SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-12

The SR Approach: a new Estimation Method for Non-Linear and Non-Gaussian Dynamic Term Structure Models

Martin M. Andreasen and Bent Jesper Christensen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

The SR Approach: a new Estimation Method for Non-Linear and Non-Gaussian Dynamic Term Structure Models

Martin M. Andreasen*
Bank of England and CREATES

Bent Jesper Christensen
Aarhus University and CREATES

March 3, 2010

Abstract

This paper suggests a new and easy approach to estimate linear and non-linear dynamic term structure models with latent factors. We impose no distributional assumptions on the factors and they may therefore be non-Gaussian. The novelty of our approach is to use many observables (yields or bonds prices) in the cross-section dimension. An important benefit of using many observables in each time period is that the latent factors can be estimated quite accurately using standard regressions, and that parameters can be estimated by standard moment matching methods.

Keywords: Bond data, GMM, Non-linear filtering, Non-linear least squares, Missing observations, SMM.

JEL: C10, C30

*We thank Juan Rubio-Ramirez, Jesus Fernandez-Villaverde, Mikhail Chernov, and George Kapetanios for useful comments and discussions. Corresponding author: martin.andreasen@bankofengland.co.uk. Telephone number: +44 207 601 3431. Both authors greatly acknowledge financial support from the Danish Center for Scientific Computation (DCSC). The authors also appreciate financial support to Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation. Finally, the views expressed in this paper are solely those of the authors and not necessarly those of Bank of England.

1 Introduction

A key property of all dynamic term structure models is their ability to price bonds with different maturities. Following the work of Vasicek (1977) and Cox, Ingersoll & Ross (1985), the dynamics in these prices are often explained by unobserved or latent factors (see for instance Duffie & Kan (1996), Dai & Singleton (2000), Duffee (2002), Diebold & Li (2006), among others). The presence of latent factors greatly simplifies dynamic term structure models, but it comes at the cost of making the models difficult and time consuming to estimate, in particular on a large and highly unbalanced panel of bond data. To make estimation feasible, it is therefore common practice to first extract or estimate 5-10 zero-coupon yields from the available bond data, and then use these yields to estimate the dynamic term structure model (see Duan & Simonato (1999), de Jong (2000), Duffee (2002), among others).

The present paper deviates from the common practice of using relatively few observables each time period when taking dynamic term structure models to the data. Instead, we suggest using a large number of observables (yields or bonds prices) for the estimation, and we argue that this, contrary to the common belief in the literature, simplifies the estimation process. This is because the latent factors can be estimated quite accurately by a sequence of standard regressions with many observables. That is, in each period we set the latent factors to minimize the distance between the observed yields/bond prices and the model implied yields/bond prices. We refer to this procedure as the "regression filter" because it like other filters such as the Kalman filter estimates the latent factors.

For non-linear models with normally distributed measurement errors we show analytically that the estimates from this regression filter converges to the optimal smoothing estimates when the number of observables in each period tends to infinity. In the case of linear and Gaussian models, this means that the regression filter converges to the Kalman smoother. We also show how to estimate parameters in dynamic term structure models using standard moment matching methods and the regression filter. Throughout this paper, we refer to the sequential use of regressions to estimate the latent factors and parameters as the Sequential Regression (SR) approach.

Our second contribution is to show consistency and asymptotic normality of the SR approach. These results hold for non-linear models with potentially non-Gaussian factor dynamics and they do so with weaker restrictions than those needed for likelihood inference. This generality and robustness is the most important advantage of the SR approach because it enables us to estimate a very wide class of dynamic term structure models in a simple, reliable, and fast manner.

Our approach has at least five additional advantages compared to existing estimation methods for dynamic term structure models. Firstly, the SR approach deals easily with large and potentially highly unbalanced panel of data. Secondly, the approach is very easy to implement because the filtering step only involves standard regressions. Thirdly, no transition function is needed in order to estimate latent factors and a subset of parameters; not even a parametric family needs to be specified. This is contrary to the requirements for the Kalman filter and all non-linear and/or non-normal filters, which require a specification of the transition function for the latent factors. Fourthly, by using many observables in each time period, we use information from financial markets more efficiently than methods estimating dynamic term structure models using just 5-10 observables. Finally, the SR approach is computationally much faster to use than any likelihood based method.

We emphasize that these advantages hold for *all* linear and non-linear dynamic term structure models and with no distributional assumptions on the factors. Two disadvantages of the SR approach are that we may need to rely on sequential identification of parameters and that the SR approach is not fully efficient. While acknowledging these disadvantages we do point out that sequential identification can be easier to implement in large models with many parameters compared to simultaneous estimation of all parameters. Moreover, simulation results show that the SR approach based on 25 observables and sequential identification is just as efficient as Maximum Likelihood (ML) with about 5-10 observables. Given these results, we argue that estimation of dynamic term structure models could benefit from using many observables instead of focusing only on how to do likelihood inference based on relatively few observables.

The SR approach requires that a large number of observables are available in each time period, and the bond market is therefore a natural place to look. Another possibility is to include the entire estimated zero-coupon curve in each period. However, this must be done with some care because when estimating the zero-coupon yield curve by a model with n parameters, this curve only has an information content corresponding to these n yields. Viewed from this perspective, the unsmoothed Fama Bliss interest rates should therefore be preferred to interest rates estimated from, for instance, the Nielson Siegel curve, because the former contains more information than the latter (Fama & Bliss (1987), Nelson & Siegel (1987)). With this minor restriction in mind, the SR approach can also be applied to estimated zero-coupon yields.

The rest of the paper is organized as follows. Section 2 presents the wide class of dynamic term structure models which can be estimated by the SR approach. The SR approach is formally presented in section 3, and asymptotic properties for the SR approach are derived and discussed in section 4. Section 5 examines the finite sample properties of the SR approach compared to the traditional ML approach for a linear and Gaussian term structure model. Concluding comments are given in section 6. All proofs are deferred to the appendix.

2 The SR approach

2.1 The state space representation

This section presents the class of dynamic term structure models considered in this paper. We adopt the notation that $\mathbf{y}_t$ with dimension $n_{y,t} \times 1$ contains all observed yields or bond prices related to the model at time t. Note that we explicitly allow for an unbalanced data panel by letting the dimension of $\mathbf{y}_t$ be time-dependent. The observable factors driving $\mathbf{y}_t$ are denoted $\mathbf{x}_{1,t}$ and has dimension $n_{x_1} \times 1$, and the latent (i.e. unobserved) factors are denoted $\mathbf{x}_{2,t}$, having dimension $n_{x_2} \times 1$. Jointly, we let $\mathbf{x}_t \equiv \left[\begin{array}{cc} \mathbf{x}'_{1,t} & \mathbf{x}'_{2,t} \end{array} \right]'$ which has dimension $n_x \times 1$ where $n_x \equiv n_{x_1} + n_{x_2}$.

Given this notation, we consider dynamic term structure models with the following representation

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right) + \mathbf{v}_t. \tag{1}$$

Here, we account for potential measurement errors in the observables by the vector $\mathbf{v}_t \equiv \{v_{t,i}\}_{i=1}^{n_{y,t}}$ which has dimension $n_{y,t} \times 1$. When bond prices are used as observables, such measurement errors can be caused by the presence of i) non-synchronous trading, ii) rounding of market prices, and/or iii) bid-ask spreads. When extracted or estimated zero-coupon yields are used as observables, the measurement errors capture any errors involved in constructing these yields. We refer to (1) as the set of measurement equations.

To introduce the remaining notation, we also specify a standard markovian law of motion for the factors even though this assumption is not needed for estimating $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$ and $\boldsymbol{\theta}_1$. That is we let

$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \mathbf{w}_{t+1}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right), \tag{2}$$

where $\mathbf{w}_t$ has dimension $n_w \times 1$ and are the $\mathcal{IID}$ disturbances to the factors.

The model is parameterized by $\boldsymbol{\theta}$ with dimension $L \times 1$. We decompose $\boldsymbol{\theta}$ into $\begin{bmatrix} \boldsymbol{\theta}_1' & \boldsymbol{\theta}_2' \end{bmatrix}'$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ have dimension $L_1 \times 1$ and $L_2 \times 1$, respectively, and $L = L_1 + L_2$. The elements in $\boldsymbol{\theta}_1$ can be identified from the measurement equations whereas $\boldsymbol{\theta}_2$ must be identified based on the law of motion for the factors. This explains why $\boldsymbol{\theta}_2$ do not appear in (1). For reduced form term structure models, $\boldsymbol{\theta}_1$ typically contains all the risk neutral parameters and $\boldsymbol{\theta}_2$ contains the parameters specifying the market price of risk. For general equilibrium models, the market price of risk is often a function of $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ can therefore be expected to be empty in this class of models.

We impose the following assumptions on the considered class of dynamic term structure models:

$$E\left[\mathbf{v}_t | \, \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right] = \mathbf{0} \quad \text{for some } \boldsymbol{\theta}^o \in \boldsymbol{\Theta} \text{ and } \mathbf{x}_{2,t}^o \in \mathcal{X}_{2,t} \text{ for all } t = 1, ..., T \tag{3}$$

$$Var\left(\mathbf{v}_t\right) = diag\left(\{Var\left(v_{j,t}\right)\}_{j=1}^{n_{y,t}}\right) \text{ where } Var\left(v_{j,t}\right) < \infty \quad \text{for all } t, \; j \tag{4}$$

$$\text{Independence between } \mathbf{v}_t \text{ and } \mathbf{w}_{t-k} \text{ for } k = 0, 1, 2, ... \tag{5}$$

The superscripts denotes "the true value" of the parameter, and the variables in $\mathbf{z}_{t,j}$ contains exogenous variables which we use below to model potential heteroscedasticity at time t across the observables. Assumption (3) means that the model is correctly specified for the conditional mean in all time periods and is clearly a very weak assumption which all existing methods also impose (see Durbin & Koopman (2001), Doucet, de Freitas & Gordon (2001), among others). Our second assumption in (4) allows for potential time-varying second moments, and this assumption is therefore much less restrictive than the normal assumption of constant second moments. The third assumption of independence between $\mathbf{v}_t$ and $\mathbf{w}_{t-k}$ is standard for term structure models and only imposed in our context to facilitate the estimation of $\boldsymbol{\theta}_2$.

We do not impose any distributional assumptions on $\mathbf{v}_t$ and $\mathbf{w}_t$ or assume that $\mathbf{v}_t$ is homoscedastic and uncorrelated across time. On the other hand, most of these assumptions have to be imposed for likelihood based inference (see Durbin & Koopman (2001), Doucet et al. (2001), among others).

2.2 An illustrative example: The one factor CIR model

We illustrate our general framework by applying it to the standard one-factor model by Cox et al. (1985). We refer to this model as the CIR model.

The instantaneous interest rate r_t is assumed to evolve according to

$$dr_t = \kappa\left(\delta - r_t\right) dt + \sigma\sqrt{r_t} dz_t, \tag{6}$$

where $\kappa\left(\delta - r_t\right)$ is the instantaneous drift and dz_t is a Brownian motion under the physical probability measure. The price of a zero-coupon bond with maturity τ at time t is given by (see Brown & Dybvig (1986))

$$\ln p\left(t, \tau\right) = \ln A\left(t, \tau\right) - B\left(t, \tau\right) r_t \tag{7}$$

where

$$A(t,\tau) \equiv \left[\frac{\phi_1 \exp\{\phi_2 \tau\}}{\phi_2 (\exp\{\phi_1 \tau\} - 1) + \phi_1} \right]^{\phi_3}$$

$$B(t,\tau) \equiv \frac{\exp\{\phi_1 \tau\} - 1}{\phi_2 (\exp\{\phi_1 \tau\} - 1) + \phi_1}$$

$$\phi_1 \equiv \sqrt{(\kappa + \lambda)^2 + 2\sigma^2} \qquad \phi_2 \equiv \frac{\kappa + \lambda + \phi_1}{2} \qquad \phi_3 \equiv \frac{2\kappa\delta}{\sigma^2}$$

The risk premium is specified as $\lambda\sqrt{r_t}/\sigma$.

When we apply our framework to this model, equation (7) is the measurement equation with $\mathbf{x}_{2,t} \equiv r_t$ and $\mathbf{x}_{1,t}$ being empty. The transition equation for factor r_t is given by (6). For the parameters we have $\boldsymbol{\theta}_1 \equiv \left[\begin{array}{ccc} \kappa\delta & (\kappa+\lambda) & \sigma \end{array} \right]$ and $\boldsymbol{\theta}_2 \equiv \kappa$. The somewhat peculiar specification of $\boldsymbol{\theta}_1$ is due to the issue of parameter identification which we return to in section 3.5.

3 The SR approach

This section presents the SR approach for estimation of $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$ and $\boldsymbol{\theta}$. In order to place the contribution of the SR approach within the literature, we begin by a brief discussion of existing methods for estimating dynamic term structure models. Here, focus is given to methods where the factors are estimated.[1] Section 3.2 turns to the estimation of the latent factors in the SR approach, and this estimator is then related to other methods in section 3.3. Sections 3.4 and 3.5 deal with estimation of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively, in the SR approach. Section 3.6 summarizes the SR approach.

3.1 Existing methods for estimating dynamic term structure models

It is well-known that the Kalman smoother is the optimal solution to the problem of estimating the latent factors in the special case where i) the functions $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ are both linear in $\mathbf{x}_t$, ii) $\mathbf{v}_t$ and $\mathbf{w}_t$ are each independent, normally distributed, and have constant variances, iii) $\mathbf{x}_0$ is normally distributed, and iv) $\mathbf{x}_0$, $\mathbf{v}_t$, and $\mathbf{w}_t$ are mutually uncorrelated at all leads and lags. Moreover, the parameters $\boldsymbol{\theta}$ can be estimated by ML based on the Kalman filter. The optimal estimator for the latent factors or the expressions for the likelihood function do not have a closed form expression when we deviate from linear and Gaussian term structure models and approximations are therefore needed in such cases.

For linear and non-Gaussian models, the Kalman filter and the Kalman smoother can still be used to estimate $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$, and $\boldsymbol{\theta}$ can in most cases be estimated consistently by quasi ML (see Hamilton (1994) and Duan & Simonato (1999)). However, an important exception is models with time-varying volatility in the factor dynamics. In this case $\boldsymbol{\theta}$ cannot be estimated consistently by quasi ML because the value of the factor in the previous period enter into the expression of the factors' conditional second moments.[2] An attractive alternative in this case is therefore to impose the ad hoc assumption that as many yields as factors are measure without errors (see for instance Chen & Scott (1993) and

[1] We refer to Carrasco & Florens (2002) for a brief presentation of simulation based methods (i.e. SMM, Indirect Inference, and EMM) where the factors are not estimated.

[2] Nevertheless, we note that the bias in the quasi ML estimator based on the Kalman filter is generally found to be small in this case (see Duan & Simonato (1999) and de Jong (2000)).

Pastorello, Patilea & Renault (2003)). A great disadvantage of this method is that the estimation results by construction depend on the set of yields which is assumed to be measured without errors.

For non-linear term structure models, a common approximation is to linearize the state space system in (1) and (2), and apply the Kalman filter to this approximated system (Jazwinski (1970)). Another and more accurate approximation is to limit the focus to linear updating rules as in the Kalman filter, but to approximate the non-linear moments in the equations up to at least second order as done in the unscented Kalman filter or the central difference Kalman filter (Julier, Uhlmann & Durrant-Whyte (1995), Norgaard, Poulsen & Ravn (2000)). However, these extensions of the Kalman filter cannot evaluate the likelihood function and quasi ML can only be expected to give consistent and normally distributed estimators in very few cases (see for instance Andreasen (2008)).

Another branch of the literature has therefore developed methods for non-linear and potentially non-Gaussian models that approximate the likelihood function based on importance sampling. The drawback of these methods is that they are quite technical to implement and very time-consuming to use, and this has so far limited their use in the context of dynamic term structure models. For instance, we are only aware of the papers by Brandt & He (2005) and Rossi (2004) which estimate dynamic term structure models based on importance sampling.

3.2 Estimation of latent factors

The SR approach is based on the observation that if a large number of observables are available in each time period, then the optimal estimator of the latent factors may be well approximated by simply ignoring the time dimension and running a sequence of regressions. That is, we suggest to do the filtering in each period by solving the following regression problem where $\boldsymbol{\theta}_1$ is constant:

$$\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1) = \arg\min_{\mathbf{x}_{2,t}} \sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \boldsymbol{\theta}_1\right)\right)^2}{2Var\left(v_{j,t}\right)}. \tag{8}$$

Here, we use the notation that $\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_1) \equiv \left[\begin{array}{cccc} g_1(\mathbf{x}_t;\boldsymbol{\theta}_1) & g_2(\mathbf{x}_t;\boldsymbol{\theta}_1) & \ldots & g_{n_{y,t}}(\mathbf{x}_t;\boldsymbol{\theta}_1)\end{array}\right]'$ and similarly for $\mathbf{y}_t$. The estimated latent factors from this regression are denoted $\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1)$, because they are a function of $\boldsymbol{\theta}_1$. We refer to this repeated use of regressions to recover the latent factors from the observed yields or bond prices as the regression filter.

Thus, the estimates of $\{\mathbf{x}_{2,t}\}_{t=1}^T$ in the SR approach do not reply on the transition equations of the factors. This has at least two interesting implications. Firstly, the estimates of $\{\mathbf{x}_{2,t}\}_{t=1}^T$ are consistent even if the factors are governed by complicated processes which are i) non-markovian, ii) fractional integrated, or iii) display jumps. Secondly, the SR approach works equally well for models set in continuous and discrete time.

Estimates of $\{Var(v_{j,t})\}_{j=1}^{n_{y,t}}$ are necessary to make the regression in (8) feasible. Using standard results from the cross-section econometric literature, $\{Var(v_{j,t})\}_{j=1}^{n_{y,t}}$ can be estimated by first running an unweighted regression where $Var(v_{j,t}) = 1$ for $j = 1, 2, ..., n_{y,t}$. This produces consistent estimates of $\mathbf{x}_{2,t}(\boldsymbol{\theta}_1)$, which can be used to calculate the estimated residuals $\hat{v}_{t,j}$. From $\hat{v}_{t,j}$, we then suggest to model potential variation in $\left\{\hat{v}_{j,t}^2\right\}_{j=1}^{n_{y,t}}$ based on i) time to maturity, ii) duration, iii) liquidity, etc. and then run the regression

$$\ln\left(\hat{v}_{j,t}^2\right) = \boldsymbol{\gamma}'\mathbf{z}_{t,j} + \varepsilon_{t,j}, \tag{9}$$

where $\mathbf{z}_{t,j}$ contains the explanatory variables and has dimension $n_\gamma \times 1$. We denote the predicted variances by $V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}}) \equiv \exp\left\{\hat{\boldsymbol{\gamma}}'\mathbf{z}_{t,j}\right\}$ for $j = 1, 2, ..., n_{y,t}$. Based on these estimates, we can then perform the weighted regression in (8) by replacing $Var(v_{j,t})$ with $V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})$.

The variance of the measurement errors in the observables at a given point in time can be estimated with standard methods if the conditional variance function is correctly specified, i.e. if

$$Var(y_{t,j}|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}) = \sigma^2_{t,o}V(\mathbf{z}_{t,j};\boldsymbol{\gamma}_o) \text{ for some } \boldsymbol{\gamma}_o \in \Gamma \text{ and } \sigma^2_{t,o} \text{ where } \hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_o. \tag{10}$$

In this case

$$\hat{\sigma}^2_t = \frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}} \frac{\hat{v}^2_{t,j}}{V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})}, \tag{11}$$

where $\hat{v}_{t,j} \equiv (y_{t,j} - g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1))$. Thus, the SR approach can easily handle potential time-varying second moments in the measurement errors.

The main advantage of the regression filter is its simplicity. The filter is very easy to implement, even with an unbalanced panel of observables, and the filter is also fast to calculate. When the function $\mathbf{g}(\cdot)$ is nonlinear in $\mathbf{x}_{2,t}$, fast optimizers such as the Levenberg-Marquardt method, the Gauss-Newton method or various modifications of these routines can be used to solve the optimization problem in (8). In this case, the estimated factors from the previous period can be used as good starting values for the optimization. When the function $\mathbf{g}(\cdot)$ is linear in $\mathbf{x}_{2,t}$, the problem in (8) reduces to an OLS regression which has a closed form solution.[3]

3.3 Relating the regression filter to other filtering methods

How does the regression filter compare with other filtering techniques? In order to illustrate this, we impose the standard assumptions that i) $\mathbf{v}_t$ and $\mathbf{w}_t$ are each independent, normally distributed, and have constant variances, ii) $\mathbf{x}_0$ is normally distributed, and iii) $\mathbf{x}_0$, $\mathbf{v}_t$, and $\mathbf{w}_t$ are mutually uncorrelated at all leads and lags. To reduce the notational burden in the argument below we assume, without loss of generality, that there are no observed factors, meaning that $\mathbf{x}_{1,t}$ is empty and $\mathbf{x}_{2,t} = \mathbf{x}_t$. In this setup, the logarithm of the conditional probability of $\mathbf{x}_{1:T} \equiv \{\mathbf{x}_t\}_{t=1}^T$ given $y_{1:T} \equiv \{\mathbf{y}_t\}_{t=1}^T$, denoted $\log p(\mathbf{x}_{1:T}|\,y_{1:T})$, is proportional to (see Durbin & Koopman (2001))

$$\begin{aligned}\tilde{Q} \;&=\; -\frac{1}{2}\sum_{t=1}^{T}\left[\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_1)\right]'(Var(\mathbf{v}_t))^{-1}\left[\mathbf{y}_t - \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_1)\right] \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\mathbf{w}'_t(Var(\mathbf{w}_t))^{-1}\mathbf{w}_t - \frac{1}{2}\tilde{\mathbf{x}}'_0\mathbf{P}_0^{-1}\tilde{\mathbf{x}}_0.\end{aligned} \tag{12}$$

Here, $\tilde{\mathbf{x}}_0$ denotes the estimation error of the initial values of the factors and $\mathbf{P}_0$ is the covariance matrix of $\tilde{\mathbf{x}}_0$. In the case of a fixed number of observables in each time period, we can scale (12) by $-1/n_y$ to

[3] Note also in relation to (8) that the regression filter is very easy to implement with mulitprocessing, because n CPU's can independently solve n different optimization problems.

get

$$Q = \frac{1}{2n_y} \sum_{t=1}^{T} \sum_{j=1}^{n_y} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)\right)^2}{Var\left(v_j\right)} + \frac{1}{2n_y} \left(\sum_{t=1}^{T} \mathbf{w}_t' \left(Var\left(\mathbf{w}_t\right)\right)^{-1} \mathbf{w}_t + \tilde{\mathbf{x}}_0' \mathbf{P}_0^{-1} \tilde{\mathbf{x}}_0 \right). \tag{13}$$

We then impose the standard assumption in cross-section regressions that

$$\frac{1}{2n_y} \sum_{j=1}^{n_y} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)\right)^2}{Var\left(v_j\right)} \overset{p}{\longrightarrow} E\left[\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)\right)^2}{2Var\left(v_j\right)}\right] > 0 \text{ for } n_y \longrightarrow \infty$$

for all $t = 1, 2, ..., T$. So, when n_y tends to infinity, (13) convergence to

$$Q = \sum_{t=1}^{T} E\left[\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)\right)^2}{2Var\left(v_j\right)}\right], \tag{14}$$

because the second part of the expression in (13) does not depend on the number of observables and therefore tends to zero for $n_y \longrightarrow \infty$.

The expression in (14) is exactly the expectation minimized each time period by the regression filter. Consequentially, the regression filter converges to the mode of $p\left(\mathbf{x}_{1:T} | y_{1:T}\right)$ as the number of observables increases.

The values of $\mathbf{x}_{1:T}$ which maximizes $p\left(\mathbf{x}_{1:T} | y_{1:T}\right)$ are the most probable values of the latent factors given the observables, and these factor estimates are therefore optimal (Durbin & Koopman (2001)). Accordingly, the estimated latent factors from the regression filter converges to the optimal estimates as the number of observables tends to infinity. In the case where the assumptions for the Kalman filter hold, then the Kalman smoother reports the mode of $p\left(\mathbf{x}_{1:T} | y_{1:T}\right)$. Hence, the regression filter converges to the Kalman smoother when the number of observables tends to infinity.

This simple argument also shows what the regression filter is missing to achieve efficiency, namely the smoothing of the estimated latent factors according to the transition equations. However, ignoring this smoothing can be justified if i) there are many observables available and/or ii) the observables are measured with a small amount of error.

We finally note for the sake of generality that the normality assumptions for $\mathbf{w}_t$ and $\mathbf{x}_0$ can be omitted without changing the key implication that the solution of the regression filter converges to the mode of $p\left(\mathbf{x}_{1:T} | y_{1:T}\right)$.

The SR approach is also related to estimation of the factors in standard or approximated factor model. For these models, Stock & Watson (2002) show how factors can be estimated consistently using the method of principal components when a large number of observables are available in each time period. This implies that the factor dynamics is estimated nonparametrically as in the SR approach. However, a key difference between the SR approach and the method of principal components is that the latter only works for linear models with stationary time-series. On the other hand, the SR approach can easily handle nonlinear models and non-stationary time series.

3.4 Estimation of $\boldsymbol{\theta}_1$

We suggest to estimate $\boldsymbol{\theta}_1$ by pooling all the estimated residuals from (8) and minimize the squared value of these residuals with respect to $\boldsymbol{\theta}_1$, i.e.

$$\hat{\boldsymbol{\theta}}_1 = \arg \min_{\boldsymbol{\theta}_1 \in \Theta_1} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1\right); \boldsymbol{\theta}_1\right)\right)^2}{2V\left(\mathbf{z}_{t,j}; \hat{\boldsymbol{\gamma}}\right)} \tag{15}$$

This estimator is thus very similar to the standard non-linear regression estimator. The only difference being that we need to account for the fact that changes in $\boldsymbol{\theta}_1$ affect the function $g_j\left(\cdot; \boldsymbol{\theta}_1\right)$ not only directly but also indirectly through the latent factors $\hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1\right)$. Hence, when solving the problem in (15), the latent factors need to be recomputed for different values of $\boldsymbol{\theta}_1$. As in the case of the estimator of $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$, the estimator of $\boldsymbol{\theta}_1$ also does not use transition equations of the factors and it is therefore robust to any form of factor dynamics.

It is instructive at this stage to compare our estimator in (15) with the backfitting estimator in Pastorello et al. (2003). One way to implement this backfitting estimator in our setting is to estimate the latent factors by the regression filter in (8), and subsequently let the objective function for $\boldsymbol{\theta}_1$ be the sum of squared residuals from these regressions. In the framework of Pastorello et al. (2003), the estimated factors then appear as the nuisance parameter in the estimation of $\boldsymbol{\theta}_1$ which is done by

$$\hat{\boldsymbol{\theta}}_1^{p+1} = \arg \min_{\boldsymbol{\theta}_1 \in \Theta_1} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1^{p}\right); \boldsymbol{\theta}_1\right)\right)^2}{2V\left(\mathbf{z}_{t,j}; \hat{\boldsymbol{\gamma}}\right)} \quad \text{for } p = 1, 2, ..., P. \tag{16}$$

The important difference between this estimator and the one in (15) is that the estimator in (16) only accounts for the direct effects of $\boldsymbol{\theta}_1$ on $g_j\left(\cdot; \boldsymbol{\theta}_1\right)$ when estimating $\hat{\boldsymbol{\theta}}_1^{p+1}$. That is, the indirect effects of $\boldsymbol{\theta}_1$ on the latent factors are not accounted for when finding $\hat{\boldsymbol{\theta}}_1^{p+1}$, and this makes the optimization of $\hat{\boldsymbol{\theta}}_1^{p+1}$ computationally simple. The estimate $\hat{\boldsymbol{\theta}}_1^{p+1}$ give rise to a new set of factors $\left\{\hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1^{p+1}\right)\right\}_{t=1}^{T}$ which can be substituted into (16) to find a new estimate of $\boldsymbol{\theta}_1$ which is denoted $\hat{\boldsymbol{\theta}}_1^{p+2}$. The idea behind the backfitting estimator is then to iterate this procedure until convergence is achieved, i.e. $\hat{\boldsymbol{\theta}}_1^{p+2} \simeq \hat{\boldsymbol{\theta}}_1^{p+1}$.

Pastorello et al. (2003) state conditions to ensure this convergence, and they show that this estimator is consistency and asymptotic normality given sufficient regularity conditions. In establishing these results, Pastorello et al. (2003) face the issue of nonadaptivity which means that the estimate of the latent factors (which depend on $\boldsymbol{\theta}_1$) prevents the econometrician from directly estimating $\boldsymbol{\theta}_1$ consistently. The issue is resolved by assuming that the mapping defined by (16), i.e.

$$\hat{\boldsymbol{\theta}}_1^{p+1} = f\left(\boldsymbol{\theta}_1^{p}\right),$$

has a unique fixed point at the true value of $\boldsymbol{\theta}_1$ and that it is contracting. As pointed out by Pastorello et al. (2003) and Sherman (2003), the conditions for consistency and asymptotic normality of iterated estimators are stronger than the conditions for optimization based estimators. In other words, the estimator suggested in (15) requires weaker assumptions to ensure consistency and asymptotic normality than the backfitting estimation in (16). This is because our estimation in (15) does not face the nonadaptivity problem as we account for the dependency of $\boldsymbol{\theta}_1$ on the latent factor during the optimization of $\boldsymbol{\theta}_1$.

3.5 Estimation of $\boldsymbol{\theta}_2$

As for the parameters $\boldsymbol{\theta}_2$, we suggest to estimate them based on the observed factors $\{\mathbf{x}_{1,t}\}_{t=1}^{T}$ and the estimated latent factors $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$. Here, we need to take account of the fact that $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ are generated regressors and therefore contains measurement errors. Ignoring this fact can easily bias the estimates of $\boldsymbol{\theta}_2$. As we will shown in section 4, the estimated latent factors are normally distributed if we have a sufficient number of observables, i.e. when $n_{y,t} \longrightarrow \infty$. By construction, we thus get the standard additive measurement error case (see for instance Fuller (1987))

$$\hat{\mathbf{x}}_{2,t} = \mathbf{x}_{2,t}^{o} + \mathbf{u}_t \qquad \mathbf{u}_t \sim \mathcal{N}\left(\mathbf{0}, Var\left(\mathbf{u}_t\right)\right) \text{ for } t = 1, 2, ..., T \tag{17}$$

where $\mathbf{x}_{2,t}^{o}$ denotes the true but unobserved factor value. We also derive the time series properties of $\mathbf{u}_t$ in section 4. For instance, we show that $\mathbf{u}_t$ is uncorrelated across time if the measurement errors in the observables, i.e. $\mathbf{v}_t$, do not display autocorrelation. At this point we simply note that we can estimate $\boldsymbol{\theta}_2$ consistently based on standard moment matching methods because we know the distribution of $\mathbf{u}_t$ and its statistical properties. We illustrate this important point by first considering a simple VAR(1) system, before we describe how to estimate systems with nonlinear factor dynamics.

3.5.1 Illustration for a VAR system

For simplicity in this example, let all the factors be unobserved, i.e. $\mathbf{x}_t \equiv \mathbf{x}_{2,t}$. We consider the following VAR(1) system

$$\mathbf{x}_{t+1} = \boldsymbol{\alpha} + \mathbf{h}_{\mathbf{x}}\mathbf{x}_t + \mathbf{w}_{t+1}, \tag{18}$$

where $\mathbf{w}_t \sim \mathcal{IID}\left(\mathbf{0}, Var\left(\mathbf{w}_t\right)\right)$. Hence, $\boldsymbol{\theta}_2 \equiv \left[\boldsymbol{\alpha}, \mathbf{h}_{\mathbf{x}}, vech\left(Var\left(\mathbf{w}_t\right)\right)\right]$ in this example. The system in (18) cannot be used for estimation of $\boldsymbol{\theta}_2$ because $\mathbf{x}_t$ is not observed. Instead, we use the following system

$$\hat{\mathbf{x}}_{t+1} = \boldsymbol{\alpha} + \mathbf{h}_{\mathbf{x}}\hat{\mathbf{x}}_t + \hat{\mathbf{w}}_{t+1} \tag{19}$$

based on the estimated factors, $\hat{\mathbf{x}}_t$. Notice that $\hat{\mathbf{w}}_t$ denotes the innovation in (19) using the true values of $\boldsymbol{\theta}_2$ and the estimated factors. Let us consider the following moments

$$\begin{bmatrix} E\left(\hat{\mathbf{w}}_t\right) \\ vec\left(E\left(\hat{\mathbf{w}}_t\hat{\mathbf{x}}_t'\right)\right) \\ vech\left(Var\left(\hat{\mathbf{w}}_t\right)\right) \end{bmatrix} \tag{20}$$

which identifies $\boldsymbol{\theta}_2$. Given (17)-(18), the population value of these moments can be readily computed given the maintained assumptions. To simplify this example, let $\mathbf{v}_t$ be uncorrelated across time which means that $Cov\left(\mathbf{u}_t, \mathbf{u}_{t-1}\right) = \mathbf{0}$. As shown in the appendix

$$\begin{bmatrix} E\left(\hat{\mathbf{w}}_t\right) \\ vec\left(E\left(\hat{\mathbf{w}}_t\hat{\mathbf{x}}_t'\right)\right) \\ vech\left(Var\left(\hat{\mathbf{w}}_t\right)\right) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ vec\left(-\mathbf{h}_{\mathbf{x}}Var\left(\mathbf{u}_t\right)\right) \\ vech\left(Var\left(\mathbf{w}_t\right) + Var\left(\mathbf{u}_{t+1}\right) + \mathbf{h}_{\mathbf{x}}Var\left(\mathbf{u}_t\right)\mathbf{h}_{\mathbf{x}}'\right) \end{bmatrix}. \tag{21}$$

Using these moments, $\boldsymbol{\theta}_2$ can now be estimated by Generalized Method of Moments (GMM) (Hansen (1982)) even though the factors are estimated.

3.5.2 Estimation of $\boldsymbol{\theta}_2$ in the general case

GMM can also be applied without restricting the factor dynamics to be linear. Let $\mathbf{m}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$ be a vector of moment conditions which has dimension $P \times 1$ where $P \geq L_2$ and identifies $\boldsymbol{\theta}_2$. A vector $\mathbf{r}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$ contains the corresponding population moments which can be computed based on the law of motion for the factors and the measurement errors in the estimated factors. Let the moments for the GMM estimation be

$$E\left[\mathbf{q}_t\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right] = 0, \tag{22}$$

where $\mathbf{q}_t\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) \equiv \mathbf{m}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) - \mathbf{r}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$. The GMM estimator for $\boldsymbol{\theta}_2$ is therefore given by

$$\hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}_2 \in \Theta_2} \left(\mathbf{q}_T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right)' \mathbf{W}\left(\mathbf{q}_T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right), \tag{23}$$

where $\mathbf{q}_T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{q}_t\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$ and $\mathbf{W}$ is some positive definite weighting matrix.

This estimator of $\boldsymbol{\theta}_2$ is very similar to the Implied State (IS) GMM estimator by Pan (2002). In the IS-GMM estimator, n_x factors are backed out from the same number of observables which all are assumed to be measured without errors. These factors are then used to evaluate a set of GMM moment conditions along the time series dimension leading to estimates of the parameters in the model. The estimator of $\boldsymbol{\theta}_2$ proposed in (23) is thus a generalization of the IS-GMM estimator to the case where we have more than n_x observables measured with errors.

For nonlinear systems we may not have closed form solutions for the population moments, i.e. we may not be able to find $\mathbf{r}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$ analytically. This situation often occurs for transition functions set in continuous time if no discretization scheme is applied. However, even if no analytical solution exist the value of the population moments can easily be simulated from the transition distribution of the factors and the measurement errors in the estimated factors. Hence, $\boldsymbol{\theta}_2$ can in this case be estimated by Method of Simulated Moments (SMM) (Duffie & Singleton (1993)). That is, the population moments $\mathbf{r}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$ in SMM are replaced by

$$\hat{\mathbf{r}}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) = \frac{1}{\tau T}\sum_{s=1}^{\tau T}\sum_{t=1}^{T} \mathbf{m}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}^s; \hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right), \tag{24}$$

where τT denotes the number of simulations. In order to compute (24), we suggest using a conditional simulator. That is, for each value of t we compute

$$\begin{bmatrix} \mathbf{x}_{1,t+1}^s \\ \hat{\mathbf{x}}_{2,t+1}^s \end{bmatrix} = \mathbf{h}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t} - \mathbf{u}_t^s, \mathbf{w}_{t+1}^s; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) + \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_{t+1}^s \end{bmatrix} \text{ for } s = 1, 2, ..., \tau, \tag{25}$$

where $\left\{\mathbf{w}_{t+1}^s\right\}_{s=1}^{\tau}$ and $\left\{\mathbf{u}_{t+1}^s\right\}_{s=1}^{\tau}$ are $\mathcal{IID}$ draws from their respective distributions (Carrasco & Florens (2002)).[4] The advantage of this simulator is that it allows us to take time-varying distributions

[4] For SMM we thus need to draw from the distribution of $\mathbf{w}_t$, and this might imply specifying a distribution for $\mathbf{w}_t$.

for the measurement errors into account. This is not possible when using an unconditional simulator of the form

$$\begin{bmatrix} \mathbf{x}^s_{1,t+1} \\ \hat{\mathbf{x}}^s_{2,t+1} \end{bmatrix} = \mathbf{h}\left(\mathbf{x}_{1,s}, \hat{\mathbf{x}}_{2,s} - \mathbf{u}_s, \mathbf{w}_{s+1}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) + \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_{s+1} \end{bmatrix} \text{ for } s = 1, 2, ..., \tau T \tag{26}$$

because we do not know the distribution of $\mathbf{u}_s$ for an arbitrary value of s. This unconditional simulator can only be used if it is reasonable to assume that the distributions of $\{\mathbf{u}_t\}_{t=1}^T$ are time-invariant and thus can be used to generate draws for $\mathbf{u}_s$.

3.6 An illustrative example: The one factor CIR-model

To illustrate our approach, let us return to standard one-factor CIR model. Here, $\boldsymbol{\theta}_1 \equiv \begin{bmatrix} \kappa\delta & (\kappa+\lambda) & \sigma \end{bmatrix}$ is estimated based on (15). As pointed out by Chen & Scott (1993), δ, κ, and λ cannot be identified because we do not use information from the factors' law of motion when estimating $\boldsymbol{\theta}_1$. This is seen directly from the corresponding law of motion for r_t under the risk neutral measure $\mathcal{Q}$

$$dr_t = \left(\kappa\delta - (\kappa+\lambda)\, r_t\right) dt + \sigma\sqrt{r_t} dz_t^{\mathcal{Q}}, \tag{27}$$

which is exactly characterized by the parameters in $\boldsymbol{\theta}_1$.

Now, let $\boldsymbol{\theta}_2 = \delta$. It is straightforward to verify that δ is the unconditional mean of r_t under the physical measure. Hence, the mean value of the estimated factor $\{\hat{r}_t\}_{t=1}^T$ can be used to estimate $\boldsymbol{\theta}_2$ because

$$E\left[\hat{r}_t\right] = E\left[r_t + u_t\right] = \delta,$$

where u_t is the measurement error in $\hat{r}_t$ from the regression filter. Thus, the estimator for $\boldsymbol{\theta}_2$ is simply

$$\hat{\boldsymbol{\theta}}_2 = \frac{1}{T}\sum_{t=1}^{T} \hat{r}_t.$$

Given this estimate, we can now identify the remaining parameters in CIR model by

$$\hat{\kappa} = \frac{\widehat{\kappa\delta}}{\hat{\boldsymbol{\theta}}_2}$$

$$\hat{\lambda} = \widehat{\kappa+\lambda} - \hat{\kappa}.$$

Standard errors for $\hat{\kappa}$ and $\hat{\lambda}$ can easily be derived by the Delta method based on the distributions for $\hat{\boldsymbol{\theta}}_1$.

3.7 Summarizing the SR approach

Before we turn to the asymptotic properties of the SR approach, let us for the sake of clarity briefly summarize the two steps in the SR approach. The steps are as follows:

Step 1:

- Use the regression filter to find the latent factors as a function of $\boldsymbol{\theta}_1$, i.e. $\{\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1)\}_{t=1}^T$.

- Minimize the objective function in (15) with respect to $\boldsymbol{\theta}_1$ while recomputing $\{\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1)\}_{t=1}^T$ for different values of $\boldsymbol{\theta}_1$. Denote the optimal value of $\boldsymbol{\theta}_1$ by $\hat{\boldsymbol{\theta}}_1$, and the estimated factors are then $\left\{\hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right)\right\}_{t=1}^T$.

Step 2:

- From $\hat{\boldsymbol{\theta}}_1$ and $\left\{\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right)\right\}_{t=1}^T$, estimate $\boldsymbol{\theta}_2$ by GMM or SMM.

The first step in the SR approach is similar to the procedure used when $\boldsymbol{\theta}_1$ is estimated by any other filter like the Kalman filter. That is, we run the filter to construct the objective function in the first step. The second step in the SR approach is new but for given values of $\hat{\boldsymbol{\theta}}_1$ and $\left\{\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right)\right\}_{t=1}^T$computationally and conceptually straightforward.

4 Asymptotic properties of the SR approach

This section derives the asymptotic distributions of the estimated factors and parameters in the SR approach. For $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^T$ and $\hat{\boldsymbol{\theta}}_1$, this is done for a fixed number of time periods where we let the number of observables tend to infinity in each period. For $\boldsymbol{\theta}_2$, inference is conducted by letting the number of time periods tend to infinity. Thus, the inference is done by sequentially letting the cross-sectional dimension tend to infinity and afterwards letting the time series dimension tend to infinity. This sequential inference approach is mainly motivated by the sequential structure of the SR approach. Connor & Korajczyk (1986) adopt the same approach when estimating a factor model based on the Arbitrage Pricing Theory.

All the derivations in this section are for dynamic term structure models which are uniquely identified. For estimation of $\{\mathbf{x}_{2,t}\}_{t=1}^T$ and $\boldsymbol{\theta}_1$, this means that

$$\sum_{t=1}^{T} E\left[\frac{\left(g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \boldsymbol{\theta}_1^o\right) - g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}; \boldsymbol{\theta}_1\right)\right)^2}{Var\left(v_{j,t}\right)}\right] > 0, \text{ for all } \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^o \text{ and } \mathbf{x}_{2,t} \neq \mathbf{x}_{2,t}^o \text{ for all } t. \quad (28)$$

This assumption is fairly weak but, for instance, rules out cases where $\mathbf{x}_{2,t}$ and $\boldsymbol{\theta}_1$ only enter as a product. That is, if $\mathbf{x}_{2,t} \equiv \left[\begin{array}{cc} x_{2,t}(1) & x_{2,t}(2) \end{array}\right]$ and $\boldsymbol{\theta}_1 \equiv \left[\begin{array}{cc} \theta_1(1) & \theta_1(2) \end{array}\right]$ and if $g = \sum_{i=1}^2 \theta_1(i)\, x_{2,t}(i)$, then we cannot identify $\mathbf{x}_{2,t}$ and $\boldsymbol{\theta}_1$. Linear factor models have the same problem, but usually impose additional assumptions on $\boldsymbol{\theta}_1$ to ensure identification (see for instance Stock & Watson (2002))

For estimation of $\boldsymbol{\theta}_2$ we assume that moments for GMM or SMM that uniquely identifies $\boldsymbol{\theta}_2$ exist.

We begin by deriving the asymptotic properties of $\hat{\boldsymbol{\theta}}_1$ and $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^T$ in section 4.1. The time series properties of the measurement errors in the estimated factors are derived in section 4.2. Asymptotic properties of $\hat{\boldsymbol{\theta}}_2$ are presented in section 4.3.

4.1 Results for $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ and $\hat{\boldsymbol{\theta}}_1$

We begin with the following important proposition:

Proposition 1 *The estimators of* $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$ *in (8) and* $\boldsymbol{\theta}_1$ *in (15) are equivalent to joint estimation of* $\left(\boldsymbol{\theta}_1, \{\mathbf{x}_{2,t}\}_{t=1}^{T}\right)$ *from*

$$Q^{joint} = \min_{\left(\boldsymbol{\theta}_1, \{\mathbf{x}_{2,t}\}_{t=1}^{T}\right)} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \boldsymbol{\theta}_1\right)\right)^2}{2V\left(z_{t,j}; \hat{\gamma}\right)}. \tag{29}$$

Proposition 1 states that the procedure described in the previous section to estimate $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$ from the regression filter and $\boldsymbol{\theta}_1$ from the output of this filter, is equivalent to a joint estimation of $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$ and $\boldsymbol{\theta}_1$. The SR approach can therefore be considered as a convenient numerical optimization procedure for optimizing a high dimensional objective function in terms of the latent factors and the parameters.

Proposition 1 also implies that consistency and normally of $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ and $\hat{\boldsymbol{\theta}}_1$ are standard and follow from properties of the M-estimator. We state conditions for consistency of $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ and $\hat{\boldsymbol{\theta}}_1$ in the next proposition

Proposition 2 ***Consistency of*** $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ ***and*** $\hat{\boldsymbol{\theta}}_1$

Impose the conditions for the uniform weak law of large numbers (UWLLN) to hold for Q^{joint} *with respect to* $\left(\boldsymbol{\theta}_1, \{\mathbf{x}_{2,t}\}_{t=1}^{T}\right)$ *as we let the number of observables tend to infinity in each period. Assumptions (3) and (28) ensure consistency of* $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ *and* $\hat{\boldsymbol{\theta}}_1$ *as* $n_{y,t} \longrightarrow \infty$ *for all* t.

Proposition 2 implies that the regression filter results in the desired value of the latent factors when there is a sufficient number of observables. Note that a corresponding consistency result for $\{\mathbf{x}_{2,t}\}_{t=1}^{T}$ as the number of time periods tend to infinity cannot be derived, because then an infinite number of factors would have to be estimated.

It is numerical challenging to apply the standard results for the M-estimator to derive the asymptotic distributions of $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ and $\hat{\boldsymbol{\theta}}_1$ jointly. For instance, in a term structure model with three latent factors and $T = 500$, the dimension of the asymptotic covariance matrix would exceed 1500. When computing the asymptotic covariance matrix, we therefore find it numerically more convenient to exploit the sparsity of this covariance matrix which is due to independence among $\hat{\mathbf{x}}_{2,1}$, $\hat{\mathbf{x}}_{2,2}$, ... ,$\hat{\mathbf{x}}_{2,T}$. This is done by first deriving the covariance matrix of $\hat{\boldsymbol{\theta}}_1$. The covariance matrices of $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ are then derived afterwards with $\hat{\boldsymbol{\theta}}_1$ as a nuisance parameter. We therefore start by deriving the asymptotic distributions of $\hat{\boldsymbol{\theta}}_1$.

4.1.1 The asymptotic distribution of $\hat{\boldsymbol{\theta}}_1$

In order to derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_1$, we start by stacking the data. For this purpose define the new index

$$\mathcal{I} = \left\{ \{j\}_{j=1}^{n_{y,1}}, \{j\}_{j=1}^{n_{y,2}}, ..., \{j\}_{j=1}^{n_{y,T}} \right\},$$

which has $N \equiv \sum_{t=1}^{T} n_{y,t}$ elements. Similarly, we define

$$\mathbf{v} \equiv \left\{ \{v_{1,j}\}_{j=1}^{n_{y,1}}, \{v_{2,j}\}_{j=1}^{n_{y,2}}, ..., \{v_{T,j}\}_{j=1}^{n_{y,T}} \right\}$$

$$\mathbf{z} \equiv \left\{ \{z_{1,j}\}_{j=1}^{n_{y,2}}, \{z_{2,j}\}_{j=1}^{n_{y,2}}, ..., \{z_{T,j}\}_{j=1}^{n_{y,T}} \right\}$$

$$\mathbf{y} \equiv \left\{ \{y_{1,j}\}_{j=1}^{n_{y,1}}, \{y_{2,j}\}_{j=1}^{n_{y,2}}, ..., \{y_{T,j}\}_{j=1}^{n_{y,T}} \right\}$$

$$\mathbf{x} \equiv \left\{ \{\mathbf{x}_{1}\}_{j=1}^{n_{y,1}}, \{\mathbf{x}_{2}\}_{j=1}^{n_{y,2}}, ..., \{\mathbf{x}_{T}\}_{j=1}^{n_{y,T}} \right\}$$

and refer to elements in these sets by v_i, $\mathbf{z}_i$, $\mathbf{y}_i$, and $\mathbf{x}_i$ for $i \in \mathcal{I}$, respectively. The asymptotic distribution of $\boldsymbol{\theta}_1$ then follows from a mean value expansion of the score function in (15). Imposing standard regularity conditions and $N \longrightarrow \infty$

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Var\left(\hat{\boldsymbol{\theta}}_1\right)\right), \tag{30}$$

where

$$Var\left(\hat{\boldsymbol{\theta}}_1\right) = \left(\mathbf{A}_o^{\boldsymbol{\theta}_1}\right)^{-1} \mathbf{B}_o^{\boldsymbol{\theta}_1} \left(\mathbf{A}_o^{\boldsymbol{\theta}_1}\right)^{-1}. \tag{31}$$

Note that the distribution in (30) is derived for by requiring that $n_{y,t} \rightarrow \infty$ in each time period, implying that $N \rightarrow \infty$. This means that that N often will be very large. For instance, with 100 observables in 100 time periods we have $N = 10.000$. Notice also that uncertainty from estimation of $\boldsymbol{\gamma}$ does not affect the asymptotic distribution of $\hat{\boldsymbol{\theta}}_1$. This is a well-known result from weighted nonlinear regression analysis with purely observed regressors (see for instance Wooldridge (2002)).

Before we present estimators of $Var\left(\hat{\boldsymbol{\theta}}_1\right)$, consider the expression for the score function in (15)

$$\begin{aligned} \mathbf{s}^{\boldsymbol{\theta}_1} \quad = \quad & -\sum_{i=1}^{N} \frac{\left(y_i - g_i\left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}\left(\boldsymbol{\theta}_1\right); \boldsymbol{\theta}_1\right)\right)}{V\left(z_i; \gamma\right)} \frac{\partial \mathbf{x}_{2,i}'\left(\boldsymbol{\theta}_1\right)}{\partial \boldsymbol{\theta}_1} \frac{\partial g_i\left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}\left(\boldsymbol{\theta}_1\right); \boldsymbol{\theta}_1\right)}{\partial \mathbf{x}_{2,i}\left(\boldsymbol{\theta}_1\right)} \\ & -\sum_{i=1}^{N} \frac{\left(y_i - g_i\left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}\left(\boldsymbol{\theta}_1\right); \boldsymbol{\theta}_1\right)\right)}{V\left(z_i; \gamma\right)} \frac{\partial g_i\left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}\left(\boldsymbol{\theta}_1\right); \boldsymbol{\theta}_1\right)}{\partial \boldsymbol{\theta}_1}. \end{aligned} \tag{32}$$

The first double sum in (32) captures the indirect effect from changes in $\boldsymbol{\theta}_1$ that leads to changes in the latent factors $\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1\right)$ which in turn leads to changes in the model implied observables, $\partial g_j / \partial \mathbf{x}_{2,t}$. The second double sum in (32) captures the direct effect of changes in $\boldsymbol{\theta}_1$ on the model implied observables, $\partial g_j / \partial \boldsymbol{\theta}_1$. Imposing conditions for uniform convergence, we therefore suggest the following heteroscedastic-robust estimator of the variance in the score function

$$\hat{\mathbf{B}}^{\boldsymbol{\theta}_1} = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{v}_i^2}{V\left(z_i; \hat{\gamma}\right)^2} [\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'}{\partial \hat{\boldsymbol{\theta}}_1} \frac{\partial \hat{g}_i}{\partial \hat{\mathbf{x}}_{2,i}}\right) \left(\frac{\partial \hat{\mathbf{x}}_{2,i}'}{\partial \hat{\boldsymbol{\theta}}_1} \frac{\partial \hat{g}_i}{\partial \hat{\mathbf{x}}_{2,i}}\right)' + 2 \left(\frac{\partial \hat{g}_i}{\partial \hat{\boldsymbol{\theta}}_1}\right) \left(\frac{\partial \hat{\mathbf{x}}_{2,i}'}{\partial \hat{\boldsymbol{\theta}}_1} \frac{\partial \hat{g}_i}{\partial \hat{\mathbf{x}}_{2,i}}\right)' + \frac{\partial \hat{g}_i}{\partial \hat{\boldsymbol{\theta}}_1} \left(\frac{\partial \hat{g}_i}{\partial \hat{\boldsymbol{\theta}}_1}\right)'], \tag{33}$$

where $\hat{g}_i \equiv g_i\left(\mathbf{x}_{1,i}, \hat{\mathbf{x}}_{2,i}; \hat{\boldsymbol{\theta}}_1\right)$. A more efficient estimator of $\mathbf{B}_o^{\boldsymbol{\theta}_1}$ can be constructed if the conditional variance function is correctly specified, i.e. if (10) holds. In this case, the variance of the score function can be estimated by

$$\hat{\mathbf{B}}^{\boldsymbol{\theta}_1}_{\text{hom}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2}{V\left(z_i;\hat{\gamma}\right)}[\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial \hat{g}_i}{\partial \hat{\mathbf{x}}_{2,i}}\right)\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial \hat{g}_i}{\partial \hat{\mathbf{x}}_{2,i}}\right)' + 2\left(\frac{\partial \hat{g}_i}{\partial \hat{\boldsymbol{\theta}}_1}\right)\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial \hat{g}_i}{\partial \hat{\mathbf{x}}_{2,i}}\right)' + \frac{\partial \hat{g}_i}{\partial \hat{\boldsymbol{\theta}}_1}\left(\frac{\partial \hat{g}_i}{\partial \hat{\boldsymbol{\theta}}_1}\right)'], \tag{34}$$

where

$$\hat{\sigma}_i^2 \equiv \left\{\left\{\hat{\sigma}_1^2\right\}_{j=1}^{n_{y,1}}, \left\{\hat{\sigma}_2^2\right\}_{j=1}^{n_{y,2}}, ..., \left\{\hat{\sigma}_T^2\right\}_{j=1}^{n_{y,T}}\right\}.$$

Here, $\hat{\sigma}_t^2$ is estimated based on (11) for $t = 1, 2, ..., T$. Given standard regularity conditions, these estimates are consistent and asymptotically normal. The asymptotic distribution is given by

$$\sqrt{n_{y,t}}\left(\hat{\sigma}_t^2 - \sigma_t^2\right) \xrightarrow{d} \mathcal{N}\left(0, E\left[\left(\frac{v_{j,t}}{\sqrt{V\left(z_{t,j};\boldsymbol{\gamma}^o\right)}}\right)^4\right] - \sigma_t^4\right) \tag{35}$$

for $n_{y,t} \to \infty$ for $t = 1, 2, ..., T$. The asymptotic variance of $\hat{\sigma}_t^2$ can be estimated by

$$\widehat{Var}\left(\hat{\sigma}_t^2\right) = \frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\left(\frac{\hat{v}_{j,t}}{\sqrt{V\left(z_{t,j};\hat{\boldsymbol{\gamma}}\right)}}\right)^4 - \left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{\hat{v}_{j,t}}{\sqrt{V\left(z_{t,j};\hat{\boldsymbol{\gamma}}\right)}}\right)^4. \tag{36}$$

Hence, the standard result for NLS also holds in our case even though $\boldsymbol{\theta}_1$ is estimated from $\{y_i\}_{i=1}^N$ whereas $\hat{\sigma}_t^2$ is estimated from $\{y_{t,j}\}_{j=1}^{n_{y,t}}$. This difference is asymptotically unimportant because $n_{y,t}$ tending to infinity for just one value of t, is sufficient to make $N = \sum_{t=1}^{T} n_{y,t}$ tend to infinity. The asymptotic distribution of $\boldsymbol{\theta}_1$ in (30) can therefore be used to derive the asymptotic properties of $\hat{\sigma}_t^2$ using standard methods.

Equation (32) shows that the Hessian matrix contains second-order derivatives of $\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1\right)$ with respect to $\boldsymbol{\theta}_1$. Since these derivatives do not have a closed form solution and must be computed using numerical procedures, we simply suggest to estimate the Hessian matrix based on numerical derivatives. One possibility is to approximate the Hessian matrix directly from the second order numerical derivatives. Another possibility is to use first order numerical derivatives of the score function in (32). In both cases, we end up with the following estimate of the Hessian matrix

$$\hat{\mathbf{A}}^{\boldsymbol{\theta}_1} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{H}}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,i}, \hat{\mathbf{x}}_{2,i}, z_i; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\gamma}}\right). \tag{37}$$

4.1.2 The asymptotic distribution of $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^T$

Let $\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_t, \mathbf{z}_{t,j}; \theta_1, \boldsymbol{\gamma}\right)$ and $\mathbf{H}^{\mathbf{x}_2}(\mathbf{x}_t, z_{t,j}\ ; \boldsymbol{\theta}_1, \boldsymbol{\gamma})$ be the score function and the Hessian matrix of the objective function in (8), respectively. Under standard regularity conditions and $n_{y,t} \longrightarrow \infty$ it holds that

$$\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Var\left(\hat{\mathbf{x}}_{2,t}\right)\right) \tag{38}$$

where

$$Var\left(\hat{\mathbf{x}}_{2,t}\right)=\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\mathbf{B}_{t,o}^{\mathbf{x}_2}\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}+\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\mathbf{D}_{t,o}Var\left(\hat{\boldsymbol{\theta}}_1\right)\mathbf{D}_{t,o}'\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}. \tag{39}$$

Here, $\mathbf{B}_{t,o}^{\mathbf{x}_2}\equiv Var\left[\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$ is the variance of the score function and $\mathbf{D}_{t,o}\equiv E\left[\frac{\partial\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)}{\partial\left(\boldsymbol{\theta}_1^o\right)'}\right]$ is the Jacobian with dimension $n_{x_2}\times L_1$. The matrix $\mathbf{A}_{t,o}^{\mathbf{x}_2}\equiv E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$ is the expected value of the Hessian matrix with respect to $\mathbf{x}_{2,t}$. Note that all these expectations are evaluated for the cross-section dimension of the set of observables at time t. Finally, $\boldsymbol{\gamma}^*$ is the limiting value of $\hat{\boldsymbol{\gamma}}$, i.e. $\hat{\boldsymbol{\gamma}}\xrightarrow{p}\boldsymbol{\gamma}^*$, where $\boldsymbol{\gamma}^*$ does not need to be the true value of $\boldsymbol{\gamma}$ (see Wooldridge (2002) for further details).

In relation to the asymptotic variance of $\hat{\mathbf{x}}_{2,t}$, the first term in (38) represents the usual uncertainty when applying regression analysis. The second term in (38) is non-standard and represents the additional uncertainty in the factor estimates due to estimation of $\boldsymbol{\theta}_1$. Note also that the uncertainty of estimating $\boldsymbol{\gamma}$ does not appear in the expression for $Var\left(\hat{\mathbf{x}}_{2,t}\right)$.

Imposing conditions for uniform convergence, we suggest the following heteroscedastic-robust estimators

$$\hat{\mathbf{B}}_t^{\mathbf{x}_2}=\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{\hat{v}_{t,j}^2}{\left(V\left(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}}\right)\right)^2}\left(\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\right)\left(\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\right)' \tag{40}$$

$$\hat{\mathbf{A}}_t^{\mathbf{x}_2}=\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V\left(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}}\right)}\left(\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\right)\left(\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\right)' \tag{41}$$

$$\hat{\mathbf{D}}_t=\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*\right)}\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\left(\frac{\partial\hat{\mathbf{x}}_{2,t}'}{\partial\hat{\boldsymbol{\theta}}_1}\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}+\frac{\partial\hat{g}_j}{\partial\hat{\boldsymbol{\theta}}_1}\right)'. \tag{42}$$

where $\hat{g}_j\equiv g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)$.

Further efficiency can be gained in the estimate of $Var\left(\hat{\mathbf{x}}_{2,t}\right)$ if the conditional variance function for the observables is correctly specified, i.e. if (10) holds. Given this assumption, the variance of the score function can be estimated by

$$\hat{\mathbf{B}}_{t,\mathrm{hom}}^{\mathbf{x}_2}=\frac{\hat{\sigma}_t^2}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V\left(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}}\right)}\left(\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\right)\left(\frac{\partial\hat{g}_j}{\partial\hat{\mathbf{x}}_{2,t}}\right)', \tag{43}$$

and the expression of $Var\left(\hat{\mathbf{x}}_{2,t}\right)$ reduces to

$$Var\left(\hat{\mathbf{x}}_{2,t}\right)=\hat{\sigma}_t^2\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}+\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\mathbf{D}_{t,o}Var\left(\hat{\boldsymbol{\theta}}_1\right)\mathbf{D}_{t,o}'\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}. \tag{44}$$

4.2 The time series properties of $\mathbf{u}_t$

The time series properties of the measurement errors in the estimated factors can be derived from the asymptotic linearity of $\hat{\mathbf{x}}_{2,t}$. As shown in the appendix,

$$\mathbf{u}_t=\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\mathbf{Z}_t\mathbf{v}_t \tag{45}$$

when $n_{y,t}$ is sufficiently large. Here

$$\mathbf{Z}_t \equiv \frac{-1}{n_{y,t}} \left[\left\{ \frac{\partial g_j \left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \hat{\boldsymbol{\theta}}_1 \right)}{\partial \mathbf{x}_{2,t}^o} \frac{1}{V\left(\mathbf{z}_{t,j}; \boldsymbol{\gamma}^*\right)} \right\}_{j=1}^{n_{y,t}} \right] \tag{46}$$

has dimension $n_{x_2} \times n_{y,t}$ and

$$\mathbf{v}_t \equiv \begin{bmatrix} v_{t,1} \\ v_{t,2} \\ \dots \\ v_{t,n_{y,t}} \end{bmatrix}.$$

The conditional autocorrelation in $\mathbf{u}_t$ at time t is therefore

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) \mathbf{Z}'_{t-k} \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1} \tag{47}$$

where $E\left[\left|\mathbf{v}_t \mathbf{v}'_{t-k}\, \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right.\right] \equiv \boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$ which may be time-varying from variation in $\mathbf{x}_{1,t}$ and/or $\mathbf{z}_{t,j}$. This implies that the conditional autocorrelation in $\mathbf{u}_t$ may be time-varying due to variation in $\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t$ across time and/or due to variation in $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$ across time. Note also that if the measurement errors in the observables do not display autocorrelation, i.e. $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) = \mathbf{0}$, then $\mathbf{u}_t$ does not display any autocorrelation as argued in section 3.5.

If we impose a homoscedasticity assumption on the second moments for the measurement errors in the observables, i.e. $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) = \boldsymbol{\Omega}_{\mathbf{vv},-k}$, then the conditional autocorrelation in $\mathbf{u}_t$ simplifies to

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \boldsymbol{\Omega}_{\mathbf{vv},-k} \mathbf{Z}'_{t-k} \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1} \tag{48}$$

Note that this conditional autocorrelation may still be time-varying if there is time-variation in $\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t$. Finally, the unconditional autocorrelation has the following expression

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k}\right) = E\left[\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \boldsymbol{\Omega}_{\mathbf{vv},-k} \mathbf{Z}'_{t-k} \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1}\right], \tag{49}$$

given $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) = \boldsymbol{\Omega}_{\mathbf{vv},-k}$.

Imposing the standard regularity conditions, the most general version of the conditional autocorrelation in (47) can be estimated consistently by

$$\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\hat{\mathbf{A}}_t^{\mathbf{x}_2}\right)^{-1} \hat{\mathbf{Z}}_t \hat{\boldsymbol{\Omega}}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) \hat{\mathbf{Z}}'_{t-k} \left(\hat{\mathbf{A}}_{t-k}^{\mathbf{x}_2}\right)^{-1}, \tag{50}$$

where $\hat{\mathbf{A}}_t^{\mathbf{x}_2}$ in (41) and $\hat{\mathbf{Z}}_t = \frac{1}{n_{y,t}} \left[\left\{ \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \frac{1}{V\left(\mathbf{z}_{t,j}; \hat{\boldsymbol{\gamma}}\right)} \right\}_{j=1}^{n_{y,t}} \right]$ are consistent estimators of $\mathbf{A}_{t,o}^{\mathbf{x}_2}$ and $\mathbf{Z}_t$, respectively. It is more difficult to get a consistent estimator of $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$, and it will in general be necessary to impose some structure on $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$ due to its large dimension. One solution may be to use a multivariate GARCH model, and yields or bonds within certain maturities ranges such as 0-2 years, 2-4 years, etc. can be assumed to have the same properties.[5]

[5] See Bauwens, Laurent & Rombouts (2006) for a survey of multivariate GARCH models.

The estimation of the conditional autocorrelation greatly simplifies if it is reasonable to impose a homoscedasticity assumption on the covariance between $\mathbf{v}_t$ and $\mathbf{v}_{t-k}$. The empirical support for this assumption can be examined using tests for multivariate GARCH effects in $\mathbf{v}_t$ (see Bauwens et al. (2006)). Given the homoscedasticity assumption of $\boldsymbol{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$, the conditional autocorrelation in (48) can be estimated consistently by

$$\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\hat{\mathbf{A}}_t^{\mathbf{x}_2}\right)^{-1} \hat{\mathbf{Z}}_t \hat{\boldsymbol{\Omega}}_{\mathbf{vv},-k} \hat{\mathbf{Z}}'_{t-k} \left(\hat{\mathbf{A}}_{t-k}^{\mathbf{x}_2}\right)^{-1}, \tag{51}$$

where $\hat{\boldsymbol{\Omega}}_{\mathbf{vv},-k} = \frac{1}{T-k}\sum_{t=k+1}^{T} \hat{\mathbf{v}}_t \hat{\mathbf{v}}'_{t-k}$.

Given $E\left[\mathbf{v}_t \mathbf{v}'_{t-k}\right] \equiv \boldsymbol{\Omega}_{\mathbf{vv},-k}$, the unconditional autocorrelation can be estimated by

$$\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t-k}\right) = \widehat{(\mathbf{A}^{\mathbf{x}_2})^{-1}\mathbf{Z}} \hat{\boldsymbol{\Omega}}_{\mathbf{vv},-k} \widehat{(\mathbf{A}^{\mathbf{x}_2})^{-1}\mathbf{Z}}', \tag{52}$$

where $\widehat{(\mathbf{A}^{\mathbf{x}_2})^{-1}\mathbf{Z}} = \frac{1}{T-k}\sum_{t=k+1}^{T}\left(\hat{\mathbf{A}}_t^{\mathbf{x}_2}\right)^{-1}\hat{\mathbf{Z}}_t$.

Finally, we note that empirical support for the independence assumption between $\mathbf{v}_t$ and $\mathbf{w}_{t-k}$ can be tested in a standard manner using $\hat{\mathbf{v}}_t$ and $\widehat{\mathbf{w}}_t$. Here, $\widehat{\mathbf{w}}_t$ is the innovation to the factors using the estimated factors and the estimated parameters in the model, i.e.

$$\hat{\mathbf{x}}_{t+1} = \mathbf{h}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}, \widehat{\mathbf{w}}_{t+1}; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2\right).$$

4.3 Results for $\hat{\boldsymbol{\theta}}_2$

When deriving the asymptotic properties of $\hat{\boldsymbol{\theta}}_2$, we must deal with two non-standard features. Firstly, measurement errors are present in the estimated latent factors. Fortunately, the SR approach provides an estimate of these errors, and it is therefore straightforward to correct for the measurement errors when setting up the moment condition as we showed in section 3.5. In other words, the first non-standard feature is dealt with in the moment conditions.

The second non-standard feature is the presence of an estimated value of $\boldsymbol{\theta}_1$ in the moment conditions instead of the true value of the parameter. Fortunately, it turns out that we do not need to correct for the fact that $\boldsymbol{\theta}_1$ is estimated. This is because inference for $\boldsymbol{\theta}_2$ is undertaken in the time-series dimension ($T \longrightarrow \infty$) whereas inference for $\boldsymbol{\theta}_1$ is carried out in the cross-section dimension ($n_{y_t} \to \infty$ for all t implying $N \equiv \sum_{t=1}^{T} n_{y_t} \to \infty$). Hence, when T tends to infinity, N tends faster to infinity, and as a result, $\boldsymbol{\theta}_1$ is estimated superconsistently. This means that we can treat $\hat{\boldsymbol{\theta}}_1$ as known when we derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_2$.

As a result, the conditions stated in Hansen (1982) and Duffie & Singleton (1993) for consistency and asymptotic normally of GMM and SMM, respectively, also apply in our case. For completeness, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_2$ are stated below.

When $\boldsymbol{\theta}_2$ is estimated by GMM and T tends to infinity

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Var\left(\hat{\boldsymbol{\theta}}_2\right)\right). \tag{53}$$

For the optimal weighting matrix, i.e. $\mathbf{W} = [Var(\mathbf{q}_t(\boldsymbol{\theta}_1^o, \boldsymbol{\theta}_2^o))]^{-1}$, it holds that

$$Var\left(\hat{\boldsymbol{\theta}}_2\right) = \left[\left(\mathbf{Q}^o_{\boldsymbol{\theta}_2}\right)' \mathbf{W}\mathbf{Q}^o_{\boldsymbol{\theta}_2}\right]^{-1}, \tag{54}$$

where $\mathbf{Q}^o_{\boldsymbol{\theta}_2} \equiv E\left[\frac{\partial \mathbf{q}_t(\boldsymbol{\theta}_1^o, \boldsymbol{\theta}_2^o)}{\partial (\boldsymbol{\theta}_2^o)'}\right]$ and has dimension $P \times L_2$.

When $\boldsymbol{\theta}_2$ is estimated by SMM and T tends to infinity

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Var\left(\hat{\boldsymbol{\theta}}_2\right)\right).$$

For the optimal weighting matrix is used, i.e. $\mathbf{W} = [Var(\mathbf{q}_t(\boldsymbol{\theta}_1^o, \boldsymbol{\theta}_2^o))]^{-1}$, then

$$Var\left(\hat{\boldsymbol{\theta}}_2\right) = \left(1 + \frac{1}{\tau}\right)\left[\left(\mathbf{Q}^o_{\boldsymbol{\theta}_2}\right)' [Var(\mathbf{q}_t(\boldsymbol{\theta}_1^o, \boldsymbol{\theta}_2^o))]^{-1} \mathbf{Q}^o_{\boldsymbol{\theta}_2}\right]^{-1}.$$

Note that these results for GMM and SMM only apply for stationary and ergodic processes. Thus, if data series are non-stationary, these series must be transformed to become stationary and moments must be set up based on the transformed series.[6]

5 A Monte Carlo study

This section studies the finite sample properties of the SR approach and compares it with the standard ML approach. We begin by outlining the study design for the Monte Carlo study in section 5.1. In section 5.2 we compare the precision of the regression filter with the optimal estimator. The finite sample distributions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are examined in sections 5.3 and 5.4, respectively.

5.1 The study design

Throughout the Monte Carlo study we focus on a linear and Gaussian dynamic term structure model because its likelihood function can be evaluated by the Kalman filter, and the optimal estimator for the latent factors is given by the Kalman smoother. In particular, we choose to consider the three factor model by Diebold, Rudebusch & Aruoba (2006) which is a dynamic interpretation of the static yield curve model by Nelson & Siegel (1987). We therefore refer to the model by Diebold et al. (2006) as the dynamic Nelson-Siegel model. Two considerations motivate our choice of model. Firstly, the dynamic Nelson-Siegel model only has one element in the parameter vector $\boldsymbol{\theta}_1$, and an extensive comparison between the properties of the SR estimator with respect to $\boldsymbol{\theta}_1$ and the corresponding properties of the ML estimator can therefore be undertaken. Such a comparison would be difficult to do in a model with many elements in $\boldsymbol{\theta}_1$because in this case the ML estimator would be numerical challenging to compute with many observables. Secondly, the dynamic Nelson-Siegel model provides a hard test of the second step in the SR approach because $\boldsymbol{\theta}_2$ contains many parameters.

[6] The case with deterministic trends is an exception (see for instance Hamilton (1994)).

In the dynamic Nelson-Siegel model the interest rate at time t with maturity τ is given by

$$y_t(\tau) = x_{1,t} + x_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + v_t(\tau), \tag{55}$$

where $\mathbf{v}_t \sim \mathcal{NID}(0, Var(\mathbf{v}_t))$. A VAR(1) model is used for the factors, i.e.

$$\mathbf{x}_{t+1} = \boldsymbol{\alpha} + \mathbf{h}_x \mathbf{x}_t + \mathbf{w}_{t+1}, \tag{56}$$

where $\mathbf{w}_t \sim \mathcal{NID}(0, Var(\mathbf{w}_t))$. The latent factors $\mathbf{x}_t \equiv \begin{bmatrix} x_{1,t} & x_{2,t} & x_{3,t} \end{bmatrix}'$ determine the level, slope, and curvature of the yield curve, respectively. Diebold et al. (2006) estimate the model on monthly US data from January 1972 to December 2000 using the 10 year yield curve. We use their estimates of λ, $\boldsymbol{\alpha}$, $\mathbf{h}_x$, and $Var(\mathbf{w}_t)$ in the Monte Carlo study.[7]

The simulated time series of bond prices are obtained from simulated values of $\mathbf{x}_t$ from (56) and adding measurement errors $v_t(\tau)$ to the value of $y_t(\tau)$ implied by $\mathbf{x}_t$. The corresponding log-transformed zero-coupon bond prices are then given by $\ln P_t(\tau) = -y_t(\tau)\,\tau$. Note that this simulation procedure induces larger measurement errors in bonds with long maturities than in bonds with short maturities. To keep the simulation study as simple as possible, we take the structure of this heteroscedasticity to be known. This implies that the subsequent results based on bonds are equivalent to using interest rates $y_t(\tau)$ directly.

To make the study design as realistic as possible, we allow the maturities of these zero-coupon bonds to vary between time periods. That is, at one point in time we may have zero-coupon bonds with maturities (5, 20, 60, 80, 120), whereas in the next time period we may have zero-coupon bonds with maturities (6, 10, 50, 80, 100). The specific maturities available at a given point in time are derivied by partitioning the 10 year yield curve into three equally sized segments according to maturity, and then sampling randomly from each of these segments. This sampling procedure ensures that we always have bonds with short, medium, and long maturities which is the case in empirical data.

The length of the simulated time series is set to 480 periods, corresponding to 40 years of monthly data. As for the number of bonds in each time period, we examine the performance of the Kalman filter by starting with a minimum of 5 bonds and then gradually increasing this number. For the SR approach, we start with a minimum of 10 bonds.

We consider two scenarios in this stimulation study. In the first scenario (Case 1), all bond prices are generated from interest rates where measurement errors have a standard deviation of 10 basis points. In the second scenario (Case 2), all interest rates have measurement errors with a standard deviation of 20 basis points. Given that the average measurement errors for interest rates in Diebold et al. (2006) have a standard deviation of 10.5 basis points, we consider Case 1 the most realistic scenario.

5.2 Factor estimation

This section examines how fast the regression filter converges to the optimal estimates as given by the Kalman smoother. The root mean squared errors (RMSE) for the regression filter and the Kalman smoother are shown in Figure 1 for each of the three factors. The red lines with a circle denote the estimates from the SR approach, and the black lines with a star denotes the estimates from the ML

[7] These estimates are reported in the appendix.

approach. 500 repetitions are used to generate each of the point estimates of the RMSE for the various number of bonds.

< Figure 1 about here >

Starting with Case 1, we see that as the number of bonds increases the regression filter converges relatively fast to the Kalman smoother. The performance of the regression filter based on just 25 bonds is close to that of the Kalman smoother, and the RMSE of the regression filter is basically identical to the RMSE of the Kalman smoother with 50 bonds.

It is even more interesting to compare the RMSE from the Kalman smoother based on 5-10 bonds with the RMSE from the regression filter based on 50-100 bonds. Such a comparison shows that the regression filter clearly outperforms the Kalman smoother, and that the gain in precision corresponds approximately to a 50% reduction in the RMSE. Hence, factor estimation in dynamic term structure models using many observables and a non-optimal estimator is clearly prefered to using just 5-10 observables *and* the optimal estimator. The latter, of course, is the current common practice in the literature.

Turning to Case 2 where all bond prices are measured less precisely than in Case 1, convergence of the regression filter to the Kalman smoother is slower than in Case 1. Now, about 100 bonds are needed for convergence of the regression filter to the Kalman smoother. Hence, the speed of convergence for the regression filter to the optimal estimates is faster when bonds are measured more precisely as argued in section 2.2.

The average number of seconds it takes to evaluate the regression filter and the Kalman filter is displayed in Figure 2.[8] Here, we choose to report the performance of the Kalman filter and not the Kalman smoother because the likelihood function is evaluated by the former.[9] The regression filter is in all cases much faster to compute than the Kalman filter. We also note that the computational gains of using the regression filter increases rapidly with 50 or more observables. This is shown by the bottom graph in Figure 2, which displays the ratio of the computing time for the Kalman filter to the computing time for the regression filter.

< Figure 2 about here >

The results reported in Figure 2 are calculated based on the standard algorithm for computing the Kalman filter (see Durbin & Koopman (2001)). However, Jungbacker & Koopman (2008) have recently shown that the computational time of a Kalman filter with a large number of observables can be substantially reduced by introducing a transformation of the state space system. Thus, the steep increase in cumputational time for the Kalman filter can probably be eliminated if the method by Jungbacker & Koopman (2008) is applied in our case.

For non-linear and/or non-Gaussian state space models, we conjecture that the computational gains associated with using the regression filter instead of importance sampling to approximate the likelihood function would be even greater than the results shown in Figure 2. This is partly because nonlinear regressions problems are very fast to solve and the estimated factors from the previous period operate as good starting values. Furthermore, the computational requirements for importance sampling increase rapidly as the number of observables increases each period.

[8] This comparision was done in Matlab on a standard desktop machine.

[9] If the likelihood function is optimized by the EM-algorithm, then it might be reasonable to report the computing time for the Kalman smoother which is higher than the computing time for the Kalman filter.

This simulation study therefore leads to the conclusion that the regression filter converges relatively fast to the optimal estimator. In the most realistic case where all bond prices are generated from interest rates with measurement errors having a standard deviation of 10 basis points (Case 1), just 50 observables are needed. We also conclude that the regression filter with just 50 observables actually outperforms a Kalman smoother based on 5-10 observables. We therefore argue that factor estimation in dynamic term structure models could benefit from using many observables instead of focusing only on how to compute the optimal estimator based on 5-10 observables.

5.3 Parameter estimation

This section examines finite sample properties when estimating parameters in the dynamic Nelson-Siegel model using the SR approach and the ML approach respectively. As mentioned earlier, the parameter λ can be identified in this model from the set of measurement equations. That is, we let $\boldsymbol{\theta}_1 = [\lambda]$ in the first estimation step of the SR approach. The remaining parameters in the dynamic Nelson-Siegel model, i.e. $[\boldsymbol{\alpha}, \mathbf{h}_x, vech\left(Var\left(\mathbf{w}_t\right)\right)]$, are elements of $\boldsymbol{\theta}_2$ which are estimated in the second step of the SR approach.

We start by studying the finite sample properties of $\boldsymbol{\theta}_1$ in the SR approach which we compare to the ML estimates of $\boldsymbol{\theta}_1$. When deriving these ML estimates, we let all parameters in $\boldsymbol{\theta}_2$ be fixed at their true values. The subsequent section extends the set of parameters to be estimated by $\boldsymbol{\theta}_2$.

5.3.1 Estimates of $\boldsymbol{\theta}_1$

Figure 3 reports biases when estimating λ by the SR approach and by the ML approach respectively. In Case 1 with small measurement errors, the SR estimates of λ are completely unbiased, even when just 10 bonds are available each period. Small biases are present in the ML estimates using 5-25 bonds in each period, but these biases disappear when the number of bonds increase. Note in relation to Figure 3 that the ML estimates are unavailable beyond 100 observables due to numerical problems when calculating the logarithm of the determinant for the one-step ahead prediction co-variance matrix.

Turning to Case 2 where measurement errors in the bond prices are larger, we see that the SR estimates of λ are still unbiased while biases in the ML estimates increase slightly compared to Case 1.

< Figure 3 about here >

The true standard errors of $\hat{\boldsymbol{\theta}}_1$ as measured by the standard deviation of the Monte Carlo estimates are reported in Figure 4. For Case 1 and 2, the precision of the SR estimates with 15 bonds or more is very close to the precision of the ML estimates.

< Figure 4 about here >

Biases from the estimation of standard errors for $\hat{\boldsymbol{\theta}}_1$ are displayed in Figure 5. The red lines with circles represents the heteroscedastic robust estimates from the SR approach, and the green lines with squares refer to the non-heteroscedastic robust estimates from the SR approach. As before, the black lines with stars denote the ML estimates. All these estimates of standard errors are basically unbiased, and this holds even when only a few bonds are available in each time period and with large measurement errors (i.e. in Case 2).

< Figure 5 about here >

To summarize, the finite sample properties of $\hat{\boldsymbol{\theta}}_1$ are well approximated by the asymptotic distribution derived in section 4. This is the case even with a small number of observables in each time period. Moreover, the SR estimator and the ML estimator exhibit a similar degree of precision. This suggests that the loss in efficiency of not using a likelihood approach is minimal in this case.

5.3.2 Estimates of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$

We now turn to estimation of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The ML estimates are obtained in the standard way by maximizing the likelihood function across all parameters. For the SR approach, $\boldsymbol{\theta}_1$ is estimated as in the previous section and $\boldsymbol{\theta}_2$ is estimated based on the moments chosen in section 3.5, equation (21). This gives 18 moments which exactly identifies the 18 parameters in $\boldsymbol{\theta}_2$. To make the simulation study numerically feasible, we only consider Case 1 where all bond prices are generated from interest rates with measurement errors having a standard deviation of 10 basis points. For a similar reason, we only compute ML estimates up to the case where 25 observables are available each time period.

The biases for the two estimators are reported in Figure 6. We first note that biases in the SR approach are small and decrease when the number of bonds increases. The latter result is very intuitive because more bonds reduce the measurement errors in the estimated factors and give more precise estimates of the size of the measurement error. We also observe that the performance of the SR approach with just 25 bonds is similar to the performance of the ML approach. On the other hand, the performance of the ML approach does not improve much when we increase the number of observables from 5 to 25 in each time period.

Figure 6 also reports the ML estimates when the factors are observed, or equivalently, when there are an infinite number of observables in each time period. These ML estimates for a standard VAR(1) model are denoted by thick pluses in Figure 6. Comparing the SR approach to these estimates, we see that the SR estimates at about 25 bonds already have converged to these optimal estimates.

< Figure 6 about here >

The true standard errors for the estimates are displayed in Figure 7. Here we find that the SR approach with just 25 bonds in each time period achieves the same precision as the ML approach based on 5-10 observables. Note also that the precision of the SR approach with about 25 bonds is very similar to the precision of the optimal (but infeasible) ML estimates in a VAR model where all factors are observed.

< Figure 7 about here >

Figure 8 illustrates biases when estimating the standard errors. When calculating the standard errors for the SR approach we use one lag in the Newey-West estimator of $E\left[\frac{\partial \mathbf{q}_t(\boldsymbol{\theta}_1^o,\boldsymbol{\theta}_2^o)}{\partial(\boldsymbol{\theta}_2^o)'}\right]$ because only $Cov\left(\hat{\mathbf{w}}_{t+1}, \hat{\mathbf{w}}_t\right) \neq \mathbf{0}$. Also in terms of estimating the standard errors does the SR approach do well with only small biases in the estimated standard errors. Compared to the ML approach with 5-10 bonds, we once again find that the SR approach is performing equally well with just 25 bonds.

< Figure 8 about here >

To summarize, the finite sample distribution of $\hat{\boldsymbol{\theta}}_2$ is well approximated by the asymptotic distribution derived in section 4.2. A second key finding is that the finite sample performance of the SR approach with just 25 bonds is similar to that of the ML approach with 5-10 bonds. These results hold in terms of unbiasness and efficiency.

6 Conclusion

This paper presents a new and simple estimation approach for a wide class of non-linear dynamic term structure models with potentially latent variables. The latent variables may have a Gaussian or non-Gaussian probability distribution. The novelty of our approach is the use of many observables (yields or bonds prices) in the cross-section dimension instead of just a few observables. We argue that this actually simplifies the estimation process, contrary to the common belief in the literature. An important benefit of using many observables in each time period is to realize that an accurate and very fast estimator of the latent factors is to minimize the distance between the observed yields/bond prices and the model implied yields/bond prices. The performance of this regression filter is shown to converge to the optimal smoothing estimator when the number of observables tends to infinity. We also show how parameters in dynamic term structure models can be estimated consistently under very weak restrictions from the output of the regression filter.

We hope that the introduction of the SR approach will generate more research in the exciting field of non-linear and non-Gaussian dynamic term structure models. More careful examining of the empirical implications of quadratic term structure models as presented by Ahn, Dittmar & Gallant (2002) and Realdon (2006) seem to be a natural starting point for future research. Further research could also focus on tests for non-linearities in the market price of risk and thus allow for strong non-linearities in the factor dynamics. We thus conjecture that the SR approach could have important applications in future research.

References

Ahn, D.-H., Dittmar, R. F. & Gallant, A. R. (2002), 'Quadratic term structure models: Theory and evidence', *The Review of Financial Studies* **15**(1), 243–288.

Andreasen, M. M. (2008), 'Non-linear DSGE models, the central difference kalman filter, and the mean shifted particle filter', *Working Paper* .

Bauwens, L., Laurent, S. & Rombouts, J. V. K. (2006), 'Multivariate GARCH models: A survey', *Journal of Applied Econometrics* **21**, 79–109.

Brandt, M. W. & He, P. (2005), 'Simulated likelihood estimation of affine term structure models from panel data', *Working Paper* .

Brown, S. J. & Dybvig, P. H. (1986), 'The empirical implications of the cox, ingersoll, ross theory of the term structure of interest rates', *The Journal of Finance* **41**(3), 617–630.

Carrasco, M. & Florens, J.-P. (2002), 'Simulation-based method of moments and efficiency', *Journal of Business and Economic Statistics* **20**, 482–492.

Chen, R.-R. & Scott, L. (1993), 'Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates', *The Journal of Fixed Income* pp. 14–31.

Connor, G. & Korajczyk, R. A. (1986), 'Performance measurement with the arbitrage pricing theory', *Journal of Financial Economics* **15**, 373–394.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**(2), 385–407.

Dai, Q. & Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *Journal of Finance* **55**, 1946–1978.

de Jong, F. (2000), 'Time series and cross-section information in affine term-structure models', *Journal of Business and Economic Statistics* **18**, 300–314.

Diebold, F. X. & Li, C. (2006), 'Forecasting the term structure of government bond yields', *Journal of Econometrics* **130**.

Diebold, F. X., Rudebusch, G. D. & Aruoba, S. B. (2006), 'The macroeconomy and the yield curve: A dynamic latent factor approach', *Journal of Econometrics* **131**, 309–338.

Doucet, A., de Freitas, N. & Gordon, N. (2001), 'An introduction to sequential monte carlo methods', *Sequential Monte Carlo Methods in Practice, Springer, editors: Arnaud Doucet, Nando de Freitas and Neil Gordon* .

Duan, J.-C. & Simonato, J.-G. (1999), 'Estimating and testing exponential-affine term structure models by kalman filter', *Review of Quantitative Finance and Accounting* **13**, 111–135.

Duffee, G. R. (2002), 'Term premia and interest rate forecasts in affine models', *The Journal of Finance* **LVII**(1), 405–443.

Duffie, D. & Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**(4), 379–406.

Duffie, D. & Singleton, K. J. (1993), 'Simulated moments estimation of markov models of asset prices', *Econometrica* **61**(4), 929–952.

Durbin, J. & Koopman, S. J. (2001), 'Time series analysis by state space methods', *Oxford University Press* .

Fama, E. F. & Bliss, R. R. (1987), 'The information in long-maturity forward rates', *The American Economic Review* **77**(4), 680–692.

Fuller, W. A. (1987), 'Measurement error models', *John Wiley and Sons* .

Hamilton, J. D. (1994), 'Time series analysis', *Princeton University Press, Princeton New Jersey* .

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054.

Jazwinski, A. H. (1970), 'Stochastic processes and filtering theory'.

Julier, S. J., Uhlmann, J. K. & Durrant-Whyte, H. F. (1995), 'A new approach for filtering nonlinear systems', *In The Proceedings of the American Control Conference* pp. 1628–1632.

Jungbacker, B. & Koopman, S. J. (2008), 'Likelihood-based analysis for dynamic factor models', *Working Paper* .

Nelson, C. R. & Siegel, A. F. (1987), 'Parsimonious modelling of yield curves', *The Journal of Business* **60**(4), 473–489.

Norgaard, M., Poulsen, N. K. & Ravn, O. (2000), 'Advances in derivative-free state estimation for nonlinear systems', *Automatica* **36:11**, 1627–1638.

Pan, J. (2002), 'The jump-risk premia implicit in options: Evidence from an integrated time-series study', *Journal of Financial Economics* **63**, 3–50.

Pastorello, S., Patilea, V. & Renault, E. (2003), 'Iterative and recursive estimation in structural nonadaptive models', *Journal of Business and Economic Statistics* **21**(4), 449–482.

Realdon, M. (2006), 'Quadratic term structure models in discrete time', *Finance Research Letters* **3**, 277–289.

Rossi, G. D. (2004), 'Maximum likelihood estimation of the cox-ingersoll-ross model using particle filters', *Working Paper* .

Sherman, R. P. (2003), 'Iterative and recursive estimation in structural nonadaptive models: Comment', *Journal of Business and Economic Statistics* **21**(4), 498–500.

Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economiecs* **5**, 177–188.

Wooldridge, J. M. (2002), 'Econometric analysis of cross section and panel data', *The MIT Press, US* .

A Condition for identification

Consider

$\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1))^2}{Var(v_{j,t})} = \frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1)\right)^2}{Var(v_{j,t})}$

$= \frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)^2+\left(g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1)\right)^2}{Var(v_{j,t})}$
$+\frac{2\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)\left(g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1)\right)}{Var(v_{j,t})}$

Thus

$E\left[\sum_{t=1}^{T}\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1))^2}{Var(v_{j,t})}\right]$

$= E\left[E\left[\sum_{t=1}^{T}\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1))^2}{Var(v_{j,t})}\middle|\mathbf{x}_{1,t},\mathbf{z}_t\right]\right]$

$= E\left[\sum_{t=1}^{T}\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)^2+\left(g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1)\right)^2}{Var(v_{j,t})}\right]$

Since the term $E\left[\frac{\left(g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)-g_j(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1)\right)^2}{Var(v_{j,t})}\right]$ is nonnegative, it follows that

$$\sum_{t=1}^{T}E\left[\frac{(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1\right))^2}{Var\left(v_{j,t}\right)}\right] \geq \sum_{t=1}^{T}E\left[\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)^2}{Var\left(v_{j,t}\right)}\right].$$

This inequality is strict and the model is uniquely identified when

$$\sum_{t=1}^{T}E\left[\frac{\left(g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\boldsymbol{\theta}_1\right)\right)^2}{Var\left(v_{j,t}\right)}\right] > 0 \text{ for all } \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^o \text{ and } \mathbf{x}_{1,t} \neq \mathbf{x}_{1,t}^o \text{ for all } t$$

B Computing moments for the illustration of GMM

This section computes the moments used in the illustration of GMM. The infeasible model for the factors is

$$\mathbf{x}_{t+1} = \boldsymbol{\alpha} + \mathbf{h}_{\mathbf{x}}\mathbf{x}_t + \mathbf{w}_{t+1}$$

where $\mathbf{w}_t \sim \mathcal{IID}\left(\mathbf{0}, Var\left(\mathbf{w}_t\right)\right)$. The observed and feasible model for the factors is

$$\hat{\mathbf{x}}_{t+1} = \boldsymbol{\alpha} + \mathbf{h}_{\mathbf{x}}\hat{\mathbf{x}}_t + \hat{\mathbf{w}}_{t+1}$$

where

$$\hat{\mathbf{x}}_t = \mathbf{x}_t^o + \mathbf{u}_t \quad \mathbf{u}_t \sim \mathcal{NID}\left(\mathbf{0}, Var\left(\hat{\mathbf{x}}_t\right)\right) \text{ for } t = 1, 2, ..., T$$

Note then that

$E\left(\hat{\mathbf{w}}_{t+1}\right) = E\left[\hat{\mathbf{x}}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\hat{\mathbf{x}}_t\right]$

$$= E\left[\mathbf{x}_{t+1}^{o} + \mathbf{u}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_{t}^{o} + \mathbf{u}_{t}\right)\right]$$

$$= E\left[\mathbf{x}_{t+1}^{o} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\mathbf{x}_{t}^{o}\right]$$

$$= E\left[\mathbf{w}_{t+1}\right]$$

$$= \mathbf{0}$$

$$E\left(\hat{\mathbf{w}}_{t+1}\hat{\mathbf{x}}_{t}'\right) = E\left[\left(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\hat{\mathbf{x}}_{t}\right)\hat{\mathbf{x}}_{t}'\right]$$

$$= E\left[\left(\mathbf{x}_{t+1}^{o} + \mathbf{u}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_{t}^{o} + \mathbf{u}_{t}\right)\right)\left(\mathbf{x}_{t}^{o} + \mathbf{u}_{t}\right)'\right]$$

$$= E\left[\left(\mathbf{x}_{t+1}^{o} + \mathbf{u}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_{t}^{o} + \mathbf{u}_{t}\right)\right)\left(\mathbf{x}_{t}^{o}\right)' + \left(\mathbf{x}_{t+1}^{o} + \mathbf{u}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_{t}^{o} + \mathbf{u}_{t}\right)\right)\mathbf{u}_{t}'\right]$$

$$= E\left[\left(\mathbf{x}_{t+1}^{o} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\mathbf{x}_{t}^{o} + \mathbf{u}_{t+1} - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)\left(\mathbf{x}_{t}^{o}\right)' + \left(\mathbf{x}_{t+1}^{o} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\mathbf{x}_{t}^{o} + \mathbf{u}_{t+1} - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)\mathbf{u}_{t}'\right]$$

$$= E\left[\left(\mathbf{w}_{t+1} + \mathbf{u}_{t+1} - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)\left(\mathbf{x}_{t}^{o}\right)' + \left(\mathbf{w}_{t+1} + \mathbf{u}_{t+1} - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)\mathbf{u}_{t}'\right]$$

$$= E\left[\mathbf{w}_{t+1}\left(\mathbf{x}_{t}^{o}\right)' + \mathbf{u}_{t+1}\left(\mathbf{x}_{t}^{o}\right)' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\left(\mathbf{x}_{t}^{o}\right)' + \left(\mathbf{w}_{t+1}\mathbf{u}_{t}' + \mathbf{u}_{t+1}\mathbf{u}_{t}' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\mathbf{u}_{t}'\right)\right]$$

$$= E\left[\mathbf{u}_{t+1}\left(\mathbf{x}_{t}^{o}\right)' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\left(\mathbf{x}_{t}^{o}\right)' + \left(\mathbf{w}_{t+1}\mathbf{u}_{t}' + \mathbf{u}_{t+1}\mathbf{u}_{t}' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\mathbf{u}_{t}'\right)\right]$$

since $\mathbf{w}_{t+1}$ is iid. and thus independent of $\mathbf{x}_{t}^{o}$

$$= E\left[-\mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\left(\mathbf{x}_{t}^{o}\right)' + \left(\mathbf{w}_{t+1}\mathbf{u}_{t}' + \mathbf{u}_{t+1}\mathbf{u}_{t}' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\mathbf{u}_{t}'\right)\right]$$

since $\mathbf{u}_{t+1}$ is a function of $\mathbf{v}_{t+1}$ and $\mathbf{x}_{t}^{o}$ is a function of $\{\mathbf{w}_{i}\}_{i=1}^{t}$,
which imply independence of $\mathbf{u}_{t+1}$ and $\mathbf{x}_{t}^{o}$ due to the independence of $\mathbf{v}_{t}$ and $\mathbf{w}_{t}$

$$= E\left[\left(\mathbf{w}_{t+1}\mathbf{u}_{t}' + \mathbf{u}_{t+1}\mathbf{u}_{t}' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\mathbf{u}_{t}'\right)\right]$$

as above because $\mathbf{v}_{t}$ and $\mathbf{w}_{t}$ are independent

$$= E\left[\mathbf{u}_{t+1}\mathbf{u}_{t}' - \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\mathbf{u}_{t}'\right]$$

because $\mathbf{u}_{t}$ is a function of $\mathbf{v}_{t}$ and $\mathbf{v}_{t}$ and $\mathbf{w}_{t}$ are independent

$$= E\left[-\mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\mathbf{u}_{t}'\right]$$

$\mathbf{u}_{t}$ is iid.

$$= -\left(\mathbf{h}_{\mathbf{x}}E\left[\mathbf{u}_{t}\mathbf{u}_{t}'\right]\right)$$

$$= -\mathbf{h}_{\mathbf{x}}Var\left(\mathbf{u}_{t}\right)$$

$$Var\left(\hat{\mathbf{w}}_{t+1}\right) = Var\left(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\alpha} - \mathbf{h}_{\mathbf{x}}\hat{\mathbf{x}}_{t}\right)$$

$$= Var\left(\mathbf{w}_{t+1} - \mathbf{u}_{t+1} + \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)$$

$$= Var\left(\mathbf{w}_{t+1}\right) + Var\left(\mathbf{u}_{t+1}\right) + \mathbf{h}_{\mathbf{x}}Var\left(\mathbf{u}_{t}\right)\mathbf{h}_{\mathbf{x}}'$$

since $Cov\left(\mathbf{w}_{t+1}, \mathbf{u}_{t+1}\right)$, $Cov\left(\mathbf{w}_{t+1}, \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)$, $Cov\left(-\mathbf{u}_{t+1}, \mathbf{h}_{\mathbf{x}}\mathbf{u}_{t}\right)$ are zero

Thus

$$\begin{bmatrix} E\left(\hat{\mathbf{w}}_{t}\right) \\ vec\left(E\left(\hat{\mathbf{w}}_{t}\hat{\mathbf{x}}_{t}'\right)\right) \\ vech\left(Var\left(\hat{\mathbf{w}}_{t}\right)\right) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ vec\left(-\mathbf{h}_{\mathbf{x}}Var\left(\mathbf{u}_{t}\right)\right) \\ vech\left(Var\left(\mathbf{w}_{t}\right) + Var\left(\mathbf{u}_{t}\right) + \mathbf{h}_{\mathbf{x}}Var\left(\mathbf{u}_{t}\right)\mathbf{h}_{\mathbf{x}}'\right) \end{bmatrix}$$

C Proof of proposition 1

The first-order conditions for $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}$ and $\hat{\boldsymbol{\theta}}_1$ are

$$\frac{\partial Q_t}{\partial \mathbf{x}_{2,t}} = -\sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)\right)}{V\left(\mathbf{z}_{t,j}; \gamma^*\right)} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} = \mathbf{0} \qquad \text{for } t = 1, 2, ..., T \tag{57}$$

$$\begin{aligned} \frac{\partial Q\left(\boldsymbol{\theta}_1\right)}{\partial \boldsymbol{\theta}_1} &= -\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right); \hat{\boldsymbol{\theta}}_1\right)\right)}{V\left(z_{t,j}; \hat{\gamma}\right)} \times \\ &\quad \left(\frac{\partial \hat{\mathbf{x}}_{2,t}'\left(\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\boldsymbol{\theta}}_1} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right)} + \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\boldsymbol{\theta}}_1}\right) \\ &= \mathbf{0} \end{aligned} \tag{58}$$

Note then that

$$\frac{\partial Q(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = -\sum_{t=1}^{T} \frac{\partial \hat{\mathbf{x}}_{2,t}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1} \sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}(\hat{\boldsymbol{\theta}}_1); \hat{\boldsymbol{\theta}}_1\right)\right)}{V(z_{t,j}; \hat{\gamma})} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}(\hat{\boldsymbol{\theta}}_1); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}(\hat{\boldsymbol{\theta}}_1)} - \sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}(\hat{\boldsymbol{\theta}}_1); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\boldsymbol{\theta}}_1}$$

$$= -\sum_{t=1}^{T} \frac{\partial \hat{\mathbf{x}}_{2,t}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1} \underbrace{\sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right); \hat{\boldsymbol{\theta}}_1\right)\right)}{V\left(z_{t,j}; \hat{\gamma}\right)} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}\left(\hat{\boldsymbol{\theta}}_1\right)}}_{\mathbf{0} \text{ using } (57)}$$

$$-\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}(\hat{\boldsymbol{\theta}}_1); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\boldsymbol{\theta}}_1}$$

$$\Updownarrow$$

$$\frac{\partial Q(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = -\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}(\hat{\boldsymbol{\theta}}_1); \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\boldsymbol{\theta}}_1} = \mathbf{0}$$

But (57) and (58) are also the first-order conditions to the joint estimation problem

$$\left(\hat{\boldsymbol{\theta}}_1, \{\hat{\mathbf{x}}_{2,t}\}_{t=1}^{T}\right) = \arg \min_{\left(\boldsymbol{\theta}_1, \{\mathbf{x}_{2,t}\}_{t=1}^{T}\right)} \sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \boldsymbol{\theta}_1\right)\right)^2}{V\left(z_{t,j}; \hat{\gamma}\right)}$$

Q.E.D.

D The asymptotic distribution of $\hat{\theta}_1$

In addition to the assumptions in Proposition 2, let: i) $\boldsymbol{\theta}_1^o$ be in the interior of $\boldsymbol{\Theta}_1$, ii) $\mathbf{s}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_t, \mathbf{z}_{t,j}; \boldsymbol{\theta}_1, \boldsymbol{\gamma}\right)$ be continuously differentiable on the interior of $\left(\mathcal{X}_{2,t}, \boldsymbol{\Theta}^1, \boldsymbol{\Gamma}\right)$ for all $\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) \in \left(\mathcal{X}_{1,t}, \mathcal{Z}_t\right)$, iii) $\left|H^{\boldsymbol{\theta}_1}\left(\mathbf{x}_t, z_{t,j}; \boldsymbol{\theta}_1, \boldsymbol{\gamma}\right)_{k,l}\right| \leq b\left(\mathbf{x}_{1,t}, z_{t,j}\right)$ for $k, l = 1, 2, ..., L_1$ where $E\left[b\left(\mathbf{x}_{1,t}, z_{t,j}\right)\right] < \infty$, iv) $\mathbf{A}_o^{\boldsymbol{\theta}_1} \equiv$

$E\left[\mathbf{H}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,z_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$ be positive definite, and v) $\mathbf{s}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)$ be i.i.d. with finite second moments. Then for $N\longrightarrow\infty$

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},Var\left(\hat{\boldsymbol{\theta}}_1\right)\right) \tag{59}$$

where

$$Var\left(\hat{\boldsymbol{\theta}}_1\right)=\left(\mathbf{A}_o^{\boldsymbol{\theta}_1}\right)^{-1}\mathbf{B}_o^{\boldsymbol{\theta}_1}\left(\mathbf{A}_o^{\boldsymbol{\theta}_1}\right)^{-1} \tag{60}$$

Proof:

Let $Q\left(\boldsymbol{\theta}_1\right)\equiv\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}}\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1))^2}{V(z_{t,j};\hat{\gamma})}$. The first-order derivative with respect to θ_1 is

$$\frac{\partial Q(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}=-\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}}\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1))}{V(z_{t,j};\hat{\gamma})}\left(\frac{\partial\hat{\mathbf{x}}_{2,t}'(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1)}+\frac{\partial g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\right)$$

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}=&-\sum_{t=1}^{T}(\sum_{j=1}^{n_{y,t}}\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1))}{V(z_{t,j};\hat{\gamma})}\frac{\partial\hat{\mathbf{x}}_{2,t}'(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}\frac{\partial g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1)}\\&+\sum_{j=1}^{n_{y,t}}\frac{(y_{t,j}-g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1))}{V(z_{t,j};\hat{\gamma})}\frac{\partial g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1})\end{aligned}$$

$$=\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}}\mathbf{s}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1\right),\mathbf{z}_{t,j};\boldsymbol{\theta}_1,\hat{\boldsymbol{\gamma}}\right)$$

where we have defined

$$\begin{aligned}\mathbf{s}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1\right),\mathbf{z}_{t,j};\boldsymbol{\theta}_1,\hat{\boldsymbol{\gamma}}\right)\equiv&-\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1\right);\boldsymbol{\theta}_1\right)\right)}{V\left(z_{t,j};\hat{\gamma}\right)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1\right);\boldsymbol{\theta}_1\right)}{\partial\boldsymbol{\theta}_1}\\&-\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1\right);\boldsymbol{\theta}_1\right)\right)}{V\left(z_{t,j};\hat{\gamma}\right)}\frac{\partial\mathbf{x}_{2,t}'\left(\boldsymbol{\theta}_1^o\right)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1^o\right);\boldsymbol{\theta}_1^o\right)}{\partial\mathbf{x}_{2,t}\left(\boldsymbol{\theta}_1^o\right)}\end{aligned} \tag{61}$$

to be the score vector for $\boldsymbol{\theta}_1$ at time t for observable number j. Note that $\frac{\partial g_j(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1}$ has dimension $L_1\times1$, $\frac{\partial\hat{\mathbf{x}}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}$ has dimension $L_1\times n_{x_2}$, and $\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\hat{\mathbf{x}}_{2,i}(\boldsymbol{\theta}_1^o)}$ has dimension $n_{x_2}\times1$. Thus, $\mathbf{s}^{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1\right),\mathbf{z}_{t,j};\boldsymbol{\theta}_1,\hat{\boldsymbol{\gamma}}\right)$ has dimension $L_1\times1$. We may therefore write the first-order condition for $\hat{\boldsymbol{\theta}}_1$ as

$$\sum_{i=1}^{N}\mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i;\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\gamma}}\right)=\mathbf{0} \tag{62}$$

where we have set up the pooled sample, i.e.

$\mathcal{I}=\left\{\{j\}_{j=1}^{n_{y,1}},\{j\}_{j=1}^{n_{y,2}},...,\{j\}_{j=1}^{n_{y,T}}\right\}$ with $N\equiv\sum_{t=1}^{T}n_{y,t}$ elements.

$\mathbf{v}\equiv\left\{\{v_{1,j}\}_{j=1}^{n_{y,1}},\{v_{2,j}\}_{j=1}^{n_{y,2}},...,\{v_{T,j}\}_{j=1}^{n_{y,T}}\right\}$ which has dimension $1\times N$

$\mathbf{z}\equiv\left\{\{z_{1,j}\}_{j=1}^{n_{y,1}},\{z_{2,j}\}_{j=1}^{n_{y,2}},...,\{z_{T,j}\}_{j=1}^{n_{y,T}}\right\}$ which has dimension $1\times N$

$\mathbf{y}\equiv\left\{\{y_{1,j}\}_{j=1}^{n_{y,1}},\{y_{2,j}\}_{j=1}^{n_{y,2}},...,\{y_{T,j}\}_{j=1}^{n_{y,T}}\right\}$ which has dimension $1\times N$

$\mathbf{x}\equiv\left\{\{\mathbf{x}_1\}_{j=1}^{n_{y,1}},\{\mathbf{x}_2\}_{j=1}^{n_{y,2}},...,\{\mathbf{x}_T\}_{j=1}^{n_{y,T}}\right\}$ which has dimension $n_x\times N$

We refer to the elements in $\mathbf{v}$, $\mathbf{z}$, $\mathbf{y}$, and $\mathbf{x}$ by v_i, $\mathbf{z}_i$, y_i, and $\mathbf{x}_i$, respectively. For notational convenience let $u_i \equiv \left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}\left(\hat{\boldsymbol{\theta}}_1\right), \mathbf{z}_i\right)$.

We now do a mean value expansion of $\sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\gamma}}\right)$ around $\boldsymbol{\theta}_1^o$ (the true value of $\boldsymbol{\theta}_1^o$) and $\boldsymbol{\gamma}^*$, where $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}^*$ but $\boldsymbol{\gamma}^*$ does not need to be the true value of $\boldsymbol{\gamma}$. This implies

$$\sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\gamma}}\right) = \sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) + \sum_{i=1}^{N} \frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \tilde{\boldsymbol{\theta}}_1}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) + \sum_{i=1}^{N} \frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1, \boldsymbol{\gamma}^*\right)}{\partial \boldsymbol{\gamma}^*}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right) \tag{63}$$

Next, let $\boldsymbol{H}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \tilde{\boldsymbol{\theta}}_1}$ be a $L_1 \times L_1$ Hessian matrix for the object function in terms of $\boldsymbol{\theta}_1$ evaluated at $\tilde{\boldsymbol{\theta}}_1$ where $\tilde{\boldsymbol{\theta}}_{1,i} = \alpha_i \hat{\theta}_{1,i} + (1-\alpha_i)\,\theta_{1,i}^o$ for $i = 1, 2, ...L_1$ and $\alpha_i \in (0,1)$. This notation indicates that each row of the Hessian matrix is evaluated at a potentially different mean, but since $\hat{\boldsymbol{\theta}}_1 \xrightarrow{p} \boldsymbol{\theta}_1^o$ it follows that $\tilde{\boldsymbol{\theta}}_1 \xrightarrow{p} \boldsymbol{\theta}_1^o$. We also let $\mathbf{J}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial (\boldsymbol{\gamma}^*)'}$ be the Jacobian of dimension $L_1 \times n_\gamma$ where n_γ is the size of $\boldsymbol{\gamma}$.

Using (62), (63) reduces to

$$\mathbf{0} = \sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) + \sum_{i=1}^{N} \mathbf{H}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) + \sum_{i=1}^{N} \mathbf{J}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right)$$

$$\Updownarrow$$

$$\mathbf{0} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) + \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{H}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)\sqrt{N}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) + \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{J}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)\sqrt{N}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right)$$

$$\Updownarrow$$

$$\left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{H}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)\sqrt{N}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) = \frac{-1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) - \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{J}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)\sqrt{N}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right)$$

$$\Updownarrow$$

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{H}^{\boldsymbol{\theta}_1}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)^{-1}\left(\frac{-1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) - \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{J}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)\sqrt{N}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right)\right)$$

For $N \to \infty$ we have $\frac{1}{N} \sum_{i=1}^{N} \mathbf{J}\left(u_i; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \xrightarrow{p} \frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right)}{\partial \boldsymbol{\gamma}^*}$ given standard regularity conditions. But

$$\frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right)}{\partial (\boldsymbol{\gamma}^*)'} = \frac{\left(y_i - g_i\left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o); \boldsymbol{\theta}_1^o\right)\right)}{V\left(\mathbf{z}_{t,j}; \boldsymbol{\gamma}^*\right)^2}\left(\frac{\partial g_i\left(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o); \boldsymbol{\theta}_1^o\right)}{\partial \boldsymbol{\theta}_1^o} + \frac{\partial \mathbf{x}_{2,t}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o} \frac{\partial g_i\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}(\boldsymbol{\theta}_1^o); \boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2,t}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial V(\mathbf{z}_i; \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}^*}\right)'$$

so
$E\left[\left.\frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)}{\partial(\boldsymbol{\gamma}^*)'}\right|\mathbf{x}_{1,i},\mathbf{z}_i\right] = \frac{E[(y_i-(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o))|\mathbf{x}_{1,i},\mathbf{z}_i]}{V(\mathbf{z}_i;\gamma^*)^2}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}+\frac{\partial \mathbf{x}'_{2,t}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,t},\mathbf{x}_{2,t}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,t}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial V(\mathbf{z}_i;\gamma^*}{\partial\gamma^*}\right.$
$\mathbf{0}$

Hence
$E\left[\frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)}{\partial(\boldsymbol{\gamma}^*)'}\right] = E\left[E\left[\left.\frac{\partial \mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)}{\partial(\boldsymbol{\gamma}^*)'}\right|\mathbf{x}_{1,i},\mathbf{z}_i\right]\right] = \mathbf{0}$

By assumption, $\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)$ is $i.i.d.$ and with $E\left[\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)\right]=\mathbf{0}$ and $E\left[\left[s^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)_l\right]^2\right]<\infty$ for $l=1,2,...,L_1$. It therefore follows by the central limit theorem (The Lindeberg-Levy) for $n_{y,t}\to\infty$ for all t, meaning $N\to\infty$, that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},Var\left(\hat{\boldsymbol{\theta}}_1\right)\right) \tag{64}$$

where

$$Var\left(\hat{\boldsymbol{\theta}}_1\right)=\left(\mathbf{A}_o^{\boldsymbol{\theta}_1}\right)^{-1}\mathbf{B}_o^{\boldsymbol{\theta}_1}\left(\mathbf{A}_o^{\boldsymbol{\theta}_1}\right)^{-1} \tag{65}$$

and $\mathbf{A}_o^{\boldsymbol{\theta}_1}\equiv E\left[\mathbf{H}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)\right]$ and $\mathbf{B}_o^{\boldsymbol{\theta}_1}=Var\left(\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)\right)$

We estimate $Var\left(\hat{\boldsymbol{\theta}}_1\right)$ as follows.

1. The variance of the score function, robust version
For notational convenience, let
$\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*) = -(\frac{(y_i-g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o))}{V(z_i;\gamma^*)}\frac{\partial \mathbf{x}'_{2,i}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}+\frac{(y_i-g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o))}{V(z_i;\gamma^*)}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o})$

$= (\mathbf{m}_{1,i}+\mathbf{m}_{2,i})$
where
$\mathbf{m}_{1,i}\equiv-\frac{(y_i-g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o))}{V(z_i;\gamma^*)}\frac{\partial \mathbf{x}'_{2,i}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)} = -\frac{v_i}{V(z_i;\gamma^*)}\frac{\partial \mathbf{x}'_{2,i}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}$
$\mathbf{m}_{2,i}\equiv-\frac{(y_i-g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o))}{V(z_i;\gamma^*)}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o} = -\frac{v_i}{V(z_i;\gamma^*)}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}$

Hence,
$E\left[\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)\,\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)'\right]$

$= E\left[(\mathbf{m}_{1,i}+\mathbf{m}_{2,i})(\mathbf{m}_{1,i}+\mathbf{m}_{2,i})'\right]$

$= E\left[\mathbf{m}_{1,i}\mathbf{m}'_{1,i}+\mathbf{m}_{1,i}\mathbf{m}'_{2,i}+\mathbf{m}_{2,i}\mathbf{m}'_{1,i}+\mathbf{m}_{2,i}\mathbf{m}'_{2,i}\right]$

$= E\left[\mathbf{m}_{1,i}\mathbf{m}'_{1,i}+2\mathbf{m}_{2,i}\mathbf{m}'_{1,i}+\mathbf{m}_{2,i}\mathbf{m}'_{2,i}\right]$

$= E[\frac{v_i^2}{V(z_i;\gamma^*)^2}\left(\frac{\partial \mathbf{x}'_{2,i}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial \mathbf{x}'_{2,i}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$
$+2\frac{v_i^2}{V(z_i;\gamma^*)^2}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\right)\left(\frac{\partial \mathbf{x}'_{2,i}(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$

$$+\frac{v_i^2}{V(\mathbf{z}_i;\gamma^*)^2}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)']$$

Imposing conditions for uniform convergence, we have

$$\hat{\mathbf{B}}^{\boldsymbol{\theta}_1}=\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{v}_i^2}{V(z_i;\hat{\gamma})^2}\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)'$$
$$+2\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{v}_i^2}{V(z_i;\hat{\gamma})^2}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\right)\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)'$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{v}_i^2}{(V(\mathbf{z}_i;\hat{\gamma}))^2}\frac{\partial g_i((\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1))}{\partial \hat{\boldsymbol{\theta}}_1}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\right)']$$

$$=\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{v}_i^2}{V(z_i;\hat{\gamma})^2}[\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)'$$
$$+2\left(\frac{\partial g(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\right)\left(\frac{\partial \hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)'$$
$$+\frac{\partial g((\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1))}{\partial \hat{\boldsymbol{\theta}}_1}\left(\frac{\partial g(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\right)']$$

$$\xrightarrow{p} Var\left(\mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right)\text{ for }N\to\infty$$

2. The variance of the score function, homoscedastic version

If the conditional variance is correctly specified, then the estimator of $\hat{\mathbf{B}}_o^{\boldsymbol{\theta}_1}$ can be simplified as follows. I.e. we assume: for some $\boldsymbol{\gamma}_o\in\Gamma$ and $\sigma_{i,o}^2$, $Var\left(v_i|\,\mathbf{x}_{1,i},\mathbf{z}_i\right)=\sigma_{i,o}^2V\left(\mathbf{z}_i;\boldsymbol{\gamma}_o\right)$. From this assumption,

$E\left[\mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\mathbf{s}^{\boldsymbol{\theta}_1}\left(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)'\big|\,\mathbf{x}_{1,i},\mathbf{z}_i\right]=$

$$=E[\frac{v_i^2}{V(z_i;\gamma^*)^2}\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+2\frac{v_i^2}{V(z_i;\gamma^*)^2}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+\frac{v_i^2}{V(\mathbf{z}_i;\gamma^*)^2}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)'|\mathbf{x}_{1,i},\mathbf{z}_i]$$

$$=E[\frac{v_i^2}{V(z_i;\gamma^*)^2}|\mathbf{x}_{1,i},\mathbf{z}_i](\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \hat{\mathbf{x}}_{2,i}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+2\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)')$$

$$=E[\frac{\sigma_{i,o}^2V(\mathbf{x}_{1,i},\mathbf{z}_i;\boldsymbol{\gamma}_o)}{V(z_i;\gamma^*)^2}|\mathbf{x}_{1,i},\mathbf{z}_i](\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \hat{\mathbf{x}}_{2,i}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+2\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \hat{\mathbf{x}}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\right)')$$

$$=\frac{\sigma_{i,o}^2}{V(z_i;\gamma^*)}(\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)\left(\frac{\partial \mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial \boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial \mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$

$$+2\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\right)\left(\frac{\partial\mathbf{x}_{2,i}'(\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o)}\right)'$$
$$+\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\mathbf{x}_{2,i}(\boldsymbol{\theta}_1^o);\boldsymbol{\theta}_1^o)}{\partial\boldsymbol{\theta}_1^o}\right)')$$

Imposing conditions for uniform convergence, we have under the additional assumption

$$\hat{\mathbf{B}}^{\boldsymbol{\theta}_1}_{\text{hom}}=\frac{1}{N}\sum_{i=1}^{N}\frac{\hat\sigma_i^2}{V(z_i;\hat\gamma)}[\left(\frac{\partial\hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)\left(\frac{\partial\hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)'$$
$$+2\left(\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\boldsymbol{\theta}}_1}\right)\left(\frac{\partial\hat{\mathbf{x}}_{2,i}'(\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\boldsymbol{\theta}}_1}\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1)}\right)'$$
$$+\frac{\partial g_i((\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1))}{\partial\hat{\boldsymbol{\theta}}_1}\left(\frac{\partial g_i(\mathbf{x}_{1,i},\hat{\mathbf{x}}_{2,i}(\hat{\boldsymbol{\theta}}_1);\hat{\boldsymbol{\theta}}_1)}{\partial\hat{\boldsymbol{\theta}}_1}\right)']$$

$\xrightarrow{p} Var\left(\mathbf{s}^{\boldsymbol{\theta}_1}(u_i;\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*)\right)$ for $N\to\infty$

where $\hat\sigma_i^2$ is estimated based on

$\hat\sigma_t^2=\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{v_{t,j}^2}{V(z_{t,j};\hat\gamma)}$ for all $t=1,2,...,T$

and $\hat\sigma_i^2=\left\{\left\{\hat\sigma_1^2\right\}_{j=1}^{n_{y,1}},\left\{\hat\sigma_2^2\right\}_{j=1}^{n_{y,2}},...,\left\{\hat\sigma_T^2\right\}_{j=1}^{n_{y,T}}\right\}$

3. The Hessian matrix

We use numerical derivatives to find $\boldsymbol{H}^{\boldsymbol{\theta}_1}\left(u_i;\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\gamma}}\right)$. Hence

$\hat{\mathbf{A}}^{\boldsymbol{\theta}_1}=\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{H}}^{\boldsymbol{\theta}_1}\left(u_i;\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\gamma}}\right)$

E The asymptotic distribution of the latent factors, $\mathbf{x}_{2,t}$

In addition to the assumptions in Proposition 2, let: i) $\mathbf{x}_{2,t}^o$ be in the interior of $\mathcal{X}_{2,t}$, ii) $\mathbf{s}^{\mathbf{x}_2}(\mathbf{x}_t,\mathbf{z}_{t,j};\boldsymbol{\theta}_1,\boldsymbol{\gamma})$ be continuous differentiable with respect to $(\mathbf{x}_{2,t},\boldsymbol{\theta}_1,\boldsymbol{\gamma})$ on the interior of $(\mathcal{X}_{2,t},\boldsymbol{\Theta}^1,\boldsymbol{\Gamma})$ for all $(\mathbf{x}_{1,t},\mathbf{z}_{t,j})\in(\mathcal{X}_{1,t},\mathcal{Z}_t)$, iii) $\left|H^{\mathbf{x}_2}(\mathbf{x}_t,z_{t,j};\boldsymbol{\theta}_1,\boldsymbol{\gamma})_{k,l}\right|\le b(\mathbf{x}_{1,t},z_{t,j})$ for $k,l=1,2,...,n_{x_2}$ where $E[b(\mathbf{x}_{1,t},z_{t,j})]<\infty$, iv) $\mathbf{A}_o^{\mathbf{x}_2}\equiv E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,z_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$ be positive definite, v) $\hat{\boldsymbol{\gamma}}\xrightarrow{p}\boldsymbol{\gamma}^*$, and vi) $\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)$ be i.i.d. with finite second moments. Then for $n_{y,t}\longrightarrow\infty$

$$\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},Var\left(\hat{\mathbf{x}}_{2,t}\right)\right)\tag{66}$$

where

$$Var\left(\hat{\mathbf{x}}_{2,t}\right)=\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\mathbf{B}_{t,o}^{\mathbf{x}_2}\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}+\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\mathbf{D}_{t,o}Var\left(\hat{\boldsymbol{\theta}}_1\right)\mathbf{D}_{t,o}'\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\tag{67}$$

Here, $\mathbf{B}_{t,o}^{\mathbf{x}_2}\equiv Var\left[\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$ and $\mathbf{D}_{t,o}\equiv E\left[\frac{\partial\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)}{\partial(\boldsymbol{\theta}_1^o)'}\right]$ is the Jacobian with dimension $n_x\times L_1$

Proof:

Let $Q_t\equiv\frac{1}{2}\sum_{j=1}^{n_{y,t}}\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\right)^2}{V(\mathbf{z}_{t,j};\gamma^*)}$. The first-order condition for the latent factors $\mathbf{x}_{2,t}$ are

$$\frac{\partial Q_t}{\partial \mathbf{x}_{2,t}} = -\sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)\right)}{V(\mathbf{z}_{t,j}; \gamma^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} = \mathbf{0}$$

$$\Updownarrow$$

$$\frac{\partial Q_t}{\partial \mathbf{x}_t} = \sum_{j=1}^{n_{y,t}} \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\hat{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\gamma}}\right) = \mathbf{0} \tag{68}$$

where we define

$$\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}, \mathbf{z}_{t,j}; \hat{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \equiv -\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)\right)}{V\left(\mathbf{z}_{t,j}; \boldsymbol{\gamma}^*\right)} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)}{\partial \mathbf{x}_{2,t}}, \tag{69}$$

which has dimension $n_{x_2} \times 1$. We now do a mean value expansion around $\mathbf{x}_{2,t}^o$, $\boldsymbol{\theta}_1^o$, and $\boldsymbol{\gamma}^*$

$$\begin{aligned}\sum_{j=1}^{n_{y,t}} \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\gamma}}\right) &= \sum_{j=1}^{n_{y,t}} \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o, \mathbf{z}_{t,j}; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) \\ &+ \sum_{j=1}^{n_{y,t}} \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \tilde{\mathbf{x}}_{2,t}'}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right) \\ &+ \sum_{j=1}^{n_{y,t}} \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \tilde{\boldsymbol{\theta}}_1'}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) \\ &+ \sum_{j=1}^{n_{y,t}} \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \left(\boldsymbol{\gamma}_1^*\right)'}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right)\end{aligned}$$

Let $\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \tilde{\mathbf{x}}_t'}$ be the $n_{x_2} \times n_{x_2}$ Hessian matrix for the latent factors in time period t. Further, let $\mathbf{D}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \tilde{\boldsymbol{\theta}}_1'}$ be the Jacobian of dimension $n_{x_2} \times L_1$, and let $\mathbf{F}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_{\mathbf{1}}, \boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)}{\partial \left(\boldsymbol{\gamma}_1^*\right)'}$ be the Jacobian of dimension $n_{x_2} \times n_\gamma$. Here, n_γ is the size of $\boldsymbol{\gamma}$. All these matrices are evaluated at $\left(\tilde{\boldsymbol{\theta}}_1, \tilde{\mathbf{x}}_{2,t}\right)$ which is on the line segment between $\left(\hat{\boldsymbol{\theta}}_1, \hat{\mathbf{x}}_{2,t}\right)$ and $\left(\boldsymbol{\theta}_1^o, \mathbf{x}_{2,t}^o\right)$. Thus for $\hat{\boldsymbol{\theta}}_1 \xrightarrow{p} \boldsymbol{\theta}_1^o$ and $\hat{\mathbf{x}}_{2,t} \xrightarrow{p} \mathbf{x}_{2,t}^o$ it follows that $\tilde{\boldsymbol{\theta}}_1 \xrightarrow{p} \boldsymbol{\theta}_1^o$ and $\tilde{\mathbf{x}}_{2,t} \xrightarrow{p} \mathbf{x}_{2,t}^o$. Using these definitions and (68), the mean value expansion then reads

$$\begin{aligned}\mathbf{0} = &\sum_{j=1}^{n_{y,t}} \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o, \mathbf{z}_{t,j}; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right) \\ &+ \sum_{j=1}^{n_{y,t}} \mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right) \\ &+ \sum_{j=1}^{n_{y,t}} \mathbf{D}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) \\ &+ \sum_{j=1}^{n_{y,t}} \mathbf{F}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_{\mathbf{1}}, \boldsymbol{\gamma}^*\right)\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right)\end{aligned}$$

$$\Updownarrow$$

$$\mathbf{0} = \frac{1}{\sqrt{n_{y,t}}} \sum_{j=1}^{n_{y,t}} \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o, \mathbf{z}_{t,j}; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right)$$

$$+\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{D}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{F}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\right)$$

$\Updownarrow$

$$\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)=$$
$$\frac{-1}{\sqrt{n_{y,t}}}\sum_{j=1}^{n_{y,t}}\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)$$
$$-\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{D}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$-\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{F}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\right)$$

$\Updownarrow$

$$\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)=$$
$$\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)^{-1}\left(\frac{-1}{\sqrt{n_{y,t}}}\sum_{j=1}^{n_{y,t}}\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right)$$
$$-\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)^{-1}\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{D}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$-\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)^{-1}\left(\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{F}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\right)\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\right)$$

For $n_{y,t}\to\infty$ we have $\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\mathbf{F}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t},\mathbf{z}_{t,j};\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\gamma}^*\right)\xrightarrow{p}E\left[\mathbf{F}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$ given standard regularity conditions. But

$$\mathbf{F}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\equiv\frac{\partial\left(-\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)}{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*\right)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial\mathbf{x}_{2,t}^o}\right)}{\partial\left(\boldsymbol{\gamma}_1^*\right)'}$$

$$=\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)}{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*\right)^2}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial\mathbf{x}_{2,t}^o}\left(\frac{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*\right)}{\partial\boldsymbol{\gamma}^*}\right)'$$

and
$E\left[\mathbf{F}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$

$$= \frac{E\left[\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)\middle|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)^2} \frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2t}^o} \left(\frac{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}^*}\right)'$$

$$= \mathbf{0}$$

Hence
$E\left[\mathbf{F}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right] = E\left[E\left[\mathbf{F}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\right] = \mathbf{0}$

Note next, when $n_{y,t} \longrightarrow \infty$ then $N \equiv \sum_{t=1}^{T} n_{y,t}$ also tends to infinity. Hence, $n_{y,t} \longrightarrow \infty$ implies $\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Var\left(\hat{\boldsymbol{\theta}}_1\right)\right)$.

By assumption, $\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)$ is *i.i.d.* and with $E\left[\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right] = \mathbf{0}$ and $E\left[\left[\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)_l\right]^2\right] < \infty$ for $l = 1, 2, ..., L_1$. By the central limit theorem (The Lindeberg-Levy) for $n_{y,t} \to \infty$, we have

$$\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Var\left(\hat{\mathbf{x}}_{2,t}\right)\right)$$

where

$$Var\left(\hat{\mathbf{x}}_{2,t}\right) = \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}\left(\mathbf{B}_{t,o}^{\mathbf{x}_2} + \mathbf{D}_{t,o} Var\left(\hat{\boldsymbol{\theta}}_1\right)\mathbf{D}_{t,o}'\right)\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1}$$

and $\mathbf{A}_{t,o}^{\mathbf{x}_2} \equiv E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right]$, $\mathbf{B}_{t,o}^{\mathbf{x}_2} \equiv Var\left(\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\right)$, and $\mathbf{D}_{t,o} \equiv \mathbf{D}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)}{\partial\left(\mathbf{x}_{2,t}^o\right)'}$.

We estimate $Var\left(\hat{\mathbf{x}}_{2,t}\right)$ as follows.

1. The Hessian matrix
$\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right) \equiv \frac{\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)}{\partial\left(\mathbf{x}_{2,t}^o\right)'}$

$$= -\frac{\partial\left(\left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2,t}^o}\right)' \frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)}{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*\right)}\right)}{\partial\left(\mathbf{x}_{2,t}^o\right)}$$

$$= -\frac{\partial^2 g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial\left(\mathbf{x}_{2,t}^o\right)'\partial\left(\mathbf{x}_{2,t}^o\right)} \frac{v_{t,j}}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} + \frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2,t}^o}\right)\left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2,t}^o}\right)'$$

and
$E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o,\mathbf{z}_{t,j};\boldsymbol{\theta}_1^o,\boldsymbol{\gamma}^*\right)\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$

$$= E\left[\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2,t}^o}\right)\left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial \mathbf{x}_{2,t}^o}\right)'\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$$

$$-E\left[v_{t,j}\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\frac{\partial^2 g_j\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{\partial\left(\mathbf{x}_{2,t}^o\right)'\partial\left(\mathbf{x}_{2,t}^o\right)}$$

$$= E\left[\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right) \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)' \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right]$$

Thus

$$E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right)\right] = E\left[E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right) \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right]\right]$$

$$= E\left[\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right) \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)' \right]$$

Imposing conditions for uniform convergence, we have

$$\hat{\mathbf{A}}^{\mathbf{x}_2}_t = \frac{1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \frac{1}{V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \right) \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \right)'$$

$$\xrightarrow{p} E\left[\mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right)\right] \text{ for } n_{y,t} \to \infty$$

2. The variance of the score function, robust version

In the general case

$$Var\left(\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right)\right)$$

$$= E\left[\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)\right)}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \left(\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)\right)}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)' \right]$$

$$= E\left[\frac{v^2_{t,j}}{\left(V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)\right)^2} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)' \right]$$

Imposing conditions for uniform convergence, we have

$$\hat{\mathbf{B}}^{\mathbf{x}_2}_t = \frac{1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \frac{\hat{v}^2_{t,j}}{\left(V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})\right)^2} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \right) \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \right)'$$

$$\xrightarrow{p} Var\left(\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right)\right) \text{ for } n_{y,t} \to \infty$$

3. The variance of the score function

If we impose the condition $Var\left(v_{t,j} \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) = \sigma^2_t$ for all $j = 1, 2, ..., n_{y,t}$, then

$$E\left[\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right) \mathbf{s}^{\mathbf{x}_t}\left(\mathbf{x}_{1,t}, \mathbf{x}^o_{2,t}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^o_1, \boldsymbol{\gamma}^*\right)' \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right]$$

$$= E\left[\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)\right)}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \left(\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)\right)}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)' \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right]$$

$$= E\left[\frac{v^2_{t,j}}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)^2} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)' \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right]$$

$$= E\left[v^2_{t,j} \middle| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right] \frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)^2} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)'$$

$$= \frac{\sigma^2_t}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)^2} \frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t}\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} \right)'$$

Imposing conditions for uniform convergence, we have

$$\hat{\mathbf{B}}^{\mathbf{x}_t}_{t,\text{hom}} = \frac{\hat{\sigma}^2_t}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \frac{1}{(V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}}))^2} \left(\frac{\partial g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \right) \left(\frac{\partial g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \right)'$$

$$\xrightarrow{p} Var\left(\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right)\right) \text{ for } n_{y,t}\to\infty$$

where $\hat{\sigma}^2_t = 1/n_{y,t}\sum_{j=1}^{n_{y,t}} \hat{v}^2_{j,t}$

4. The Jacobian

Recall that $\mathbf{D}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right) \equiv \frac{\partial \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right)}{\partial(\boldsymbol{\theta}^o_1)'}$

$$= -\frac{\partial\left(\frac{\left(y_{t,j}-g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)\right)}{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*\right)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}}\right)}{\partial(\boldsymbol{\theta}^o_1)'}$$

$$= \frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}}\left(\frac{\partial \mathbf{x}'_{2,t}(\boldsymbol{\theta}^o_1)}{\partial \boldsymbol{\theta}^o_1}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} + \frac{g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \boldsymbol{\theta}^o_1}\right)'$$
$$-\frac{v_{t,j}}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\frac{\partial^2 g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial(\boldsymbol{\theta}^o_1)'\partial \mathbf{x}^o_{2,t}}$$

$$E\left[\mathbf{D}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right)\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right] = -\frac{\partial^2 g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial(\boldsymbol{\theta}^o_1)'\partial \mathbf{x}^o_{2,t}} E\left[\frac{v_{t,j}}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$$
$$+E\left[\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}}\left(\frac{\partial \mathbf{x}'_{2,t}(\boldsymbol{\theta}^o_1)}{\partial \boldsymbol{\theta}^o_1}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} + \frac{g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \boldsymbol{\theta}^o_1}\right)'\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$$

$$= E\left[\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}}\left(\frac{\partial \mathbf{x}'_{2,t}(\boldsymbol{\theta}^o_1)}{\partial \boldsymbol{\theta}^o_1}\frac{\partial g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \mathbf{x}^o_{2,t}} + \frac{g_j\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)}{\partial \boldsymbol{\theta}^o_1}\right)'\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$$

and

$$E\left[\mathbf{D}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right)\right] = E\left[E\left[\mathbf{D}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right)\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\right]$$

Imposing conditions for uniform convergence, we have

$$\hat{\mathbf{D}}_t = \frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\frac{\partial g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}}\left(\frac{\partial \mathbf{x}'_{2,t}(\hat{\boldsymbol{\theta}}_1)}{\partial \hat{\boldsymbol{\theta}}_1}\frac{\partial g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial(\hat{\mathbf{x}}_{2,t})} + \frac{g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\boldsymbol{\theta}}_1}\right)'$$
$$\xrightarrow{p} E\left[\mathbf{D}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t},\mathbf{z}_{t,j};\boldsymbol{\theta}^o_1,\boldsymbol{\gamma}^*\right)\right] \text{ for } n_{y,t}\to\infty$$

F The asymptotic properties of $\hat{\sigma}^2_t$

This section shows consistency and asymptotic normality of $\hat{\sigma}^2_t$. In addition to the conditions ensuring consistency and normality of of $\hat{\boldsymbol{\theta}}_1$ and $\{\hat{\mathbf{x}}_{2,t}\}_{t=1}^T$, we impose:

1. $Var\left(y_{t,j}\middle|\,\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right) = \sigma^2_{t,o}V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}_o\right)$ for some $\boldsymbol{\gamma}_o\in\Gamma$ and $\sigma^2_{t,o}$ where $\hat{\boldsymbol{\gamma}}\xrightarrow{p}\boldsymbol{\gamma}_o$
2. $E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}^o_{2,t};\boldsymbol{\theta}^o_1\right)\right] < \infty$

3. $E\left[\frac{1}{V(z_{t,j}:\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]<\infty$

4. $E\left[\frac{1}{V(z_{t,j}:\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]<\infty$

5. $E\left[\left(\frac{v_{j,t}}{\sqrt{V(z_{t,j}:\boldsymbol{\gamma}^o)}}\right)^4\right]<\infty$

F.1 Proof of consistency for $\hat{\sigma}_t^2$

We begin by defining:

$\mathbf{v}_t'\equiv\{v_{t,j}\}_{j=1}^{n_{y,t}}$ which has dimension $1\times n_{y,t}$

$\mathbf{z}_t'\equiv\{z_{t,j}\}_{j=1}^{n_{y,t}}$ which has dimension $1\times n_{y,t}$

$\mathbf{y}_t'\equiv\{y_{t,j}\}_{j=1}^{n_{y,t}}$ which has dimension $1\times n_{y,t}$

$\mathbf{x}_t'\equiv\{\mathbf{x}_t\}_{j=1}^{n_{y,t}}$ which has dimension $n_x\times n_{y,t}$

We also need the mean value expansion of $\mathbf{g}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)$ in $\mathbf{x}_{2,t}$ and $\boldsymbol{\theta}_1$, that is

$$\mathbf{g}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)=\mathbf{g}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\mathbf{g}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\mathbf{g}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

Here, $\mathbf{g}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\equiv\frac{\partial\mathbf{g}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)}{\partial\mathbf{x}_{2,t}'}$ has dimension $n_{y,t}\times n_{x_2}$ and $\mathbf{g}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\equiv\frac{\partial\mathbf{g}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)}{\partial\boldsymbol{\theta}_1'}$ has dimension $n_{y,t}\times L_1$. All these matrices are evaluated at $\left(\tilde{\boldsymbol{\theta}}_1,\tilde{\mathbf{x}}_{2,t}\right)$ which is on the line segment between $\left(\hat{\boldsymbol{\theta}}_1,\hat{\mathbf{x}}_{2,t}\right)$ and $\left(\boldsymbol{\theta}_1^o,\mathbf{x}_{2,t}^o\right)$. Thus for $\hat{\boldsymbol{\theta}}_1\xrightarrow{p}\boldsymbol{\theta}_1^o$ and $\hat{\mathbf{x}}_{2,t}\xrightarrow{p}\mathbf{x}_{2,t}^o$ it follows that $\tilde{\boldsymbol{\theta}}_1\xrightarrow{p}\boldsymbol{\theta}_1^o$ and $\tilde{\mathbf{x}}_{2,t}\xrightarrow{p}\mathbf{x}_{2,t}^o$. Variables scaled by $\sqrt{V\left(\mathbf{z}_{t,j};\hat{\gamma}\right)}$ are denoted by a bar, i.e. $\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\equiv\mathbf{g}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)/\sqrt{V\left(\mathbf{z}_t;\hat{\gamma}\right)}=\sum_{j=1}^{n_{y,t}}g_j\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)/\sqrt{V\left(\mathbf{z}_{t,j};\boldsymbol{\gamma}^o\right)}$, etc.

Now consider

$$\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t=\left(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)'\left(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right)$$

$$=\left(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\right)'$$
$$\times\left(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\right)$$

$$=(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right))'$$
$$\times(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right))$$

using the mean value expansion from above

$$=\left(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\right)'$$

$$\times\left(\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\right)$$

$$=\left(\widehat{\bar{\mathbf{v}}}_t+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\right)'$$
$$\times\left(\widehat{\bar{\mathbf{v}}}_t+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\right)$$

because $\widehat{\bar{\mathbf{v}}}_t=\bar{\mathbf{y}}_t-\bar{\mathbf{g}}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)$

$$=\left(\widehat{\bar{\mathbf{v}}}_t'+\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'+\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\right)$$
$$\times\left(\widehat{\bar{\mathbf{v}}}_t+\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)\right)$$

$$=\widehat{\bar{\mathbf{v}}}_t'\widehat{\bar{\mathbf{v}}}_t+\widehat{\bar{\mathbf{v}}}_t'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)+\widehat{\bar{\mathbf{v}}}_t'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\widehat{\bar{\mathbf{v}}}_t+\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\widehat{\bar{\mathbf{v}}}_t+\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$=\widehat{\bar{\mathbf{v}}}_t'\widehat{\bar{\mathbf{v}}}_t$$
$$+2\widehat{\bar{\mathbf{v}}}_t'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+2\widehat{\bar{\mathbf{v}}}_t'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+2\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$\Updownarrow$$

$$\frac{1}{n_{y,t}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t=\frac{1}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t'\widehat{\bar{\mathbf{v}}}_t$$
$$+\frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\frac{1}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$+\frac{2}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$+\frac{1}{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$\Updownarrow$$

$$\frac{1}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t'\widehat{\bar{\mathbf{v}}}_t=\frac{1}{n_{y,t}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t$$

$$-\frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$-\frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$-\frac{1}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$
$$-\frac{2}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$
$$-\frac{1}{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

We now consider each of the terms on the left hand side in turn

Term 1

$\frac{1}{n_{y,t}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t=\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{v_{t,j}^2}{V(z_{t,j};\boldsymbol{\gamma}^o)}$

where $\left\{\frac{v_{t,j}^2}{V(z_{t,j};\boldsymbol{\gamma}^o)}\right\}_{j=1}^{n_{y,t}}$ is a sequence with mean $\sigma_{t,o}^2$. Hence the law of large numbers imply $\frac{1}{n_{y,t}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t\xrightarrow{p}\sigma_{t,o}^2$ for $n_{y,t}\longrightarrow\infty$.

Term 2

$$\frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)=\frac{2}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{\hat{v}_{t,j}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$

$$\xrightarrow{p}2E\left[\frac{v_{t,j}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]\left(\mathbf{x}_{2,t}^o-\mathbf{x}_{2,t}^o\right)=\mathbf{0}\text{ for }n_{y,t}\longrightarrow\infty$$

because

$$E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$$

$$=E\left[v_{j,t}\,|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\frac{\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{V(z_{t,j};\boldsymbol{\gamma}^o)}$$

$$=\mathbf{0}$$

So $E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]=E\left[E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\right]=\mathbf{0}$

Term 3

$$\frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)=\frac{2}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{\hat{v}_{t,j}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$\xrightarrow{p}2E\left[\frac{v_{t,j}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]\left(\boldsymbol{\theta}_1^o-\boldsymbol{\theta}_1^o\right)=\mathbf{0}$$

for $n_{y,t}\longrightarrow\infty$ because this implies $N\longrightarrow\infty$ and

$$E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]$$

$$=E\left[v_{j,t}\,|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\frac{\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)}{V(z_{t,j};\boldsymbol{\gamma}^o)}$$

$$= \mathbf{0}$$

So $E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{\boldsymbol{\theta}_1,j}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right] = E\left[E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{\boldsymbol{\theta}_1,j}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)|\mathbf{x}_{1,t},\mathbf{z}_{t,j}\right]\right] = \mathbf{0}$

Term 4

$$\frac{1}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$

$$= \frac{1}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\sum_{j=1}^{n_{y,t}}\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)$$

$$\xrightarrow{p}\left(\mathbf{x}_{2,t}^o-\mathbf{x}_{2,t}^o\right)E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]\left(\mathbf{x}_{2,t}^o-\mathbf{x}_{2,t}^o\right) = \mathbf{0}$$

for $n_{y,t}\longrightarrow\infty$ provided $E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right] < \infty$

Term 5

$$-\frac{2}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$= \frac{2}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^o\right)'\sum_{j=1}^{n_{y,t}}\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$\xrightarrow{p}2\left(\mathbf{x}_{2,t}^o-\mathbf{x}_{2,t}^o\right)E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]\left(\boldsymbol{\theta}_1^o-\boldsymbol{\theta}_1^o\right) = \mathbf{0}$$

for $n_{y,t}\longrightarrow\infty$ because this implies $N\longrightarrow\infty$ provided
$E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right] < \infty$

Term 6

$$\frac{1}{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$= \frac{1}{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)'\sum_{j=1}^{n_{y,t}}\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^o\right)$$

$$\xrightarrow{p}2\left(\boldsymbol{\theta}_1^o-\boldsymbol{\theta}_1^o\right)E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right]\left(\boldsymbol{\theta}_1^o-\boldsymbol{\theta}_1^o\right) = \mathbf{0}$$

for $n_{y,t}\longrightarrow\infty$ because this implies $N\longrightarrow\infty$, and provided
$E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^o;\boldsymbol{\theta}_1^o\right)\right] < \infty$.

This proves consistency of $\frac{1}{n_{y,t}}\widehat{\mathbf{v}}_t{}'\widehat{\mathbf{v}}_t$. We then note that

$$\hat{\sigma}_t^2 = \frac{1}{n_{y_t}}\sum_{j=1}^{n_{y_t}}\frac{\hat{v}_{t,j}^2}{V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})}$$

$$= \frac{1}{n_{y,t}}\sum_{j=1}^{n_{y_t}}\frac{\hat{v}_{t,j}^2}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^o)}\frac{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^o)}{V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})}$$

$\xrightarrow{p} \sigma^2_{t,o}$

for $n_{y,t} \longrightarrow \infty$ due to consistency of $\hat{\boldsymbol{\gamma}}$. This proves consistency of $\hat{\sigma}^2_t$.

F.2 Proof of normality for $\hat{\sigma}^2_t$

To prove asymptotic normality, we scale the expression for $\frac{1}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\widehat{\bar{\mathbf{v}}}_t$ by $\sqrt{n_{y,t}}$. Thus

$$\begin{aligned}\frac{1}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\widehat{\bar{\mathbf{v}}}_t - \sigma^2_{t,o} &= \frac{1}{n_{y,t}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t - \sigma^2_{t,o} \\ &- \frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right) \\ &- \frac{2}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right) \\ &- \frac{1}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right) \\ &- \frac{2}{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right) \\ &- \frac{1}{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right)\end{aligned}$$

$\Updownarrow$

$$\begin{aligned}\sqrt{n_{y,t}}\left(\frac{1}{n_{y,t}}\widehat{\bar{\mathbf{v}}}_t{}'\widehat{\bar{\mathbf{v}}}_t - \sigma^2_{t,o}\right) &= \frac{1}{\sqrt{n_{y,t}}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t - \sqrt{n_{y,t}}\sigma^2_{t,o} \\ &- \frac{2}{\sqrt{n_{y,t}}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right) \\ &- \frac{2}{\sqrt{n_{y,t}}}\widehat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right) \\ &- \frac{1}{\sqrt{n_{y,t}}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right) \\ &- \frac{2}{\sqrt{n_{y,t}}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}^o_{2,t}\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right) \\ &- \frac{1}{\sqrt{n_{y,t}}}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^o_1\right)\end{aligned}$$

We now consider each of the terms on the right hand side in turn.

Term 1:

$\frac{1}{\sqrt{n_{y,t}}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t - \sqrt{n_{y,t}}\sigma^2_{t,o}$

$$= \frac{1}{\sqrt{n_{y,t}}}\sum_{i=1}^{n_{y,t}}\frac{v^2_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)} - \sqrt{n_{y,t}}\left(\sigma^o_t\right)^2$$

$$= \frac{1}{\sqrt{n_{y,t}}}\sum_{i=1}^{n_{y,t}}\frac{v^2_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)} - \frac{n_{y,t}}{\sqrt{n_{y,t}}}\left(\sigma^o_t\right)^2$$

$$= \frac{1}{\sqrt{n_{y,t}}}\sum_{i=1}^{n_{y,t}}\left(\frac{v^2_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)} - \sigma^2_{t,o}\right)$$

where $\left\{\frac{v^2_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)} - \sigma^2_{t,o}\right\}_{j=1}^{n_{y,t}}$ is an iid sequence with mean value of zero and a variance of $E\left[\left(\frac{v^2_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)} - \sigma^2_{t,o}\right)^2\right]$

$$= E\left[\left(\frac{v_{j,t}^2}{V(z_{t,j};\boldsymbol{\gamma}^o)}\right)^2 + \sigma_{t,o}^4 - \frac{2\hat{v}_{j,t}^2\sigma_{t,o}^2}{V(z_{t,j};\boldsymbol{\gamma}^o)}\right]$$

$$= E\left[\left(\frac{v_{j,t}}{\sqrt{V(z_{t,j};\boldsymbol{\gamma}^o)}}\right)^4\right] + \sigma_{t,o}^4 - 2\sigma_{t,o}^4$$

$$= E\left[\left(\frac{v_{j,t}}{\sqrt{V(z_{t,j};\boldsymbol{\gamma}^o)}}\right)^4\right] - \sigma_{t,o}^4$$

Thus, by the central limit theorem we have $\frac{1}{\sqrt{n_{y,t}}}\bar{\mathbf{v}}_t'\bar{\mathbf{v}}_t - \sqrt{n_{y,t}}\sigma_{t,o}^2 \xrightarrow{d} \mathcal{N}\left(0, E\left[\left(\frac{v_{j,t}}{\sqrt{V(z_{t,j};\boldsymbol{\gamma}^o)}}\right)^4\right] - \sigma_{t,o}^4\right)$ for $n_{y,t} \longrightarrow \infty$.

Term 2:
$\frac{2}{\sqrt{n_{y,t}}}\hat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right)$

$$= 2\left[\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{\hat{v}_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\right]\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right)$$

$$\longrightarrow 2E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \boldsymbol{\theta}_1^o\right)\right]\mathcal{N}\left(\mathbf{0}, Var\left(\hat{\mathbf{x}}_{2,t}\right)\right)$$
for $n_{y,t} \longrightarrow \infty$

$$= \mathbf{0}$$
because $E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \boldsymbol{\theta}_1^o\right)\right] = \mathbf{0}$

Term 3:
$\frac{2}{\sqrt{n_{y,t}}}\hat{\bar{\mathbf{v}}}_t{}'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right)$

$$= 2\left[\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{\hat{v}_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}; \tilde{\boldsymbol{\theta}}_1\right)\right]\sqrt{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right)$$

$$\longrightarrow 2E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \boldsymbol{\theta}_1^o\right)\right]\mathcal{N}\left(\mathbf{0}, Var\left(\hat{\boldsymbol{\theta}}_1\right)\right)$$
for $n_{y,t} \longrightarrow \infty$

$$= \mathbf{0}$$
because $E\left[\frac{v_{j,t}}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \boldsymbol{\theta}_1^o\right)\right] = \mathbf{0}$

Term 4:

$\frac{1}{\sqrt{n_{y,t}}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^{o}\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^{o}\right)$

$= \sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^{o}\right)'\left[\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\right]\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^{o}\right)$

$\longrightarrow \mathcal{N}\left(\mathbf{0},Var\left(\hat{\mathbf{x}}_{2,t}\right)\right)E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)\right]\left(\mathbf{x}_{2,t}^{o}-\mathbf{x}_{2,t}^{o}\right)$
for $n_{y,t}\longrightarrow\infty$

$= \mathbf{0}$
provided $E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)'\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)\right]<\infty$

Term 5:
$\frac{2}{\sqrt{n_{y,t}}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^{o}\right)'\bar{\mathbf{g}}_{\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^{o}\right)$

$= 2\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t}-\mathbf{x}_{2,t}^{o}\right)'\left[\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\right]\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^{o}\right)$

$\longrightarrow 2\mathcal{N}\left(\mathbf{0},Var\left(\hat{\mathbf{x}}_{2,t}\right)\right)E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)\right]\left(\boldsymbol{\theta}_1^{o}-\boldsymbol{\theta}_1^{o}\right)$
for $n_{y,t}\longrightarrow\infty$

$= 0$
provided $E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\mathbf{x}_2}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)\right]<\infty$

Term 6:
$\frac{1}{\sqrt{n_{y,t}}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^{o}\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)'\bar{\mathbf{g}}_{\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\hat{\mathbf{x}}_{2,t};\hat{\boldsymbol{\theta}}_1\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^{o}\right)$

$= \sqrt{n_{y,t}}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^{o}\right)'\left[\frac{1}{n_{y,t}}\sum_{j=1}^{n_{y,t}}\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\tilde{\mathbf{x}}_{2,t};\tilde{\boldsymbol{\theta}}_1\right)\right]\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1^{o}\right)$

$\longrightarrow \mathcal{N}\left(\mathbf{0},Var\left(\hat{\boldsymbol{\theta}}_1\right)\right)E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)\right]\left(\boldsymbol{\theta}_1^{o}-\boldsymbol{\theta}_1^{o}\right)$
for $n_{y,t}\longrightarrow\infty$

$= 0$
provided $E\left[\frac{1}{V(z_{t,j};\boldsymbol{\gamma}^o)}\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)'\mathbf{g}_{j,\boldsymbol{\theta}_1}\left(\mathbf{x}_{1,t},\mathbf{x}_{2,t}^{o};\boldsymbol{\theta}_1^{o}\right)\right]<\infty$

In conclusion, we thus have
$\sqrt{n_{y,t}}\left(\frac{1}{n_{y,t}}\widehat{\mathbf{v}}_t{}'\widehat{\mathbf{v}}_t-\sigma_{t,o}^2\right)\xrightarrow{d}\mathcal{N}\left(0,E\left[\left(\frac{v_{j,t}}{\sqrt{V(z_{t,j};\boldsymbol{\gamma}^o)}}\right)^4\right]-\sigma_{t,o}^4\right)\qquad$ for $n_{y,t}\longrightarrow\infty$

From above

$$\hat{\sigma}_t^2 - \sigma_{t,o}^2 = \frac{1}{n_{y_t}} \sum_{j=1}^{n_{y_t}} \frac{\hat{v}_{t,j}^2}{V(\mathbf{z}_{t,j};\hat{\boldsymbol{\gamma}})} - \sigma_{t,o}^2 \xrightarrow{p} 0 \qquad \text{for } n_{y,t} \longrightarrow \infty$$

So by the asymptotic equivalence lemma

$$\sqrt{n_{y,t}}\left(\hat{\sigma}_t^2 - \sigma_{t,o}^2\right) \xrightarrow{d} \mathcal{N}\left(0, E\left[\left(\frac{v_{j,t}}{\sqrt{V(z_{t,j};\boldsymbol{\gamma}^o)}}\right)^4\right] - \sigma_{t,o}^4\right) \qquad \text{for } n_{y,t} \longrightarrow \infty$$

We suggest to estimate the asymptotic variance of $\hat{\sigma}_t^2$ by

$$\widehat{Var}\left(\hat{\sigma}_t^2\right) = \frac{1}{n_{y,t}} \sum_{j=1}^{n_{y_t}} \left(\frac{\hat{v}_{j,t}}{\sqrt{V(z_{t,j};\hat{\boldsymbol{\gamma}})}}\right)^4 - \left(\frac{1}{n_{y,t}} \sum_{j=1}^{n_{y_t}} \frac{\hat{v}_{j,t}}{\sqrt{V(z_{t,j};\hat{\boldsymbol{\gamma}})}}\right)^4$$

G The time series properties of $\mathbf{u}_t$

This section derives the time series properties of $\mathbf{u}_t$ which denotes the measurement errors in the estimated latent factors. We have from above

$$\sqrt{n_{y,t}}\left(\hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o\right) =$$

$$\left(\frac{1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)^{-1} \left(\frac{-1}{\sqrt{n_{y,t}}} \sum_{j=1}^{n_{y,t}} \mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o, \mathbf{z}_{t,j}; \boldsymbol{\theta}_1^o, \boldsymbol{\gamma}^*\right)\right)$$

where we can ignore the term with $\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o\right)$ because $\hat{\boldsymbol{\theta}}_1$ is estimated superconsistently when $T \longrightarrow \infty$. Recall that

$$\mathbf{s}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}\mathbf{x}_{2,t}, \mathbf{z}_{t,j}; \hat{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right) \equiv -\frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)\right)}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)}{\partial \mathbf{x}_{2,t}}$$

Hence

$$\mathbf{u}_t \equiv \hat{\mathbf{x}}_{2,t} - \mathbf{x}_{2,t}^o = \left(\frac{1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)^{-1} \left(\frac{-1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \frac{\left(y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \hat{\boldsymbol{\theta}}_1\right)\right)}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \hat{\boldsymbol{\theta}}_1\right)}{\partial \mathbf{x}_{2,t}^o}\right)$$

$$= \left(\frac{1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \tilde{\mathbf{x}}_{2,t}, \mathbf{z}_{t,j}; \tilde{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)\right)^{-1} \left(\frac{-1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \frac{\partial g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \hat{\boldsymbol{\theta}}_1\right)}{\partial \mathbf{x}_{2,t}^o} \frac{v_{t,j}}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\right)$$

where $v_{t,j} = y_{t,j} - g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \hat{\boldsymbol{\theta}}_1\right)$

$$= \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \mathbf{v}_t$$

which is asymptotically equivalent because

$$\mathbf{A}_{t,o}^{\mathbf{x}_2} \equiv \frac{1}{n_{y,t}} \sum_{j=1}^{n_{y,t}} \mathbf{H}^{\mathbf{x}_2}\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}^o, \mathbf{z}_{t,j}; \hat{\boldsymbol{\theta}}_1, \boldsymbol{\gamma}^*\right)$$

$$\mathbf{Z}_t \equiv \frac{-1}{n_{y,t}} \left[\left\{\frac{\partial g_j\left(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}^o; \hat{\boldsymbol{\theta}}_1\right)}{\partial \mathbf{x}_{2,t}^o} \frac{1}{V(\mathbf{z}_{t,j};\boldsymbol{\gamma}^*)}\right\}_{j=1}^{n_{y,t}}\right] \text{ which is } n_{x_2} \times n_{y,t}$$

$$\mathbf{v}_t \equiv \begin{bmatrix} v_{t,1} \\ v_{t,2} \\ \dots \\ v_{t,n_{y,t}} \end{bmatrix}$$

Thus

$$\mathbf{u}_t = \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \mathbf{v}_t$$

The mean and the variance of $\mathbf{u}_t$ are stated above. The expression for conditional autocorrelation in $\mathbf{u}_t$ is

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = E\left[\left| \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \mathbf{v}_t \mathbf{v}_{t-k}' \mathbf{Z}_{t-k}' \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right.\right]$$

$$= \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t E\left[\left| \mathbf{v}_t \mathbf{v}_{t-k}' \right. \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right] \mathbf{Z}_{t-k}' \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1}$$

$\Updownarrow$

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \mathbf{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) \mathbf{Z}_{t-k}' \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1}$$

where $E\left[\left| \mathbf{v}_t \mathbf{v}_{t-k}' \right. \mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right] \equiv \mathbf{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$. if $\mathbf{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) = \mathbf{\Omega}_{\mathbf{vv},-k}$ then

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \mathbf{\Omega}_{\mathbf{vv},-k} \mathbf{Z}_{t-k}' \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1}$$

The unconditional conditional autocorrelation is (given $E\left[\mathbf{v}_t \mathbf{v}_{t-k}'\right] \equiv \mathbf{\Omega}_{\mathbf{vv},-k}$)

$$Cov\left(\mathbf{u}_t, \mathbf{u}_{t-k}\right) = E\left[\left(\mathbf{A}_{t,o}^{\mathbf{x}_2}\right)^{-1} \mathbf{Z}_t \mathbf{\Omega}_{\mathbf{vv},-k} \mathbf{Z}_{t-k}' \left(\mathbf{A}_{t-k,o}^{\mathbf{x}_2}\right)^{-1}\right]$$

The conditional autocorrelation can be estimated by

$$\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\hat{\mathbf{A}}_t^{\mathbf{x}_2}\right)^{-1} \hat{\mathbf{Z}}_t \hat{\mathbf{\Omega}}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) \hat{\mathbf{Z}}_{t-k}' \left(\hat{\mathbf{A}}_{t-k}^{\mathbf{x}_2}\right)^{-1}$$

where $\hat{\mathbf{Z}}_t \equiv \frac{1}{n_{y,t}} \left[\left\{ \frac{\partial g_j\left(\mathbf{x}_{1,t}, \hat{\mathbf{x}}_{2,t}; \hat{\boldsymbol{\theta}}_1\right)}{\partial \hat{\mathbf{x}}_{2,t}} \frac{1}{V\left(\mathbf{z}_{t,j}; \hat{\boldsymbol{\gamma}}\right)} \right\}_{j=1}^{n_{y,t}}\right]$. $\hat{\mathbf{\Omega}}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$ is some consistent estimator of $\mathbf{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right)$, for instance by a multivariate GARCH model. If we impose the assumption of $\mathbf{\Omega}_{\mathbf{vv},-k}\left(\mathbf{x}_{1,t}, \mathbf{z}_{t,j}\right) = \mathbf{\Omega}_{\mathbf{vv},-k}$ then

$$\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t-k} \left| \mathbf{x}_{1,t}, \mathbf{z}_{t,j} \right.\right) = \left(\hat{\mathbf{A}}_t^{\mathbf{x}_2}\right)^{-1} \hat{\mathbf{Z}}_t \hat{\mathbf{\Omega}}_{\mathbf{vv},-k} \hat{\mathbf{Z}}_{t-k}' \left(\hat{\mathbf{A}}_{t-k}^{\mathbf{x}_2}\right)^{-1}$$

where $\hat{\mathbf{\Omega}}_{\mathbf{vv},-k} = \frac{1}{T-k} \sum_{t=k+1}^{T} \hat{\mathbf{v}}_t \hat{\mathbf{v}}_{t-k}'$. Finally, the unconditinal autocorrelation can be estimated by (given $E\left[\mathbf{v}_t \mathbf{v}_{t-k}'\right] \equiv \mathbf{\Omega}_{\mathbf{vv},-k}$)

$$\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t-k}\right) = \widehat{(\mathbf{A}^{\mathbf{x}_2})^{-1}\mathbf{Z}} \hat{\mathbf{\Omega}}_{\mathbf{vv},-k} \widehat{(\mathbf{A}^{\mathbf{x}_2})^{-1}\mathbf{Z}}'$$

where $\widehat{(\mathbf{A}^{\mathbf{x}_2})^{-1}\mathbf{Z}} = \frac{1}{T-k} \sum_{t=k+1}^{T} \left(\hat{\mathbf{A}}_t^{\mathbf{x}_2}\right)^{-1} \hat{\mathbf{Z}}_t$

H The dynamic term structure model for the Monte Carlo study

We use the following linear and Gaussian term structure model for our Monte Carlo study

$$y_t(\tau) = x_{1,t} + x_{2,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + x_{3,t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + v_{t,\tau}$$

$$\mathbf{x}_{t+1} = \boldsymbol{\alpha} + \mathbf{h}_x \mathbf{x}_t + \mathbf{w}_{t+1}$$

where $v_{t,\tau} \sim \mathcal{NID}\left(0, Var\left(\mathbf{v}_{t,\tau}\right)\right)$ and $\mathbf{w}_t \equiv \begin{bmatrix} w_{1,t} & w_{2,t} & w_{3,t} \end{bmatrix}' \sim \mathcal{NID}\left(0, Var\left(\mathbf{w}_t\right)\right)$. Moreover, $\mathbf{x}_0$, $\mathbf{v}_t$, and $\mathbf{w}_t$ are mutually uncorrelated at all leads and lags. Diebold et al. (2006) estimate this model based on US data (Jan. 1972 - 2000 Dec.) for 15 zero-coupon yields with maturities between 3 and 120 months. They find $\lambda = 0.077$ and

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0.115 \\ 0.171 \\ -0.279 \end{bmatrix}$$

$$\mathbf{h}_x = \begin{bmatrix} 0.99 & 0.03 & -0.02 \\ -0.03 & 0.94 & 0.04 \\ 0.03 & 0.02 & 0.84 \end{bmatrix}$$

$$Var\left(\mathbf{w}_t\right) = \begin{bmatrix} 0.09 & -0.01 & 0.04 \\ -0.01 & 0.38 & 0.01 \\ 0.04 & 0.01 & 0.80 \end{bmatrix}$$

Figure 1: The RMSE for the estimates of the latent factors

This figure reports the RMSE for the estimates of the three latent factors in the dynamic Nelson-Siegel model. These RMSE are calculated based on 500 repetitions of a sample of 480 observations. The black lines with stars refer to the Kalman smoother and are computed for 5, 10, 15, 20, 25, 50, 100, 150, and 200 bonds. The red lines with circle refer to the regression filter and are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. Case 1 refers to the scenario with measurement errors of 10 basis points along the yield curve, and Case 2 refers to the scenario with measurement errors of 20 basis points along the yield curve.

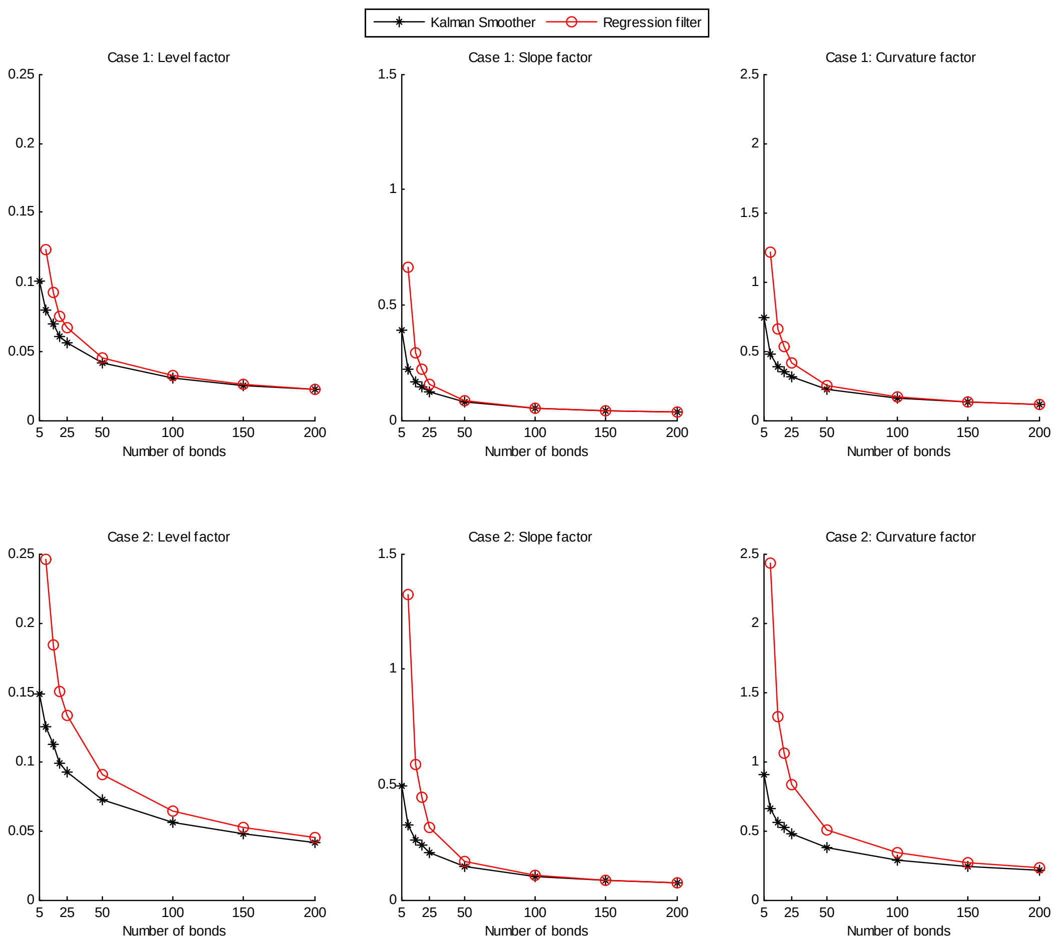


Figure 2: The computing time

This top graph reports the average number of seconds used to calculate the Kalman smoother (the black line with stars) and the regression filter (the red line with circles). The buttom graph displays the computational gain of using the regression filter instead of the Kalman smoother. That is we report the ratio of the average number of seconds for the Kalman smoother to the average number of seconds for the regression filter.

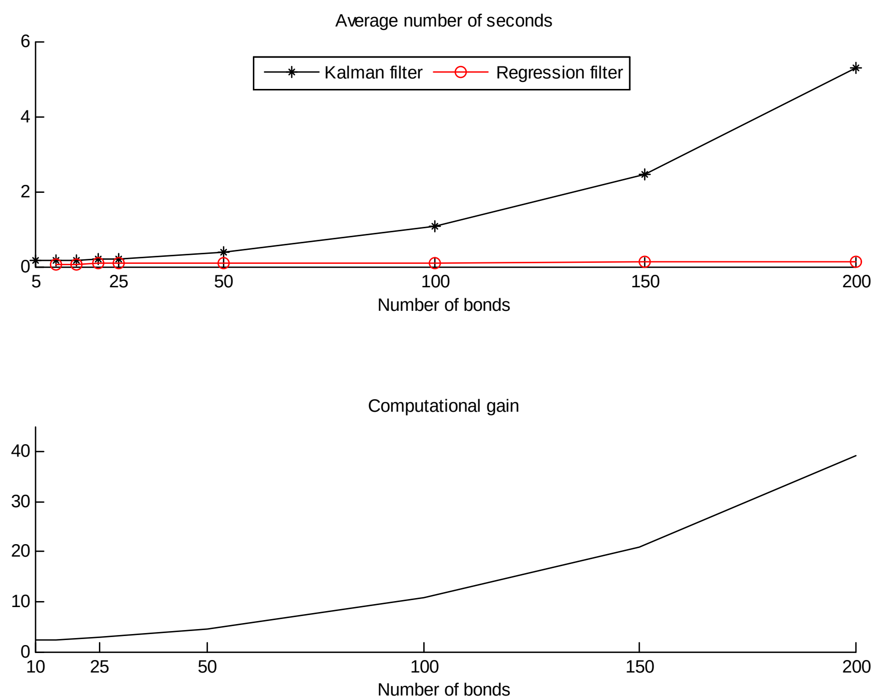


Figure 3: Biases when estimating lambda

This figure reports the biases when estimating lambda in the dynamic Nelson-Siegel model. These results are calculated based on 1000 repetitions of a sample of 480 observations. The black lines with stars refer to the ML estimates and are computed for 5, 10, 15, 20, 25, 50, and 100 bonds. The red lines with circles refer to the SR estimates and are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. Case 1 refers to the scenario with measurement errors of 10 basis points along the yield curve, and Case 2 refers to the scenario with measurement errors of 20 basis points along the yield curve.

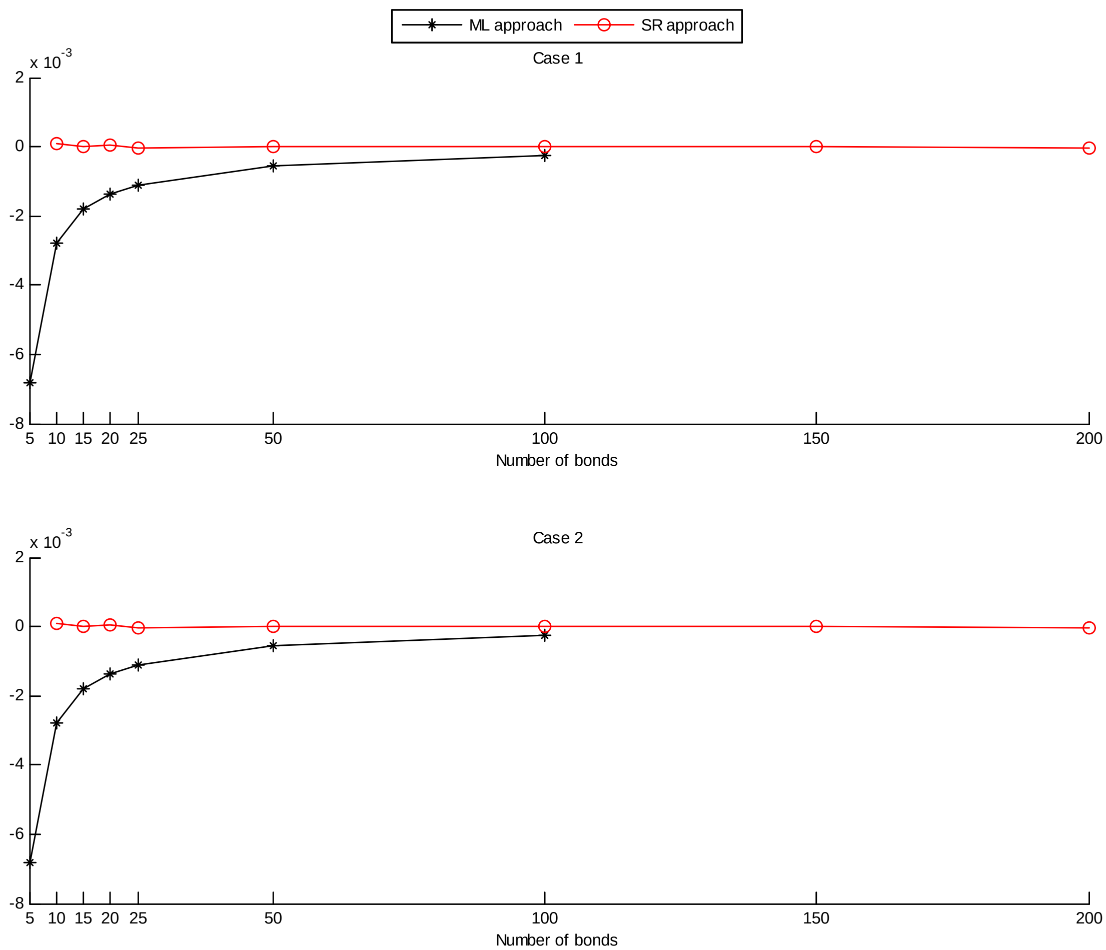


Figure 4: The true standard error when estimating lambda

This figure reports the true standard errors when estimating lambda in the dynamic Nelson-Siegel model. These results are calculated based on 1000 repetitions of a sample of 480 observations. The black lines with stars refer to the ML estimates and are computed for 5, 10, 15, 20, 25, 50, and 100 bonds. The red lines with circles refer to the SR estimates and are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. Case 1 is to the scenario with measurement errors of 10 basis points along the yield curve, and Case 2 is to the scenario with measurement errors of 20 basis points along the yield curve.

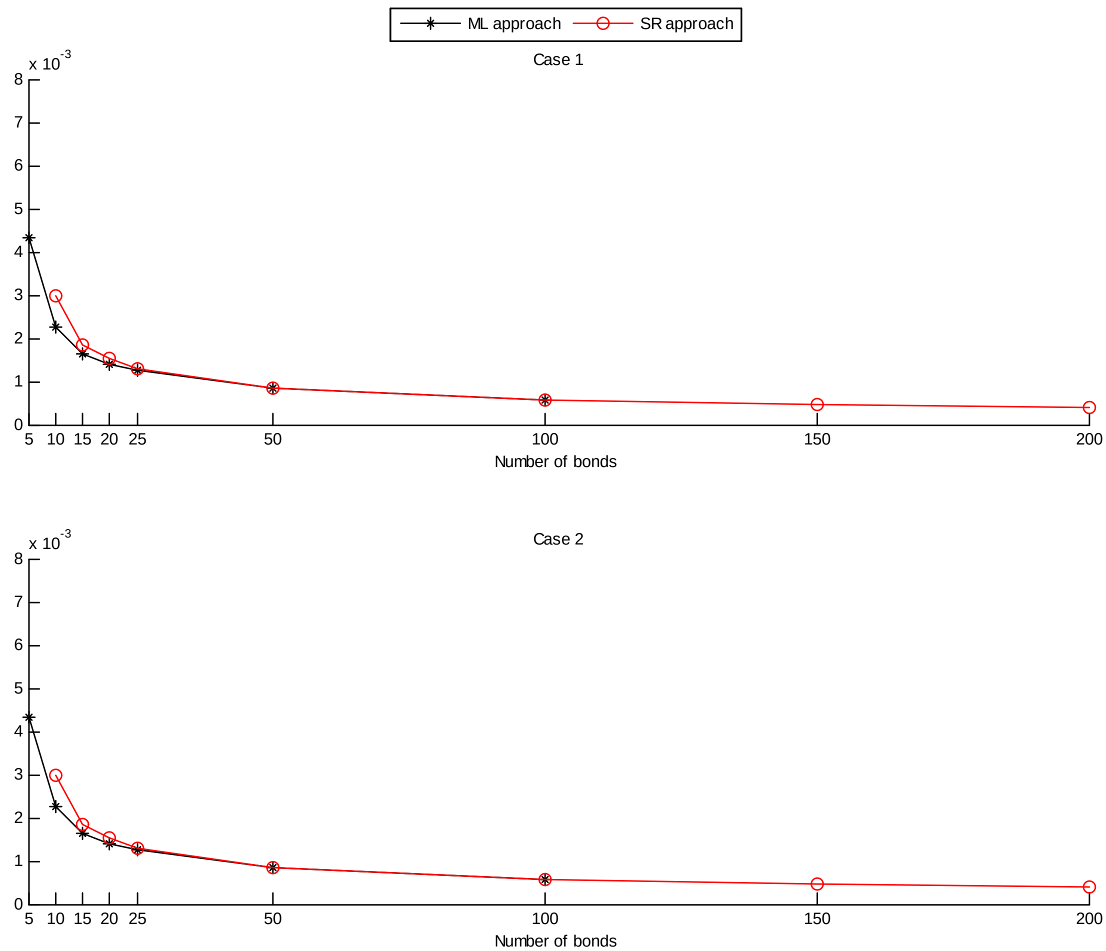


Figure 5: The biases in estimating the standard error of lambda

This figure reports biases in estimating the standard error of lambda in the dynamic Nelson-Siegel model. These results are calculated based on 1000 repetitions of a sample of 480 observations. The black lines with star refer to the ML estimates using the outer product of the score function and are computed for 5, 10, 15, 20, 25, 50, and 100 bonds. The red lines with circles refer to the heteroskedastic robust estimates in the SR approach. The green lines with squares refer to the non-heteroskedastic robust estimates in the SR approach. The SR estimates are are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. Case 1 is to the scenario with measurement errors of 10 basis points along the yield curve, and Case 2 is to the scenario with measurement errors of 20 basis points along the yield curve.

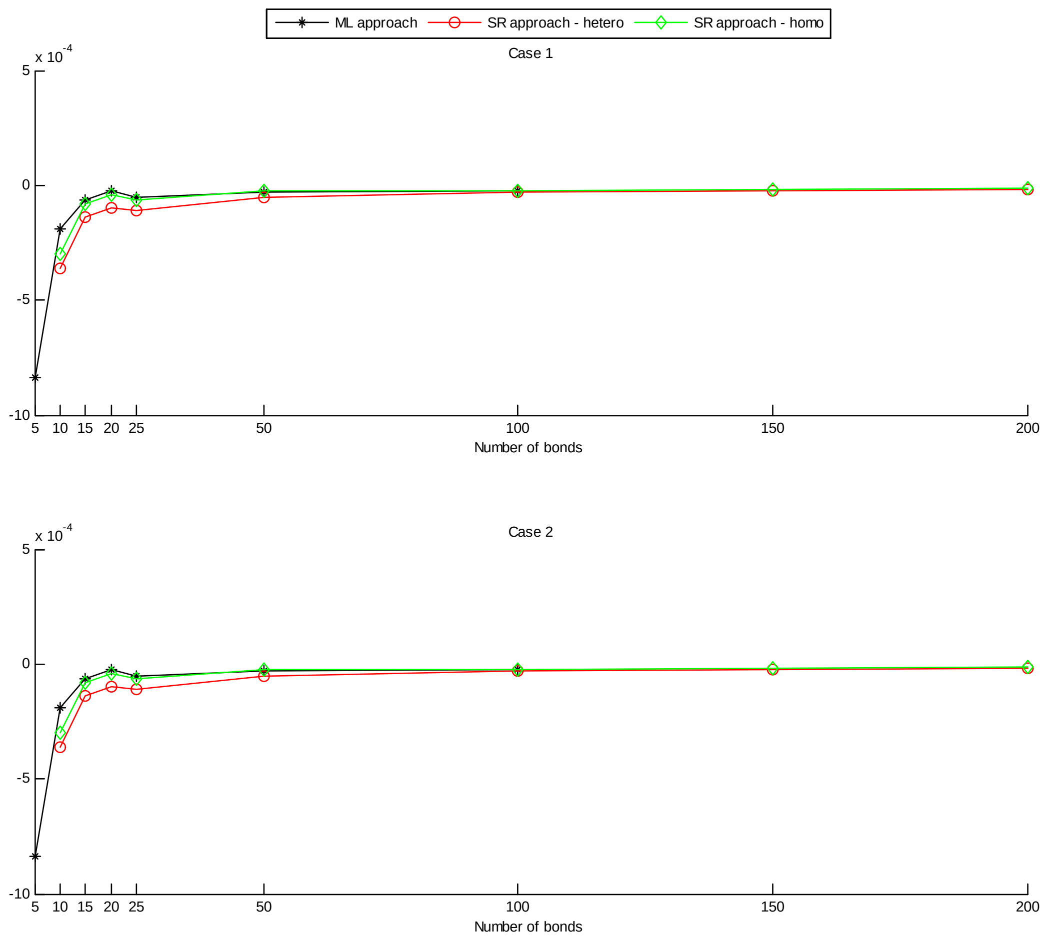


Figure 6: Biases when estimating the Dynamic Nelson-Siegel model

This figure reports biases when estimating all the parameters in the dynamic Nelson-Siegel model. These results are calculated based on 1000 repetitions of a sample of 480 observations and with measurement errors of 10 basis points along the yield curve. The black lines with stars refer to the ML estimates and are computed for 5, 10, 15, 20, and 25 bonds. The red lines with circles refer to the SR estimates and are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. The black pluses refer to the infeasible ML estimates when the factors are observed, or equivalently, when there are an infinite number of observables each time period.

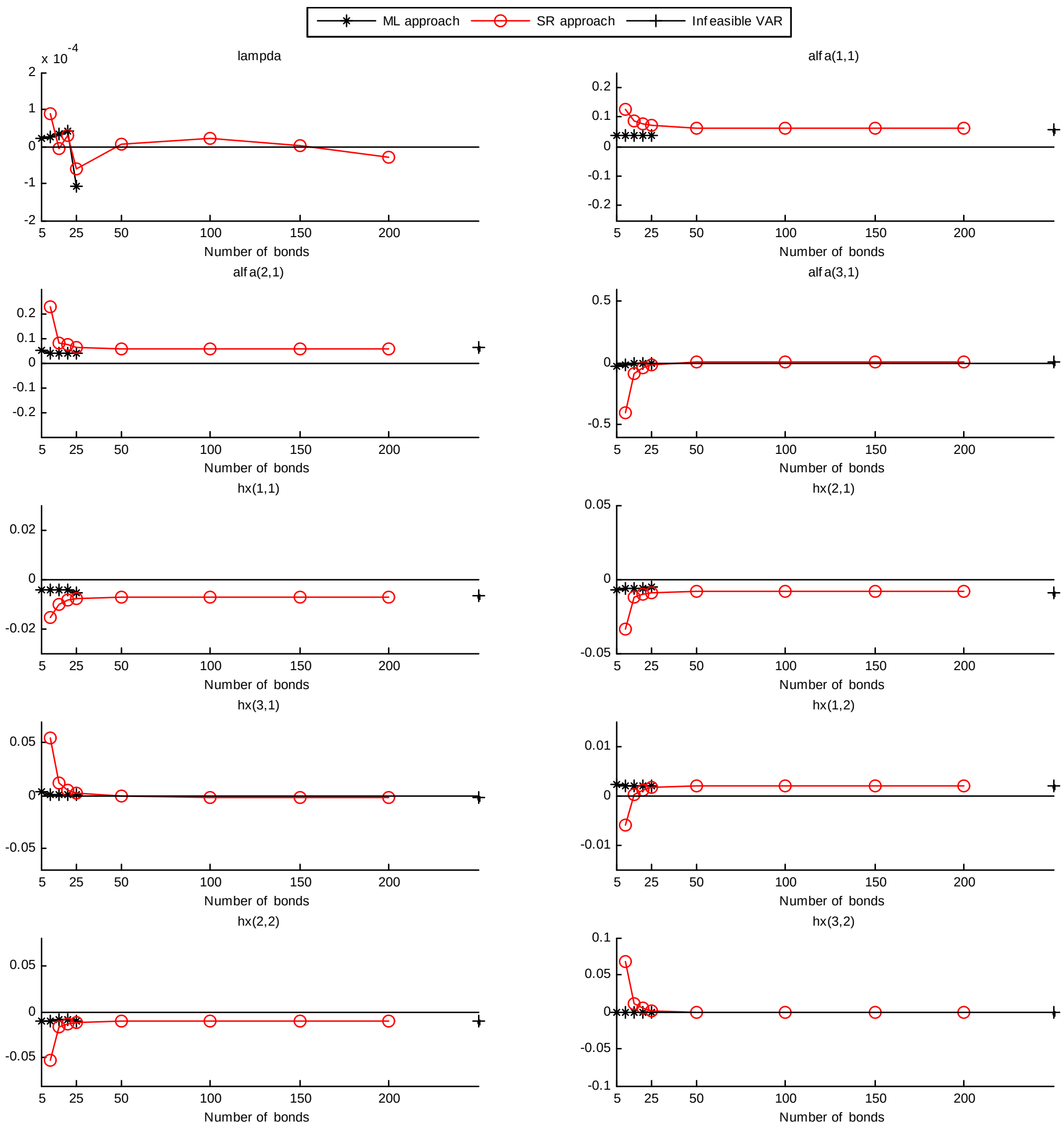


Figure 6: Continued

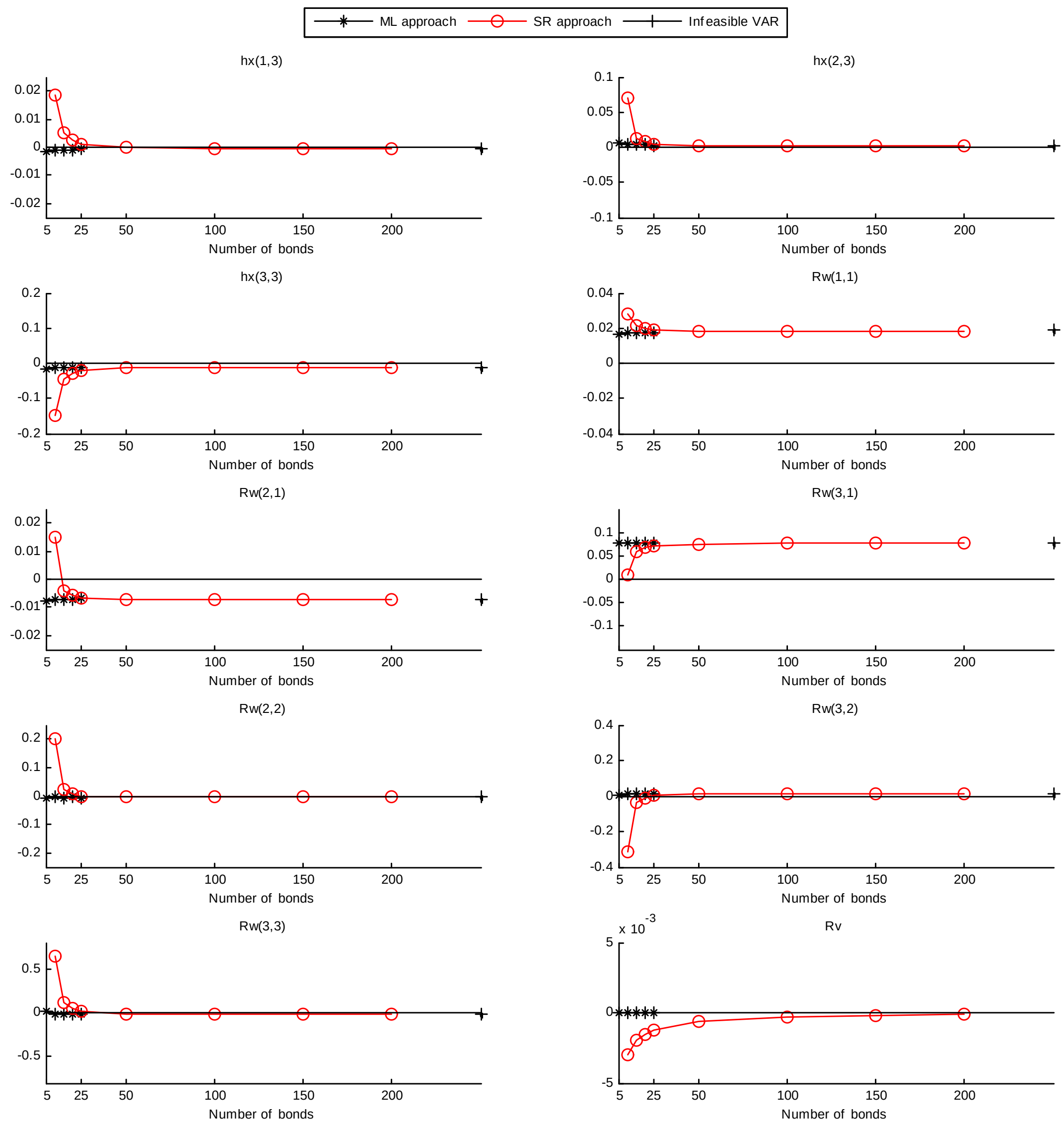


Figure 7: The true standard errors for the estimates in the Dynamic Nelson-Siegel model
This figure reports true standard errors for the estimates in the dynamic Nelson-Siegel model. These results are calculated based on 1000 repetitions of a sample of 480 observations and with measurement errors of 10 basis points along the yield curve. The black lines with stars refer to the ML estimates and are computed for 5, 10, 15, 20, and 25 bonds. The red lines with circles refer to the SR estimates and are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. The black pluses refer to the infeasible ML estimates when the factors are observed, or equivalently, when there are an infinite number of observables each time period.

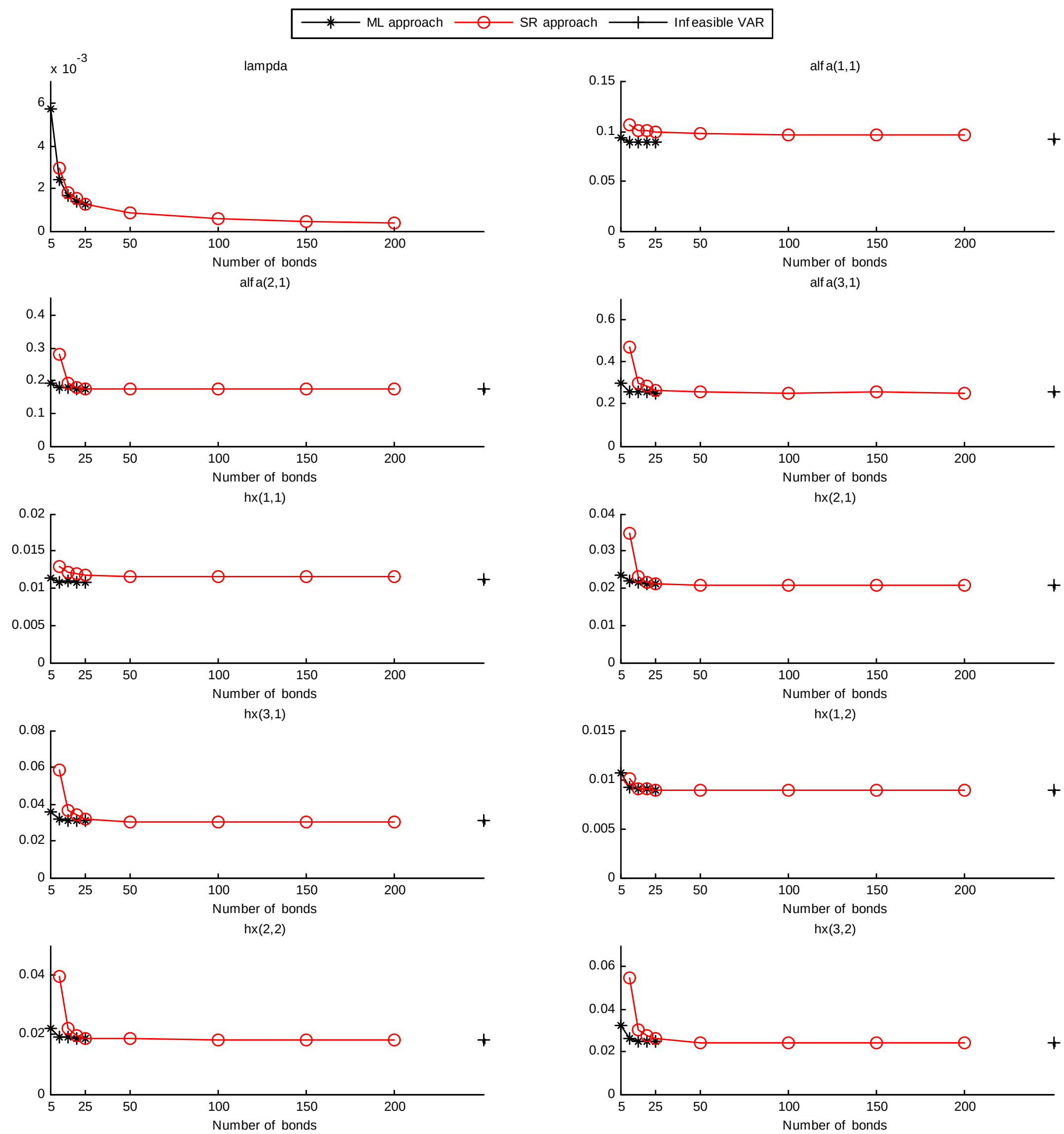


Figure 7: Continued

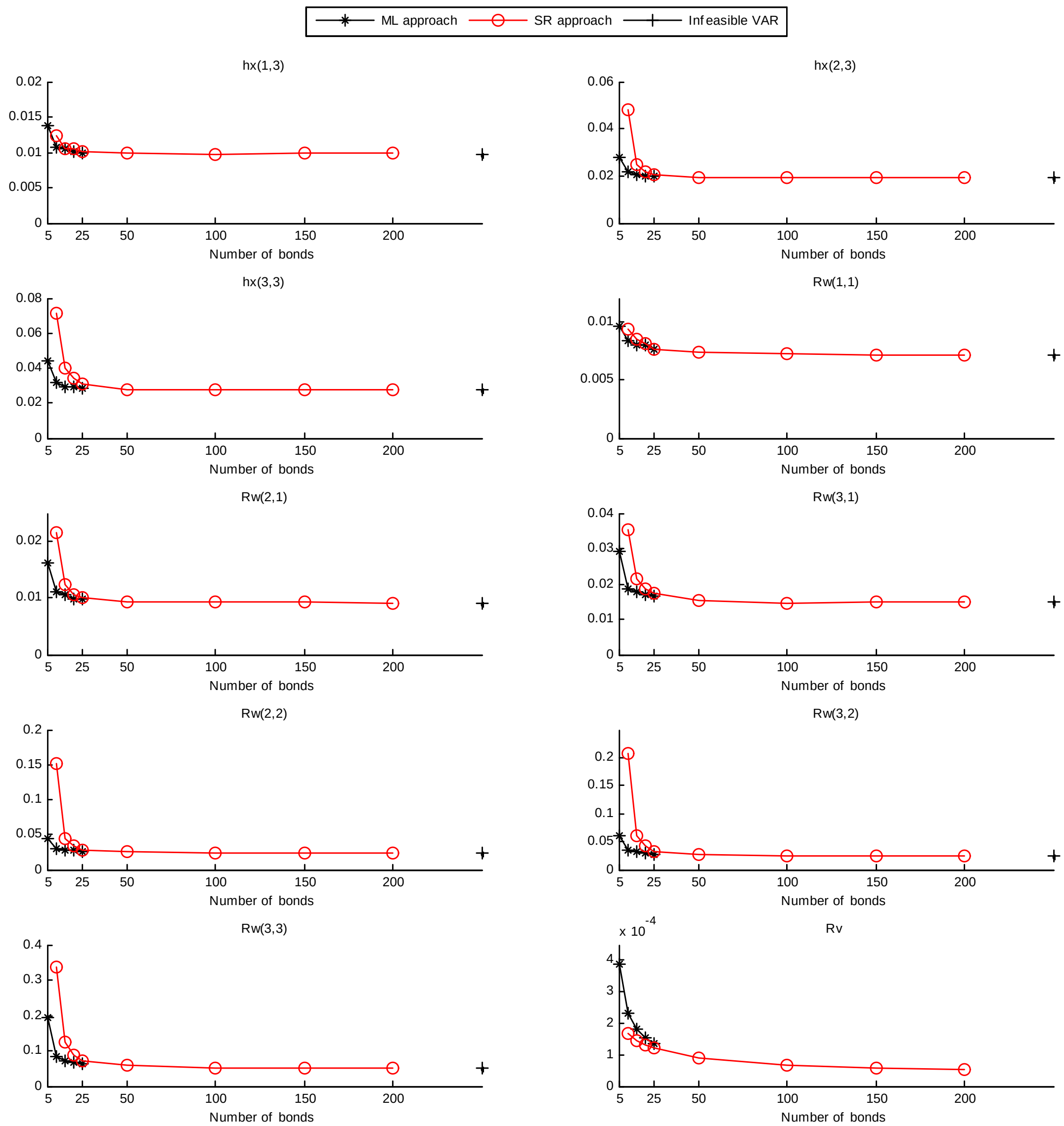


Figure 8: Biases in the estimates of standard errors in the Dynamic Nelson-Siegel model
This figure reports biases when estimating the standard errors for the estimates in the dynamic Nelson-Siegel model. These results are calculated based on 1000 repetitions of a sample of 480 observations and with measurement errors of 10 basis points along the yield curve. The black lines with stars refer to the ML estimates and are computed for 5, 10, 15, 20, and 25 bonds. The red lines with circles refer to the SR estimates and are computed for 10, 15, 20, 25, 50, 100, 150, and 200 bonds. The black pluses refer to the infeasible ML estimates when the factors are observed, or equivalently, when there are an infinite number of observables each time period.

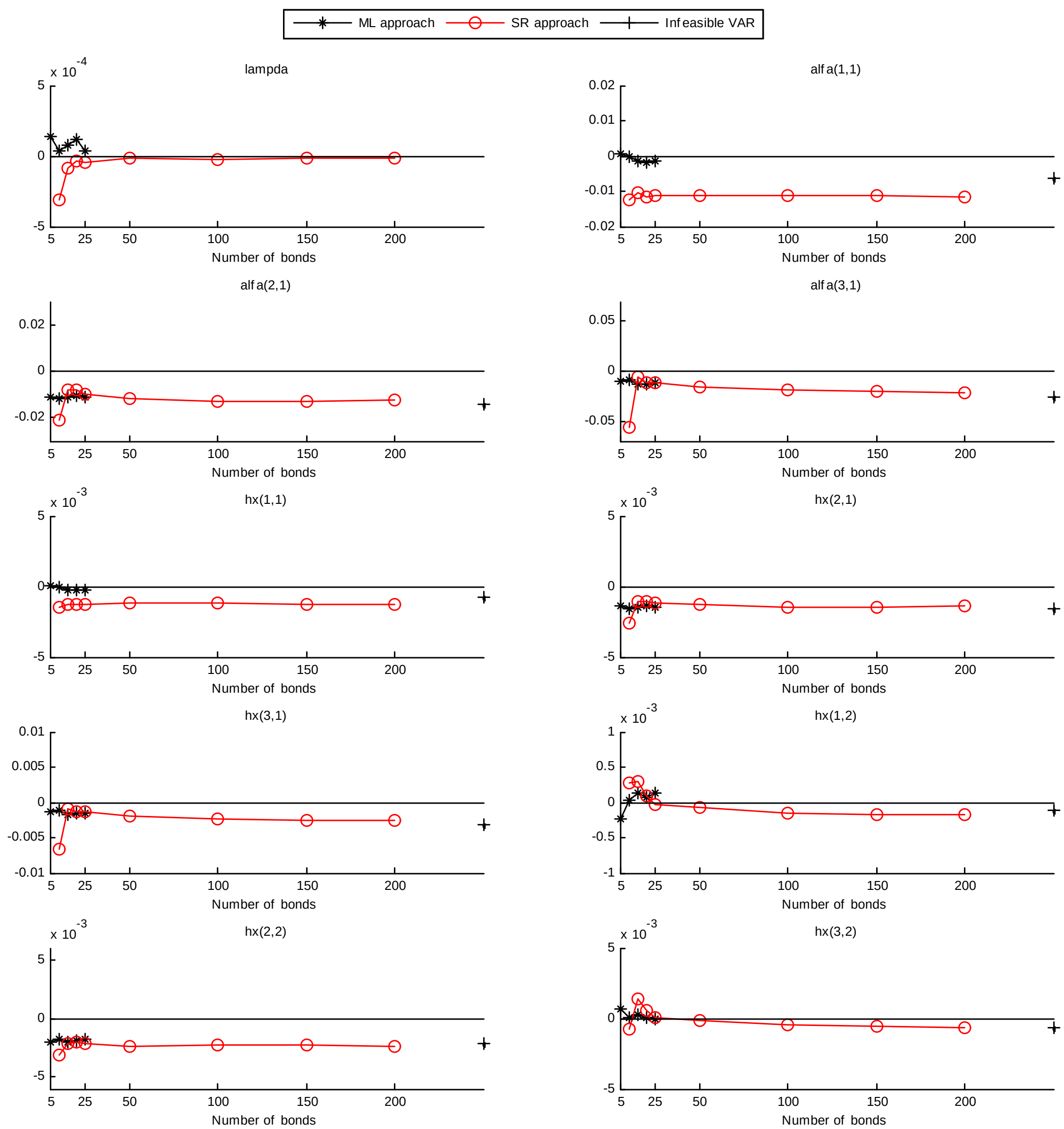
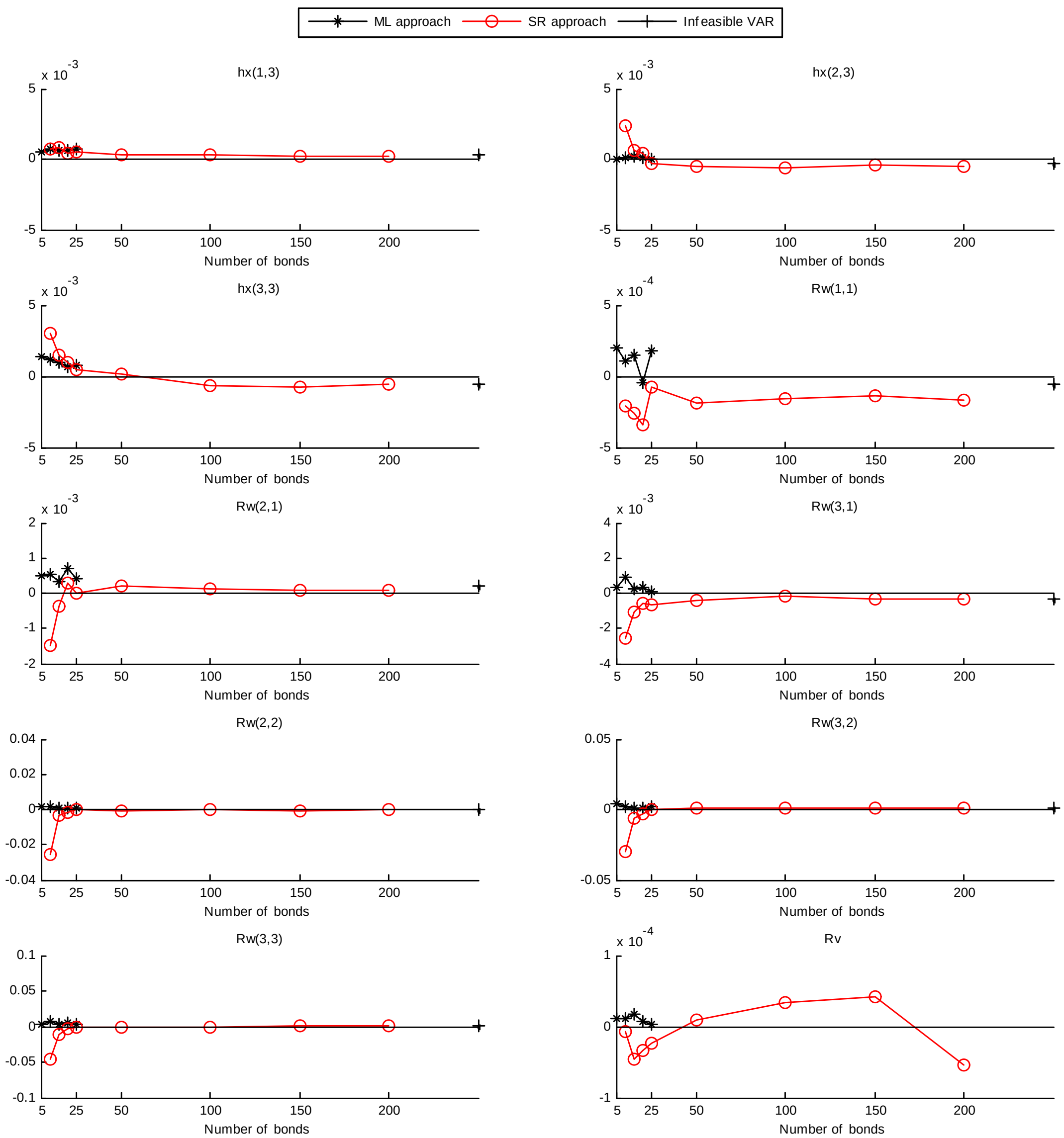


Figure 8: Continued



Research Papers 2010

Center for Research in Econometric Analysis of Time Series

2009-58: Olaf Posch: Risk premia in general equilibrium

2009-59: Christian M. Dahl and Emma M. Iglesias: Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary

2009-60: Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij: Limit theorems for functionals of higher order differences of Brownian semi-stationary processes

2010-01: Anders Bredahl Kock and Timo Teräsvirta: Forecasting with nonlinear time series models

2010-02: Gunnar Bårdsen, Stan Hurn and Zoë McHugh: Asymmetric unemployment rate dynamics in Australia

2010-03: Jesper Rangvid, Maik Schmeling and Andreas Schrimpf: Cash Flow-Predictability: Still Going Strong

2010-04: Helle Bunzel and Walter Enders: The Taylor Rule and "Opportunistic" Monetary Policy

2010-05: Martin M. Andreasen: Non-linear DSGE Models and The Optimized Particle Filter

2010-06: Søren Johansen and Bent Nielsen: Discussion of The Forward Search: Theory and Data Analysis by Anthony C. Atkinson, Marco Riani, and Andrea Ceroli

2010-07: Giuseppe Cavaliere, Anders Rahbek and A.M.Robert Taylor: Bootstrap Sequential Determination of the Co-integration Rank in VAR Models

2010-08: Peter R. Hansen and Asger Lunde: Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error

2010-09: Tom Engsted, Thomas Q. Pedersen and Carsten Tanggaard: Pitfalls in VAR based return decompositions: A clarification

2010-10: Torben G. Andersen and Luca Benzoni: Stochastic Volatility

2010-11: Torben B. Rasmussen: Affine Bond Pricing with a Mixture Distribution for Interest Rate Time-Series Dynamics

2010-12: Martin M. Andreasen and Bent Jesper Christensen: The SR Approach: a new Estimation Method for Non-Linear and Non-Gaussian Dynamic Term Structure Models