Marcelo Moura Jardim Teixeira Sena

Instrument Selection and Identification of Macroeconomic Equilibrium Conditions

DISSERTAÇÃO DE MESTRADO

DEPARTAMENTO DE ECONOMIA

Programa de Pós–graduação em Economia

Rio de Janeiro
March 2016

Pontifícia Universidade Católica
do Rio de Janeiro

Marcelo Moura Jardim Teixeira Sena

Instrument Selection and Identification of Macroeconomic Equilibrium Conditions

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–Graduação em Economia of the Departamento de Economia, PUC–Rio as partial fulfillment of the requirements for the degree of Mestre em Economia.

Orientador : Prof. Tiago Berriel
Co–Orientador: Prof. Marcelo Cunha Medeiros

Rio de Janeiro
March 2016

Pontifícia Universidade Católica do Rio de Janeiro

Marcelo Moura Jardim Teixeira Sena

Instrument Selection and Identification of Macroeconomic Equilibrium Conditions

Dissertation presented to the Programa de Pós–Graduação em Economia of the Departamento de Economia, PUC–Rio as partial fulfillment of the requirements for the degree of Mestre em Economia.

Prof. Tiago Berriel
Orientador
Departamento de Economia — PUC–Rio

Prof. Marcelo Cunha Medeiros
Co–Orientador
Departamento de Economia — PUC–Rio

Prof. Carlos Viana de Carvalho
Departamento of Economia — PUC–Rio

Prof. João Victor Issler
Fundação Getúlio Vargas

Prof. Monica Herz
Coordenadora Setorial do Centro de Ciências Sociais – PUC–Rio

Rio de Janeiro, March 31st, 2016.

Marcelo Moura Jardim Teixeira Sena

The author graduated in Economics from Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

Ficha Catalográfica

Moura Jardim Teixeira Sena, Marcelo

Instrument Selection and Identification of Macroeconomic Equilibrium Conditions / Marcelo Moura Jardim Teixeira Sena; orientador: Tiago Berriel; co–orientador: Marcelo Cunha Medeiros. — Rio de Janeiro : PUC–Rio, Departamento de Economia, 2016.

v., 72 f: il. ; 29,7 cm

1. Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui referências bibliográficas.

1. Economia – Tese. 2. Identificação Fraca. 3. Seleção de Instrumentos. 4. Estimadores de Encolhimento. I. Berriel, Tiago. II. Cunha Medeiros, Marcelo. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. IV. Título.

CDD: 330

Acknowledgments

To my advisors, Professor Tiago Couto Berriel and Professor Marcelo Cunha Medeiros, for the attention and friendliness. Their teachings were crucial, specially showing the subtleties of good economic research. The excited way with which they received the latest developments of my work were always source of extra motivation. I hope to continue working and contributing to my, their and our research, in any way possible, in the future.

I would like to thank Professor Carlos Viana de Carvalho and Professor João Victor Issler for the valuable comments and ideas which enhanced this work and that will contribute to future research agenda.

I thank all of the other professors in the Department of Economics of PUC-Rio, whom I learn from since my undergraduate studies. I specially mention Professor Gustavo Gonzaga and Professor Carlos Viana de Carvalho, who always motivated me to continue and deepen my studies, an experience that certainly enriched and gratified me.

To all my colleagues, from all the different years since undergraduate to the master's, that made this long period at PUC-Rio very pleasing and that I will certainly miss. Specially during the Master's years, Tiago and Tepedino (which I had the pleasure to study together since undergraduate years) and Pedro, Bruno, Moisés, Leandro and John. The discussions, be them academic or not, which fortunately were frequent, be it in the bus, department, favelinha, Pilotis or during lunch were certainly very enjoyable.

To all my family and friends, for understanding why on earth I got so involved with my studies, especially during the weekends. For my brothers, Gui and Dudu, my mother Beatriz and my grandfather Rubem, who passed away a few months before I finished and which would certainly be very proud of this achievement, I would like to express my thankfulness: without them anything I accomplished would not be possible. I certainly do not have the competence to do it with due fairness, so please allow me simplicity: thank you.

Finally, I would like to thank the whole staff from the Department of Economics of PUC-Rio, especially Graça, for quickly responding to all of my concerns. I gratefully acknowledge financial support from CAPES, CNPq and FAPERJ, which allowed me the peace of mind to work freely on my studies.

Abstract

Moura Jardim Teixeira Sena, Marcelo; Berriel, Tiago; Cunha Medeiros, Marcelo. **Instrument Selection and Identification of Macroeconomic Equilibrium Conditions**. Rio de Janeiro, 2016. 72p. MSc. Dissertation — Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Mavroeidis (2005) alerted that equilibrium conditions from typical rational expectations macroeconomic models could be weakly identified due to the use of poor instruments. I argue that, although such concerns are legitimate, they are not empirically severe, provided instruments are properly selected. I use an estimated medium scale *DSGE* model as a laboratory to assess single-equation estimation of macroeconomic equilibrium conditions. I present *LASSO*-based estimators to select instruments that perform well in finite samples, which I argue have a better chance of performing for forward-looking relationships, such as the New Keynesian Phillips Curve. Finally, I provide an empirical application of the estimators for the *US* economy's Phillips Curve and show that it validates a dominant forward-looking behavior.

Keywords

Weak Identification; Instrument Selection; Shrinkage Estimators;

Resumo

Moura Jardim Teixeira Sena, Marcelo; Berriel, Tiago; Cunha Medeiros, Marcelo. **Seleção de Instrumentos e Identificação de Equações de Equilíbrio Macroeconômico**. Rio de Janeiro, 2016. 72p. Dissertação de Mestrado — Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Mavroeidis (2005) alertou que equações de equlíbrio motivadas por modelos macroeconômicos com expectatvas racionais poderiam ser fracamente identficados devido ao uso de instrumentos fracos. Eu argumento que, embora tais preocupações sejam legítimas, elas não são empiricamente graves, contanto que instrumentos sejam devidamente selecionados. Eu utilizo um modelo *DSGE* estimado de média escala como laboratório para avaliar estimação uniequacional de condições de equilíbrio macroeconômicas. Apresento estimadores baseados no *LASSO* que selecionam instrumentos e tem boa performance em amostra finita, que argumento funcionam melhor em relações que incluem termos de expectativa, como a Curva de Phillips Novo Keynesiana. Por último, faço uma aplicação empírica para a Curva de Phillips da economia dos Estados-Unidos e as estimativas validam um componente de expectativa predominante.

Palavras–chave

Identificação Fraca; Seleção de Instrumentos; Estimadores de Encolhimento;

Sumário

Lista de figuras

Lista de tabelas

1
Introduction

Rational expectations micro-founded models have been the dominant framework in macroeconomic theory, linking aggregate phenomena to fundamental individual decision-making (through preferences and technology). They yield the so-called structural relations. More concretely, these relations are usually expectational difference equations derived from first order conditions of forward-looking economic agents and since they are part of a fully specified general equilibrium model, they are also equilibrium conditions. They imply few restrictions on the set of moment conditions, which poses a challenge to confront them with data. In such a setting, some of these moments may not be very informative, yielding weak identification. This is typically a manifestation of weak instruments, as seminally outlined by Mavroeidis (2005). I propose estimators to select instruments and show how they are relevant for proper identification of structural relationships.

For concreteness, I consider the estimation of a Phillips Curve, which has become an important example in the literature (see Mavroeidis et al. (2014)). The Phillips Curve represents the trade-off between inflation and some measure economic activity. In the canonical specification of New Keynesian models, the Phillips Curve describes the relationship of inflation (π_t) with inflation expectations of forward-looking agents ($E_t[\pi_{t+1}]$), lagged inflation motivated by some source of indexation (π_{t-1}), firm's marginal cost as a measure of economic activity (mc_t) and an unpredictable component, which may be thought as cost-push shock or data measurement error (ε_t):

$$\pi_t = w_f \mathbb{E}_t[\pi_{t+1}] + w_b \pi_{t-1} + \kappa mc_t + \varepsilon_t \tag{1-1}$$

Under this setting, I aim to answer the following questions. Is weak identification, as suggested by Mavroeidis (2005), empirically relevant? Can we select instruments to aid identification, avoiding the weak ones to obtain tolerable errors in finite samples? The two questions are connected, since weak identification may be a symptom of bad instrument choice. By addressing these questions empirically, I acknowledge the potential of weak identification and ask how relevant it is. I show that a rich enough data-generating process may lead to identification of forward-looking equilibrium conditions in limited information methods, provided instruments are properly selected. I use as laboratory a medium scale $DSGE$ and discipline the exercise by evaluating the model at its maximum likelihood estimate. I will also argue what makes such selection non-trivial, present estimators and show its good properties for a realistic macroeconomic sample size through Monte Carlo experiments.

How does the (empirically relevant) macroeconomic data generating process here considered corroborate identification and the selection methods? $DSGE$ models incorporate frictions that generally induce greater persistence in variables dynamics; persistent shocks, indexation and interesting rate smoothing are some examples. When variables are more persistent, the variance explained by the selected instruments is greater vis-à-vis unobservables; identification is strengthened. Equivalently, prediction of the endogenous variables is enhanced; given an information set, it is easier to predict a persistent variable than a random *i.i.d.* realization. The selection procedures are motivated by their documented abilities as good predictors.

I consider Generalized Method of Moments (GMM) estimation strategy. This choice allows the econometric framework to be seen as quite general. To quote Cochrane (2009), *"[it] is all GMM; the issue is the choice of moments."* I will consider a particular dimension of moment selection, which are those induced by the choice of instruments; the linearity of the equation being estimated makes the two stage least squares ($TSLS$) framework arise naturally. Even though GMM opens the possibility for more generality, through nonlinear estimation or a more general moment selection procedure, this work confines itself to linear and single equation estimation, the latter meaning that no auxiliary model structure is directly imposed.

Following the main exposition, I provide two further applications. In Section 7 I apply the estimator for another important macroeconomic relationship, the Taylor Rule. This case exemplifies how instruments selection methods possess advantages for forward-looking equations. Finally, in Section 8 I undertake a proper empirical exercise of the estimators I suggest and show that the instrument selection estimators agrees with the seminal estimation of Gali & Gertler (1999) for the US economy, where the forward-looking component dominates. In my concluding remarks, I point to possible interesting extensions.

2
Literature Review

2.1
Identification

Estimation of structural relationships is challenging and the literature has thoroughly explored the different ways to do it, but the debate on how to properly estimate these equations is ongoing. These were initially calibrated without any econometric estimation (Kydland & Prescott (1982)). Computational power and development of new techniques (GMM and Bayesian estimation, for example) allowed econometric estimation and inference. More recently identification became a concern: Canova & Sala (2009) alerted to identification issues in fully specified $DSGE$ models and Mavroeidis (2005) to the potential problem of weak identification of limited-information estimation, in particular caused by the use of weak instruments. This work focuses on the latter topic.

Mavroeidis (2005) concern was that, even though estimation aimed at using partial information (typically a single equation), identification would ultimately rest on extra information of the macroeconomic model, after all it determines the dynamics of all variables in equilibrium. A data-generating processes could, for example, induce weak enough instruments that would produce estimates of a dominant forward-looking component for a predominantly backward-looking Phillips Curve, even for unrealistic large samples. The purpose here is to take a step back on this critique; a data generating process which yields weak identification of the Phillips Curve may not be representative of the data that econometricians actually collects. A more realistic data generating, featuring commonly used macro frictions that has been shown to reproduce actual data with some degree of accuracy may provide a better identification environment. In this sense, the motivation is very similar to that presented in Krogh (2015). Frictions incorporated in macroeconomic models typically induce persistent dynamics, similar to what is observed empirically. This may quantitatively allow for identification of single-equation estimation.

Instrument selection becomes relevant in this environment since poor identification may be a symptom of bad instrument choice. When considering a medium-scale $DSGE$ model, choosing instruments is non trivial. This is so for two main reasons: there is a plethora of variables available and many of them may be weak predictors. This is a reasonable description of the challenges faced by an empirical macroeconomist; data availability has increased over the past years and macroeconomic variables are empirically hard to predict, see Medeiros & Vasconcelos (2016) and Stock & Watson (2007) respectively. Note that the predictability problem is directly linked to the search for instruments; classic simultaneity issues and/or rational expectations restricts candidate instruments to be lagged variables. I consider here factors and mainly, shrinkage estimators, since they have a documented good performance for forecasting, see Li & Chen (2014).

2.2
Limited Information vs Full Information

Limited-information, in particular single equation methods which is the focus here, has the advantage of incorporating minimum information possible which translates to estimates less susceptible to model misspecification. When identification is weak, one possibility is carrying out robust inference. An example is Dufour et al. (2006) who inverts a Anderson & Rubin (1949) type test to obtain a confidence interval valid whether the model is identified or not, but not necessarily size correct with weak identification. Dufour et al. (2013) compares single equation with (weak) identification-robust analysis for full and other limited information methods. Kleibergen & Mavroeidis (2009) reviews other identification robust tests. The disadvantage of carrying out such methods of inference is that robustness come at the cost of wider intervals, which may not be very informative. For this work I do not provide formal inference theory here and distributions are obtained through simulations.

An alternative estimation strategy considers a full information environment. These procedures may overcome weak identification more easily since it places further restrictions on dynamics of the observables; the drawback is the greater risk of model misspecification. Leeper & Sims (1994) is an early attempt to this task. An & Schorfheide (2007) and Smets & Wouters (2007) are more recent examples using Bayesian methods; the former incorporates nonlinear filtering while the later estimates a model of similar scale as the one used here and shows that it predicts time-series competitively with reduced forms. Lindé (2005) argues via Monte-Carlo exercise that a full information maximum likelihood is preferable for estimation of the New Keynesian Phillips Curve and that it is robust to some misspecification. Recent work by Andrews & Mikusheva (2015) however shows that even such methods may suffer from weakly identified parameters, which manifests itself via a near-flat likelihood function, inducing distortions for calculations of the Fisher information.

There are other alternative estimation strategies which departs from the framework here considered. One is using survey data, such as Roberts (1995), which allows rational expectations hypothesis to be abandoned. Another one is considered by Mavroeidis et al. (2014) who uses different data vintages as “external instruments". Anyhow, both approaches are still potentially exposed to weak instruments and the selection procedures I outline can still, in principle, be applied to these cases.

2.3 Variable Selection

Another possibility to overcome weak identification would be to, vaguely speaking, provide a judicious choice of variables. Fuhrer & Olivei (2005) suggests improving this selection by imposing dynamic constraints of an auxiliary model structure; they choose a parsimonious reduced form for such. In a similar fashion, Dees et al. (2009) motivates instruments from a global vector autoregression ($GVAR$) structure that alleviate the weak instruments problem.

I consider factors and shrinkage estimators for selection. The former was motivated Ng & Bai (2009), who obtains instruments by assuming that variables possess a non-observable factor structure. In this environment, they propose estimating the principal components from a rich panel of instruments to construct a feasible instrumental variable estimator. Shrinkage estimators have become popular recently by the introduction of computationally efficient algorithms (for example, the Least Angle Regression($LARS$)). I use variants of the Least Absolute Shrinkage and Selection Operator ($LASSO$), originally proposed by Tibshirani (1996). The estimator will shrink coefficients to exactly zero in a linear model. The two variants considered here are the adaptive $LASSO$ and the Group $LASSO$. The former was proposed by Zou (2006), who showed that it attains the Oracle property[1]. The latter was introduced by Yuan & Lin (2006) when there is a natural grouping of variables; here, these will be lags of the same observable. Medeiros & Mendes (2015) shows how the adaptive $LASSO$ estimator performs well in a more general non Gaussian and heteroskedastic time-series environment.

I also consider the Ridge, an analytically convenient and parsimonious form of regularization (Tikhonov Regularization). These have been considered as an alternative in a many instruments environment, with potentially weak ones. Carrasco & Tchuente (2015) proposes such type of regularization to avoid finite sample deterioration of an excessive number of instruments. Carrasco & Tchuente (2015) proposes a similar approach for a many weak in-

[1]The penalized estimator is asymptotically equivalent to the Oracle estimator, which is the unpenalized estimation obtained as if the relevant variable were known beforehand.

struments environment. Hansen & Kozbur (2014) proposes a jackknife instrumental variable estimator with regularization, when the number of instruments is potentially larger than sample size.

2.4 Relationship to Literature - Main Differences

Krogh (2015) also highlights the importance of macro frictions. His results are via simulations to a stylized model; I obtain some analytical, albeit simple, results in the context of the Basic New Keynesian model. I relegate simulations to an empirically disciplined medium-scale model which provides empirical support for proper identification of the New Keynesian Phillips Curve, while his model is calibrated without formal likelihood methods. I show that this has important consequences and that his results do not generalize to the larger model considered here. Inference-wise he builds on the robust procedure based on the S-statistic of Stock & Wright (2000) while I use instrument selection where I conjecture classical asymptotic approximations have better chance of being reliable.

Hall (2005) contains an initial discussion on moment and instrument selection. There has already been applications with the $LASSO$; for example, Cheng & Liao (2015) adds a slackness parameter to moment conditions to select the valid ones. My application is not a matter of choosing valid moments; I assume that the candidate instruments induce valid moment conditions. There is no obvious "Oracle" benchmark; rational expectations places very few restrictions on the set of possible instruments. Selection is made on a relevance (efficiency) criterion and ideally, inference can be obtained in the spirit of Belloni et al. (2010) who derives estimators based on the $LASSO$ and $post-LASSO$ that selects optimal instruments in a linear instrumental variable and high dimensional setting. I use $post-Lasso$ methods to form first-stage predictions and apply these to a GMM framework, obtaining finite sample distributions via simulations.

3
Identification Strength in the Basic New Keynesian Model

In ths section I consider the identification strength of the Phillips Curve parameters when the data generating process is the Basic New Keynesian model of Galí (2008). This model is a parsimonious micro founded model of an economy and allows for analytical solution once linearized. Moreover, it contains the core equations and mechanisms embedded in any type of New Keynesian model: a Consumption Euler Equation, a Phillips Curve and a Taylor Rule. The log-linearized equations, which are necessary conditions for the equilibrium dynamics of the variables, are presented below:

$$\pi_t = \beta E_t[\pi_{t+1}] + \kappa y_t \tag{3-1}$$

$$y_t = -\frac{1}{\sigma}(i_t - E_t[\pi_{t+1}] - r_t^n) + E_t[y_{t+1}] \tag{3-2}$$

$$i_t = \rho_i + \phi_\pi \pi_t + \phi_y y_t + v_t \tag{3-3}$$

where π_t is the inflation rate, y_t is output gap, i_t is the nominal interest rate, r_t^n is the natural rate of interest linked to technology shocks, v_t a monetary policy shock and E_t the mathematical expectations operator.

3.1
Ordinary Least Squares (OLS) Estimation

To see the root of the problem, consider first the inconsistent OLS estimator. If $E_t[\pi_{t+1}]$ was observed, then there would be no endogeneity and $\begin{pmatrix}\beta & \kappa\end{pmatrix}$ could be easily recovered by OLS (it would actually be deterministic equation). Assume then that $E_t[\pi_{t+1}]$ is not observable and interest lies in the parameter β. Define the expectational error η_{t+1}:

$$\pi_{t+1} = E_t[\pi_{t+1}] + \eta_{t+1} \tag{3-4}$$

which by construction satisfies $E_t[\eta_{t+1}] = 0$. The OLS estimator is:

$$\boldsymbol{x}_t = \begin{pmatrix}\pi_{t+1} & y_t\end{pmatrix}' \tag{3-5}$$

$$\begin{pmatrix}\hat{\beta}_{OLS} \\ \hat{\kappa}_{OLS}\end{pmatrix} = \left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \pi_t\right) \tag{3-6}$$

Proposition 3.1 *In the Basic New Keynesian model, the OLS estimator of the forward-looking component of the Phillips Curve β is asymptotically attenuated.*

$$\hat{\beta}_{OLS} \xrightarrow{p} \beta \left(\frac{\sigma^2_{r^*}}{\sigma^2_{r^*} + \sigma^2_{\eta}} \right) < \beta \tag{3-7}$$

where $\sigma^2_\eta = E(\eta^2_{t+1})$ and $\sigma^2_{r^*}$ is the projection error variance from π_{t+1} on y_t.

The proof is given in Appendix A and is a direct application of an errors-in-variables argument. It is straightforward to extend the argument with more exogenous variables (for example, a backward-looking component) and a cost-push shock, which encompasses the more general specification (1-1). The result remains even in finite-sample simulations, as I show in Section 6. Note that if expectational errors are not very variable, in particular, if agents rarely get their predictions wrong, then an economy's structural Phillips Curve can be precisely estimated by OLS regressions. Transparent economic environment has the additional benefit of making econometric estimation easier! This relates to the more general point that OLS may not be as bad an estimator as one might think; so long as the degree of endogeneity (here measured by σ^2_η) is not so big, OLS provides a good approximation to the true parameter, even though it is inconsistent. For some macroeconomic conditions, this does seem to be corroborated empirically (see Taylor Rule case in Section 7). Still, this is a matter to be settled empirically and not known in advance. Instruments are potentially needed, to which I turn next.

3.2 Instrumental Variables

The typical strategy to estimate the Phillips curve uses GMM framework, but here I will specialize to the $TSLS$ procedure which allows for a more intuitive interpretation of instruments as a signal in the first stage. The reduced form of the model will indicate relevant instruments, since a lagged variable w_{t-1} will necessarily satisfy the exogeneity restriction:

$$E[w_{t-1}\eta_{t+1}] = 0 \tag{3-8}$$

It is in this spirit that Mavroeidis (2005) alerted how identification would rest on the dynamics of what he called "driving variables", which is ultimately determined by the joint equilibrium dynamics of all variables. For this model, substituting the interest rate rule:

$$\pi_t = \beta E_t[\pi_{t+1}] + \kappa y_t \tag{3-9}$$

$$y_t = -\frac{1}{\sigma}(\rho_i + \phi_\pi \pi_t + \phi_y y_t + v_t - E_t[\pi_{t+1}] - r_t^n) + E_t[y_{t+1}] \tag{3-10}$$

In the Basic New Keynesian Model, technology and monetary policy shocks cannot be identified separately. Define the composite shock $\varepsilon_t \equiv r_t^n - v_t - \rho_i$. Undetermined coefficients (Christiano (2002)) guess yields:

$$\pi_t = \psi_\pi \varepsilon_t \tag{3-11}$$

$$y_t = \psi_y \varepsilon_t \tag{3-12}$$

The solution to this guess is given in Appendix A.2.1. Identification analysis of β is more promptly done from these solutions, which may be viewed as the data generating process (or reduced form) for the observed endogenous variables. Identification and its strength will rest on the instruments predictive power. What follows is similar in spirit to that done in Nason & Smith (2008).

If these shocks are assumed $i.i.d.$, then no consistent estimation procedure is possible since there are no instruments that satisfy the relevance condition for π_{t+1}. Identification may be recovered under assumptions on non-observables, as usual in econometrics. To be more specific about this case, it rests upon assumptions we are willing to impose on ε_t.[1] Assume now that ε_t follows an $AR(1)$ process. In this case, the solution still has the form given in (3-11), with coefficients given in Appendix A.2.2:

$$\pi_t = \psi_\pi \varepsilon_t \tag{3-13}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \tag{3-14}$$

$$\implies \pi_t = \rho \pi_{t-1} + \psi_\pi u_t \tag{3-15}$$

Leading (3-15) one period and substituting for ε_t:

$$\pi_{t+1} = \rho^2 \pi_{t-1} + \rho \psi_\pi u_t + \psi_\pi u_{t+1} \tag{3-16}$$

Equation (3-16) motivates π_{t-1} as an instrument for π_{t+1}. Identification

[1]It is in fact true that the shock ε_t is tied to more structural relationships from the deeper microfoundations of the model (it is a linear combination of the shocks v_t and r_t^n, where the latter is a function of the "even more structural" technology shock a_t). To simplify exposition we will at this point allow ourselves to impose some arbitrary dynamics into this shocks, reminding that the more proper way to proceed would be design the model that would "naturally" allow those dynamics.

strength is an empirical matter, in this case determined by the persistence parameter ρ. Quantitatively, there are reasons to believe that macroeconomic variables potentially generate enough persistence. As a matter of fact, $DSGE$ models that fit data as well as reduced forms models achieves this by mimicking real world phenomena with frictions, which typically introduce persistence: price and wage indexation, consumption habits, interest rate smoothing and adjustment costs are common examples in the modern $DSGE$ model literature. Moreover, some residual dynamics is usually assumed in the processes. A typical example is calibration of the exogenous technology shocks, which is very hard to model.

3.3 Strength of Identification

Stock & Yogo (2005) use the concentration parameter to motivate statistical tests of weak identification for linear instrumental variables model. The concentration parameter μ^2 is a measure of signal to noise ratio of the instrument for the endogenous variable and I use the following definition (see Kleibergen & Mavroeidis (2009)):

$$\mu^2 \equiv mineval(T\mathbf{\Sigma}_{VV}^{-\frac{1}{2}}\mathbf{\Pi}'\mathbf{\Sigma}_{ZZ\perp X}\mathbf{\Pi}\mathbf{\Sigma}_{VV}^{-\frac{1}{2}}) \tag{3-17}$$

where $mineval(\mathbf{A})$ obtains the minimum eigenvalue of matrix $\mathbf{A}$ and the reduced form for the endogenous variables $\boldsymbol{y}_t$ is written as:

$$\boldsymbol{y}_t = \boldsymbol{z}_t'\mathbf{\Pi} + \boldsymbol{x}_t'\mathbf{\Phi} + \boldsymbol{\nu}_t \tag{3-18}$$

with the moment $\mathbf{\Sigma}_{ZZ\perp X}$ defined as:

$$\mathbf{\Sigma}_{ZZ\perp X} = \mathbb{E}[\boldsymbol{z}_t\boldsymbol{z}_t'] - \mathbb{E}[\boldsymbol{z}_t\boldsymbol{x}_t']\mathbb{E}[\boldsymbol{x}_t\boldsymbol{x}_t']^{-1}\mathbb{E}[\boldsymbol{x}_t\boldsymbol{z}_t'] \tag{3-19}$$

Here, $\boldsymbol{z}_t$ are the instruments and $\boldsymbol{x}_t$ are other exogenous variables of interest in the structural relationship. For the Basic New Keynesian model example we consider, there are no other exogenous variables in the reduced form for π_{t+1}.[2]

Proposition 3.2 *In the Basic New Keynesian model, the forward-looking parameter β is identified if the unobservable ε_t follows an $AR(1)$ process. Identification strength is increasing with persistence $\rho \in (0,1)$:*

$$\mu^2 = T\frac{\rho^4}{1-\rho^4} \tag{3-20}$$

[2] In a typical $TSLS$ procedure, one would include the other exogenous variable of interest in the structural relationship to estimate the reduced form; in this case y_t potentially. For this example I assume that it is known in advance that y_t has no predictive power for π_{t+1} once π_{t-1} is controlled for, i.e., it has a zero coefficient in the reduced form.

The proof is given in Appendix A.3. This is actually quite intuitive for an AR process: as persistence increases, the variance of lagged variable dominates the noise. Note that trying to instrument a leading endogenous variable makes the search for strong instruments more challenging; this why the quartic rate appears in the expression. Since $\rho \in (0,1)$, it actually makes a force against identification strength. Identification becomes weaker the further in the future is the endogenous variable. This is also quite intuitive: lagged inflation becomes a weaker signal for inflation in the distant future. For completeness I provide the proof in Appendix A.4.

For a single endogenous variable and instrument, $\mu^2 \geq 9$ would suffice.[3] For a sample size of $T = 100$, which for quarterly data represents 25 years, this occurs at approximately $\rho = 0.54$. Note also that here the exogenous persistence is the *only* model mechanism contributing to single-equation identification. Also, the fact that identification improves substantially as $\rho \to 1$ suggests that non-stationarity can be an important source of identification. Some work in this direction are Magnusson & Mavroeidis (2013) who proposes a GMM estimator that exploits instability in the data and Gorodnichenko & Ng (2010) proposes $DSGE$ estimation for persistent data, having near or exact unit roots.

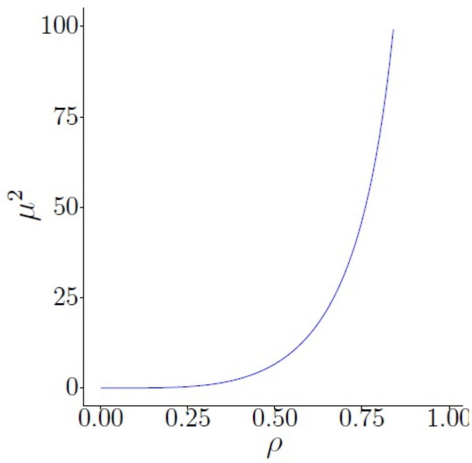


Figure 3.1: Concentration Parameter (μ^2) and Persistence (ρ). $T = 100$

As a first experiment, if we empirically project inflation on its lagged component with quarterly data, we obtain:

[3]The concentration parameter divided by the number of instruments is approximately equal to $E(F) - 1$ where F is the first stage F statistic. When instruments are completely irrelevant, then the first stage F statistic is equal to 1. Staiger & Stock (1997) establishes a general threshold of $F = 10$ for the case of a single endogenous regressor. See Kleibergen & Mavroeidis (2009) for a summary.

$$\pi_t = \underset{(17.49)}{0.787}\pi_{t-1} + \hat{e}_t$$

Quarterly % Change, Average Aggregation Method. 01/01/1980 to 01/07/2015. US. Bureau of Labor Statistics, *Consumer Price Index for All Urban Consumers: All Items [CPIAUCSL]*, retrieved from FRED, Federal Reserve Bank of St. Louis. t-statistic in parenthesis.

I do not intend take the above result as conclusive evidence of strong identification; the threshold $\rho = 0.54$ is obviously contingent on the Basic New Keynesian stylized example. Nevertheless, the above result gives us some quantitative hint not to be too pessimistic about macroeconomic limited information estimates. I end this section by positing that persistence as a necessary condition for identification is not a new result. This was shown in Mavroeidis (2005): identification requires "enough" persistence, in a well-defined sense. The point here is that "some" persistence might just be enough. The same type of state variables that are introduced in larger models by the inclusion of frictions to better fit the data will generate persistence, which may also indirectly provide support for identification of limited information methods. Reminding that persistence improves identification by making variables "more predictable", in the next section I turn to the related problem of how to select these predictors.

4
Instrument Selection

4.1
Challenges

I have argued that for a rich enough data generating process there are variables the econometrician can use as instruments in order to properly estimate and identify structural relationships. The Basic New Keynesian Model example can be misleading nonetheless, since the econometrician will never know exactly which instruments to use and usually is not willing provide structural relationships which would corroborate a specific one. If a full structural model were to be postulated, there would no need to carry out limited information estimation in the first place.

While interesting to highlight the mechanisms of monetary policy, the Basic New Keynesian Model is stylized and unfit to make quantitative assertions regarding monetary policy. In order to do that, one would require a larger and more realistic model. Naturally, for larger models there is a larger number of variables which increases the set of candidate instruments. An excessive number of instruments consumes valuable degrees of freedom, especially for the small samples that may be typically encountered in macroeconomic time-series.

Finally, within a large number of candidate variables, some of them may actually be valid, but produce a very weak signal as an instrument. This weak instrument problem is common in macroeconomics and it may induce very high variance in estimators or severe finite simple bias. Weak instruments occurs when where there is so little predictive content that it is not even practical to frame it as a finite sample issue; estimators behave very poorly even for a huge sample size, as in Bound et al. (1995). A bad instrument choice is sufficient to conclude that a parameter is weakly identified.

Taking these issues into consideration, one must choose an instrument selection procedure that satisfies the following criteria:

i) Does not require full model specification.

ii) Chooses a parsimonious number of variables, especially when sample size is small.

iii) Chooses relevant instruments and discards the weak ones, producing consistent estimators with tolerable variance.

4.2 Estimation

I consider GMM estimation strategy. I now use the more general specification (1-1). Using the expectational error definition (3-4):

$$\pi_t - w_f\pi_{t+1} - w_b\pi_t - \kappa mc_t + (w_f\eta_{t+1} - \varepsilon_t) = 0 \tag{4-1}$$

Rational expectations implies:

$$\mathbb{E}\left[\pi_t - w_f\pi_{t+1} - w_b\pi_{t-1} - \kappa mc_t | \boldsymbol{\mathcal{F}}_t\right] = 0 \tag{4-2}$$

$$\implies \mathbb{E}\left[(\pi_t - w_f\pi_{t+1} - w_b\pi_{t-1} - \kappa mc_t)\, \boldsymbol{z}_{t,s}\right] = 0, \quad \forall \boldsymbol{z}_{t,s} \in \boldsymbol{\mathcal{F}}_t \tag{4-3}$$

where $\boldsymbol{\mathcal{F}}_t$ is the information set at period t and s indexes the selection procedure of the instruments $\boldsymbol{z}_{t,s}$. The parameter estimate is:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{g}_t(\boldsymbol{\theta})\right)' \hat{\boldsymbol{W}} \left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{g}_t(\boldsymbol{\theta})\right) \tag{4-4}$$

$$u_t(\boldsymbol{\theta}) = \pi_t - w_f\pi_{t+1} - w_b\pi_{t-1} - \kappa mc_t \tag{4-5}$$

$$\boldsymbol{g}_t(\boldsymbol{\theta}) = u_t(\boldsymbol{\theta}) \otimes \boldsymbol{z}_{t,s} \tag{4-6}$$

$$\boldsymbol{\theta} \equiv \begin{pmatrix} w_f & w_b & \kappa \end{pmatrix}' \sim p \times 1 \tag{4-7}$$

with $\hat{\boldsymbol{W}}$ an estimated weighting matrix.

This is the *Generalized Instrumental Variables* estimation environment, following Hansen & Singleton (1982). Nothing up to this point has been said about the identity of the instruments $\boldsymbol{z}_{t,s}$ or how they are selected. In the $TSLS$ framework, the instruments must satisfy an exogeneity and a relevance condition. Note that the model information used by the econometrician places only the very weak condition (4-2) on the former. Since the instruments $\boldsymbol{z}_t$ may include not only lagged variables, but also arbitrary transformations thereof, exclusion restrictions are not very informative. Using relevance as a choice criterion is more fruitful, as has been tackled by the weak instruments literature. Near irrelevant instruments renders the estimation exercise futile; regular tests will yield misleading sizes or if robust procedures are used, intervals may be too wide to be informative.

A relevant instrument set, vaguely defined, will provide enough information about the “likelihood" of the moment (how close it is to being satisfied) given a range of parameter values. Asymptotically, if $g_t(\boldsymbol{\theta})$ is a martingale difference sequence then the optimal instrument set[1] $\boldsymbol{z}^o_{t-1}$ of moment condition $g_t(\boldsymbol{\theta})$ is given by (Theorem 7.2, Hall (2005))

$$\boldsymbol{z}^o_{t-1} = K\mathbb{E}\left[\frac{\partial u_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\middle| \boldsymbol{\mathcal{F}}_t\right]\Sigma^{-1}_{u|\boldsymbol{\mathcal{F}}_t} \tag{4-8}$$

where K is a $p \times p$ constant matrix and $\Sigma^{-1}_{u|\boldsymbol{\mathcal{F}}_t} = \mathbb{E}[u_t(\boldsymbol{\theta})u_t(\boldsymbol{\theta})|\boldsymbol{\mathcal{F}}_t]$. This result, however, has limited practical implications: it potentially depends on true parameter values and may require further assumptions or knowledge of the data generating process to construct a feasible counterpart. Moreover, it is an asymptotic results and there is no guarantee that it performs well in finite samples. A thorough discussion is provided in Hall (2005).

I take a more pragmatic stand as a motivation to instrument selection. As (4-8) indicates and observed in Belloni et al. (2010): "*Optimal instruments are conditional expectations... LASSO and post-LASSO [obtain] estimates of nonparametric conditional expectations functions...* ". In my application, non parametrics are not needed, since I have a linear moment condition and with a rich enough set of predetermined variables in $\boldsymbol{\mathcal{F}}_t$ the dynamics of the $DSGE$ model can be properly approximated by a finite order VAR. The proper conditions for an exact VAR representation is given in Fernández-Villaverde et al. (2005); a sufficient condition is that all state variables are observed. This is then analogous to choosing variables from what would be a first stage regression of the endogenous variables on the instruments. In the spirit of $TSLS$, a linear predictive model for the endogenous variable. For the sake of parsimony, the conditioning information I consider, the candidate instruments, are lags of observables. An interesting extension would be to see if enlarging this candidate set with further transformations, as in Belloni et al. (2010), could enhance estimation.

4.3 Selection Methods

I cosider 9 methods do select instruments, 5 of which are data-driven. The other 4 will serve as benchmarks. Denote $\boldsymbol{Z}$ the $T \times K$ design matrix of candidate instruments $\boldsymbol{z}_t$ (this encompasses lags as well, so if there are N observables and L lags, $K = NL$).

4.3.1 Benchmark Procedures

I) *Ad Hoc*

[1]The estimator attains the minimum asymptotic variance.

This is the most common procedure. The instrument set is usually motivated by a specific economic model or prior economic knowledge. Following the what has been typically considered in the literature (especially in simulations exercises), I consider a similar set:

$$Ad\ Hoc \to \{\pi_{t-l}, mc_{t-l}, i_{t-l}\}_{l=1}^{3} \tag{4-9}$$

where i_t is the nominal interest rate. Note that prior knowledge of the New Keynesian data generating process heavily influenced this choice, even without any formal statistical procedure.

II) *All*

All instruments $\boldsymbol{z}_t$ are used. This is a minimal benchmark that any selection procedures should beat. Note that here this amounts to a total of 27 instruments (9 observables and 3 lags of each). As a reference, the seminal estimation of Gali & Gertler (1999) uses a similar number of instruments, a total of 24 (6 observables and 4 lags of each).

$$All \to \{\boldsymbol{z}_t\} \tag{4-10}$$

III) *Random*

A number N_R of instruments are selected at random (uniformly) from the columns of $\boldsymbol{Z}$. Note that this is only supposed to be a benchmark to lend support for the importance of proper selection. One could try and devise a selection method that would take advantage of randomness, in the spirit of bagging for example, but this is not the aim here. I set $N_R = 5$.

IV) *OLS*

Ordinary least squares allows us to see the “net" gains from each selection procedure. In principle, a bad instrument set could lead to less reliable estimates, with finite sample error even larger than that induced by endogeneity. The instrument set is trivial:

$$OLS \to \{\pi_{t+1}, \pi_{t-1}, mc_t\} \tag{4-11}$$

4.3.2 Data Driven Procedures

I) Adaptive *LASSO* (*adaLASSO*)

The adaptive Least Absolute Shrinkage and Selection Operator (*adaLASSO*) was proposed by Zou (2006). The selection is made by shrinking coefficients of the reduced form for each endogenous variable, with parameter specific penalty added to the usual *OLS* objective function:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{j=1}^{p} w_j \|\beta_j\|_1, \tag{4-12}$$

where $\|\|_p$ is the ℓ^p norm and $\boldsymbol{y}$ is $T \times 1$ stacking the endogenous variable. λ controls the general shrinkage and $w_j = |\tilde{\beta}_j|^{-\tau}$ is a parameter specific penalty weight determined by a preliminary estimator $\tilde{\beta}_j$. I set $\tau = 1$ and use the Ridge as the preliminary estimator $\tilde{\beta}_j$. The selected instruments are those associated with the nonzero coefficients. I employ two forms of selection via the *adaLASSO*:

- *adaLASSO Observables*: the candidate set is restricted to only the first lag of observables; once the selection is made, 3 lags are used for estimation.
- *adaLASSO Lags*: the candidate set is composed of 3 lags of all observables.

This separation allows us to more properly distinguish the estimator's performance regarding observables and lag selection.

II) *Group LASSO*

The *Group LASSO* was proposed by Yuan & Lin (2006) and is similar to the *adaLASSO*, except that selection is made on a group level. The estimator is given by:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \sum_{j=1}^{J} \boldsymbol{Z}_j\boldsymbol{\beta}_j\|_2^2 + \lambda \sum_{j=1}^{J} \|\boldsymbol{\beta}_j\|_{K_j} \tag{4-13}$$

where j indexes a group of variables, $\boldsymbol{\beta}_j$ is $p_j \times 1$ and the norm $\|\boldsymbol{\beta}_j\|_{K_j}$ is defined as:

$$\|\boldsymbol{\beta}_j\|_{K_j} = (\boldsymbol{\beta}'\boldsymbol{K}_j\boldsymbol{\beta})^{\frac{1}{2}} \tag{4-14}$$

Following Yuan & Lin (2006), I set $\boldsymbol{K}_j = p_j \boldsymbol{I}_{p_j}$ with $\boldsymbol{I}_n$ being the $n \times n$ identity matrix. Groups are determined as lags per observable. For example:

$$Group\ j : \boldsymbol{z}_{j,t} \rightarrow \{mc_{t-1}, mc_{t-2}, mc_{t-3}\} \tag{4-15}$$

III) *Ridge*

ℓ^2-regularization was considered by Carrasco & Tchuente (2015) to circumvent finite sample deterioration due to the inclusion of an excessive number of instruments. I consider here a particular case, the *Ridge* estimator (Tikhonov Regularization).

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{j=1}^{p} \beta_j^2, \tag{4-16}$$

Coefficients are shrunk to zero, but will never be exactly zero. The selected instruments are the fitted values of the endogenous variables induced by the estimates (4-16) (as well as π_{t-1}, otherwise the equation is under identified).

IV) *Factors*

Ng & Bai (2009) assume that the instrument set share a common factor structure. In this setting, the ideal instruments are the common components, which may be obtained by factor analysis.

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{\Lambda} + \boldsymbol{u} \tag{4-17}$$

where $\boldsymbol{F}$ is $T \times N_F$ and $\boldsymbol{\Lambda}$ is the $T \times K$ matrix of factor loadings. The selected instruments are the estimated factors $\hat{\boldsymbol{F}}$, the first N_F principal components of $\boldsymbol{Z}$. I set $N_F = 3$.

5
Monte-Carlo Exercise

I now describe the Monte-Carlo exercise and its configurations. The idea is simple: to estimate the Phillips Curve from a $DSGE$ model using (4-4) and the selection procedures described in the previous sections.

5.1
Model

The $DSGE$ model is based on Fernández-Villaverde (2010) and described more thoroughly in Appendix B.1. It is as medium scale model, with the following commonly employed frictions:

- Consumption habits.
- Wage rigidity.
- Endogenous labor supply.
- Price and wage indexation.
- Interest rate smoothing.
- Capital (and depreciation).
- Investment adjustment costs.
- Persistent shocks.

This data generating process emulates the estimation exercise of a typical econometrician more accurately. I have argued in Section 3 that these commonly used macroeconomic frictions may provide support for identification due to the persistent behavior. Even more importantly, it is empirically corroborated, in the sense that their introduction allows the $DSGE$ model to mimic the real world data generating more precisely; Smets & Wouters (2007) showed how a similar model could predict macroeconomic series competitively.

5.1.1
Phillips Curve

For simulations I consider the linearized version of the model. Consequences of nonlinear data generating process on identification and instrument selection is an interesting possible extension. Once log-linearized, the model yields the same hybrid specification of (1-1) with the following relationship with the structural parameters (I use lowercase for variables which are written as log deviations from steady state):

$$\pi_t = w_f \mathbb{E}_t[\pi_{t+1}] + w_b \pi_{t-1} + \kappa mc_t + \varepsilon_t \tag{5-1}$$

$$Var(\varepsilon_t) = \sigma_c^2 \tag{5-2}$$

$$w_f = \frac{\beta}{1+\beta\chi} \tag{5-3}$$

$$w_b = \frac{\chi}{1+\beta\chi} \tag{5-4}$$

$$\kappa = \frac{1-\theta_p-\beta\theta_p\left(1-\theta_p\right)}{\left(1+\beta\chi\right)\theta_p} \tag{5-5}$$

The original model in Fernández-Villaverde (2010) actually does not have the cost-push shock ε_t, in which case $\sigma_c^2 = 0$. In order to conform the exercise with the typical specification used in the literature, I add a cost-push shock to the model and discipline it by calibrating σ_c^2 using Smets & Wouters (2007) estimate. Robustness exercises have shown that altering this parameter will, naturally, quantitatively affect strength of identification and estimators errors, but does not alter significantly the qualitative ranking amongst them. Future work could extend this exercise for actual policy-making models, making use for example of Cwik et al. (2012), allowing for more general specifications.

5.2 Calibration

There is a notorious trade-off for identification regarding data variance. If variance from shocks driving the model are too small, observed data has very little variation if the steady state is reached quickly and estimation will be either very sensitive to initial conditions; in the limit if there is no variance at all, there is no identification. If it is very high, there is very little signal coming from observables and estimation is very sensitive to the particular history of shocks. In summary, strength of identification of the internal mechanisms of the model ultimately rests on the signal-to-noise ratio coming from the data.

With this in mind and to make the exercise consistent with the above mentioned fact that $DSGE$ models are a good approximation of an economy's data generating process, the chosen parameter values are taken from Fernández-Villaverde (2010), which took the model to data in a full-information procedure. The calibration generates quarterly time series and allows us to have a quantitative feel whether estimation with limited information are in fact reliable given currently available samples sizes. I provide the calibrated values in Appendix B.2. In what follows, variables with a tilde (for example $\widetilde{\chi}$) refer to these calibrated values.

5.3 Monte-Carlo Setup

I assume that the following 9 time-series are observed:

- c - consumption
- y^d - aggregate product
- i - nominal interest rates
- Π - inflation
- x - investment
- mc - marginal cost
- k - capital
- w - real wage
- v^w - wage dispersion

There is nothing special about the choice of these series; it is arguably a reasonable set of observables inside the information set of econometrician. A natural extension at this point would be to confront these estimators with varying information sets, in particular, in a more data rich or high dimensional environment for example. Although this is in fact interesting, for the estimation exercise here considered, the econometrician usually has enough prior knowledge to restrict the observables to a more concise set. Unless stated otherwise, the candidate instruments $\boldsymbol{Z}$ is composed of 3 lags of each observable.

One important observation is that this choice does *not* allow the econometrician to observe all relevant state variables from policy functions. Hence, following Fernández-Villaverde et al. (2005), a well specified finite order VAR cannot be estimated for the endogenous variables reduced forms. This places estimators in the realm of misspecified models since I do not estimate moving averages.

I run 500 replications. Data simulations are done in `Dynare` (Adjemian et al. (2011)). A preliminary portion of the time series is burned to allow for a stochastic initial state. Estimation is carried out with two sample sizes: $T = 150$ and $T = 1000$, which are interpreted as a "realistic" samples size (equivalent to 37.5 years of quarterly data) and the "large sample size", respectively. All cross validations schemes for the penalty parameter λ regarding shrinkage estimators are done using the *Bayesian Information Criterion* (BIC) using a Gaussian likelihood. GMM estimators are obtained using the optimal weighting matrix iterated to convergence as in Hansen et al. (1996), using the HAC covariance matrix estimator with the *Bartlett* kernel and bandwidth b_T set according to the common *Bartlett* value

$$b_T = int\left[4\left(\frac{T}{100}\right)^{\frac{1}{4}}\right], \tag{5-6}$$

where $int[.]$ is the integer part.

Finally, when searching for instruments, I assume that π_{t+1} and mc_t are potentially endogenous, while π_{t-1} is known to be exogenous.

6
Results

The results section is broken in two further sections, which follows from the discussion of Section 3 and Section 4 respectively. The first part shows how the more realistic data generating process corroborates proper identification of the Phillips Curve whilst the second part of the simulations links to this discussion by showing the importance of proper choice of instruments.

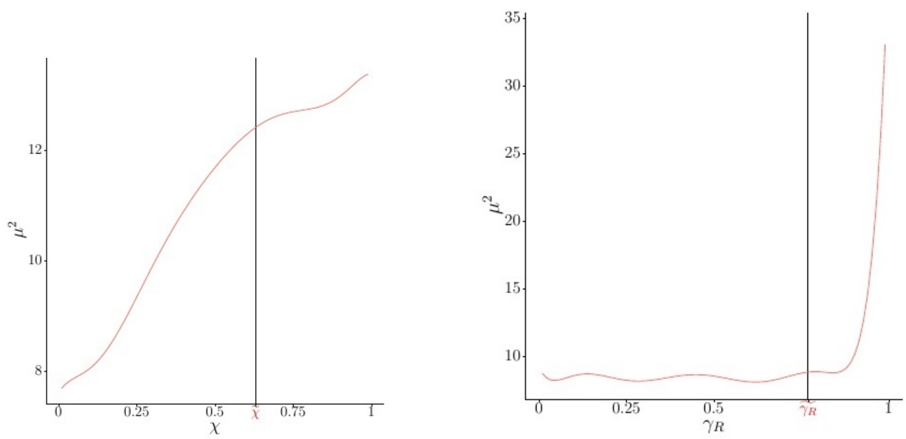
6.1
Frictions and Identification

6.1.1
Indexation, Interest Rate Smoothing and Consumption Habits

The first result is the analogue of Figure 3.1, where I alter different frictions of the model and plot them against the population concentration parameter.[1].

As inflation indexation and interest rate smoothing intensity increases, identification becomes stronger. This is the same result as that in Krogh (2015) to a more general model. Note that for the interest rate smoothing case, the

[1] Since this done via simulation, one will note slightly different values for the population concentration parameter when evaluated the calibrated values. This is due to a less than optimal number of Monte Carlo samples.



6.1(a): Inflation Indexation (χ). 6.1(b): Interest Rate Smoothing (γ_R).

Figure 6.1: Concentration Parameter μ^2 and Frictions, *Ad Hoc* Instruments, $T = 150$.

figure displays the same characteristics of Figure 3.1 where small perturbations may dramatically boost identification strength. Parameter estimates are on or near stronger identification regions; for example taking into account error estimate in $\tilde{\gamma}_R$, the empirical validity of weak identification can be fragile.

I calculate below the autocorrelation function for inflation and interest rates, which are used as instrument in the *Ad Hoc* set, with “strong" and “weak" frictions for each case. These are estimated autocorrelations, but I do so with huge sample size to ensure it is near the population value.[2]

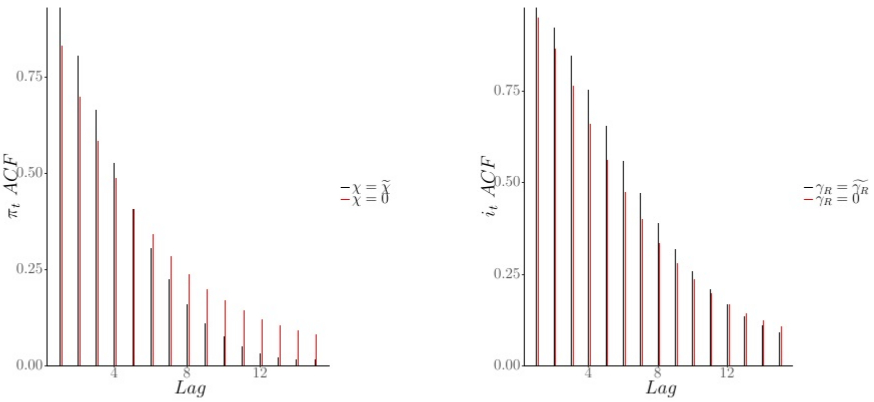


Figure 6.2: ACFs for π_t

In conformity with the intuition provided by (3-20), persistence increases for the shorter lags, which are more relevant for instrument use. Hence, as suggested by Krogh (2015), taking into account relevant macroeconomic frictions into the data generating process raises the possibility of good identification.

However, there are pitfalls with this argument in isolation. First, frictions may induce *less* persistence in other variables. This is the case with marginal cost for example, which is very commonly used as instrument:

[2]I simulate time series with $T = 100000$ and a few replications to ensure estimates have converged.

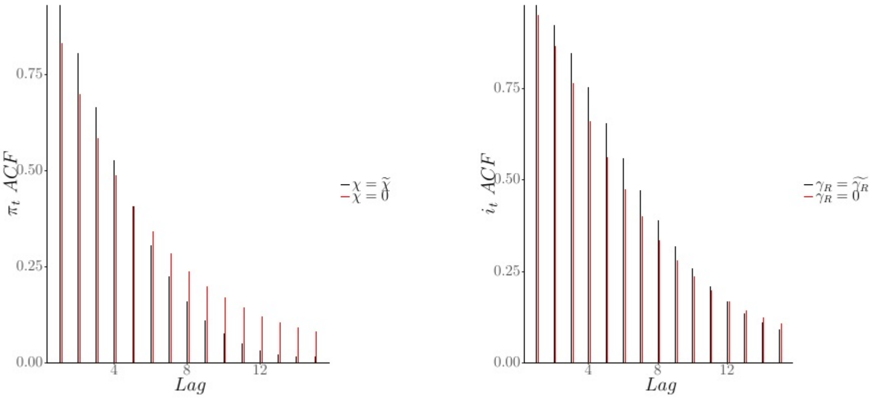


Figure 6.3: ACFs for mc_t

Moreover, in the same way one can design a data generating to mimic weak identification, one can also calibrate a $DSGE$ model to mimic strong identification. The exercise here is disciplined by a formal estimation procedure, which brings the problem to an empirical domain. Krogh (2015) uses standard calibration, but they do not come from formal likelihood methods. As an example, Krogh (2015) results do not extend for this model for the Consumption Habits friction.

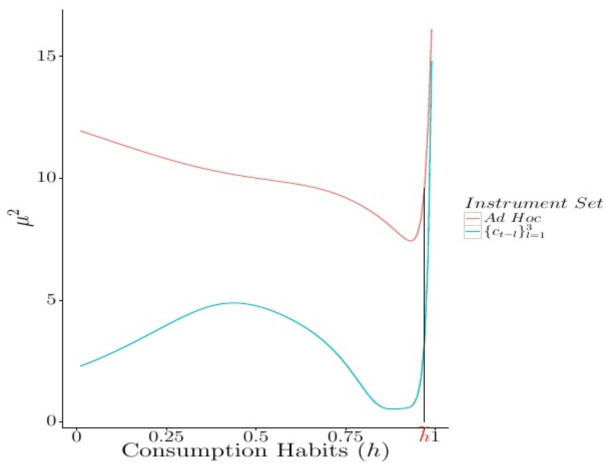


Figure 6.4: Concentration Parameter (μ^2) and Consumption Habits (h). $T = 150$.

Differently from Krogh (2015), the *Ad Hoc* instrument set leads shows a non-monotonic relationship between the consumption habits friction and identification strength, although we can still verify that it improves dramatically once habits is sufficiently strong. Moreover, changing the instrument can bring about different conclusions. At the estimated value $\tilde{h}$ the *Ad Hoc* concentration parameter more than doubles with respect to the alternative set. In this context, the fact that the estimate $\tilde{h}$ falls in the increasing portion of the concentration parameter is of interest; for the stylized example in Krogh (2015) this is irrelevant, as the relationship is everywhere increasing. Even more importantly, the nonlinear mapping of μ^2 into h may change for the whole range of parameter values. This and acknowledging uncertainty in the estimate $\tilde{h}$ underscores the importance of proper instrument selection. Differently from a more stylized example, the nonlinear relationships between the deep parameters of the model yields different conclusions regarding identification. This is in some sense a manifestation of what Hansen & Sargent (1980) called the *"hallmark of rational expectations models...[the] restrictions across parameters in agents' decision rules...are an important source of identification..."*.

6.1.2 Forward-Looking Phillips Curve vs Backward-Looking Phillips Curve

In a similar spirit, the graph below answers the following question: is a forward-looking Phillips Curve easier/harder to identify than a backward-looking Phillips Curve? This is relevant from a methodological point of view; how testable are theories about forward-looking/backward-looking behavior? It turns out that identification is strongest when both components are equally important. The graphs below represents the level curves of the concentration parameter as the parameters of interest w_f and w_b are altered.[3]

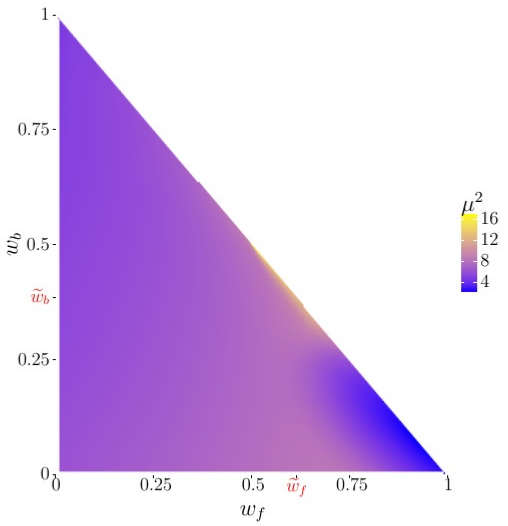


Figure 6.5: Mapping of (w_f, w_b) onto μ^2. *Ad Hoc* Instruments, $T = 150$.

The concentration parameter is decreasing as the economy becomes purely backward-looking or forward-looking, with a more dramatic case for

[3]Mechanically, I actually alter the structural parameters $\{\beta, \chi, \theta\}$ and map those values into $\{w_f, w_b, \kappa\}$ following (5-3), (5-4) and (5-5).

the latter. In a purely forward-looking economy, the intuition is best understood in the context of the Basic New-Keynesian Model. Here, the central Bank achieves the “divine coincidence"; inflation and output gap are perfectly stabilized forever and there is no variability for the econometrician to explore. Cochrane (2011) provides this intuition for the Taylor Rule case. The backward-looking case intuition may be developed in a deterministic economy. So long as inflation has not reached its steady state, the econometrician can still find variability to estimate the backward-looking process. If a deterministic process reaches its steady state, then no identification is possible.

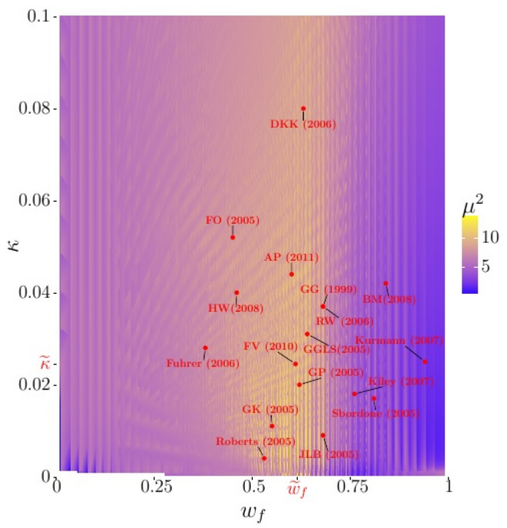
Note nonetheless that the concentration parameter value is rather low in the whole parameter space, suggesting that Mavroeidis (2005) concerns are legitimate. The counterargument is that calibrated values fall exactly on the stronger identification region, suggesting that this concern may not be empirically severe. This is, nonetheless, only one estimate, while the literature has provided other values for a plethora of different techniques. A natural follow up question, in order to acknowledge such uncertainty, is: in what regions are the estimates obtained in the literature?

6.1.3
Literature Estimates

The following graph plots some estimates obtained in the literature on the contour plot above. I now do it on the (w_f, κ) space. This is actually a more interesting and informative plot, since most literature estimates for Figure 6.5 will lie on the boundary of the space restricted by:

$$w_f + w_b = 1 \tag{6-1}$$

Some papers actually impose this restriction, but even when free estimates are allowed (as I do) estimators will lie near the boundary (structural models usually imply (6-1) or comes very close to it).



Note: The choice follows Mavroeidis et al. (2014), which use as criteria papers with more than twenty-five Google Scholar citations as of mid-September 2012. The only different is my inclusion of Fernández-Villaverde (2010). The labels are: DKK (Dufour et al. (2013)), FO (Fuhrer & Olivei (2005)), AP (Adam & Padula (2011)), HW (Henzel & Wollmershäuser (2008)), BM (Brissimis & Magginas (2008)), GG (Gali & Gertler (1999)), RW (Rudd & Whelan (2007)), Fuhrer (Fuhrer (2006)), GGLS (Gali et al. (2005)), Kurmann (Kurmann (2007)), GP (Guay & Pelgrin (2004)), Kiley (Kiley (2007)), GK (Gagnon & Khan (2005)), Sbordone (Sbordone (2005)), Roberts (M. (2005)), JLB (Jondeau & Le Bihan (2005)), FV (Fernández-Villaverde (2010)).

Figure 6.6: Mapping of (w_f, κ) onto μ^2. *Ad Hoc* Instruments, $T = 150$.

Once again, minor perturbations reveal regions of stronger identification. The extremes values for w_f are zones of very weak identification, with concentration parameters below 5. It improves substantially for the region $(w_f, \kappa) \in [0.5, 0.7] \times [0, 0.06]$ (especially when κ approaches zero), where the majority of literature estimates are placed. In this region estimates are is still somewhat informative: the analogue Cragg-Donald statistic is in the interval $(1, 2)$, indicating over 30% bias relative to *OLS* (see Stock & Yogo (2005) for critical values, weak instruments tests and boundary sets). This is not so extreme as the possibility raised by Mavroeidis (2005) where estimators would be completely uninformative. Examples of such are *BM* (2008) and *Kurmann* (2007) which are indeed in non informative, weak identification zone.

I conclude that Mavroeidis (2005) concerns are legitimate, in that a large region is weakly identified. However, maximum likelihood estimates $(\tilde{w}_f, \tilde{\kappa})$ suggest that these regions are less plausible. More importantly, this analysis has considered a fixed instrument set. As a matter of fact, the literature point estimates above have not used the instrument set I used to build the graph and as such it should be viewed as an approximation to the empirical strength of identification.[4] Still, it is exactly the fact that instruments set are distinct that I wish to explore: the *Ad Hoc* instrument set might be suboptimal. For

[4]Although the literature generally agrees on a choice similar to the *Ad Hoc* set used here.

example, Jondeau & Le Bihan (2005), which is placed in a region of stronger identification, explores a range of different specifications and variables; its place in this region might not be a coincidence. The selection methods I proposed boosts the concentration parameter.

6.2 Instrument Selection

I now turn to the proper Monte Carlo results regarding instrument selection estimators. I focus on estimates for the parameter w_f, but results are similar for w_b and κ. I *do not* want to implicitly imply that estimating w_f is special in any manner. Forward-looking equations do present a particular estimation environment, but this has nothing to do with the parameter associated with expectations component. The parameter κ, for example, is also very important in determining the sluggishness of the response of inflation and output gap to a monetary policy shock; greater κ and smaller w_f are associated with more sluggish responses. I present results for those in Appendix C. I relegate histograms to the Appendix C.1; it suffices to say that estimators possess bell-shaped distributions and does not present bi modality.

6.2.1 Finite Sample Performance

Table 6.1: w_f Estimators Descriptive Statistics, $T = 150$.

Estimator	*Mean Bias*	*Median Bias*	*MAE*	*RMSE*
adaLASSO Observables	-0.16	-0.1646	0.1843	0.2212
adaLASSO Lags	-0.119	-0.1314	0.2179	0.3921
Group LASSO	-0.1808	-0.1767	0.1858	0.2105
Ad Hoc	-0.1771	-0.1804	0.1938	0.2297
All	-0.9612	-0.1731	1.2006	12.7138
Random	-0.1159	-0.1054	0.1863	0.2596
Ridge	-0.2822	-0.1603	0.2888	2.3076
Factors	-0.2979	-0.0444	8.8302	118.14
OLS	-0.2854	-0.2854	0.2854	0.2862

Note: True value $w_f = 0.6127$.

Note that not only *OLS* underestimates the true parameter value, but so does the selection estimators, confirming that (3-7) seems to extend to this more general setting. *OLS* does not have such a bad relative performance. The best selection methods are *Group LASSO*, *adaLASSO Observables* and

Ad Hoc choice. Their performance is very similar, with a slight advantage to the former. The performance of *adaLASSO Lags* is disappointing, since it does achieve a small bias but has large mean errors, suggesting that lag selection is a more difficult task. Recall that *Group LASSO* was a compromise between observables and lag selection and as such, is a more data-driven way and possibly slightly more reliable way to impose lags in comparison to *adaLASSO Observables*.

It might seem that the advantage of selection methods over the *Ad Hoc* is not quantitatively meaningful. Although it is indeed true for this case, I would not interpret results in this fashion. First, I shall argue that for a big enough sample size (I quantify this in Section 6.3) the data-driven procedures are more precise. The more subtle and important point is that the *Ad Hoc* selection procedure is indirectly imposing model structure, while the the *LASSO*-based ones are not. The New Keynesian data generating process is of the same type as the implied economic model that would suggest the *Ad Hoc* set used here. Taking into account potential for model misspecification, *LASSO*-based have an extra edge. I provide a Bayesian interpretation of this fact and show how it relates to identification gains in Section 6.3.

Regarding the other estimators, their performances are unsatisfactory for the setup here considered. Using *All* estimators is not a wise choice; its bad performance reminds of the general principle in data analysis that using excessive information is suboptimal and extracting the bare essential is of first order. *Factors* has easily the worst performance, even though it captures more than 99% of the candidate set variation. *Ridge* is also unsatisfactory since it does not even beat *OLS*. The *Random* choice performance is surprising, suggesting that the candidate set is comprised for good instruments and estimation is good so long as a parsimonious number of instruments is chosen, although it does have an approximate 7% disadvantage relative to *Group LASSO* and *adaLASSO Observables* in the *RMSE* criteria. Moreover, randomly picking instruments is a risky strategy even if simulations show otherwise! This can be seen in the empirical application of the estimators in Section 8, where the random instruments deliver an implausible and insignificant value.

Some qualifications are in order. There is no “one-size-fits-all" estimator and a bad performance might be a sign of using it in an inappropriate environment. Carrasco & Tchuente (2015) emphasizes the use of regularization in a weak identification and many instruments environment; I have argued throughout how the use of an empirically plausible data-generating process is important for the methods I use here exactly to provide an opportunity to alleviate the weak instruments problem. In a similar fashion, Ng & Bai (2009) emphasized that $Factors$ perform well in data-rich environment, which may not be the case with the candidate set here used. A more data-rich environment would improve its relative performance since it would also make the environment more challenging for $LASSO$-based methods. The degree by which the econometrician can restrict the candidate set is ultimately a matter of how much prior economic knowledge can be imposed, which varies by applications. This naturally brings the problem into the realm of model misspecification, which is a much more challenging task and which I will not undertake here.

Robustness to Cost-Push Shock Variance

Since the simulated model used here is supposed to provide some empirical credibility to the methods here proposed, one might be concerned that the cost-push shock added in an ad hoc fashion and calibration according to Smets & Wouters (2007) might be a cause of concern. To tackle this issue, I provide the mapping of σ_c to the estimators $RMSE$ in Appendix C.2. I do this for the good performing estimators $adaLASSO\ Observables$ and $Ad\ Hoc$. With this, one can verify that that, naturally, estimators errors will alter with noise variance. Nevertheless, calibrated variance is still below the one which would minimize the $RMSE$. Moreover, the $adaLASSO\ Observables$ still seems to perform relatively well even when the noise achieves very high and implausible values.

6.2.2 Asymptotic Performance

For the large sample performance, I restrict attention to the good-performing estimators.

Table 6.2: w_f Estimators Descriptive Statistics, $T = 1000$.

$Estimator$	$Mean\ Bias$	$Median\ Bias$	MAE	$RMSE$
$adaLASSO\ Observables$	-0.0788	-0.0793	0.0934	0.1156
$Group\ LASSO$	-0.0856	-0.0831	0.095	0.128
$Ad\ Hoc$	-0.1424	-0.1448	0.1661	0.1997
OLS	-0.2846	-0.2845	0.2846	0.2848

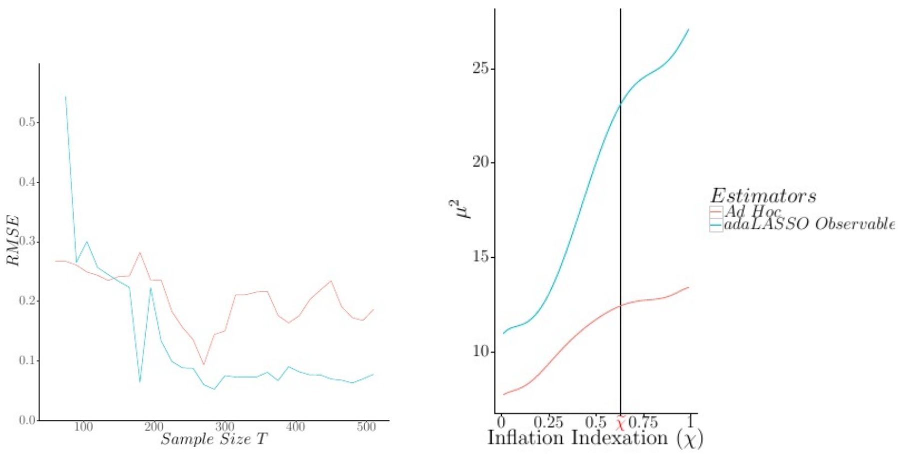
Note: True value $w_f = 0.6127$

The main message is that the data-driven selection procedure enjoys the greatest gains from the larger sample. The advantage over the $Ad\ Hoc$ procedure is now even greater. Note in (3-17) that there is a one-to-one mapping between sample size and the concentration parameter[5] so this result may also be seen as the shrinkage estimators enjoying the largest identification gains. In light of the fact that the medium scale model provides a more hospitable environment identification-wise, with small perturbations from that model being even more strongly identified (see Figure 6.1(b) and/or Figure 6.6) then there is reason to believe that data-driven methods may indeed be informative of structural parameters. Moreover, given the increase in length of macroeconomic time-series, this results provide support for the use of such methods in empirical macroeconomic estimation.

6.3 Selection, Identification and Sample Size

These two figures may be (informally) seen as two sides of the same coin, since sample size and concentration parameter have a one-to-one mapping. Figure 6.7(a) shows that for small sample size the *Ad Hoc* instruments set perform better, but as sample size increases the data driven methods overtakes. I interpret this in a Bayesian fashion: when sample size is very small, shrinkage estimators have little information to explore, while the arbitrary choice is imposing heavy prior knowledge over the data generating process. Note how bad *adaLASSO Observables* performs for the smaller sample sizes. As sample size increase, the *Ad Hoc* choice does not “learn" from the data, while shrinkage estimators manage to capture the more efficient instruments. It seems sample size should be as big $T = 100$, which would represent 25 years of data

[5]Conceptually this is obvious, since the greater the sample size the easier it should be identify parameters.



6.7(a): Sample Size and Estimators Errors.

6.7(b): Concentration Parameter μ^2 and Instrument Selection, $T = 150$.

Figure 6.7: Selection, Identification and Sample Size.

which nowadays is not a very stringent requirement for many macroeconomic time series at quarterly frequency. Asymptotically the difference does in fact diminish, although the practical content of this convergence is small: $T = 400$ represents 100 years of data, quite hard to obtain.

Figure 6.7(b) superimposes the population concentration parameter of the *adaLASSO Observables* instrument set on Figure 6.1(a). There are two main characteristics as to how the concentration parameter is boosted, regarding level and slope. The former is greater for the data-driven procedure for the whole range of values for χ, suggesting that indeed the selection procedures is making a more clever choice of instruments. What is more revealing is greater slope of the concentration parameter. Identification does improve for all instrument sets, but they do so in greater magnitude for *adaLASSO Observables*.

6.4 Selection Diagnostics

Since the selection procedure for instruments performed well, it is worth investigating which variables are being selected. In fact, one of the advantages of *LASSO*-type procedures over other data selection methods is that it allows a more transparent interpretation of the instruments being used (which is not the case for example of an ℓ^2-penalized first stage regression or principal components). The table below summarizes selection for both sample sizes for *adaLASSO Observables* and *Group LASSO*. Since both methods selects 3 lags, I restrict attention only to the observables identity of the instrument.

	% of Times Chosen			
	$T = 150$		$T = 1000$	
Observable	*adaLASSO Observables*	*Group LASSO*	*adaLASSO Observables*	*Group LASSO*
c	4.2	7.6	7.2	10.6
y^d	17.8	25.6	55.2	57.4
R	19.8	59	20	68
Π	96.2	74.6	100	100
x	10.2	24.8	9.8	9.6
w	94.8	83.6	100	100
mc	9.4	18.4	3.4	8
k	9	46.8	0.2	19.6
v^w	52.6	30.4	91.4	5.8

Table 6.3: Selection Proportions

Inflation is the only variable that coincides consistently with the *Ad Hoc* choice. Marginal cost is rarely chose. It is in fact true that the interest rate is also selected by the *Group LASSO*, but in smaller proportion. Also, the fact that *adaLASSO Observables* does not include it makes the benefits less clear cut. Both methods also choose w, which is reassuring since inflation and wage processes are generally thought to be intertwined. The main difference between the two procedures lie in the choice of R and v^w, where *Group LASSO* gives preferences for the former and *adaLASSO Observables* for the latter.

The $T = 1000$ case here is of special interest. If one were to prove that *LASSO* methods attains some type of efficiency bound, then these selections proportions could provide some type of “Oracle" benchmark for the set of instruments (not so much) restricted by (4-2). I have not, evidently, approached this to any degree, but the dominant performance of these *LASSO*-based estimators in these simulations suggests this is a viable conjecture. Under this hypothesis, inflation and wages seem to be optimal instruments.

7 Taylor Rule

I have focused on parameters of the Phillips Curve, but the ideas can be applied more generally to other important macroeconomic equilibrium conditions. In this section I apply the methods for estimation of the model's Taylor Rule. I choose this as it is another widely studied macroeconomic equation, see for example Cochrane (2011). The log-linearized Taylor Rule supplied by the model is:

$$i_t = \gamma_R i_{t-1} + (1-\gamma_R)(\gamma_\Pi \pi_t + \gamma_y \Delta \widetilde{y}_t^d) + \varepsilon_{m,t} \tag{7-1}$$

where variables are written as deviations from steady state and tildes ($\sim$) indicate normalized variables due to growth in the model.

Taylor rules is also a thoroughly used macroeconomic equation. Following the seminal work by Taylor (1993), they are good approximations of how central banks pursue interest targeting monetary policy. Note that for this economy's Central Bank there is no forward-looking component, even though one could postulate a Taylor Rule where monetary authority reacts to expected inflation instead of contemporaneous inflation. Nevertheless, variables are still potentially endogenous via classic simultaneity issues.

Interest lies in estimating:

$$\phi_R = \gamma_R \tag{7-2}$$

$$\phi_\pi = (1-\gamma_R)\gamma_\Pi \tag{7-3}$$

$$\phi_y = (1-\gamma_R)\gamma_y \tag{7-4}$$

The table below summarizes estimators errors for estimation of ϕ_π. I restrict attention to this parameter and the finite sample size. The *Ad Hoc* instrument set here is composed of:

$$Ad\ Hoc \rightarrow \{\pi_{t-l}, i_{t-l}, \Delta \widetilde{y}_{t-l}^d\}_{l=1}^3 \tag{7-5}$$

Table 7.1: ϕ_π Estimators Descriptive Statistics, $T = 150$.

Estimator	*Mean Bias*	*Median Bias*	*MAE*	*RMSE*
adaLASSO Observables	-0.034	-0.0342	0.0345	0.0365
adaLASSO Lags	-0.0346	-0.0342	0.0352	0.0374
Group LASSO	-0.0371	-0.0342	0.0353	0.0671
Ad Hoc	-0.0332	-0.0335	0.0332	0.0345
All	-0.0333	-0.035	0.0942	0.6123
Random	-0.0336	-0.0336	0.0336	0.0350
Ridge	-0.033	-0.0334	0.033	0.0339
Factors	-0.0399	-0.0382	0.0604	0.1127
OLS	-0.0338	-0.0339	0.0338	0.0343

Note: True parameter value $\phi_\pi = 0.2967$.

Note that there is little difference between *adaLASSO Observables* and the *Ad Hoc* procedure. *Group LASSO* has a worst performance. The more revealing point is the performance of *OLS* which is essentially indistinguishable from those using instruments. What occurs here is that the degree of endogeneity, empirically determined by *DSGE* estimated value of $\varepsilon_{m,t}$ variances is small vis-à-vis the extra variance induced by the use of instruments. In this case *OLS*, although inconsistent, still delivers a competitive approximation to the true parameter value for finite samples. This parallels the analytical case for the Basic New Keynesian Model Phillips Curve in (3-7); the analogous σ_η^2 object is small. Instrumental variables and in particular, the selection procedures, would be more competitive were the Taylor Rule composed of an expectational component, as in the Phillips Curve, inflating the degree of endogeneity by the extra expectational error variance term. In the limiting case, were the equation deterministic, *OLS* would be able to recover the parameters *exactly* for any sample size. This intuition is developed in Carvalho & Nechio (2014), where they determine how tolerable are *OLS* estimates errors as the monetary policy shock variance increases. In summary, the Taylor Rule provides a case where not only instrument selection may irrelevant, but an instrumental variables strategy in general may be useless, not due to a weak instruments problem, but due to the low degree of endogeneity empirically measured.

8
Empirical Application

I now apply the instrument selection procedures to empirically estimate a New Keynesian Phillips Curve specification for the US economy, in the spirit of Gali & Gertler (1999). I do this with two datasets. First I use a sample which parallels that of Gali & Gertler (1999) in order to allow for a proper comparison with this seminal literature estimate. I also use data from 1980 onwards until 2014, which updates the sample relative to that in Gali & Gertler (1999) and is also parallels the simulation done here: it amounts to approximately the same sample size as that used for the benchmark simulation in Section 6.2 and uses analogous variables. The datasets used and treatment is described in Appendix D.1. Once again I focus on w_f for concreteness, but provide the other estimates in Appendix D.2.2.

8.1
Gali and Gertler (1999) Analogous Sample

Table 8.1: Empirical Estimation of w_f for the US Economy

$Estimator$	$EstimatedCoefficient$	$StandardError$
adaLASSO Observables	0.84	0.12
adaLASSO Lags	0.67	0.06
Group LASSO	0.60	0.14
Ad Hoc	0.84	0.12
All	0.57	0.04
Random	0.52	0.14
Ridge	0.42	0.11
Factors	0.53	1.08
OLS	0.50	0.06

Note: Optimal weighting matrix used. Covariance matrix for GMM calculated using HAC estimator matrix with *Bartlett* kernel and bandwidth set according to Andrews (1991). OLS standard error obtained assuming homoskedasticity.

Note that I proceed with classical inference, calculating the usual standard error without proving that the estimators (mainly the $LASSO$-based ones) actually converge in distribution accordingly. I do this based solely on the conjecture that inference will proceed as such. Belloni et al. (2010) proves in a similar $TSLS$ framework that such approach is valid.

The *All* is the instrument set that most closely resembles that used by Gali & Gertler (1999) and it does in fact achieve a similar value; the analogous estimated value in the paper is 0.59. Using *adaLASSO Observables* and

Ad Hoc, two instrument set which performed well in the simulations, reveals identical point estimates (0.84) showing that the forward-looking component might have been even greater. *Group LASSO* divergent point estimate may pose doubts on such a high number, but it still delivers a larger forward-looking component than that in Gali & Gertler (1999).

8.2 Updated Sample

Table 8.2: Empirical Estimation of w_f for the US Economy

Estimator	*Estimated Coefficient*	*Standard Error*
adaLASSO Observables	0.63	0.07
adaLASSO Lags	0.65	0.08
Group LASSO	0.61	0.13
Ad Hoc	0.68	0.13
All	0.72	0.11
Random	0.65	0.20
Ridge	0.08	0.28
Factors	-0.30	1.02
OLS	0.39	0.08

Note: Optimal weighting matrix used. Covariance matrix for *GMM* calculated using *HAC* estimator matrix with *Bartlett* kernel and bandwidth set according to Andrews (1991). *OLS* standard error obtained assuming homoskedasticity.

Using the updated sample (it ends in 2014; about 40% is different) provides a more general agreement of values with respect to the *adaLASSO Observables*, *Ad Hoc* and *Group LASSO*. Moreover this agreement comes with the reduction of the former two estimators point estimates, suggesting that maybe w_f has reduced recently. This is a superficial conjecture, since the specification here used does not contemplate time varying parameters. It is nonetheless an interesting avenue to explore with instrument selection; structural breaks may limit available sample size.

The bad performing estimators, *Factors*, *Ridge* and *Random* show economically implausible and non-significant point estimates.

8.3 Selected Instruments

I now compare the selected instruments once again restricting attention to the observables identity. I do not analyze the *adaLASSO Lags* since its performance throughout has been dubious. Also, recalling that selection is made on a regularized reduced form for each possible endogenous variable, π_{t+1} and mc_t, I concentrate on the former. Once again, I only indicate the observables identity, since three lags are imposed for the estimators.

8.3.1 Gali and Gertler (1999) Analogous Sample

Estimator	y	c	x	l	π	r	mc	G	w	*Spread*	*comm*
adaLASSO Observables					■						
Group LASSO		■			■						■
Ad Hoc					■	■	■				

Table 8.3: Instruments for π_{t+1}. Black box indicates a selected variable.

The number of variables selected is much more parsimonious than that in Gali & Gertler (1999). *adaLASSO Observables* selects only inflation as observable, which is even more conservative relative to other literature attempts. *Group LASSO* selects an alternative set of instruments, which reveals why its point estimate is so divergent and therefore, must be taken with caution. It also indicates once again how lag selection is a harder task.

8.3.2 Updated Sample

Estimator	y	c	x	u	l	Π	r	mc	w	G
adaLASSO Observables						■	■			
Group LASSO	■		■	■	■	■	■	■		
Ad Hoc						■	■	■		

Table 8.4: Instruments for π_{t+1}. Black box indicates a selected variable.

Once again, *adaLASSO Observables* is the most conservative in the number of instruments, selecting two of them. *Group LASSO* selection encompasses both the *Ad Hoc* and *adaLASSO Observables*, selecting this time an instrument set similar in size to that used by Gali & Gertler (1999). This once suggests difficulty in lag selection. Still, the fact that point estimates agree for this case is rather reassuring, suggesting that indeed the forward-looking component may be dominant as suggested by Gali & Gertler (1999). Differently from the simulations case, no selection procedure chose wages w and it seems that data-driven procedure rely instead on the usual *Ad Hoc* instruments, with the exception of mc for *adaLASSO Observables*.

9
Conclusion

The choice of instruments in estimation of structural relationship is often made arbitrarily and relegated to secondary considerations in an empirical exercise. I have argued that proceeding this way imposes model structure, albeit indirectly. Not only this prior economic knowledge may be wrong, but even what may seem as intuitive instruments may actually not be the optimal choice. This becomes even more important in possibly weak instruments environment. I have applied these ideas in the context of an estimated $DSGE$ model, aiming to estimate its structural relationship, mainly the New Keynesian Phillips Curve. I show how a more realistic data generating process may corroborate identification from a limited-information setting, which partially reverts Mavroeidis (2005) where estimation would be completely uninformative. Slight perturbations of this data generating process may yield even stronger identification, provided instruments are properly selected. More importantly, I have applied data driven methods to select instruments and have shown that $LASSO$-based ones are well suited for the task, obtaining tolerable errors in finite sample, boosting the concentration parameter and enjoying the greatest identification gains. An empirical application of the estimators have shown that Gali & Gertler (1999) dominant forward-looking estimate is validated by these selection procedures.

There are a number of interesting avenues by which this work may be extended. The first and probably most important is development of a formal theory of the estimators here proposed. The simulation results provide enough basis to conjecture a sound econometric theory behind, but not less challenging. Belloni et al. (2010) develops an analog landmark theoretical framework for the $LASSO$ in a linear IV setting.

Also of interest would be approach the problem of model misspecification closer. I have argued heuristically and only sketched the surface between this topic and the ideas here developed. For example, one could link this to the issue of weak identification by *marginally* adding model structure. How would identification improve when estimating the New Keynesian Phillips Curve and the Taylor Rule jointly? Note that the limit estimator which uses all equilibrium conditions attains the efficient maximum likelihood estimator. The problem of moment selection arises naturally in this setting. How to select these equilibrium conditions?

The selection procedures may be extended in a few directions. I have highlighted throughout the challenges regarding lag selection. Finding a more reliable way to do data driven lag selection would be desirable. Also, augmenting the candidate set may improve identification even more, either by increasing the number of observables or using nonlinear transformations, as in Belloni et al. (2010) and Carrasco & Tchuente (2015). What would happen if candidate set is larger than the sample size? In this setting, $LASSO$-based estimators are even more desirable. Also, the setting here considers linear moment conditions. For the nonlinear case, optimal instruments potentially depends on true parameter values. Variable selection is still feasible, possibly through iterative procedures over the parameter space. Nonlinear moment conditions would also estimation of the so-called deep parameters of the model (for example, the discount factor β).

Regarding the data-generating process, one could try and implement these ideas for even more realistic models, as those used for policy making in Central Banks. Cwik et al. (2012) provides a unified framework of models. It is a brute force way of corroborating the efficiency of $LASSO$-based estimators, but may still be relevant from a practitioner's point of view. Moreover, it may provide further insight between the role of frictions and identification. For example, I have mentioned that minor perturbations to the model here considered could dramatically improve identification. Mapping these standard errors into concentration parameter confidence intervals could be such a way to summarize this information. To end, recent developments have allowed $DSGE$ models to be solve in nonlinear fashion, which is a potentially more accurate description of an economy's data generating process. Moreover, macroeconomic time-series frequently suffer structural breaks, which may impose restrictions on usable sample size. Investigating consequences of these nonlinearities into identification and the selection estimators would also be interesting.

Bibliography

J. Li and W. Chen. **Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models**. *International Journal of Forecasting*, 30(4):996 – 1015, 2014.

T. Cwik, G. Mueller, S. Schmidt, V. Wieland and M. H. Wolters. **A new comparative approach to macroeconomic modeling and policy analysis**. *Journal of Economic Behavior & Organization*, 83(3):523–541, 2012.

J. Fernández-Villaverde, J. Rubio-Ramirez and T. J. Sargent. **A, b, cs (and d)s for understanding vars**, 2005.

K. Adam and M. Padula. **Inflation dynamics and subjective expectations in the united states**. *Economic Inquiry*, 49(1):13–25, 2011.

A. Hall. **Generalized method of moments**. Oxford University Press, 2005.

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

T. W. Anderson and H. Rubin. **Estimation of the parameters of a single equation in a complete system of stochastic equations**. *Ann. Math. Statist.*, 20(1):46–63, 03 1949.

D. W. K. Andrews. **Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation**. *Econometrica*, 59(3):817–58, May 1991.

J. Bound, D. A. Jaeger and R. M. Baker. **Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak**. *Journal of the American Statistical Association*, 90(430):443–450, 1995.

A. Belloni, D. Chen, V. Chernozhukov and C. Hansen. **Sparse models and methods for optimal instruments with an application to eminent domain**. *Econometrica 2012 (ArXiv 2010)*, 80(6):2369–2429, 2010.

S. N. Brissimis and N. S. Magginas. **Inflation Forecasts and the New Keynesian Phillips Curve**. *International Journal of Central Banking*, 4(2):1–22, June 2008.

G. A. Calvo. **Staggered prices in a utility-maximizing framework**. *Journal of Monetary Economics*, 12(3):383–398, September 1983.

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

C. Carvalho and F. Nechio. **Do people understand monetary policy?** *Journal of Monetary Economics*, 66(C):108–123, 2014.

M. Carrasco and G. Tchuente. **Regularized liml for many instruments**. *Journal of Econometrics*, 186(2):427–442, 2015.

M. Carrasco and G. Tchuente. **Efficient estimation with many weak instruments using regularization techniques**. *Econometric Reviews*, 2015.

X. Cheng and Z. Liao. **Select the valid and relevant moments: An information-based lasso for gmm with many moments**. *Journal of Econometrics*, 186(2):443–464, 2015.

L. J. Christiano. **Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients**. *Computational Economics*, 20(1-2):21–55, October 2002.

J. Cochrane. **Asset pricing: (revised edition)**. Princeton University Press, 2009.

J. H. Cochrane. **Determinacy and Identification with Taylor Rules**. *Journal of Political Economy*, 119(3):565 – 615, 2011.

J.-M. Dufour, L. Khalaf and M. Kichian. **Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis**. *Journal of Economic Dynamics and Control*, 30(9-10):1707–1727, 2006.

J.-M. Dufour, L. Khalaf and M. Kichian. **Identification-robust analysis of DSGE and structural macroeconomic models**. *Journal of Monetary Economics*, 60(3):340–350, 2013.

S. Adjemian, H. Bastani, M. Juillard, F. Karamé, F. Mihoubi, G. Perendia, J. Pfeifer, M. Ratto and S. Villemot. **Dynare: Reference manual, version 4**. *Dynare Working Papers*, 1, 2011.

J. C. Fuhrer and G. P. Olivei. **Estimating forward-looking Euler equations with GMM estimators: an optimal-instruments approach**. *Proceedings*, pages 87–114, 2005.

J. C. Fuhrer. **Intrinsic and Inherited Inflation Persistence**. *International Journal of Central Banking*, 2(3), September 2006.

E. Gagnon and H. Khan. **New phillips curve under alternative production technologies for canada, the united states, and the euro area**. *European Economic Review*, 49(6):1571 – 1602, 2005.

J. Gali, M. Gertler and J. David Lopez-Salido. **Robustness of the estimates of the hybrid New Keynesian Phillips curve**. *Journal of Monetary Economics*, 52(6):1107–1118, September 2005.

J. Galí. **Monetary policy, inflation, and the business cycle: An introduction to the new keynesian framework**. Princeton University Press, 2008.

J. Gali and M. Gertler. **Inflation dynamics: A structural econometric analysis**. *Journal of Monetary Economics*, 44(2):195–222, October 1999.

J. H. Stock and J. Wright. **GMM with Weak Identification**. *Econometrica*, 68(5):1055–1096, September 2000.

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

A. Guay and F. Pelgrin. **The U.S. New Keynesian Phillips Curve: An Empirical Assessment**. Staff Working Papers 04-35, Bank of Canada, 2004.

S. Dees, M. H. Pesaran, L. V. Smith and R. P. Smith. **Identification of New Keynesian Phillips Curves from a Global Perspective**. *Journal of Money, Credit and Banking*, 41(7):1481–1502, October 2009.

L. P. Hansen and T. J. Sargent. **Formulating and estimating dynamic linear rational expectations models**. *Journal of Economic Dynamics and Control*, 2(1):7–46, May 1980.

L. P. Hansen and K. J. Singleton. **Generalized instrumental variables estimation of nonlinear rational expectations models**. *Econometrica*, 50(5):1269–1286, 1982.

L. P. Hansen, J. Heaton and A. Yaron. **Finite-sample properties of some alternative gmm estimators**. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

C. Hansen and D. Kozbur. **Instrumental variables estimation with many weak instruments using regularized {JIVE}**. *Journal of Econometrics*, 182(2):290 – 308, 2014.

S. Henzel and T. Wollmershäuser. **The new keynesian phillips curve and the role of expectations: evidence from the cesifo world economic survey**. *Economic Modelling*, 25(5):811–832, 2008.

R. J. Hodrick and E. C. Prescott. **Postwar u.s. business cycles: An empirical investigation**. *Journal of Money, Credit and Banking*, 29(1):pp. 1–16, 1997.

E. Jondeau and H. Le Bihan. **Testing for the new keynesian phillips curve. additional international evidence**. *Economic Modelling*, 22(3):521–550, 2005.

M. T. Kiley. **A Quantitative Comparison of Sticky-Price and Sticky-Information Models of Price Setting**. *Journal of Money, Credit and Banking*, 39(s1):101–125, 02 2007.

F. Kleibergen and S. Mavroeidis. **Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve**. *Journal of Business & Economic Statistics*, 27(3):293–311, 2009.

T. S. Krogh. **Macro frictions and theoretical identification of the New Keynesian Phillips curve**. *Journal of Macroeconomics*, 43(C):191–204, 2015.

A. Kurmann. **VAR-based estimation of Euler equations with an application to New Keynesian pricing**. *Journal of Economic Dynamics and Control*, 31(3):767–796, March 2007.

F. E. Kydland and E. C. Prescott. **Time to Build and Aggregate Fluctuations**. *Econometrica*, 50(6):1345–70, November 1982.

J. Lindé. **Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach**. *Journal of Monetary Economics*, 52(6):1135–1149, September 2005.

L. M. Magnusson and S. Mavroeidis. **Identification using stability restrictions**. *Econometrica, forthcoming*, 2013.

S. Mavroeidis. **Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve**. *Journal of Money, Credit and Banking*, 37(3):421–48, June 2005.

S. Mavroeidis, M. Plagborg-Møller and J. H. Stock. **Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve**. *Journal of Economic Literature*, 52(1):124–88, March 2014.

M. C. Medeiros and E. F. Mendes. **ℓ_1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors**. *Journal of Econometrics*, 2015.

M. C. Medeiros and G. F. Vasconcelos. **Forecasting macroeconomic variables in data-rich environments**. *Economics Letters*, 138:50 – 52, 2016.

I. Andrews and A. Mikusheva. **Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models**. *Quantitative Economics*, 6(1):123–152, 2015.

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

J. M. Roberts. **New keynesian economics and the phillips curve**. *Journal of Money, Credit and Banking*, 27(4):975–984, 1995.

R. J. M. **How Well Does the New Keynesian Sticky-Price Model Fit the Data?** *The B.E. Journal of Macroeconomics*, 5(1):1–39, September 2005.

J. Rudd and K. Whelan. **Modeling inflation dynamics: A critical review of recent research**. *Journal of Money, Credit and Banking*, 39:155–170, 2007.

A. M. Sbordone. **Do expected future marginal costs drive inflation dynamics?** *Journal of Monetary Economics*, 52(6):1183–1197, September 2005.

S. Ng and J. Bai. **Selecting Instrumental Variables in a Data Rich Environment**. *Journal of Time Series Econometrics*, 1(1):1–34, April 2009.

E. M. Leeper and C. A. Sims. **Toward a Modern Macroeconomic Model Usable for Policy Analysis**. In *NBER Macroeconomics Annual 1994, Volume 9, NBER Chapters*, pages 81–140. National Bureau of Economic Research, Inc, 1994.

D. Staiger and J. H. Stock. **Instrumental Variables Regression with Weak Instruments**. *Econometrica*, 65(3):557–586, May 1997.

J. H. Stock and M. Yogo. **Testing for weak instruments in linear iv regression**. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, 2005.

J. H. Stock and M. W. Watson. **Why Has U.S. Inflation Become Harder to Forecast?** *Journal of Money, Credit and Banking*, 39(s1):3–33, 02 2007.

F. Smets and R. Wouters. **Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach**. *American Economic Review*, 97(3):586–606, June 2007.

J. B. Taylor. **Discretion versus policy rules in practice**. *Carnegie-Rochester conference series on public policy*, 39:195–214, 1993.

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

J. Fernández-Villaverde. **The econometrics of dsge models**. *SERIEs*, 1(1-2):3–49, 2010.

J. Wooldridge. **Econometric analysis of cross section and panel data**. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2002.

M. Yuan and Y. Lin. **Model selection and estimation in regression with grouped variables**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

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

A
Appendix for Chapter 3

A.1
Proof of Proposition 3.1 - Attenuation Bias in the Basic New Keynesian Model

Proof. The environment here is analogous to that of the Classic Errors in Variable (Wooldridge (2002)). By definition:

$$Cov(\eta_{t+1}, E_t[\pi_{t+1}]) = 0 \tag{A-1}$$

The measurement error η_{t+1} is uncorrelated with the unobserved measure and therefore, *must* be correlated with its observed measure π_{t+1}.

$$\begin{aligned} Cov(\eta_{t+1}, \pi_{t+1}) &= Cov(\eta_{t+1}, E_t[\pi_{t+1}] + \eta_{t+1}) && \text{(A-2)} \\ &= Cov(\eta_{t+1}, \eta_{t+1}) && \text{(A-3)} \\ &= \sigma^2_\eta \neq 0 && \text{(A-4)} \end{aligned}$$

Now, the OLS estimator $\hat{\beta}_{OLS}$ may be obtained with partitioned projections. For the observed (o) and unobserved (u) measures, we project on the other variables of interest:

$$\begin{aligned} \pi_{t+1} &= \gamma_{o0} + \gamma_{o1} y_t + r_{t+1} && \text{(A-5)} \\ E_t[\pi_{t+1}] &= \gamma_{u0} + \gamma_{u1} y_t + r^*_{t+1} && \text{(A-6)} \\ \implies r_{t+1} &= r^*_{t+1} + \eta_{t+1} && \text{(A-7)} \end{aligned}$$

By the law of iterated projections (Property LP.7 from Wooldridge (2002)):

$$L(\pi_t | r_{t+1}) = \beta r^*_{t+1} \tag{A-8}$$

where $L(.|.)$ is the linear projection operator.

The OLS estimator is:

$$\hat{\beta}_{OLS} = \left(\sum_{t=1}^{T} r_{t+1} r_{t+1}\right)^{-1} \left(\sum_{t=1}^{T} r_{t+1}\pi_t\right) \tag{A-9}$$

$$= \left(\sum_{t=1}^{T} r_{t+1} r_{t+1}\right)^{-1} \left(\sum_{t=1}^{T} (r^*_{t+1} + \eta_{t+1})\pi_t\right) \tag{A-10}$$

$$= \left(\frac{1}{T}\sum_{t=1}^{T} r_{t+1} r_{t+1}\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} (r^*_{t+1} + \eta_{t+1})\left(\beta E_t[\pi_{t+1}] + \kappa \tilde{y}_t\right)\right) \tag{A-11}$$

Hence by a Law of Large Numbers:

$$\hat{\beta}_{OLS} \xrightarrow{p} E[r_{t+1}r_{t+1}]^{-1}\left(\beta E[r^*_{t+1}E_t[\pi_{t+1}]] + \kappa E[r^*_{t+1}\tilde{y}_t] + \beta E[\eta_{t+1}E_t[\pi_{t+1}]] + \kappa E[\eta_{t+1}\tilde{y}_t]\right) \tag{A-12}$$

Now $E(r^*_{t+1}y_t) = 0$ following (A-6) and $E(r^*_{t+1}E_t[\pi_{t+1}]) = \sigma^2_{r^*}$. Also, $E[\eta_{t+1}E_t[\pi_{t+1}]] = E(\eta_{t+1}y_t) = 0$ by definition of a rational expectations error.

$$plim\ \hat{\beta}_{OLS} = \beta\frac{\sigma_{r^*}}{\sigma^2_r} \tag{A-13}$$

Since $E(E_t[\pi_{t+1}]\eta_{t+1}) = 0$, $\sigma^2_r = \sigma^2_{r^*} + \sigma^2_\eta$ and we obtain the desired result:

$$\hat{\beta}_{OLS} \xrightarrow{p} \beta\left(\frac{\sigma^2_{r^*}}{\sigma^{2*}_r + \sigma^2_\eta}\right) < \beta. \tag{A-14}$$

■

A.2
Undetermined Coefficients Solution of the Basic New Keynesian Model

A.2.1
$i.i.d.$ Shocks

There are two solutions to this guess. One of them imply implausible or unwanted parameter combinations. For example, a calvo parameter θ that would imply prices are sticky forever and $\pi_t = 0, \forall t$. This would render the estimation exercise infeasible.

More precisely, this solution is:

$$\left\{\psi_\pi = 0,\, \psi_y = \frac{1}{\sigma + \phi_y}, \kappa = 0\right\} \tag{A-15}$$

where we remind that κ is a function of structural parameters:

$$\kappa = \frac{(1-\theta)(1-\beta\theta)}{\theta}\frac{1-\alpha}{1-\alpha+\alpha\varepsilon} \tag{A-16}$$

$\kappa = 0$ could imply $\theta = 0$ as stated in the text. Another possibility is that labor does not enter the production function, in which case $\alpha = 1$.

Hence, I focus on the more "conventional" solution, which is used and estimated in the literature, since no additional restriction is placed on the deep parameters of the model.

$$\left\{\psi_\pi = \frac{\kappa}{\kappa\phi_\pi + \sigma + \phi_y}, \psi_y = -\frac{1}{-\kappa\phi_\pi - \sigma - \phi_y}\right\} \tag{A-17}$$

A.2.2 Persistent Shocks

Once again, unwanted solutions imply the following restrictions:

$$\left\{\kappa = 0, \sigma = \frac{1}{\psi_y(1-\rho)}, \psi_\pi = 0\right\} \tag{A-18}$$

$$\left\{\phi_\pi = \frac{1 - \phi_y\psi_\pi + \psi_\pi\rho}{\psi_\pi}, \beta = \frac{1}{\rho}, \psi_y = 0\right\} \tag{A-19}$$

The solution used is given by:

$$\left\{\psi_\pi = \frac{\kappa}{(\rho-1)\sigma(\beta\rho-1) + \kappa(\phi_\pi + \phi_y - \rho)},\ \psi_y = \frac{1-\beta\rho}{(\rho-1)\sigma(\beta\rho-1) + \kappa(\phi_\pi + \phi_y - \rho)}\right\} \tag{A-20}$$

A.3 Concentration Parameter in the Basic New Keynesian Model

Proof.

Following the notation in the text:

$$\boldsymbol{z}_{t-1} = \pi_{t-1} \tag{A-21}$$

$$\boldsymbol{\nu}_t = \rho\psi_\pi u_t + \psi_\pi u_{t+1} \tag{A-22}$$

Then:

$$\mathbb{E}[\boldsymbol{z}_{t|\boldsymbol{x}}] = \mathbb{E}[\pi_{t-1}^2] = \mathbb{E}[\pi_t^2] = \psi_\pi^2\mathbb{E}[\varepsilon_t^2] = \psi_\pi^2\sigma_\varepsilon^2 \tag{A-23}$$

where $\sigma_\varepsilon^2 = \frac{\sigma_u^2}{1-\rho^2}$. Now:

$$\begin{aligned}\mathbf{\Sigma}_{VV} &= \mathbb{E}[(\rho\psi_\pi u_t + \psi_\pi u_{t+1})^2] \\ &= \psi_\pi^2(1+\rho^2)\sigma_u^2\end{aligned} \tag{A-24}$$

$$\implies \mathbf{\Sigma}_{VV}^{-\frac{1}{2}} = \frac{1}{\psi_\pi\sqrt{(1+\rho^2)\sigma_u^2}} \tag{A-25}$$

Hence:

$$\mu^2 = T\left(\frac{1}{\psi_\pi\sqrt{(1+\rho^2)\sigma_u^2}}\right)\left(\rho^2\right)\psi_\pi^2\sigma_\varepsilon^2\left(\rho^2\right)\left(\frac{1}{\psi_\pi\sqrt{(1+\rho^2)\sigma_u^2}}\right) \tag{A-26}$$

$$= T\left(\frac{1}{\psi_\pi\sqrt{(1+\rho^2)\sigma_u^2}}\right)\left(\rho^2\right)\psi_\pi^2\frac{\sigma_u^2}{1-\rho^2}\left(\rho^2\right)\left(\frac{1}{\psi_\pi\sqrt{(1+\rho^2)\sigma_u^2}}\right) \tag{A-27}$$

$$= T\left(\frac{1}{\psi_\pi^2(1+\rho^2)\sigma_u^2}\right)\left(\rho^2\right)\psi_\pi^2\frac{\sigma_u^2}{1-\rho^2}\left(\rho^2\right) \tag{A-28}$$

$$= T\left(\frac{1}{(1+\rho^2)}\right)\left(\rho^2\right)\frac{1}{1-\rho^2}\left(\rho^2\right) \tag{A-29}$$

$$= T\frac{\rho^4}{1-\rho^4} \tag{A-30}$$

It is increasing for $\rho \in (0,1)$:

$$\frac{\partial\mu^2}{\partial\rho} = T\frac{4\rho^3}{(1-\rho^4)^2} > 0. \tag{A-31}$$

■

A.4 Instruments for Future Inflation: General Exponent

If $\pi_{t+\tau}$ was the endogenous variable, then:

$$\mu^2 = T\frac{\rho^{2(\tau+1)}}{1-\rho^{2(\tau+1)}} \tag{A-32}$$

Then identification strength decreases with horizon τ with $\rho \in (0,1)$:

$$\frac{\partial\mu^2}{\partial\tau} = \frac{2\rho^{2\tau+2}\log(\rho)}{1-(\rho^{2\tau+2})^2} < 0 \tag{A-33}$$

B
Appendix for Chapter 5

The model is essentially that of Fernández-Villaverde (2010) with an added cost-push shock. For simplicity, I alter the inflation steady state to zero inflation, whereas in the original model it is set to 1%.

B.1
Model

B.1.1
Consumers

A continuum of consumers indexed by j maximize stream of period utility, choosing consumption c_{jt}, labor l_{jt}, real money demand $\frac{m_{jt}}{p_t}$, bonds b_{jt+1}, state contingent claims a_{jt+1}, intensity of use of capital u_{jt} and investment x_{jt} subject to Calvo (1983) wage rigidity.

$$U = \max_{\{c_{jt},b_{jt},u_{jt},k_{jt},x_{jt},w_{jt},l_{jt},a_{jt+1}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t d_t \left\{ \log(c_{jt} - hc_{jt-1}) + \upsilon \log\left(\frac{m_{jt}}{p_t}\right) - \varphi_t \psi \frac{l_{jt}^{1+\gamma}}{1+\gamma} \right\} \tag{B-1}$$

$$\begin{aligned} \text{s.t. } & c_{jt} + x_{jt} + \frac{m_{jt}}{p_t} + \frac{b_{jt+1}}{p_t} + \int q_{jt+1,t} a_{jt+1} d\omega_{j,t+1,t} \\ & = w_{jt} l_{jt} + \left(r_t u_{jt} - \mu_t^{-1} a[u_{jt}]\right) k_{jt-1} + \frac{m_{jt-1}}{p_t} + R_{t-1}\frac{b_{jt}}{p_t} + a_{jt} + T_t + F_t \\ & k_{jt} = (1-\delta)k_{jt-1} + \mu_t \left(1 - S\left[\frac{x_{jt}}{x_{jt-1}}\right]\right) x_{jt} \end{aligned} \tag{B-2}$$

where:

$$a[u] = \gamma_1(u-1) + \frac{\gamma_2}{2}(u-1)^2 \tag{B-3}$$

$$S\left[\frac{x_t}{x_{t-1}}\right] = \frac{\kappa}{2}\left(\frac{x_t}{x_{t-1}} - \Lambda_x\right)^2 \tag{B-4}$$

B.1.2

Firms

Firms choose prices subject to Calvo (1983) lottery with non-optimization probability θ_p. Non-optimized prices are indexed with intensity χ.

$$\max_{p_{it}} \mathbb{E}_t \sum_{t=0}^{\infty} (\beta\theta_p)^{\tau} \frac{\lambda_{t+\tau}}{\lambda_t} \left\{ \left(\prod_{s=1}^{\tau} \Pi_{t+s-1}^{\chi} \frac{p_{it}}{p_{t+\tau}} - \exp(\varepsilon_{t+\tau}) mc_{t+\tau} \right) y_{it+\tau} \right\} \tag{B-5}$$

$$\text{s.t.} \quad \begin{aligned} y_{it+\tau} &= \left(\prod_{s=1}^{\tau} \Pi_{t+s-1}^{\chi} \frac{p_{it}}{p_{t+\tau}} \right)^{-\varepsilon} y_{t+\tau}^{d} \\ y_{it} &= A_t k_{it-1}^{\alpha} \left(l_{it}^{d} \right)^{1-\alpha} - \phi z_t \end{aligned} \tag{B-6}$$

B.1.3 Government

The fiscal side of the government is passive, running a balanced budget every period, using transfers T_t that finances monetary policy.

$$T_t = \frac{\int_0^1 m_{jt} dj}{p_t} - \frac{\int_0^1 m_{jt-1} dj}{p_t} + \frac{\int_0^1 b_{jt+1} dj}{p_t} - R_{t-1} \frac{\int_0^1 b_{jt} dj}{p_t} \tag{B-7}$$

Monetary policy is determined by a Taylor-type rule.

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R} \right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi} \right)^{\gamma_\Pi} \left(\frac{\frac{y_t^d}{y_{t-1}^d}}{\Gamma_{y^d}} \right)^{\gamma_y} \right)^{1-\gamma_R} \exp(m_t) \tag{B-8}$$

B.1.4 Shocks

Shocks driving the model are preference shocks (d_t), labor supply shock (φ_t), a cost-push shock (ε_t), monetary policy shock (m_t), growth rate of investment-specific technology (μ_t) and growth rate of neutral technology (A_t).

$$\log d_t = \rho_d \log d_{t-1} + \sigma_d \varepsilon_{d,t} \tag{B-9}$$

$$\log \varphi_t = \rho_\varphi \log \varphi_{t-1} + \sigma_\varphi \varepsilon_{\varphi,t} \tag{B-10}$$

$$\log \varepsilon_t = \rho_c \log \varepsilon_{t-1} + \sigma_c \varepsilon_{c,t} \tag{B-11}$$

$$\log m_t = \rho_m \log m_{t-1} + \sigma_m \varepsilon_{m,t} \tag{B-12}$$

$$\mu_t = \mu_{t-1} \exp(\Lambda_\mu + \sigma_\mu \varepsilon_{\mu,t}) \tag{B-13}$$

$$A_t = A_{t-1} \exp(\Lambda_A + \sigma_A \varepsilon_{A,t}) \tag{B-14}$$

$$\varepsilon_{s,t} \sim \mathcal{N}(0,1), s \in \{d, \varphi, c, m, \mu, A\} \tag{B-15}$$

B.1.5 Phillips Curve

The nonlinear Phillips Curve in this model is characterized recursively by:

$$g_t^1 = \tilde{\lambda}_t \exp(\varepsilon_t) mc_t \tilde{y}_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}} \right)^{-\varepsilon} g_{t+1}^1 \tag{B-16}$$

$$g_t^2 = \tilde{\lambda}_t \Pi_t^* \tilde{y}_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}} \right)^{1-\varepsilon} \left(\frac{\Pi_t^*}{\Pi_{t+1}^*} \right) g_{t+1}^2 \tag{B-17}$$

$$\varepsilon g_t^1 = (\varepsilon - 1) g_t^2 \tag{B-18}$$

B.2

Model Calibration

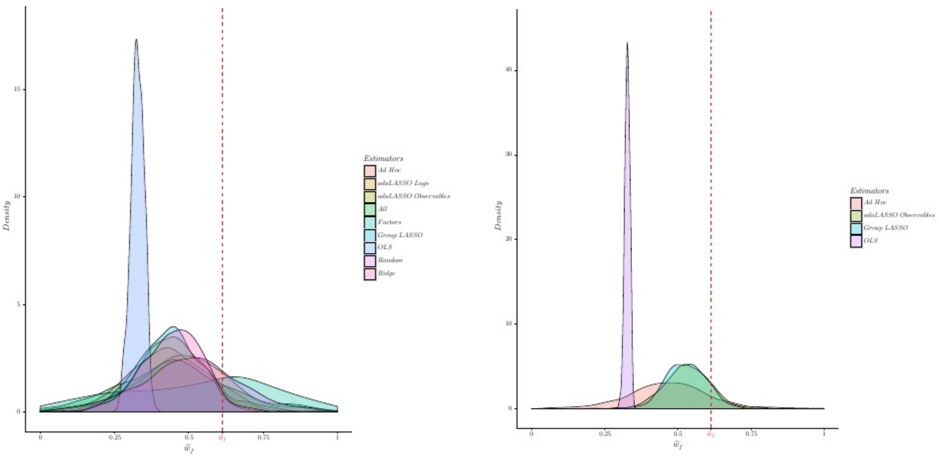
Table B.1: Calibration

Parameter Description	Parameter	Value
Elasticity of Substitution Among Consumption Varieties	ε	10
Elasticity of Substitution Among Different Labor	η	10
Discount Factor	β	0.998
Capital Share in Production Function	α	0.21
Steady State Inflation	Π	1
Fixed Cost of Production Constant	ϕ	0
Parameter 2 for Physical Cost of Use of Capital Function	γ_2	0.001
Consumption Habits	h	0.97
Calvo Price Rigidity	θ_p	0.82
Degree of Price Indexation	χ	0.63
Calvo Wage Rigidity	θ_p	0.68
Degree of Wage Indexation	χ_w	0.62
Taylor Rule Interest Rate Smoothing Weight	γ_R	0.9
Taylor Rule Inflation Weight	γ_Π	1.29
Taylor Rule Product Weight	γ_y	0.19
Inverse of Frisch Labor Supply Elasticity	γ	1.17
Depreciation Rate	δ	0.025
Investment Adjustment Cost Scalar	κ	9.51
Labor Disutility	φ	8.92
Monetary Policy Shock Persistence	ρ_r	0
Preference Shock Persistence	ρ_d	0.12
Labor Supply Shock Persistence	ρ_φ	0.93
Cost-Push Shock Persistence	ρ_c	0
Preference Shock Variance Scalar	σ_d	-1.51
Labor Supply Variance Scalar	σ_φ	-2.36
Cost-Push Shock Variance Scalar	σ_c	-1.99
Investment Specific Growth Rate Variance Scalar	σ_μ	-5.43
Monetary Policy Shock Variance Scalar	σ_m	-5.85
Steady State Investment-Specific Technology Growth Rate	Λ_μ	3.4E-3
Steady State Neutral Technology Growth Rate	Λ_A	2.8E-3

Note: Other unlisted parameters are function of the deep parameters above, which follows from the model solution. For example γ_1 is in equilibrium equal to the normalized (growth-wise) rental rate of capital $\tilde{r}_t = r_t \mu_t$.

C
Appendix for Chapter 6

C.1
Estimators Histograms



C.1(a): $T = 150$. C.1(b): $T = 1000$.

Figure C.1: Estimators Histograms.

C.2
Cost-Push Shock Variance and $RMSE$

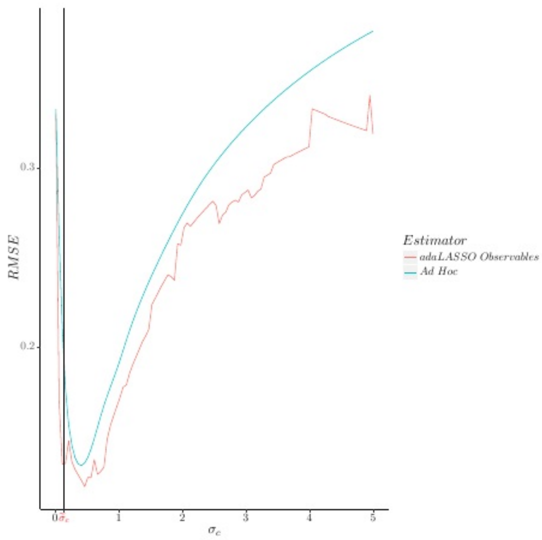


Figure C.2: σ_c and $RMSE$. $T = 150$.

C.3 Estimation of Other Parameters

Table C.1: w_b Estimators Descriptive Statistics, $T = 150$.

Estimator	*Mean Bias*	*Median Bias*	*MAE*	*RMSE*
adaLASSO Observables	0.0701	0.0725	0.0811	0.0964
adaLASSO Lags	0.057	0.0635	0.0864	0.1191
Group LASSO	0.0869	0.0877	0.0903	0.1019
Ad Hoc	0.0698	0.07	0.0788	0.0949
All	0.0055	0.0847	0.6281	6.9266
Random	0.0506	0.0536	0.0817	0.1126
Ridge	0.136	0.0833	0.14	1.1061
Factors	0.3595	0.0898	2.2583	29.3155
OLS	0.1607	0.1605	0.1607	0.1615

Note: True value $w_b = 0.3868$.

Table C.2: κ Estimators Descriptive Statistics, $T = 150$.

Estimator	*Mean Bias*	*Median Bias*	*MAE*	*RMSE*
adaLASSO Observables	0.0295	0.0308	0.0367	0.0464
adaLASSO Lags	0.0194	0.0213	0.0497	0.1062
Group LASSO	0.0293	0.0295	0.0326	0.0392
Ad Hoc	0.0363	0.0365	0.0418	0.0504
All	0.3567	0.0281	0.415	6.7653
Random	0.0203	0.0171	0.0426	0.0596
Ridge	0.048	0.0258	0.0512	0.4173
Factors	0.1052	-0.0121	2.474	32.1602
OLS	0.0343	0.0341	0.0343	0.0346

Note: True value $\kappa = 0.0245$.

Table C.3: w_b Estimators Descriptive Statistics, $T = 1000$.

Estimator	*Mean Bias*	*Median Bias*	*MAE*	*RMSE*
adaLASSO Observables	0.0269	0.0279	0.0339	0.0412
Group LASSO	0.0307	0.0313	0.0359	0.0478
Ad Hoc	0.0433	0.045	0.0521	0.0632
OLS	0.1601	0.16	0.1601	0.1603

Note: True value $w_b = 0.3868$.

Table C.4: κ Estimators Descriptive Statistics, $T = 1000$.

Estimator	*Mean Bias*	*Median Bias*	*MAE*	*RMSE*
adaLASSO Observables	0.018	0.0177	0.022	0.0273
Group LASSO	0.019	0.0187	0.0218	0.0293
Ad Hoc	0.0351	0.0359	0.041	0.0495
OLS	0.034	0.034	0.034	0.034

Note: True value $\kappa = 0.0245$.

D
Appendix for Chapter 8

D.1
Data

The two samples used are not exactly the same due to data availability. Also, for the updated sample case, I only use variables that allow for a mapping with the model variables used in the simulations. This is to allow for a better interpretation of the results vis-à-vis the simulations.

Where necessary, data is aggregated to quarterly frequency by taking the period mean, before filtering. All data is used as percentage deviation of its sample mean (this is analogous to the log-linearization done in the model).

D.1.1
Gali and Gertler (1999) Analogous Sample

Table D.1: Observables for Gali and Gertler (1999) Analogous Sample

Data Code	Data Description	Data Source
GDPC96	Real Gross Domestic Product	US. Bureau of Economic Analysis
PCECC96	Real Personal Consumption Expenditures	US. Bureau of Economic Analysis
RINV	Real Investment	DiCecio, Riccardo
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	US. Bureau of Labor Statistics
GDPDEF	Gross Domestic Product: Implicit Price Deflator	US. Bureau of Economic Analysis
FF	Effective Federal Funds Rate	Board of Governors of the Federal Reserve System (US)
ULCNFB	Nonfarm Business Sector: Unit Labor Cost	US. Bureau of Labor Statistics
GCEC96	Real Government Consumption Expenditures and Gross Investment	US. Bureau of Economic Analysis
GS10	10-Year Treasury Constant Maturity Rate	Board of Governors of the Federal Reserve System (US)
A576RC1	Compensation of Employees, Received: Wage and Salary Disbursements	US. Bureau of Economic Analysis
PPIIDC	Producer Price Index by Commodity Industrial Commodities	US. Bureau of Labor Statistics

Table D.2: Data Treatment

Variable	Data Mapping	Original Data Frequency	Data Transformation	HP Filter?
$y^d_{t,obs}$	GDPC96	Quarterly	Level	Yes
$c_{t,obs}$	PCECC96	Quarterly	Level	Yes
$x_{t,obs}$	RINV	Quarterly	Level	Yes
$l_{t,obs}$	AHWMAN	Monthly	Level	No
$\pi_{t,obs}$	GDPDEF	Monthly	% Change	No
$r_{t,obs}$	FF	Weekly	Level	No
$mc_{t,obs}$	ULCNFB	Quarterly	Level	Yes
$G_{t,obs}$	GCEC96	Quarterly	% Change	No
$Spread_{t,obs}$	GS10 - FF	Monthly - Weekly	Level - Level	No
$w_{t,obs}$	A576RC1	Monthly	Level	No
$comm_{t,obs}$	PPIIDC	Monthly	Level	Yes

D.1.2
Updated Sample

For clarity, I provide a mapping of the model variables and data collected, in the spirit of measurement equations. This is not necessary since I undertake a limited information estimation, which has as its main advantage exactly to obviate such a mapping with a particular structural model. All data was retrieved from FRED, Federal Reserve Bank of St. Louis

Table D.3: Observables for Updated Sample

Observed Variable	Data Description	Data Source	Data Code
$y^d_{t,obs}$	Real Gross Domestic Product	US. Bureau of Economic Analysis	GDPC96
$c_{t,obs}$	Real Personal Consumption Expenditures	US. Bureau of Economic Analysis	PCECC96
$x_{t,obs}$	Real Investment	DiCecio, Riccardo	RINV
$u_{t,obs}$	Capacity Utilization: Total Industry	Board of Governors of the Federal Reserve System (US)	TCU
$l_{t,obs}$	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	US. Bureau of Labor Statistics	AWHMAN
$\pi_{t,obs}$	Gross Domestic Product: Implicit Price Deflator	US. Bureau of Economic Analysis	GDPDEF
$r_{t,obs}$	Effective Federal Funds Rate	Board of Governors of the Federal Reserve System (US)	FF
$mc_{t,obs}$	Nonfarm Business Sector: Unit Labor Cost	US. Bureau of Labor Statistics	ULCNFB
$w_{t,obs}$	Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private	US. Bureau of Labor Statistics	AHETPI
$G_{t,obs}$	Real Government Consumption Expenditures and Gross Investment	US. Bureau of Economic Analysis	GCEC96

Variables in the model are normalized to stationarity using the variable z_t. To obtain stationary variables from the observables I consider here, I remove trends by using the Hodrick & Prescott (1997) (HP) filter using a smoothing parameter of 1600 for quarterly data. The table below summarizes data treatment:

Table D.4: Data Treatment

Model Variable	Original Data Frequency	Data Transformation	HP Filter?	"Measurement Equation"
$y^d_{t,obs}$	Quarterly	Level	Yes	$y^d_{t,obs} = y^d_t$
$c_{t,obs}$	Quarterly	Level	Yes	$c_{t,obs} = c_t$
$x_{t,obs}$	Quarterly	Level	Yes	$x_{t,obs} = x_t$
$u_{t,obs}$	Monthly	Level	No	$u_{t,obs} = u_t$
$l_{t,obs}$	Monthly	Level	No	$l_{t,obs} = l_t$
$\pi_{t,obs}$	Monthly	% Change	No	$1 + \frac{\pi_{t,obs}}{100} = \Pi_t$
$r_{t,obs}$	Weekly	Level	No	$1 + \frac{r_{t,obs}}{100} = R_t$
$mc_{t,obs}$	Quarterly	Level	Yes	$mc_{t,obs} = mc_t$
$w_{t,obs}$	Monthly	Level	Yes	$w_{t,obs} = w_t$
$G_{t,obs}$	Quarterly	Level	Yes	$G_{t,obs} = G_t$

D.2
Estimates for Other Parameters

.

D.2.1

Gali and Gertler (1999) Analogous Sample

Table D.5: Empirical Estimation of w_b for the US Economy

Estimator	*Estimated Coefficient*	*Standard Error*
adaLASSO Observables	0.17	0.11
adaLASSO Lags	0.33	0.05
Group LASSO	0.40	0.15
Ad Hoc	0.17	0.11
All	0.42	0.04
Random	0.53	0.16
Ridge	0.56	0.11
Factors	0.51	1.32
OLS	0.47	0.06

Note: Optimal weighting matrix used. Covariance matrix for *GMM* calculated using *HAC* estimator matrix with *Bartlett* kernel and bandwidth set according to Andrews (1991). *OLS* standard error obtained assuming homoskedasticity.

Table D.6: Empirical Estimation of κ for the US Economy

Estimator	$10^3\times$ *Estimated Coefficient*	$10^3\times$ *Standard Error*
adaLASSO Observables	0.56	0.25
adaLASSO Lags	0.41	0.16
Group LASSO	0.19	0.33
Ad Hoc	0.56	0.25
All	0.17	0.11
Random	-0.12	0.35
Ridge	-0.11	0.22
Factors	0.06	2.16
OLS	0.00	0.00

Note: Optimal weighting matrix used. Covariance matrix for *GMM* calculated using *HAC* estimator matrix with *Bartlett* kernel and bandwidth set according to Andrews (1991). *OLS* standard error obtained assuming homoskedasticity.

D.2.2
Updated Sample

Table D.7: Empirical Estimation of w_b for the US Economy

Estimator	*Estimated Coefficient*	*Standard Error*
adaLASSO Observables	0.22	0.06
adaLASSO Lags	0.27	0.07
Group LASSO	0.31	0.11
Ad Hoc	0.17	0.08
All	0.19	0.08
Random	0.36	0.28
Ridge	0.02	0.13
Factors	-0.10	0.43
OLS	0.43	0.07

Note: Optimal weighting matrix used. Covariance matrix for *GMM* calculated using *HAC* estimator matrix with *Bartlett* kernel and bandwidth set according to Andrews (1991). *OLS* standard error obtained assuming homoskedasticity.

Table D.8: Empirical Estimation of κ for the US Economy

Estimator	$10^3\times$ *Estimated Coefficient*	$10^3\times$ *Standard Error*
adaLASSO Observables	-0.22	0.07
adaLASSO Lags	-0.10	0.10
Group LASSO	-0.22	0.13
Ad Hoc	-0.26	0.35
All	-0.24	0.13
Random	-0.17	0.21
Ridge	-0.76	0.46
Factors	-3.47	5.33
OLS	-0.00	0.00

Note: Optimal weighting matrix used. Covariance matrix for *GMM* calculated using *HAC* estimator matrix with *Bartlett* kernel and bandwidth set according to Andrews (1991). *OLS* standard error obtained assuming homoskedasticity.