (.17)	.87 (.17)	.93 (.17)	.58 (.21)
DI-AR	.78 (.28)	.62 (.16)	.81 (.15)	.66 (.11)	.76 (.19)	.70 (.17)	.97 (.28)	.52 (.19)
DI	.73 (.24)	.69 (.15)	.81 (.15)	.66 (.11)	.68 (.19)	.81 (.17)	.95 (.26)	.53 (.18)
Full dataset; $m = 1$, $p =$ *BIC,* k *fixed*								
DI-AR, $k = 1$	.97 (.15)	.58 (.33)	.91 (.07)	.80 (.23)	.99 (.11)	.52 (.29)	.94 (.12)	.60 (.19)
DI-AR, $k = 2$	.67 (.22)	.77 (.15)	.76 (.11)	.90 (.14)	.64 (.18)	.86 (.16)	.84 (.13)	.71 (.16)
DI-AR, $k = 3$	.64 (.23)	.81 (.15)	.75 (.12)	.89 (.14)	.64 (.18)	.88 (.17)	.88 (.14)	.66 (.17)
DI-AR, $k = 4$	.64 (.23)	.80 (.15)	.74 (.13)	.87 (.14)	.63 (.18)	.87 (.15)	.91 (.16)	.60 (.18)
RMSE, AR Model	.030		.016		.028		.008	
B. Horizon = 24 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	1.09 (.28)	.45 (.14)	1.29 (.31)	.30 (.20)	1.08 (.21)	.45 (.14)	1.07 (.31)	.47 (.15)
VAR	1.01 (.10)	.44 (.48)	.98 (.06)	.63 (.34)	1.03 (.06)	.13 (.85)	1.06 (.13)	.35 (.31)
Full dataset ($N = 215$)								
DI-AR, Lag	.57 (.24)	.88 (.13)	.70 (.20)	.94 (.23)	.66 (.18)	.95 (.18)	.82 (.15)	.88 (.26)
DI-AR	.59 (.25)	.88 (.15)	.76 (.22)	.80 (.26)	.70 (.20)	.89 (.19)	.74 (.19)	.97 (.24)
DI	.55 (.26)	.91 (.14)	.76 (.22)	.80 (.25)	.70 (.20)	.89 (.19)	.74 (.19)	.97 (.24)
Balanced panel ($N = 149$)								
DI-AR, Lag	.57 (.25)	.87 (.14)	.76 (.19)	.86 (.23)	.64 (.20)	.94 (.18)	.74 (.17)	1.06 (.25)
DI-AR	.58 (.25)	.87 (.14)	.83 (.20)	.74 (.24)	.67 (.19)	.93 (.18)	.76 (.18)	.94 (.25)
DI	.58 (.25)	.87 (.14)	.83 (.20)	.74 (.24)	.67 (.20)	.94 (.19)	.75 (.18)	.94 (.24)
Full dataset; $m = 1$, $p =$ *BIC,* k *fixed*								
DI-AR, $k = 1$	1.12 (.19)	.10 (.46)	1.07 (.09)	.81 (1.00)	.97 (.04)	.90 (.62)	1.03 (.07)	.33 (.46)
DI-AR, $k = 2$	.76 (.19)	.68 (.11)	.88 (.13)	.68 (.17)	.65 (.20)	.87 (.14)	.72 (.16)	.99 (.17)
DI-AR, $k = 3$	.58 (.24)	.89 (.13)	.72 (.19)	.90 (.18)	.70 (.17)	.89 (.14)	.79 (.16)	.95 (.24)
DI-AR, $k = 4$	.56 (.24)	.90 (.14)	.70 (.20)	.93 (.23)	.67 (.18)	.95 (.18)	.78 (.16)	.96 (.24)
RMSE, AR Model	.075		.046		.070		.031	

Table 3. Simulated Out-of-Sample Forecasting Results: Price Inflation, 12-Month Horizon

Forecast method	CPI		Consumption deflator		CPI exc. food & energy		Producer price index	
	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	.79 (.15)	.76 (.15)	.95 (.12)	.58 (.17)	1.00 (.16)	.50 (.21)	.82 (.15)	.75 (.19)
Phillips Curve	.82 (.13)	.95 (.20)	.92 (.10)	.72 (.23)	.79 (.18)	.80 (.22)	.87 (.14)	.96 (.30)
VAR	.91 (.09)	.74 (.20)	1.02 (.06)	.45 (.20)	.99 (.05)	.56 (.21)	1.29 (.14)	.25 (.12)
Full dataset ($N = 215$)								
DI-AR, Lag	.72 (.14)	.91 (.14)	.90 (.09)	.65 (.13)	.84 (.15)	.76 (.20)	.83 (.13)	.78 (.21)
DI-AR	.71 (.16)	.83 (.13)	.90 (.10)	.62 (.13)	.85 (.15)	.74 (.20)	.82 (.14)	.75 (.20)
DI	1.30 (.16)	.34 (.08)	1.40 (.16)	.25 (.08)	1.55 (.31)	.24 (.06)	2.40 (.88)	.13 (.07)
Balanced panel ($N = 149$)								
DI-AR, Lag	.70 (.14)	.94 (.12)	.90 (.08)	.67 (.15)	.84 (.15)	.77 (.21)	.86 (.11)	.77 (.21)
DI-AR	.69 (.15)	.88 (.13)	.87 (.10)	.66 (.12)	.85 (.15)	.73 (.20)	.85 (.14)	.71 (.19)
DI	1.30 (.16)	.32 (.08)	1.34 (.13)	.26 (.09)	1.57 (.33)	.20 (.07)	2.44 (.87)	.14 (.06)
Stacked balance panel								
DI-AR	.73 (.15)	.82 (.12)	.87 (.09)	.65 (.12)	.85 (.15)	.77 (.21)	.81 (.14)	.75 (.20)
DI	1.54 (.31)	.28 (.08)	1.51 (.18)	.25 (.08)	1.55 (.32)	.23 (.06)	3.06(1.89)	.11 (.06)
Full dataset; $m = 1$, $p = BIC$, *k fixed*								
DI-AR, $k = 1$	.64 (.15)	1.14 (.14)	.77 (.12)	.96 (.16)	.71 (.17)	1.25 (.23)	.76 (.16)	.95 (.24)
DI-AR, $k = 2$	.67 (.14)	1.07 (.13)	.83 (.09)	.83 (.14)	.72 (.17)	.97 (.19)	.77 (.15)	.93 (.23)
DI-AR, $k = 3$	.76 (.13)	.91 (.15)	.94 (.07)	.61 (.14)	.86 (.14)	.73 (.20)	.86 (.11)	.78 (.21)
DI-AR, $k = 4$	.74 (.14)	.89 (.15)	.91 (.09)	.64 (.14)	.87 (.15)	.72 (.21)	.82 (.13)	.79 (.21)
Full dataset; $m = 1$, $p = 0$, *k fixed*								
DI, $k = 1$	1.60 (.34)	.25 (.07)	1.56 (.20)	.22 (.09)	1.55 (.31)	.23 (.06)	2.76(1.61)	.12 (.07)
DI, $k = 2$	1.56 (.31)	.26 (.07)	1.58 (.20)	.21 (.08)	1.62 (.39)	.22 (.07)	2.72(1.56)	.13 (.07)
DI, $k = 3$	1.57 (.32)	.24 (.08)	1.60 (.20)	.17 (.08)	1.69 (.43)	.18 (.07)	2.68(1.49)	.13 (.07)
DI, $k = 4$	1.56 (.25)	.25 (.07)	1.56 (.19)	.21 (.08)	1.67 (.40)	.19 (.07)	2.55 (.99)	.16 (.06)
RMSE, AR Model	.021		.015		.019		.033	

Table 4. Simulated Out-of-Sample Forecasting Results: Price Inflation, 6- and 24-Month Horizons

Forecast method	CPI Rel. MSE	CPI $\hat{\alpha}$	Consumption deflator Rel. MSE	Consumption deflator $\hat{\alpha}$	CPI exc. food & energy Rel. MSE	CPI exc. food & energy $\hat{\alpha}$	Producer price index Rel. MSE	Producer price index $\hat{\alpha}$
A. Horizon = 6 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	.82 (.12)	.78 (.16)	1.04 (.09)	.42 (.16)	1.10 (.16)	.32 (.27)	1.00 (.09)	.51 (.19)
Phillips Curve	.90 (.11)	.80 (.27)	.99 (.06)	.54 (.23)	.90 (.11)	.68 (.19)	1.02 (.04)	.34 (.37)
VAR	1.04 (.08)	.41 (.16)	1.15 (.07)	.08 (.20)	1.00 (.05)	.50 (.21)	1.34 (.16)	.19 (.12)
Full dataset ($N = 215$)								
DI-AR, Lag	.73 (.14)	1.05 (.18)	.91 (.08)	.71 (.17)	.83 (.13)	.89 (.25)	.87 (.11)	.87 (.26)
DI-AR	.74 (.14)	1.01 (.19)	.89 (.08)	.79 (.18)	.83 (.13)	.89 (.25)	.87 (.10)	.87 (.26)
DI	1.57 (.25)	.21 (.08)	1.68 (.26)	.10 (.08)	1.74 (.43)	.13 (.07)	2.42 (.74)	.05 (.07)
Balanced panel ($N = 149$)								
DI-AR, Lag	.79 (.13)	1.00 (.22)	.97 (.07)	.59 (.18)	.85 (.13)	.85 (.25)	.91 (.09)	.78 (.27)
DI-AR	.78 (.13)	.94 (.21)	.96 (.08)	.60 (.18)	.85 (.13)	.85 (.25)	.91 (.09)	.82 (.29)
DI	1.59 (.26)	.19 (.08)	1.64 (.21)	.09 (.08)	1.73 (.43)	.13 (.07)	2.42 (.70)	.07 (.07)
Full dataset; $m = 1$, $p = BIC$, k *fixed*								
DI-AR, $k = 1$	.71 (.14)	1.15 (.19)	.85 (.09)	.91 (.20)	.85 (.11)	1.13 (.29)	.85 (.12)	.90 (.26)
DI-AR, $k = 2$	.72 (.14)	1.03 (.18)	.88 (.08)	.78 (.17)	.80 (.13)	1.00 (.24)	.86 (.12)	.86 (.26)
DI-AR, $k = 3$	.76 (.13)	.97 (.18)	.93 (.08)	.66 (.17)	.86 (.12)	.82 (.25)	.91 (.10)	.76 (.26)
DI-AR, $k = 4$	.76 (.13)	.96 (.19)	.93 (.08)	.65 (.17)	.88 (.12)	.79 (.25)	.90 (.10)	.75 (.25)
RMSE, AR Model	.010		.007		.009		.017	
B. Horizon = 24 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	.70 (.21)	.76 (.12)	.70 (.20)	.78 (.11)	.99 (.29)	.51 (.25)	.65 (.22)	.84 (.19)
Phillips Curve	.84 (.12)	.77 (.08)	.81 (.15)	.80 (.09)	.72 (.21)	.93 (.19)	.77 (.19)	1.00 (.06)
VAR	.92 (.08)	.80 (.22)	.98 (.06)	.57 (.18)	1.00 (.06)	.49 (.34)	1.18 (.12)	.29 (.10)
Full dataset ($N = 215$)								
DI-AR, Lag	.74 (.23)	.74 (.18)	.75 (.16)	.79 (.13)	.92 (.26)	.58 (.28)	.82 (.14)	.68 (.12)
DI-AR	.75 (.25)	.67 (.16)	.71 (.21)	.73 (.12)	.96 (.33)	.53 (.27)	.77 (.17)	.68 (.13)
DI	1.18 (.22)	.40 (.12)	1.21 (.18)	.38 (.12)	1.40 (.22)	.30 (.07)	2.09 (.72)	.19 (.09)
Balanced panel ($N = 149$)								
DI-AR, Lag	.59 (.22)	.95 (.12)	.67 (.18)	.84 (.10)	.84 (.22)	.69 (.24)	.76 (.14)	.78 (.13)
DI-AR	.70 (.24)	.72 (.13)	.70 (.20)	.75 (.12)	.87 (.29)	.61 (.25)	.86 (.15)	.62 (.11)
DI	1.07 (.20)	.46 (.12)	1.08 (.18)	.45 (.12)	1.43 (.22)	.27 (.07)	2.10 (.70)	.19 (.08)
Full dataset; $m = 1$, $p = BIC$, k *fixed*								
DI-AR, $k = 1$	.63 (.20)	1.04 (.18)	.68 (.17)	.97 (.15)	.60 (.25)	1.12 (.20)	.73 (.17)	.93 (.22)
DI-AR, $k = 2$	.61 (.21)	1.07 (.17)	.72 (.16)	.92 (.13)	.64 (.24)	.96 (.17)	.68 (.19)	.97 (.20)
DI-AR, $k = 3$	.80 (.17)	.82 (.23)	.80 (.12)	.83 (.13)	.94 (.25)	.56 (.29)	.81 (.11)	.80 (.14)
DI-AR, $k = 4$	.76 (.20)	.81 (.21)	.74 (.15)	.83 (.14)	.92 (.26)	.59 (.29)	.78 (.14)	.78 (.14)
RMSE, AR Model	.052		.038		.046		.077	