Discussion

"Robust forecasting with many predictors"
by Dobrev and Schaumburg

Domenico Giannone
Université libre de Bruxelles, ECARES and CEPR

Seventh ECB Workshop on Forecasting Techniques
New directions for forecasting

Frankfurt am Main May 2012

The paper

- The goal: Forecasting in a data-rich environment
  - Many predictors
  - Many targets
- The problem: over-fitting due to high collinearity
- The proposed solution
  - Shrinkage estimator
  - Reduced rank regression
- Empirics: Forecasting the Macroeconomy and bonds returns
  - Shrinkage: from the curse to the blessing of dimensionality
  - Additional accuracy gains from Reduced Rank Regression
- Setting the degree of shrinkage and the rank of the regression
  - Assume a Dynamic Factor Structure
  - Analyze the distribution of the eigenvalues
- Interpreting the factors by imposing group membership

The discussion

- General comments
    - Very broad paper
    - Interesting and policy relevant problem
    - Competently executed empirical exercise
- The structure of the discussion
    - Shrinkage is indeed a powerful forecasting tool
    - Shrinkage, Reduced Rank and Dynamic Factor Models
    - Empirical issues

Forecasting with Many Predictors

Forecast y_t using a large information set

$$\Omega_T = \text{span}\left[Z_{T-s}; s = 0, 1, 2, ...\right]$$

where $Z_t = (z_{1t}, ..., z_{nt})'$.

$$\boxed{\hat{y}_{T+h|T} = \text{proj}\left[y_{T+h}|\Omega_T\right]}$$

The forecast

$$\hat{y}_{T+h|T} = \hat{\beta}_0' Z_T + ... + \hat{\beta}_p' Z_{T-p} = \hat{\beta}' X_T$$

where $\hat{\beta}$ is estimated using sample information

$$\{y_t, Z_t; t = 1, ..., T\}$$

$X_t = (Z_t', ..., Z_{t-p}')'$ $\quad$ $\hat{\beta} = (\hat{\beta}_0', ..., \hat{\beta}_p')'$

Forecasting with Many Predictors

Forecast y_t using a large information set

$$\Omega_T = \mathrm{span}\left[Z_{T-s}; s = 0, 1, 2, ...\right]$$

where $Z_t = (z_{1t}, ..., z_{nt})'$.

$$\boxed{\hat{y}_{T+h|T} = \mathrm{proj}\left[y_{T+h}|\Omega_T\right]}$$

The forecast

$$\hat{y}_{T+h|T} = \hat{\beta}_0' Z_T + ... + \hat{\beta}_p' Z_{T-p} = \hat{\beta}' X_T$$

where $\hat{\beta}$ is estimated using sample information

$$\{y_t, Z_t; t = 1, ..., T\}$$

$X_t = (Z_t', ..., Z_{t-p}')'$ $\quad$ $\hat{\beta} = (\hat{\beta}_0', ..., \hat{\beta}_p')'$

Forecasting with Many Predictors

Forecast y_t using a large information set

$$\Omega_T = \mathrm{span}\left[Z_{T-s}; s = 0, 1, 2, ...\right]$$

where $Z_t = (z_{1t}, ..., z_{nt})'$.

$$\boxed{\hat{y}_{T+h|T} = \mathrm{proj}\left[y_{T+h}|\Omega_T\right]}$$

The forecast

$$\hat{y}_{T+h|T} = \hat{\beta}_0' Z_T + ... + \hat{\beta}_p' Z_{T-p} = \hat{\beta}' X_T$$

where $\hat{\beta}$ is estimated using sample information

$$\{y_t, Z_t; t = 1, ..., T\}$$

$X_t = (Z_t', ..., Z_{t-p}')'$ $\quad \hat{\beta} = (\hat{\beta}_0', ..., \hat{\beta}_p')'$

Forecasting with Many Predictors

Forecast y_t using a large information set

$$\Omega_T = \mathrm{span}\left[Z_{T-s}; s = 0, 1, 2, ...\right]$$

where $Z_t = (z_{1t}, ..., z_{nt})'$.

$$\boxed{\hat{y}_{T+h|T} = \mathrm{proj}\left[y_{T+h}|\Omega_T\right]}$$

The forecast

$$\hat{y}_{T+h|T} = \hat{\beta}_0' Z_T + ... + \hat{\beta}_p' Z_{T-p} = \hat{\beta}' X_T$$

where $\hat{\beta}$ is estimated using sample information

$$\{y_t, Z_t; t = 1, ..., T\}$$

$X_t = (Z_t', ..., Z_{t-p}')'$ $\quad$ $\hat{\beta} = (\hat{\beta}_0', ..., \hat{\beta}_p')'$

Traditional time series methods

Estimate $\hat{\beta}$ via OLS, i.e. minimize the in-sample fit of the model:

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T-h} \left(y_{t+h} - \beta' X_t\right)^2$$

$$\Longrightarrow \boxed{\hat{\beta} = (X'X)^{-1}X'y} \Rightarrow \boxed{\hat{y}^{OLS}_{T+h|T} = \hat{\beta}' X_T}$$

where $X = (X_1, ..., X_{T-h})'$; $y = (y_{h+1}, ..., y_T)'$

Problem!! If the size information set (n) is too large relative to the sample size (T) then OLS forecasts are poor or unfeasible: *curse of dimensionality.*

Traditional time series methods

Estimate $\hat{\beta}$ via OLS, i.e. minimize the in-sample fit of the model:

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T-h} \left(y_{t+h} - \beta' X_t\right)^2$$

$$\Longrightarrow \boxed{\hat{\beta} = (X'X)^{-1}X'y} \Rightarrow \boxed{\hat{y}^{OLS}_{T+h|T} = \hat{\beta}' X_T}$$

where $X = (X_1, ..., X_{T-h})'$; $y = (y_{h+1}, ..., y_T)'$

Problem!! If the size information set (n) is too large relative to the sample size (T) then OLS forecasts are poor or unfeasible: *curse of dimensionality.*

Traditional time series methods

Estimate $\hat{\beta}$ via OLS, i.e. minimize the in-sample fit of the model:

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T-h} \left(y_{t+h} - \beta' X_t\right)^2$$

$$\Longrightarrow \boxed{\hat{\beta} = (X'X)^{-1}X'y} \Rightarrow \boxed{\hat{y}_{T+h|T}^{OLS} = \hat{\beta}' X_T}$$

where $X = (X_1, ..., X_{T-h})'$; $y = (y_{h+1}, ..., y_T)'$

Problem!! If the size information set (n) is too large relative to the sample size (T) then OLS forecasts are poor or unfeasible: *curse of dimensionality*.

Curse of dimensionality (solutions)

- Principal components regression
  (Forni, Hallin, Lippi and Reichlin; Stock and Watson; Bai and Ng...)
- Bayesian regression
  (De Mol, Giannone and Reichlin, 2008)

Solution 1: Principal components regression

PCA aims at building r uncorrelated linear combinations $z_{1t}, ..., z_{rt}$ of a set of observable random variables $x_{1t}, ..., x_{nt}$, which explain most of the variance.

$$\hat{z}_1, \hat{\alpha}_1 = \arg\min_{z_1,\alpha_1} \sum_{i=1} \|X - \alpha_1' z_1\|^2$$
$$s.t. \;\; z_1' z_1 / T = 1$$

$$\hat{z}_2, \hat{\alpha}_2 = \arg\min_{z_2,\alpha_2} \|X - \hat{\alpha}_1' \hat{z}_1 - \alpha_2' z_2\|^2$$
$$s.t. \;\; z_2' \hat{z}_1 / T = 0 \text{ and } z_2' z_2 / T = 1$$

$\dots$

Sample covariance matrix: $\hat{\Sigma} = \frac{1}{T-h-p} X'X$

Spectral decomposition: $\hat{\Sigma} v_j = v_j d_j \quad d_1 \geq d_2 \geq ... \geq d_{n(p+1)}$

Normalized principal components: $\hat{z}_{jt} = \frac{1}{\sqrt{d_j}} v_j' X_t$

$\frac{d_1 + ... + d_r}{d_1 + ... + d_r + d_{r+1} + ... + d_n}$: percentage of variance explained by $\hat{z}_1, ..., \hat{z}_r$ (average R^2)

Principal components regression

Forecast based on the first r PC of the predictors:

$$\hat{y}^{PC}_{T+h|T} = \sum_{j=1}^{n(p+1)} w_j \hat{\alpha}_j \hat{z}_{jT}$$

$\hat{\alpha}_j$ OLS coefficients of the y over z_j

$$w_j = \begin{cases} 1 & \text{if } j \leq r \\ 0 & \text{otherwise} \end{cases}$$

Remark: $w_j = 1, \forall j \implies$ OLS
$\Rightarrow$ PC regression give weight only to linear combinations of the predictors that account for most of their fluctuations
$\Rightarrow$ it captures the **large/pervasive** driving forces

Bayesian regression with Gaussian prior

Idea: shrink regression coefficients to zero via priors (limit length β) + estimate coefficients as the posterior mode to compute forecast

$$\boxed{y_{t+h} = \beta' X_t + u_{t+h}}$$

Gaussian prior: u_t i.i.d. $\mathcal{N}(0, \sigma_u^2)$, $\beta \sim \mathcal{N}(0, \Phi_0)$

$$\Rightarrow \boxed{\hat{\beta}^{bay} = \left(X'X + \sigma_u^2 \Phi_0^{-1}\right)^{-1} X'y} \Rightarrow \boxed{\hat{y}_{T+h|T}^{bay} = \hat{\beta}'^{bay} X_T}$$

Solution 2: Bayesian regression with Gaussian prior

A simple case:

i.i.d prior on β: $\Phi_0 = \sigma_\beta^2 I$

$\Leftrightarrow$ ridge: penalized regression (L_2 norm)

$$\hat{\beta}^{bay} = \arg\min_{\beta} \sum_{t=1}^{T-h} \left(y_{t+h} - \beta' X_t\right)^2 + \nu \sum_{i=1}^{n(p+1)} \beta_i^2$$

- penalization parameter: $\nu = \frac{\sigma_u^2}{\sigma_\beta^2}$

$$\boxed{\hat{\beta}^{bay} = \left(X'X + \nu I\right)^{-1} X'y}$$

For the special case of iid gaussian prios, there is a simple relation between OLS, PC and Bayesian regression $\Rightarrow$ Bayesian regression is a weighted sum of projections on PC

Bayesian regression with Gaussian prior

Rewriting the Ridge in terms of principal components, we have:

$$\hat{y}^{bay}_{T+h|T} = \underbrace{\sum_{j=1}^{n(p+1)} w_j \hat{\alpha}_j \hat{z}_{jT}}_{\mathrm{proj}\left[y_{T+h}|\Omega_T\right]}$$

where the weights are:

$$w_j = \frac{d_j}{d_j + \frac{\nu}{T-h-p}}$$

⇒ Like PC regression, Bayesian regression give more weight to linear combination of the data that explain most of the overall variance

⇒ Like PC regression, Bayesian regressions capture the **large/pervasive** driving forces

A model for collinearity: The Dynamic Factor Model

$$X_t = \Lambda f_t + e_t$$

- f_t: (stationary) common factors, $\mathrm{E}[f_t f_t'] = I_r$
- e_t: (stationary) idiosyncratic component, $\mathrm{E}[e_t e_t'] = \Psi$
- $\mathrm{E}[f_t e_t'] = 0$
- $\Rightarrow \mathrm{E}[X_t X_t'] = \Lambda\Lambda' + \Psi$

Assumption (Approximate factor model)

$$0 < \underline{\lambda} < \liminf_{n\to\infty} \frac{1}{n}\lambda_{min}(\Lambda'\Lambda) \leq \limsup_{n\to\infty} \lambda_{max}\frac{1}{n}(\Lambda'\Lambda) < \bar{\lambda} < \infty$$

$$0 < \underline{\psi} < \liminf_{n\to\infty} \lambda_{min}(\Psi) \leq \limsup_{n\to\infty} \lambda_{max}(\Psi) < \bar{\psi} < \infty$$

$$\Rightarrow \quad \lambda_j(\Lambda\Lambda' + \Psi) \sim \begin{cases} n & \text{if } j \leq r \\ \text{bounded} & \text{otherwise} \end{cases}$$

A model for collinearity: The Dynamic Factor Model

$$X_t = \Lambda f_t + e_t$$

- f_t: (stationary) common factors, $\mathrm{E}[f_t f_t'] = I_r$
- e_t: (stationary) idiosyncratic component, $\mathrm{E}[e_t e_t'] = \Psi$
- $\mathrm{E}[f_t e_t'] = 0$
- $\Rightarrow \mathrm{E}[X_t X_t'] = \Lambda\Lambda' + \Psi$

Assumption (Approximate factor model)

$$0 < \underline{\lambda} < \liminf_{n\to\infty} \frac{1}{n}\lambda_{min}\left(\Lambda'\Lambda\right) \leq \limsup_{n\to\infty} \lambda_{max}\frac{1}{n}\left(\Lambda'\Lambda\right) < \bar{\lambda} < \infty$$

$$0 < \underline{\psi} < \liminf_{n\to\infty} \lambda_{min}\left(\Psi\right) \leq \limsup_{n\to\infty} \lambda_{max}\left(\Psi\right) < \bar{\psi} < \infty$$

$$\Rightarrow \quad \lambda_j(\Lambda\Lambda' + \Psi) \sim \quad \begin{cases} n & \text{if } j \leq r \\ \text{bounded} & \text{otherwise} \end{cases}$$

Bayesian Shrinkage with Collinear Regressors

Characterizing asymptotic collinearity (DFM)

$$\lambda_j(\Sigma) \sim \begin{cases} n & \text{if } j \leq r \\ \text{bounded} & \text{otherwise} \end{cases}$$

$$\Rightarrow \quad d_j \sim \begin{cases} n & \text{if } j \leq r \\ \frac{n}{\sqrt{T}} & \text{otherwise} \end{cases}$$

Asymptotic behavior of shrinkage estimator:

$$\boxed{\text{Recall: } y^{bay}_{T+h|T} = \sum_{j=1}^{n} w_j \hat{\alpha}_j f_{jT}; \quad w_j = \frac{d_j}{d_j + \frac{\nu}{T}}}$$

setting $\frac{\nu}{nT} \to 0$ and $\frac{\nu}{n\sqrt{T}} \to \infty \Rightarrow y^{bay}_{T+h|T} \to y^{pc}_{T+h|T}$

$\Rightarrow \nu = \frac{\sigma_u^2}{\sigma_\beta^2} \sim cnT^{1/2+\delta}$ does the job!!

[As $n \uparrow, \nu \uparrow$ (or $\sigma_\beta \downarrow$) $\to$ more shrinkage]

Factor models and Shrinkage: additional insights

$X_t = \Lambda f_t + e_t$
$Y_{t+h} = \Gamma f_t + u_{t+h}$
Assume:
$\mathrm{E}[f_t f_t'] = I_r \ \mathrm{E}[e_t e_t'] = \Psi \ , \ \mathrm{E}[u_t u_t'] = \Phi, \ u_t \perp e_t, \ u_t \perp f_t \ e_t \perp f_t$
This implies:

$$\mathrm{Proj}[Y_t | X_t] = BX_t$$

with $B = \Gamma(\Lambda'\Psi^{-1}\Lambda + I_r)^{-1}\Lambda'\Psi^{-1} = \Gamma(\Lambda\Lambda' + \Psi)^{-1}\Lambda$

Two observations:

- $\|B_i\|_2^2 = O\left(\frac{1}{n}\right)$ if $(\Lambda'\Psi^{-1}\Lambda)^{-1} = O\left(\frac{1}{n}\right)$: this motivates shrinkage
- $B = \alpha\beta'$: this might motivate Reduced Rank regression

Factor models and Shrinkage:... Intuition

Large PC capture common pervasive forces driving macroeconomic fluctuations.

• Ridge gives more weight to large/pervasive forces underlying the predictors

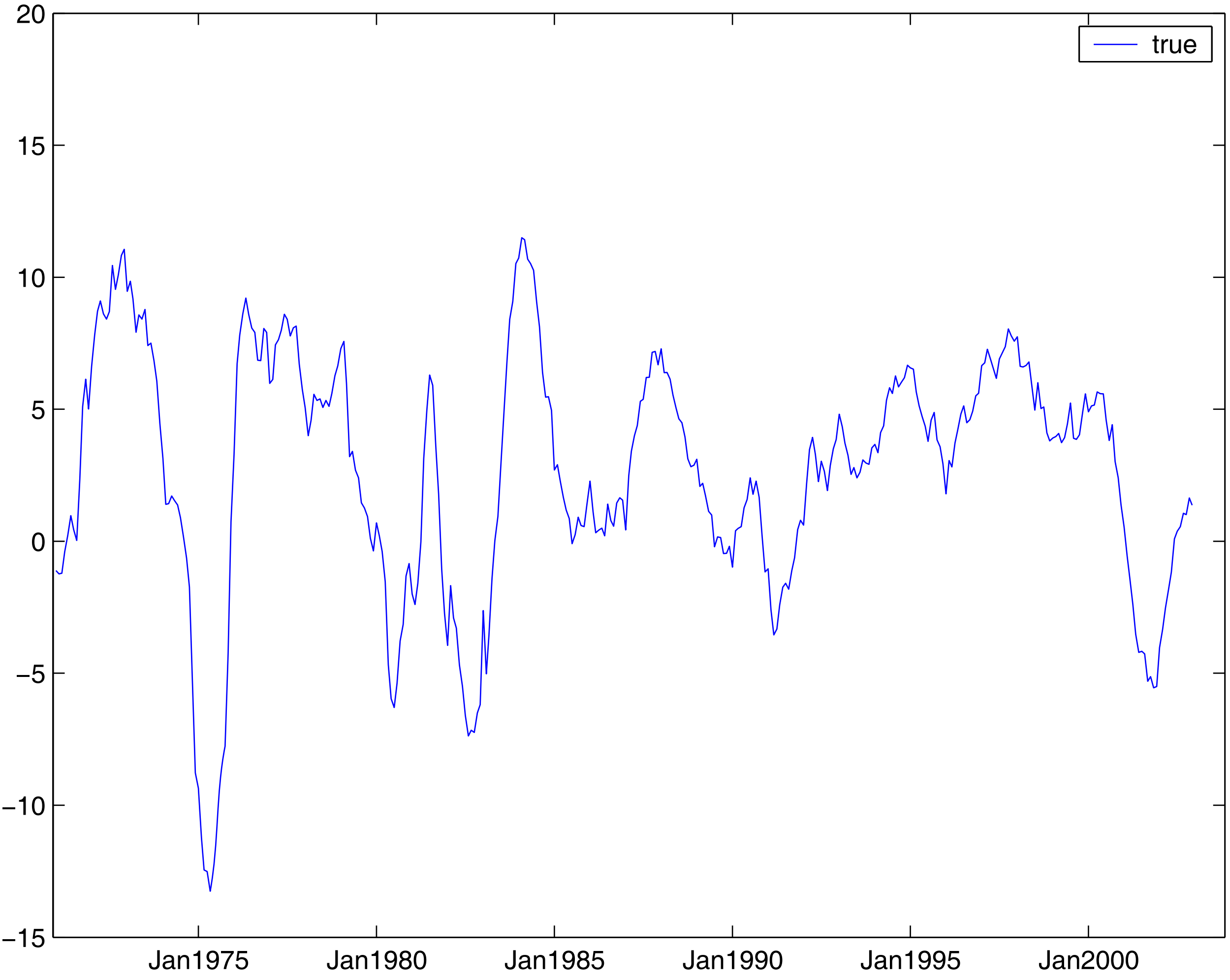
⇒ If there are only few large/pervasive factors

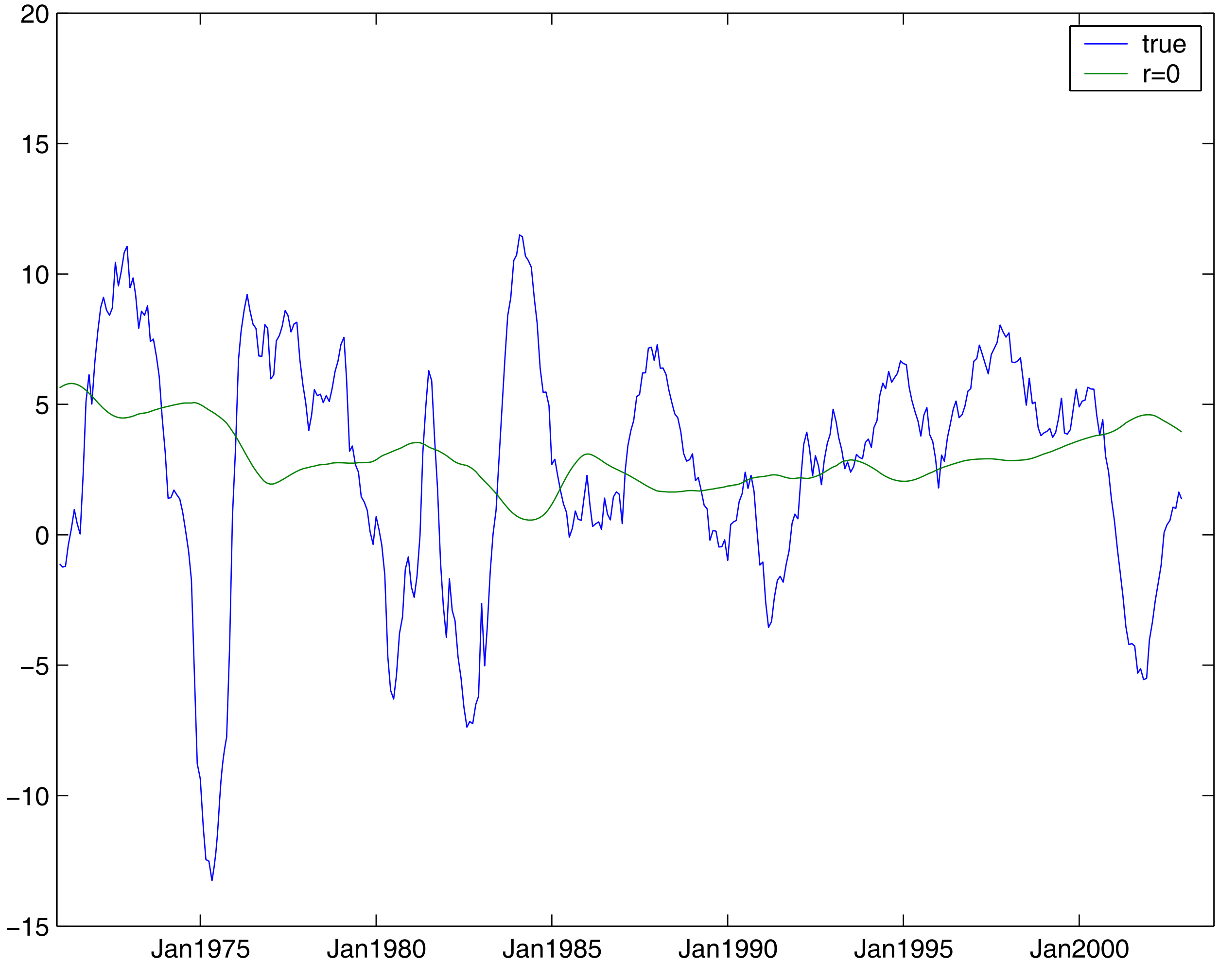
⇒ consistency

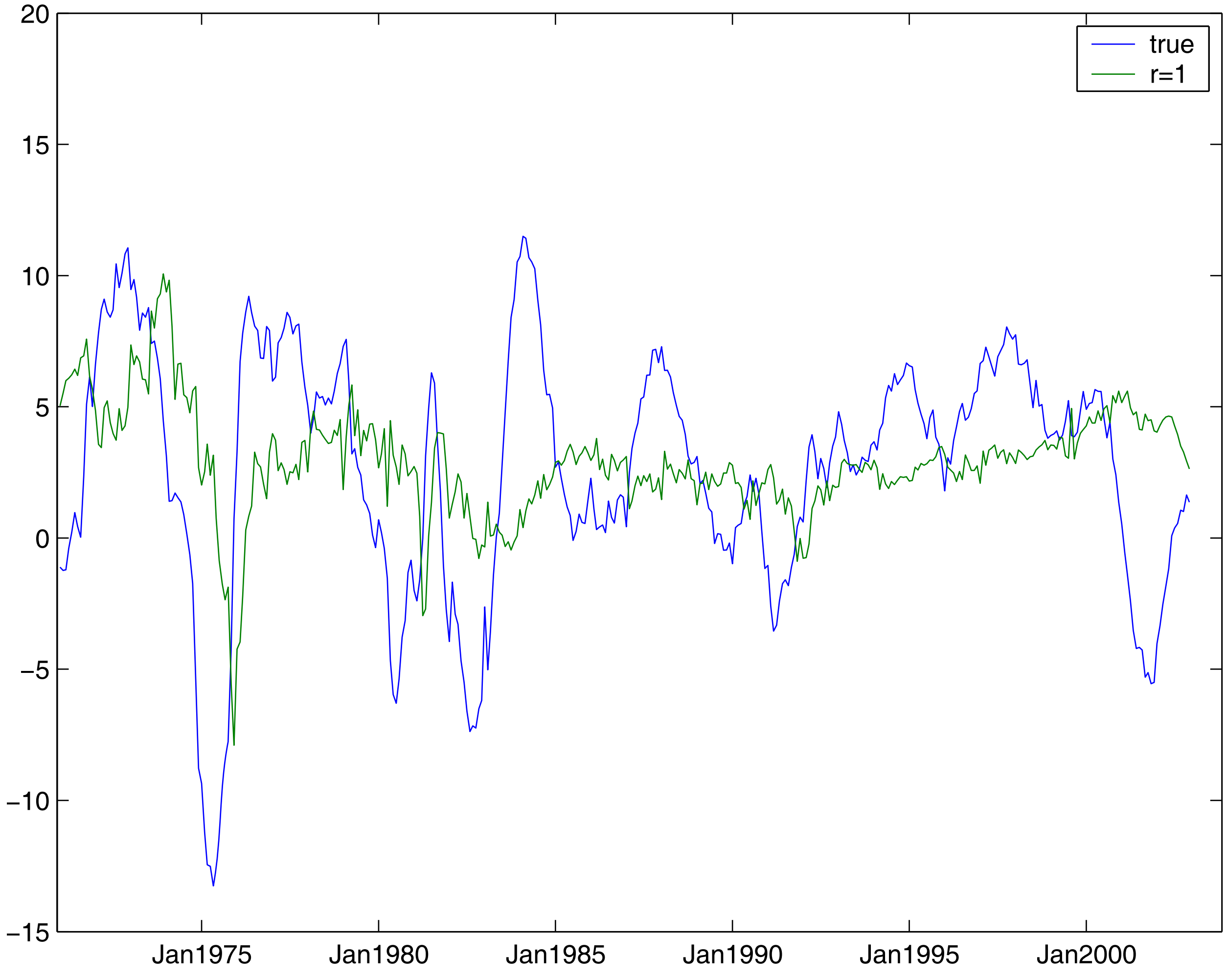
• If the common forces are pervasive, all variables contain relevant info since they are all affected by the factors

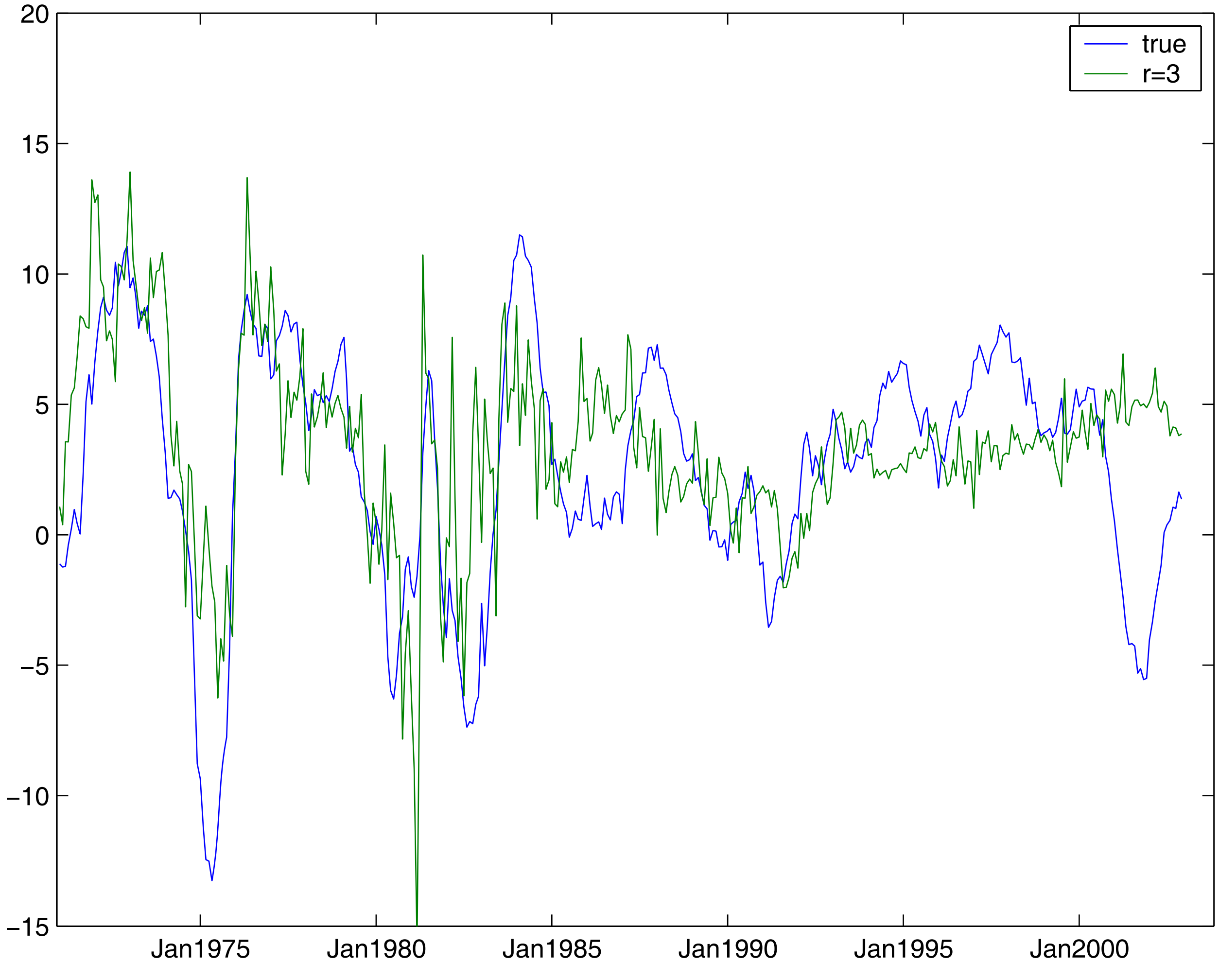
⇒ we should weight all of them, but should **use a prior that shrinks increasingly more coefficients as *n* increases**

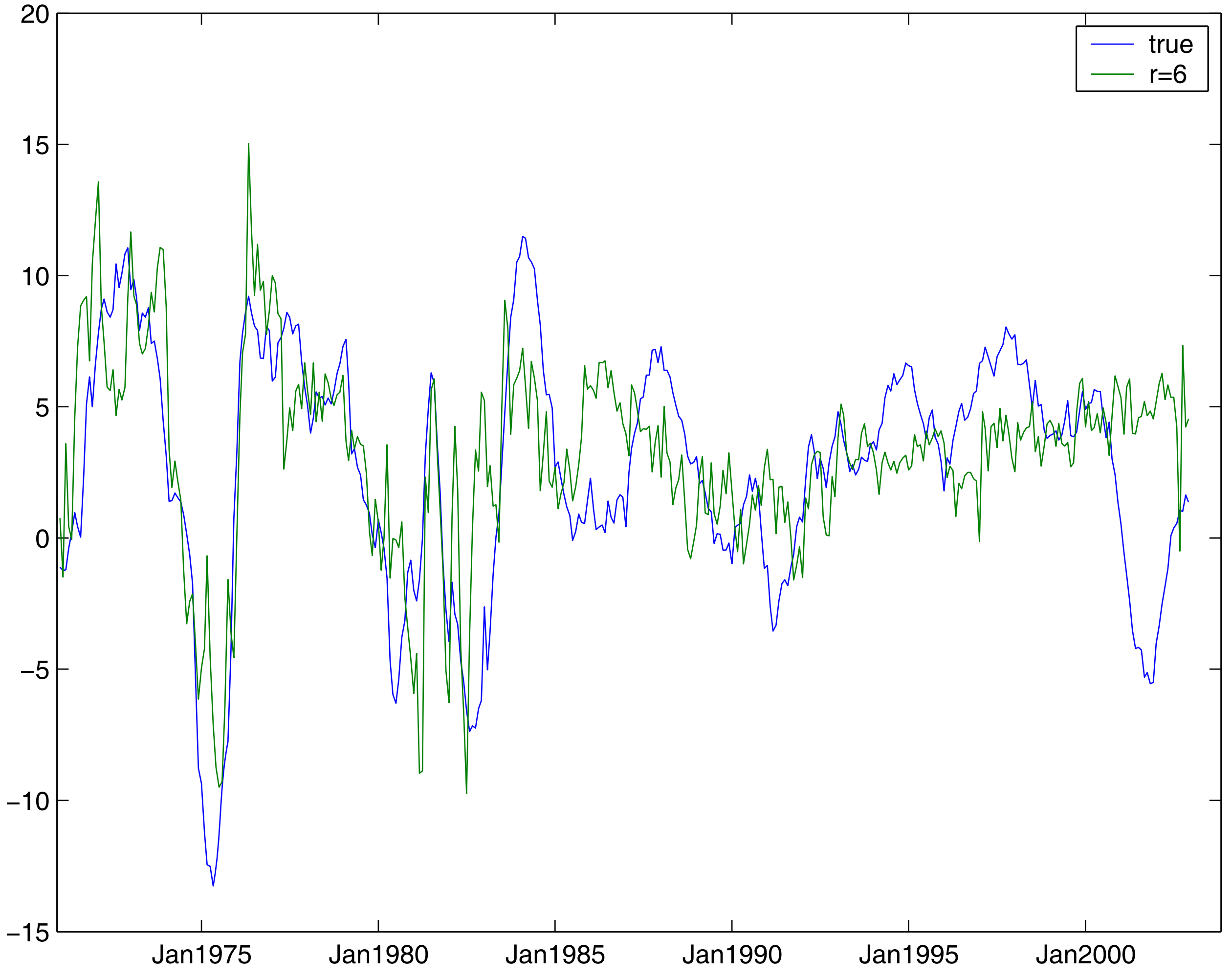
For asymptotics, see De Mol, Giannone and Reiclin, JoE 2008.

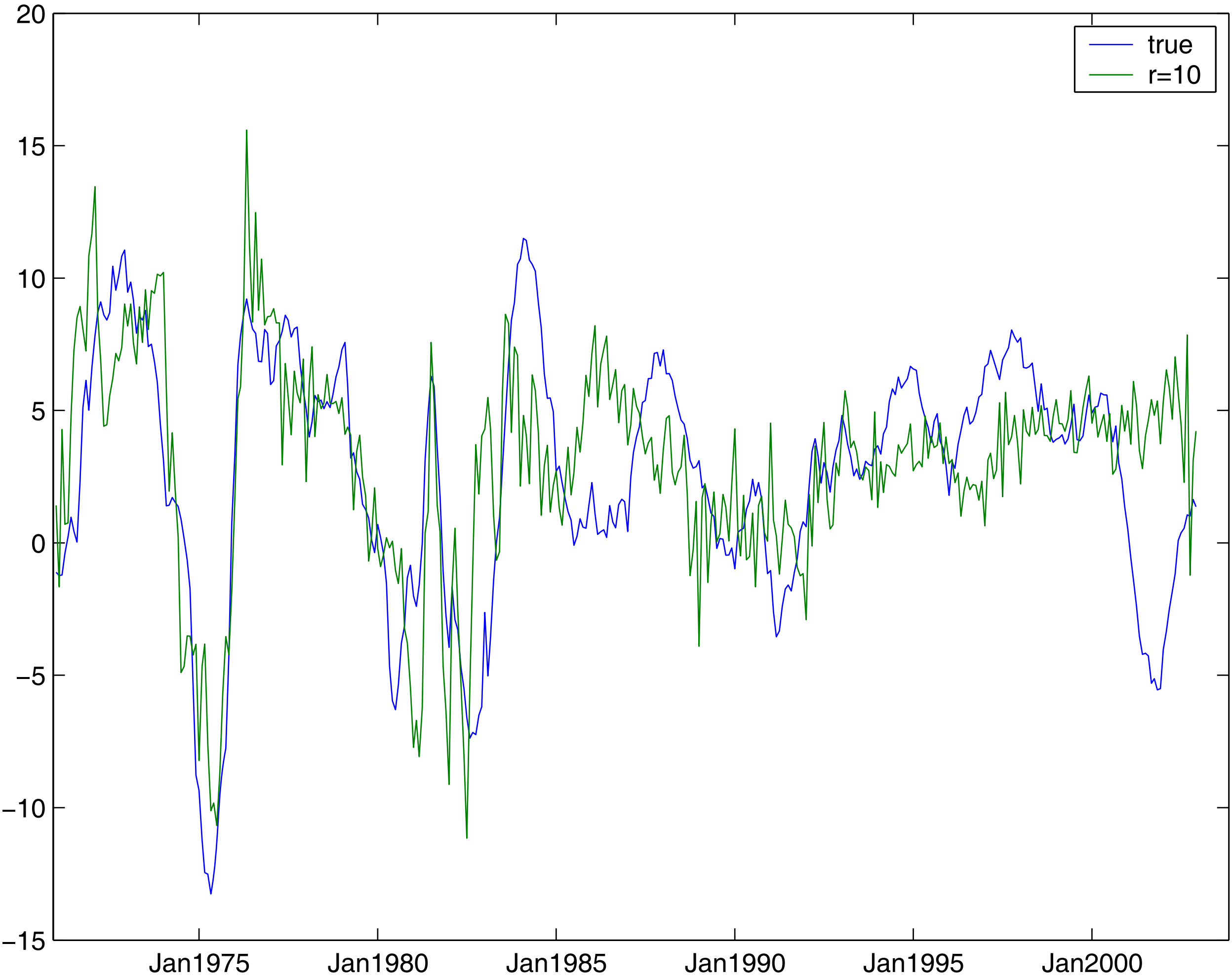

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true

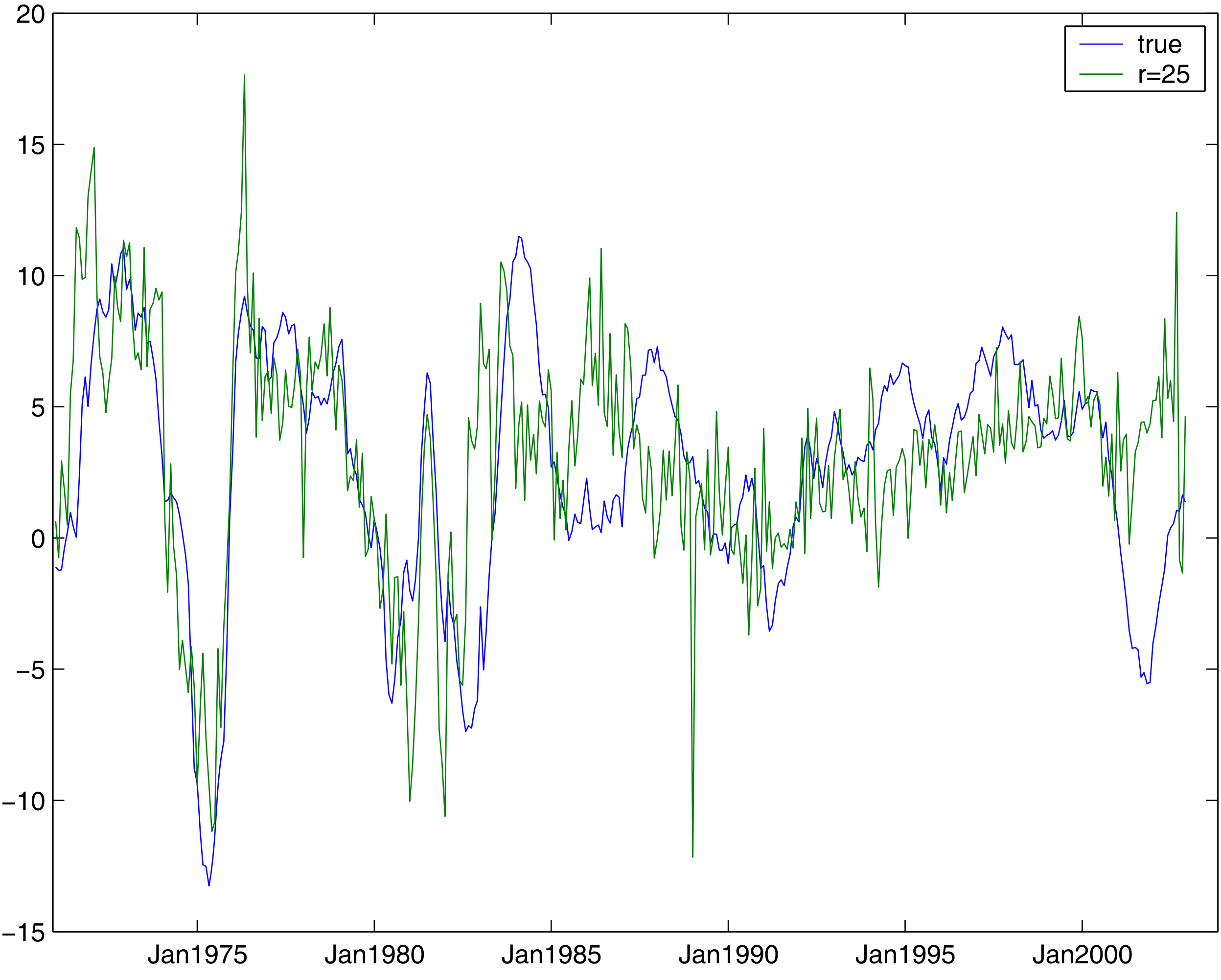

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=0

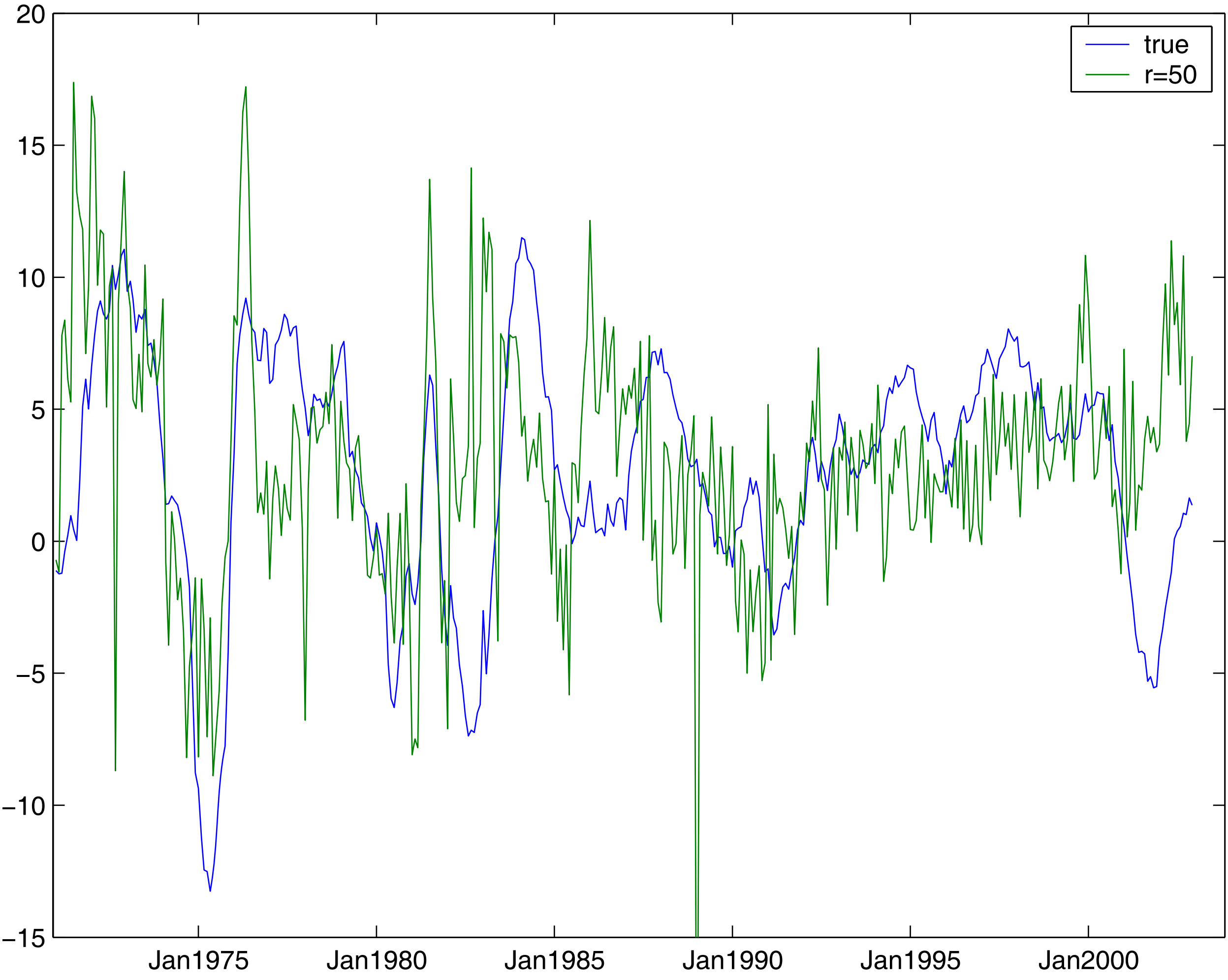

true
r=1
20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000

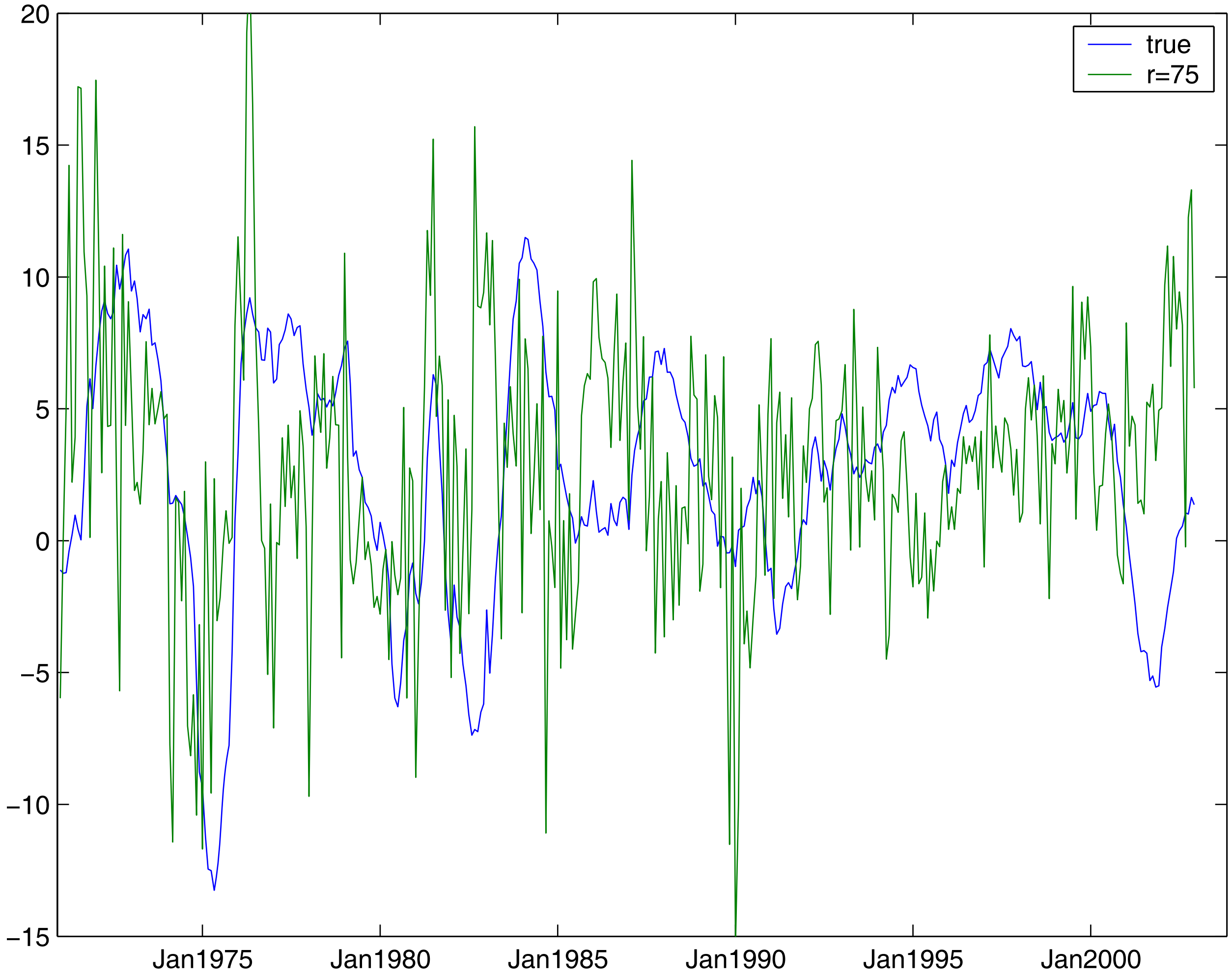

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=3

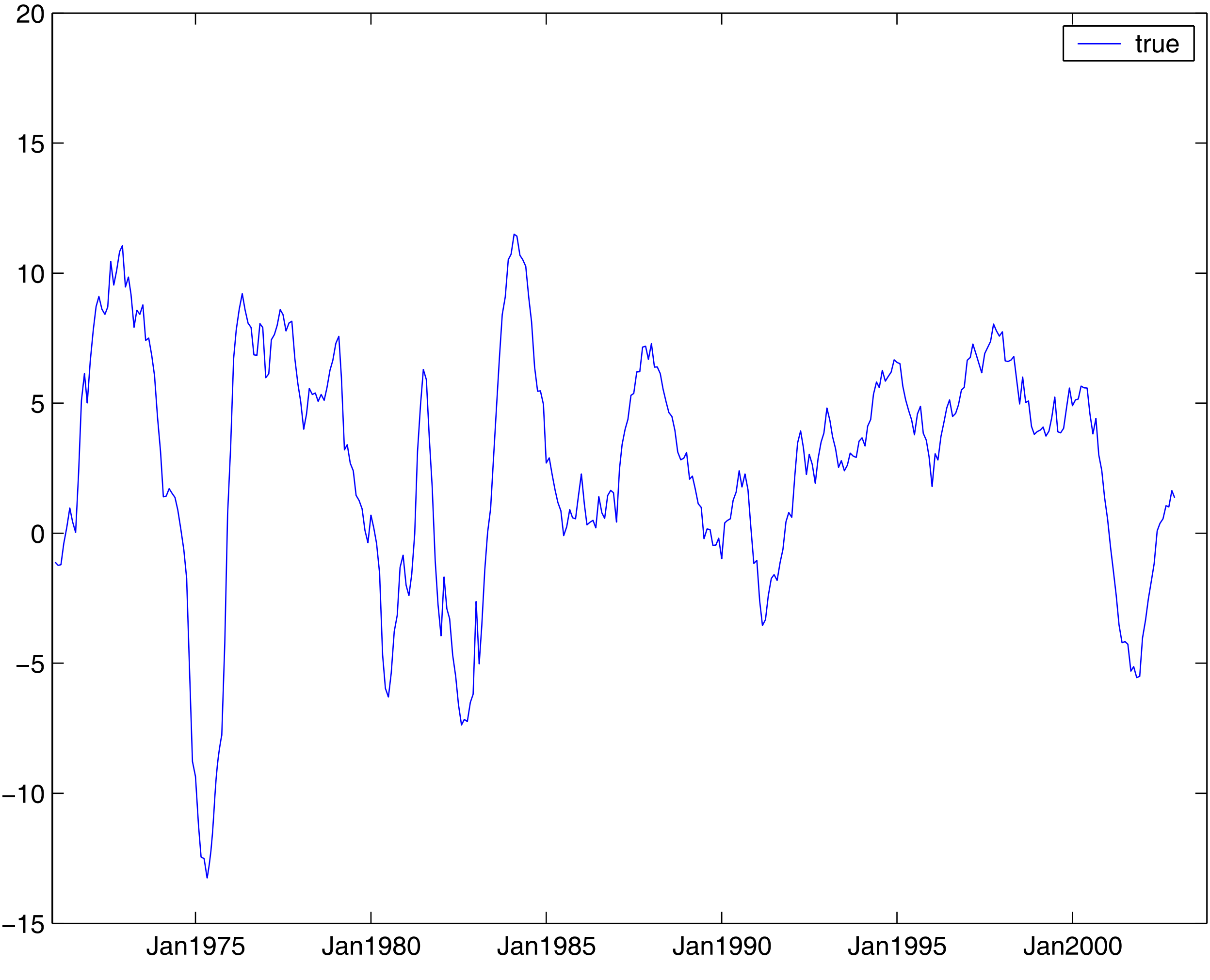

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=6

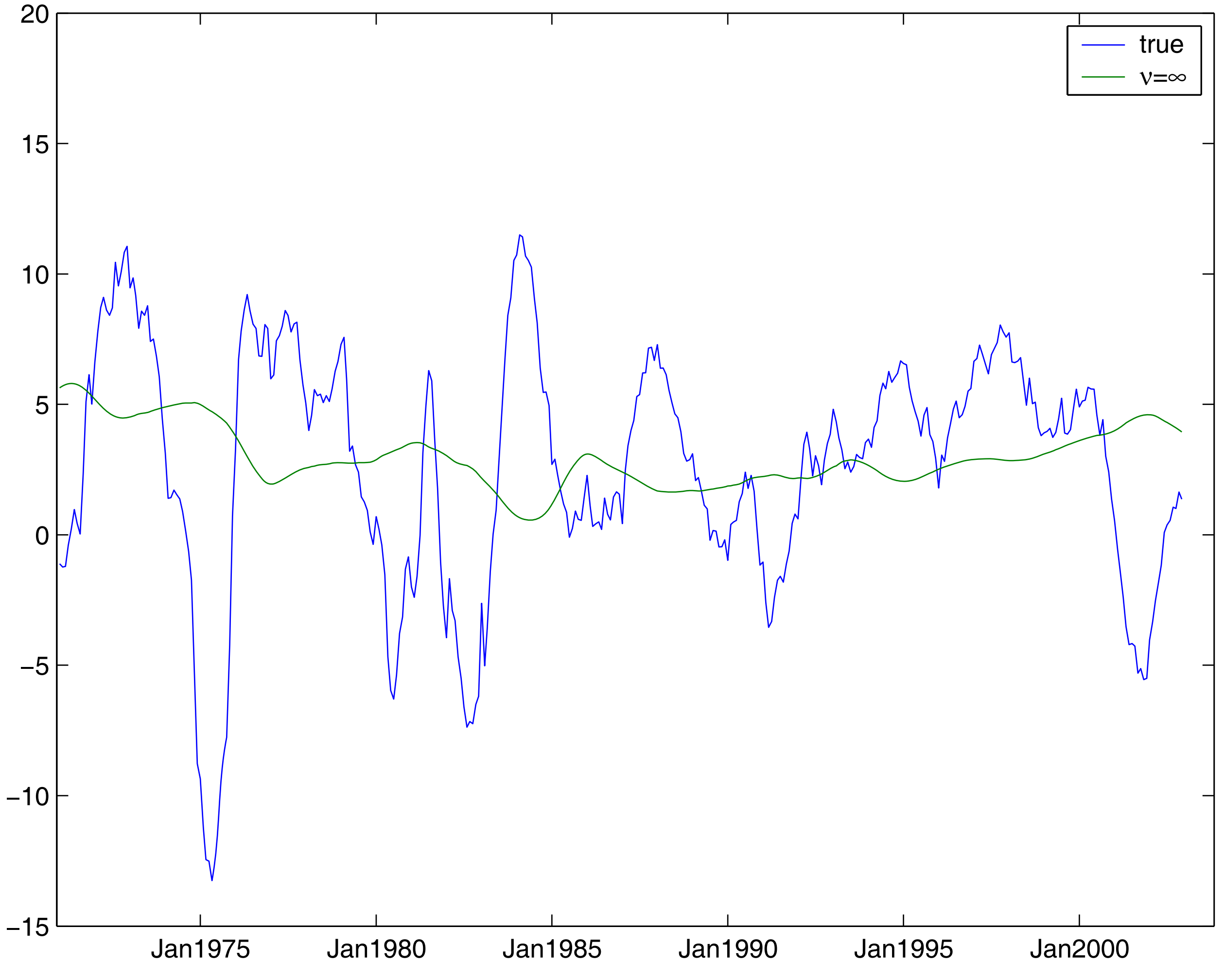

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=10


20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=25

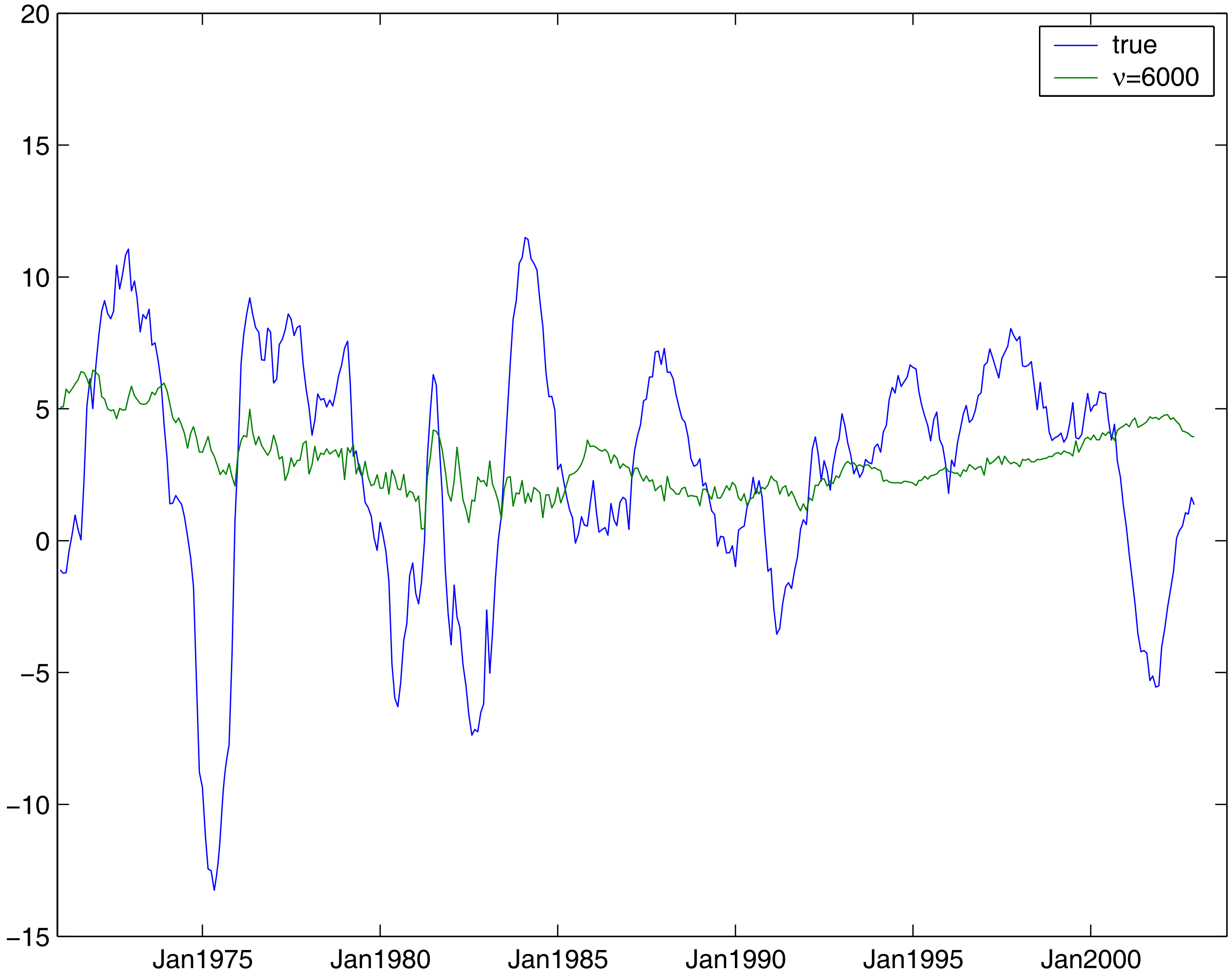

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=50

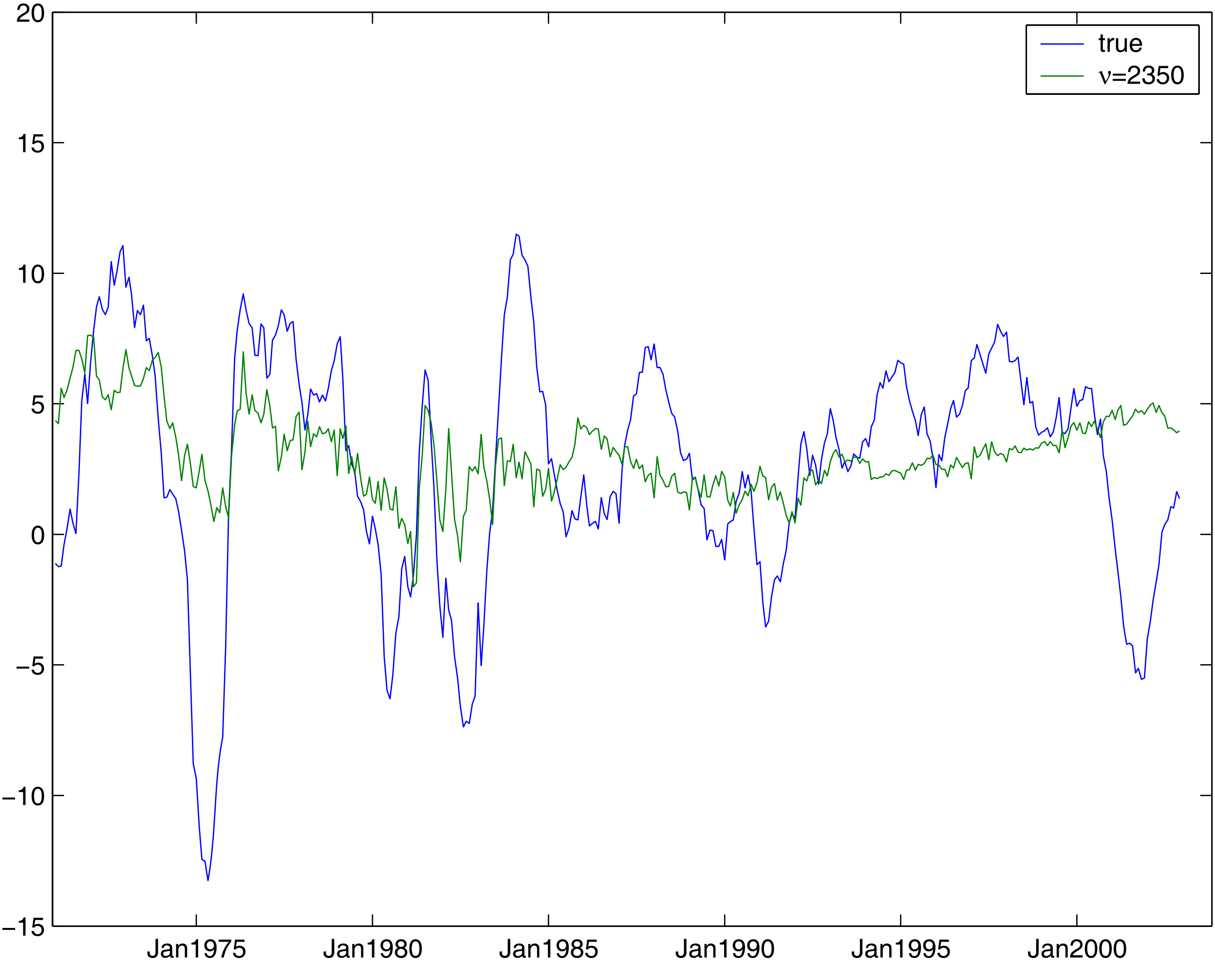

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
r=75

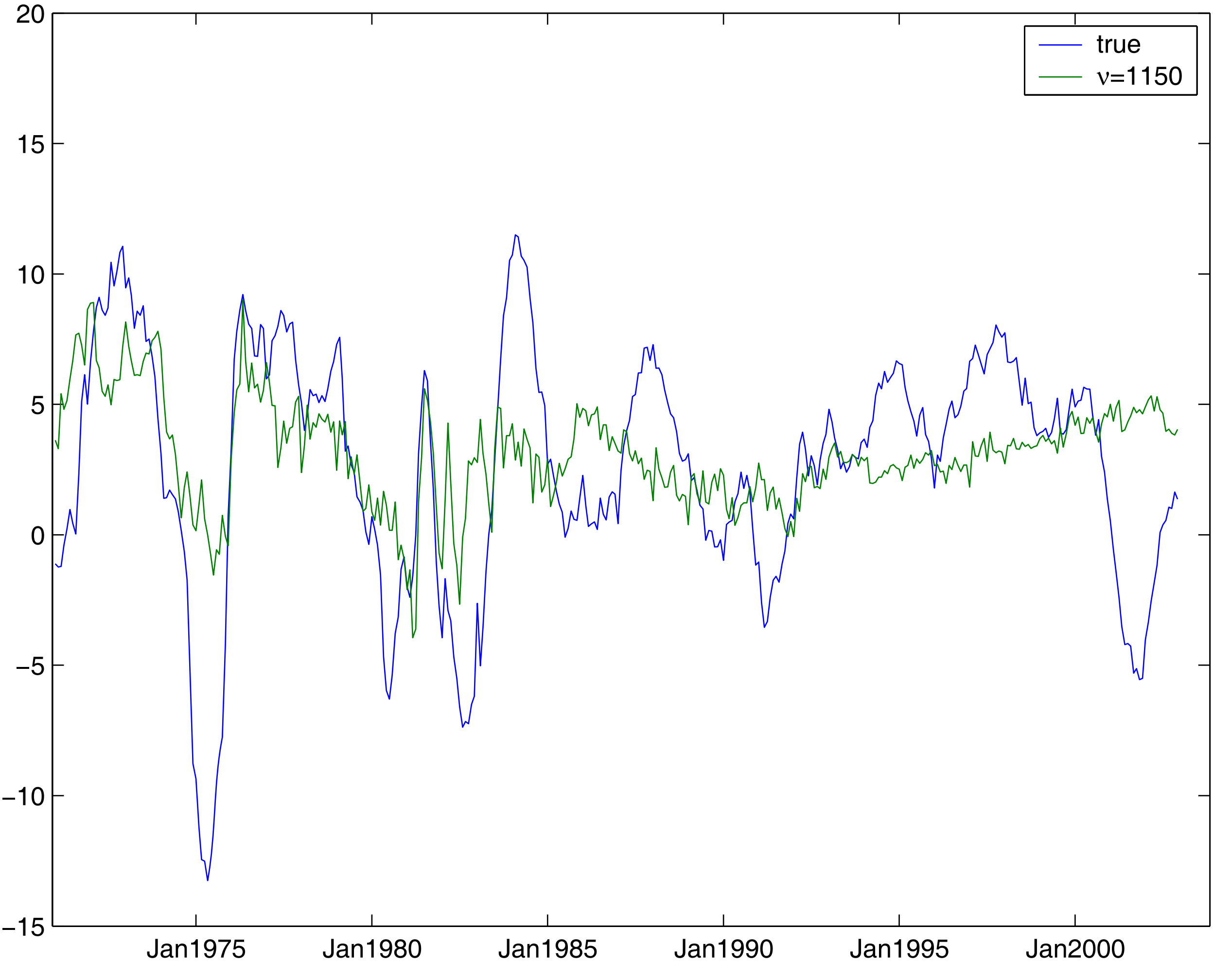

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true



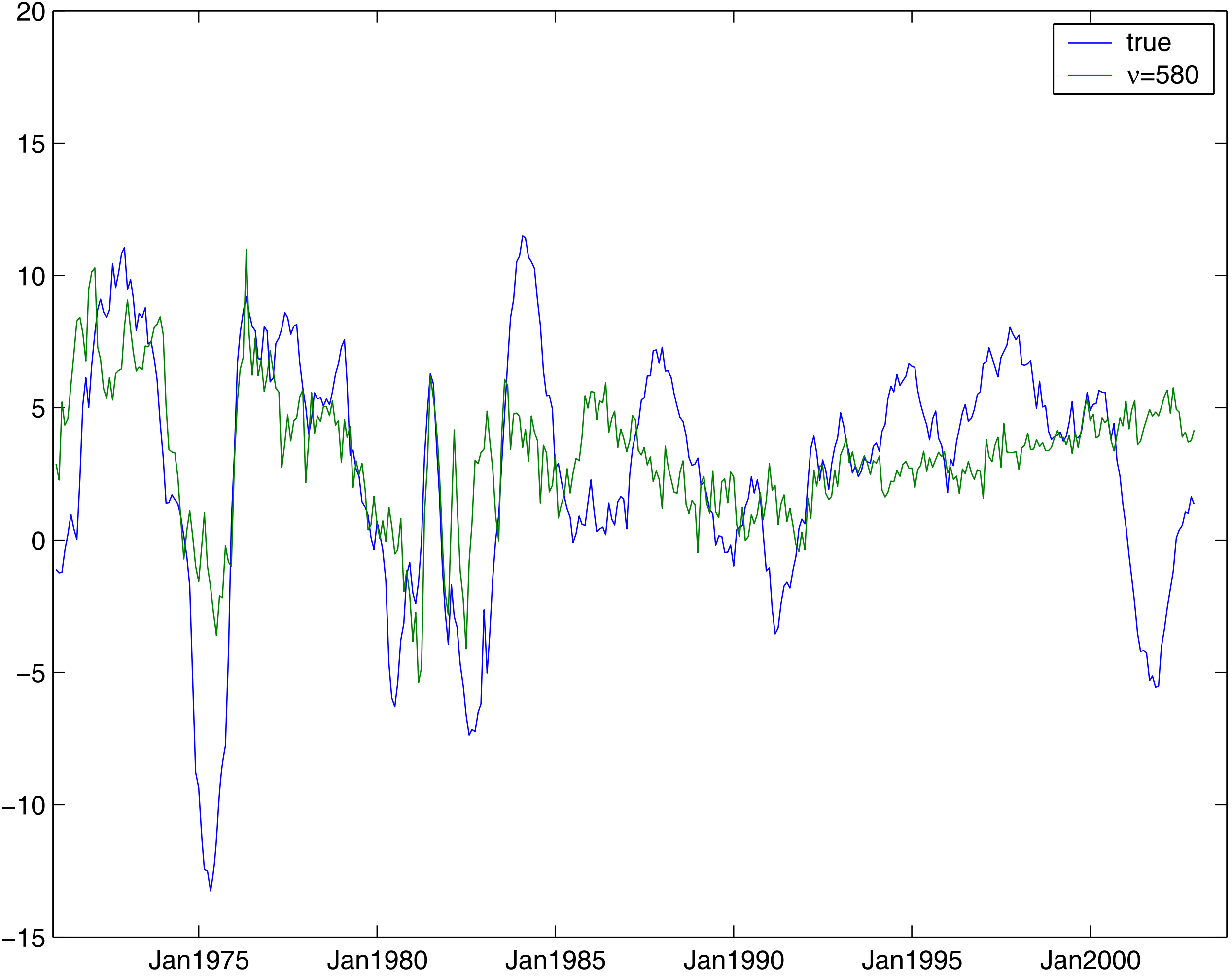
true
ν=∞
20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000

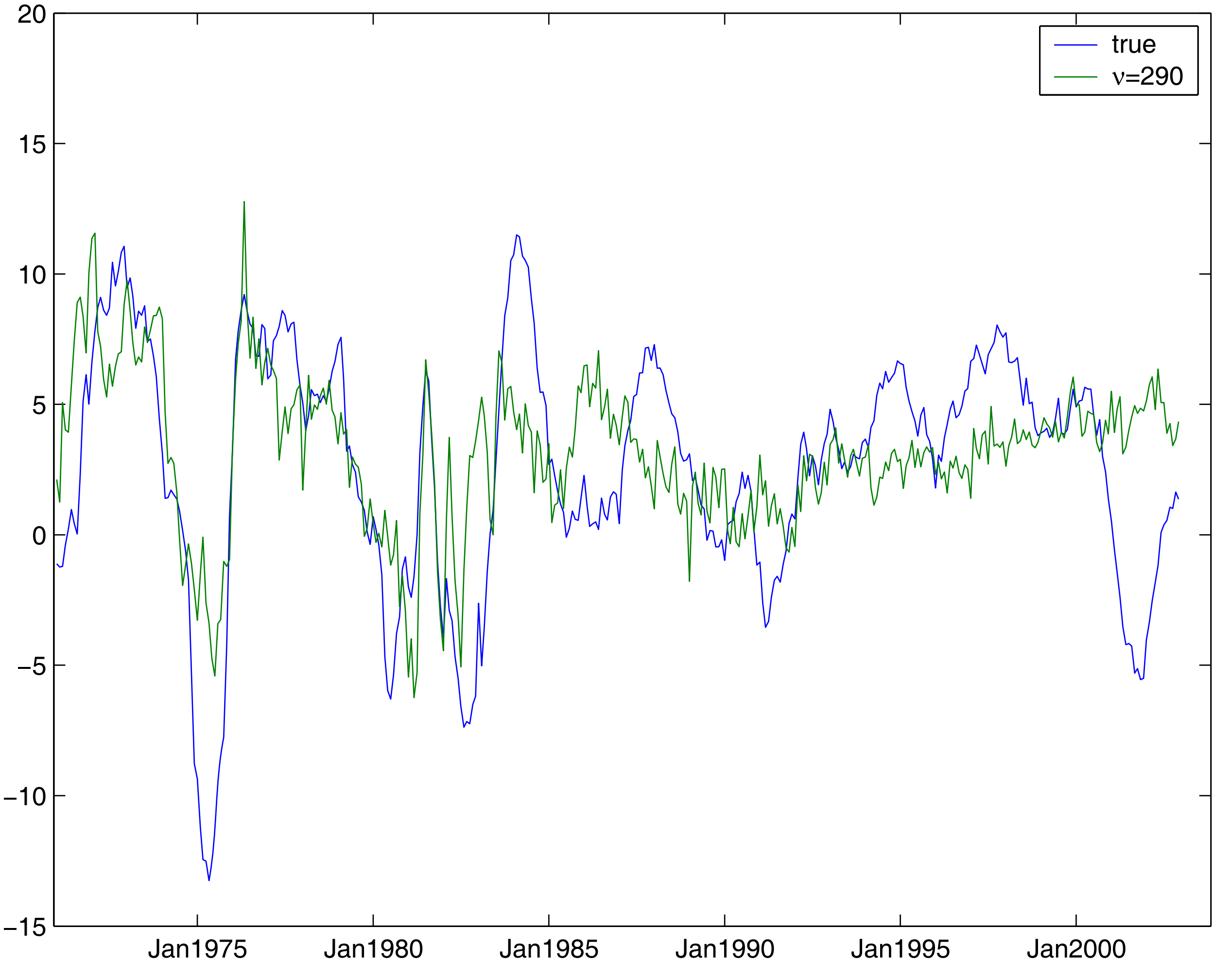

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=6000

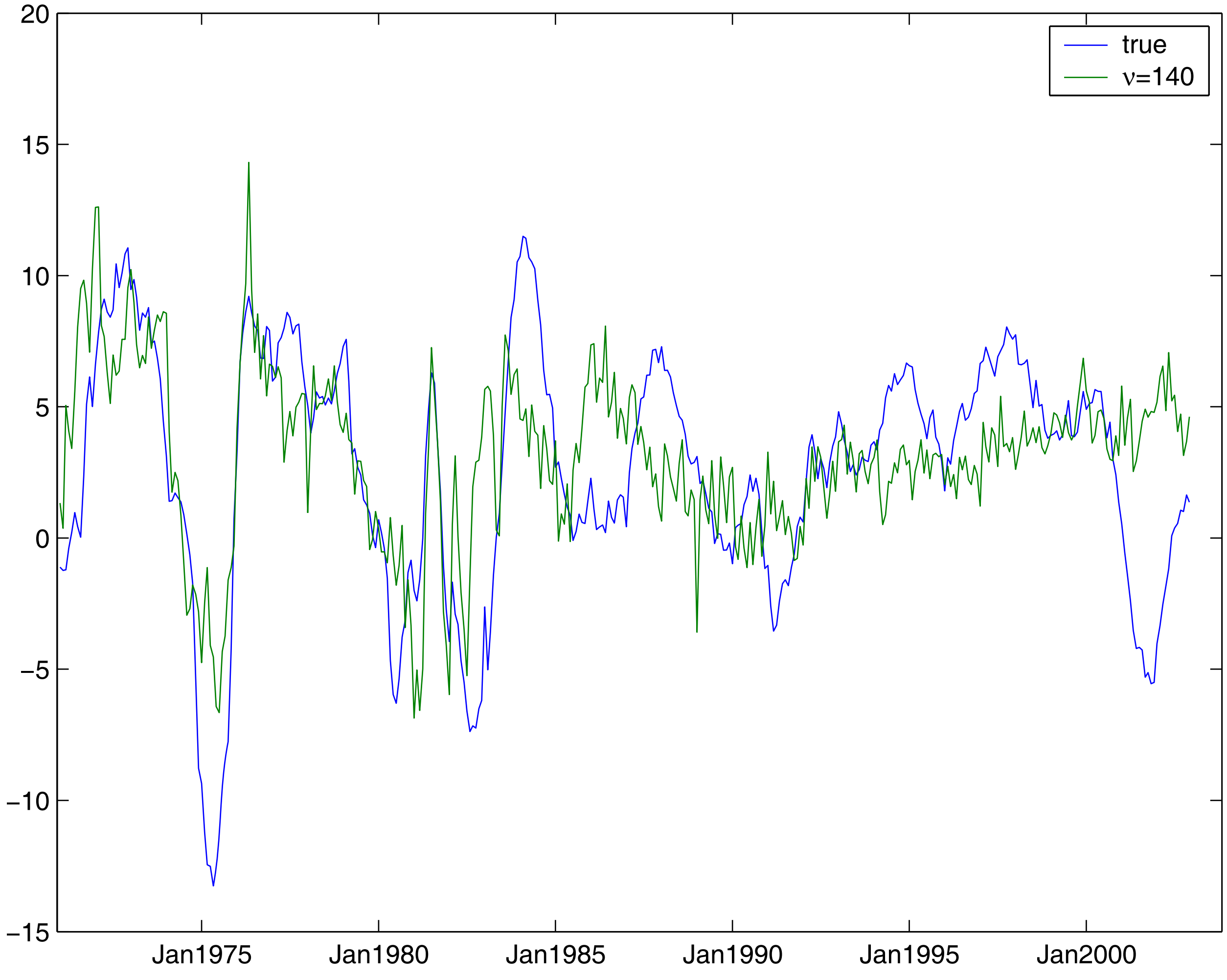

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=2350

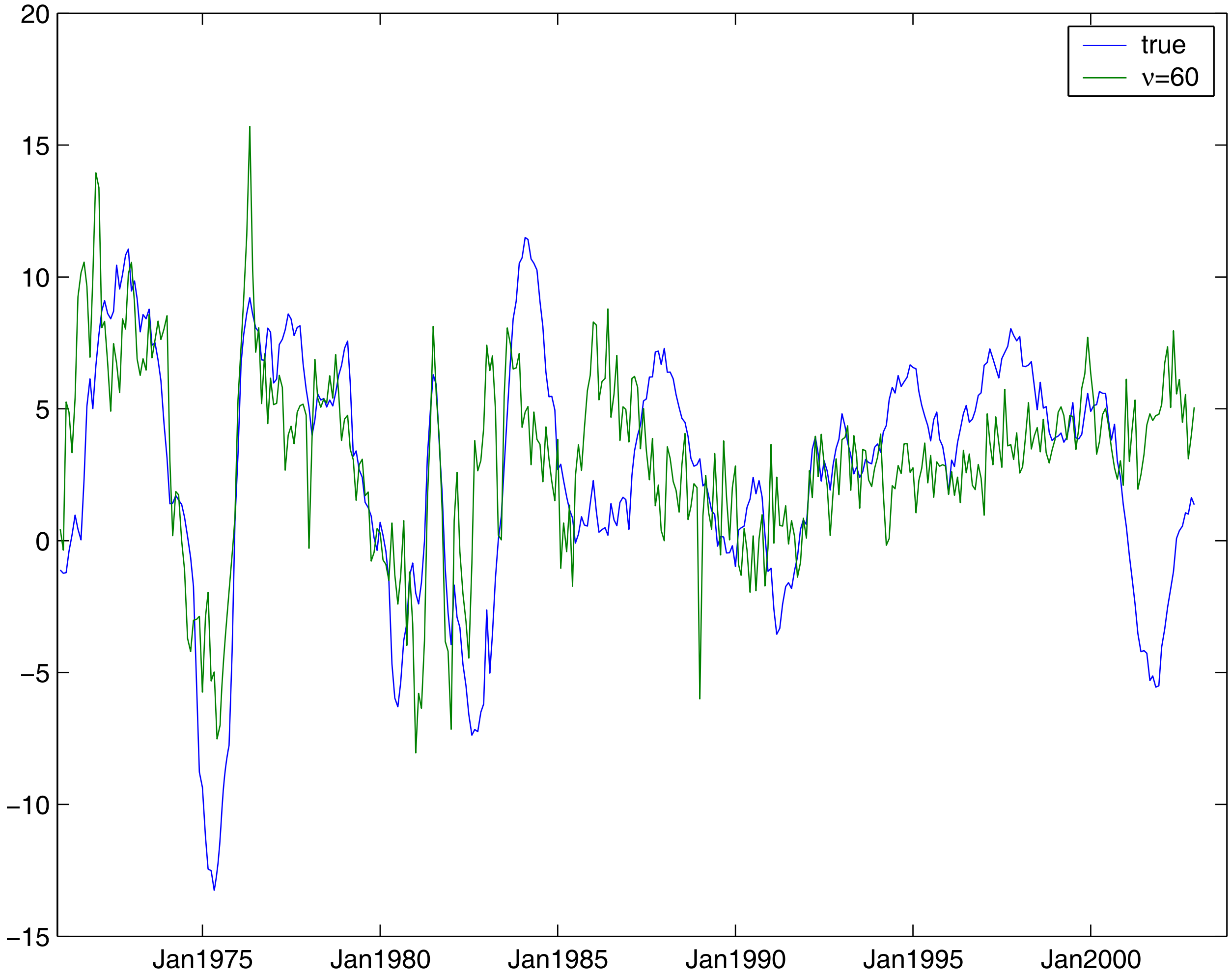


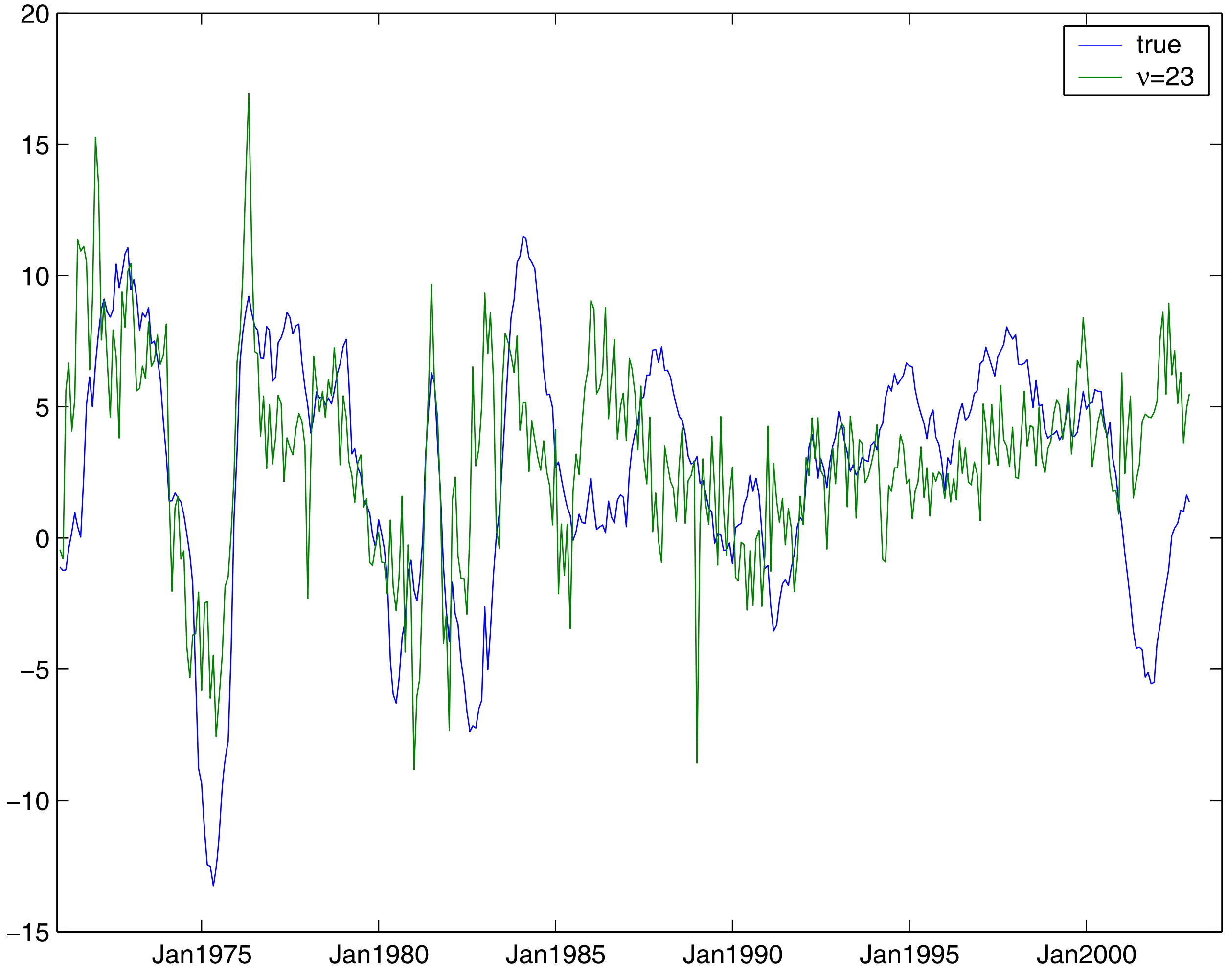
20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=1150

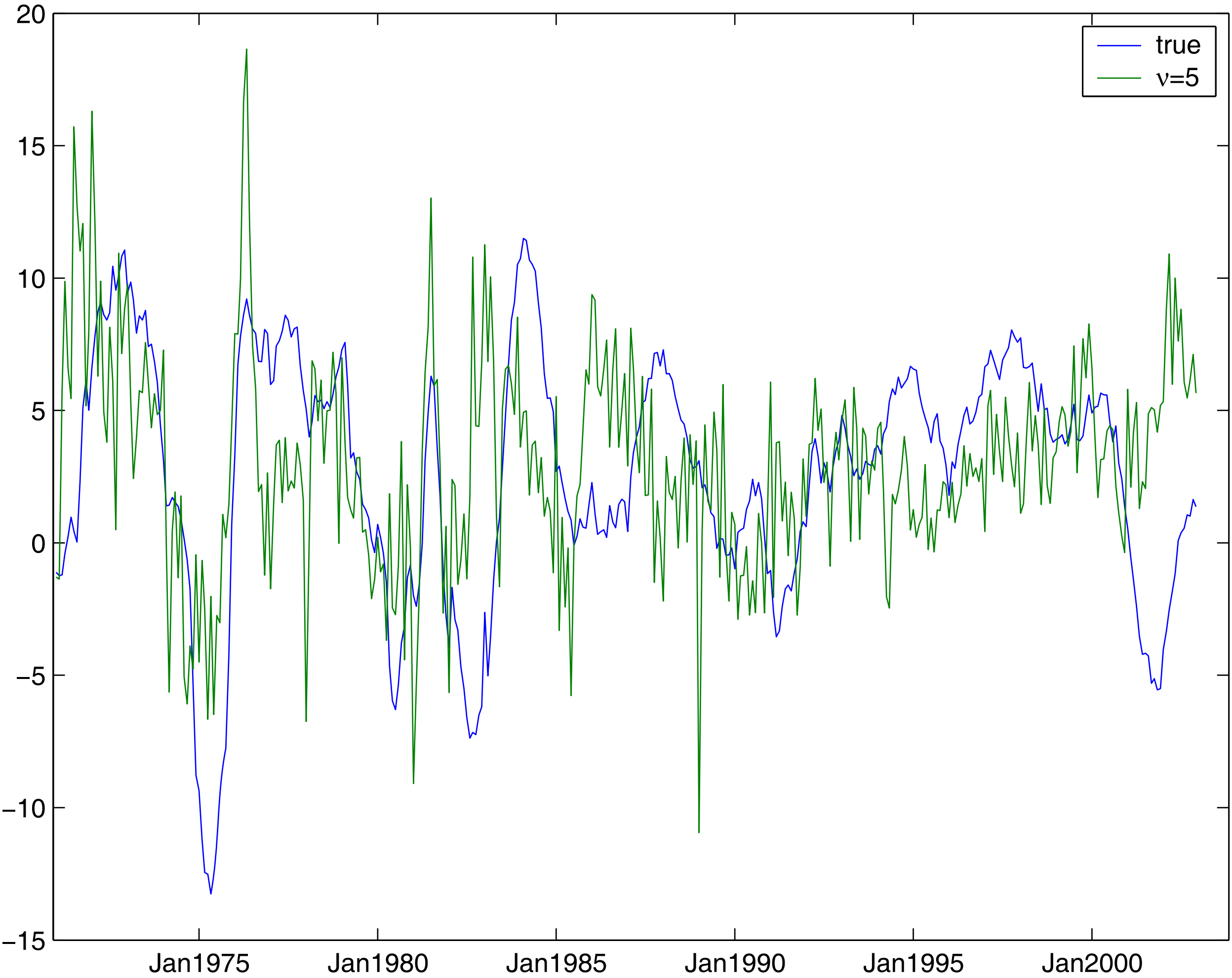

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=580

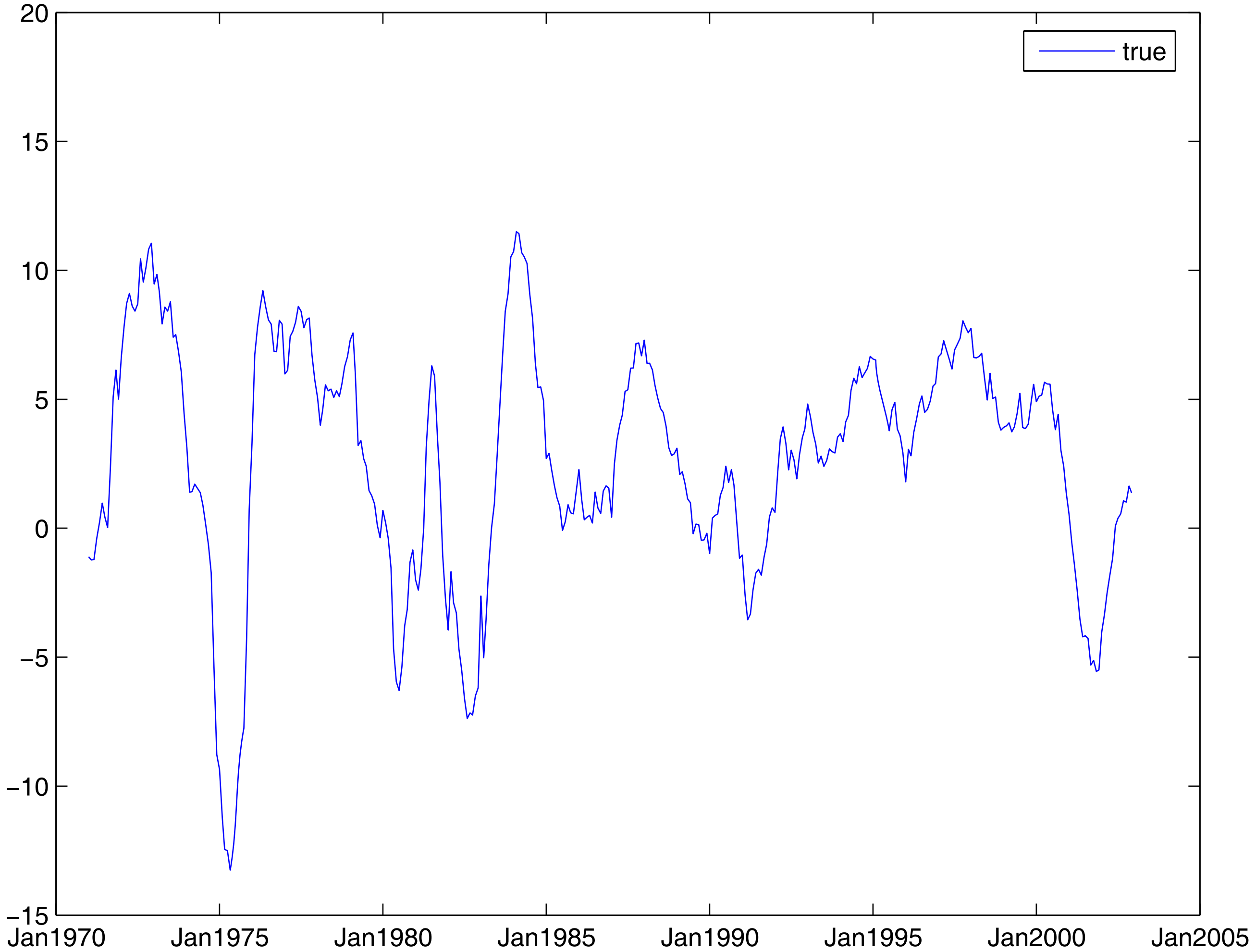

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=290

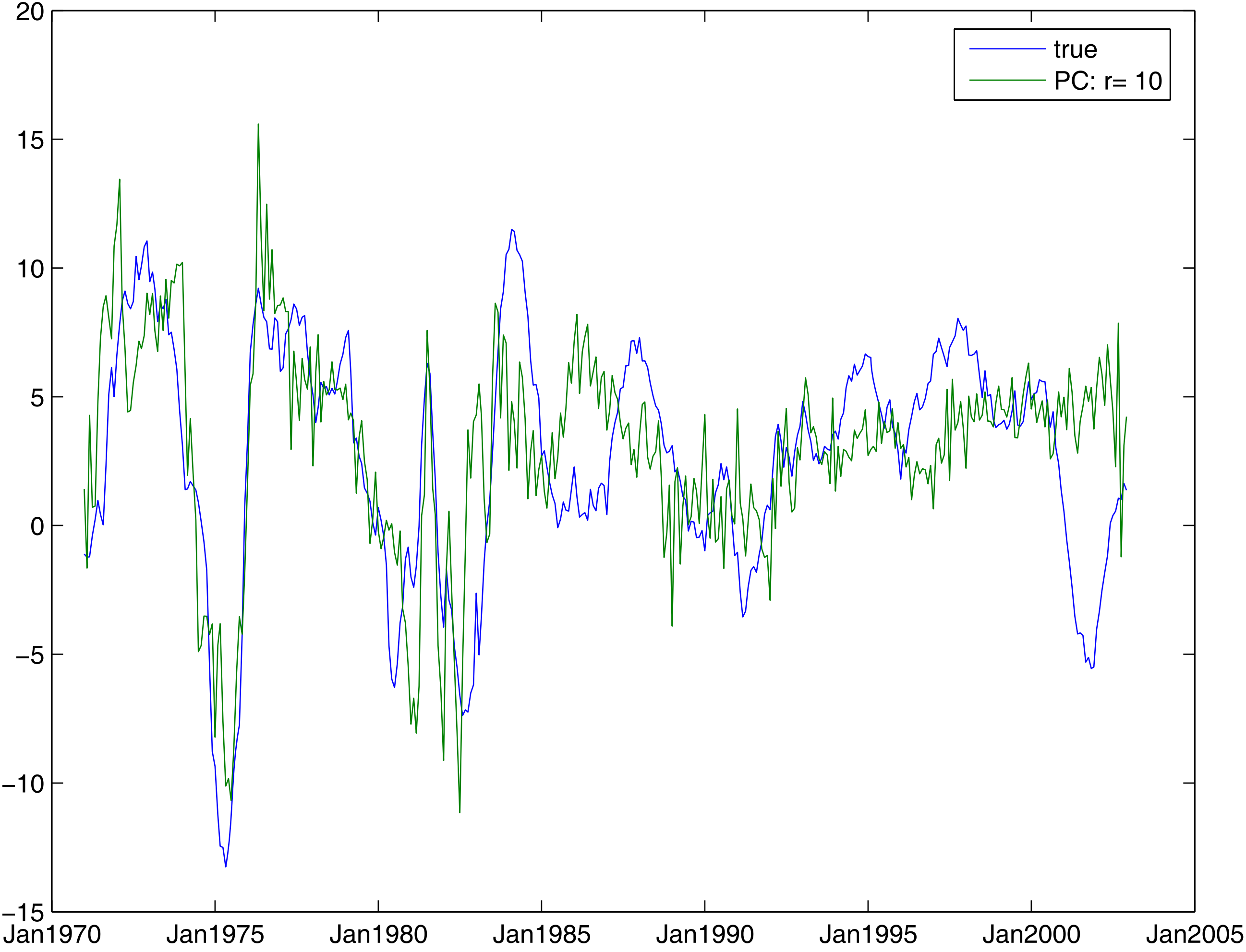

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=140

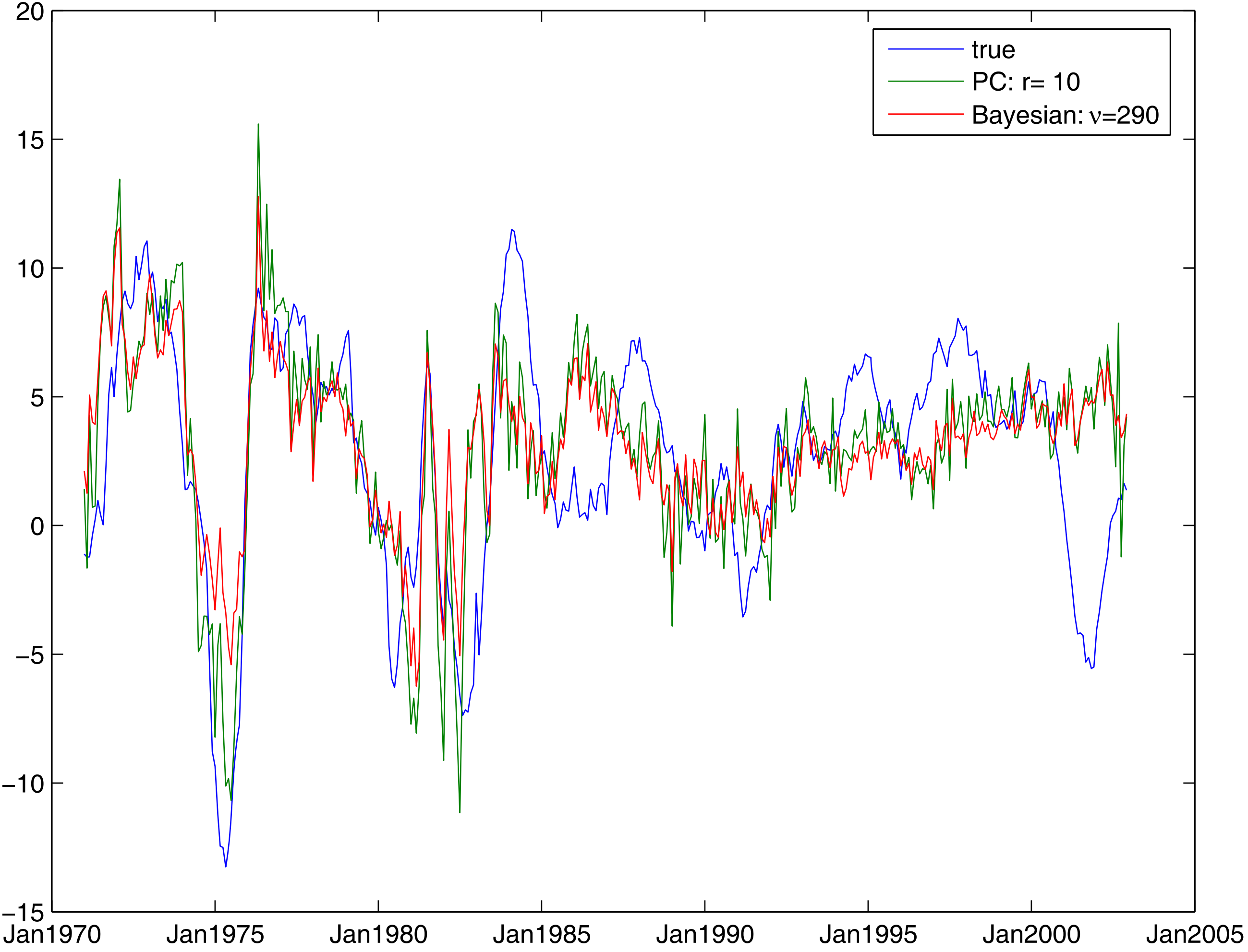

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=60

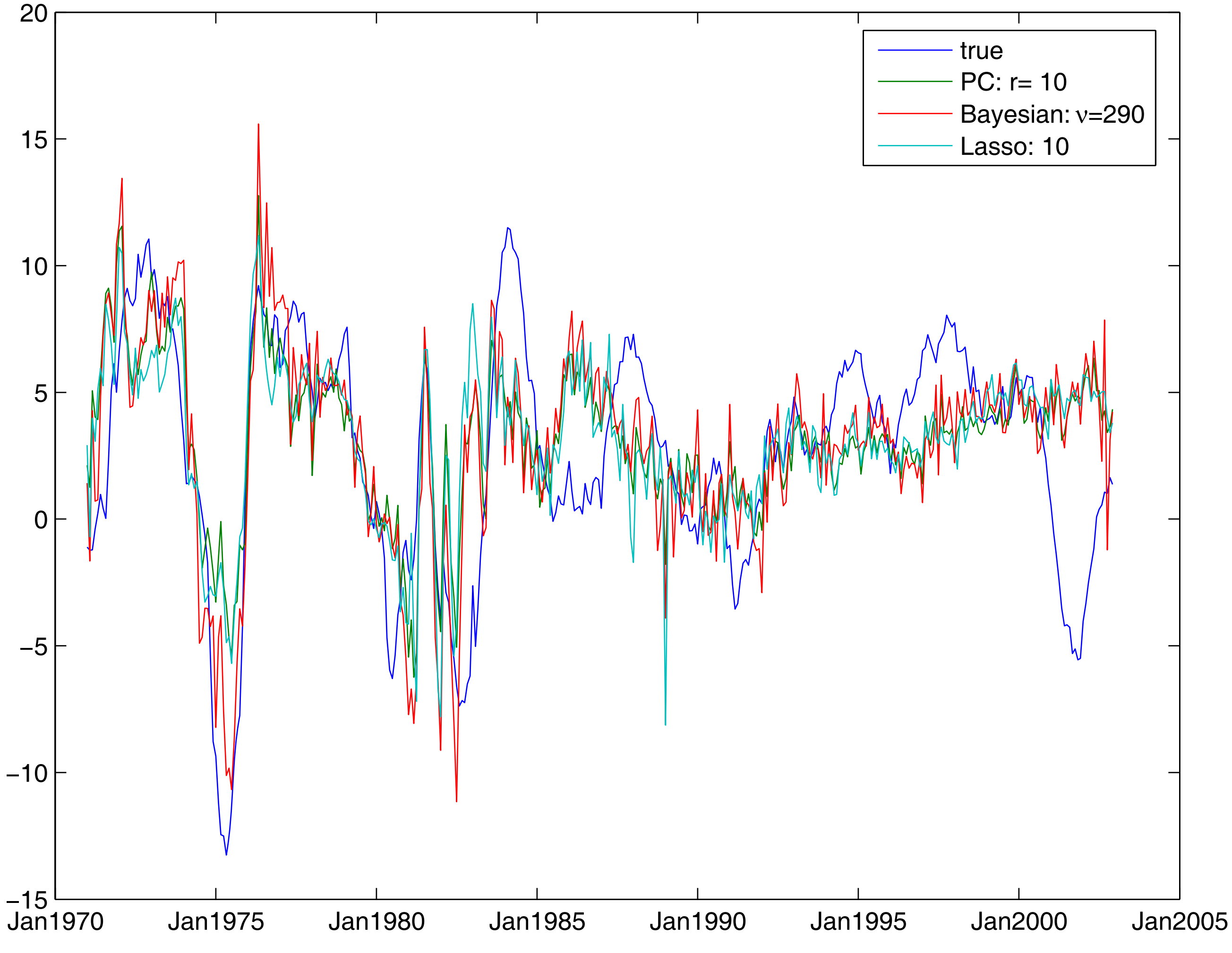

20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=23


20
15
10
5
0
−5
−10
−15
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
true
ν=5


20
15
10
5
0
−5
−10
−15
Jan1970
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
Jan2005
true


true
PC: r= 10
20
15
10
5
0
−5
−10
−15
Jan1970
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
Jan2005


20
15
10
5
0
−5
−10
−15
Jan1970
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
Jan2005
true
PC: r= 10
Bayesian: ν=290


true
PC: r= 10
Bayesian: ν=290
Lasso: 10
20
15
10
5
0
−5
−10
−15
Jan1970
Jan1975
Jan1980
Jan1985
Jan1990
Jan1995
Jan2000
Jan2005

Growing evidence on the power of shirnkage

- Large Bayesian VARs
  - Banbura, Giannone and Reichlin (2010), Bickel and Song (2011), Carriero, Clark and Marcellino (2012), Carriero, Kapetanios and Marcellino (2010a) Christoffel, Coenen and Warne (2011), Giannone, Lenza and Primiceri (2011), Koop (2010); Koop and Korobolis (2010), Lenza, Pill and Reichlin (2010), Matheson (2010), Stock and Watson (2009),...
- Sparse and stable portfolio selection:
  - Brodie et al. (2009), Carrasco and Noumon (2012), De Miguel et al., (2009)...
- Optimal pooling of forecasts:
  - Conflitti, De Mol and Giannone (2012)
- Combining shrinkage and reduced rank regression
  - Carriero, Kapetanios and Marcellino, 2011 and 2012

Growing evidence on the power of shirnkage

- Large Bayesian VARs
  - Banbura, Giannone and Reichlin (2010), Bickel and Song (2011), Carriero, Clark and Marcellino (2012), Carriero, Kapetanios and Marcellino (2010a) Christoffel, Coenen and Warne (2011), Giannone, Lenza and Primiceri (2011), Koop (2010); Koop and Korobolis (2010), Lenza, Pill and Reichlin (2010), Matheson (2010), Stock and Watson (2009),...
- Sparse and stable portfolio selection:
  - Brodie et al. (2009), Carrasco and Noumon (2012), De Miguel et al., (2009)...
- Optimal pooling of forecasts:
  - Conflitti, De Mol and Giannone (2012)
- Combining shrinkage and reduced rank regression
  - Carriero, Kapetanios and Marcellino, 2011 and 2012

Growing evidence on the power of shirnkage

- Large Bayesian VARs
  - Banbura, Giannone and Reichlin (2010), Bickel and Song (2011), Carriero, Clark and Marcellino (2012), Carriero, Kapetanios and Marcellino (2010a) Christoffel, Coenen and Warne (2011), Giannone, Lenza and Primiceri (2011), Koop (2010); Koop and Korobolis (2010), Lenza, Pill and Reichlin (2010), Matheson (2010), Stock and Watson (2009),...
- Sparse and stable portfolio selection:
  - Brodie et al. (2009), Carrasco and Noumon (2012), De Miguel et al., (2009)...
- Optimal pooling of forecasts:
  - Conflitti, De Mol and Giannone (2012)
- Combining shrinkage and reduced rank regression
  - Carriero, Kapetanios and Marcellino, 2011 and 2012

Should we reduce the rank?

This is an empirical question depending on how strong is the factor structure

This paper

- Not really if the focus is on macroeconomic forecasting
  See also: Carriero, Kapetanios and Marcellino (JAE 2010)
- Yes we should when forecasting bond returns
  - The yield data are indeed very well characterized by a factor structure.

Table II. Relative WTMSFE vs. AR(p^*) benchmark

	RR	SW	BVAR	MB	RRP	BRR
Hor: 1						
Rel. WTMSFE	1.36	1.70	1.18	2.08	**1.15**	1.22
IPS10	1.10***	1.09	**0.90**	0.99	0.97	**0.90**
PUNEW	1.31	**1.08**	1.11*	**1.08**	1.12*	1.24***
FYFF	1.09*	1.01	0.94	1.02	**0.93**	0.99
Hor: 2						
Rel. WTMSFE	1.17	1.35	1.06	1.67	1.06	**1.05**
IPS10	1.14	1.05	0.86	1.05	0.90	**0.81**
PUNEW	1.08	1.02	**0.99**	1.10	**0.99**	1.05
FYFF	1.01	0.98	0.91	1.01	**0.86***	0.95
Hor: 3						
Rel. WTMSFE	1.07	1.17	0.99	1.40	0.98	**0.97**
IPS10	1.06	1.03	0.80	1.06	0.81	**0.78***
PUNEW	0.93	0.96	0.87*	1.06	**0.86***	0.91
FYFF	1.01	0.99	0.92	1.01	**0.89***	0.94
Hor: 6						
Rel. WTMSFE	0.94	1.11	0.88	1.09	**0.87**	0.88
IPS10	0.87	1.04	0.69**	1.01	**0.67****	0.71***
PUNEW	0.76***	0.95	**0.71*****	1.06	**0.71*****	0.74***
FYFF	1.00	1.12	0.89*	0.99	**0.83*****	0.91***
Hor: 9						
Rel. WTMSFE	0.91	1.13	0.85	1.01	**0.84**	0.87
IPS10	0.82	1.07	0.68***	1.02	**0.66*****	0.72***
PUNEW	0.76*	0.97	**0.67*****	1.11*	**0.67*****	0.69***
FYFF	0.97	1.14	0.91*	1.00	**0.79*****	0.90***
Hor: 12						
Rel. WTMSFE	0.90	1.20	0.85	0.97	**0.84**	0.87
IPS10	0.87	1.10	0.68***	1.01	**0.65*****	0.74***
PUNEW	0.72***	0.99	**0.64*****	1.06	0.65***	**0.64*****
FYFF	0.95	1.20	0.90***	1.00	**0.84*****	0.91***

Why do rank reduction helps forecasting bond returns?

The data are indeed very well characterized by a factor structure because of the way they have been constructed.

They have been constructed by using the Nelson and Siegel model to interpolate between the existing limited number of securities with different maturities and coupons (see Gurkaynak et al., 2006)

$$y_t^\tau = \lambda_L^\tau L_t + \lambda_S^\tau S_t + \lambda_C^\tau C_t + u_t^\tau$$

A fact or an artifact?
Check with appropriate data, for example constructed using unsmoothed Fama-Bliss method

A not on reduced rank regression

This paper is about minimizing:
$\|Y - X\beta\alpha'\|_2^2$ + penalty
$\Rightarrow$ Spectral decomposition of

$$(Y'X/T)(X'X/T + \nu I)^{-1}(X'Y/T)$$

This amount at computing the PC (SVD) of the fit $X\hat{B}$

Carriero, Kapetanios and Marcellino look at SVD of $\hat{B}$ only

Maximum likelihood implies minimizing
$\det[(Y - X\beta\alpha')'(Y - X\beta\alpha')]$
$\Rightarrow$ Spectral decomposition of

$$(Y'Y/T)^{-1}(Y'X/T)(X'X/T + \nu I)^{-1}(X'Y/T)$$

The two approaches coincide when the target variables are not correlated

The term $(Y'Y/T)^{-1}$ is necessary to exploit the collinearity features of Y that has motivated the paper.

These incoherencies can be avoided by resorting to a properly defined bayesian framework (Geweke, 1996)

Additional comments

- Setting the degree of shrinkage and the rank of the regression: very restrictive assumptions and not completely developed implications. You might exploit the more general results derived by Onatsky, 2010.
- Group membership and other linear restrictions can also be imposed by QML estimation of factor models (Doz et al., 2011)