Penalized Indirect Inference

Francisco Blasques*
Artem Duplinskiy

* Tinbergen Institute

Faculty of Economics and Business Administration, VU University Amsterdam, the Netherlands.

Penalized Indirect Inference[*]

Francisco Blasques[(a,b)] and Artem Duplinskiy[(a)]

(a) VU University Amsterdam

(b) Tinbergen Institute

January 18, 2015

Abstract

Parameter estimates of structural economic models are often difficult to interpret at the light of the underlying economic theory. Bayesian methods have become increasingly popular as a tool for conducting inference on structural models since priors offer a way to exert control over the estimation results. This paper proposes a penalized indirect inference estimator that allows researchers to obtain economically meaningful parameter estimates in a frequentist setting. The asymptotic properties of the estimator are established for both correctly and incorrectly specified models. A Monte Carlo study reveals the role of the penalty function in shaping the finite sample distribution of the estimator. The advantages of using this estimator are highlighted in the empirical study of a state-of-the-art dynamic stochastic general equilibrium model.

JEL Codes: C15; C13; D58; E32.
Keywords: Penalized estimation, Indirect Inference, Simulation-based methods, DSGE.

1 INTRODUCTION

In economics, it is often difficult to reconcile the estimates obtained for parameters of structural models with the underlying economic theory. This problem is especially evident when employing frequentist estimation techniques that leave the researcher unable to exert some control over the estimator, at least within the parameter space bounds. Such problems are well known in the dynamic stochastic general equilibrium (DSGE) literature. Canova and Sala (2009) show that identification problems are pervasive in New Keynesian DSGE models; see also Ma (2002), Beyer and Farmer (2004), Nason and Smith (2008). Model misspecification also leads to difficulties since parameter estimates are biased and multiple pseudo-true parameters may exist.

In the case of DSGE models, the problem of model misspecification and the challenge posed by parameters that are unidentified, or only weakly identified, has lead to the widespread adoption of Bayesian methods. The introduction of priors in Bayesian estimation allows the researcher to exert control over the estimation procedure, in the effort to obtain economically meaningful parameter estimates, by including information that is not contained in the data sample.

For example, An and Schorfheide (2007) point out that *"Any estimation and evaluation method is confronted with the following challenges: potential model misspecification and possible lack of identification of parameters of interest."* and that Bayesian methods become useful since *"prior distributions can be used to incorporate additional information into the parameter estimation.".* Similarly, Fernandez-Villaverde (2010) defends the use of additional structure in estimation since *"Pre-sample information is often amazingly rich and considerably useful and not taking advantage of it is an unforgivable omission".* He further concludes that *"Yes, our inference would have depended heavily on the prior, but why is this situation any worse than not being able to say anything of consequence?".*

This paper provides a frequentist tool for incorporating pre-sample information in estimation. The added structure is provided by a penalty function that plays a role that is similar to that played by the prior in Bayesian estimation. Penalized estimation is not new in the frequentist literature. On the contrary, the basic idea of adding penalties to well known estimation criteria such as least squares, maximum likelihood or the method of moments has been present for a long time in many statistical ap-

plications; see e.g. Tibshirani (1996) and Zou (2006) and Liao (2013).[1] This paper 'brings' the use of penalties to a simulation-based estimation setting, which allows us to estimate complex, high-dimensional, nonlinear dynamic models with unobserved variables, like DSGE models. In particular, we introduce a penalty in the criterion function of the indirect inference (II) estimator proposed by Gourieroux et al. (1993) and Smith (1993). We believe that, compared to Bayesian estimators, the penalized indirect inference (PII) estimator offers a number of important advantages.

First, while Bayesian estimation can only rely on the likelihood function, the PII estimator allows for a wide range of criterion functions. Specifically, the PII estimator encompasses both limited information estimation methods, such as the simulated method of moments of Duffie and Singleton (1990) and Lee and Ingram (1991), as well as likelihood-based full information estimation methods, such as the efficient method of Gallant and Tauchen (1996). Popular indirect inference criteria include moment matching, matching VAR parameters, and matching impulse response functions at certain periods; see e.g. Christiano et al. (2005), Ruge-Murcia (2007) and Dupor et al. (2009) for examples of these different approaches. The ability to choose among different criteria is a crucial advantage for estimation of potentially misspecified models. Indeed, Fukac and Pagan (2010) argue for the use of limited information estimation techniques in DSGE models since the maximum likelihood (ML) estimator requires the entire probabilistic structure of the model to be well specified, rather than just a few features of interest. Furthermore, ML estimators may have poor robustness properties; see e.g. the seminal work of Huber (1967, 1974).

Second, unlike the prior in Bayesian methods, the penalty function is not restricted to being a density. The penalty function can take a wide range of forms and thus it is easy to incorporate various forms of prior information in a flexible way. As we shall see, the asymptotic properties of PII estimators can be established under very mild regularity conditions on the nature of the penalty function.

Third, the influence of the penalty is allowed to vanish asymptotically at any pre-specified rate. Depending on this rate, the penalty may or may not influence the asymptotic distribution of the estimator. These considerations are in stark contrast with Bayesian estimation where the influence of the prior is always vanishing

---

[1]In some sense, penalized estimation also plays a crucial role in many non-parametric; semi-parametric and semi-nonparametric estimation methods; see e.g. Chen (2007), Dalalyan et al. (2006), Green (1996).

asymptotically at a rate that is not controlled by the researcher.

The remainder of the article is organized as follows. Section 2 introduces the PII estimator. Section 3 establishes its asymptotic properties. Section 4 analyses finite-sample properties by means of a Monte Carlo exercise. Section 5 applies the new PII estimator to a state-of-the-art New-Keynesian DSGE model considered in Smets and Wouters (2007). Section 6 concludes.

2 The Penalized Indirect Inference Estimator

Following Gourieroux et al. (1993), we let $\widehat{\boldsymbol{\beta}}_T$ denote a q-dimensional vector of auxiliary statistics describing the sample of observed data $\mathbf{x}_1, ..., \mathbf{x}_T$. Similarly, $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ is the vector of auxiliary statistics that describes the artificial data simulated from the structural model of interest. The structural model is parameterized by the p-dimensional vector $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. The auxiliary statistics vector $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ is defined as an average of S vectors $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ obtained from S streams $\{\widetilde{x}_{1,s}, ..., \widetilde{x}_{T,s}\}_{s=1}^{S}$ of simulated data,

$$\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^{S} \widetilde{\boldsymbol{\beta}}_{T,s}(\boldsymbol{\theta}).$$

The penalized indirect inference (PII) estimator $\widehat{\boldsymbol{\theta}}_{T,S}$ is then defined as,

$$\widehat{\boldsymbol{\theta}}_{T,S} \in \arg\min_{\boldsymbol{\theta} \in \Theta} \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right)' \Omega \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right) + \pi_T(\boldsymbol{\theta}) ,$$

where $\pi_T : \Theta \to [0, \infty)$ is a penalty function that depends on sample size T; and Ω is a weighting matrix. Note that the additive nature of the penalty can be easily generalized. In particular, the theory considered here extends naturally to the case of a multiplicative penalty function by means of a log transformation.

Consider the case of a scalar parameter $\boldsymbol{\theta} = \theta \in [0, 1]$. A trivially simple example of a penalty function is given by,

$$\pi_T(\theta) = \frac{1}{T}\left(\theta - \frac{1}{2}\right)^2.$$

This quadratic penalty function penalizes deviations of $\widehat{\boldsymbol{\theta}}_T$ from the 'central' point $1/2$ of the parameter space $[0, 1]$. This penalty function vanishes at speed T asymptotically uniformly in the parameter space $[0, 1]$ since $\sup_{\theta \in [0,1]} \pi_T(\theta) = O(T^{-1})$ as

$T \to \infty$.

Section 2 provides simple examples of penalty functions for parameters in $\mathbb{R}$ and $\mathbb{R}^2$. Penalty (b) is similar to that in (a) but uses its explosive behavior near $\boldsymbol{\theta} = 1$ to effectively restrict the parameter space from above to values $\boldsymbol{\theta} < 1$. Numerical optimization routines, which specify the admissible range for parameters or set the value of the objective function to a very large number (when parameters reach 'unreasonable' values), do exactly the same thing: introduce a penalty to restrict the parameter space (see e.g. Gorodnichenko and Ng, 2010). While penalty (c) disapproves of estimates near zero, penalty (d) excludes the region near zero altogether. Penalty (e) shows how to center the estimates at the point $\boldsymbol{\theta}^* = (1,1)$; and penalty (f) shows how to introduce 'soft' cross-restrictions in the parameter space that penalize deviations from the relation $\theta_1 = a\theta_2$ for some $a \in \mathbb{R}$ and $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$. Darker shades represent lower values for the penalty (closer to zero), whereas lighter shades indicate larger values for the penalty. Naturally, these considerations extend to arbitrary dimensions.
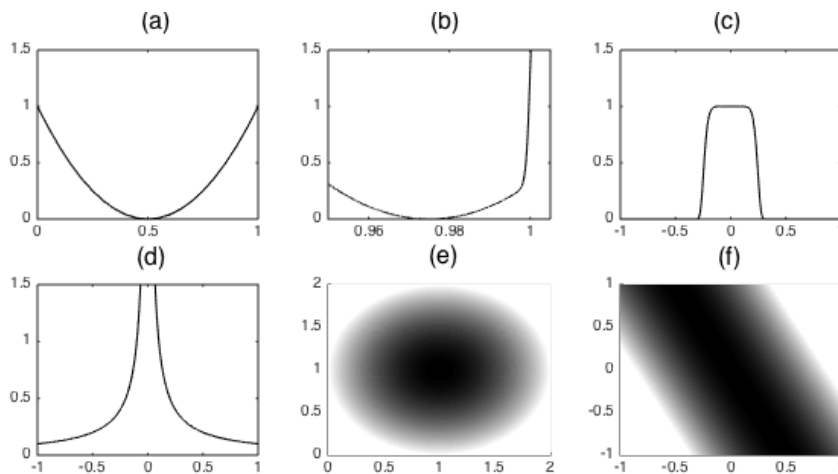
Figure 1: This figure presents examples of penalty functions in $\mathbb{R}$ and $\mathbb{R}^2$. Penalty (a) and (b) center the parameter estimate at the point $\boldsymbol{\theta}^* = 0.5$. Penalties (c) and (d) penalize values close to zero. Finally, penalties (e) and (f) show how to center the estimates at the point in $\mathbb{R}^2$.

Penalty functions π_T that vanish asymptotically, are interesting when the researcher wishes to impose penalties only for small sample sizes. This can be desirable when the estimator $\widehat{\boldsymbol{\theta}}_{T,S}$ has poor finite sample properties. Poor finite sample behavior of $\widehat{\boldsymbol{\theta}}_{T,S}$ can occur, for example, when the structural parameter is weakly identified in small samples. This is often the case in structural economic models; see Canova

and Sala (2009) for an analysis of a new-Keynesian model where parameters are identified but only weakly so. As we shall see, under appropriate conditions, when π_T vanishes asymptotically at an appropriate speed one can still obtain consistency of $\widehat{\boldsymbol{\theta}}_{T,S}$ to the true parameter $\boldsymbol{\theta}_0$.

3 ASYMPTOTIC PROPERTIES

In this section, it is convenient to divide PII estimators into two main classes: those with penalty function that vanishes as $T \to \infty$, and those for which penalty function does not vanish. For PII estimators featuring penalty functions that do not vanish asymptotically, we establish the convergence of $\widehat{\boldsymbol{\theta}}_{T,S}$ as $T \to \infty$ to a pseudo-true parameter $\boldsymbol{\theta}_0^*$ that minimizes the limit criterion function. For PII estimators with penalty functions that vanish asymptotically, we establish the convergence of $\widehat{\boldsymbol{\theta}}_{T,S}$ to the true parameter $\boldsymbol{\theta}_0$ as $T \to \infty$.

The class of PII estimators with vanishing penalty functions can be further categorized according the rate at which the penalty function vanishes. In general, if the penalty function vanishes sufficiently fast, then the finite sample distribution of the the PII estimator is affected by the penalty, but the asymptotic distribution is not. If the penalty vanishes at a slow rate, then the asymptotic distribution of the PII estimator is also affected.

3.1 EXISTENCE AND CONSISTENCY

Let Θ and $\mathcal{B}$ denote the structural parameter space, and auxiliary parameter space, respectively. Below, given a topological space $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$, we let $\mathfrak{B}(\mathbb{A})$ denote the Borel σ-algebra generated by the topology $\mathcal{T}_{\mathbb{A}}$ of $\mathbb{A}$.

ASSUMPTION 1. $(\Theta, \mathfrak{B}(\Theta))$ *and* $(\mathcal{B}, \mathfrak{B}(\mathcal{B}))$ *are measurable spaces and* Θ *is compact.*

In order to ensure the existence and measurability of the PII estimator we assume also that the penalty function π_T is continuous in $\boldsymbol{\theta} \in \Theta$ for every T and that the auxiliary estimator $\widetilde{\boldsymbol{\beta}}_{T,S}$ is almost surely (a.s.) continuous in $\boldsymbol{\theta}$ for every S and T.

ASSUMPTION 2. $\pi_T \in \mathbb{C}(\Theta, \mathbb{R})$ *and* $\widetilde{\boldsymbol{\beta}}_{T,S} \in \mathbb{C}(\Theta, \mathcal{B})$ *a.s.* $\forall (T, S) \in \mathbb{N} \times \mathbb{N}$.

The following theorem establishes the existence and $\mathfrak{B}(\Theta)/\mathcal{F}$-measurability of the PII estimator where $\mathcal{F}$ denotes the σ-algebra in the underlying complete probability space of interest $(\mathcal{E}, \mathcal{F}, \mathbb{P})$.[2]

THEOREM 1. *Let Assumptions Assumptions [1] and [2] hold. Then there exists an $\mathfrak{B}(\Theta)/\mathcal{F}$-measurable map $\widehat{\boldsymbol{\theta}}_{T,S} : \mathcal{E} \to \Theta$ satisfying*

$$\widehat{\boldsymbol{\theta}}_{T,S} \in \arg\min_{\boldsymbol{\theta} \in \Theta} \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\right)' \Omega\left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\right) + \pi_T(\boldsymbol{\theta}) \ \forall \ (T, S) \in \mathbb{N}^2.$$

Following Gourieroux et al. (1993) we obtain the consistency of the PII estimator $\widehat{\boldsymbol{\theta}}_{T,S}$ by building on the consistency of the auxiliary estimators. This is established in the following assumption.

ASSUMPTION 3. $\|\widehat{\boldsymbol{\beta}}_T - \boldsymbol{b}_0\| \xrightarrow{a.s.} 0$ *and* $\sup_{\boldsymbol{\theta} \in \Theta} \|\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{b}(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$ *as* $T \to \infty \ \forall \ S \in \mathbb{N}$,

where $\boldsymbol{b} : \Theta \to \mathcal{B}$ is the binding function. The PII estimator requires further that the penalty function converge asymptotically to a well defined limit function $\pi : \Theta \to [0, \infty)$.

ASSUMPTION 4. $\sup_{\boldsymbol{\theta} \in \Theta} \|\pi_T(\boldsymbol{\theta}) - \pi(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$ *as* $T \to \infty$.

Finally, we follow the extremum estimation literature in assuming the identifiable uniqueness of the pseudo-true parameter of interest; see e.g. White (1999). In other words, we assume that there exists a well separated unique minimizer of the limit criterion function. Under an axiom of correct specification this assumption can be substituted by the usual injectivity of the binding function; see e.g. Gourieroux et al. (1993).

ASSUMPTION 5. *There exists a unique point $\boldsymbol{\theta}_0^* \in \Theta$ that minimizes the limit criterion function; i.e.*

$$\left(\boldsymbol{b}_0 - b(\boldsymbol{\theta}_0^*)\right)' \Omega\left(\boldsymbol{b}_0 - b(\boldsymbol{\theta}_0^*)\right) + \pi(\boldsymbol{\theta}_0^*) \ < \ \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta})\right)' \Omega\left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta})\right) + \pi(\boldsymbol{\theta}) \quad \forall \ \boldsymbol{\theta} \in \Theta.$$

The following theorem establishes the consistency of the PII estimator w.r.t. the pseudo-true parameter $\boldsymbol{\theta}_0^*$ that provides the best approximation of the structural model to the DGP as judged by the limit indirect inference criterion.

---

[2]An implicit assumption is maintained that the auxiliary estimators are well defined random variables.

THEOREM 2. *Let Assumptions 1 to 5 hold. Then $\widehat{\boldsymbol{\theta}}_{T,S} \xrightarrow{a.s.} \boldsymbol{\theta}_0^*$ as $T \to \infty \; \forall \; S \in \mathbb{N}$.*

As pointed out above, under an axiom of correct specification, the consistency of the PII estimator $\widehat{\boldsymbol{\theta}}_{T,S}$ w.r.t. the true parameter $\boldsymbol{\theta}_0$ can be obtained using the injectivity of the binding function instead of the assumed uniqueness in Assumption 5. The following assumption imposes the correct specification of the structural model by assuming that there exists a point $\boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{b}_0 = \boldsymbol{b}(\boldsymbol{\theta}_0)$ and ensures identification of the true parameter $\boldsymbol{\theta}_0$ by imposing the injectivity of the binding function $\boldsymbol{b} : \Theta \to \mathcal{B}$.

ASSUMPTION 6. *$\exists \boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{b}_0 = \boldsymbol{b}(\boldsymbol{\theta}_0)$ and $\boldsymbol{b} : \Theta \to \mathcal{B}$ is injective.*

Consistency w.r.t. the true parameter $\boldsymbol{\theta}_0$ can now be obtained as long as the penalty function behaves appropriately. Assumption 7 below considers three different cases of interest.

ASSUMPTION 7. *The penalty function π_T satisfies one of the following conditions*

(i) $\pi_T(\boldsymbol{\theta}_0) = 0 \; \forall \; T \in \mathbb{N}$; or

(ii) $\pi_T(\boldsymbol{\theta}_0) = o(1)$ as $T \to \infty$; or

(iii) $\sup_{\boldsymbol{\theta} \in \Theta} \|\pi_T(\boldsymbol{\theta})\| = o(1)$ as $T \to \infty$.

Condition *(i)* in Assumption 7 describes the case where the penalty is well centered; i.e. the case where the penalty has a minimum precisely at $\boldsymbol{\theta}_0 \in \Theta$. Note that the condition allows for several minima, or even a continuum of minima containing in a region of Θ containing $\boldsymbol{\theta}_0$. As we shall see, this case offers a good benchmark to analyze the various properties of the penalized estimator. Condition *(ii)* in Assumption 6 considers the more general case where the penalty function may be incorrectly centered on small samples but is nonetheless asymptotically correct. Clearly, condition *(i)* is just a special case of condition *(ii)*. Finally, condition *(iii)* considers the case where the criterion vanishes asymptotically uniformly on the parameter space.

The following theorem establishes the consistency of the PII estimator w.r.t. the true parameter $\boldsymbol{\theta}_0$.

THEOREM 3. *Let Assumptions 1 to 4, 6 and 7 hold. Then $\widehat{\boldsymbol{\theta}}_{T,S} \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $T \to \infty \; \forall \; S \in \mathbb{N}$.*

3.2   Asymptotic Normality

Following Gourieroux et al. (1993) we obtain the asymptotic normality of the PII estimator by building on the asymptotic normality of the auxiliary estimators.

Assumption 8 ensures several important ingredients used for obtaining asymptotic normality of II estimators. First, it imposes typical smoothness assumptions on the auxiliary estimators. Second, it assumes the asymptotic normality of the auxiliary estimators. Third, it ensures that the second derivative of the auxiliary estimators converges uniformly over Θ. These conditions are further complemented by similar smoothness and convergence requirements on the penalty function. We denote the asymptotic variance of $\widehat{\boldsymbol{\beta}}_T$ by $\boldsymbol{\Sigma}$ and note that the asymptotic variance of $\widetilde{\boldsymbol{\beta}}_{T,S}$ is given by $S^{-1}\boldsymbol{\Sigma}$ for any $S \in \mathbb{N}$. In Assumption 8, and throughout, we let ∇^i denote the i-th derivative operator.

ASSUMPTION 8. *The following smoothness conditions hold a.s. $\forall\ (T,S) \in \mathbb{N} \times \mathbb{N}$:*

(i) $\widetilde{\boldsymbol{\beta}}_{T,S} \in \mathbb{C}^2(\Theta, \mathcal{B})$;

(ii) $\pi_T \in \mathbb{C}^2(\Theta, \mathbb{R})$.

The auxiliary estimators are asymptotically normal for every $S \in \mathbb{N}$:

(iii) $\sqrt{T}(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{b}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma})$ *as $T \to \infty$;*

(iv) $\sqrt{T}\left(\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0) - b(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N(0, S^{-1}\boldsymbol{\Sigma})$ *as $T \to \infty$.*

The second derivatives of both the auxiliary estimators and the penalty function converge uniformly to some deterministic limit for every $S \in \mathbb{N}$:

(v) $\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) - \nabla^2 \boldsymbol{b}(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$ *as $T \to \infty$;*

(vi) $\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \pi_T(\boldsymbol{\theta}) - \nabla^2 \pi(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$ *as $T \to \infty$.*

The asymptotic variance $\boldsymbol{\Sigma}$ in Assumption 8 can be estimated in different ways depending on the nature of auxiliary statistics. Gourieroux et al. (1993) consider the case of extremum estimators

$$\widehat{\boldsymbol{\beta}}_T = \arg\min_{\beta \in \mathcal{B}} \ell_T(\boldsymbol{\beta})$$

where ℓ_T denotes a random criterion that depends on observed data. The asymptotic variance $\boldsymbol{\Sigma}$ is thus naturally obtained from the classical expansion

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{b}_0) = \left(\nabla^2\ell_T(\boldsymbol{\beta}_T^*)\right)^{-1}\sqrt{T}\nabla\ell_T(\boldsymbol{b}_0)$$

where $\nabla^2\ell_T(\boldsymbol{\beta}_T^*)$ denotes the Hessian of ℓ_T evaluated (row-wise) at some point $\boldsymbol{\beta}_T^*$ between $\widehat{\boldsymbol{\beta}}_T$ and $\boldsymbol{b}_0$ and $\sqrt{T}\nabla\ell_T(\boldsymbol{b}_0)$ is the scaled Jacobian. Typically, application of a CLT to the scaled Jacobian yields $\sqrt{T}\nabla\ell_T(\boldsymbol{b}_0) \xrightarrow{d} N(0, I(\boldsymbol{b}_0)^{-1})$ and, together with the uniform convergence of the Hessian to some limit deterministic matrix J; i.e. $\sup_\beta \|\nabla^2\ell_T(\boldsymbol{\beta}_T^*) - J\| \xrightarrow{a.s.} 0$, we obtain the usual extremum estimation result (see e.g. White (1994))

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{b}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}) \quad \text{as} \quad T \to \infty \quad \text{with} \quad \boldsymbol{\Sigma} = J(\boldsymbol{b}_0)I(\boldsymbol{b}_0)^{-1}J(\boldsymbol{b}_0).$$

The relation between the two asymptotic variances stated in Assumption 8 can also be easily understood in this setting by noting that $S^{-1}\boldsymbol{\Sigma}$ is the asymptotic variance obtained for the estimator

$$\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0) = \frac{1}{S}\sum_{i=1}^{S}\widetilde{\boldsymbol{\beta}}_{T,i}(\boldsymbol{\theta}_0) \quad \text{where} \quad \widetilde{\boldsymbol{\beta}}_{T,i}(\boldsymbol{\theta}_0) = \arg\max_{\beta\in\mathcal{B}}\widetilde{\ell}_T(\boldsymbol{\beta}; \boldsymbol{\theta}_0)$$

where $\widetilde{\ell}_T(\boldsymbol{\beta}; \boldsymbol{\theta}_0)$ denotes the counterpart of $\ell_T(\boldsymbol{\beta})$ obtained from data simulated under the structural parameter vector $\boldsymbol{\theta}_0 \in \Theta$.

The asymptotic normality of the PII estimator requires also that the penalty function converges in an appropriate manner to some limit. Assumption 9 states alternative conditions that render the PII estimator asymptotically normal. Each condition yields a different limit result in terms of asymptotic bias and asymptotic variance.

ASSUMPTION 9. *The penalty function satisfies one of the following conditions:*

(i) $\nabla\pi_T(\boldsymbol{\theta}_0) = o(T^{-\frac{1}{2}})$ *and* $\nabla^2\pi(\boldsymbol{\theta}_0) = o(1)$ *as* $T \to \infty$;

(ii) $\nabla\pi_T(\boldsymbol{\theta}_0) = O(T^{-\frac{1}{2}})$ *and* $\nabla^2\pi(\boldsymbol{\theta}_0) = o(1)$ *as* $T \to \infty$;

(iii) $\nabla\pi_T(\boldsymbol{\theta}_0) = o(T^{-\frac{1}{2}})$ *and* $\nabla^2\pi(\boldsymbol{\theta}_0) = O(1)$ *as* $T \to \infty$;

(iv) $\nabla\pi_T(\boldsymbol{\theta}_0) = O(T^{-\frac{1}{2}})$ *and* $\nabla^2\pi(\boldsymbol{\theta}_0) = O(1)$ *as* $T \to \infty$.

The role of Assumption 9 can be easily understood by analyzing an expansion of the PII estimator. Consider, for simplicity, the case of scalar parameters $\boldsymbol{\theta} = \theta$ and $\boldsymbol{\beta} = \beta$, then it is easy to show that

$$\sqrt{T}(\widehat{\theta}_{T,S} - \theta_0) = \frac{\nabla\widetilde{\beta}_{T,S}(\theta_0)\ \sqrt{T}\Big(\widehat{\beta}_T - \widetilde{\beta}_{T,S}(\theta_0)\Big)\ -\ \frac{1}{2}\sqrt{T}\nabla\pi_T(\theta_0)}{\nabla\widetilde{\beta}_{T,S}(\theta_0)^2\ +\ \frac{1}{2}\nabla^2\pi_T(\theta_0)\ +\ o_{a.s.}(1)}.$$

Compared to the standard II estimator, there are two additional terms in this expansion that relate to the penalty function. These are the scaled derivative of the penalty function $\frac{1}{2}\sqrt{T}\nabla\pi_T(\theta_0)$, and the term $\frac{1}{2}\nabla^2\pi_T(\theta_0)$, which is proportional to the second derivative of the penalty function. Statistical inference on $\boldsymbol{\theta}_0$ can thus be affected by both the slope and curvature of the penalty function at $\boldsymbol{\theta}_0$, even when the penalty function vanishes asymptotically. In particular, note that when the curvature vanishes asymptotically, $\nabla^2\pi_T(\theta_0) = o(1)$, and the slope vanishes sufficiently fast, $\nabla\pi_T(\theta_0) = o(T^{-\frac{1}{2}})$, then we recover the asymptotic distribution of the standard indirect inference estimator since

$$\sqrt{T}(\widehat{\theta}_{T,S} - \theta_0) = \frac{\nabla\widetilde{\beta}_{T,S}(\theta_0)\ \sqrt{T}\Big(\widehat{\beta}_T - \widetilde{\beta}_{T,S}(\theta_0)\Big)\ +\ o_{a.s.}(1)}{\nabla\widetilde{\beta}_{T,S}(\theta_0)^2\ +\ o_{a.s.}(1)}.$$

In contrast, if the curvature of the penalty function at θ_0 does not vanish asymptotically, but instead converges to some constant $\nabla^2\pi_T(\theta_0) = O(1)$ as $T \to \infty$, then the asymptotic variance of the penalized indirect inference estimator will be different from that of the standard indirect inference estimator. If the limit curvature is positive at θ_0, then the asymptotic variance of the PII estimator will be smaller than that of the unpenalized II estimator. If the limit curvature is negative, then the asymptotic variance of the PII estimator will be larger. Note that if the penalty's curvature $\nabla^2\pi_T(\theta_0)$ converges to a negative value that cancels out with the quadratic term $\nabla\widetilde{\beta}_{T,S}(\theta_0)^2$ as $T \to \infty$

$$\nabla\widetilde{\beta}_{T,S}(\theta_0)^2\ +\ \frac{1}{2}\nabla^2\pi_T(\theta_0) = o_{a.s.}(1),$$

then, the PII criterion is asymptotically flat at θ_0 and, as a result, the $\sqrt{T}(\widehat{\theta}_{T,S} - \theta_0)$ diverges. Inversely, if the curvature diverges, then $\sqrt{T}(\widehat{\theta}_{T,S} - \theta_0) \xrightarrow{a.s.} 0$.

Similarly, if the penalty's slope at θ_0 does not vanish fast enough, i.e. if $\sqrt{T}\nabla\pi_T(\theta_0)$

11

converges to a constant, so that $\nabla \pi_T(\theta_0) = O(T^{-\frac{1}{2}})$, then an asymptotic bias is introduced and asymptotic distribution of the PII estimator differs from that of the standard II estimator. If the scaled slope $\sqrt{T}\nabla \pi_T(\theta_0)$ diverges, then $\sqrt{T}(\widehat{\theta}_{T,S} - \theta_0)$ diverges a.s. as $T \to \infty$.

Table 1 shows the four different limiting cases for which a non-degenerate limiting distribution is well defined. These are the cases considered in Assumption 9.

Table 1: Different limit cases for the asymptotic distribution of the PII estimator.

	$\nabla \pi_T(\theta_0) = o(T^{-\frac{1}{2}})$	$\nabla \pi_T(\theta_0) = O(T^{-\frac{1}{2}})$
$\nabla^2 \pi_T(\theta_0) = o(1)$	Standard II Asymptotics	Asymptotic Bias
$\nabla^2 \pi_T(\theta_0) = O(1)$	Different Asymptotic Variance	Asymptotic Bias + Different Asymptotic Variance

Theorem 4 establishes the asymptotic normality of the PII estimator $\widehat{\theta}_{T,S}$, and describes its asymptotic mean and variance, under the four different limit cases described by Assumption 9. For notational simplicity, Theorem 4 adopts the notation

$$A := \frac{\partial b(\theta_0)'}{\partial \theta} \Omega \frac{\partial b(\theta_0)}{\partial \theta'} \quad , \quad B := \frac{\partial b(\theta_0)'}{\partial \theta} \Omega \Sigma \Omega \frac{\partial b(\theta_0)}{\partial \theta'} \; ,$$

and $\quad \Pi = \lim_{T \to \infty} \sqrt{T}\nabla \pi_T(\theta_0) \quad$ under Assumption 9(ii) and 9(iv).

THEOREM 4. *Let Assumptions 1 to 4 and 6 to 9 hold. Then $\sqrt{T}\left(\widehat{\theta}_{T,S} - \theta_0\right) \xrightarrow{d} N(\mu, W) \; \forall \; S \in \mathbb{N}$ as $T \to \infty$ for some $p \times 1$ vector μ and $p \times p$ matrix W.*

1. *If Assumption 9(i) holds then:*

$$\mu = \mathbf{0} \quad and \quad W = \left(1 + \frac{1}{S}\right)A^{-1}BA^{-1} \; .$$

2. *If Assumption Assumption 9(ii) holds then:*

$$\mu = -\frac{1}{2}A^{-1}\Pi \quad and \quad W = \left(1 + \frac{1}{S}\right)A^{-1}BA^{-1} \; .$$

12

3. *If Assumption Assumption 9(iii) holds then:*

$$\boldsymbol{\mu} = \boldsymbol{0} \quad and \quad W = \left(1 + \frac{1}{S}\right)\left(A + \frac{1}{2}\nabla^2\pi(\boldsymbol{\theta}_0)\right)^{-1} B\left(A + \frac{1}{2}\nabla^2\pi(\boldsymbol{\theta}_0)\right)^{-1} .$$

4. *If Assumption Assumption 9(iv) holds then:*

$$\boldsymbol{\mu} = -\frac{1}{2}\left(A + \frac{1}{2}\nabla^2\pi(\boldsymbol{\theta}_0)\right)^{-1}\Pi$$

$$and \quad W = \left(1 + \frac{1}{S}\right)\left(A + \frac{1}{2}\nabla^2\pi(\boldsymbol{\theta}_0)\right)^{-1} B\left(A + \frac{1}{2}\nabla^2\pi(\boldsymbol{\theta}_0)\right)^{-1} .$$

Understanding the limit behavior of the PII estimator is useful since it reveals how statistical inference may be influenced by the introduction of the penalty. Furthermore, it can help provide some guidance for the construction of penalties that are asymptotically innocuous from an inferential perspective. For example, when $\boldsymbol{\theta}_0$ is believed to be only weakly identified in small samples, and the penalty is only designed to give some additional 'finite sample structure' to the estimation problem, then one may legitimately wish to construct penalties that are void of any asymptotic inferential influence. Theorem 4 provides us with the necessary tools for constructing such penalties.

4   FINITE SAMPLE BEHAVIOR

This section provides a Monte Carlo analysis of the finite sample behavior of the PII estimator $\widehat{\boldsymbol{\theta}}_{T,S}$. In particular, it shows how the penalty function can be used to obtain several desirable properties for the estimator.

We consider a the following prototypical RBC model

$$\max \mathbb{E}_t \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\psi} - 1}{1 - \psi}\right) ,$$

$$k_{t+1} = (1 - \delta)k_t + \exp(z_t)k_t^\alpha - c_t ,$$

$$z_t = \rho z_{t-1} + \epsilon_t \quad , \quad \{\epsilon_t\} \sim NID(0, \boldsymbol{\Sigma}^2) ,$$

where c_t stands for consumption, k_t denotes the capital stock, z_t denotes total factor productivity, ϵ_t is normal independently distributed (NID) productivity shock, and

$(\psi, \delta, \alpha, \rho, \sigma^2)$ is a vector of structural parameters. According to this model, the economic agent maximizes consumption over time subject to a capital accumulation constraint. The economy is described by an AK-production function and the level of technology follows a stationary AR(1) process with no drift. If $\psi \to 1$, this model has a log utility function and can be solved analytically. The following two equations characterize the solution:

$$\log k_t = \log(\alpha\beta) + \alpha \log k_{t-1} + z_{t-1}$$

$$\log c_t = \log(1 - \alpha\beta) + \alpha \log k_t + z_t$$

Below we investigate the small sample behavior of the PII estimator by means of a Monte Carlo study. In particular, we generate $N = 1000$ streams of data of length $T = 250$ from this model under the 'true' parameter values: $\alpha = 0.33$, $\beta = 0.99$, $\delta = 0.023$, $\psi = 1.75$, $\rho = 0.95$, $\sigma = 0.0104$. These are standard parameter values taken from Heer and Maussner (2005). We also follow the literature in fixing σ^2, ψ and δ and focusing on the estimation (α, β, ρ). The binding function stacks variances of the data normalized by the variance of the output and correlations of the elements of the data vector.

Figure 2 compares the finite sample distribution of the II estimator (top row) and the PII estimator (bottom row) under correct model specification. The PII estimator imposes a quadratic penalty that centers the estimation of α at the value 0.33, and centers the estimate of β at 0.9785. This penalty is also designed to explode as $\beta \to 1$ so as to ensure that the estimate of β is bounded above by 1. The penalty is flat in the parameter ρ.

Since the distribution of the II estimator of α is already reasonably well centered at 0.33, the PII estimator has a similar distribution in terms of location. However, as expected, the distribution of the PII estimator of α appears more concentrated at the true value. Figure 2 also shows that the PII estimator is successful in producing estimates of β that avoid the 'unacceptable' values $\beta \geq 1$. In contrast, the II estimator avoids a large number of parameter estimates above 1, which are 'impossible' from an economic perspective, and render the dynamic optimization problem underlying the RBC model ill defined and invalid. The distribution of the ρ estimates are similar.

Figure 3 reveals the finite sample distribution of the II and PII estimators under incorrect specification. This is done by generating the 'observed data' from our RBC
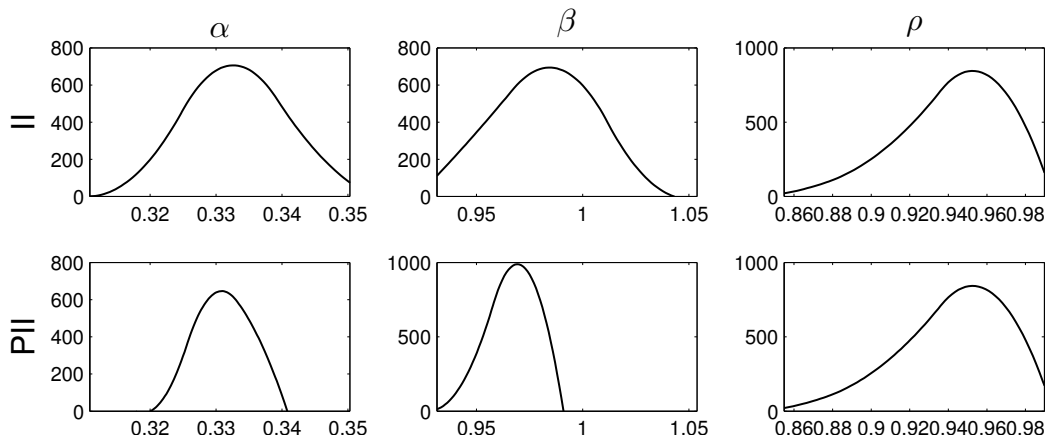
Figure 2: Distribution of the estimates for the correctly specified model. Note: This figure shows the distribution of the estimates across 1000 replications for II and PII estimators. The sample size is $T = 250$. α is the elasticity of capital in production function, β is the discount rate and ρ is the persistency of technological progress.

model with a CRRA utility function with $\psi = 2$, but attempting to estimate our RBC model with log utility function obtained by setting $\psi \to 1$. Despite all other features of the model remaining intact, Figure 3 shows that even this small misspecification in the utility function leads to considerable problems for the classical II estimator. In particular, the distribution of the II estimator (in the top row), reveals that (i) the estimates of α become incomprehensible low, implying marginal productivities of capital that are hard to justify from an economic stand point; and (ii) the estimates of β become simply unacceptable from an economic theoretic stand point. It is simply impossible for an economist to justify point estimates of β that are close to zero. In comparison, the PII estimates of both α and β are 'shifted' towards more 'acceptable' values. Just as in a Bayesian setting, the degree to which the point estimates are shifted to acceptable values will depend on the relative strength of the penalty. The PII estimate of ρ is also affected in this case although the estimate of ρ is not penalized.

In Bayesian econometrics, the effect of the prior vanishes as the sample size increases, as the sample information contained in the likelihood function becomes increasingly dominating in large samples. In Section 3 we have noted that a similar structure can be imposed in the context of PII estimation by considering penalties whose effect vanishes asymptotically. In particular, Theorem 4 highlighted that the penalty may play an important role in shaping the asymptotic distribution. As such, it is important to understand the influence that the penalty may have in hypothesis
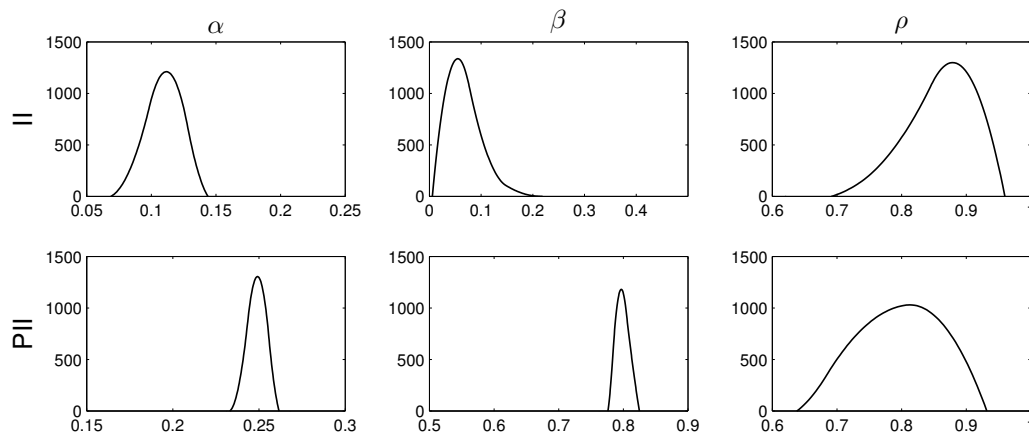
Figure 3: Distribution of the estimates for the incorrectly specified model: CRRA utility with $\psi = 2$ instead of log utility. Note: This figure shows the distribution of the estimates across 1000 replications for Indirect inference (II) and Penalized Indirect Inference(PII). The sample size is $T = 250$. α is the elasticity of kapital in production function, β is the discount rate and ρ is the persistency of technological progress.

testing and parameter inference when conducting statistical inference based on the asymptotic distribution of the estimator.

Figure 4 investigates the effect of the vanishing rate of the penalty function on the finite-sample and asymptotic distributions of the PII estimator. The sample sizes are $T = 100$, 250, 500 and 1000. For simplicity, we focus on estimating the parameter $\alpha_0 = 0.33$ while fixing the remaining parameters to their true values. In particular, we focus on testing the null hypothesis $H_0 : \alpha = 0.33$ at the 5% level.

We impose a quadratic penalty function of the form

$$\pi_T(\alpha) = b_T\Big(\alpha - (\alpha_0 + a_T)\Big)^2 \alpha_0 \ ,$$

where $\{a_T\}$ and $\{b_T\}$ are vanishing sequences. For illustrative purposes we consider two alternative rates for the sequences:

- A slow-vanishing penalty where $b_T = O(1)$ and $a_T = O(T^{-1/2})$;

- A fast-vanishing penalty where $b_T = O(T^{-\frac{1}{4}})$ and $\sim a_T = O(T^{-1/2})$.

Note that since the first and second derivatives of the penalty function are given by

$$\nabla \pi_T(\boldsymbol{\theta}_0) := \frac{\partial \pi_T(\alpha_0)}{\partial \alpha} = 2b_T\Big(\alpha - (\alpha_0 + a_T)\Big)\Big|_{\alpha_0} = 2b_T a_T,$$

$$\text{and} \quad \nabla^2 \pi_T(\boldsymbol{\theta}_0) := \frac{\partial^2 \pi_T(\alpha_0)}{\partial \alpha^2} = 2b_T,$$

16

it follows that the fast-vanishing penalty satisfies Assumption 9(i) with $\nabla\pi_T(\boldsymbol{\theta}_0) = O(T^{-1/2})$ and $\nabla^2\pi_T(\boldsymbol{\theta}_0) = O(1)$, while the slow-vanishing penalty satisfies Assumption 9(iv) with $\nabla\pi_T(\boldsymbol{\theta}_0) = O(T^{-3/4})$ and $\nabla^2\pi_T(\boldsymbol{\theta}_0) = O(T^{-1/4})$.
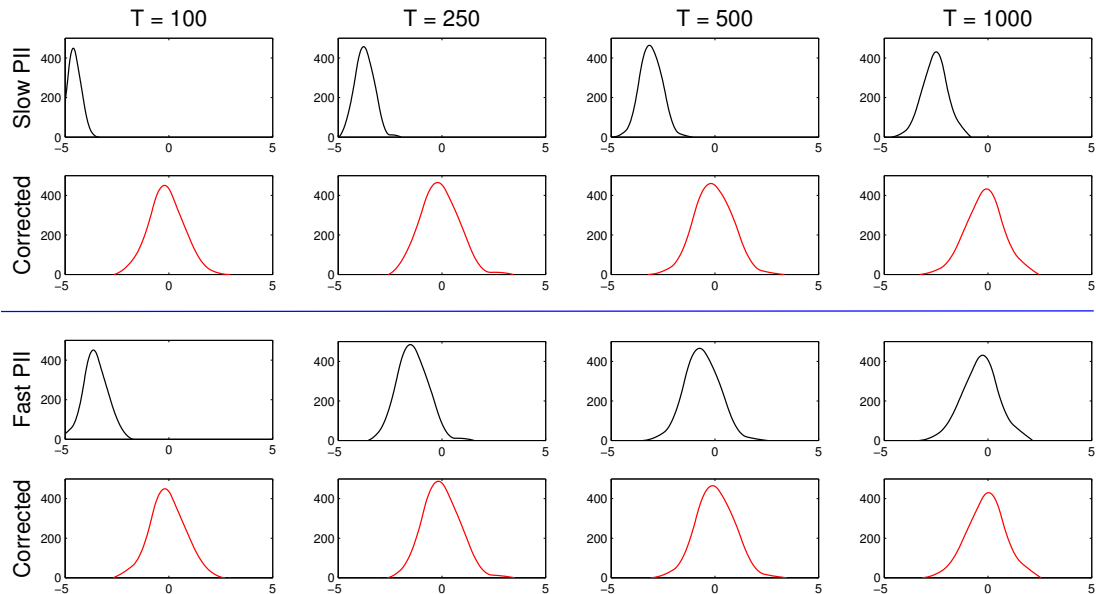

Figure 4: Corrections to the distributions. Note: This figure shows the distribution of the standardized Penalized Indirect Inference(PII) estimator with fast and slow penalties and their corrected versions across 1000 replications. The sample sizes are 100, 250, 500 and 1000.

As expected, the two bottom rows of Figure 4 show that the fast vanishing penalty changes the location and scale of the finite-sample distribution of PII estimator, but leaves the asymptotic distribution unchanged. Indeed, for $T = 100$ (left column), the density of the (standardized) PII estimator contrasts significantly with the 'corrected' distribution (in red) that is obtained using Theorem 4 to 'eliminate' the effect of the penalty. For the larger sample size of $T = 1000$, this discrepancy with respect to the corrected distribution is no longer identifiable by visual inspection of the graph. In contrast, the top two rows of Figure 4 show that the slow-vanishing penalty affects not only the small sample distribution of the PII estimator, but also, its asymptotic distribution. This is in accordance to Theorem 4, which predicts an asymptotic location and scale effect of penalties satisfying $\nabla\pi_T(\boldsymbol{\theta}_0) = O(T^{-1/2})$ and $\nabla^2\pi_T(\boldsymbol{\theta}_0) = O(1)$.

Table 2 provides more detailed insight into the tail behavior of the distribution of the PII estimator under the slow vanishing penalty. In particular, it reports rejection

frequencies, under the null, for a Wald test of nominal size of 5%, that relies on the asymptotic distributions derived in Theorem 4. The column labeled II, reports rejection frequencies for the II estimator using the standard II asymptotic distribution derived in Theorem 4(i). The columns under 'Slow PII' report rejection frequencies for the PII estimator featuring a slow vanishing penalty. For the PII estimator, we report rejection frequencies using both the standard asymptotic distribution derived in Theorem 4(i), as well as, the distribution derived in Theorem 4(iv). As expected, the rejection frequencies are closer to their nominal value of 5% when using the distribution derived in derived in Theorem 4(iv).

Table 2: Rejection frequencies for PII under the null

| T | II | Slow PII | |
	Theo. 4(i)	Theo. 4(i)	Theo. 4(iv)
100	0.023	1.000	0.023
250	0.030	1.000	0.031
500	0.036	0.983	0.037
1000	0.037	0.844	0.038

Note: This table reports rejection frequencies for II and PII estimators with fast and slow penalties, at a nominal size of 5%.

5 ECONOMETRIC ANALYSIS OF A DSGE MODEL

In this section, we consider the model of Smets and Wouters (2007), from now on denoted SW for brevity, and use their data set to estimate the parameters of the model using II and PII estimators.

The SW model is a medium-scale monetary business cycle model that allows for 'sticky' prices and 'sticky' wages. As pointed out by the authors, the model accommodates *"for backward inflation indexation, habit formation in consumption and investment adjustment costs that create hump-shaped responses of aggregate demand, and variable capital utilization and fixed costs in production"*.

The stochastic dynamics are driven by seven orthogonal structural shocks: total factor productivity shocks; risk premium shocks and investment-specific technology shocks, which affect the intertemporal margin; wage and price mark-upshocks, which

influence the intratemporal margin; and two policy shocks: exogenous spending and monetary policy shocks.

The source of the original data is the U.S. Department of Commerce, Bureau of Economic Analysis. We take the data from the supplementary materials of SW. The interested reader could find the details on the data composition in those materials. The vector of the observed time series consists of output, consumption, investment, real wages, hours worked, inflation, and the interest rate. The first four series are taken in the first differences following SW, whereas the last three are in levels. The sample size is $T = 230$. We use the Anderson and Moore (1985) algorithm to solve the model.

We consider two different sets of auxiliary statistics for the II and PII estimation of the model: (i) the vector of autocovariances of the data up to the order 1, as proposed, e.g., by Gorodnichenko and Ng, 2010; and (ii) the vector that stacks the autocorrelations of the data up to the order 1 and the ratios of the variances to the variance of the output, as suggested, e.g., by Dave and Dejong, 2007.

For simplicity, we introduce a simple quadratic penalty function for our PII estimator,

$$\pi_T(\boldsymbol{\theta}) = b_T \left(\boldsymbol{\theta} - \boldsymbol{\theta}_{SW}\right)' \left(\boldsymbol{\theta} - \boldsymbol{\theta}_{SW}\right)$$

where $\boldsymbol{\theta}$ is the stacked vector of the parameters of interest, $\boldsymbol{\theta}_{SW}$ is the vector of the posterior modes of SW. The parameter $b_T \geq 0$ determines the curvature of the penalty function.

5.1 PENALTY SENSITIVITY ANALYSIS

In the first part of this study, we estimate only four parameters of the model: persistence of technology shocks ρ_a; investment adjustment cost ϕ; external habit formation in consumption λ; and Calvo's probability of wage adjustment ξ_w. All remaining parameters fixed to the posterior modes obtained in SW. Figure 5 reports the estimates obtained by the PII estimator. Despite the identical structure of the penalty function for every parameter, the results suggest that parameter estimates react on changes in the curvature of the penalty differently. Figure 5 suggests that the estimate of ϕ is quite sensitive to the penalty. This occurs despite the absolute value of ϕ being larger than those of the remaining parameters. The parameters λ and ξ_w demonstrate a similar sensitivity to the penalty. Namely, gradual deviation from the penalty min-

imizer as the strength of the penalty decreases. ρ_a is close to 1 for all considered curvature values. This parameter is thus less sensitive to alternative specifications of our quadratic penalty. Figure 5 also shows that the estimates of λ, ξ_w and ρ_a should move away from the posterior modes in SW if the objective is to match the vector of autocovariances of the data as close as possible. Indeed, note that for $b_T = 0$, the PII estimates are considerably different from those reported in SW (dashed lines).
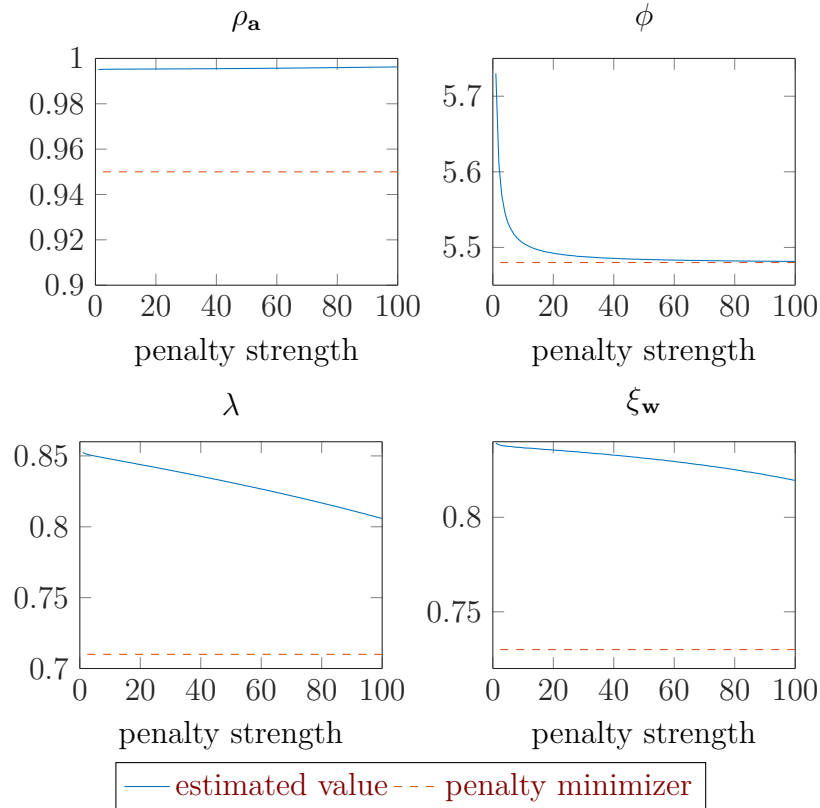
Figure 5: Estimated values of different parameters depending on the penalty strength. Note: This figure shows the estimated values of various parameters of Smets and Wouters (2007) model for Penalized Indirect Inference estimator. ρ_a is persistence in technology shocks, ϕ is investment adjustment cost, λ is habit formation, ξ_w is wage adjustment probability. Penalty function is the same for all parameters: $\pi_T(\theta) = b_T (\theta - \theta_{SW})' (\theta - \theta_{SW})$; b_T varies from 0 to 100.

5.2 Estimation Results

In this section we report the estimation for the full $p = 34$ parameters of the model.[3] Table 3 reports the obtained parameter estimates. To speed-up computations we calculate the autocovariance matrix implied by the model using the approach discussed in Gorodnichenko and Ng (2010). In particular, the moments are computed analytically making use of the VAR structure of the reduced form and the fact that all the (filtered) variables assumed to be weakly stationary. To take into account the different scales of the variables, we adjust the curvature of the penalties for every parameter, leaving the functional form of the penalty function fixed,

$$\pi_T(\boldsymbol{\theta}) = b_T \, \boldsymbol{w}' \boldsymbol{\kappa}(\boldsymbol{\theta}),$$

where $\boldsymbol{w}$ is a $p \times 1$ vector of weights; $\boldsymbol{\kappa}(\boldsymbol{\theta})$ is a $p \times 1$ vector, such that $\boldsymbol{\kappa}_i(\boldsymbol{\theta}) = (\theta_i - \theta_{i,SW})^2$. The weight of ϕ is normalized to 1, and the remainder weights are set relative to that of ϕ according to the relative magnitude of each parameter. Specifically, we let the elements of $\boldsymbol{w}$ be defined as the ratios of the squares of the posterior modes of SW to the square of ϕ,

$$\boldsymbol{w}_i = \frac{\theta_{i,SW}^2}{\phi_{SW}^2}.$$

We find that a large number of parameters are considerably sensitive to the specification of the penalty function. A detailed analysis of this sensitivity can be found in a supplementary appendix, which is available upon request. Given the high-dimensional parameter vector, we also find that the point estimate of the II estimator exhibits considerable sensitivity to the choice of initial parameter in the optimization of the criterion function. The PII criterion function has fewer local minima. Hence the estimates are more robust. Tables 5 and 6 in Appendix B provide a detailed analysis of the sensitivity of the estimates to starting values.

Overall, the parameter estimates obtained by the II and PII estimators are considerably different, and they are also quite far from the posterior modes obtained by SW. The II estimator provides the largest estimate of γ. The estimate of 1.5% for

---

[3]SW claim that certain parameters are not identified, so their values are calibrated: (i) the depreciation rate δ is fixed at 0.025; (ii) the exogenous spending-GDP ratio g_y is set at 18 percent; (iii) the steady-state mark-up in the labor market λ_w is given the value of 1.5; (iv) and the curvature parameters of the Kimball aggregators in the goods and labor market (ϵ, and ϵ_s) are both set at 10.

this parameter, which determines the steady-state growth rate, is quite far from the values usually reported in the literature. In contrast, the SW and PII estimators deliver similar estimates of 0.4% and 0.3%, respectively. The II estimator is also alone in suggesting almost no serial correlation in the markup shock (values of ρ_p and μ_p are close to zero). In contrast, both the SW and PII estimators report positive and persistent values. Interestingly, despite the smallest estimate of ϕ, the II point estimate of ψ is very close to 1, implying that it is extremely costly to change the utilization of capital. As a result, despite that a more responsive supply side of the capital market, utilization of capital is less responsive to shocks for the model implied by II estimates. In contrast, the II and PII estimates of ρ_a are quite similar, suggesting that total factor productivity might have a unit root. The posterior mode of SW is 0.95 and implies considerably different dynamic behavior for productivity. The estimate of the share of capital in production α decreases by 46% from the SW's value 0.190 to the PII estimate 0.101.

In general, estimates of the parameters with asterisk differ considerably between the II and PII. As expected, the penalty function keeps the PII estimates close to the posterior mode values reported by SW. It is possible that the 'loose' II estimates of the non-identified parameters influence the estimation of the remaining parameters. Cogley (2001) reports similar findings estimating a misspecified DSGE model: a unit-root autoregressive TFP process is estimated with the stationarity restriction on the persistence parameter, as a consequence, the GMM estimate of the variance of the error term increases above the true value.

Table 4 reports the moments implied by the model under the parameters estimated by the different methods. It also displays the values obtained for a number of criteria that describe the ability of the model to match the first order autocovariance of the observed data. In particular, *PII obj fun* is the value of the PII criterion function; *II obj fun* is the value of the PII criterion without penalty, i.e. it denotes the sum of the squared differences of all the elements; *maxnorm* stands for the maximal absolute difference element by element; and *SAV* is the sum of the absolute values of all the elements.

It is interesting how SW provide a poor approximation to the covariance between hours worked, the real interest rate, as well as the autocovariance between the real wage and the real interest rate. Naturally, if the goal of the researcher is to match certain moments, then a likelihood-based estimator is not necessarily a good choice.

Table 3: Parameter estimates of Smets and Wouters (2007) model

	SW	II	PII		SW	II	PII
γ^*	1.004	1.015	1.003	ρ_w	0.820	0.918	0.723
σ_C	1.390	1.727	1.306	μ_w	0.620	0.688	0.633
σ_l^*	1.920	2.131	1.409	ρ_r	0.290	0.673	0.298
λ	0.710	0.630	0.717	ρ_a	0.950	0.997	0.997
β^*	0.990	0.966	0.983	ρ_b	0.180	0.585	0.177
ϕ^*	5.480	2.001	6.164	ρ_p^*	0.740	0.001	0.885
ϕ_p	1.610	1.454	1.488	μ_p^*	0.590	0.013	0.750
α	0.190	0.141	0.101	ρ_i^*	0.640	0.634	0.636
ψ	0.540	1.000	0.637	ρ_g^*	0.910	0.985	0.757
ι_p^*	0.220	0.202	0.267	ρ_{ga}^*	0.390	0.302	0.438
ξ_p^*	0.650	0.702	0.708	var_a^*	0.450	0.256	0.347
ι_w^*	0.590	0.079	0.442	var_b^*	0.240	0.000	0.246
ξ_w^*	0.730	0.728	0.912	var_g^*	0.240	0.198	0.254
ρ^*	0.810	0.686	0.618	var_i^*	0.140	0.245	0.150
r_π^*	2.030	1.692	2.800	var_r^*	0.450	0.531	0.439
r_y^*	0.080	0.116	0.091	var_p^*	0.520	0.990	0.435
$r_{\Delta y}$	0.220	0.131	0.213	var_w^*	0.240	0.155	0.269

Note: SW denotes posterior modes from Smets and Wouters (2007). II is the Indirect Inference estimator without penalties. PII is the Penalized Indirect inference estimator with quadratic loss penalties centered around the posterior modes from SW. The data suggest less information on the parameters with asterisk, so penalty function has a positive curvature for them.

The large value of the objective function for the II is expected as this estimator does not take the penalty function into account. As a result, it is more likely to move away from SW estimates, than PII. The higher value of SSD for PII compared to II is arguably limited. Both II and PII improve a lot in comparison with SW. Neither of the estimators capture the variation in output $var(\Delta y_t)$ well, suggesting that it is not the penalty function that explains the poor matching. In any case, both the II and PII estimator improve considerably on SW. On the contrary, the II estimator matches $var(\Pi_t)$ better than PII and SW. This is due to the large estimate of variance of the price mark-up shock.

5.3   POLICY IMPLICATIONS

Impulse response functions (IRF) provide important tools for governments and central banks to conduct macroeconomic policy analysis. In this section we study the IRFs implied by the different estimators for the model of SW. Figures 6 and 7 high-

Table 4: Moments match for different estimators

	Data	SW	II	PII
$\mathrm{var}(\Delta c_t)$	0.71	0.70	0.73	0.91
$\mathrm{var}(\Delta y_t)$	1.03	0.19	0.54	0.54
$\mathrm{var}(\Delta i_t)$	6.82	4.32	6.33	6.28
$\mathrm{var}(\Delta w_t)$	0.39	0.38	0.44	0.32
$\mathrm{var}(l_t)$	8.42	2.86	8.41	8.39
$\mathrm{var}(\Pi_t)$	0.45	0.17	0.43	0.20
$\mathrm{var}(r_t)$	0.75	0.21	0.54	0.55
$\mathrm{cov}(l_t, r_t)$	-1.06	-0.06	-1.10	-1.16
$\mathrm{cov}(w_t, r_{t-1})$	-1.02	0.01	-0.98	-1.03
PII obj fun	—	78.95	641.31	3.59
II obj fun	—	78.95	2.75	3.59
maxnorm	—	5.56	0.60	0.77
SAV	—	32.68	9.80	11.86

Note: A few selected moments and fit measures
implied by SW, II and PII parameter estimates.

light the difference between the impulse responses of key macroeconomic variables to technology and government spending shocks.

Figure 6 reveals that the magnitude of the responses of output, consumption, inflation and wages is significantly smaller for SW than for II or PII. This can be due to the fact that the persistence parameter of the technological shock ρ_a reported by SW is considerably smaller than that obtained by II or PII. The speed at which output converges back to equilibrium is also substantially faster for SW compared to the II or PII estimators. Similarly, the consumption responses implied by all approaches are qualitatively similar, but the response implied by SW is noticeably different quantitatively.

Figure 7 shows that the different estimation methods also predict different IRFs for government spending shocks. In particular, it highlights that the penalty function is crucial for obtaining a sizable reaction of aggregate investment to this kind of shock. Indeed, under the II estimates, investment is essentially non responsive. This is due to the large estimated value of ψ, which implies that government spending crowds out consumption. Similarly, the IRF for hours worked predicted by II differs substantially from that obtained under SW and PII.

A supplementary appendix containing a more thorough analysis of the IRFs im-

plied by different shocks is available upon request. For example, we find that, in response to the risk premium shock, the II estimates produce an IRF that different from that implied by SW and the PII estimator. In particular, it is more persistent and its magnitude is larger in the short-run. Overall, the IRFs implied by the II estimator are often an outlier.
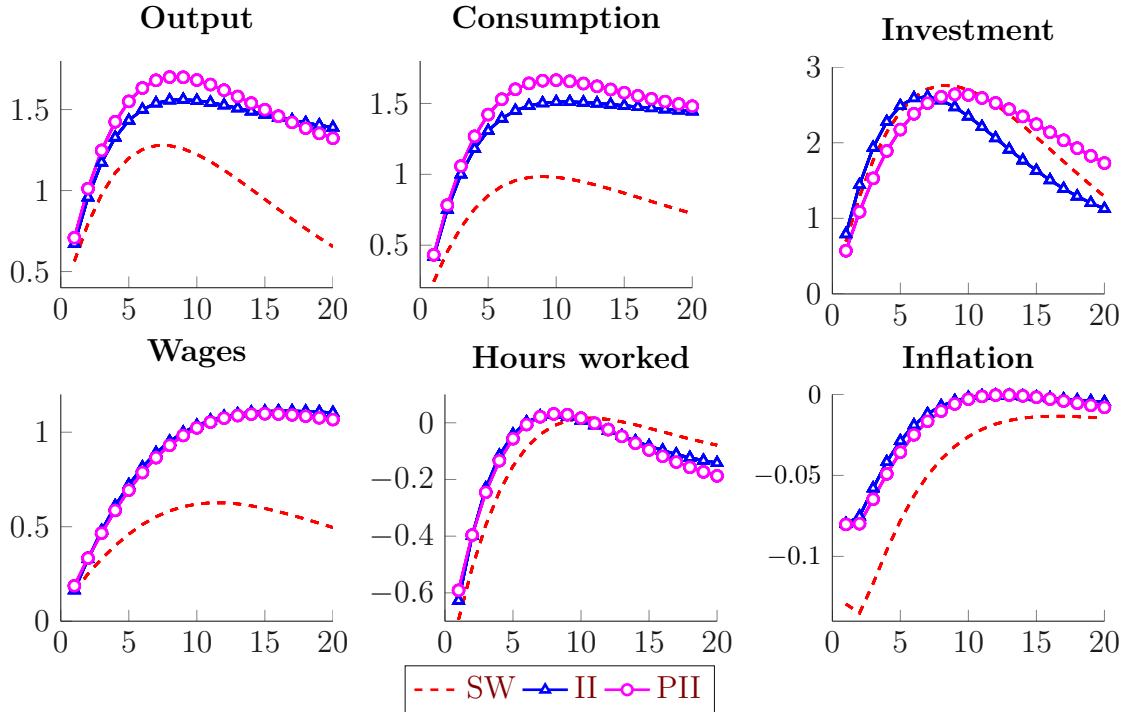

Figure 6: Estimated impulse responses functions to a technology shock. Note: This figure shows impulse response functions of various parameters of Smets and Wouters (2007) model based on different estimators.

5.4 SUMMARY OF EMPIRICAL FINDINGS

Our results suggest that the PII estimator is capable of estimating a state-of-the-art DSGE model. Both the II and PII estimators are able to produce match well the observed sample moments. However, the estimates produced by the II estimator are less robust than those of the PII estimator as they are more dependent on the initial values of the parameters; see Appendix B. Furthermore, the usual II estimator fails to provide estimates that are not convincing from an economic point of view. For example, the demand for capital implied by the II estimator becomes inelastic. As a result, aggregate investment does not react to a government spending shock.
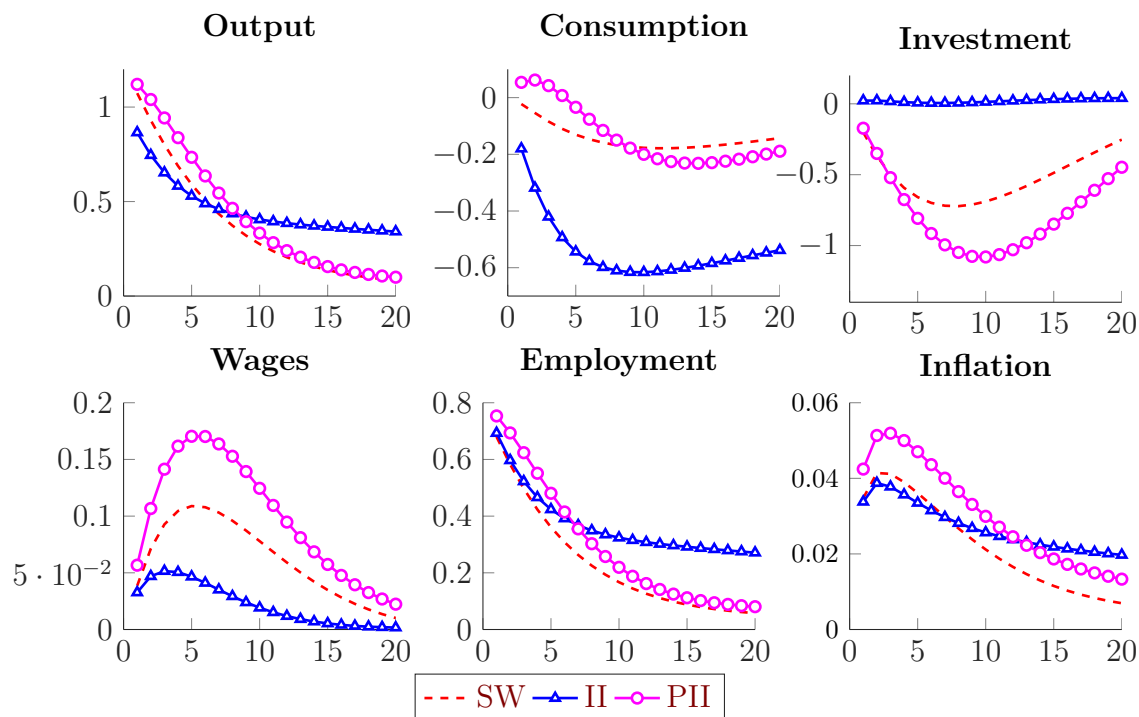
Figure 7: Estimated impulse responses functions to a government spending shock. Note: This figure shows impulse response functions of various parameters of Smets and Wouters (2007) model based on different estimators.

The Bayesian posterior mode estimates reported in Smets and Wouters (2007) produce responses that are acceptable to macroeconomists. However, they are not appealing from a moment matching point of view. In particular, they underestimate variances of consumption, investment and inflation, missing the negative covariance between the real interest rate and hours worked.

The PII estimator can be used to combine the positive aspects of each of the estimation methods discussed above. Specifically, the PII estimator is able to: (i) be considerably more robust than the II estimator in terms of criterion optimization; (ii) match well the observed sample moments; and (iii) produce parameter estimates that are compatible with economic theory and evidence.

6   FINAL REMARKS

This paper proposed an indirect inference estimator that offers a way for the researcher to add information not included in the sample of data, and in this way, exert control over the estimator and obtain parameter estimates that are sensible from an economic

perspective. The estimator was obtained by introducing a penalty in the indirect inference objective function. The penalty may vanish asymptotically or not. The asymptotic properties of the penalized indirect inference estimator were established for both correctly and incorrectly specified models.

A Monte Carlo study revealed the role of the penalty function in shaping the finite sample distribution of the estimator. The Monte Carlo study also illustrated how the penalized indirect inference estimator may be useful in dealing with problems of parameter identification and model misspecification. Finally, the empirical application suggested that this estimator can be used for obtaining economically relevant estimates of the parameters of a state-of-the-art dynamic general equilibrium model.

The theory developed in this paper can be further extended by investigating general classes of penalties. For example, one may consider penalties that depend not only on the sample size and the structural parameters, but also, one the number of simulations or the values of the auxiliary statistics.

References

An, S. and Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews*, 26(2-4):113–172.

Anderson, G. and Moore, G. (1985). A linear algebraic procedure for solving linear perfect foresight models. *Economics letters*, 17(3):247–252.

Beyer, A. and Farmer, R. E. A. (2004). On the Indeterminacy of New-Keynesian Economics. Computing in Economics and Finance 2004 152, Society for Computational Economics.

Canova, F. and Sala, L. (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Engle, R. F. and Fadden, D. M., editors, *Handbook of Econometrics*, volume 6b, chapter 76, pages 5549–5632. Elsevier.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Cogley, T. (2001). Estimating and testing rational expectations models when the trend specification is uncertain. *Journal of Economic Dynamics and Control*, 25(10):1485–1525.

Dalalyan, A. S., Golubev, G. K., and Tsybakov, A. B. (2006). Penalized maximum likelihood and semiparametric second-order efficiency. *ANN. STATIST*, pages 169–201.

Dave, C. and Dejong, D. N. (2007). *Structural Macroeconometrics*. Princeton University Press.

Duffie, D. and Singleton, K. J. (1990). Simulated Moments Estimation of Markov Models of Asset Prices. NBER Technical Working Papers 0087, National Bureau of Economic Research, Inc.

Dupor, B., Han, J., and Tsai, Y.-C. (2009). What do technology shocks tell us about the New Keynesian paradigm? *Journal of Monetary Economics*, 56(4):560–569.

Fernandez-Villaverde, J. (2010). The econometrics of DSGE models. *SERIEs*, 1(1):3–49.

Fukac, M. and Pagan, A. (2010). Limited information estimation and evaluation of dsge models. *Journal of Applied Econometrics*, 25(1):55–70.

Gallant, A. R. and Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, 12(04):657–681.

Gallant, R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.

Gorodnichenko, Y. and Ng, S. (2010). Estimation of dsge models when the data are persistent. *Journal of Monetary Economics*, 57(3):325–340.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:85–118.

Green, P. J. (1996). Penalized likelihood. In *In Encyclopedia of Statistical Sciences, Update Volume 2*, pages 578–586. John Wiley & Sons.

Heer, B. and Maussner, A. (2005). *Dynamic General Equilibrium Modelling: Computational Methods and Applications ; with 24 Tables*. Springer-Verlag GmbH.

Huber, P. (1974). *Robust Statistics*. Wiley, New York.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the berkeley symposium on mathematical statistics an*.

Lee, B.-S. and Ingram, B. F. (1991). Simulation estimation of time-series models. *Journal of Econometrics*, 47(2-3):197–205.

Liao, Z. (2013). Adaptive gmm shrinkage estimation with consistent moment selection. *Econometric Theory*, 29(05):857–904.

Ma, A. (2002). GMM estimation of the new Phillips curve. *Economics Letters*, 76(3):411–417.

Nason, J. M. and Smith, G. W. (2008). Identifying the new Keynesian Phillips curve. *Journal of Applied Econometrics*, 23(5):525–551.

Ruge-Murcia, F. (2007). Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, 31(8):2599–2636.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.

Smith, A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:563–584.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

White, H. (1994). *Estimation, Inference and Specification Analysis.* Cambridge Books. Cambridge University Press.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

A Proofs

Proof of Theorem 1

Note first that the criterion function

$$Q_{T,S}(\boldsymbol{\theta}) := \big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\big)'\Omega\big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\big) + \pi_T(\boldsymbol{\theta})$$

is a.s. continuous in $\boldsymbol{\theta} \in \Theta$ by continuity of the criterion in $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ and $\pi_T(\boldsymbol{\theta})$ and the a.s. continuity of $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$ and continuity of $\pi_T(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$ stated in Assumption 2. Furthermore, the criterion is $\mathfrak{B}(\mathbb{R})/\mathcal{F}$-measurable for every $\boldsymbol{\theta} \in \Theta$ since it is continuous in $\widehat{\boldsymbol{\beta}}_T$ and $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ and both $\widehat{\boldsymbol{\beta}}_T$ and $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ are $\mathfrak{B}(\mathcal{B})/\mathcal{F}$-measurable for every $\boldsymbol{\theta} \in \Theta$. The measurability of $\widehat{\boldsymbol{\theta}}_T$ follows from (White, 1994, Theorem 2.11) or (Gallant and White, 1988, Lemma 2.1, Theorem 2.2).

Proof of Theorem 2

Following a classical consistency argument (found, e.g., in White, 1994, Theorem 3.4 or Theorem 3.3 in Gallant and White, 1988), we obtain $\widehat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0^*$ from the uniform convergence of the criterion function $Q_{T,S}(\boldsymbol{\theta}) := \big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\big)'\Omega\big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\big) + \pi_T(\boldsymbol{\theta})$ to the limit $Q_\infty(\boldsymbol{\theta}) := \big(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta})\big)'\Omega\big(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta})\big) + \pi(\boldsymbol{\theta})$ as $T \to \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \ \forall \ S \in \mathbb{N} \ \text{ as } \ T \to \infty \tag{1}$$

and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \Theta$ introduced in White (1994),

$$\sup_{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \epsilon} Q_\infty(\boldsymbol{\theta}) < Q_\infty(\boldsymbol{\theta}_0) \ \forall \ \epsilon > 0. \tag{2}$$

The uniform convergence of the criterion in (eq. (1)) follows from the uniform convergence of the auxiliary statistics $\widehat{\boldsymbol{\beta}}_T$ and $\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ in Assumption 3 and the uniform convergence of $\pi_T(\boldsymbol{\theta})$ in Assumption 4,

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \Theta} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| &= \sup_{\boldsymbol{\theta} \in \Theta} \left| \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right)' \Omega \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right) + \pi_T(\boldsymbol{\theta}) \right. \\
&\quad \left. - \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right)' \Omega \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right) - \pi(\boldsymbol{\theta}) \right| \\
&\leq \sup_{\boldsymbol{\theta} \in \Theta} \left| \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right)' \Omega \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right) - \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right)' \Omega \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right) \right| \\
&\quad + \sup_{\boldsymbol{\theta} \in \Theta} \left| \pi_T(\boldsymbol{\theta}) - \pi(\boldsymbol{\theta}) \right| \\
&= o_{a.s.}(1) + o_{a.s.}(1) = o_{a.s.}(1),
\end{aligned}$$

where $\sup_{\boldsymbol{\theta} \in \Theta} \left| \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right)' \Omega \left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) \right) - \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right)' \Omega \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right) \right| = o_{a.s.}(1)$ follows by Assumption 3 and $\sup_{\boldsymbol{\theta} \in \Theta} \left| \pi_T(\boldsymbol{\theta}) - \pi(\boldsymbol{\theta}) \right| = o(1)$ holds by Assumption 4.

The identifiable uniqueness of $\boldsymbol{\theta}_0^*$ (see, e.g., White, 1994) follows from the assumed uniqueness of $\boldsymbol{\theta}_0^*$ in Assumption 5, the continuity of the limit criterion Q_∞ on Θ (implied by the uniform convergence of $\{Q_T\}$ and the Arzella-Ascolli Theorem) and the compactness of Θ.

PROOF OF THEOREM 3

We follow again the classical argument found e.g. in (White, 1994, Theorem 3.4) or Theorem 3.3 in Gallant and White (1988).

The uniform convergence of the criterion function

$$\sup_{\boldsymbol{\theta} \in \Theta} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \ \forall \ S \in \mathbb{N} \ \text{as} \ T \to \infty \tag{3}$$

is again implied by Assumptions 1 to 4, by the same argument as in the Proof of Theorem 2.

The identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \Theta$ follows from the continuity of the limit criterion Q_∞ on Θ and, compactness of Θ and the uniqueness of $\boldsymbol{\theta}_0$ as a maximizer of Q_∞, i.e. $Q_\infty(\boldsymbol{\theta}_0) < Q_\infty(\boldsymbol{\theta}) \ \forall \ \boldsymbol{\theta} \in \Theta$. The compactness of Θ is directly assumed. The continuity of Q_∞ is obtained by the same argument as in the proof of Theorem 2. The uniqueness of $\boldsymbol{\theta}_0$ follows from Assumptions 6 and 7. In particular, note that Assumption 6 ensures that $\boldsymbol{\theta}_0$ minimizes the quadratic term

$$\left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right)' \Omega \left(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta}) \right) = \left(\boldsymbol{b}(\boldsymbol{\theta}_0) - \boldsymbol{b}(\boldsymbol{\theta}) \right)' \Omega \left(\boldsymbol{b}(\boldsymbol{\theta}_0) - \boldsymbol{b}(\boldsymbol{\theta}) \right)$$

by setting it to zero. Assumption 7 (i) and (ii) ensure that the limit penalty $\pi : \Theta \to [0, \infty)$ has a

minimum at $\boldsymbol{\theta}_0$. As a result Q_∞ is uniquely minimized at $\boldsymbol{\theta}_0$. Finally, under Assumption 7 (iii), the penalty vanishes uniformly from limit criterion, and hence

$$Q_\infty(\boldsymbol{\theta}) = \big(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta})\big)'\Omega\big(\boldsymbol{b}_0 - \boldsymbol{b}(\boldsymbol{\theta})\big)$$

and the uniqueness of $\boldsymbol{\theta}_0$ follows from Assumption 6.

PROOF OF THEOREM 4

Following Gourieroux et al. (1993), we note that the first order condition for the PII estimator $\widehat{\boldsymbol{\theta}}_{T,S}$ is given by

$$-2\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\widehat{\boldsymbol{\theta}}_{T,S})}{\partial\boldsymbol{\theta}}'\Omega\Big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\widehat{\boldsymbol{\theta}}_{T,S})\Big) + \nabla\pi_T(\widehat{\boldsymbol{\theta}}_{T,S}) = \boldsymbol{0}$$

where $\boldsymbol{0}$ denotes a vector of zeros. Application of a mean value theorem at $\boldsymbol{\theta}_0$ yields

$$\begin{aligned}
\boldsymbol{0} = & -2\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}'\Omega\Big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0)\Big) + \nabla\pi_T(\boldsymbol{\theta}_0) \\
& + \left[2\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}}'\Omega\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}} - 2\frac{\partial^2\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}'\Omega\Big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_{T,S}^*)\Big) + \nabla^2\pi_T(\boldsymbol{\theta}_{T,S}^*)\right](\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)
\end{aligned}$$

and hence,

$$\begin{aligned}
\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0) = & \left[\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}}'\Omega\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}} - \frac{\partial^2\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}'\Omega\Big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_{T,S}^*)\Big) + \frac{1}{2}\nabla^2\pi_T(\boldsymbol{\theta}_{T,S}^*)\right]^{-1} \\
& \times\left(\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}'\Omega\sqrt{T}\Big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0)\Big) - \frac{1}{2}\sqrt{T}\nabla\pi_T(\boldsymbol{\theta}_0)\right).
\end{aligned}$$

The consistency of $\widehat{\boldsymbol{\theta}}_T$ and the uniform convergence of $\nabla^2\widetilde{\boldsymbol{\beta}}_{T,S}$ as $T \to \infty$ ensures that

$$-\frac{\partial^2\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}'\Omega\Big(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_{T,S}^*)\Big) = o_{a.s.}(1) \quad \text{as} \quad T \to \infty.$$

Furthermore, the uniform convergence of $\nabla\widetilde{\boldsymbol{\beta}}_{T,S}$ and $\nabla^2\pi_T$ ensures also that

$$\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}}'\Omega\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_{T,S}^*)}{\partial\boldsymbol{\theta}} - \frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}'\Omega\frac{\partial\widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = o_{a.s.}(1) \quad \text{as} \quad T \to \infty,$$

$$\text{and} \quad \nabla^2\pi_T(\boldsymbol{\theta}_{T,S}^*) - \nabla^2\pi_T(\boldsymbol{\theta}_0) = o(1) \quad \text{as} \quad T \to \infty.$$

31

As a result, we can re-write the expression for $\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0)$ as

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0) = \left[\frac{\partial \widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}' \Omega \frac{\partial \widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{2}\nabla^2 \pi_T(\boldsymbol{\theta}_0)\right]^{-1}$$
$$\times \left(\frac{\partial \widetilde{\boldsymbol{\beta}}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}' \Omega \sqrt{T}\left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0)\right) - \frac{1}{2}\sqrt{T}\nabla \pi_T(\boldsymbol{\theta}_0)\right).$$

The expressions for $\boldsymbol{\mu}$ and W now follow immediately from the fact that Assumption 8(iii) and (iv) imply for every $S \in \mathbb{N}$, and as $T \to \infty$, the weak convergence

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_T - \widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0)\right) = \sqrt{T}\left(\widehat{\boldsymbol{\beta}}_T - b(\boldsymbol{\theta}_0)\right) - \sqrt{T}\left(\widetilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}_0) - b(\boldsymbol{\theta}_0)\right) \stackrel{d}{\to} N\left(\mathbf{0}, (1 + S^{-1})\boldsymbol{\Sigma}\right).$$

B Robustness of Estimation Results

The following Tables report the means and the standard deviations of the Penalized II and II estimators across 1000 replications. For all parameters with asterisk, the PII has smaller standard deviation, for example, the standard deviation of σ_l is 44 times larger for II estimators compared with PII. The mean values of all estimators differ from the values reported in Table 3: for example the mean of ι_w, the degree of wage indexation in labor markets, estimated by II is larger by a factor of 7.50. However, the mean deviation from the value in Table 3: is more than 20% smaller for PII estimators. Given that the values of some parameters do not vary much across the replications, but their mean is different from the value reported in Table 3; it seems that the values in Tables 5 and 6 govern more information and more credible, whereas conclusions regarding relative performance of the estimators drawn from Table 3 could be misleading.

Table 5: Estimators sensitiviy to the starting values

	SW	II	PenII		SW	II	PenII
γ^*	1.004	1.009	1.004	ι_p^*	0.220	0.325	0.243
		0.033	0.003			0.188	0.106
σ_C	1.390	1.887	1.780	ξ_p^*	0.650	0.693	0.671
		0.586	0.663			0.171	0.097
σ_l^*	1.920	1.916	1.917	ι_w^*	0.590	0.596	0.573
		1.295	0.029			0.253	0.071
λ	0.710	0.571	0.622	ξ_w^*	0.730	0.823	0.806
		0.096	0.107			0.162	0.118
β^*	0.990	0.961	0.975	ρ^*	0.810	0.722	0.803
		0.047	0.023			0.183	0.068
ϕ^*	5.480	2.788	5.479	r_π^*	2.030	1.809	2.025
		1.971	0.012			0.584	0.034
ϕ_p	1.610	1.386	1.433	r_y^*	0.080	0.114	0.090
		0.182	0.218			0.048	0.041
α	0.190	0.148	0.103	$r_{\Delta y}$	0.220	0.111	0.106
		0.088	0.066			0.089	0.062
ψ^*	0.540	0.880	0.549	ρ_w	0.820	0.704	0.695
		0.205	0.067			0.219	0.233

Note: This table reports estimators sensitivity to the starting values that drawn randomly in the interval between the posterior mode and the prior mean reported in Smets and Wouters (2007). SW denotes posterior modes from Smets and Wouters (2007). II is the Indirect Inference estimator without penalties. PII is the Penalized Indirect inference estimator with quadratic loss penalties centered around SW. We report means and the standard deviations of the estimates across 1000 replications. The data suggest less information on the parameters with asterisk, so penalty function has a positive curvature for them.

Table 6: Estimators sensitiviy to the starting values

	SW	II	PenII		SW	II	PenII
μ_w	0.620	0.571	0.545	ρ_{ga}^*	0.390	0.430	0.383
		0.198	0.231			0.234	0.070
ρ_r	0.290	0.461	0.376	var_a^*	0.450	0.267	0.333
		0.184	0.190			0.272	0.115
ρ_a	0.950	0.602	0.516	var_b^*	0.240	0.157	0.152
		0.319	0.403			0.150	0.079
ρ_b	0.180	0.462	0.277	var_g^*	0.240	0.166	0.154
		0.299	0.235			0.159	0.078
ρ_p^*	0.740	0.549	0.745	var_i^*	0.140	0.117	0.098
		0.279	0.097			0.105	0.061
μ_p^*	0.590	0.491	0.551	var_r^*	0.450	0.261	0.330
		0.270	0.074			0.280	0.119
ρ_i^*	0.640	0.514	0.646	var_p^*	0.520	0.355	0.392
		0.180	0.059			0.437	0.137
ρ_g^*	0.910	0.778	0.905	var_w^*	0.240	0.188	0.152
		0.215	0.088			0.198	0.078

Note: This table reports estimators sensitivity to the starting values that drawn randomly in the interval between the posterior mode and the prior mean reported in Smets and Wouters (2007). SW denotes posterior modes from Smets and Wouters (2007). II is the Indirect Inference estimator without penalties. PII is the Penalized Indirect inference estimator with quadratic loss penalties centered around SW. We report means and the standard deviations of the estimates across 1000 replications. The data suggest less information on the parameters with asterisk, so penalty function has a positive curvature for them.