Estimating DSGE Models with Stochastic Singularity using Likelihood-based Dimensionality Reduction☆

Michal Andrle[1]

International Monetary Fund, Research Dept., 700 19th St. N.W., Washington, D.C., 20431, USA

Abstract

This paper demonstrates a method of estimating and testing stochastically singular state-space models using likelihood-based methods. The approach uses dimensionality reduction of the observed data and rotates the model into directions of maximum variance, using the (dynamic) principal component analysis. The model is estimated in the space determined by a range (input space) of principal componetns. The principal components are then used as observables for transformed model, without any change of its economic structure. The method endogenously determines the structure of measurement errors. The approach can be interpreted as implementing identical principal-component filter to both data and the model before estimation.

Keywords: stochastic singularity, maximum likelihood, dynamic principal components, state space, term structure
JEL Codes: C10, E50

1. Introduction

In paper I propose a method for estimating a stochastically singular model using information on more time series than stochastic shocks in the model. Instead of specifying pseudo-structural shocks or ad-hoc measurement noise processes to link the model to data, I map the model onto a subspace of maximum variance of the data available, using a (dynamic) principal component analysis – (D)PCA. This approach makes use of the dimensionality reduction and the fact that many data sets in economics and finance are readily by large explained by only few static and even less dynamic factors.

A well known (and a bit of extreme) example of a stochastic singular model is a standard real business cycle (RBC) model with only one source of stochastic disturbance – a technology shock. Variability of all variables is driven by one shock. In finance, affine term structure models are stochastically singular, having few factors driving wide set of maturities. This affects seriously a maximum likelihood approach to estimation, as one can use only one observable series for estimation, unless more stochastic shocks are added into the system. The spectral density of model observables is singular, which is not always the case in the data.

I argue bellow that allowing for 'an error' in the estimation guided by principal components is more convenient than specifying pseudo-structural shocks or specifying measurement errors.

In order to bring on board information from other time series, I suggest to calculate static or dynamic principal components (factors) of observed data and associated principal angles – directions of maximum variance in the data. After transforming the observable equation of the model, the input to estimation is the number of principal components coincident with the number of stochastic shocks of the model. The use of dynamic principal component provides sharper dissection of lead-lag structure in data, yet commands use of spectral domain methods.

In most cases, the dimensionality reduction amounts to loss of 'information' and it depends on the application at hand how much information and at what frequencies is lost, yet a signal-to-noise ratio of the data can increase. The key to understand the method is to understand that it amounts to transforming (filtering) the data and searching in the input space of the transform for a proper parameterization.

I focus on time series analysis, where the source of redundancy is not just contemporaneous correlation, but serial cross-correlation and thus the idea of dynamic principal component analysis (PCA across frequencies) arises quite naturaly. The approach is not identical, but related to theory of dynamic factor models.

For instance, even in case of one structural shocks and only first static principal component used the model is makes use of information on all series and their relative variances. Such combination of parameters is sought for that produces dynamics which, when mapped onto a subspace of first principal component, generates the result as much close to the data as possible.

First, I review the problem associated with stochastic singularity and most commonly used solutions. Then I propose a method how to implement dimensionality-reduction parameter and unobserved component estimation in time and frequency domain and explain its properties. Computational experiment

☆The views expressed herein are those of the author and should not be attributed to the International Monetary Fund, its Executive Board, or its management.

[1]I would like to thank Wouter den Haan, Jan Brůha, Mika Kortelainen and Anders Warne for discussions. First version: Sept 23, 2011

on parameter estimation and identification follows, before I conclude.

2. Stochastic Singularity and Likelihood-based Estimation

2.1. The Model

I assume a model can be cast in terms of linear, time invariant state-space form

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{X}_t \quad (1)$$

$$\mathbf{X}_t = \mathbf{T}\mathbf{X}_{t-1} + \mathbf{R}\varepsilon_t, \quad (2)$$

where $\mathbf{X}_t$ is a $(n_x \times 1)$ vector of transition variables, $\mathbf{Y}_t$ is a $(n_y \times 1)$ vector of measurement variables (observables) and ε_t is $(n_e \times 1)$ vector of stochastic shocks, assmed to follow Gaussian distribution, i.e. $\varepsilon \sim N(0, \mathbf{\Sigma}_\varepsilon)$. With no loss of generality, I assume the stochastic dimension of the model is n_e. The form of the model does not specify if ε_t are structural shocks or measurement errors.

For further reference, we define a spectral density of Y_t to be

$$\mathbf{S}_{yy}(z) = (1/2\pi) \times \mathbf{S}(z)\mathbf{\Sigma}_\varepsilon \mathbf{S}(z)^H \qquad \mathbf{S}(z) = \mathbf{Z}(\mathbf{I} - \mathbf{T}z)^{-1}\mathbf{R} \quad (3)$$

with z being a complex variable $z = e^{-i\lambda}$, where $\lambda \in [-\pi, \pi]$ is the angular frequency.[2]

The process $\mathbf{Y}(t) := \{\mathbf{Y}_t\}$ can be rewritten in a form of a moving average representations $\mathbf{Y}_t = \mathbf{S}(L)\varepsilon_t$ with $\mathbf{S}(z) = \sum_{i=0}^{\infty} \mathbf{S}_j z^j$.

Due to equivalence of time-domain and frequency-domain (spectral) analysis of time series, the k-th order autocovariance of $\mathbf{Y}(t)$ process is given by

$$\mathbf{\Gamma}_y^k = \int_{-\pi}^{\pi} \mathbf{S}_{yy}(z) e^{-i\lambda k} d\lambda. \quad (4)$$

2.2. Stochastic Singularity

Stochastic singularity is best understood by investigating $\mathbf{S}_{yy}(z)$, with diagonal square matrix $\mathbf{\Sigma}_\varepsilon$ and $(n_y \times n_e)$ matrix polynomial $\mathbf{S}(z)$. Stochastic singularity implies that the spectral density matrix $\mathbf{S}_{yy}(z)$ is rank-defficient at almost all frequencies.

For maximum likelihood estimation (MLE) or a Bayesian likelihood-based estimation one requires that $n_e \geq n_y$, i.e. one cannot use more observables than is the stochastic dimension of the model. When $n_e < n_y$, given any state $\mathbf{X}_{t-1}$, it is impossible to find ε_t that would be able to explain data, $\mathbf{Y}_t$.

Not only likelihood-based estimation is affected by stochastic singularity, but any other estimator. In case *methods of moments* estimation one can employ more time series than stochastic rank, but there is only a limited set of moments that are linearly independent – the number of estimable parameters is limited to number of moment conditions.

[2]Given a matrix A, I denote A^H to be its conjugate-transpose.

Common solutions of the singularity problem. The issue is not knew and it is well known. Sargent (1989) solved the problem by augmenting the model with 'measurement errors', which can be interpreted as noise in releases of the statistical office, for instance. Altug (1989) estimates a real business cycle (RBC) model driven by one shock by augmenting variables with uncorrelated measurement errors and explicitely acknowledges models implicit factor structure. In case of uncorrelated, zero-mean measurement errors all cross-correlations between the data are in principle unaffected, but their respective variances are not.

Ireland (2004) follows Sargent (1989) but makes the measurement errors serially correlated, constructing thus a hybrid model where ad-hoc dynamics could potentially explain a lot of data dynamics and seriously affect estimation of 'structural' model's parameters and render the underlying economic model uninteresting.

Ingram, Kocherlakota, and Savin (1994) solve the stochastic singularity by adding structural shocks with economic interpretation. This approach may, however, lead to proliferation of pseudo-structural shocks in order to make use of multiple time series for the analysis. Further some shocks then seriously affect the explanatory power of the model, e.g. various cost-push shocks used in presence of high-frequency variation in the price data are mostly used to explain the erratic high frequency movement, but have often non-trivial medium-frequency response in other real economic variables, that need to be counteracted with other shocks to 'fit' the data. Fitting high frequencies with misspecified shocks can also lead to sever bias of inflation persistence estimates. Essentially one can obtain a purely conditioned state estimation problem.

In addition, evidence from principal component analysis of data in many fields, including economics and finance suggest that only relatively small number of principal components is needed to explain most of the data variance. A chance of needing large variety of shocks is thus small.

Measurment errors approach is chosen often also in finance, see e.g. Piazessi (2010, Sect. 6.1.) for a discussion of stochastic singularity in relation to affine term structure models and its problematic nature.

2.2.1. Problem of using measurement errors as a solution

Are there any problems of estimating structural parameters using measurement erors? Yes, there are – both in terms of economics and estimation properties.

In *dynamic macroeconomics* models the decision between *structural shocks* and *measurement errors* is absolutely crucial from theory point of view. An improper choice of measurement error can have vast consequences, e.g. adding a measurement error to growth rate of a variable can lead to divergence in levels of the observed data and the model concept of a variable. Another example is the measurement error on exports and imports, creating a distorted view of trade balance, etc. In case of cross- or auto-correlated measurement errors, see Ireland (2004) the particular structure seriously affects parameter estimates.

In *finance*, namely in term structure models stochastic singularity and measurement error solution is a big issue, see e.g.

Piazessi (2010), Hamilton and Wu (2011) or insightful comment on measurement errors by Sims (2003).

Affine models of term structure convey the theory that there is a small number of driving forces (factors) that drive the yield curve. Augmenting selected or all maturities with a stationary process for a measurement error is far from innocent – are these serrially correlated, are these correlated across maturities, etc?

Sims (2003, pp. 502) comments: "*Identification . . . requires that we find a way to distinguish a priori the properties of the measurement error process from that of the process for the unobserved component that satisfies the theory. Because the theory is precisely not about the measurement error, however, we have little guidance as to what to assume about its properties.* . . . And continues: *Is the "error" to be interpreted as reflecting failure of no-arbitrage conditions, so that its existence implies arbitrage opportunities?*"

3. Dimensionality Reduction and Directions of Maximum Variance

I assume that there are observed data $\{\mathcal{Y}_t\}_{t=1}^{T}$ of dimension n_y, corresponding by its structure to model variables in $\mathbf{Y}_t$. Ideally, one would like to use all data for estimation, but only number smaller or equal to stochastic rank n_e of the model can be actually used in likelihood-based estimation.

I consider using *dimensionality reduction* to facilitate the estimation and analysis of models with stochastic singularity, i.e. transformation of the model (rotation) of the model into a smaller dimensional space from which the original information can be reconstructed with no or 'minimal' loss.

Specifically, I consider rotating the model in the direction of maximum variance using non-parametric *principal component analysis.* Given the redundancy in the model, the dimensionality-reduction approach explores redundancy in the observed data. Bellow, the static and dynamic principal components are used with different implications and interpretation of the stochastic singularity.

Principal component analysis is well established in multivariate analysis. In the data without temporal structure, the contemporaneous cross-correlation is a source of redundancy. In case of time series, the additional potential for redundancy stems from temporal cross-correlation. For a $(n_y \times 1)$ process $\mathbf{Y}(t)$ I consider situations when either a matrix transformation $\mathbf{P}$ or filter transformation $\mathbf{P}(L)$ of the form

$$\mathbf{F}_t = \mathbf{P}\mathbf{Y}_t \qquad \xi_t = \mathbf{P}(L)\mathbf{Y}_t \qquad \text{with } \mathbf{P}(z) = \sum_{i=-\infty}^{\infty} \mathbf{P}_i z^i \tag{5}$$

is appropriate for finding components $\mathbf{F}_t$ or ξ_t, where sub-vectors can be used for parameter estimation of the structural model. Using a contemporaneous transformation $\mathbf{F}_t = \mathbf{P}\mathbf{Y}_t$ is easier to implement.

Consider a well-known example, a simple process $\mathbf{U}(t)$ driven by serially uncorrelated stationary process $\nu(t) \sim N(0, \sigma^2)$

$$\mathbf{U}_t = \begin{bmatrix} \nu_t \\ \alpha\nu_{t-1} \end{bmatrix} \; \mathbf{\Gamma}_U^0 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \alpha^2\sigma^2 \end{bmatrix} \; \mathbf{\Gamma}_U^{+1} = \begin{bmatrix} 0 & 0 \\ \alpha\sigma^2 & 0 \end{bmatrix}, \tag{6}$$

with rank-defficient spectral density $\mathbf{S}_U(\lambda)$

$$\mathbf{S}_U(\lambda) = \begin{bmatrix} 1 & \alpha e^{-i\lambda} \\ \alpha e^{i\lambda} & \alpha^2 \end{bmatrix} \tag{7}$$

see Pourahmadi (1994) for further discussion of existence conditions for reduced-rank transfrom $\mathbf{P}$.[3] This simple example demonstrates clearly the difference between a static rank of the process, determined by rank of its covariance matrix and dynamic stochastic rank of the proces as a rank of spectral density.

There is an established literature on relationship of dynamic factor models and principal component analysis, see e.g. Forni, Hallin, Lippi, and Reichlin (2000) and Bai and Ng (2007). In terms of this literature in the example above (6)–(7) a process $\nu(t)$ would be interpreted as dynamic factor and $\mathbf{U}(t)$ is irreducible static factor in this specific case.

It is important to unerstand that I do not suggest to *identify* a factor model or a dynamic factor model. I propose to use as many principal component of the observed data as there are stochastic shocks (or less) and essentially estimate the models using the information in sample spectral density of the data.

4. Rotating the Model in a Maximum Variance Direction

This section demonstrates how to use dimensionality reduction for estimation of structural parameters of a state-space model. First a simpler via *static principal components* is presented to make the idea clear and then a more powerful implemenation using *dynamic principal components* is demonstrated.

4.1. Subspace of 'Static' Principal Components

I am searching for 'best' conteporaneous (static) transformation $\mathbf{P}$ that will allow to capture as much information in the data as possible with a set of orthogonal components. There are multiple options and I will work with principle component analysis (PCA).

Let the observed data vector be $\mathcal{Y}_t$, then it's sample covariance matrix $\Sigma_{\mathcal{Y}}$ and the data can be decomposed into $k \leq r$ principal components where r is a static rank of the data, i.e.

$$\mathbf{\Sigma}_{\mathcal{Y}} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}' \qquad r = \text{rank}(\mathbf{\Sigma}_{\mathcal{Y}}), \tag{8}$$

where $\mathbf{P}, \mathbf{\Lambda}$ denote matrix of eigenvectors and diagonal matrix of eigenvalues, respectively.

I define $\mathbf{P}^k$ to be a projection matrix on a subspace spanned by $k \leq r$ first principal components of the data, hence

$$\mathbf{F}_t^k = \mathbf{P}^k \mathcal{Y}_t, \qquad \hat{\mathcal{Y}}_t = \bar{\mathbf{P}}^k \mathbf{F}_t^k = \bar{\mathbf{P}}^k \mathbf{P}^k \mathcal{Y}_t \tag{9}$$

where $\mathbf{F}_t^k$ is a $(k \times 1)$ vector and $\mathbf{P}^k$ is a $(k \times n_y)$ projection matrix. $\hat{\mathcal{Y}}_t$ is a 'recovered' signal based on low-rank dimensionality transform.

[3]Pourahmadi (1994) demonstrates that for a rank-degenerate stationary process a contemporaneous transform $\mathbf{P}$ produces a lower-dimensional full-rank time series only if the canonical correlation between past and future of the series is smaller than one, i.e. the spectral density rank is constant and the range of the spectral density is constat across frequencies.

It is feasible to view the proces as applying a PCA-filter. The error of approximation $\mathbf{e}_t = \mathcal{Y}_t - \hat{\mathcal{Y}}_t$ has a spectral density $\mathbf{S}_e(\lambda) = \mathbf{L}\mathbf{S}_{\mathcal{Y}}(\lambda)\mathbf{L}'$ where $\mathbf{L} = (\mathbf{I} - \bar{\mathbf{P}}^k\mathbf{P}^k)$. Understanding this PCA-filter induced determination of 'measurment errors' is crucial as it is importnat not to shuffle the assumption under the rug.

For $k = 1$, for instance, the projection in a direction of the first principal component, direction of maximum variance, is just a weighted average of available observables and F_t^k is a univariate time series.

Model Rotation in Directions of Maximum Variance. Having extracted at most $k = n_e$ principal components from the data, I condition the estimation of the model on k and $\mathbf{P}^k$. The models observables are projected onto subspace of k principal components and the model is thus given by

$$\mathbf{F}_{S,t}^k = \mathbf{K}\mathbf{X}_t \tag{10}$$

$$\mathbf{X}_t = \mathbf{T}\mathbf{X}_{t-1} + \mathbf{R}\varepsilon_t, \tag{11}$$

where $\mathbf{F}_{m,t}^k \equiv \mathbf{P}^k\mathbf{Y}_t$, with $\dim(F_{S,t}) = k \leq n_e$, and $\mathbf{K} = \mathbf{P}^k\mathbf{Z}$, which is $(k \times n_x)$. The observables for the conditional likelihood-based estimation, conditional on k, are the k principal components associated with the data $\mathcal{Y}$.

Essentially, one would include first k principal components, as these are determined by orthogonal directions in decreasing order of data variation, but any combinations of principal components is possible, if it provides sharper identification of the model's paramters.

The likelihood of the model can be thus expressed in terms of observable non-singular components $\mathbf{F}_S^k$, conditioned on $\mathbf{\Phi} = \{k, \mathbf{P}\}$

$$\log \mathcal{L}(\{F_t\}_{t=1}^T | \mathbf{\Phi}) = -\sum_{t=1}^{T} \frac{1}{2}\left[k \log 2\pi + \log |\mathbf{\Sigma}_{t,t-1}| + \mathbf{v}_t'\mathbf{\Sigma}_{t,t-1}^{-1}\mathbf{v}_t\right],$$

with $\mathbf{v}_t$ being the prediction-error with Kalman filter implementation of the likelihood.

It is important to recall several key characteristics of PCA. First, it is a *non-parametric* algorithm. Second, principal components are mutually orthogonal. Third, principal components are not defined uniquely, only the associated sub-space is uniquely defined. Finally, PCA is very sensitive to units of measurement, i.e. scaling.

The sensitivity to scaling is not likely to be an issue in the current context as long as one takes care that the units of the data and of the model agree, which is key in case of any estimation procedure. Obviously, some data transformation may result in loss of identification on some parameters, e.g. detrending.

Finally, the method of 'static' principal components is only one of the possible choices of arriving at a transform $\mathbf{P}$. A method closely related to it and to the next section is to follow Forni, Hallin, Lippi, and Reichlin (2005) and use a projection matrix as a result of generalized principal component on a factor space obtained by a dynamic principal components of Brillinger (1981), that takes more into account serial correlation of the series.

4.2. *Space of Dynamic Principal Components*

In the previous section I introduced computationally very simple rotation of the model using a conteporaneous transform $\mathbf{P}^k$ in a direction of *static principal components* obtained using eigenvalue decomposition of contemporaneous covariance matrix.

Importantly, there is another space that the model can be mapped to – *space of dynamic principal components.* This method should in principle capture the dynamic correlations among the components of $\{\mathcal{Y}_t\}_{t=t_0}^T$ more efficiently, see Brillinger (1981) and Forni et al. (2000).

Let's denote $(k \times 1)$ vector $\mathbf{F}_{D,t}$ be the vector of dynamic principal components obtained using a projection matrix polynomial $\mathbf{P}^k(z)$ if

$$\mathbf{F}_{D,t}^k = \mathbf{P}^k(L)\mathcal{Y}_t \qquad \hat{\mathcal{Y}} = \bar{\mathbf{P}}^k(L)\mathbf{F}_{D,t}^k \qquad \mathbf{P}^k(z) = \sum_{i=-\infty}^{\infty} \mathbf{P}_i^k z^i. \tag{12}$$

The major difference between (9) and (12) is that dynamic principal component are not simply a weighted average of contemporaneous elements of $\mathcal{Y}_t$, but can involve their leads and lags to account for dynamic relationships.

Rotating in Directions of Dynamic Principal Components. The state-space model can be rotated in direction of k dynamic principal component redefining as

$$\mathbf{F}_{D,t}^k = \mathbf{P}^k(L)\mathbf{Z}\mathbf{X}_t \tag{13}$$

$$\mathbf{X}_t = \mathbf{T}\mathbf{X}_{t-1} + \mathbf{R}\varepsilon_t. \tag{14}$$

In case of finite order matrix-valued polynomial $\mathbf{P}^k(z)$ it is very easy to expand the state-space of the model and put the model into standard form and carry on with the analysis using dynamic factors $\{\mathbf{F}_{D,t}^k\}_{t=1}^T$ implied by the data, conditioning on k.

In general $\mathbf{P}^k(z)$ may not be finite, hence approximation is needed in time domain to calculate factors and the model state-space form. A more convenient approach is to stick to *frequency domain.*

Likelihood-based estimation of the transformed model. Following Hannan (1970), inter alia, the Gaussian log-likelihood can be approximated using Whittle-likelihood in frequency domain

$$\log \mathcal{L} \propto -\frac{1}{2}\sum_{j=0}^{T-1} \log[\det \mathbf{S}_F(\lambda_j)] \tag{15}$$

$$-\pi \times \operatorname{tr} \sum_{j=0}^{T-1} [\mathbf{S}_F(\lambda_j)^{-1}\mathbf{I}_{\hat{\mathcal{Y}}}(\lambda_j)],$$

evaluated at $\lambda_j = 2\pi j/T$, where $\mathbf{I}_{\hat{\mathcal{Y}}}(\lambda) = \mathbf{P}^k(\lambda)\mathbf{I}_{\mathcal{Y}}\mathbf{P}^k(\lambda)^H$ is a periodogram of of $\mathcal{Y}_t$ and $\mathbf{S}_F(\lambda) = \mathbf{P}^k(\lambda)\mathbf{S}_{yy}(\lambda)\mathbf{P}^k(\lambda)^H$ is a spectral density of model observables after applying DPCA-filter $\mathbf{P}^k(L)$. Note that except special degenerate cases $(k \times k)$ matrices $\mathbf{S}_F(\lambda)$ is full rank for almost all frequencies λ.

Dynamic principal component filter is obtained using a procedure proposed by Brillinger (1981, Ch. 9, Theorem 9.3.1).

The problem is to determine $(k \times n_y)$ linear filter $\mathbf{P}^k(L)$ and $(n_y \times k)$ filter $\bar{\mathbf{P}}^k(L)$ such as to minimize the approximation error

$$\mathbb{E}||\mathcal{E}|| = \mathbb{E}||\mathcal{Y}_t - \hat{\mathcal{Y}}_t|| = \mathbb{E}||\mathcal{Y}_t - \bar{\mathbf{P}}^k(L)\mathbf{P}^k(L)\mathcal{Y}_t|| \qquad (16)$$

arising from (12). The solution relies on spectral analysis, computing principal component of spectral density at individual frequency bands.

Given an estimate of $(n_y \times n_y)$ spectral density matrix $\mathbf{S}_\mathcal{Y}(\lambda)$ of $\mathcal{Y}(t)$ and its eigenvalue (spectral) decomposition

$$\mathbf{S}_\mathcal{Y} = \mathbf{V}(\lambda)\mathbf{\Lambda}(\lambda)\mathbf{V}(\lambda)^H \qquad (17)$$

frequency-domain filter $\mathbf{P}^k(\lambda)$ is given by first k (out of n_y) transposed conjugates of eigenvectors $\mathbf{V}(\lambda)$, i.e. first k rows of $\mathbf{V}(\lambda)^H$. It follows that $\bar{\mathbf{P}}^k(\lambda) = \mathbf{P}^k(\lambda)^H$ from symmetry of spectral density matrix. Coefficients of $\mathbf{P}^k(L)$ result from a inverse-Fourirer transform of $\mathbf{P}^k(\lambda)$ as these are Fourier pairs.[4]

In case of dynamic principal components the analogy with pre-filtering is much clearer – original data and the model are both transformed using a DPCA-filter $\mathbf{P}^k(L)$. The 'measurement' error $\mathcal{E}_t$ has zero mean and spectral density given by

$$\mathbf{S}_\mathcal{E}(\lambda) = [I - \mathbf{A}(\lambda)]\mathbf{S}_\mathcal{Y}(\lambda)[I - \mathbf{A}(\lambda)]^H, \qquad (18)$$

where $\mathbf{A}(\lambda) = \bar{\mathbf{P}}^k(\lambda)\mathbf{P}^k(\lambda)$ is the 'purifying filter'.

The estimation in frequency domain is very flexible and fast. The only nuisance of the procedure suggested is the estimation of unobserved variates $\{\mathbf{X}(t), \varepsilon(t)\}$ associated with stochastically singular system (1) as convenient Kalman filter recursions are not available, unless it is possible to truncate the filter or when it converges 'sufficiently fast'.

4.2.1. Estimating state variables

Unobserved variables implied by the singular model can be estimated using the dynamic principal filter. Due to infinite-dimensional transformation by $\mathbf{P}^k(z)$ the Kalman filter and smoother algorithms are not available for economic analysis of the historical data.

The problem is well defined in terms of classical approach of Wiener-Kolmogorov theory by employing spectral densities of the unobservables $\mathbf{S}_X(\lambda)$ and model-implied dynamic principal components $\mathbf{S}_F(\lambda)$, see e.g. lucid explanation in Whittle (1983) for details. In case of doubly-infinite sample a relationship

$$\hat{\mathbf{X}}_{t|T} = \mathbf{\Omega}(L)\mathbf{P}^k(L)\mathcal{Y}_t \qquad \mathbf{\Omega}(z) = \mathbf{S}_{X,F_D}(z)\mathbf{S}_{F_D}(z)^{-1}, \qquad (19)$$

where $\mathbf{S}_{X,F_D}(\lambda)$ is cross-spectral density matrix and $\mathbf{S}_{F_D}(\lambda)$ is spectral density matrix implied by the parameterized infinite-dimensional state-space model (13).

The practical implementation of the filter I choose relies on finite analogue to stochastic process as represented by circulant matrices and their relationship to discrete Fourier transform (DFT), see e.g. **?** for a similar approach. All computations are carried out at frequency domain and then transformed back to time domain.

[4] Actual implementation of frequency-domain likelihood requires the filter to be estimated using a sample-estimate of cross-spectrum. Cross-spectrum estimation (smoothed periodogram) is sensitive to a selection of a a smoothing kernel.

5. Controlled Experiments

The purpose of controlled experiments is to test if the method provides sufficient identification and delivers more precise parameter estimates, than using limited number of time series.

I consider first a simplistic stochastic process with a factor structure and simple version of the dynamic general equilibrium model as a data generating process for testing the procedure.

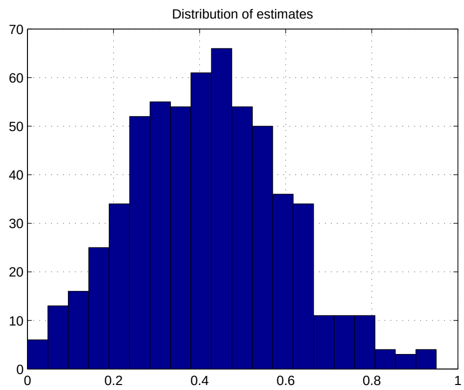
5.1. Simple stochastic process

The most simple setup is a stochastic process defined as

$$x_t = \rho_x x_{t-1} + \varepsilon_x \qquad (20)$$
$$u_t = \rho_u u_{t-1} + (1-\rho_u)\alpha_{x,u} x_t, \qquad (21)$$

where I assume that $\rho_x = 0.9, \rho_u = 0.4$ are known (conditioned upon) and try to estimate only the parameter $\alpha_{x,u}$, taking a true value of $\alpha_{x,u} = 0.4$. The setup is very simplistic. The parameter is not identified when only x_t is observed and can be estimated using only u_t. The results using a dynamic principal component filter suggest that the estimate has smaller standard error than estimates using u_t only. Based on a sample of $T = 200$ and $N = 600$ replications, using the Bartlett kernel with window lag $M = 20$ (not optimized) the distribution of the estimate is depicted at Fig. 1.

Figure 1: Distribution of estimates $\hat{\alpha}_{x,u}$



5.2. Simple dynamic economic model

The model is a one sector real business cycle model in a decentralized version, with representative households and firms. Households derive the utility from habit adjusted consumption C_t and suffer disutility from work L_t and rent their capital stock K_{t-1} and labor L_t to labor unions and firms at competive prices $\Delta_{k,t}$ and W_t. Profits of trade unions and monopolistic final good firms aggregates into total profits $\Pi = \Pi^U_t + \Pi^P_t$. In order to accumulate the capital stock household buy investments J_t. Household i maximizes their lifetime utility subject to their budget constraint

$$\mathfrak{W}_t = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \mathcal{U}(C_t, L_t) \qquad (22)$$

subject to

$$P_tC(i)_t + P_tJ(i)_t + B(i)_t \tag{23}$$
$$= i^l_{t-1}B(i)_{t-1} + P_t\Delta_{k,t}K_{t-1} + W_tL_t + \Pi. \tag{24}$$

Monopolistically competitive final good firms rent labor from trade unions at wage rate V_t and capital stock from household, facing quadratic adjustment costs of re-setting the price of the good. The firm maximizes $\mathbb{E}_t\sum_{s=0}^{\infty}\Xi_{t,s}\Pi^P_{t+s}$

$$\Pi^P_t = P_tY_t - V_tL^U_t - P_t\Delta_{k,t}k_t - P_t\frac{\phi}{2}\left(\frac{\pi_t(j)}{\pi_{t-1}} - 1\right)^2, \tag{25}$$

subject to demand for their goods based on usual CES aggregate, with elasticity of substitution among individual varieties σ_p. The productin function available to every firm is a standard Cobb-Douglas one with labor-augmenting technology shock Z_t

$$Y_t = k^\alpha[Z_tL_t]^{1-\alpha}. \tag{26}$$

The labor augmenting technology shock Z_t is assumed to be stationary AR(1) process defined by $\log Z_t = \rho\log Z_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0,\sigma_z^2)$.

Labor unions solve analogous problem as final good firms with the instantaneous profits given by

$$\Pi(u)^U_t = V_tL(u)^d_t - W_tL(u)_t - V_t\frac{\phi_v}{2}\left(\frac{\pi(u)^V_t}{\pi^V_{t-1}} - 1\right)^2, \tag{27}$$

where $L(u)^d_t)$ denotes the variety sold by a particular trade union branch and $L(u)_t$ is a demand of the branch of the hours worked by households.

Aggregation implies that $k_t = K_{t-1}$. The functional form of momentary utility is chosen to imply separability of consumption and leisure and would not be consistent with balanced growth expansion of the economy,

$$\mathcal{U}_t \equiv \frac{[C_t - hC_{t-1}]^{1-\sigma}}{1-\sigma} - \chi\frac{L_t^{1+\eta}}{1+\eta}. \tag{28}$$

I assume perfectly competitive financial sector which can be source of financial shock due to intermediation and risk-preference changes, giving rise to interet rate premium ζ_t on top of the central bank policy rate that serves as marginal refinancing cost for financial intermediary, hence

$$i^l_t = i_t\zeta_t \tag{29}$$

The policy rule determines the central bank nominal interest rate

$$4\log i_t = \rho_i 4\log i_{t-1} + (1-\rho_i)[4\log\bar{i}_t + \rho_\pi\pi^4_{t+3}/\pi^*_{t+3}], \tag{30}$$

i.e. monetary policy targets year over year inlfation forecast deviation from the target three periods ahead.

The parameter vector of interest consists of eight parameters and is defined as $\theta = \{\alpha,\beta,\sigma,\eta,\delta,\rho,\sigma_z\}$. The parameter χ is always chosen to keep agents working constant portion of time endowment in a steady state. True values for data generating process are in Tab. 1.

The model is solved by log-linearization in the vicinity of a deterministic steady state.

Table 1: True parameter values of the data generating process

Parameter	α	β	σ	η	δ	ρ	σ_z
Value	0.6	0.99	2	1	0.025	0.95	0.01

5.2.1. *Design of the Experiment*

For the dynamic model I use the following set of six observed variables: $c_t, j_t, y_t, i_t, \pi_t$ and π^v_t with the model having up to three stochastic innovations – u^z_t, u^{prem}_t, u^g_t to stationary TFP, risk premium and government spending.

I design a controlled experiments as follows. A data generating process (DGP) is simulated for $T = 100$ periods using true parameter values, with $N = 200$ replications. For each replication a projection matrix $\mathbf{P}^k$ is calculated, though the DGP implies a distribution for elements of $\mathbf{P}^k$.

Variables are measured in percentage deviation from a steady-state and I consider observing the following: (i) output, (ii) consumption, (iii) investment, (iv) hours worked, (v) real wages and (vi) rental rate. Inclusion of the capital stock, a slow moving state variable, is analyzed separately. A PCA-conditioned estimation is contrasted with estimation using only data for consumption.

I check the identification of the parameters first. For each case I calculate Fisher Information Matrix (FIM) for the model evaluated at true parameter values and evaluate its condition number and implied standard errors of classical estimate. The lower the condition number, the smaller is the degree of identification of the model, given particular set of observables, see Andrle (2010), inter alia.

Second, I construct a grid for the parameter space, which includes a true parameter value.

6. Results

First, I discuss the results of simulation experiment and check identification of the parameters. Second, I use macroeconomic data from th United States to demonstrate (a well known) fact that it has very parsimonious factor structure, indicating the usefullness of the method.

6.1. *Simulation Results*

7. Conclusion

This paper suggest a method of estimating parameters in stochastically singular models.

TBW

References

Altug, S. (1989). Time-to-build and aggregate fluctuations: Some new evidence. *American Economic Review 30*(4), 889–920.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics 25*, 52–60.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics & Statistics 82*(4), 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of American Statistical Association 100*(September), 830–840.

Hamilton, J. and J. Wu (2011, April). Testable implications of affine term structure models. Working Paper 16931, NBER, Cambridge, MA.

Hannan, E. (1970). *Multiple Time Series*. New York: Wiley.

Ingram, B., N. Kocherlakota, and N. Savin (1994). Explaining business cycles: A multiple-shock approach. *Journal of Monetary Economy 34*, 415–428.

Ireland, P. (2004). A method for taking models to the data. *Journal of Economic Dynamics and Control 28*(6), 1205–1226.

Piazessi, M. (2010). *Affine Term Structure Models, in: Handbook of Financial Econometrics: Volume 1 – Tools and Techniques , eds. Y. Ait-Sahalia and L.P. Hansen*. Amsterdam: North-Holland.

Pourahmadi, M. (1994). Canonical correlation and reduction of multiple time series. *Ann. Inst. Statist. Math. 46*(4), 625–631.

Sargent, T. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy 97*(2), 251–287.

Sims, C. (2003). [iterative and recursive estimation in structural nonadaptive models]: Comment. *Journal of Business & Economic Statistics 21*(4), 500–503.

Whittle, P. (1983). *Prediction and Regulation by Linear Least Squares Methods*. Minneapolis: University of Minnesota Press.

8. Appendix

8.1. Dynamic principal components

The definition of dynamic principal components follows (Brillinger, 1981, pp. 344, Theorem 9.3.1)

8.2. DGE Model Formulation

The model is a simple dynamic general equilibrium model of a closed economy with simplistic financial sector, whose only role is to introduce risk premium into the lending rate. I omitt capacity utilisation and other features used in business cycle literature for the sake of simplicity.

$$\hat{c}_t = \frac{h}{1+h}\hat{c}_{t-1} + \frac{1}{1+h}\hat{c}_{t+1} - \frac{1-h}{1+h}\frac{1}{\sigma}\hat{r}_t \quad (31)$$

$$\hat{q}_t = \kappa[\hat{j}_t - \hat{j}_{t-1}] - \beta\kappa[\hat{j}_{t+1} - \hat{j}_t] \quad (32)$$

$$\hat{q}_t = -\hat{r}_t + (\bar{\Delta}/\bar{r})\hat{\Delta}_{k,t+1} + [(1-\delta)/\bar{r}]\hat{q}_{t+1} \quad (33)$$

$$\hat{K}_t = (1-\delta)\hat{K}_{t-1} + \delta\hat{j}_t \quad (34)$$

$$\hat{y}_t = (C/Y)\hat{c}_t + (J/Y)\hat{j}_t + (G/Y)\hat{g}_t \quad (35)$$

$$\hat{y}_t = \alpha\hat{k}_t + (1-\alpha)[\hat{Z}_t + \hat{L}_t] \quad (36)$$

$$\hat{k}_t = \hat{K}_{t-1} \quad (37)$$

$$\hat{v}_t = \widehat{rmc} + \hat{y}_t - \hat{L}_t \quad (38)$$

$$\hat{\Delta}_{k,t} = \widehat{rmc} + \hat{y}_t - \hat{k}_t \quad (39)$$

$$\hat{\lambda}_t = -[\sigma/(1-h)] \times [\hat{c}_t - h\hat{c}_{t-1}] \quad (40)$$

$$-\eta\hat{L}_t = \hat{\lambda}_t + \hat{w}_t \quad (41)$$

$$r_t = i^l_t - \pi_{t+1} \quad (42)$$

$$\hat{i}^l_t = \hat{i}_t + \hat{\zeta}_t \quad (43)$$

$$4\hat{i}_t = 4\rho_i\hat{i}_{t-1} + (1-\rho_i)[\rho_\pi(\hat{\pi}^4_{t+3})] \quad (44)$$

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi_P}\hat{\pi}_{t+1} + \frac{\xi_P}{1+\beta\xi_P}\hat{\pi}_{t-1} + \kappa_P\widehat{rmc}_t \quad (45)$$

$$\hat{\pi}^W_t = \frac{\beta}{1+\beta\xi_W}\hat{\pi}^W_{t+1} + \frac{\xi_W}{1+\beta\xi_W}\hat{\pi}^W_{t-1} + \kappa_W[\hat{w}_t - \hat{v}_t] \quad (46)$$

$$\hat{\pi}^W_t = \hat{\pi}_t + (\hat{w}_t - \hat{w}_{t-1}) \quad (47)$$